Math Bootcamp · MSSP Students Aleks Goeva

Probability – the formal language of uncertainty which is th[e] basis of statistical inference

Introduction

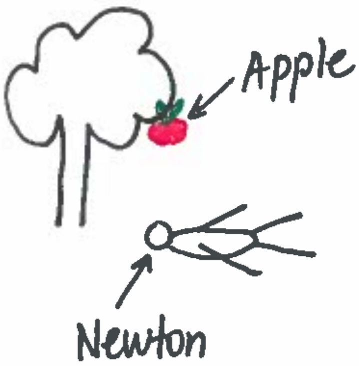
Data generating process ← inference — Observed data

Data generating process → probability — Observed data

probability - given a data generating process what are the properties of the outc[ome]

- **Deterministic experiment** - experiment performed under some conditions that lead to same outcome

Ex.



$$\underline{F = m \cdot a}$$

Newton's 2^{nd} Law of motion

- If we know that the apple will fall, the force is <u>deterministic</u>,

but whether the apple will fall or not at a given moment is not certain.

- **Random experiment**

experiment = { throw a die } ← fixed outcome? probabilities

get $1, 2, 3, \ldots, 6$

Framework: "Do not reinvent the wheel"

↳ use the framework of set theory.

Steps for describing a random experiment:

1) Specify experiment $\mathcal{E}$ → Ex. $\mathcal{E}$ = throw a die

2) Specify the sample space Ω = list of all possible (elementary) outcomes → $\Omega = \{1, 2, 3, 4, 5, 6\}$

3) Focus on subsets of Ω → Example: $A = \{$get even #
$= \{2 \text{ or } 4 \text{ or } 6\}$
lists the elementary events we are interested in

subset = event = outcome

4) Assign prob. to each event → Equally likely:
$\mathbb{P}(1) = \dots = \mathbb{P}(6) = \frac{1}{6}$

↳ assign it to elementary events

↳ derive it for more complicated events

Then: $\mathbb{P}(\text{even } \#)$
$= \mathbb{P}(\{2, 4, 6\})$
$= \frac{3}{6} = \frac{1}{2}$

Venn diagram

Ω

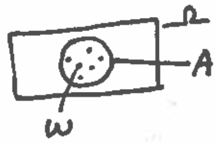
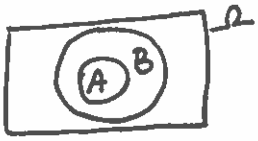
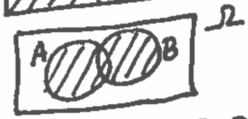
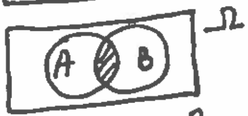
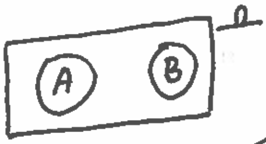
1	2	3
4	5	6

Once the experiment is performed, only one elementary outcome will result.

Notation: Elementary outcome : ω

Ex. $\omega = \{6\}$ = get a 6.

Dictionary

Language of events (English Language)	Language of sets
event A occurs	A = subset of Ω 
the sure event (the outcome is certain)	Ω
the outcome cannot happen (the impossible event)	$\emptyset$ (empty set) (null set) Ex. $A = \{$ get a 3 and a 4 $\} = \emptyset$
A implies B (if A occurs then B occurs)	$A \subset B$ 
A does not occur	A^c 
either A or B occurs	$A \cup B$ 
both A and B occur	$A \cap B$ 
at least one of $A_1, \dots, A_n$ occurs	$A_1 \cup \dots \cup A_n = \bigcup_{i=1}^{n} A_i$ finite union
all of $A_1, \dots, A_n$ occurs	$A_1 \cap \dots \cap A_n = \bigcap_{i=1}^{n} A_i$
A and B are mutually exclusive (disjoint)	$A \cap B = \emptyset$ 

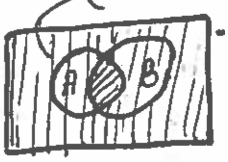
Useful Relations:

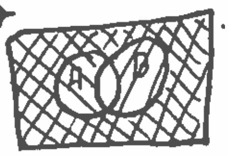
- De Morgan Laws:

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

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

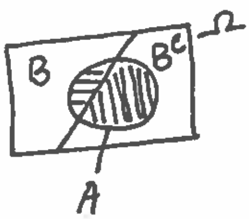
$(A\cap B)^c$





$(A \cup B)^c$

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



Distributive Laws:

$$A \cap \left(\bigcup_{j=1}^{n} B_j \right) = \bigcup_{j=1}^{n} (A \cap B_j)$$

$$A \cup \left(\bigcap_{j=1}^{n} B_j \right) = \bigcap_{j=1}^{n} (A \cup B_j)$$

2.6 An urn contains 10 balls, numbered 0, 1, 2, ..., 9. Three balls are removed, one at a time, without replacement.
a) Obtain the sample space for this random experiment.
b) Determine, as a subset of the sample space, the event that an even number of odd-numbered balls are removed from the urn.

2.7 Refer to Example 2.3 on page 27 where two dice are rolled, one black and one gray. For $i = 2, 3, \ldots, 12$, determine explicitly as a subset of the sample space the event A_i that the sum of the faces is i.

2.8 Consider the following random experiment: First a die is rolled and you observe the number of dots facing up; then a coin is tossed the number of times that the die shows and you observe the total number of heads.
a) Determine the sample space for this random experiment.
b) Determine the event that the total number of heads is even.

2.9 George and Laura take turns tossing a coin. The first person to get a tail wins. George goes first. *Note:* You may assume that eventually a tail will be tossed.
a) Describe the sample space for this random experiment.
b) Determine, as a subset of the sample space, the event that Laura wins.

2.10 This exercise considers two random experiments involving the repeated tossing of a coin. *Note:* You may assume that eventually a head will be tossed.
a) If the coin is tossed until the first time a head appears, find the sample space.
b) If the coin is tossed until the second time a head appears, find the sample space.
c) For the experiment in part (a), express the event that the coin is tossed exactly six times in the form $\{\ldots\}$, where in place of "..." you list all of the outcomes in that event.
d) Repeat part (c) for the experiment described in part (b).

• **2.11** From 10 men and 8 women in a pool of potential jurors, 12 are chosen at random to constitute a jury. Suppose that you observe the number of men who are chosen for the jury. Let A be the event that at least half of the 12 jurors are men and let B be the event that at least half of the 8 women are on the jury.
a) Determine the sample space for this random experiment.
b) Find $A \cup B$, $A \cap B$, and $A \cap B^c$, listing all the outcomes for each of those three events.
c) Are A and B mutually exclusive? A and B^c? A^c and B^c? Explain your answers.

• **2.12** Let A and B be events of a sample space.
a) Show that, if A and B^c are mutually exclusive, then B occurs whenever A occurs.
b) Show that, if B occurs whenever A occurs, then A and B^c are mutually exclusive.

• **2.13** Let A, B, and C be events of a sample space. Write a mathematical expression for each of the following events.
a) A occurs, but B doesn't occur.
b) Exactly one of A and B occurs.
c) Exactly one of A, B, and C occurs.
d) At most two of A, B, and C occur.

2.14 Refer to Example 2.17 on page 34, but now suppose that two cards are selected at random, one after the other, without replacement.
a) What is Ω for this random experiment?
b) Let A be the event that at least one of the cards is a face card and let B be the event that at least one of the cards is an ace. Are A and B mutually exclusive? Why or why not?

2.11)

10 m 8 w

12 chosen at random

observe the number of men

a) $\Omega = \{4, 5, \dots, 10\}$

A: at least half are men

$A = \{6, 7, 8, 9, 10\}$

B: at least half of the 8 w are on the jury

$B = \{4, 5, 6, 7, 8\}$

b) $A \cup B = \{4, \dots, 10\} = \Omega$

$A \cap B = \{6, 7, 8\}$

$A \cap B^c = \{9, 10\}$

c) $A \cap B \neq \emptyset \Rightarrow$ not mutually exclusive

$A \cap B^c \neq \emptyset \Rightarrow$ –//–

$A^c \cap B^c = \{4, 5\} \cap \{9, 10\} = \emptyset \Rightarrow$ mutually exclusive

Kolmogorov axioms of probability

We want to attach to each event A a probability.
What rules should that probability satisfy?

measure its "area" $\quad \mathbb{P}(A)$
measure its likelihood

Def $\mathbb{P}$ is a probability measure on the events of Ω if:

nonnegativity 1) $\mathbb{P}(A) \geq 0, \quad \forall A \in \Omega$

certainty 2) $\mathbb{P}(\Omega) = 1$

countable additivity 3) $A_1, A_2, \dots$ mutually exclusive

$$\Rightarrow \mathbb{P}\left(\bigcup_{k=1}^{\infty} A_k \right) = \sum_{k=1}^{\infty} \mathbb{P}(A_k), \quad \forall A_1, A_2, A_3, \dots$$

In particular:

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

follow the rule for area.

Consequences: ① $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$ ← complement rule

Pf. $\Omega = A \cup A^c$ disjoint

$\mathbb{P}(\Omega) = \mathbb{P}(A \cup A^c)$

$1 = \mathbb{P}(A) + \mathbb{P}(A^c)$

② $\mathbb{P}(\emptyset) = 0 \qquad \mathbb{P}(A) = \mathbb{P}(A \cup \emptyset) \stackrel{3)}{=} \mathbb{P}(A) + \mathbb{P}(\emptyset)$

③ if $A \subset B \Rightarrow \mathbb{P}(A) \leq \mathbb{P}(B)$ ← domination rule

if A occurs then B occurs

B is at least as probable as A

by picture Ⓐ B

④ $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$
← addition rule

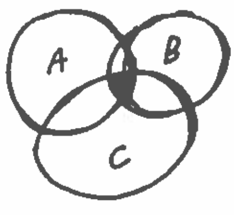
Pf. $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(A^c \cap$

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

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

↳ Extension:

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



Exercises:

2.27 p.47

•2.27 An urn contains four balls numbered 1, 2, 3, and 4. A ball is chosen at random, its number noted, and the ball is replaced in the urn. This process is repeated one more time.

a) Determine the sample space Ω.

b) If each outcome is assigned the same probability, what is that common probability?

c) Using the probability assignment in part (b), find the probability that the two numbers chosen are different.

a) $\Omega = \{(x, y) : x, y \in \{1, 2, 3, 4\}\}$ ← 16 outcomes

b) $\boxed{p = \frac{1}{16}}$

common probability

c) $\mathbb{P}$(the two chosen numbers are different) =

= 1 − $\mathbb{P}$(the two chosen numbers are the same) =

$= 1 - \mathbb{P}((1,1), \text{or } (2,2) \text{ or } (3,3) \text{ or } (4,4)) =$

$= 1 - (p + p + p + p) = 1 - 4p = 1 - 4 \cdot \frac{1}{16} = \frac{12}{16} = \boxed{\frac{3}{4}}$

from pre-assessment test

2.66 p.75 Let A and B be events such that $\mathbb{P}(A) = \frac{1}{4}$, $\mathbb{P}(B) = \frac{1}{3}$, and $\mathbb{P}(A \cup B) = \frac{1}{2}$.

a) Are events A and B mutually exclusive? Explain your answ

$\mathbb{P}(A) + \mathbb{P}(B) = \frac{1}{4} + \frac{1}{3} = \frac{7}{12} \neq \frac{1}{2} = \mathbb{P}(A \cup B)$

⇒ not mutually exclusive

b) Determine $\mathbb{P}(A \cap B)$

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

$\frac{1}{2} = \frac{1}{4} + \frac{1}{3} - \mathbb{P}(A \cap B) \Rightarrow \mathbb{P}(A \cap B) = \frac{7}{12} - \frac{1}{2} = \frac{1}{1}$

Additionally, find • $\mathbb{P}(A \cap B^c)$: $\mathbb{P}(A) = \mathbb{P}(B \cap A) + \mathbb{P}(B^c \cap A)$

$\frac{1}{4} = \frac{1}{12} + \mathbb{P}(A \cap B^c) \Rightarrow \mathbb{P}(A \cap B^c) =$

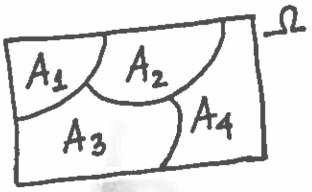
• and $\mathbb{P}(A^c \cap B^c) \underset{\text{De Morgan}}{=} \mathbb{P}((A \cup B)^c) = 1 - \mathbb{P}(A \cup B) =$

$= 1 - \frac{1}{2} = \boxed{\frac{1}{2}}$.

Law of partition

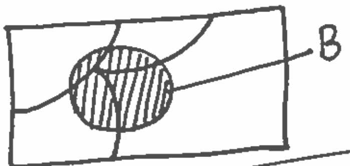
Def Events $A_1, A_2, \dots$ from a partition of Ω if they are mutually disjoint and if their union is all of Ω.

Ex.



Proposition: Let $A_1, A_2, \dots$ be a partition of Ω.

Then $\quad \mathbb{P}(B) = \sum_n \mathbb{P}(A_n \cap B)$



Pf. $B = \Omega \cap B = \left(\bigcup_n A_n\right) \cap B$

$= \bigcup_n (A_n \cap B)$

$\Rightarrow \mathbb{P}(B) = \sum_n \mathbb{P}(A_n \cap B)$

axiom 3, since disjoint

$B = B \cap \underbrace{\Omega}_{\bigcup_n A_n} = B \cap \left(\bigcup_n A_n\right)$

Distributive Law

$\overset{\text{Distributive Law}}{=} \bigcup_n \underbrace{(B \cap A_n)}_{\text{disjoint}}$

Examples

2.28 p.68

The US National Center for Education Statistics compiles information on institutions of higher education and publishes its findings in the Digest of Education Statis
According to that document, 8.1% of institutions of higher education are public schools in the Northeast,
11.0% are public schools in the Midwest,
16.3% are public schools in the South, and
9.6% are public schools in the West.
If a US institution of higher education is selected at random, determine the probability that it is public.

Solution:

First, $\Omega = ?$

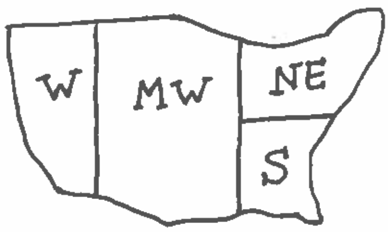
$\Omega = \Omega_1 \times \Omega_2$

Ω_1: list of colleges in the country

Ω_2: {public, private}

B = event that the colleg selected is public

$\mathbb{P}(B) = ?$



partition in 4:

A_1 = event college selected in NE

A_2 = — " — MW

A_3 = — " — S

A_4 = — " — W

$A_1, \dots, A_4$ are a partition of Ω

$$\Rightarrow \mathbb{P}(B) = \sum_{i=1}^{4} \mathbb{P}(B \cap A_i) = \mathbb{P}(B\cap A_1) + \mathbb{P}(B \cap A_2) + \mathbb{P}(B \cap A_3) + \mathbb{P}(B$$

$$= 0.081 + 0.110 + 0.163 + 0.09$$

$$= \boxed{0.45}$$

$\Rightarrow$ 45% of institutions of higher education in the United States are public.

Counting Methods

- Multiplication Rule (Basic Counting Rule)

  r actions must be performed

  action 1 can be done in m_1 ways,

  $\vdots$

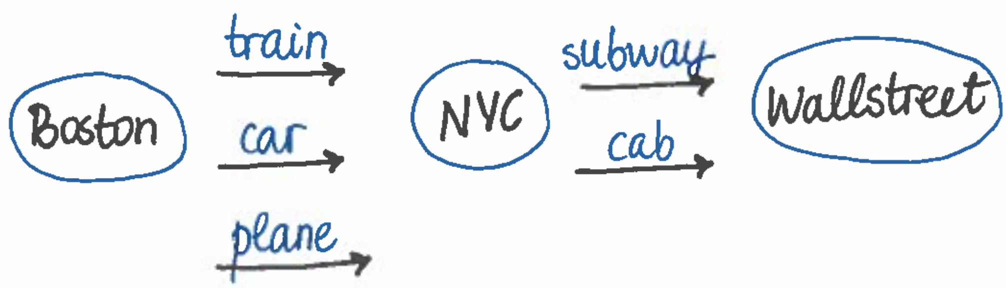
  action r can be done in m_r ways.

  The r actions can be done in

  $m_1 \cdot \ldots \cdot m_r$ ways.

  $$\underbrace{m_1}_{1}\ \underbrace{m_2}_{2}\ \cdots\ \underbrace{m_r}_{r}$$

Ex.



In how many different ways can we go from Boston, to NYC, and then to Wallstreet?

2 actions : 1^{st} action : Boston → NYC, can be done in 3 w

2^{nd} action : NYC → Wallstreet, can be done in 2 w

$$\underbrace{3}_{1}\ \underbrace{2}_{2} \Rightarrow 6 \text{ ways.}$$

n distinct objects

- **Permutation**

  # of ways we can permute (rearrange) them

$$\underset{①}{\underline{n}}\ \underset{②}{\underline{n-1}}\ \underset{③}{\underline{n-2}}\ \cdots\ \underset{ⓝ}{\underline{1}}$$

$n \cdot (n-1) \cdot \ldots \cdot 1 = n!$ factorial

$n = 3$

$n! = 3 \cdot 2 \cdot 1 = 6$

and $0! = 1$ (convention)

- **Arrangement** $\underset{①}{\underline{n}}\ \underset{②}{\underline{n-1}}\ \cdots\ \underset{ⓚ}{\underline{n-(k-1)}}$

  Place the n distinct objects into $k \leq n$ places (order counts) & (no replacement

$${}_nP_k = n \cdot (n-1) \cdot \ldots \cdot (n-(k-1)) =$$

$$= n \cdot (n-1) \cdot \ldots \cdot (n-k+1) =$$

$$= \frac{n!}{(n-k)!} \qquad \underset{①}{\underline{n}}\ \underset{②}{\underline{n-1}}\ \cdots\ \underset{ⓚ}{\underline{n-(k-1)}}$$

if $k = n \Rightarrow {}_nP_n = n!$

Ex. $n = 3$ a, b, c

$k = 2$ ab, ac, bc,

ba, ca, cb.

$${}_3P_2 = \frac{3!}{(3-2)!} = \frac{3!}{1!} = \frac{3 \cdot 2 \cdot 1}{1} = 6$$

- <u>Combination</u>

Place n distinct objects in $k \leq n$ positions, but order does not count (& no replacement)

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

since order does not count, with ${}_nP_k$ we are countin too many things, which would count as 1 whi order does not count $\Rightarrow$ divide by the number of ways in which we can permute the k chosen objects.

<u>Ex.</u> $n = 3$ a, b, c

$k = 2$ ab, bc, ac

$$ {}_3C_2 = \binom{3}{2} = \frac{3!}{(3-2)!\,2!} = \frac{3.2.1}{1.2.1} = 3 $$

<u>Notes:</u> $\binom{n}{k} = \binom{n}{n-k}$

- n choose k
- binomial coefficient

$$ \binom{n}{0} = \binom{n}{n} = 1 $$

Ex. Poker $\quad$ Example 3.14 $\quad$ p.101

Draw 5 cards of 52.

a) # of possibilities $= \binom{52}{5} = 2,598,960$

b). # hands with 3 Kings and 2 Queens

$$\underbrace{\binom{4}{3}}_{\text{3 Kings out of 4}} \cdot \underbrace{\binom{4}{2}}_{\text{2 Queens out of 4}} = 24$$

c) full house: 3 cards with same face value
2 cards with same face value

# of full houses

$$\underbrace{\binom{13}{1}}_{\text{1 out of 13 different face values}} \boxed{\underbrace{\binom{12}{1}}_{\text{1 out of 12 diff. face values left}}} \underbrace{\binom{4}{3}}_{\text{3 out of 4 cards with the first face value}} \boxed{\underbrace{\binom{4}{2}}_{\text{2 out of 4 cards with the second face value}}} = 13 \cdot 12 \cdot 24 = 3774$$

d) $\mathbb{P}(\text{full house}) = \dfrac{3774}{2,598,960} \approx 0.00145212$

form the ratio of $\dfrac{\text{\# of hands with full house}}{\text{\# of all possible hands}}$

to get $\mathbb{P}$(full house).

- n objects not all distinct

EX. # of words we can write with the letters of the word MISSISSIPPI (11 letters)

- if they were distinct $11!$
- but they are not all distinct, so

$$\frac{11!}{\underset{M}{1!}\ \underset{I}{4!}\ \underset{S}{4!}\ \underset{P}{2!}}$$

In general, n objects, r of them are distinct,

$$\frac{n!}{n_1!\, n_2! \, \ldots \, n_r!} = \binom{n}{n_1 \ldots n_r}$$

n_1 # 1st obj. is repeated is

⋮

r^{th} obj. is repeated is n_r

$(n_1 + n_2 + \cdots + n_r = n)$

Let $r = 2$

$$\binom{n}{n_1 \ n_2} = \frac{n!}{n_1!\, n_2!} = \frac{n!}{n_1!(n-n_1)!} = \binom{n}{n_1} = \binom{n}{n_2}$$

(do it) Binomial theory

$$(a+b)^2 = a^2 + 2ab + b^2$$

$$= \underset{\text{choose 0 } b\text{'s}}{\binom{2}{0}a^2} + \underset{\text{choose 1 } b}{\binom{2}{1}ab} + \underset{\text{choose 2 } b\text{'s}}{\binom{2}{2}b^2}$$

$$(a+b)^m = (a+b)(a+b)\ldots(a+b)$$

$$= \sum_{i=0}^{m} \binom{m}{i} a^i \cdot b^{m-i}$$

EXERCISES 3.3 Basic Exercises

3.56 Provide the details for the solution of Example 3.18(b) on page 111.

• **3.57** Four cards are dealt from an ordinary deck of 52 playing cards. What is the probability that the denominations (face values) of the cards are
a) all the same? b) all different?

$1 - \frac{\binom{8}{2}}{\binom{10}{2}} \approx 0.378$

• **3.58** In a small lottery, 10 tickets—numbered 1, 2, ..., 10—are sold. Two numbers are drawn at random for prizes. You hold tickets numbered 1 and 2. What is the probability that you win at least one prize?

• **3.59** An ordinary deck of 52 playing cards is shuffled and dealt. What is the probability that
a) the seventh card dealt is an ace? b) the first ace occurs on the seventh card dealt?

3.60 From an urn containing M red balls and $N - M$ black balls, a random sample of size n is taken without replacement. Find the probability that exactly j black balls are in the sample.

X **3.61** **The birthday problem:** A probability class has 38 students.
a) Find the probability that at least 2 students in the class have the same birthday. For simplicity, assume that there are always 365 days in a year and that birth rates are constant throughout the year. *Hint:* Use the complementation rule.
b) Repeat part (a) if the class has N students.
c) Evaluate the probability obtained in part (b) for $N = 1, 2, \ldots, 70$. Use a computer or calculator to do the number crunching.
d) What is the smallest class size for which the probability that at least 2 students in the class have the same birthday exceeds 0.5?

• **3.62** An urn contains four red balls and six black balls. Balls are drawn one at a time at random until three red balls have been drawn. Determine the probability that a total of seven balls is drawn if the sampling is
a) without replacement. b) with replacement.

3.63 Four mathematicians, three chemists, and five physicists are seated randomly in a row. Find the probability that all the members of each discipline sit together.

3.64 Suppose that a random sample of size n without replacement is taken from a population of size N. For $k = 1, 2, \ldots, n$, determine the probability that k specified members of the population will be included in the sample.

3.65 Suppose that a random sample of size n with replacement is taken from a population of size N.
a) Determine the probability that no member of the population is selected more than once.
b) Show that the probability in part (a) approaches 1 as $N \to \infty$. Interpret this result.

• **3.66** Refer to Example 3.19 on page 111. Determine the probability that the number of defective TVs selected is
a) exactly one. b) at most one. c) at least one.

3.67 Refer to Example 3.20(b) on page 113. In this exercise, you are to obtain $P(E)$—the probability that a specified member of the population will be included in the sample—in two additional ways.
a) Compute $N(E)$ directly and then apply Equation (3.4) on page 110 to determine $P(E)$.
b) For $k = 1, 2, \ldots, n$, let A_k denote the event that the kth member selected is the specified member. Without doing any computations, explain why $P(A_k) = 1/N$, for $k = 1, 2, \ldots, n$. Conclude that $P(E) = n/N$. Explain your reasoning.

Ex. 3.57 p.116 order doesn't matter

4 cards are dealt from an ordinary deck of 52.

a) $\mathbb{P}$(face values of the cards are all the same) = ?

$$|\Omega| = \binom{52}{4} = 270,725$$

$$|A| = \underbrace{\binom{13}{1}}_{\text{pick which face value}} \cdot \binom{4}{4} = 13$$

$$\Rightarrow \mathbb{P}(A) = \frac{13}{270,725} \approx 0.000048$$

b) B = face values are all different

$\underline{13}\ \underline{12}\ \underline{11}\ \underline{10}$ $\quad \binom{13}{4} = \binom{13}{9}$

$$|B| = \underbrace{\binom{13}{4}}_{\text{pick four different face values}} \cdot \underbrace{4^4}_{\text{four possibilities for the suit}}$$

♡♤◇♧

$$\mathbb{P}(B) = \frac{183,040}{270,725} \approx 0.676.$$

compare to the problem with the full house

Q: Why here we do not do $\underbrace{\binom{13}{1}\binom{12}{1}\binom{11}{1}\binom{10}{1}}_{\frac{13!}{9!}}$ instead of $\underbrace{\binom{13}{4}}_{\frac{13!}{9!\,4!}}$

$\frac{13!}{9!}$: order matters if calculated this way

$\frac{13!}{9!\,4!}$: in our problem order does not matter

(order matters in full house ex. since one face value is repeated 3 times, the other 2 times).

3.31 An economics professor is using a new method to teach a junior-level course with an enrollment of 42 students. The professor wants to conduct in-depth interviews with the students to get feedback on the new teaching method but doesn't want to interview all 42 of them. She decides to interview a sample of 5 students from the class. How many different samples are possible?

3.32 The Powerball® is a multistate lottery that was introduced in April 1992 and is now sold in 24 states, the District of Columbia, and the Virgin Islands. To play the game, a player first selects five numbers from the numbers 1–53 and then chooses a Powerball number, which can be any number between 1 and 42, inclusive. How many possibilities are there?

3.33 In the game of *keno*, there are 80 balls, numbered 1–80. From these 80 balls, 10 are selected at random.

a) How many different outcomes are possible?

b) If a player specifies 20 numbers, in how many ways can he get all 10 numbers selected?

3.34 A club has 14 members.

a) How many ways can a governing committee of size 3 be chosen?

b) How many ways can a president, vice president, and treasurer be chosen?

c) How many ways can a president, vice president, and treasurer be chosen if two specified club members refuse to serve together?

3.35 How many license plates are there consisting of three digits and ~~three~~ two letters if there is no restriction on where the digits and letters are placed? $\binom{5}{2}10^3 26^2$

3.36 A five-card draw poker hand consists of 5 cards dealt from an ordinary deck of 52 playing cards. The order in which the cards are received is unimportant. Note that, in sequence, an ace can play as either the lowest or highest card. In other words, the hierarchy of card denominations, from lowest to highest, is ace, 2, 3, . . . , 10, jack, queen, king, ace. Determine the number of possible hands of the specified type.

a) Straight flush: five cards of the same suit in sequence

b) Four of a kind: $\{w, w, w, w, x\}$, where w and x are distinct denominations

c) Full house: $\{w, w, w, x, x\}$, where w and x are distinct denominations

d) Flush: five cards of the same suit, not all in sequence

e) Straight: five cards in sequence, not all of the same suit

f) Three of a kind: $\{w, w, w, x, y\}$, where w, x, and y are distinct denominations

g) Two pair: $\{w, w, x, x, y\}$, where w, x, and y are distinct denominations

h) One pair: $\{w, w, x, y, z\}$, where w, x, y, and z are distinct denominations

3.37 Repeat Exercise 3.36 for the game of five-card stud, where the order in which the cards are received matters.

3.38 Refer to the inclusion–exclusion principle, Proposition 2.10 on page 73. Determine the number of summands in each sum.

3.39 The U.S. Senate consists of 100 senators, 2 from each state. A committee consisting of 5 senators is to be formed.

a) How many different committees are possible?

b) How many are possible if no state can have more than 1 senator on the committee?

3.40 Refer to Example 3.17 on page 105. How many possible results are there in which the United States has exactly two finishers in the top three and one in the bottom three?

3.41 Without doing any calculations, explain why

a) $\binom{n}{k} = \binom{n}{n-k}$. **b)** $\binom{n}{k} = \binom{n}{k,\, n-k}$.

Ex. 3.59 p. 116

order matters

a) $\mathbb{P}$(the seventh card dealt is an ace) = ?

$$= \frac{4 \cdot 51 \cdot 50 \cdot 49 \cdot 48 \cdot 47 \cdot 46}{52 \cdot 51 \cdot \ldots \cdot 46} = \frac{4}{52} = \frac{1}{13} \approx 0.0769$$

four possibilities for the seventh card

$A\heartsuit, A\spadesuit, A\diamondsuit, A\clubsuit$

b) $\mathbb{P}$(the first ace occurs on the seventh card dealt)

$$= \frac{4 \cdot \underset{48}{(52-4)} \cdot 47 \cdot 46 \cdot 45 \cdot 44 \cdot 43}{52 \ldots 46} = \frac{45 \cdot 44 \cdot 43}{13 \cdot 51 \cdot 50 \cdot 49}$$

$\approx 0.0524.$

Conditional Probability

Ex. A box contains d defective items
and g good ones.

Pick 2 items (pick one, then pick a second one) → order matters

Let $D_1 = \{ 1^{st}$ item is defective $\}$

$D_2 = \{ 2^{nd}$ item is defective $\}$

Q1 : $\mathbb{P}(D_1) = ?$

$$\mathbb{P}(D_1) = \frac{d}{d+g}$$

Q2 : $\mathbb{P}(D_2) = ?$ Answer depends on how the sampling is done :

- sampling with replacement $\Rightarrow \mathbb{P}(D_2) = \dfrac{d}{d+g}$.

- sampling without replacement :

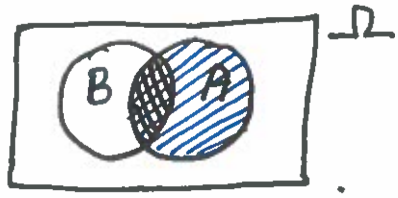
  consider $\mathbb{P}(D_2 \mid D_1) =$ prob. that D_2 occurs given that D_1 occured

  and $\mathbb{P}(D_2 \mid D_1^c) =$ prob. that D_2 occurs given that D_1 did not occur.

More generally,

Def Let B and A be two events,

then $$\underbrace{\mathbb{P}(B \mid A)}_{\text{conditional probability of } B \text{ given } A} = \frac{\mathbb{P}(B \cap A)}{\mathbb{P}(A)}$$



compare the area of the intersection to the area of A.

Useful formulas

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

2. $$\mathbb{P}(A\cap B) = \mathbb{P}(B|A).\mathbb{P}(A)$$ multiplication rule

$$\mathbb{P}(\text{both items are defective}) = \mathbb{P}(D_1 \cap D_2) = \mathbb{P}(D_2 | D_1).\mathbb{P}(D_1)$$
$$= \left(\frac{d-1}{d+g-1}\right)\cdot\left(\frac{d}{d+g}\right)$$

Extension

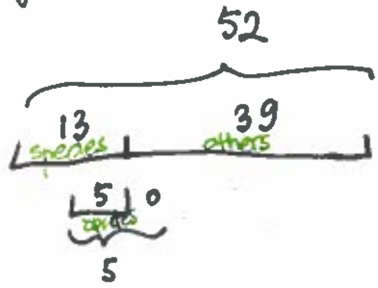
$$\mathbb{P}(A\cap B\cap C) = \mathbb{P}(A | B\cap C).\mathbb{P}(B\cap C) =$$
$$= \mathbb{P}(A|B\cap C).\mathbb{P}(B|C).\mathbb{P}(C)$$

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

$$\Rightarrow \mathbb{P}(A\cap B) = \mathbb{P}(A|B).\mathbb{P}(B)$$ multiplication ru

Ex. $\mathbb{P}(5 \text{ spade hand})$

5 cards are dealt from a regular 52-card deck



Method 1: Combinatorics

$$\frac{\binom{13}{5} \cdot \binom{39}{0}}{\binom{52}{5}}$$

Method 2: Conditional probability

$$\mathbb{P}(5 \text{ spades}) = \mathbb{P}(A_1 \cap A_2 \cap A_3 \cap A_4 \cap A_5) =$$

$$= \mathbb{P}(A_5 \mid A_1 \cap A_2 \cap A_3 \cap A_4) \cdot \mathbb{P}(A_4 \mid A_1 \cap A_2 \cap A_3) \cdot \mathbb{P}(A_3 \mid A_1 \cap A_2) \cdot$$

$$\cdot \mathbb{P}(A_2 \mid A_1) \cdot \mathbb{P}(A_1) = \frac{9}{48} \cdot \frac{10}{49} \cdot \frac{11}{50} \cdot \frac{12}{51} \cdot \frac{13}{52}$$

Independence

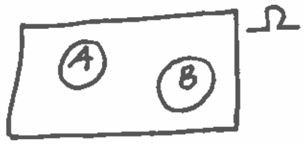
Def A and B are independent if knowing that A occured does not change the probability of B. $\mathbb{P}(B|A) = \mathbb{P}(B)$

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

Def A and B are independent if $\mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B)$

Note:

- A and B are mutually exclusive $\Leftrightarrow$ $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$



- A and B are independent $\Leftrightarrow \mathbb{P}(A \cap B) = \mathbb{P}(A) \cdot \mathbb{P}(B)$

- If $\mathbb{P}(A) > 0$, $\mathbb{P}(B) > 0$, then
  mutually exclusive $\Rightarrow$ not independent
  (independent $\not\Rightarrow$ not mutually exclusive)

Ex. Toss Two Coins

Case 1

Ω

HH	HT
TH	TT

prob.

1/4	1/4
1/4	1/4

Case 2

Ω

HH	HT
TH	TT

prob.

1/10	2/10
3/10	4/10

Let A = H on 1st toss

B = H on 2nd toss

Are A and B independent?

$$\underbrace{\mathbb{P}(A \cap B)}_{\{HH\}} \stackrel{?}{=} \mathbb{P}(A) \cdot \mathbb{P}(B)$$

Case 1 $\mathbb{P}(A \cap B) = \frac{1}{4} \stackrel{?}{=} \left(\frac{1}{4}+\frac{1}{4}\right) \cdot \left(\frac{1}{4}+\frac{1}{4}\right) = \frac{1}{4}$ ✓

yes, independent

Case 2 $\mathbb{P}(A \cap B) = \frac{1}{10} \stackrel{?}{=} \left(\frac{1}{10}+\frac{2}{10}\right)\left(\frac{1}{10}+\frac{3}{10}\right) = \frac{3}{10} \cdot \frac{4}{10} = \frac{12}{100}$

not independent

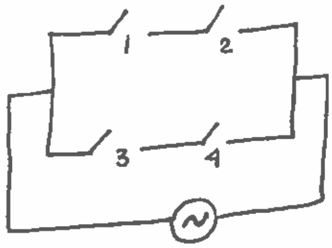
$\Rightarrow$ Independence deals with probability assignment.

Def $A_1, A_2, \ldots, A_n$ are mutually independent if the prob. of the intersection of any number of them equals the product of their probabilities.

Ex. A, B, C are mutually independent if

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

Ex.



A_i : switch i is closed

$\mathbb{P}(A_i) = p$, $i = 1, 2, 3, 4$

A_1, A_2, A_3, A_4 are mutually indep.

Let C = current goes through

$\mathbb{P}(C) = ?$ not mutually exclusive

$$C = (A_1 \cap A_2) \cup (A_3 \cap A_4)$$

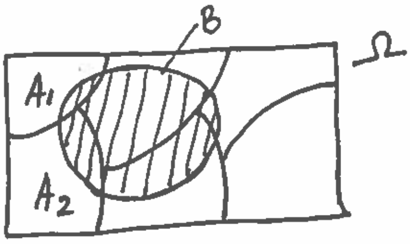
$$\mathbb{P}(C) = \mathbb{P}(A_1 \cap A_2) + \mathbb{P}(A_3 \cap A_4) - \mathbb{P}(A_1 \cap A_2 \cap A_3 \cap A_4)$$
$$= p^2 + p^2 - p^4 = 2p^2 - p^4$$

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

$$\Rightarrow \mathbb{P}(C) = 2 \cdot \frac{1}{4} - \frac{1}{16} = \frac{7}{16} < \frac{1}{2}.$$

Law of Total Probability

$A_1, A_2, \dots$ partition of Ω $\left(\begin{array}{l}\text{all } A_j \text{ are disjoint} \\ \text{and their union} = \Omega\end{array}\right)$

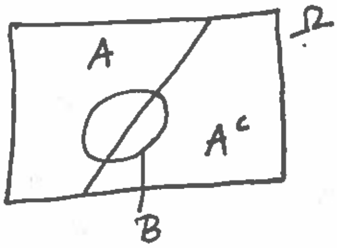


$$B = \bigcup_{j=1}^{\infty} \underbrace{(A_j \cap B)}_{\text{disjoint}}$$

$$\Rightarrow \mathbb{P}(B) = \sum_{j=1}^{\infty} \mathbb{P}(A_j \cap B) =$$

$$\overset{\text{multiplication rule}}{=} \sum_{j=1}^{\infty} \mathbb{P}(B \mid A_j) \cdot \mathbb{P}(A_j)$$

Law of total probability



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

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

Application to our example

Let G_i = i^{th} item is good

D_i = i^{th} item is defective , $i = 1, 2$

$\mathbb{P}(2^{nd}$ item is defective$) = \mathbb{P}(D_2) =$

$$= \mathbb{P}(D_2 \cap G_1) + \mathbb{P}(D_2 \cap D_1) =$$

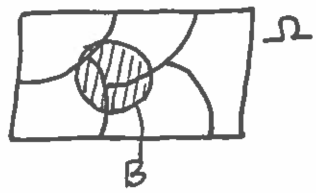
$$= \mathbb{P}(D_2 \mid G_1) \cdot \mathbb{P}(G_1) + \mathbb{P}(D_2 \mid D_1) \cdot \mathbb{P}(D_1) =$$

$$= \frac{d}{d+g-1} \cdot \frac{g}{d+g} + \frac{d-1}{d+g-1} \cdot \frac{d}{d+g} = \frac{d\,\cancel{(g+d-1)}}{\cancel{(d+g-1)}(d+g)} = \frac{d}{d+g}$$

Bayes's Rule

Let $A_1, A_2, \ldots$ be a partition of Ω

$$\mathbb{P}(B) = \sum_{j=1}^{\infty} \mathbb{P}(B|A_j) \cdot \mathbb{P}(A_j)$$



Problem Choose one of the A's, say A_n,

and compute $\mathbb{P}(A_n | B)$ we know B happened

If B is observed what is the probability than A_n was the cause.

$$\underbrace{\mathbb{P}(A_n|B)} = \frac{\mathbb{P}(A_n \cap B)}{\mathbb{P}(B)} = \underbrace{\frac{\mathbb{P}(B|A_n) \cdot \mathbb{P}(A_n)}{\sum_j \mathbb{P}(B|A_j) \cdot \mathbb{P}(A_j)}}_{\text{Bayes formula}}$$

$$\mathbb{P}(A_n|B) = \frac{\mathbb{P}(B|A_n) \cdot \mathbb{P}(A_n)}{\sum_{j=1}^{\infty} \mathbb{P}(B|A_j) \cdot \mathbb{P}(A_j)}$$

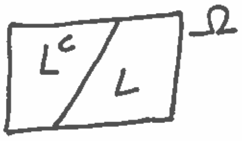
Ex. $S = \{\text{person is a smoker}\}$

$L = \{\text{person has lung disease}\}$

Data: $\left.\begin{array}{l} \mathbb{P}(L) = 0.07 \\ \mathbb{P}(S|L) = 0.9 \\ \mathbb{P}(S|L^c) = 0.253 \end{array}\right\}$ given

Find $\mathbb{P}(\text{person has lung disease} \mid \text{he is a smoker}) =$

$$\mathbb{P}(L|S) = \frac{\mathbb{P}(L \cap S)}{\mathbb{P}(S)} = \frac{\mathbb{P}(S|L).\mathbb{P}(L)}{\mathbb{P}(S|L).\mathbb{P}(L) + \mathbb{P}(S|L^c).\mathbb{P}(L^c)}$$



$$= \frac{(0.900)(0.070)}{(0.900).(0.070) + (0.253)(1-0.070)} = 0.211 \approx 21\%$$

$\mathbb{P}(L)$ = prior probability

$\mathbb{P}(L|S)$ = posterior probability

L = she loves me

S = has supper with me

$\mathbb{P}(L) = 0.070$ (small)

$\mathbb{P}(S|L) = 0.900$

$\mathbb{P}(S|L^c) = 0.253$ (indifferent)

$\Rightarrow \mathbb{P}(L|S) \approx 21\%$

Ex. 4.74 p. 165

4.74 If you toss a balanced coin, in the long run, you get a head half the time. If you toss a certain unbalanced coin, in the long run, you get a head two-thirds of the time. First you choose one of these two coins at random, then you toss the chosen coin twice. Find the conditional probability that the balanced coin is chosen, given that

a) the first toss is a head and the second toss is a tail.
b) the second toss is a head and the first toss is a tail.
c) exactly one of the two tosses is a head.

1/2 H	1/2 T

balanced

2/3 H	1/3 T

unbalanced

① choose a coin $\rightarrow \mathbb{P}(\text{Balanced}) = \mathbb{P}(\text{Unbalanced}) = \frac{1}{2}$

② toss it twice

a) $$\mathbb{P}(\text{Balanced} \mid \{H,T\}) \overset{\text{Bayes}}{=} \frac{\mathbb{P}(\{H,T\} \mid B) \cdot \mathbb{P}(B)}{\mathbb{P}(\{H,T\} \mid B) \cdot \mathbb{P}(B) + \mathbb{P}(\{H,T\} \mid B^c) \cdot \mathbb{P}(\underbrace{B^c}_{\substack{U \\ \text{(unbalanced)}}})}$$

$$= \frac{\frac{1}{2}\left(1-\frac{1}{2}\right)\cdot\frac{1}{2}}{\frac{1}{2}\left(1-\frac{1}{2}\right)\frac{1}{2} + \left(\frac{2}{3}\right)\cdot\left(1-\frac{2}{3}\right)\cdot\frac{1}{2}} = \frac{9}{17}$$

b) $$\mathbb{P}(B \mid \{T,H\}) = \frac{\frac{1}{2}\cdot\frac{1}{2}\cdot\frac{1}{2}}{\frac{1}{2}\cdot\frac{1}{2}\cdot\frac{1}{2} + \frac{1}{3}\cdot\frac{2}{3}\cdot\frac{1}{2}} = \frac{9}{17}$$

c) $$\mathbb{P}(B \mid \underbrace{\text{exactly one } H}_{C}) = \mathbb{P}(B \mid C) = \frac{\mathbb{P}(C \mid B)\cdot\mathbb{P}(B)}{\mathbb{P}(C \mid B)\cdot\mathbb{P}(B) + \mathbb{P}(C \mid U)\cdot\mathbb{P}}$$

$$\mathbb{P}(C \mid B) = \mathbb{P}(\{HT\} \cup \{TH\} \mid B) =$$
$$= \mathbb{P}(\{HT\} \mid B) + \mathbb{P}(\{TH\} \mid B) = \frac{1}{2}\cdot\frac{1}{2} + \frac{1}{2}\cdot\frac{1}{2} = \frac{1}{2}$$

$$\mathbb{P}(C \mid U) = \mathbb{P}(\{HT\} \mid U) + \mathbb{P}(\{TH\} \mid U) = \frac{2}{3}\cdot\frac{1}{3} + \frac{1}{3}\cdot\frac{2}{3} = \frac{4}{9}$$

$$\Rightarrow \mathbb{P}(B \mid C) = \frac{\frac{1}{2}\cdot\frac{1}{2}}{\frac{1}{2}\cdot\frac{1}{2} + \frac{4}{9}\cdot\frac{1}{2}} = \frac{9}{17}.$$

Now suppose that the person selected is a smoker. On the basis of this additional information, we can revise the probability that the person has lung disease. We do so by determining the conditional probability that a randomly selected person has lung disease, given that the person selected is a smoker: $P(L \mid S) = 0.211$ (from Example 4.21). This revised probability is called a **posterior probability** because it represents the probability that the person selected has lung disease *after* we know that the person is a smoker.

EXERCISES 4.4 Basic Exercises

4.66 The National Sporting Goods Association collects and publishes data on participation in selected sports activities. For Americans 7 years old or older, 17.4% of males and 4.5% of females play golf. And, according to the U.S. Census Bureau's *Current Population Reports*, of Americans 7 years old or older, 48.6% are male and 51.4% are female. From among those Americans who are 7 years old or older, one is selected at random. Find the probability that the person selected

a) plays golf, given that the person is a female.
b) is a female, given that the person plays golf.
c) Interpret your answers in parts (a) and (b) in terms of percentages.

4.67 A survey conducted by TELENATION/Market Facts, Inc., combined with information from the U.S. Census Bureau's *Current Population Reports*, yielded the table given in Exercise 4.30 on page 145. What percentage of adult moviegoers are between 25 and 34 years old?

4.68 In a certain population of registered voters, 40% are Democrats, 32% are Republicans, and 28% are Independents. Sixty percent of the Democrats, 80% of the Republicans, and 30% of the Independents favor increased spending to combat terrorism. If a person chosen at random from this population favors increased spending to combat terrorism, what is the probability that the person is a Democrat?

4.69 An insurance company classifies people as *normal* or *accident prone*. Suppose that the probability that a normal person has an accident in a specified year is 0.2 and that for an accident prone person this probability is 0.6. Further suppose that 18% of the policyholders are accident prone. A policyholder had no accidents in a specified year. What is the probability that he or she is accident prone?

4.70 Refer to Example 4.18 on page 155. If you win a game of craps, what is the probability that your first toss resulted in a sum of 4?

4.71 An urn contains one red marble and nine green marbles; a second urn contains two red marbles and eight green marbles; and a third urn contains three red marbles and seven green marbles. An urn is chosen at random, and then one marble is randomly selected from the chosen urn.

a) Given that the marble obtained is red, what is the probability that the chosen urn was the first one; the second one; the third one?
b) Modify the problem by adding one million additional green marbles to each urn. Given the unlikely result that the marble selected is red, what are the posterior probabilities of the three urns? How is the result affected by the addition of the extra green marbles?
c) In part (a), find the posterior probabilities of the three urns, given that the marble selected is green. Then do the same after the additional green marbles of part (b) have been introduced. How is this result affected by the addition of the extra green marbles?

4.69 For a randomly selected policyholder, let

A = accident prone

B = no accidents in a specified year.

$\mathbb{P}(A) = 0.18$

$\mathbb{P}(B|A) = 1 - \mathbb{P}(B^c|A) = 1 - 0.6 = 0.4$

$\mathbb{P}(B|A^c) = 1 - \mathbb{P}(B^c|A^c) = 1 - 0.2 = 0.8$

Bayes' Rule:

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(B|A) \cdot \mathbb{P}(A)}{\mathbb{P}(B|A) \cdot \mathbb{P}(A) + \mathbb{P}(B|A^c) \cdot \mathbb{P}(A^c)}$$

$$= \frac{0.18 \cdot 0.4}{0.18 \cdot 0.4 + 0.82 \cdot 0.8} \approx 0.0989.$$

2.6 An urn contains 10 balls, numbered 0, 1, 2, ..., 9. Three balls are removed, one at a time, without replacement.
a) Obtain the sample space for this random experiment.
b) Determine, as a subset of the sample space, the event that an even number of odd-numbered balls are removed from the urn.

2.7 Refer to Example 2.3 on page 27 where two dice are rolled, one black and one gray. For $i = 2, 3, \ldots, 12$, determine explicitly as a subset of the sample space the event A_i that the sum of the faces is i.

2.8 Consider the following random experiment: First a die is rolled and you observe the number of dots facing up; then a coin is tossed the number of times that the die shows and you observe the total number of heads.
a) Determine the sample space for this random experiment.
b) Determine the event that the total number of heads is even.

2.9 George and Laura take turns tossing a coin. The first person to get a tail wins. George goes first. *Note:* You may assume that eventually a tail will be tossed.
a) Describe the sample space for this random experiment.
b) Determine, as a subset of the sample space, the event that Laura wins.

2.10 This exercise considers two random experiments involving the repeated tossing of a coin. *Note:* You may assume that eventually a head will be tossed.
a) If the coin is tossed until the first time a head appears, find the sample space.
b) If the coin is tossed until the second time a head appears, find the sample space.
c) For the experiment in part (a), express the event that the coin is tossed exactly six times in the form $\{\ldots\}$, where in place of "..." you list all of the outcomes in that event.
d) Repeat part (c) for the experiment described in part (b).

2.11 From 10 men and 8 women in a pool of potential jurors, 12 are chosen at random to constitute a jury. Suppose that you observe the number of men who are chosen for the jury. Let A be the event that at least half of the 12 jurors are men and let B be the event that at least half of the 8 women are on the jury.
a) Determine the sample space for this random experiment.
b) Find $A \cup B$, $A \cap B$, and $A \cap B^c$, listing all the outcomes for each of those three events.
c) Are A and B mutually exclusive? A and B^c? A^c and B^c? Explain your answers.

2.12 Let A and B be events of a sample space.
a) Show that, if A and B^c are mutually exclusive, then B occurs whenever A occurs.
b) Show that, if B occurs whenever A occurs, then A and B^c are mutually exclusive.

2.13 Let A, B, and C be events of a sample space. Write a mathematical expression for each of the following events.
a) A occurs, but B doesn't occur.
b) Exactly one of A and B occurs.
c) Exactly one of A, B, and C occurs.
d) At most two of A, B, and C occur.

2.14 Refer to Example 2.17 on page 34, but now suppose that two cards are selected at random, one after the other, without replacement.
a) What is Ω for this random experiment?
b) Let A be the event that at least one of the cards is a face card and let B be the event that at least one of the cards is an ace. Are A and B mutually exclusive? Why or why not?

The probability that a head will eventually be tossed is 1. In other words, we will eventually get a head when we repeatedly toss a balanced coin. As we demonstrate in Example 4.17, this result holds regardless of whether the coin is balanced, provided only that the probability is not 0 of getting a head when the coin is tossed once. ■

EXERCISES 2.4 Basic Exercises

2.58 Let A and B be events of a sample space. Provide an example where, as sets, A is a proper subset of B, but $P(A) = P(B)$.

2.59 Give an example to show that the converse of the domination principle fails.

2.60 A person is selected at random from among the inhabitants of a state. Which is more probable: that the person so chosen is a lawyer, or that the person so chosen is a Republican lawyer? Explain your answer.

2.61 Refer to Exercise 2.41 on page 62. Use the complementation rule to find the probability that at least one Republican will be on the subcommittee. Why would use of the complementation rule for this problem make things easier than if that rule weren't used?

2.62 Refer to Exercise 2.20 on page 45. Determine the probability that an oil spill in U.S. navigable and territorial waters doesn't occur in the Gulf of Mexico
a) without use of the complementation rule **b)** by using the complementation rule.
c) Compare the work done in your solutions in parts (a) and (b).

2.63 Refer to Exercise 2.39 on page 61. Suppose that a player on the New England Patriots is selected at random. Determine the probability that the player obtained
a) has at least 1 year of experience. **b)** weighs at most 300 lb.
c) is either a rookie or weighs more than 300 lb. Solve this problem both with and without use of the general addition rule and compare your work.

2.64 If a point is selected at random from the unit square $\{ (x, y) : 0 \le x \le 1, 0 \le y \le 1 \}$, find the probability that the magnitude of the difference between the x and y coordinates of the point obtained is at most 1/4. Solve this problem both with and without use of the complementation rule and compare your work.

2.65 According to *Current Population Reports*, published by the U.S. Bureau of the Census, 51.0% of U.S. adults are female, 7.1% are divorced, and 4.1% are divorced females. Determine the probability that a U.S. adult selected at random is
a) either female or divorced. **b)** a male.
c) a female but not divorced. **d)** a divorced male.

• **2.66** Let A and B be events such that $P(A) = 1/4$, $P(B) = 1/3$, and $P(A \cup B) = 1/2$.
a) Are events A and B mutually exclusive? Explain your answer.
b) Determine $P(A \cap B)$.

2.67 Let A and B be events such that $P(A) = 1/3$, $P(A \cup B) = 5/8$, and $P(A \cap B) = 1/10$. Determine
a) $P(B)$. **b)** $P(A \cap B^c)$. **c)** $P(A \cup B^c)$. **d)** $P(A^c \cup B^c)$.

2.68 Gerald Kushel, Ed.D., was interviewed by *Bottom Line/Personal* on the secrets of successful people. To study success, Kushel questioned 1200 people, among whom were lawyers, artists, teachers, and students. He found that 15% enjoy neither their jobs nor their

HW 3

✓ 1) 4.40 p.146

✓ 2) 4.44 p. 157

X 3) 4.46 p. 157

✓ 4) 4.69 p. 164

Extra credit

x 1) 4.47 p. 157

x 2) 4.49 p. 158

divorced and that "... compared with moms whose offspring had died, nearly twice the percentage of females that raised their youngsters to the fledgling stage moved out of the family flock and took mates elsewhere the next season—81% versus 43%." For the females in this study, find the percentage

a) whose offspring died.

b) that divorced and whose offspring died.

c) whose offspring died among those that divorced.

4.38 Urn I contains two red balls and three green balls, Urn II contains four red balls and one green ball, and Urn III contains two red balls and two green balls. A ball is chosen at random from Urn I and placed in Urn II. Next, a ball is chosen at random from Urn II and placed in Urn III. Then a ball is chosen at random from Urn III. What is the probability that it is red?

4.39 In Pólya's urn scheme (page 141), use mathematical induction to prove that the probability is $r/(b+r)$ that the nth ball drawn is red.

4.40 Suppose that, for each $k \in \{0, 1, 2, \dots, n\}$, there is a bowl containing k red marbles and $n-k$ green marbles. One bowl is chosen at random from these $n+1$ bowls and then two marbles are selected at random without replacement from the chosen bowl. Determine the probability that both marbles are red.

4.41 Tom, Dick, and Harry compete in a game that consists of a succession of rounds. In each round, two of the players compete, the winner being decided by the toss of a balanced coin; the third player sits on a bench and waits. In each round after the first, the loser of the previous round sits on the bench while the other two compete. The first player to win two consecutive rounds wins the game. The problem is to determine the probability that Tom wins the game, the probability that Dick wins the game, and the probability that Harry wins the game. Proceed as follows.

a) In each round after the first, call the winner of the previous round the incumbent, the person against whom the incumbent competes in that round the challenger, and the player who sits out that round the benchwarmer. Let p, q, and r denote the probabilities that the current incumbent, challenger, and benchwarmer, respectively, win the game. Determine p, q, and r. *Hint:* Use the law of total probability, conditioning on whether the current incumbent does or does not win the current round.

b) Suppose that, in the first round, Tom and Dick play. Determine the probability that Tom wins the game, the probability that Dick wins the game, and the probability that Harry wins the game. *Hint:* Use the law of total probability, conditioning on whether Tom does or does not win the first round.

c) Now suppose that the initial benchwarmer is obtained by choosing one of Tom, Dick, and Harry at random. Use two different methods to determine the probability that Tom wins the game, the probability that Dick wins the game, and the probability that Harry wins the game: (i) a symmetry argument; (ii) conditioning on whether the specified contestant is chosen as the benchwarmer.

. **4.42** An ordinary deck of 52 playing cards is randomly divided into two equal stacks of 26 cards each. A card randomly drawn from the first stack turns out to be the queen of spades. That queen is placed in the second stack and then a card is randomly drawn from that stack. What is the probability that the card obtained is a queen? *Note:* There is an efficient way to do this problem by conditioning suitably and using symmetry.

Mutually Exclusive Versus Independent Events

The terms *mutually exclusive* and *independent* refer to different concepts. Mutually exclusive events are those that can't occur simultaneously. Independent events are those for which the occurrence of some doesn't affect the probabilities of the others occurring. If two or more events are mutually exclusive, the occurrence of one precludes the occurrence of the others. Hence, two or more events with positive probabilities can't be both mutually exclusive and independent. See Exercise 4.55 for more on this issue.

EXERCISES 4.3 Basic Exercises

4.43 Verify the following statements made on page 148.

a) If event B is independent of event A in the sense of Definition 4.3, then the two events are independent in the sense of Definition 4.4.

b) If $P(A) > 0$ and events A and B are independent in the sense of Definition 4.4, then event B is independent of event A in the sense of Definition 4.3.

4.44 The U.S. National Center for Health Statistics compiles data on injuries and publishes the information in *Vital and Health Statistics*. A contingency table for injuries in the United States, by circumstance and sex, is as follows. Frequencies are in millions.

		Circumstance			
		Work C_1	Home C_2	Other C_3	Total
Sex	Male S_1	8.0	9.8	17.8	35.6
	Female S_2	1.3	11.6	12.9	25.8
	Total	9.3	21.4	30.7	61.4

a) Are events C_1 and S_2 independent? Explain.

b) Is the event that an injured person is male independent of the event that an injured person was hurt at home? Explain.

4.45 Refer to the joint probability distribution in Exercise 4.8 on page 135 for living arrangement and age of U.S. citizens 15 years of age and older. Are events A_2 and L_1 independent? Interpret your answer.

4.46 In the game of Yahtzee, five balanced dice are rolled. What is the probability

a) of rolling all 2s?

b) that all the dice come up the same number?

c) of getting a full house—three of one number and two of another?

4.47 Let A be an event of a sample space. Verify the following statements.

a) If $P(A) = 0$ or $P(A) = 1$, then, for each event B of the sample space, A and B are independent events.

b) If A and A are independent events, then $P(A) = 0$ or $P(A) = 1$.

HW 2

v 1) 3.35 p. 107

v 2) 3.58 p. 116

x 3) 3.66 p. 116

Extra credit

x 1) 3.49 p. 109

v 2) 3.61 a), b) p. 116

Solution a) An unordered sample of size n without replacement from a population of size N is a combination of n objects from a collection of N objects. Hence, by the combinations rule, there are

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

possible unordered samples of size n without replacement from a population of size N.

b) Substituting $N = 52$ and $n = 5$ into the formula in part (a), we find that

$$\binom{52}{5} = \frac{52!}{5!\,(52-5)!} = 2{,}598{,}960.$$

There are 2,598,960 possible unordered samples of size 5 without replacement from a population of size 52. ■

When we refer simply to "a sample" from a finite population, we mean an unordered sample without replacement unless specifically stated otherwise. As we showed in Example 3.13(a), there are $\binom{N}{n}$ possible samples of size n from a population of size N.

Combining Counting Rules

In Example 3.14, we illustrate how counting rules are sometimes combined to obtain a required result.

EXAMPLE 3.14 ***The Combinations Rule***

Five-Card Draw Poker A five-card draw poker hand consists of 5 cards dealt from an ordinary deck of 52 playing cards. The order in which the cards are dealt is unimportant.

a) How many possible five-card draw poker hands are there?
b) How many different hands are there consisting of three kings and two queens?
c) The hand in part (b) is an example of a full house: three cards of one denomination (face value) and two of another. How many different full houses are there?

Solution a) A five-card draw poker hand is a combination of 5 objects (the hand) from a collection of 52 objects (the deck). Hence, by the combinations rule,

$$\binom{52}{5} = \frac{52!}{5!\,(52-5)!} = 2{,}598{,}960$$

hands are possible.

b) For this part, we use both the combinations rule and the BCR. There are three actions: the choice of which three of the four kings are to appear in the hand, the choice of which two of the four queens are to appear in the hand, and the choice of which zero of the other (nonking and nonqueen) 44 cards are to appear in the hand. By the combinations rule, we have $\binom{4}{3}$ possibilities for the first action, $\binom{4}{2}$ possibilities for the second action, and $\binom{44}{0}$ possibilities for the third action. Hence, by the BCR,

Random Variables

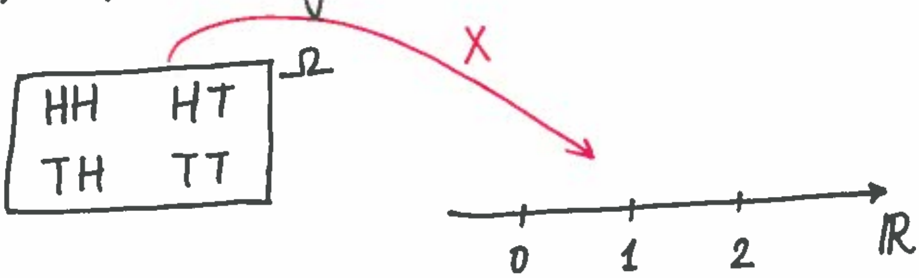
Map from the sample space Ω onto the real line $\mathbb{R}$.

Ex. $\mathcal{E} = \{\text{toss two coins}\}$

$\Omega = \{HH, HT, TH, TT\}$

Suppose we are interested in

X = number of heads we obtain



$\underbrace{X}_{\text{function}} : \Omega \to \mathbb{R}$

range of X: $R_X = \{0, 1, 2\}$

$\underbrace{\{TT\}}_{\text{event}} = \{X = 0\} = \{\omega : X(\omega) = 0\}$

$\{HT, TH\} = \{X = 1\} = \{\omega : X(\omega) = 1\}$

$\{HH\} = \{X = 2\} = \{\omega : X(\omega) = 2\}$

$\mathbb{P}(X = 0) = \mathbb{P}(TT) = \frac{1}{4}$

since disjoint

$\mathbb{P}(X = 1) = \mathbb{P}(HT, TH) = \mathbb{P}(HT) + \mathbb{P}(TH) = \frac{2}{4} = \frac{1}{2}$

$\mathbb{P}(X = 2) = \mathbb{P}(HH) = \frac{1}{4}$

$\mathbb{P}(X = x)$ = prob. that X takes on a specific value x

$$\underset{\substack{\uparrow \\ \text{random variable}}}{\mathbb{P}(X = x)} = \begin{cases} 1/4, & \text{if } x = 0 \\ 1/2, & \text{if } x = 1 \\ 1/4, & \text{if } x = 2 \\ 0, & \text{otherwise} \end{cases}$$

$$\sum_{x \in R_X} \mathbb{P}(X = x) = 1.$$

How to describe a random variable

Specify 1) R_X = range of X

↳ list of values of X

In the example $R_X = \{0, 1, 2\}$

2) $\mathbb{P}(X = x)$ for all $x \in R_X$

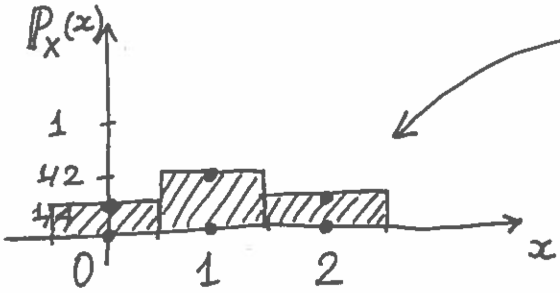
such that $\mathbb{P}(X = x) \geq 0$

$$\sum_{x \in R_X} \mathbb{P}(X = x) = 1$$

<u>Note:</u> $\mathbb{P}_X(x) = \mathbb{P}(X = x), \quad x \in R_X$

is called
- <u>the probability mass function of X</u> (PMF)
- <u>probability distribution of X</u>

$$\mathbb{P}_X(x) = \begin{cases} 1/4, & \text{if } x = 0 \\ 1/2, & \text{if } x = 1 \\ 1/4, & \text{if } x = 2 \end{cases}$$



area is probability

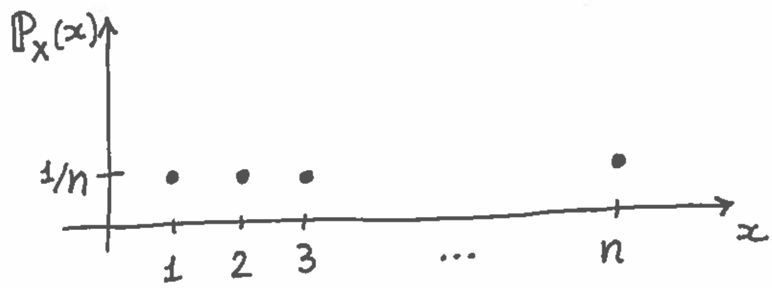
Typical Models of rv's

1) Discrete Uniform

Let $n \geqq 1$ be given.

Def $R_X = \{1, 2, \dots, n\}$

Specify $\mathbb{P}_X(x) = \mathbb{P}(X = x) = \frac{1}{n}$ for any $x \in R_X$



n = parameter of the distribution

Notation: $X \sim \text{Unif}(n)$

or

$\left(X \sim \text{Unif}(a, b) \right)$

Ex. Rolling a die

$X \sim \text{Unif}(6)$

the rv X has the discrete uniform distribution with parameter $n = 6$.

2) Bernoulli

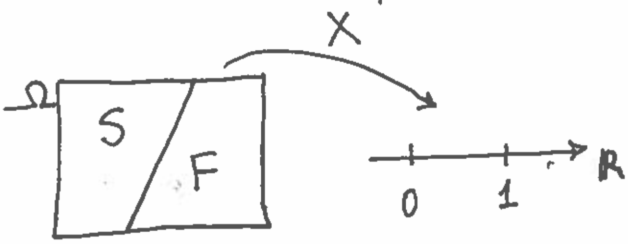
$\mathcal{E}$ has only two elementary outcomes

the Bernoulli random experiment

$\Omega = \{ S, F \}$

$\mathbb{P}(S) = p \quad , \quad 0 \leqq p \leqq 1$

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



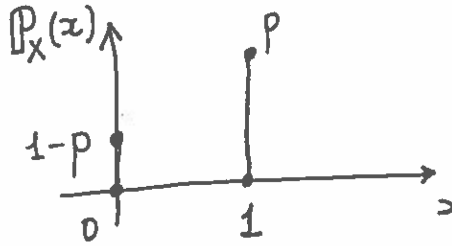
X = number of successes

$X(S) = 1$

$X(F) = 0$

So, $R_X = \{0, 1\}$

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

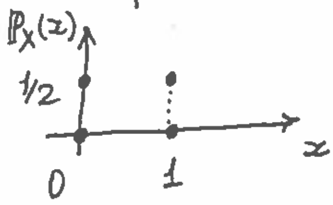


not necessarily the same prob. for success & failure

X is called a Bernoulli r.v.

$X \sim \text{Bern}(p)$ ← parameter of the distribution

* Special case of Discrete Uniform, where $p = \frac{1}{2}$



$$\mathbb{P}_X(k) = \mathbb{P}(X=k) = p^k (1-p)^{1-k}, \quad k = 0, 1.$$

3) Binomial

Perform $\underline{\underline{n}}$ independent Bernoulli trials

Call X = number of successes.

$$\Omega = \{ \overbrace{(S, S, \dots, S)}^{\text{length } n}, (S, S, \dots, F), (S, F, S, \dots, S), \dots \}$$

$R_X = \{0, 1, 2, \dots, n\}$

- prob. of success is p from Bernoulli

$\mathbb{P}(X = k) = ?$, $k = 0, 1, \dots, n$

$$\mathbb{P}(X = n) = p^n = \mathbb{P}(\underbrace{S, S, \dots, S}_{n \text{ times}})$$

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

$$\mathbb{P}(X = 1) = \mathbb{P}(1 \text{ success } \& \ (n-1) \text{ failures}) = \mathbb{P}(S, F, \dots, F) + \mathbb{P}(F, S, F, \dots, F) + \dots$$

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

$$\mathbb{P}(X = k) = \mathbb{P}(k \text{ successes and } (n-k) \text{ failures})$$

$$= \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, \dots, n$$

permute the S's

$$\sum_{k=0}^{n} \mathbb{P}(X=k) = \sum_{k=0}^{n} \binom{n}{k} \underbrace{p}_{a}^k \underbrace{(1-p)}_{b}^{n-k}$$

Binomial theory $(a+b)^n = \sum_{k=0}^{n} \binom{n}{k} a^k \cdot b^{n-k}$

$$= (\cancel{p} + (1-\cancel{p}))^n = 1^n = 1$$

Ex. Test 20 questions (True / False)

You drank too much last night

$\rightarrow$ guess answers.

p = prob. of correct answer $1/2$

To pass the test you need 18 correct answers.

$\mathbb{P}(\text{pass test}) = ?$

repeating bernoulli 20 times

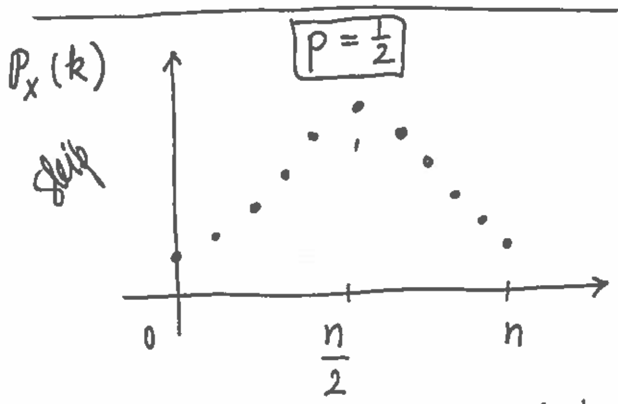
X = number of correct answers

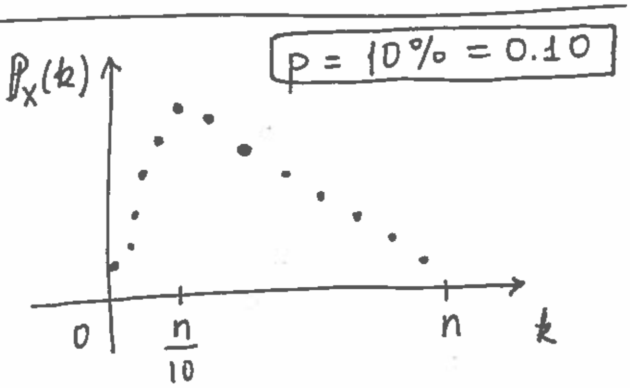
$$\mathbb{P}(X \geqq 18) = \mathbb{P}(X=18) + \mathbb{P}(X=19) + \mathbb{P}(X=20) =$$

$$= \binom{20}{18} p^{18} (1-p)^2 + \binom{20}{19} \left(\tfrac{1}{2}\right)^{19} \left(1-\tfrac{1}{2}\right)^1 + \binom{20}{20} \left(\tfrac{1}{2}\right)^{20} \cdot \left(\tfrac{1}{2}\right)^0$$

$$= \left[\binom{20}{18} + \binom{20}{19} + \binom{20}{20}\right] \cdot \left(\tfrac{1}{2}\right)^{20}$$

$$\mathbb{P}(\underbrace{X \leqq 17}_{\text{fail}}) = 1 - \mathbb{P}(X \geqq 18)$$





Mode of distribution is around $\underline{\underline{pn}}$

Notation $X \sim Bin(\underbrace{n, p}_{\text{parameters of the distribution}})$

4) Geometric

Keep performing Bernoulli independent trials until 1^{st} time we get a success.

Let X = number of trials until the 1^{st} success.

$R_X = \{1, 2, \dots\}$ range of X is not finite

$\mathbb{P}(X=k) = ?$ $\quad k \geq 1$

$$\mathbb{P}(\underbrace{FF\dots F}_{k-1}\ \underset{k^{th}}{S}) = (1-p)^{k-1} \cdot p$$

$$\sum_{k=1}^{\infty} \mathbb{P}(X=k) = \sum_{k=1}^{\infty} p(1-p)^{k-1} = p \underbrace{\sum_{k=1}^{\infty}(1-p)^{k-1}}_{} = p \cdot \frac{1}{1-(1-p)} = p \cdot \frac{1}{p} =$$

geometric series $r < 1$

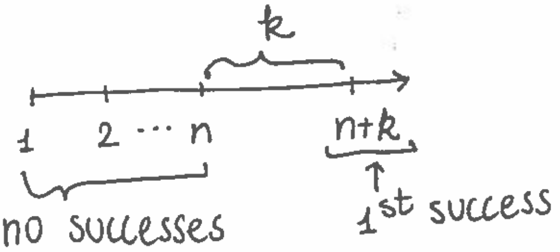
$a + ar + ar^2 + \dots = \dfrac{a}{1-r}$

parameter of the distribution

$X \sim \text{Geom}(p)$, $p > 0$

No memory property

$X \sim \text{Geom}(p)$



$$\mathbb{P}\left(\begin{matrix}\text{first success at}\\ (n+k)^{th} \text{ trial}\end{matrix} \,\middle|\, \begin{matrix}\text{no successes in}\\ \text{first } n \text{ trials}\end{matrix}\right) =$$

$$= \mathbb{P}(X = n+k \mid X > n) =$$

$$= \frac{\mathbb{P}(X = n+k \text{ and } X > n)}{\mathbb{P}(X>n)} = \frac{\mathbb{P}(X=n+k)}{\mathbb{P}(X>n)} = \frac{(1-p)^{n+k-1} \cdot p}{(1-p)^n} =$$

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

$$\Rightarrow \boxed{\mathbb{P}(X = n+k \mid X > n) = \mathbb{P}(X=k)} \text{ no memory}$$

Ex. Two players A and B, keep rolling a die until one of them wins: 6 = A wins

1 or 2 = B wins

$\mathbb{P}(\text{A wins}) = ?$

$$\mathbb{P}(\text{A wins}) = \sum_{n=1}^{\infty} \mathbb{P}(\text{A wins in } n \text{ rolls}) =$$

$$= \sum_{n=1}^{\infty} [\underbrace{\mathbb{P}(\text{no one wins})}_{\substack{\mathbb{P}(3,4,\text{or } 5) \\ = \frac{3}{6} = \frac{1}{2}}}]^{n-1} \cdot \underbrace{\mathbb{P}(\text{A wins at } n^{\text{th}} \text{ roll})}_{\substack{\mathbb{P}(6) \\ = \frac{1}{6}}} =$$

$$= \sum_{n=1}^{\infty} \left(\frac{1}{2}\right)^{n-1} \cdot \left(\frac{1}{6}\right) = \frac{1}{6} \cdot \frac{1}{1 - 1/2} = \boxed{\frac{1}{3}}.$$

5) Pascal (Negative Binomial)

Repeat Bernoulli independent trials until we get r successes

$X \sim \text{Pascal}(p, r)$ (not just 1 success but r)

$R_X = \{r, r+1, \dots\}$

for $k \geq r$

$$P(X = \underline{\underline{k}}) = P(\underbrace{\overbrace{\dots SF \dots}^{(k-1)\text{ trials}}}_{(r-1)\text{ successes}} \underset{r^{th}\text{ success}}{\overset{k^{th}\text{ trial}}{S}}) = \binom{k-1}{r-1} p^{r-1} (1-p)^{k-r} \cdot p$$

$(k-r)$ failures

$$= \binom{k-1}{r-1} p^{r} \cdot (1-p)^{k-r}.$$

skip

f) Identify the probability distribution of X as right-skewed, symmetric, or left-skewed without consulting its probability distribution or drawing its probability histogram.

g) Draw a probability histogram for X.

5.43 Sickle cell anemia is an inherited blood disease that occurs primarily in blacks. In the United States, about 15 of every 10,000 black children have sickle cell anemia. The red blood cells of an affected person are abnormal; the result is severe chronic anemia (inability to carry the required amount of oxygen), which causes headaches, shortness of breath, jaundice, increased risk of pneumococcal pneumonia and gallstones, and other severe problems. Sickle cell anemia occurs in children who inherit an abnormal type of hemoglobin, called hemoglobin S, from both parents. If hemoglobin S is inherited from only one parent, the person is said to have sickle cell trait and is generally free from symptoms. There is a 50% chance that a person who has sickle cell trait will pass hemoglobin S to an offspring.

a) Obtain the probability that a child of two people who have sickle cell trait will have sickle cell anemia.

b) If two people who have sickle cell trait have five children, determine the probability that at least one of the children will have sickle cell anemia.

c) If two people who have sickle cell trait have five children, find the probability distribution of the number of those children who will have sickle cell anemia.

d) Construct a probability histogram for the probability distribution in part (c).

5.44 If all sex distributions are equally likely, what proportion of families with five children have three girls and two boys?

5.45 A baseball player has a batting average of .260. Suppose that you observe successive at-bats of the player and note for each at-bat whether the player gets a hit.

a) Under what conditions is the assumption of Bernoulli trials appropriate?

If successive tr are indep. and w prob. hit = the player's batting averag

b) Assuming your conditions in part (a), what is the probability that the player will get two or more hits in his next four times at-bat?

X = # of hits in next four times at-bat

$X \sim Bin(4, 0.26)$

$P(X \geq 2)$

$= 1 - P(X < 2)$

$= 1 - (P(X=0) + P(X=1))$

$= 0.279$

5.46 Sixty percent of all voters in a state intend to vote "yes" in a referendum. An opinion poll took a sample of 20 voters. No precaution was taken against the unlikely event of choosing the same voter more than once. Find the probability that in the sample of 20 there are more voters who intend to vote "yes" than voters who don't intend to vote "yes."

5.47 Consider 15 Bernoulli trials with success probability p.

a) Assuming that $p = 0.5$, what is the probability that the number of successes is between six and nine, inclusive? What is the most likely number of successes?

b) Repeat part (a) for $p = 0.4$.

5.48 Availability of statistical software or binomial tables is useful for this problem. Suppose that you throw a balanced die 100 times.

a) What is the probability that you get a 1 at most 4 times? What would you think if that happened?

b) What is the probability that you get a 1 more than 32 times? What would you think if that happened?

5.49 Simultaneously and independently, each of n people make a single toss of a coin with probability p of a head. What is the probability of an "odd man"—that is, one person gets a different result from all the other people?

5.50 How long must a sequence of random decimal digits be so that the probability of getting a 6 or a 7 will exceed 0.95?

trials. The associated PMF is

$$p_X(x) = p(1-p)^{x-1}, \qquad x = 1, 2, \ldots, \tag{5.37}$$

and $p_X(x) = 0$ otherwise. Here, of course, p denotes the success probability.

Instead of considering the number of trials, X, up to and including the first success, we can consider the number of trials, Y, before the first success or, equivalently, the number of failures before the first success. We can determine the PMF of Y directly, or we can use Equation (5.37) and the fact that $Y = X - 1$. In any case, we find that the PMF of the random variable Y is

$$p_Y(y) = p(1-p)^y, \qquad y = 0, 1, \ldots, \tag{5.38}$$

and $p_Y(y) = 0$ otherwise.

Some researchers and textbook authors allude to the geometric distribution as that given by Equation (5.38) instead of Equation (5.37). That is, they define a geometric random variable as the number of failures before the first success rather than the number of trials until the first success. Be sure that you know which form is being considered when you see a reference to a geometric distribution or a geometric random variable.

EXERCISES 5.6 Basic Exercises

5.93 According to the *Daily Racing Form*, the probability is about 0.67 that the favorite in a horse race will finish in the money (first, second, or third place). Suppose that you always bet the favorite "across the board," which means that you win something if the favorite finishes in the money. Let X denote the number of races that you bet until you win something.

a) Determine and identify the PMF of the random variable X.
b) Find the probability that the number of races that you bet until you win something is exactly three; at least three; at most three.
c) How many races must you bet to be at least 99% sure of winning something?

X = # of at-bats up to and including the first hit

$X \sim Geom(0.26)$

5.94 A baseball player has a batting average of .260. Suppose that you observe successive at-bats of the player and note for each at-bat whether the player gets a hit. Presuming that the assumption of Bernoulli trials is appropriate, what is the probability that the first hit by the player occurs

a) on his fifth at-bat? $P(X=5) = (0.26)(1-0.26)^4 = 0.0780$
b) after his fifth at-bat? $P(X>5) = (1-0.26)^5 = 0.222$
c) between his third and tenth at-bats, inclusive? $P(3 \le X \le 10) = P(2 < X \le 10)$ $= P(X>2) - P(X>10)$ $= 0.74^2 - 0.74^{10} = 0.498$

5.95 Let $X \sim \mathcal{G}(p)$. Determine

a) $P(X \text{ is even})$. **b)** $P(X \text{ is odd})$. **c)** $P(2 \le X \le 9 \mid X \ge 4)$.
d) $P(X = k \mid X \le n)$ for $k = 1, 2, \ldots, n$.
e) $P(X = n - k \mid X < n)$ for $k = 1, 2, \ldots, n-1$.

5.96 Let X be a positive-integer valued random variable. If X has the lack-of-memory property, which of these two equations must hold:

$$P(X = 10 \mid X > 6) = P(X = 4) \qquad \text{or} \qquad P(X = 10 \mid X > 6) = P(X = 10)?$$

Explain.

5.112 Suppose that $A_1, A_2, \ldots$ is a countable collection of mutually exclusive events of a sample space. Show that $I_{\bigcup_n A_n} = \sum_n I_{A_n}$.

5.113 Suppose that X has the binomial distribution with parameters n and p.
a) Express X as the sum of n indicator random variables. Interpret your answer.
b) Express X as the sum of n Bernoulli random variables.

5.114 Let $\Omega = \{\omega_1, \ldots, \omega_N\}$ be a finite sample space with equally likely outcomes (i.e., a classical probability model). Define the random variable X on Ω by $X(\omega_k) = k$ for $k = 1, 2, \ldots, N$. Determine and identify the PMF of X.

5.115 Let S consist of the 10 decimal digits. Suppose that a number X is chosen according to the discrete uniform distribution on S and then a number Y is chosen according to the discrete uniform distribution on S with X removed. Determine and identify the PMF of the random variable Y
a) by conditioning on the value of X.
b) by using a symmetry argument.

5.116 According to the *Daily Racing Form*, the probability is about 0.67 that the favorite in a horse race will finish in the money (first, second, or third place). Suppose that you always bet the favorite "across the board," which means that you win something if the favorite finishes in the money. Let X denote the number of races that you bet until you win something three times.
a) Determine and identify the PMF of the random variable X.
b) Find the probability that the number of races that you bet until you win something three times is exactly four; at least four; at most four.

5.117 A baseball player has a batting average of .260. Suppose that you observe successive at-bats of the player and note for each at-bat whether the player gets a hit. Presuming that the assumption of Bernoulli trials is appropriate, what is the probability that the second hit by the player occurs
a) on his fifth at-bat?
b) after his fifth at-bat?
c) between his third and tenth at-bats, inclusive?

5.118 Let X have the negative binomial distribution with parameters r and p. For what values of r does X have the lack-of-memory property? Explain your answer.

5.119 Two balanced dice are rolled until the fourth time a sum of 7 or 11 occurs. What is the probability that it will take more than six rolls?

5.120 Suppose that $X \sim \mathcal{NB}(r, p)$ and that $Y \sim \mathcal{B}(n, p)$.
a) Give a probabilistic argument to show that $P(X > n) = P(Y < r)$.
b) Use the FPF to express the equality in part (a) in terms of PMFs.
c) Using the complementation rule, how many terms of the PMF of X must be evaluated to determine $P(X > n)$?
d) How many terms of the PMF of Y must be evaluated to determine $P(Y < r)$?
e) Use your answers from parts (c) and (d) to comment on the computational savings of using the result of part (a) to evaluate $P(X > n)$ when n is large relative to r.

5.121 Verify that the PMF of a negative binomial random variable with parameters r and p can be expressed in the form given in Equation (5.44) on page 241.

5.122 Refer to Example 5.26 on page 242. Find the probability that the telemarketer makes
a) the second sale by the fifth call and the fifth sale on the fifteenth call.
b) the second sale by the fifth call and the fifth sale by the fifteenth call.

X = # of at-bats up to and including the second hit

$X \sim NB(2, 0.26)$

$\mathbb{P}(X = x) = \binom{x-1}{1} \cdot (0.26)^2 \cdot (0.74)^{x-2}, \quad x = 2, 3, \ldots$

a) $\mathbb{P}(X=5) = \binom{4}{1}(0.26)^2 \cdot (0.74)^3 = 0.11$

b) $\mathbb{P}(X>5) = 1 - \mathbb{P}(X \le 5) = 1 - \sum_{i=2}^{5} \mathbb{P}(X=i) = 0.612$

or alternatively, $X > 5 \Leftrightarrow$ the # of hits in first five at-bats is at most 1

c) $\mathbb{P}(3 \le X \le 10) = \sum_{i=3}^{10} \mathbb{P}(X=i)$

$= 0.71$

5.51 Three numbers are selected independently and randomly from the interval $(0, 1)$.

a) Explain why the probability that the median of the three numbers exceeds 0.7 is the same as the probability of two or more successes in three Bernoulli trials, with a certain success probability p.

b) In part (a), what is the value of p?

✓ **5.52** Six marbles are randomly selected with replacement from an urn. The urn is equally likely to be Urn I (which contains 1 red marble and 1 green marble) or Urn II (which contains 11 red marbles and 9 green marbles). Find the conditional probability that the urn from which the marbles are selected is Urn II, given that the number of red marbles selected is x.

Theory Exercises

5.53 Let n be a positive integer and let $0 < p < 1$.

a) Use the binomial theorem to prove that $\sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} = 1$.

b) Establish the identity in part (a) without using the binomial theorem or doing any computations, but rather by using a probabilistic argument.

5.54 Let $X \sim \mathcal{B}(n, p)$. We say that X is *symmetric* if $p_X(x) = p_X(n-x)$ for all x. Prove that X is symmetric if and only if $p = 1/2$.

5.55 Let $n \in \mathcal{N}$, let $0 \le k \le n$ be a nonnegative integer, and let $0 < p < 1$. Set $q = 1 - p$.

a) Prove the combinatorial identities

$$(n-k)\binom{n}{k} = n\binom{n-1}{k} \quad \text{and} \quad k\binom{n}{k} = n\binom{n-1}{k-1}.$$

b) Use integration by parts to verify that, for $0 \le k \le n-1$,

$$\int_0^q t^{n-k-1}(1-t)^k\,dt = \frac{1}{n-k}p^k q^{n-k} + \frac{k}{n-k}\int_0^q t^{n-k}(1-t)^{k-1}\,dt.$$

c) Let $X \sim \mathcal{B}(n, p)$. Prove that

$$P(X \le k) = n\binom{n-1}{k}\int_0^q t^{n-k-1}(1-t)^k\,dt = n\binom{n-1}{k}B_q(n-k, k+1),$$

where $B_x(\alpha, \beta) = \int_0^x t^{\alpha-1}(1-t)^{\beta-1}\,dt$, the so-called *incomplete beta function*.

d) Let $X \sim \mathcal{B}(n, p)$. Prove that

$$P(X > k) = n\binom{n-1}{k}\int_0^p t^k(1-t)^{n-k-1}\,dt = n\binom{n-1}{k}B_p(k+1, n-k).$$

Hint: Consider the random variable $Y = n - X$ and refer to part (c).

Advanced Exercises

5.56 An urn contains k red marbles and $12 - k$ green marbles. Five times, a marble is chosen randomly and then replaced.

a) What value of k maximizes the probability that the number of red marbles among the 5 obtained is exactly 2?

b) Referring to part (a), what is that maximum probability?

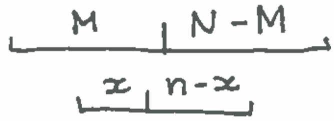
5.57 One rocket has two engines and another has four. All engines are identical. Each rocket will achieve its mission if and only if at least half its engines work. Both rockets have the same nonzero probability of achieving its mission. What is the probability than any particular engine works?

skip 6) Hypergeometric sampling without replacement

N = population size

p = proportion of population with some attribute

$M = Np$ = number of people in population with attribute $\left(p = \frac{M}{N}\right)$



n = sample size $\leq N$

Sample without replacement

X = # of people in sample with attribute

$R_X = \{0, 1, \dots, \min(n, M)\}$

$$\mathbb{P}_X(x) = \mathbb{P}(X = x) = \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}} = \frac{\binom{Np}{x}\binom{N(1-p)}{n-x}}{\binom{N}{n}}$$

Fact: As $N \to \infty$

$$\mathbb{P}_X(x) \to \binom{n}{x} p^x (1-p)^{n-x},$$

i.e. Hypergeometric becomes Binomial when $n \to \infty$.

* if $n < 5\% N$, good approx.

Ex. Firm receives package of $N = 100$ items

Select a sample of $n = 3$ items. if all are OK, package is accepted; if any one is not OK, rejected

Is this a good policy?

Suppose there are $M = 5$ bad items in the package.

X = # of bad items in sample

Policy: accept sample if $X = 0$

$$\mathbb{P}(X=0) = \frac{\binom{95}{3}\binom{5}{0}}{\binom{100}{3}} = \underline{0.856}$$

Compare with Binomial: p = proportion of bad items in population

$$p = \frac{M}{N} = \frac{5}{100} = 0.05$$

$$\mathbb{P}(X=0) = \binom{3}{0}(0.05)^0 \cdot (0.95)^3 = (0.95)^3 = \underline{0.8574}$$

not far from the true answer $n = 3 < 5\% N = 5$.

skip

Example 5.15 p. 212

Lotto 1-42

choose six of them = \$1 → ticket

without repetition → 6 winning # are drawn

To win a prize → three or more of the winning #'s

a) Determine PMF of X = # of winning numbers on ticket

$R_X = \{0, 1, \ldots, 6\}$

- distr. of X = hypergeometric (because no repetition)

$$\mathbb{P}(X = x) = \frac{\binom{6}{x}\binom{42-6}{6-x}}{\binom{42}{6}} \quad \text{for } x = \{0, 1, \ldots, 6\}$$

b) $\mathbb{P}(\text{winning ticket}) = \mathbb{P}(X \geqq 3) = \boxed{0.029}$

c) play once a week for a year

Let Y = # times you win in a year

$Y \sim \text{Binomial}(n = 52, p = 0.029)$

$$\mathbb{P}(Y \geqq 1) = 1 - \mathbb{P}(Y = 0) = 1 - (1-p)^n = 1 - (1-0.029)^{52} =$$

$$= \boxed{0.7948}$$

skip

Poisson distribution and rare events

Ex. X = # accidents in Kenmore Square in a month.

Claim X has Poisson distribution

0   each interval = 1 second   1 month

In one second
- no accident
- 1 accident $\rightarrow$ very small prob. p
- >1 accident $\rightarrow$ ≈ 0 prob.

Let n = total number of seconds in the month

$\Rightarrow X \sim \text{Bin}(n, p)$, but n very large, p very small

$\Rightarrow$ approx. by $\text{Poiss}(\lambda = np)$

esert Samaritan Hospital is Mesa, Arizona, keeps records of its emergency-room traffic. From those records, we find that the number of patients arriving between 6pm and 7pm ha Poisson distr. with par. $\lambda = 6.9$

Example 5.19   p. 223   Emergency room traffic

X = # of arrivals in ER from 6 pm to 7 pm

$X \sim \text{Poiss}(\lambda = 6.9)$, $\quad x = 0, 1, 2, \ldots$

$$\mathbb{P}(X = x) = \frac{e^{-6.9} (6.9)^x}{x!}$$

Determine the prob. that, the # patients arriving at the ER between 6pm and 7pm will be

exactly four:

a) $\mathbb{P}(X = 4) = \dfrac{e^{-6.9}(6.9)^4}{4!} = 0.095$

at least two

b) $\mathbb{P}(X \geqq 2) = 1 - \mathbb{P}(X \leqq 1) = 1 - e^{-6.9} \sum_{x=0}^{1} \dfrac{(6.9)^x}{x!} = 0.992$

between four and ten, inclusive.

c) $\mathbb{P}(4 \leqq X \leqq 10) = e^{-6.9} \sum_{x=4}^{10} \dfrac{(6.9)^x}{x!} = 0.821$

7) Poisson Distribution

$X \sim Poiss(\lambda)$, rate parameter $\lambda > 0$

$R_X = \{0, 1, \dots\}$

$$\mathbb{P}_X(x) = \mathbb{P}(X = x) = \frac{e^{-\lambda} \cdot \lambda^x}{x!}, \quad x = 0, 1, 2, \dots$$

Use
1. As an approx. to Binomial as $n \to \infty$, $p \to 0$.
2. To model rare occurences.

Poisson approximation to the Binomial (n, p)

Theorem: $$\lim_{\substack{p \to 0 \\ n \to \infty \\ p \cdot n \to \lambda}} \binom{n}{x} p^x q^{n-x} \longrightarrow e^{-\lambda} \cdot \frac{\lambda^x}{x!}$$

understanding

1) $\underbrace{p \cdot n}_{\text{mean of Binomial}} \to \underbrace{\lambda}_{\text{mean of Poisson}}$

2) $p \to 0$, $n \to \infty$, but n should not increase too fa
The speed of $p \to 0$ is as $\frac{sth}{n}$ (not $\frac{sth}{n^2}$),
and $n \to \infty$ with speed $\frac{sth}{p}$.

Ex. X = # typos in a book
n = # pages (very large)
p = prob. of a mistake on any given page (very smal

ex. $n = 10^5$
$p = 10^{-5}$

Ex. X = # of people within an age group (20-25) that will die next year

$n = 10^5$ (large population)

$p = 10^{-5}$ (unlikely to die)

$np = 1 = \lambda$ is neither too large nor too small

$\Rightarrow \mathbb{P}(X = x)$

$$= \binom{n}{x} p^x (1-p)^{n-x} \xrightarrow[\substack{n \to \infty \\ p \to 0 \\ np \to \lambda}]{} \frac{e^{-\lambda} \cdot \lambda^x}{x!} = \frac{e^{-1} \cdot 1^x}{x!} = \frac{1}{(e)(x!)}$$

$x = 0, 1, \ldots, n$

The maximum error in the approximation is $\boxed{np^2}$

$$np^2 = np \cdot p = \lambda p = \underline{10^{-5}}$$

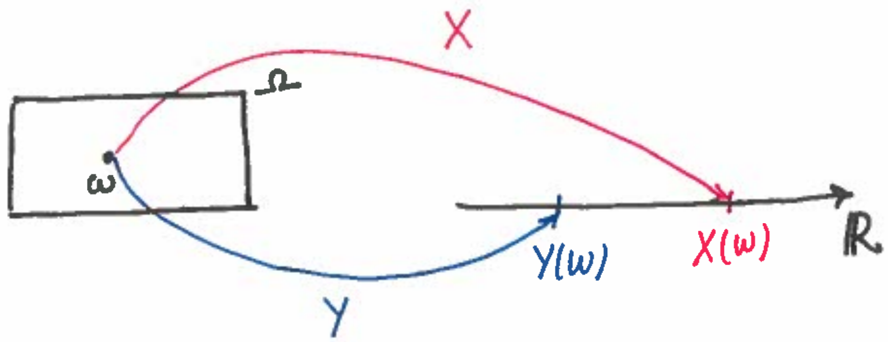
skip

<u>Proof of Theorem</u>: $\binom{n}{x} p^x (1-p)^{n-x} = \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x} =$

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

$$= \frac{\lambda^x}{x!} \left(\underbrace{\frac{n!}{(n-x)!} \cdot \left(\frac{1}{n}\right)^x}_{\underbrace{\frac{n(n-1)\ldots(n-x+1)}{n \cdot n \cdots n}}_{\underbrace{1\left(1-\frac{1}{n}\right)\cdots\left(1-\frac{x+1}{n}\right)}_{1 \text{ as } n \to \infty}}} \right) \underbrace{\left(1 - \frac{\lambda}{n}\right)^n}_{= e^{-\lambda} \text{ as } n \to \infty} \underbrace{\left(1 - \frac{\lambda}{n}\right)^{-x}}_{= 1 \text{ as } n \to \infty} \xrightarrow[\substack{n \to \infty \\ np \to \lambda}]{} \frac{\lambda^x \cdot e^{-\lambda}}{x!}$$

- <u>Jointly Discrete Random Variables</u>



We are interested in the pair (X,Y).

<u>Joint PMF</u>: $\mathbb{P}_{XY}(x,y) = \mathbb{P}(X=x, Y=y)$

and

- $\mathbb{P}_{XY}(x,y) \geq 0, \quad \forall x,y \in \mathbb{R}$
- $\sum_{x\in R_X} \sum_{y \in R_Y} \mathbb{P}_{XY}(x,y) = 1$

- same ideas as in one dimension

   <u>plus</u> two new concepts:

   1. Marginal Distribution
   2. Conditional Distribution

Ex. An urn contains three balls, numbered 1, 2, 3.

Draw two balls without replacement.

Let $X = \#$ of 1^{st} ball

$Y = \#$ of 2^{nd} ball

Find the PMF $\mathbb{P}_{XY}(x, y) \rightarrow$ Best to use a chart.

Y \ X	1	2	3	
1	0	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{3} = \mathbb{P}(Y=1)$
2	$\frac{1}{6}$	0	$\frac{1}{6}$	$\frac{1}{3} = \mathbb{P}(Y=2)$
3	$\frac{1}{6}$	$\frac{1}{6}$	0	$\frac{1}{3} = \mathbb{P}(Y=3)$
	$\frac{1}{3}$	$\frac{1}{3}$	$\frac{1}{3}$	1
	$= \mathbb{P}(X=1)$	$= \mathbb{P}(X=2)$	$= \mathbb{P}(X=3)$	

marginal distr. of Y $\downarrow$

marginal distr. of X $\rightarrow$

$\{X=1\} =$

$= \{X=1, Y=1\} \cup$

$\{X=1, Y=2\} \cup$

$\{X=1, Y=3\}$

Conditional Distribution of X given $Y=y$

$X | Y=y$

	1	2	3
$y=1$	0	$1/6 \div 1/3 = 1/2$	$1/2$
$y=2$	$1/2$	0	$1/2$
$y=3$	$1/2$	$1/2$	0

$$\mathbb{P}_{X|Y}(x|y) = \mathbb{P}(X=x \mid Y=y) = \frac{\mathbb{P}(X=x, Y=y)}{\mathbb{P}(Y=y)} \begin{array}{l} \leftarrow \text{joint} \\ \leftarrow \text{marginal} \end{array}$$

Marginal Distribution

$$\mathbb{P}_X(x) = \sum_y \mathbb{P}_{XY}(x,y)$$

$$\mathbb{P}_Y(y) = \sum_x \mathbb{P}_{XY}(x,y)$$

$$\mathbb{P}(X=x, Y=y) = \mathbb{P}(X=x \mid Y=y).\mathbb{P}(Y=y)$$

Def The random variables X and Y are independent if

$$\mathbb{P}(X\in A, Y\in B) = \mathbb{P}(X\in A).\mathbb{P}(Y\in B).$$

(If X and Y are independent, then the joint is the product of the marginals.)

Ex. X and Y are the $\underbrace{\text{lifetimes}}_{\text{measured in discrete time units}}$ of two electrical components

Let success = death of the electrical component, $p = \mathbb{P}(\text{success})$

$\Rightarrow$ $X \sim \text{Geom}(p)$ $\Rightarrow$ $\mathbb{P}(X=x) = p(1-p)^{x-1}$, $x = 1,2,\dots$

$Y \sim \text{Geom}(p)$ $\Rightarrow$ $\mathbb{P}(Y=y) = p(1-p)^{y-1}$, $y = 1,2,\dots$

Assume X and Y are independent.

a) Find the joint distr. of (X,Y)

$$\mathbb{P}_{XY}(x,y) = \mathbb{P}_X(x).\mathbb{P}_Y(y) = p^2(1-p)^{x+y-2}, \text{ for all } x,y = 1,2,\dots$$

b) Knowing the joint, recover the marginal for X

$$\mathbb{P}_X(x) = \sum_{y=1}^{\infty} \mathbb{P}_{XY}(x,y) = \sum_{y=1}^{\infty} p^2(1-p)^{x+y-2} = p^2(1-p)^{x-2}\underbrace{\sum_{y=1}^{\infty}(1-p)^{y}}_{\frac{1-p}{1-(1-p)} = \frac{1-p}{p}}$$

$$= p(1-p)^{x-1}, \quad x = 1,2,\dots$$

c) Find $\mathbb{P}(X>4, Y>4)$

Two methods:

<u>Method 1</u>: $\sum_{x=5}^{\infty} \sum_{y=5}^{\infty} \mathbb{P}_{XY}(x,y)$

<u>Method 2</u>: $\mathbb{P}(X>4, Y>4) = \mathbb{P}(X>4) \cdot \mathbb{P}(Y>4)$

$= (1-p)^4 \cdot (1-p)^4 \quad = (1-p)^8$

d) Find the conditional distr. of Y given $X=x$

<u>Method 1</u>: $\mathbb{P}_{Y|X}(y|x) = \dfrac{\mathbb{P}_{X,Y}(x,y)}{\mathbb{P}_X(x)} =$

$= \dfrac{p^2(1-p)^{x+y-2}}{p(1-p)^{x-1}} \quad = p(1-p)^{y-1}, \quad y = 1, 2, \dots$

<u>Method 2</u>: $\mathbb{P}_{Y|X}(y|x) = \underbrace{\mathbb{P}(Y=y \mid X=x)}_{Y \text{ and } X \text{ are indep.}} = \mathbb{P}(Y=y) =$

$= p(1-p)^{y-1}, \quad y = 1, 2, \dots$

skip

Binomial Distribution Revisited

n indep. trials

p = prob. of success

$q = 1 - p$ = prob. of failure

X = # of successes

Y = # of failures

$X + Y = n$

Joint PMF of X and Y

$$\mathbb{P}_{XY}(x,y) = \mathbb{P}(X = x, Y = y) \quad \text{where } x+y=n$$

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

$$= \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x} =$$

$$= \frac{n!}{x!\,y!} p^x \cdot q^y, \quad \underline{\text{where } x+y=n}$$

<u>Ex.</u> $n = 5$

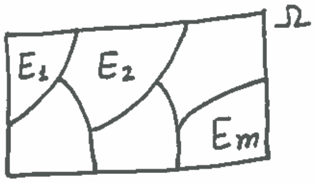
$$\mathbb{P}(X = 2, Y = 3) = \frac{5!}{2!\,3!} p^2 q^3$$

$$\mathbb{P}(X = 3, Y = 3) = 0.$$

skip

Multinomial Distribution

Let $E_1, E_2, \dots, E_m$ be a partition of Ω.



$$p_1 = \mathbb{P}(E_1)$$
$$p_2 = \mathbb{P}(E_2)$$
$$\vdots$$
$$p_m = \mathbb{P}(E_m)$$

$$p_1 + p_2 + \dots + p_m = 1$$

Repeat the experiment n times.

Let $X_1 = \#$ of times E_1 occurs

$\vdots$

$X_m = \#$ of times E_m occurs

$$X_1 + \dots + X_m = n$$

Find the PMF of $(X_1, X_2, \dots, X_m)$

$$\mathbb{P}_{X_1 \dots X_m}(x_1, \dots, x_m) = \mathbb{P}(X_1 = x_1, \dots, X_m = x_m)$$
$$x_1 + \dots + x_m = n$$

$$= \underbrace{\frac{n!}{x_1! \dots x_m!}}_{\binom{n}{x_1 \dots x_m}} p_1^{x_1} p_2^{x_2} \cdots p_m^{x_m}, \quad \begin{array}{l} x_1 + \dots + x_m = n \\ x_1, \dots, x_m \geqq 0 \end{array}$$

multinomial coefficient

skip

Example 6.9 p. 276 Roulette

38 numbers $\rightarrow$ 18 red, 18 black, 2 green

when the roulette is spun, the ball is equally likely to land on any of the 38 numbers

$p_1 = \mathbb{P}(R) = \frac{18}{38}$

$p_2 = \mathbb{P}(B) = \frac{18}{38}$

$p_3 = \mathbb{P}(G) = \frac{2}{38}$

Play $n = 10$ times

X_i = # of times R / B / G occur, $i=1$ / $i=2$ / $i=3$

a) $\mathbb{P}(X_1 = 3, X_2 = 6, X_3 = 5) = 0$

b) $\mathbb{P}(X_1 = 3, X_2 = 6, X_3 = 1) = \frac{10!}{3!\,6!\,1!}\left(\frac{18}{38}\right)^3\left(\frac{18}{38}\right)^6\left(\frac{2}{38}\right)^1$

c) Suppose that getting a green is a success,
not getting a green is a failure

$\mathbb{P}(X_3 = 1) = \binom{10}{1}\left(\frac{2}{38}\right)^1\left(\frac{36}{38}\right)^9$

skip

Multinomial has replacement.

Multivariate Hypergeometric Distribution

sampling without replacement

Population of size N

M_1 of type 1 $\qquad M_1 + \dots + M_m = N$

$\vdots$

M_m of type m

n = sample size

X_j = # in sample of type j, $j = 1, \dots, m$

$$\mathbb{P}_{X_1 \dots X_m}(x_1, \dots, x_m) = \mathbb{P}(X_1 = x_1, \dots, X_m = x_m) =$$

$$\boxed{x_1 + \dots + x_m = n}$$

$$= \frac{\binom{M_1}{x_1}\binom{M_2}{x_2}\cdots\binom{M_m}{x_m}}{\binom{N}{n}} = \qquad p_j = \frac{M_j}{N}, \ j = 1, \dots, m$$

$$\Rightarrow M_j = p_j N$$

$$= \frac{\binom{Np_1}{x_1}\binom{Np_2}{x_2}\cdots\binom{Np_m}{x_m}}{\binom{N}{n}} \xrightarrow[n \to \text{small}]{N \to \infty} \frac{n!}{x_1! \dots x_m!} p_1^{x_1} \cdots p_m^{x_m}$$

Functions of a Discrete Random Variable

$$X = \begin{cases} 1 & \text{with prob. } 1/3 \\ 0 & \text{with prob. } 1/3 \\ -1 & \text{with prob. } 1/3 \end{cases}.$$

Interested in U (a function of X).

$$U = X^2 = \begin{cases} 0 & \text{with prob. } 1/3 \\ 1 & \text{with prob. } 2/3 \end{cases}$$

$$\mathbb{P}(U = 0) = \mathbb{P}(X = 0) = 1/3$$

$$\mathbb{P}(U = 1) = \mathbb{P}(\underbrace{X = \pm 1}_{\text{two disjoint events}}) = 2/3$$

Functions of Several Discrete Random Variables

Let $(X, Y) \sim \mathbb{P}_{XY}(x, y)$

We are interested in the distribution of Z:

$$Z = X + Y$$

$$\mathbb{P}_Z(z) = \mathbb{P}(Z = z) = \sum_{\substack{(x,y) \\ \text{such that} \\ x+y=z}} \mathbb{P}_{XY}(x, y)$$

Example 6.22   p. 302, 303

$Z = X+Y$

find $\mathbb{P}_Z(z)$

$X \sim \text{Geom}(p)$, $Y \sim \text{Geom}(p)$ independent

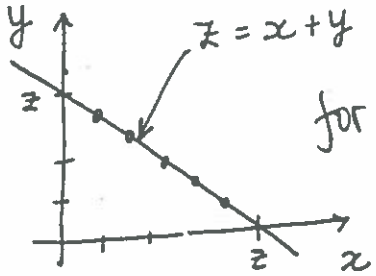
$$\mathbb{P}_{XY}(x,y) = \mathbb{P}(X=x, Y=y) \overset{\text{by indep.}}{=} \mathbb{P}(X=x)\cdot\mathbb{P}(Y=y) =$$

$$= (1-p)^{x-1}\cdot p \cdot (1-p)^{y-1}\cdot p =$$

$$= p^2(1-p)^{x+y-2}, \quad \text{for } x,y = 1,2,\ldots$$

$Z = X+Y$

$R_Z = 2,3,\ldots$



for a fixed z

$$\mathbb{P}_Z(z) = \mathbb{P}(Z=z) = \mathbb{P}(X+Y=z)$$

$$\mathbb{P}_{X+Y}(z) = \sum_{\substack{\text{all } (x,y) \\ \text{such that} \\ x+y=z}}\sum p^2(1-p)^{x+y-2} = \sum_{\substack{x=1 \\ y=z-x}}^{z-1} p^2(1-p)^{\cancel{x}+z-\cancel{x}-2} =$$

$$= p^2(1-p)^{z-2}\sum_{x=1}^{z-1} 1 = (z-1)(1-p)^{z-2}p^2 =$$

$$= \binom{z-1}{2-1}p^2(1-p)^{z-2}, \quad z = 2,3,\ldots$$

Negative Binomial

Pascal $(p, r=2)$

number of trials until the 2^{nd} succe

Convolution

↳ sum of two <u>indep.</u> random variables

$$\left.\begin{array}{l} X \sim \mathbb{P}_X(x) \\ Y \sim \mathbb{P}_Y(y) \end{array}\right\rangle \text{indep.}$$

$$Z = X + Y$$

$$\mathbb{P}_{X+Y}(z) = \sum_{\substack{(x,y) \text{ s.t.} \\ x+y=z}} \mathbb{P}_X(x) \cdot \mathbb{P}_Y(y) = \sum_x \mathbb{P}_X(x)\mathbb{P}_Y(z-x)$$

$$\sum_{\substack{y=z-x \\ z \in \{0,\dots,n+m\}}} \binom{n}{x} p^x (1-p)^{n-x} \cdot \binom{m}{y} p^y (1-p)^{m-y} =$$

$$= \binom{n}{x}\binom{m}{z-x} \cdot p^{x+z-x} \cdot (1-p)^{n-x+}$$

$$= \underbrace{\binom{n}{x}\binom{m}{z-x}}_{\binom{n+m}{z}} \cdot p^z \cdot (1-p)^{n+}$$

$$\binom{n+m}{z} p^z (1-p)^{n+m}$$

$$z = \{0, \dots, n+m$$

In previous example:

$$\mathbb{P}_{X+Y}(z) = \sum_x \underbrace{p(1-p)^{x-1} \cdot p(1-p)^{z-x-1}}_{p^2(1-p)^{z-2}}$$

<u>Ex.</u> $\left.\begin{array}{l} X \sim \text{Bin}(m,p) \\ Y \sim \text{Bin}(n,p) \end{array}\right\rangle \text{indep.}$ $\quad Z = X+Y \sim \text{Bin}(m+n, p)$

<u>Ex.</u> $\left.\begin{array}{l} X \sim \text{Geom}(p) \\ Y \sim \text{Geom}(p) \end{array}\right\rangle \text{indep.}$ $\quad Z = \min(X, Y)$ $\quad R_Z = 1, 2, \dots$ <u>p. 303</u>

$$Z > z \iff \min(X,Y) > z \iff X > z \text{ and } Y > z$$

$$\mathbb{P}(Z > z) = \mathbb{P}(X > z, Y > z) = \mathbb{P}(X > z) \cdot \mathbb{P}(Y > z) =$$

$$= (1-p)^z \cdot (1-p)^z = (1 - 2p + p^2)^z =$$

$$= (1 - (2p - p^2))^z, \quad z = 1, 2, \dots$$

$$\sim \text{Geom}(2p - p^2)$$

6.131 A point (X, Y) is randomly chosen from the unit square, $[0, 1] \times [0, 1]$. Let

$$U_{0.6} = \begin{cases} 0, & \text{if neither } X \text{ nor } Y \text{ is less than } 0.6; \\ 1, & \text{if either } X \text{ or } Y\text{, but not both, is less than } 0.6; \\ 2, & \text{if both } X \text{ and } Y \text{ are less than } 0.6. \end{cases}$$

Define $U_{0.3}$ similarly by putting "0.3" in place of "0.6." Obtain the
a) PMF of $U_{0.6}$.
b) PMF of $U_{0.3}$.
c) joint PMF of $U_{0.3}$ and $U_{0.6}$.
d) conditional PMF of $U_{0.3}$ given $U_{0.6} = v$.
e) conditional PMF of $U_{0.6}$ given $U_{0.3} = u$.

6.132 Refer to the discussion of the multiple hypergeometric distribution in Exercise 6.31 on page 280. In a group of 100 voters, there are 50 Republicans, 30 Democrats, and 20 Independents. A committee of 10 people is randomly selected without replacement from the group to represent it at a state political convention.
a) What is the probability that the committee will be perfectly representative in the sense that the proportion of people on the committee in each political party is the same as that of the group as a whole?
b) What is the probability that the committee will be close to perfectly representative in the sense that the number of people on the committee in each political party is within 1 of that which would be perfect representation?

• **6.133** Refer to the discussion of the multiple hypergeometric distribution in Exercise 6.31 on page 280. The candy N&Ns is produced in seven colors: red, orange, yellow, green, blue, violet, and brown. A package contains 70 N&Ns: 10 red, 7 orange, 12 yellow, 15 green, 10 blue, 8 violet, and 8 brown. You randomly take 7 N&Ns from the package, without replacement. What is the probability that you get
a) all seven colors? b) only one color?
c) 2 red, 2 yellow, 2 green, and 1 blue?

6.134 During a weekday, the number of accidents at a dangerous intersection is a Poisson random variable with parameter $\lambda = 2.4$, and each accident results in serious damage to at least one vehicle with probability $p = 0.6$. On weekends, the situation is the same except that $\lambda = 1.3$ and $p = 0.4$. What is the distribution of the total number of accidents in a week that result in serious damage to at least one vehicle?

✓ **6.135** Let X and Y be discrete random variables with joint PMF

$$p_{X,Y}(x, y) = \frac{x^y e^{-x}}{y!M}, \qquad x = 1, 2, \dots, M, \quad y = 0, 1, \dots,$$

and $p_{X,Y}(x, y) = 0$ otherwise.
a) Obtain and identify the marginal PMF of X.
b) Obtain the marginal PMF of Y.
c) Obtain and identify the conditional PMF of Y given $X = x$.
d) Are X and Y independent random variables? Explain your answer.

• **6.136** In a factory, each manufactured item is defective with probability p and, independently, inspected with probability q. If both defective status and inspection status are taken into account, there are four categories: nondefective-uninspected, nondefective-inspected, defective-uninspected, and defective-inspected. Determine the joint PMF of the numbers of items in the four categories for a random sample of n items.

6.131 A point (X, Y) is randomly chosen from the unit square, $[0, 1] \times [0, 1]$. Let

$$U_{0.6} = \begin{cases} 0, & \text{if neither } X \text{ nor } Y \text{ is less than } 0.6; \\ 1, & \text{if either } X \text{ or } Y \text{, but not both, is less than } 0.6; \\ 2, & \text{if both } X \text{ and } Y \text{ are less than } 0.6. \end{cases}$$

Define $U_{0.3}$ similarly by putting "0.3" in place of "0.6." Obtain the

a) PMF of $U_{0.6}$.
b) PMF of $U_{0.3}$.
c) joint PMF of $U_{0.3}$ and $U_{0.6}$.
d) conditional PMF of $U_{0.3}$ given $U_{0.6} = v$.
e) conditional PMF of $U_{0.6}$ given $U_{0.3} = u$.

6.132 Refer to the discussion of the multiple hypergeometric distribution in Exercise 6.31 on page 280. In a group of 100 voters, there are 50 Republicans, 30 Democrats, and 20 Independents. A committee of 10 people is randomly selected without replacement from the group to represent it at a state political convention.

a) What is the probability that the committee will be perfectly representative in the sense that the proportion of people on the committee in each political party is the same as that of the group as a whole?
b) What is the probability that the committee will be close to perfectly representative in the sense that the number of people on the committee in each political party is within 1 of that which would be perfect representation?

X **6.133** Refer to the discussion of the multiple hypergeometric distribution in Exercise 6.31 on page 280. The candy N&Ns is produced in seven colors: red, orange, yellow, green, blue, violet, and brown. A package contains 70 N&Ns: 10 red, 7 orange, 12 yellow, 15 green, 10 blue, 8 violet, and 8 brown. You randomly take 7 N&Ns from the package, without replacement. What is the probability that you get

a) all seven colors? b) only one color?
c) 2 red, 2 yellow, 2 green, and 1 blue?

6.134 During a weekday, the number of accidents at a dangerous intersection is a Poisson random variable with parameter $\lambda = 2.4$, and each accident results in serious damage to at least one vehicle with probability $p = 0.6$. On weekends, the situation is the same except that $\lambda = 1.3$ and $p = 0.4$. What is the distribution of the total number of accidents in a week that result in serious damage to at least one vehicle?

• **6.135** Let X and Y be discrete random variables with joint PMF

$$p_{X,Y}(x, y) = \frac{x^y e^{-x}}{y!M}, \qquad x = 1, 2, \ldots, M, \quad y = 0, 1, \ldots,$$

and $p_{X,Y}(x, y) = 0$ otherwise.

a) Obtain and identify the marginal PMF of X.
b) Obtain the marginal PMF of Y.
c) Obtain and identify the conditional PMF of Y given $X = x$.
d) Are X and Y independent random variables? Explain your answer.

6.136 In a factory, each manufactured item is defective with probability p and, independently, inspected with probability q. If both defective status and inspection status are taken into account, there are four categories: nondefective-uninspected, nondefective-inspected, defective-uninspected, and defective-inspected. Determine the joint PMF of the numbers of items in the four categories for a random sample of n items.

skip

The Multinomial Distribution

One of the most important multivariate distributions is the *multinomial distribution.* Before discussing this distribution, we briefly review the binomial distribution, which we examined in Chapter 5.

The binomial distribution gives the probability distribution of the number of successes in a finite sequence of Bernoulli trials. Recall that, in Bernoulli trials, we have independent repetitions (trials) of a random experiment in which the result of each trial is classified as either a success or a failure, depending on whether or not a specified event occurs; the success probability (i.e., the probability that the specified event occurs on any particular trial) is assumed to remain the same from trial to trial and is denoted p.

The number of successes, X, in n Bernoulli trials has PMF

$$p_X(x) = \binom{n}{x} p^x (1-p)^{n-x}, \qquad x = 0, 1, \ldots, n, \tag{6.9}$$

and $p_X(x) = 0$ otherwise. The random variable X is called a binomial random variable and is said to have the binomial distribution with parameters n and p. The "bi" in "binomial" reflects the fact that the result of each trial is classified as one of two mutually exclusive possibilities—success (the specified event occurs) or failure (the specified event doesn't occur).

The binomial setting can also be viewed in terms of a bivariate distribution. Indeed, in n Bernoulli trials, let X denote the number of successes and let Y denote the number of failures. Recalling that the failure probability is $q = 1 - p$, the joint PMF of X and Y is

$$p_{X,Y}(x, y) = \binom{n}{x,\ y} p^x q^y, \tag{6.10}$$

if x and y are nonnegative integers whose sum is n, and $p_{X,Y}(x, y) = 0$ otherwise. Because $\binom{n}{x} = \binom{n}{x,\ n-x}$, the right side of Equation (6.10) can be obtained by simply replacing $1 - p$ by q and $n - x$ by y in the right side of Equation (6.9).

We can generalize the binomial distribution by considering repeated independent trials of a random experiment in which the result of each trial is classified as one of r mutually exclusive possibilities. Example 6.9 introduces this idea.

EXAMPLE 6.9 ***Introduces the Multinomial Distribution***

Roulette A U.S. roulette wheel contains 38 numbers of which 18 are red, 18 are black, and 2 are green. When the roulette wheel is spun, the ball is equally likely to land on any of the 38 numbers. Here the random experiment consists of a play at the roulette wheel; each play constitutes one trial and the trials are independent.

To analyze this random experiment with regard to the color of the number on which the roulette ball lands, the result of each trial is classified as one of three mutually exclusive possibilities: red, black, or green. The probabilities of these three events are 18/38, 18/38, and 2/38, respectively. In 10 plays at the roulette wheel, let X_1, X_2, and X_3 denote the number of times that red, black, and green, respectively, occur.

a) Determine $P(X_1 = 3,\ X_2 = 6,\ X_3 = 1)$.

b) Determine the joint PMF of X_1, X_2, and X_3.

PMF of a Function of Two Discrete Random Variables

Skip

Several techniques are available for determining the PMF of a function of two (or more) discrete random variables. The most straightforward method, although not necessarily the easiest, is presented in Proposition 6.16.

◆◆◆ **Proposition 6.16 PMF of a Function of Two Discrete Random Variables**

Let X and Y be two discrete random variables defined on the same sample space and let g be a real-valued function of two variables defined on the range of (X, Y). Then the PMF of the random variable $Z = g(X, Y)$ is

$$p_Z(z) = \sum\sum_{(x,y)\in g^{-1}(\{z\})} p_{X,Y}(x, y), \tag{6.26}$$

for z in the range of Z, and $p_Z(z) = 0$ otherwise. In words, if z is in the range of Z, we obtain the probability that $Z = z$—that is, the probability that $g(X, Y) = z$—by summing the joint PMF of X and Y over all points (x, y) in the plane such that $g(x, y) = z$.

Proof Let z be in the range of Z. From the FPF for two discrete random variables,

$$\begin{aligned} p_Z(z) &= P(Z = z) = P\big(g(X, Y) = z\big) \\ &= P\big((X, Y) \in g^{-1}(\{z\})\big) = \sum\sum_{(x,y)\in g^{-1}(\{z\})} p_{X,Y}(x, y), \end{aligned}$$

as required. ◆

Note: We can express Equation (6.26) in the alternate form

$$p_{g(X,Y)}(z) = \sum\sum_{g(x,y)=z} p_{X,Y}(x, y), \tag{6.27}$$

where $\sum\sum_{g(x,y)=z}$ indicates that the double sum is taken over all x and y such that $g(x, y) = z$.

EXAMPLE 6.22 ***PMF of a Function of Two Discrete Random Variables***

Let X and Y be independent random variables, both having the geometric distribution with parameter p. Determine and identify the PMF of the random variable
a) $X + Y$. b) $\min\{X, Y\}$.

Solution We apply Proposition 6.16. In doing so, we first need to obtain the joint PMF of X and Y. By assumption, both X and Y have the geometric distribution with parameter p. Therefore,

$$p_X(x) = \begin{cases} p(1-p)^{x-1}, & \text{if } x \in \mathcal{N}; \\ 0, & \text{otherwise.} \end{cases} \qquad p_Y(y) = \begin{cases} p(1-p)^{y-1}, & \text{if } y \in \mathcal{N}; \\ 0, & \text{otherwise.} \end{cases}$$

HW 4

✓ 1) 5.43 a), b) p. 206

✓ 2) 5.45 p. 206

✓ 3) 5.52 p. 207

✓ ~~4~~) 5.63 p. 217 hypergeometric

✓ 5) 5.94 p. 235 geometric (first hit)

✓ 6) 5.117 p. 244 negative binomial (second hit)

Extra credit:

x 1) 5.50 p. 206 seq. of random decimal digits

x 2) 5.95 p. 235 a), b), c)

5.61 Five cards are selected at random without replacement from an ordinary deck of 52 playing cards.

a) What is the probability that exactly 3 face cards are obtained?

b) Identify and provide a formula for the probability distribution of the number of face cards obtained.

5.62 Estimating a population proportion: Many statistical studies are concerned with the proportion of members of a finite population that have a specified attribute, called the *population proportion*. In practice, we mostly rely on sampling and use the sample data to estimate the population proportion. Suppose that a random sample of size n is taken without replacement from a population of size N in which the proportion of members having the specified attribute is p. Intuitively, it makes sense to estimate the population proportion, p, by the *sample proportion*, $\hat{p} = X/n$, where X denotes the number of members sampled that have the specified attribute. Determine the PMF of the random variable $\hat{p}$.

X **5.63** From 10 pills, of which 5 are placebos, you are to randomly select and take 5 as part of an experiment on the effectiveness of a new treatment. Find the probability that

a) you select at least 2 placebos.

b) the first three pills you select are placebos.

5.64 As reported by Television Bureau of Advertising, Inc., in *Trends in Television*, 84.2% of U.S. households have a VCR. If six U.S. households are randomly selected without replacement, what is the (approximate) probability that the number of households sampled that have a VCR will be

a) exactly four? b) at least four? c) at most four?

d) between two and five, inclusive?

e) Determine a formula that approximates the PMF of the random variable Y, the number of households of the six sampled that have a VCR.

f) Strictly speaking, why is the PMF that you obtained in part (e) only approximately correct?

g) Determine a formula that provides the exact PMF of the random variable Y in terms of the number, N, of U.S. households.

5.65 Bin I contains 20 parts, of which 5 are defective. Bin II contains 15 parts, of which 4 are defective. One of these two bins is chosen at random and 3 parts are randomly selected from the bin chosen. If 2 of the 3 parts obtained are defective, what is the probability that Bin I was chosen?

5.66 An upper-level probability class has six undergraduate students and four graduate students. A random sample of three students is taken from the class. Let X denote the number of undergraduate students selected. Identify, obtain a formula for, and tabulate the PMF of the random variable X if the sampling is

a) without replacement. b) with replacement.

c) Compare your results in parts (a) and (b).

5.67 Refer to Example 5.14 on page 208, where X denotes the number of defective TVs obtained when 5 TVs are randomly selected without replacement from a lot of 100 TVs in which 6 are defective.

a) Construct a table for the PMF of the random variable X similar to Table 5.14 on page 213. Round each probability to eight decimal places.

b) Approximate the PMF by the appropriate binomial distribution. Construct a table similar to Table 5.15 on page 215.

c) Comment on the accuracy here of the binomial approximation to the hypergeometric distribution.

HW 5

1) 5.80 p.226 (poisson approx. to the Binomial)
2) 5.82 p.226 (poisson)
? ✓ 3) 6.2 p.269 (joint)
x ~~4)~~ 6.45 p.288 (joint)
✓ 5) 6.135 p.320 (joint)
x * ~~6)~~ 6.136 p.320 (multinomial)

Extra credit:

x 1) 6.3 p.269 (joint)
x 2) 6.46 p.288 (joint)

Let's apply Proposition 5.8 to the Poisson distribution shown in Figure 5.6. In this case, $\lambda = 6.9$ and, so, according to Proposition 5.8, the probabilities increase until $x = \lfloor 6.9 \rfloor = 6$ and then decrease thereafter. This fact is borne out by both Table 5.17 and Figure 5.6.

EXERCISES 5.5 Basic Exercises

5.77 Suppose that X has the Poisson distribution with parameter $\lambda = 3$. Determine
a) $P(X = 3)$. **b)** $P(X < 3)$. **c)** $P(X > 3)$. **d)** $P(X \leq 3)$. **e)** $P(X \geq 3)$.

5.78 A paper by L. F. Richardson, published in the *Journal of the Royal Statistical Society,* analyzed the distribution of wars over time. The data indicate that the number of wars that begin during a given calendar year has approximately the Poisson distribution with parameter $\lambda = 0.7$. If a calendar year is selected at random, find the probability that the number of wars that begin during that calendar year will be
a) zero. **b)** at most two. **c)** between one and three, inclusive.

5.79 M. F. Driscoll and N. A. Weiss discussed the modeling of motel reservation networks in "An Application of Queuing Theory to Reservation Networks" (*TIMS,* 1976, Vol. 22, pp. 540–546). They defined a Type 1 call to be a call from a motel's computer terminal to the national reservation center. For a certain motel, the number of Type 1 calls per hour has a Poisson distribution with parameter $\lambda = 1.7$. Find the probability that the number of Type 1 calls made from this motel during a period of 1 hour will be
a) exactly one. **b)** at most two. **c)** at least two.

Let X denote the number of Type 1 calls made by the motel during a 1-hour period.

d) Construct a table of probabilities for the random variable X. Compute the probabilities until they are zero to three decimal places.

e) Draw a histogram of the probabilities in part (d).

X✓ **5.80** The second leading genetic cause of mental retardation is Fragile X Syndrome, named for the fragile appearance of the tip of the X chromosome in affected individuals. One in 1500 males are affected worldwide, with no ethnic bias. For a sample of 10,000 males, use the Poisson approximation to the binomial distribution to determine the probability that the number who have Fragile X Syndrome
a) exceeds 7. **b)** is at most 10.

c) What is the maximum possible error that you made in parts (a) and (b) by using the Poisson approximation to the binomial distribution?

5.81 A most amazing event occurred during the second round of the 1989 U.S. Open at Oak Hill in Pittsford, New York. Four golfers—Doug Weaver, Mark Wiebe, Jerry Pate, and Nick Price—made holes in one on the sixth hole. According to the experts, the odds against a PGA golfer making a hole in one are 3708 to 1—that is, the probability of making a hole in one is 1/3709. Determine the probability to nine decimal places that at least four of the 155 golfers playing the second round would get a hole in one on the sixth hole by using
a) the binomial distribution. **b)** the Poisson approximation to the binomial.

c) What assumptions are you making in obtaining your answers in parts (a) and (b)? Do you think that those assumptions are reasonable? Explain your reasoning.

5.82 At a service counter, arrivals occur at an average rate of 20 per hour and follow a Poisson distribution. Let X denote the number of arrivals during a particular hour.

• **6.2** In Example 5.1 on page 176, we considered the number of siblings for each of 40 students in one of Professor Weiss's classes. The following contingency table cross classifies the number of siblings and number of sisters of the 40 students.

Siblings, x \ Sisters, y	0	1	2	3	Total
0	8	0	0	0	8
1	8	9	0	0	17
2	1	8	2	0	11
3	0	1	2	0	3
4	0	0	0	1	1
Total	17	18	4	1	40

Suppose that one of the 40 students is selected at random. Let X and Y denote the number of siblings and number of sisters, respectively, of the student obtained.

a) Determine $p_{X,Y}(2, 1)$ and interpret your answer.
b) Obtain the joint PMF of the random variables X and Y.
c) Construct a table similar to Table 6.2 on page 262.
d) Explain why the marginal PMF of X obtained in part (c) is the same as the PMF in Table 5.5 on page 186.
e) Use random variable notation to express the event that the student obtained has no brothers and then determine the probability of that event.
f) Use random variable notation to express the event that the student obtained has no sisters and then determine the probability of that event.

• **6.3** Let S consist of the 10 decimal digits. Suppose that a number X is chosen according to the discrete uniform distribution on S and then a number Y is chosen according to the discrete uniform distribution on S with X removed.

a) Obtain the joint PMF of X and Y. *Hint:* Use the general multiplication rule.
b) Determine $P(X = Y)$ first by using the FPF and then without doing any computations.
c) Determine $P(X > Y)$ first by using the FPF and then without doing any computations.
d) Find the marginal PMF of Y by using Proposition 6.2 on page 263.
e) Find the marginal PMF of Y by using a symmetry argument.

6.4 Let E and F be events of a sample space.

a) Determine the joint and marginal PMFs of I_E and I_F. Construct a table similar to Table 6.2.
b) You can apply Proposition 6.2 on page 263 to conclude that the sum of the values of this joint PMF in a row or column equals the value of the marginal PMF at the end (right or bottom) of that row or column. However, in this case, you can apply another result to get that conclusion. What is that result?
c) On page 264, we noted that, in general, the marginal PMFs of two discrete random variables don't determine the joint PMF. What does that mean in the context of indicator random variables?
d) Under what condition is the joint PMF of two indicator random variables determined by the marginal PMFs?

EXERCISES 6.3 Basic Exercises

Note: Several of the exercises in this section are continuations of exercises presented in Sections 6.1 and 6.2.

6.44 Refer to Example 6.12 on page 283.
a) Obtain the conditional PMF of the number of bedrooms for each number of bathrooms. Construct a table similar to Table 6.6 on page 284.
b) Given that a randomly selected home (from among the 50 homes) has two bathrooms, what is the probability that it has at least three bedrooms?
c) Interpret your answer in part (b) in terms of percentages.

✓ •**6.45** In Exercise 6.2 on page 269, you considered the number of siblings, X, and the number of sisters, Y, for a randomly selected student from one of Professor Weiss's classes.
a) For each number of siblings, determine and interpret the conditional PMF of the number of sisters.
b) Construct a table like Table 6.6 on page 284 that displays the marginal PMF of the number of sisters and the conditional PMF of the number of sisters for each number of siblings.
c) What percentage of students in the class with two siblings have at least one sister?
d) For each number of sisters, determine and interpret the conditional PMF of the number of siblings.
e) Construct a table that displays the marginal PMF of the number of siblings and the conditional PMF of the number of siblings for each number of sisters.
f) What percentage of students in the class with one sister have at least two siblings?

• **6.46** Let X be a number chosen according to the discrete uniform distribution on the set S of 10 decimal digits and let Y be a number chosen according to the discrete uniform distribution on S with X removed.
a) Determine and identify the conditional PMF of Y given $X = x$.
b) Determine and identify the conditional PMF of X given $Y = y$.
c) Find $P(3 \leq X \leq 4 \mid Y = 2)$.

6.47 Let E and F be events of a sample space.
a) Determine the conditional PMF of I_F given $I_E = x$ for each possible value x of I_E. Construct a table similar to Table 6.6 on page 284 that displays these conditional PMFs and the marginal PMF of I_F.
b) In this special case of indicator random variables, what result guarantees that the values in each row of your table in part (a) sum to 1?

6.48 Let X and Y denote the smaller and larger of the two faces, respectively, when a balanced die is tossed twice.
a) Determine each conditional PMF of Y given $X = x$ and construct a table similar to Table 6.6 on page 284 that displays these conditional PMFs and the marginal PMF of Y.
b) Repeat part (a) for the conditional PMFs of X given $Y = y$ and the marginal PMF of X.

6.49 Consider a sequence of Bernoulli trials with success probability p. Let X denote the number of trials up to and including the first success and let Y denote the number of trials up to and including the second success.
a) Determine the conditional PMF of Y given $X = x$ without doing any computations. Explain your reasoning.
b) Determine the conditional PMF of Y given $X = x$ by referring to Exercise 6.9 on page 270 and applying Definition 6.3 on page 283. Compare your result with that in part (a).

HW6

✓ p. 320 ~~p~~ 6.133

✓ p. 320 6.136

Functions of random variables

The problem: $X \sim F_X(x)$ known

$Y = g(X)$

Y is a function of X

Find $\mathbb{F}_Y(y)$ and/or $f_Y(y)$.

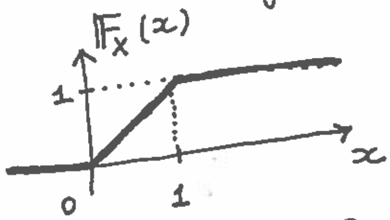
2 methods:

1) The CDF method

Ex. $X \sim \text{Unif}(0,1)$

$Y = 3X + 2$

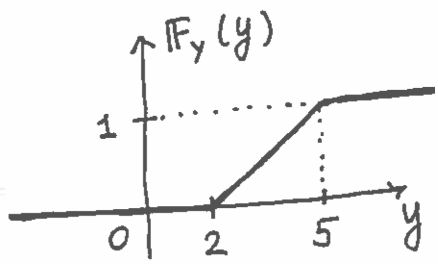
Recall: $\mathbb{F}_X(x) = \begin{cases} 0 & \text{if } x < 0 \\ x & \text{if } x \in [0,1] \\ 1 & \text{if } x > 1 \end{cases}$



$$\mathbb{F}_Y(y) = \mathbb{P}(Y \leq y) = \mathbb{P}(3X + 2 \leq y) =$$

$$= \mathbb{P}\left(X \leq \frac{y-2}{3}\right) = \begin{cases} 0 & \text{, if } \frac{y-2}{3} < 0 \text{, i.e.} \\ \frac{y-2}{3} & \text{, if } \frac{y-2}{3} \in [0,1] \quad y \in [2, \\ 1 & \text{, if } \frac{y-2}{3} > 1, \ y > \end{cases}$$

So, $\mathbb{F}_Y(y) = \begin{cases} 0 & \text{if } y < 2 \\ \frac{y-2}{3} & \text{if } 2 \leq y \leq 5 \\ 1 & \text{if } y > 5 \end{cases}$



$\Rightarrow Y \sim \text{Unif}(2,5)$

Q: $\sqrt{4} = ?$

Q: Find all numbers whose square is 4?

A: $-2, 2$

Q: Find all numbers whose square is π?

A: $-\sqrt{\pi}, \sqrt{\pi}$

So, $\sqrt{x} \geq 0$

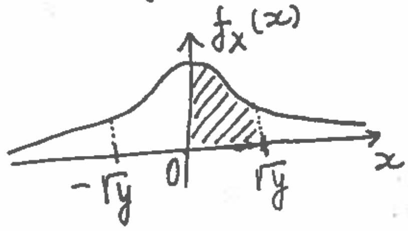
Ex. $X \sim N(0, 1)$

$Y = X^2$

$$\mathbb{F}_Y(y) = \mathbb{P}(Y \leq y) = \underbrace{\mathbb{P}(X^2 \leq y)}_{\text{this is 0 if } y<0, \text{ so assume } y \geq 0,}$$

then $\mathbb{P}(X^2 \leq y) = \mathbb{P}(-\sqrt{y} \leq X \leq \sqrt{y}) = 2\,\mathbb{P}(0 \leq X \leq \sqrt{y}) =$

$$= 2\int_0^{\sqrt{y}} \frac{1}{\sqrt{2\pi}} \cdot e^{-\frac{x^2}{2}} dx =$$



let $x = \sqrt{u}$

$\Rightarrow u = x^2$

$x = \sqrt{y} : u = y$

$x = 0 : u = 0$

$du = 2x\,dx$

$\Rightarrow dx = \frac{1}{2x} du = \frac{1}{2\sqrt{u}} du$

$$= \int_0^y \frac{1}{\sqrt{2\pi u}} e^{-\frac{u}{2}} du$$

$$\Rightarrow \underline{\text{CDF}} \text{ is } \mathbb{F}_Y(y) = \begin{cases} \int_0^y \frac{1}{\sqrt{2\pi u}} e^{-\frac{u}{2}} du, & y \geq 0 \\ 0 & \text{, otherwise} \end{cases}$$

Note: ~~This is Gamma($\frac{1}{2}$, 1)~~, also called χ_1^2 distribution.

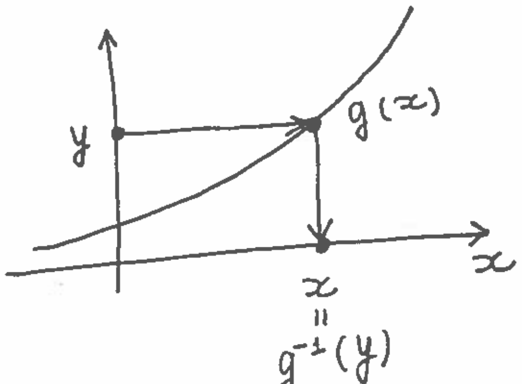
chi-squared with 1 df

2) The PDF method

Let $X \sim f_X(x)$

$Y = g(X)$, monotone (i.e. either $\nearrow$ or $\searrow$ and differentiabl on the R_X, i.e. where $f_X(x) \neq 0$

Then
$$\boxed{f_Y(y) = f_X(x) \cdot \left| \frac{dx}{dy} \right|, \quad \text{where } x = g^{-1}(y)} \qquad (*)$$



Ex. $y = x^3 \quad (y = g(x) = x^3)$

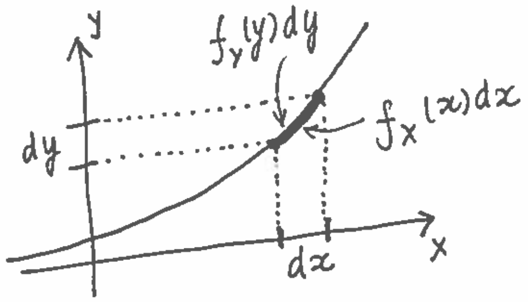
$\Rightarrow x = y^{1/3} \quad (x = g^{-1}(y) = y^{1/3})$

Q: $g(g^{-1}(y)) \stackrel{?}{=} y$

Note: If $y = g(x)$, $\frac{dy}{dx} = g'(x) \Rightarrow \frac{dx}{dy} = \frac{1}{g'(x)}$

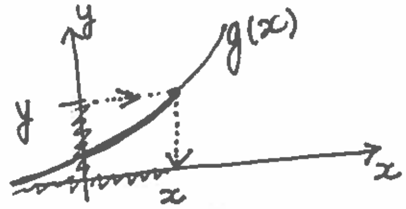
So, $(*)$ can be written (as in Weiss):
$$\boxed{f_Y(y) = f_X(x) \cdot \left| \frac{1}{g'(x)} \right|}$$

Physically:



skip **Proof of (*):**

1) Suppose $g \nearrow$



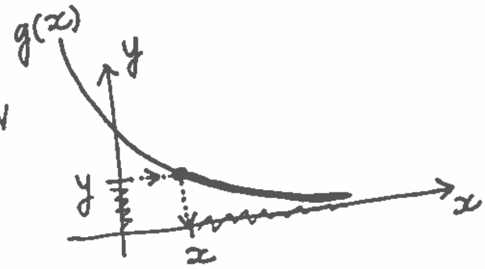
$$\mathbb{F}_Y(y) = \mathbb{P}(Y \leq y) = \mathbb{P}(g(X) \leq y) = \mathbb{P}(X \leq x) = \mathbb{F}_X(x)$$

where $x = g^{-1}(y)$

$$f_Y(y) = \frac{d}{dy}\mathbb{F}_X(x) \underset{\text{chain rule}}{=} \frac{d\,\mathbb{F}_X(x)}{dx} \cdot \frac{dx}{dy} = f_X(x) \cdot \underbrace{\frac{dx}{dy}}_{>0}$$

$$= f_X(x) \cdot \left|\frac{dx}{dy}\right|$$

2) Suppose $g \searrow$



$$\mathbb{F}_Y(y) = \mathbb{P}(Y \leq y) = \mathbb{P}(X \geq x) = 1 - \mathbb{F}_X(x)$$

where $x = g^{-1}(y)$

$$f_Y(y) = \frac{d}{dy}(1 - \mathbb{F}_X(x)) = -\frac{d}{dy}\mathbb{F}_X(x) = -\frac{d\,\mathbb{F}_X(x)}{dx} \cdot \underbrace{\frac{dx}{dy}}_{<0}$$

$$= f_X(x) \cdot \left|\frac{dx}{dy}\right|$$

So, in either case

$$f_Y(y) = f_X(x) \cdot \left|\frac{dx}{dy}\right|, \text{ where } x = g^{-1}(y).$$

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

$Y = a + bX, \quad b \neq 0$

Find $f_Y(y) = ?$

Solution: $y = a + bx$

- $\left|\frac{dy}{dx}\right| = |b|$
- inverse function: $x = \frac{y-a}{b} \; (\equiv g^{-1}(y))$

$\Rightarrow \left|\frac{dx}{dy}\right| = \left|\frac{1}{b}\right| = \frac{1}{|b|}$

$$f_Y(y) = f_X(x) \cdot \left|\frac{dx}{dy}\right| = \frac{1}{|b|} \cdot \frac{1}{\sqrt{2\pi}\cdot\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \qquad \frac{x-\mu}{\sigma} = \frac{\frac{y-a}{b} - \mu}{\sigma} = \frac{y-a-b\mu}{b\sigma}$$

$$= \frac{1}{\sqrt{2\pi}\cdot\sigma|b|} \cdot \exp\left\{-\frac{1}{2}\left(\frac{y-(a+b\mu)}{b\sigma}\right)^2\right\}, \quad -\infty < y < \infty$$

i.e. $\boxed{Y \sim N(a+b\mu, \; \sigma^2 b^2)}$

Special case: $Z = \frac{X-\mu}{\sigma} = a + bX$ with $a = -\frac{\mu}{\sigma}$, $b = \frac{1}{\sigma}$

mean: $a + b\mu = -\frac{\mu}{\sigma} + \frac{\mu}{\sigma} = 0$

Z variance: $\sigma^2 b^2 = \sigma^2 \cdot \left(\frac{1}{\sigma}\right)^2 = 1$

$\Rightarrow Z \sim N(0,1)$

variation—if the variation is too large, too many of the bolts produced will be unusable. The manufacturer has set the tolerance specifications for the 10 mm bolts at ± 0.3 mm; that is, a bolt's diameter is considered satisfactory if it is between 9.7 mm and 10.3 mm. Furthermore, the manufacturer has decided that only 1 in 1000 bolts produced should be defective. Assuming that the diameters of bolts produced are normally distributed with $\mu = 10$ mm, what must σ be to insure that the manufacturer's production criteria are met?

8.104 Blood alcohol concentration (BAC) is the amount of alcohol in the bloodstream, measured in percentages. In many states, a driver is considered legally intoxicated if his or her BAC is 0.10% (i.e., 1 part alcohol per 1000 parts blood in the body) or higher. When a suspected DUI driver is stopped, police request that the person take a breathalyzer test to determine his or her BAC. Such tests are imperfect and exhibit a certain amount of measurement error. Suppose that the measured BAC is a normal random variable with μ equal to the person's actual BAC and $\sigma = 0.005\%$.

a) What is the probability that a driver with a BAC of 0.11% will pass the breathalyzer test?

b) What is the probability that a driver with a BAC of 0.095% will incorrectly be determined to be DUI?

maybe Hint: use CDF method

8.105 Let $X \sim \mathcal{N}(0, 1/\alpha^2)$, where α is a positive real number. Determine the PDF of the random variable $1/X^2$. *Note:* The distribution of $1/X^2$ is called a *one-sided stable distribution of index 1/2*.

8.106 The diameters of ball bearings made by the Acme Ball Bearing Company are normally distributed with $\mu = 1.4$ cm and $\sigma = 0.025$ cm. The bearings are fully inspected and those that have diameters either less than 1.35 cm or greater than 1.48 cm are discarded. Determine the PDF of the diameters of the remaining ball bearings.

8.107 At a bottling plant, two machines are used for filling 16 oz bottles of soda. Machine I has an average fill (μ) of 16.21 oz with $\sigma = 0.14$ oz; Machine II has an average fill (μ) of 16.12 oz with $\sigma = 0.07$ oz. Both fill amounts are normally distributed. Machine I fills twice as many bottles per day as Machine II. What percentage of bottles that contain less than 15.96 oz of soda are filled by Machine I?

8.108 As reported by a spokesperson for Southwest Airlines, the no-show rate for reservations is 16%—that is, the probability is 0.16 that a person making a reservation will not take the flight. For a certain flight, 42 people have reservations. For each part, determine and compare the exact probability by using the appropriate binomial PMF and an approximate probability by using the normal approximation to the binomial as given in Proposition 8.12 on page 445. The probability that the number of people who don't take the flight is

a) exactly 5. **b)** between 9 and 12, inclusive. **c)** at least 1. **d)** at most 2.

8.109 According to *USA TODAY*, Anchorage, Alaska, is the city with the highest rate of cell phone ownership, with 56% of the residents owning cell phones. Of 500 randomly selected (without replacement) Anchorage residents, let X denote the number who own a cell phone.

a) Identify the exact probability distribution of the random variable X.

b) Identify the binomial distribution that should be used to approximate the probability distribution of X.

c) Identify the normal distribution that should be used to approximate the probability distribution of X.

Use the local De Moivre–Laplace theorem to approximate the probability that, of 500 randomly selected Anchorage residents, the number who own a cell phone is

d) exactly 280. **e)** between 278 and 280, inclusive.

Read Ch 9.

Jointly Continuous Random Variables

(X, Y)

Same ideas as in the discrete case:

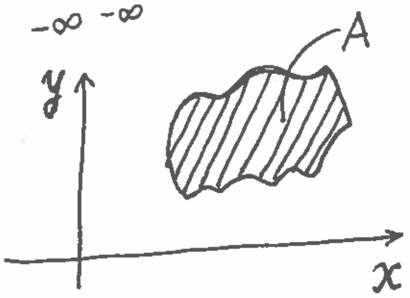
PDF $\quad f_X(x) \qquad f_{XY}(x,y)$

CDF $\quad \mathbb{F}_X(x) = \mathbb{P}(X \leq x)$

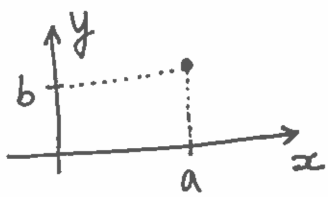
$\mathbb{F}_{XY}(x,y) = \mathbb{P}(X \leq x,\ Y \leq y)$

Rules for $f_{XY}(x,y)$:

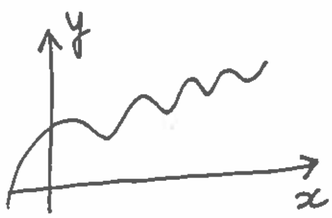
1. $f_{XY}(x,y) \geq 0$
2. $\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{XY}(x,y)\,dx\,dy = 1$



$$\mathbb{P}((X,Y) \in A) = \iint_A f_{XY}(x,y)\,dx\,dy$$



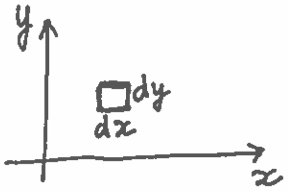
$$\iint f_{XY}(a,b)\,dx\,dy = 0$$



$\mathbb{P}($falling on the curve$) = 0$

If $[X] = $ inches $\quad \underbrace{[Y]}_{\text{dimension}} = $ pounds

$$\Rightarrow [f_{XY}(x,y)] = \frac{1}{\text{inch} \times \text{lb}}$$

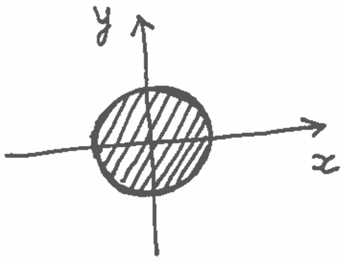


$$f_{XY}(x,y)\,dx\,dy = \mathbb{P}(x < X < x+dx,\ y < Y < y+dy)$$

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

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

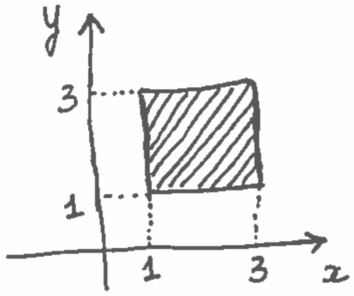
Ex. $(X,Y) \sim \text{Unif (unit disk)}$



$$f_{XY}(x,y) = \begin{cases} \text{const} & \text{on unit disk} \\ 0 & \text{otherwise} \end{cases}$$

$$= \begin{cases} \frac{1}{\pi}, & \text{if } x^2+y^2<1 \\ 0, & \text{otherwise} \end{cases}$$

Ex.



$(X,Y) \sim \text{Unif (on square)}$

PDF $$f_{XY}(x,y) = \begin{cases} \frac{1}{4}, & \begin{cases} 1 \le x \le 3 \\ 1 \le y \le 3 \end{cases} \\ 0, & \text{otherwise} \end{cases}$$

$$\mathbb{F}_{XY}\left(\tfrac{1}{2}, 6\right) = \mathbb{P}\left(X \le \tfrac{1}{2}, Y \le 6\right) = 0$$

$$\mathbb{F}_{XY}(4,5) = \mathbb{P}(\underbrace{X \le 4, Y \le 5}_{\text{encompasses whole square}}) = 1$$

$$\mathbb{F}_{XY}(2,2) = \int_1^2 \int_1^2 \tfrac{1}{4}\, dx\, dy = \tfrac{1}{4} \int_1^2 dx \int_1^2 dy = \tfrac{1}{4}.$$

Find PDF of X, and PDF of Y.

$$X \sim f_X(x) = \int_{-\infty}^{\infty} f_{XY}(x,y)\, dy = \underbrace{\begin{cases} \frac{1}{2}, & 1 \le x \le 3 \\ 0, & \text{if } x<1 \text{ or } x>3 \end{cases}}_{\text{Unif}(1,3)}$$

$$Y \sim f_Y(y) = \begin{cases} \frac{1}{2}, & 1 \le y \le 3 \\ 0, & \text{otherwise} \end{cases} \quad \leftarrow \text{because of symmetry}$$

Note: $$f_{XY}(x,y) = f_X(x) \cdot f_Y(y) = \begin{cases} \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4}, & \begin{cases} 1 \le x \le 3 \\ 1 \le y \le 3 \end{cases} \\ 0, & \text{otherwise} \end{cases}$$

$\Rightarrow$ X and Y are independent.

Rules for the CDF:

$$\mathbb{F}_{XY}(x,y) = \mathbb{P}(X \leqq x, Y \leqq y)$$

$\boxed{=0\,?}$

$$\mathbb{F}_{XY}(-\infty, y) = 0$$

$$\mathbb{F}_{XY}(x, -\infty) = 0$$

$\boxed{=1\,?}$

$$\mathbb{F}_{XY}(\infty, \infty) = 1$$

PDF $\rightarrow$ CDF:

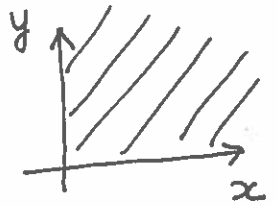
$$\mathbb{F}_{XY}(x,y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{XY}(u,v)\, dv\, du$$

CDF $\rightarrow$ PDF:

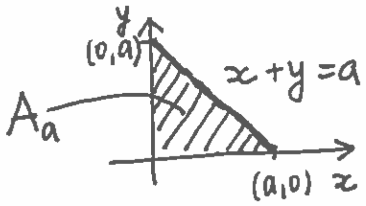
$$\frac{\partial}{\partial x} \frac{\partial}{\partial y} \mathbb{F}_{XY}(x,y) = f_{XY}(x,y)$$

Ex. X = lifetime of 1^{st} device $\sim \mathcal{E}xp(\lambda)$
Y = lifetime of 2^{nd} device $\sim \mathcal{E}xp(\lambda)$ $\Big\rangle$ indep.

($\Rightarrow$ X and Y are iid)



Find $\mathbb{P}(X+Y \leqq a)$, $a>0$

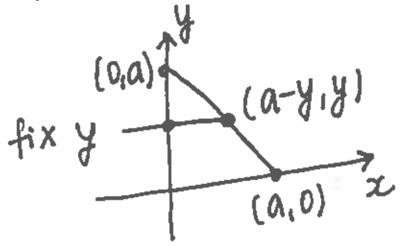


$$\mathbb{P}(X+Y \leqq a) = \iint_{A_a} f_{XY}(x,y)\,dx\,dy =$$

$$= \iint_{A_a} \lambda e^{-\lambda x} \cdot \lambda e^{-\lambda y}\,dx\,dy =$$

fix y
$\int$ over x first

$$= \int_0^a \left[\lambda e^{-\lambda y} \int_0^{a-y} \lambda e^{-\lambda x}\,dx \right] dy =$$



$$= \int_0^a \lambda e^{-\lambda y} \left(\frac{-\lambda}{\lambda} e^{-\lambda x} \Big|_0^{a-y} \right) dy =$$

$$= \int_0^a \lambda e^{-\lambda y} \left(1 - e^{-\lambda(a-y)} \right) dy = \int_0^a \left(\lambda e^{-\lambda y} - \lambda e^{-\lambda a} \right) dy =$$

$$= -\frac{\lambda}{\lambda} e^{-\lambda y} \Big|_0^a - \lambda e^{-\lambda a} \cdot y \Big|_0^a = \boxed{1 - e^{-\lambda a} - \lambda a e^{-\lambda a}}$$

Conditional Densities

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

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

If X and Y are independent

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

HW 9

1) 8.49 p. 426

2) 8.33 p. 415

Hint: for the min, start with $\mathbb{P}(X > x)$ and ask yourself what $\{X > x\}$ implies about $X_1, \dots, X_m$; then use independence.

Hint: for the max, start with $\mathbb{P}(X \leq x)$ and ask yourself what $\{X \leq x\}$ implies about $X_1, \dots, X_m$; then use independence.

3) 8.80 p. 436

4) 8.101 p. 448

5) 8.102 p. 448

6) 8.105 p. 449 Hint: use CDF method

7) 8.154 p. 474 Hint: use PDF method

Extra credit:

1) 8.50 p. 426

2) 8.160 p. 474

morning class HW 10

All of HW 9

\+ 8.30 p. 415

c) Let a and b be real numbers with $a < b$. Determine $P(a < X \leq b)$, $P(a < X < b)$, $P(a \leq X < b)$, and $P(a \leq X \leq b)$ in terms of f.
d) Why are all four answers in part (c) the same?

X **8.30** For each function f, let X be as in Exercise 8.29. Graph f and determine and graph F_X.
a) $f(x) = \lambda e^{-\lambda x}$ if $x > 0$ and $f(x) = 0$ otherwise, where λ is a positive real number.
b) $f(x) = 1/(b-a)$ if $a < x < b$ and $f(x) = 0$ otherwise, where a and b are real numbers with $a < b$.
c) $f(x) = b^{-1}(1 - |x|/b)$ if $-b < x < b$ and $f(x) = 0$ otherwise, where b is a positive real number.

8.31 Decide whether each function F is the CDF of a random variable by checking properties (a)–(d) of Proposition 8.1 on page 411. For each F that is the CDF of a random variable, classify the random variable as discrete, continuous, or mixed.
a) $F(x) = 0$ if $x < 0$ and $F(x) = 1$ if $x \geq 0$.
b) $F(x) = 0$ if $x < 0$, $F(x) = 1 - p$ if $0 \leq x < 1$, and $F(x) = 1$ if $x \geq 1$. Here p is a real number with $0 < p < 1$.
c) $F(x) = 0$ if $x < 0$ and $F(x) = \lfloor x \rfloor$ if $x \geq 0$.
d) $F(x) = 0$ if $x < 0$ and $F(x) = \sum_{n=0}^{\lfloor x \rfloor} a_n$ for $x \geq 0$, where $\{a_n\}_{n=0}^{\infty}$ is a sequence of nonnegative real numbers whose sum is 1.
e) $F(x) = 0$ if $x < 0$ and $F(x) = 1 - e^{-\lambda x}\sum_{j=0}^{r-1}(\lambda x)^j/j!$ if $x \geq 0$. Here λ is a positive real number and r is a positive integer.
f) $F(x) = 0$ if $x < 0$ and $F(x) = x$ if $x \geq 0$.
g) $F(x) = 0$ if $x < -1$, $F(x) = \frac{1}{2} + \frac{3}{8}x$ if $-1 \leq x < 1$, and $F(x) = 1$ if $x \geq 1$.

8.32 Let X be a random variable and let m be a real number. Determine the CDF of each of the following random variables in terms of the CDF of X.
a) $Y = \max\{X, m\}$ b) $Z = \min\{X, m\}$

V **8.33** Let $X_1, \ldots, X_m$ be independent random variables, each having the same probability distribution as a random variable X. Determine the CDF of each of the following random variables in terms of the CDF of X.
a) $Y = \max\{X_1, \ldots, X_m\}$ b) $Z = \min\{X_1, \ldots, X_m\}$

8.34 A function F can be the CDF of a continuous random variable X and still have "flat spots." What is the probabilistic meaning of $F(x) = c$ for all $x \in [a, b]$, where c is a constant and a and b are real numbers with $a < b$?

Theory Exercises

8.35 Prove Equation (8.7) on page 412: If X is a random variable, then $F_X(x-) = P(X < x)$ for all $x \in \mathcal{R}$.

8.36 Prove parts (a), (b), and (d) of Proposition 8.2 on page 412.

Advanced Exercises

8.37 Let X be a random variable. Prove that the CDF of X has a countable number of discontinuities. *Hint:* For each $n \in \mathcal{N}$, consider the set $D_n = \{x \in \mathcal{R} : F_X(x) - F_X(x-) \geq 1/n\}$.

8.38 For each $n \in \mathcal{N}$, let the random variable X_n have the discrete uniform distribution on the set $\{0, 1/n, \ldots, (n-1)/n\}$.
a) Determine the CDF of X_n.

8.76 A trucker drives between fixed locations in Los Angeles and Phoenix. The duration, in hours, of a round trip has an exponential distribution with parameter 1/20. Determine the probability that a round trip

a) takes at most 15 hours. b) takes between 15 and 25 hours. c) exceeds 25 hours.

d) Without using the conditional probability rule, find the probability that a round trip takes at most 40 hours, given that it exceeds 15 hours.

e) Assuming that round-trip durations are independent from one trip to the next, find the probability that exactly two of five round trips take more than 25 hours.

8.77 Ten years ago at a certain insurance company, the size of claims under homeowner insurance policies had an exponential distribution. Furthermore, 25% of claims were less than \$1000. Today, the size of claims still has an exponential distribution but, owing to inflation, every claim made today is twice the size of a similar claim made 10 years ago. Determine the probability that a claim made today is less than \$1000.

8.78 Let X be a positive continuous random variable. Consider the following two mathematical relations:

$$P(X > s + t \mid X > s) = P(X > t), \qquad s, t \geq 0.$$

$$P(X > s + t \mid X > s) = P(X > s + t), \qquad s, t \geq 0.$$

a) Which of the two mathematical relations describes the lack-of-memory property?

b) Describe in words the other mathematical statement.

c) Is it possible for the second relation to hold?

8.79 Let $X \sim \mathcal{U}(0, 1)$ and let $0 < s < s + t < 1$.

a) If X had the lack-of-memory property, what would be $P(X > s + t \mid X > s)$?

b) Determine $P(X > s + t \mid X > s)$ and compare your answer to that in part (a).

✓ **8.80** Suppose that $X_1, \ldots, X_m$ are independent exponential random variables with parameters $\lambda_1, \ldots, \lambda_m$, respectively. Determine and identify the PDF of the random variable $X = \min\{X_1, \ldots, X_m\}$.

8.81 Suppose that the time T, in minutes, for a customer service representative to respond to 10 telephone inquiries is uniformly distributed on the interval (8, 12). Let R denote the average rate, in customers per minute, at which the representative responds to inquiries. Determine the PDF of the random variable R.

8.82 Beginning at 6:00 P.M. on any given day, the number of patients, $N(t)$, that arrive at an emergency room within the first t hours has a Poisson distribution with parameter $6.9t$. Using this fact only, determine and identify the probability distribution of the elapsed time X until the first patient arrives.

8.83 Median of a random variable: A *median* of a random variable X is any number M such that $P(X \leq M) \geq 1/2$ and $P(X \geq M) \geq 1/2$. Every random variable has at least one median, but medians are not necessarily unique.

a) Show that M is a median of X if and only if $F_X(M-) \leq 1/2 \leq F_X(M)$.

b) Show that M is a median of X if and only if $P(X < M) \leq 1/2$ and $P(X > M) \leq 1/2$.

c) Determine the median(s) for an indicator random variable.

d) Suppose that X is a continuous random variable whose range is an interval and whose CDF is strictly increasing on its range. Show that X has a unique median given by the number M that satisfies $F_X(M) = 1/2$.

e) Suppose that $X \sim \mathcal{U}(a, b)$. Without doing any calculations, make an educated guess for the median of X.

8.44 A point is chosen at random from the interior of a triangle with base b and height h. Let Y denote the distance from the point chosen to the base of the triangle.

a) Determine a PDF of the random variable Y. *Note:* Exercise 8.17 on page 413 asks for the CDF of Y.
b) Interpret the PDF of Y obtained in part (a) with regard to which values of Y are more likely than others.

8.45 Refer to Example 8.2 on page 403. Let X and Y denote the x and y coordinates, respectively, of the center of the first spot (visible bacteria colony) to appear.

a) Obtain a PDF of X. *Hint:* First use calculus to express the CDF of X as an integral.
b) Interpret the PDF of X obtained in part (a) with regard to which values of X are more likely than others.
c) Without doing any calculations, obtain a PDF of Y. Explain your reasoning.

8.46 Let X be a continuous random variable with a PDF. Express a PDF of each of the following functions of X in terms of f_X.

a) $a + bX$, where a and $b \neq 0$ are real numbers   b) X^2

8.47 Let f be a nonnegative real-valued function such that $\int_{-\infty}^{\infty} f(x)\,dx = 1$. Suppose that X is a random variable that satisfies $P(X \in A) = \int_A f(x)\,dx$ for all subsets A of $\mathcal{R}$. Determine a PDF of X.

8.48 Construct a PDF from each of the following functions. Explain your reasoning.

a) $g(x) = x^3(1-x)^2$ for $0 < x < 1$ and $g(x) = 0$ otherwise
b) $g(x) = e^{-4x}$ for $0 < x < \infty$ and $g(x) = 0$ otherwise
c) $g(x) = x^2 e^{-6x}$ for $0 < x < \infty$ and $g(x) = 0$ otherwise
d) $g(x) = 1/(1+x^2)$ for $-\infty < x < \infty$
e) $g(x) = \sin x$ for $0 < x < \pi$ and $g(x) = 0$ otherwise

X **8.49** Decide whether each function f is a PDF of a random variable by checking properties (a) and (b) of Proposition 8.6 on page 423. Obtain the CDF corresponding to each function f that is a PDF.

a) $f(x) = \lambda e^{-\lambda x}$ for $x > 0$ and $f(x) = 0$ otherwise, where λ is a positive real number
b) $f(x) = 1/(b-a)$ for $a < x < b$ and $f(x) = 0$ otherwise, where a and b are real numbers with $a < b$
c) $f(x) = x/4$ for $-1 < x < 3$ and $f(x) = 0$ otherwise
d) $f(x) = x^2$ for $-1 < x < 3$ and $f(x) = 0$ otherwise
e) $f(x) = b^{-1}(1 - |x|/b)$ for $-b < x < b$ and $f(x) = 0$ otherwise, where b is a positive real number

8.50 Provide an example of an unbounded PDF. What is the corresponding CDF? Is an unbounded CDF possible?

8.51 Give an example of a PDF that is positive for all $x \in \mathcal{R}$. Find the corresponding CDF.

8.52 Define $F: \mathcal{R} \to \mathcal{R}$ by

$$F(x) = \begin{cases} 0, & \text{if } x < 0; \\ x/2, & \text{if } 0 \le x < 1; \\ (x+2)/6, & \text{if } 1 \le x < 4; \\ 1, & \text{if } x \ge 4. \end{cases}$$

a) Is F the CDF of a continuous random variable? Explain your answer.
b) If your answer to part (a) is "yes," determine the corresponding PDF.

8.94 Let $X \sim \mathcal{N}(\mu, \sigma^2)$. Show that, for all $t > 0$,
a) $P(|X - \mu| \le t) = 2\Phi(t/\sigma) - 1$. b) $P(|X - \mu| \ge t) = 2\big(1 - \Phi(t/\sigma)\big)$.

8.95 Directly verify Equation (8.36) on page 443. That is, show that

$$\frac{1}{\sigma}\phi\left(\frac{x-\mu}{\sigma}\right) = \frac{1}{\sqrt{2\pi}\,\sigma}\,e^{-(x-\mu)^2/2\sigma^2}$$

directly from the definition of ϕ.

8.96 Let $X \sim \mathcal{N}(\mu, \sigma^2)$ and let $z > 0$.
a) Without evaluation, explain why a probability of the form $P(\mu - z\sigma \le X \le \mu + z\sigma)$ doesn't depend on either μ or σ.
b) Determine $P(\mu - z\sigma \le X \le \mu + z\sigma)$ for $z = 1, 2,$ and 3.

8.97 Two normal random variables, X and Y, have the same μ and σ^2 parameters. What can you say about the probability distributions of X and Y? Explain your answer.

8.98 Students in an introductory statistics course at the U.S. Air Force Academy participated in Nabisco's "Chips Ahoy! 1,000 Chips Challenge" by confirming that there were at least 1000 chips in every 18-ounce bag of cookies that they examined. As part of their assignment, they concluded that the number of chips per bag is approximately normally distributed. [Source: Brad Warner and Jim Rutledge, "Checking the Chips Ahoy! Guarantee," *Chance*, 1999, Vol. 12(1), pp. 10–14] Give two reasons why the number of chips in a bag couldn't be exactly normally distributed.

8.99 As reported in *Runner's World* magazine, the times of the finishers in the New York City 10 km run are normally distributed with $\mu = 61$ minutes and $\sigma = 9$ minutes. Let X be the time, in minutes, of a randomly selected finisher. Find
a) $P(X > 75)$. b) $P(X < 50 \text{ or } X > 70)$.

8.100 In 1905, R. Pearl published the article "Biometrical Studies on Man. I. Variation and Correlation in Brain Weight" (*Biometrika*, Vol. 4, pp. 13–104). According to the study, brain weights of Swedish men are normally distributed with $\mu = 1.40$ kg and $\sigma = 0.11$ kg. Obtain the percentage of Swedish men who have brain weights
a) between 1.50 kg and 1.70 kg. b) less than 1.6 kg.

X **8.101** Refer to Example 8.14 on page 444.
a) What percentage of pregnant women give birth before 300 days?
b) Among those women with a longer than average gestation, what percentage give birth within 300 days? *Note:* As we show in Chapter 10, the parameter μ of a normal random variable is its average (expected) value.
c) In a court case, the prosecuting attorney claims that the defendant is the father of a child who was born on July 6, 2002. The defendant can prove that he was out of town from September 1, 2001, to April 3, 2002. Can the defendant use this information to refute the prosecuting attorney's claim? Explain your answer.

✓ **8.102** When you put your money into a soft drink machine at the Student Union, a paper cup comes down, and some cola is put into the cup. You are supposed to get 8 oz of cola. However, the actual amount of cola dispensed is random, having a normal distribution with μ equal to the machine setting and $\sigma = 0.25$ oz. What should the machine setting be so that, in the long run, only 2% of the drinks will contain less than 8 oz?

8.103 A hardware manufacturer produces 10 mm bolts. The manufacturer knows that the diameters of the bolts produced vary somewhat from 10 mm and also from each other. But even if he is willing to accept some variation in bolt diameters, he can't tolerate too much

EXERCISES 8.7 Basic Exercises

In Exercises 8.141–8.149, we cite exercises that appeared earlier in this chapter. Basically, their solutions involved the application of the CDF method, Procedure 8.2 on page 465. In each case, decide whether the transformation method, Procedure 8.3 on page 467, is appropriate; if it is, apply it.

8.141 Exercise 8.67(b) on page 435.

8.142 Exercise 8.72(a) on page 435.

8.143 Exercise 8.73(a) on page 435.

8.144 Exercise 8.105 on page 449.

8.145 Exercise 8.118 on page 461.

8.146 Exercise 8.120 on page 461.

8.147 Exercise 8.122 on page 462.

8.148 Exercise 8.128(b) on page 462.

8.149 Exercise 8.130(b) on page 462.

8.150 Solve Example 8.20 on page 468 by using the CDF method, Procedure 8.2 on page 465.

8.151 Use the transformation method to verify Equation (8.64): If X is a continuous random variable with a PDF and a and $b \neq 0$ are real numbers, then $f_{a+bX}(y) = |b|^{-1} f_X\big((y-a)/b\big)$.

8.152 An actuary modeled the lifetime of a device by the random variable $Y = 10X^{0.8}$, where X is exponentially distributed with parameter 1. Find a PDF of Y.

8.153 Let $X \sim \mathcal{C}(\eta, \theta)$ and let ψ denote the standard Cauchy PDF.
a) Verify Equation (8.66) on page 471: $f_X(x) = \theta^{-1}\psi\big((x-\eta)/\theta\big)$.
b) Use part (a) to conclude that $(X - \eta)/\theta$ has the standard Cauchy distribution.

Hint: Use PDF method

maybe **8.154** **Lognormal random variable:** Let $X \sim \mathcal{N}(\mu, \sigma^2)$.
a) Determine a PDF of the random variable $Y = e^X$. *Note:* A random variable with this PDF is called a *lognormal random variable* and is said to have the *lognormal distribution with parameters μ and σ^2*.
b) Explain why the term "lognormal" is used for a random variable with a PDF as found in part (a).

8.155 Find a PDF of the random variable $Y = \sin X$ when
a) $X \sim \mathcal{U}(-\pi/2, \pi/2)$. b) $X \sim \mathcal{U}(-\pi, \pi)$.

8.156 A point is selected at random from the unit circle (i.e., boundary of the unit disk).
a) Determine a PDF of the x coordinate of the point chosen.
b) Without doing any calculations, determine a PDF of the y coordinate of the point chosen. Explain your reasoning.
c) Determine a PDF of the distance from the point chosen to the point $(1, 0)$.

8.157 Let X have the beta distribution with parameters α and β and set $Y = a + (b-a)X$, where a and b are real numbers with $a < b$.
a) Identify the range of Y. b) Obtain a PDF of Y.

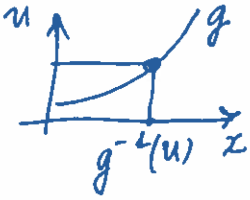
8.158 Show that, if X has the standard Cauchy distribution, so does $1/X$.

Bivariate Transformation Theorem

Recall the PDF method in the univariate case

$X \sim f_X(x)$

$\underbrace{U}_{\text{function of } X} = g(X)$, $g \nearrow$ (or $\searrow$)



Find $U \sim f_U(u) = ?$

$$f_U(u) = f_X(x) \cdot \left|\frac{dx}{du}\right| = \frac{1}{\left|\frac{du}{dx}\right|} \cdot f_X(x),$$

where x is the point corresponding to u

i.e. $x = g^{-1}(u)$.

Bivariate case

$(X, Y) \sim f_{XY}(x, y)$

$\begin{cases} U = g(X, Y) \\ V = h(X, Y) \end{cases}$ the transformation is $\underbrace{(X, Y)}_{\text{old}} \to \underbrace{(U, V)}_{\text{new}}$

Assume the transformation is 1 to 1,

i.e. given (u, v) we can recover (x, y) uniquely.

Then $\boxed{f_{UV}(u, v) = \frac{1}{|J(x, y)|} \cdot f_{XY}(x, y), \text{ where } (x, y) \text{ corresponds to } (u, v).}$

· absolute value

· $J(x, y) = \det \begin{pmatrix} \frac{\partial g}{\partial x} & \frac{\partial g}{\partial y} \\ \frac{\partial h}{\partial x} & \frac{\partial h}{\partial y} \end{pmatrix} = \frac{\partial g}{\partial x} \cdot \frac{\partial h}{\partial y} - \frac{\partial g}{\partial y} \cdot \frac{\partial h}{\partial x}$.

Example 9.25 p. 547

The Bivariate Transformation Theorem

Quality Assurance For purposes of quality assurance, an expensive item produced at a manufacturing plant is independently inspected by two engineers. The amount of time it takes each engineer to inspect the item has the exponential distribution with parameter λ. Let S denote the proportion of the total inspection time attributed to the first engineer, and let T denote the total inspection time by both engineers.

a) Use the bivariate transformation theorem to obtain a joint PDF of S and T.
b) Use the result of part (a) to obtain and identify a marginal PDF of S.
c) Use the result of part (a) to obtain and identify a marginal PDF of T.
d) Show that S and T are independent random variables.

a) Let X and Y denote the inspection times for the first and second engineers, respectively.

X, Y iid $Exp(\lambda)$

$$\Rightarrow f_{XY}(x,y) = \begin{cases} \lambda e^{-\lambda x} \cdot \lambda e^{-\lambda y} = \lambda^2 \cdot e^{-\lambda(x+y)}, & \text{if } \begin{cases} x>0 \\ y>0 \end{cases} \\ 0, & \text{otherwise} \end{cases}$$

Note that $S = \dfrac{X}{X+Y}$, $S \in (0,1)$

and $T = X+Y$, $T \in (0,\infty)$

$$\left.\begin{aligned} s &= g(x,y) = \frac{x}{x+y} \\ t &= h(x,y) = x+y \end{aligned}\right| \Rightarrow x = t-y \Rightarrow s = \frac{t-y}{t-\cancel{y}+\cancel{y}} \Rightarrow \boxed{y = t(1-}$$

$$\Rightarrow x = t-y = \cancel{t} - \cancel{t} + ts = ts$$

$$\Rightarrow \boxed{x = ts}$$

$$\Rightarrow J(x,y) = \det\begin{bmatrix} \frac{\partial g}{\partial x} & \frac{\partial g}{\partial y} \\ \frac{\partial h}{\partial x} & \frac{\partial h}{\partial y} \end{bmatrix} = \det\begin{bmatrix} \frac{y}{(x+y)^2} & \frac{-x}{(x+y)^2} \\ 1 & 1 \end{bmatrix} = \frac{y+x}{(x+y)^2} = \frac{1}{x+y}$$

$$\Rightarrow f_{ST}(s,t) = (x+y) \cdot \lambda^2 \cdot e^{-\lambda(x+y)} = \begin{cases} \lambda^2 \cdot t \cdot e^{-\lambda t}, & \begin{matrix} 0<s<1 \\ t>0 \end{matrix} \\ 0, & \text{otherwise} \end{cases}$$

b) for $0 < s < 1$:

$$f_S(s) = \int_{-\infty}^{\infty} f_{ST}(s,t)\,dt = \int_0^{\infty} \lambda^2 \cdot t e^{-\lambda t}\,dt \overset{\text{re-write}}{=}$$

$$= \underbrace{\Gamma(2)}_{1} \cdot \underbrace{\int_0^{\infty} \frac{\lambda^2}{\Gamma(2)} \cdot t^{2-1} \cdot e^{-\lambda t}\,dt}_{1}$$

$$\Rightarrow f_S(s) = \begin{cases} 1, & 0 < s < 1 \\ 0, & \text{otherwise} \end{cases} \Rightarrow \underline{S \sim \text{Unif}(0,1)}$$

c) for $t > 0$:

$$f_T(t) = \int_{-\infty}^{\infty} f_{ST}(s,t)\,ds = \int_0^{1} \lambda^2 \cdot t e^{-\lambda t}\,ds = \lambda^2 \cdot t \cdot e^{-\lambda t}$$

$$\Rightarrow f_T(t) = \begin{cases} \frac{\lambda^2}{\Gamma(2)} \cdot t^{2-1} \cdot e^{-\lambda t}, & t > 0 \\ 0, & \text{otherwise} \end{cases} \Rightarrow \underline{T \sim \text{Gamma}(2, \ldots}$$

d) $f_{ST}(s,t) = f_S(s) \cdot f_T(t) \Rightarrow$ S and T are independent.

Ex. Distribution of the sum of two indep. RV's

Convolution

Let $X \sim f_X(x)$, $Y \sim f_Y(y)$ indep.

Find the pdf of $U = X+Y$

$$\begin{cases} U = X+Y \\ V = X \end{cases}$$

Jacobian: $J(x,y) = \det \begin{pmatrix} \frac{\partial(x+y)}{\partial x} & \frac{\partial(x+y)}{\partial y} \\ \frac{\partial x}{\partial x} & \frac{\partial x}{\partial y} \end{pmatrix} = \det \begin{pmatrix} 1 & 1 \\ 1 & 0 \end{pmatrix} = -1$ ASK

Inverses: $x = v$, $y = u - x = u - v$ $\Rightarrow$ Joint: $f_{UV}(u,v) = \frac{1}{|-1|} f_{XY}(x,y)$

$$= f_{XY}(v, u-v)$$

Marginal of U: $\boxed{f_U(u)} = \int_{-\infty}^{\infty} f_{UV}(u,v)\,dv$

$$= \int_{-\infty}^{\infty} f_{XY}(v, u-v)\,dv$$

by indep. $= \boxed{\int_{-\infty}^{\infty} f_X(v) \cdot f_Y(u-v)\,dv}$.

Ex. Distribution of a product of two indep. RVs

$X \sim f_X(x)$
$Y \sim f_Y(y)$ $\quad$ X and Y indep.

Find the distribution of $V = X \cdot Y$

Make a 1 to 1 transformation between (X, Y) and (U, V):

$$U = X \qquad X = U$$
$$V = XY \quad \Rightarrow \quad Y = \frac{V}{U}$$

$$J(x,y) = \det \begin{bmatrix} 1 & 0 \\ y & x \end{bmatrix} = x$$

$$\Rightarrow f_{UV}(u,v) = \frac{1}{|x|} \cdot f_X(x) \cdot f_Y(y) = \frac{1}{|u|} f_X(u) f_Y\left(\frac{v}{u}\right)$$

$$\Rightarrow \boxed{f_V(v) = \int_{-\infty}^{\infty} \frac{1}{|u|} f_X(u) f_Y\left(\frac{v}{u}\right) du}$$

✓ **9.122** A device containing two key components fails when and only when both components fail. The lifetimes, T_1 and T_2, of these components are independent with common density function $f(t) = e^{-t}$ for $t > 0$, and $f(t) = 0$ otherwise. The cost of operating the device until failure is $2T_1 + T_2$. Obtain a PDF of this cost.

9.123 Let X and Y be independent random variables with PDFs given by $f_X(x) = f_Y(y) = 0$ for negative x and y, and $f_X(x) = ax^{\alpha}e^{-\lambda x}$ and $f_Y(y) = by^{\beta}e^{-\lambda y}$ for positive x and y, where a and b are constants. Without doing any calculations, identify the probability distribution of the random variable $X + Y$.

9.124 Let X and Y be independent random variables with $X \sim \mathcal{N}(2, 1)$ and $Y \sim \mathcal{N}(3, 2)$. Identify the probability distribution of each of the following random variables.
a) $X + Y$ b) $X - Y$ c) $2 - 3X + 4Y$

9.125 Let $X_1, \ldots, X_n$ be a random sample from a $\mathcal{N}(\mu, \sigma^2)$ distribution.
a) Determine the standardized version of the random variable $X_1 + \cdots + X_n$ and identify its probability distribution.
b) Identify the probability distribution of the sample mean, $\bar{X}_n$.

9.126 Two instruments are used to measure the height, h, of a tower. The error made by the less accurate instrument is normally distributed with parameters 0 and $(0.0056h)^2$. The error made by the more accurate instrument is normally distributed with parameters 0 and $(0.0044h)^2$. If the two measurements are independent random variables, what is the probability that their average value is within $0.005h$ of the height of the tower?

9.127 A company manufactures a brand of light bulb with a lifetime, in months, that is normally distributed with parameters 3 and 1. A consumer buys a number of these bulbs with the intention of replacing them successively as they burn out. The light bulbs have independent lifetimes. What is the smallest number of bulbs to be purchased so that the succession of light bulbs produces light for at least 40 months with probability at least 0.9772?

9.128 In Exercise 8.154 on page 474, we presented the definition of a lognormal random variable. Let $X_1, \ldots, X_m$ be independent lognormal random variables and let $a_1, \ldots, a_m$ be nonzero real numbers.
a) Show that $\prod_{j=1}^{m} X_j^{a_j}$ is lognormal.
b) Deduce from part (a) that the product of a finite number of independent lognormal random variables is a lognormal random variable.

9.129 A company uses batteries of two types, A and B. The lifetimes of type A and type B batteries are exponential random variables with parameters λ and μ, respectively. For a life test, n batteries of type A and m batteries of type B are used. What is the probability that the first battery to fail is of type A?

9.130 An insurance company that offers earthquake insurance models annual premiums and annual claims by exponential distributions with parameters $\frac{1}{2}$ and 1, respectively. Premiums and claims are independent. Obtain a density function for the ratio of claims to premiums.

9.131 Simulation: This exercise requires access to a computer or graphing calculator.
a) Use a basic random number generator to conduct two simulations of 10,000 observations each of a $\mathcal{U}(0, 1)$ random variable.
b) Add the first observations from the two simulations, the second observations from the two simulations, and so on.
c) Roughly, what would you expect a histogram of the 10,000 sums found in part (b) to look like? Explain your answer.
d) Obtain a histogram of the 10,000 sums found in part (b).

Multivariate Independent Continuous Random Variables

As in the bivariate continuous case, a necessary and sufficient condition for several continuous random variables with a joint PDF to be independent is that their joint PDF equals the product of their marginal PDFs. We leave the precise statement of this result and its proof to you as Exercise 9.108.

EXERCISES 9.5 Basic Exercises

9.88 In the petri-dish illustration of Example 9.8 on page 511, let X and Y denote the x and y coordinates, respectively, of the center of the first spot (visible bacteria colony) to appear. We showed, in Example 9.14 on page 524, that the random variables X and Y aren't independent. Here you are asked to provide three other arguments to establish that result.

a) Argue heuristically that X and Y aren't independent by considering the possible values of Y among different specified values of X.
b) Use the results of Examples 9.8(b) and 9.10 (pages 512 and 515, respectively) and Proposition 9.10 (page 527) to show that X and Y aren't independent.
c) Use Proposition 9.4 on page 499—specifically, Equation (9.22)—to show that X and Y aren't independent. *Hint:* Assume that the product of the marginals of X and Y is a joint PDF of X and Y. Obtain a contradiction.

9.89 Provide another verification that the random variables X and Y in Example 9.15 on page 525 are independent. *Hint:* Refer to Example 9.11 on page 516.

9.90 Solve Example 9.16 on page 526 if the arrival times are independent uniform random variables on the interval $(-5, 5)$, where time is measured in minutes relative to noon.

√ **9.91** Let X and Y be the x and y coordinates, respectively, of a point selected at random from the unit square. Determine whether X and Y are independent random variables.

9.92 Let X and Y be the x and y coordinates, respectively, of a point selected at random from the upper half of the unit disk, that is, from the set $\{ (x, y) \in \mathcal{R}^2 : x^2 + y^2 < 1,\ y > 0 \}$. Determine whether X and Y are independent random variables. *Note:* Exercise 9.65 on page 520 provides a joint PDF of X and Y and asks for marginal and conditional PDFs.

9.93 Refer to the regression analysis illustration of Example 9.13 on page 517. Use Proposition 9.10 on page 527 to determine necessary and sufficient conditions for X and Y to be independent random variables.

9.94 Let X and Y be independent random variables, each uniform on the interval $(-1, 1)$. Find the probability that the roots of the random quadratic equation $x^2 + Xx + Y = 0$ are real.

9.95 Let X and Y be continuous random variables with a joint PDF. Suppose that there are nonnegative functions g and h defined on $\mathcal{R}$ such that $f_{X,Y}(x, y) = g(x)h(y)$ for all $x, y \in \mathcal{R}$. Show that X and Y are independent by proceeding as follows.

a) Obtain a marginal PDF of X in terms of g and h.
b) Obtain a marginal PDF of Y in terms of g and h.
c) Explain why $\left(\int_{-\infty}^{\infty} g(x)\,dx\right)\left(\int_{-\infty}^{\infty} h(y)\,dy\right) = 1$.
d) Verify that X and Y are independent random variables.

9.96 In Exercise 9.95, is it necessarily true that g is a marginal PDF of X and h is a marginal PDF of Y? If not, find conditions when that is the case.

9.22 Let X and Y be the x and y coordinates of a point selected at random from the unit square. Determine $P(1/2 \leq X \leq 3/4,\ 1/4 \leq Y \leq 3/4)$ by using
a) Equation (9.4) on page 487.
b) the joint CDF of X and Y, obtained in Example 9.1 on page 487.
c) the joint PDF of X and Y, obtained in Example 9.3 on page 496.
d) Repeat parts (a)–(c) for $P(X > 0.6 \text{ or } Y < 0.2)$.

9.23 Let X and Y be the x and y coordinates, respectively, of a point selected at random from the upper half of the unit disk—that is, from the set $\{(x, y) \in \mathcal{R}^2 : x^2 + y^2 < 1,\ y > 0\}$.
a) Determine a joint PDF of X and Y. *Note:* Exercise 9.4(c) on page 491 asks for the joint CDF of X and Y.
b) Explain why the joint PDF obtained in part (a) is a nonzero constant on the upper half of the unit disk and is zero elsewhere.

9.24 Let X and Y be independent exponential random variables with parameters λ and μ, respectively.
a) Obtain a joint PDF of X and Y. *Note:* Exercise 9.7(b) on page 491 asks for the joint CDF of X and Y.
b) What is the relationship between the joint PDF of X and Y found in part (a) and the (standard) individual PDFs of X and Y?
c) Does the relationship in part (b) surprise you? Explain your answer.

9.25 Let X and Y be the minimum and maximum, respectively, of a random sample of size n taken from an exponential distribution with parameter λ. Determine a joint PDF of X and Y.

maybe **9.26** Let X and Y have joint PDF given by $f(x, y) = x + y$ for $0 < x < 1$ and $0 < y < 1$, and $f(x, y) = 0$ otherwise.
a) Use the joint PDF to determine $P(1/4 < X < 3/4,\ 1/2 < Y < 1)$.
b) Determine the joint CDF of the random variables X and Y.
c) Use the joint CDF to determine $P(1/4 < X < 3/4,\ 1/2 < Y < 1)$. Compare your answer with that obtained in part (a).

Theory Exercises

9.27 Prove Proposition 9.3 on page 496 when the partial derivatives of $F_{X,Y}$, up to and including those of the second order, exist and are continuous everywhere.

9.28 Prove Proposition 9.4 on page 499, which provides an equivalent condition for the existence of a bivariate joint PDF.

9.29 Let $X_1, \ldots, X_m$ be random variables defined on the same sample space.
a) Define *joint probability density function* for these random variables. *Hint:* Refer to Definition 9.2 on page 495.
b) State the m-variate analogue of Proposition 9.3 on page 496.
c) State the m-variate analogue of Proposition 9.4 on page 499.

Advanced Exercises

9.30 Refer to Exercise 9.14 on page 492, where X and Y are the x and y coordinates, respectively, of a point selected at random from the diagonal of the unit square—that is, from $\{(x, y) \in \mathcal{R}^2 : y = x,\ 0 < x < 1\}$.
a) Show that X and Y are continuous random variables.
b) Show that X and Y can't possibly have a joint PDF.

b) Compare the work entailed in part (a) to that in Example 9.3 on page 496 (which also depended on the work done in Example 9.1 on page 487).
c) Find the probability that the magnitude of the difference of the x and y coordinates of the point obtained is at most 1/4 by using the joint PDF of X and Y; by using Exercise 9.40(b).

9.42 Let X and Y be the x and y coordinates, respectively, of a point selected at random from the upper half of the unit disk—that is, from the set $\{ (x, y) \in \mathcal{R}^2 : x^2 + y^2 < 1,\ y > 0 \}$.
a) Use Exercise 9.40(a) to obtain a joint PDF of X and Y.
b) Compare the work entailed in part (a) to that in Exercise 9.23 on page 500 (which also depended on the work done in Exercise 9.4(c) on page 491).
c) Find the probability that the point obtained lies in the triangle with vertices $(-1, 0)$, $(0, 1)$, and $(1, 0)$ by using the joint PDF of X and Y; by using Exercise 9.40(b).
d) Determine the probability that the x coordinate of the point obtained is at least 1/2 unit from the origin.

9.43 Multivariate uniform random variables: Let S be a subset of $\mathcal{R}^m$ with finite nonzero m-dimensional volume. Suppose that a point is selected at random (i.e., uniformly) from S and, for $1 \leq j \leq m$, let X_j denote the x_j coordinate of the point obtained.
a) Determine a joint PDF of $X_1, \ldots, X_m$.
b) Determine a simple formula for $P\big((X_1, \ldots, X_m) \in A\big)$, where $A \subset \mathcal{R}^m$.

9.44 A point is selected a random from the unit cube, $\{ (x, y, z) \in \mathcal{R}^3 : 0 < x, y, z < 1 \}$. Let X, Y, and Z be the x, y, and z coordinates, respectively, of the point obtained.
a) Use the result of Exercise 9.43(a) to determine a joint PDF of X, Y and Z.
b) Use the result of part (a) to find the probability that $Z = \max\{X, Y, Z\}$—that is, that Z is the largest among X, Y and Z.
c) Use a symmetry argument to solve part (b).
d) Find the probability that the point obtained lies in the sphere of radius 1/4 centered at the point $(1/2, 1/2, 1/2)$.
e) Find the probability that, for the point obtained, the sum of the x and y coordinates exceeds the z coordinate.

maybe **9.45** A company is reviewing tornado damage claims under a farm insurance policy. Let X be the portion of a claim representing damage to the house and let Y be the portion of the same claim representing damage to the rest of the property. A joint density function of X and Y is $f_{X,Y}(x, y) = 6(1 - x - y)$ in the triangle with vertices $(0, 0)$, $(1, 0)$, and $(0, 1)$, and $f_{X,Y}(x, y) = 0$ otherwise.
a) Determine the probability that the portion of a claim representing damage to the house exceeds the portion of the same claim representing damage to the rest of the property.
b) Determine the probability that the portion of a claim representing damage to the house is less than 0.2.

9.46 In Example 9.5 on page 503, we considered two electrical components, A and B, with respective lifetimes X and Y whose joint PDF is $f_{X,Y}(x, y) = \lambda\mu e^{-(\lambda x + \mu y)}$ for $x > 0$ and $y > 0$, and $f_{X,Y}(x, y) = 0$ otherwise. Suppose that components A and B constitute an electrical unit.
a) Find a PDF of this electrical unit's lifetime if it's a parallel system—that is, if it functions when at least one of the components is working. *Hint:* First obtain the CDF of the electrical unit's lifetime by using the FPF.
b) Find a PDF of this electrical unit's lifetime if it's a series system—that is, if it functions only when both components are working. Identify the lifetime distribution in this case.
c) Determine the probability that exactly one of the two components is working at time t.

However, when considering more than two continuous random variables, there are many more marginal PDFs.

For instance, suppose that X, Y, and Z are three continuous random variables with a joint PDF. In this case, there are $\binom{3}{1} = 3$ univariate marginal PDFs—namely, f_X (the PDF of X), f_Y (the PDF of Y), and f_Z (the PDF of Z). Additionally, there are $\binom{3}{2} = 3$ bivariate marginal PDFs—namely, $f_{X,Y}$ (the joint PDF of X and Y), $f_{X,Z}$ (the joint PDF of X and Z), and $f_{Y,Z}$ (the joint PDF of Y and Z). Each marginal PDF is obtained by integrating the joint PDF of X, Y, and Z on the unwanted variable(s). To illustrate, the univariate marginal PDF of Y and the bivariate marginal PDF of Y and Z are obtained as follows:

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

and

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

In general, if $X_1, \ldots, X_m$ are m continuous random variables with a joint PDF, there are $\binom{m}{j}$ j-variate marginal PDFs $(1 \leq j \leq m-1)$. Each such j-variate marginal PDF is obtained by integrating the joint PDF of $X_1, \ldots, X_m$ on the other $m - j$ variables.

We can also define conditional PDFs in the general multivariate case. The formulas for these conditional PDFs are analogous to those in the bivariate case. For instance, if X, Y, and Z are continuous random variables with a joint PDF, then

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

and

$$f_{Z\,|\,X,Y}(z \,|\, x, y) = \frac{f_{X,Y,Z}(x, y, z)}{f_{X,Y}(x, y)}.$$

From the previous formula and the general multiplication rule for the joint PDF of two continuous random variables—Equation (9.34) on page 517—we easily obtain the general multiplication rule in the trivariate case:

$$f_{X,Y,Z}(x, y, z) = f_X(x) f_{Y\,|\,X}(y \,|\, x) f_{Z\,|\,X,Y}(z \,|\, x, y). \tag{9.36}$$

Exercise 9.80 asks you to state the general multiplication rule for the joint PDF of m continuous random variables.

EXERCISES 9.4 Basic Exercises

9.63 Let X and Y denote the x and y coordinates, respectively, of a point selected at random from the unit square. In Example 9.7 on page 510, we obtained marginal PDFs of X and Y.

a) Obtain and identify all conditional PDFs.

b) Compare the conditional PDFs of Y given $X = x$ to each other and to the marginal PDF of Y. Interpret the results.

maybe

9.64 For each part, determine $P(X > 0.9 \,|\, Y = 0.8)$ for the specified joint PDF of a random point (X, Y) in the unit square.

a) $f_{X,Y}(x, y) = 1$ b) $f_{X,Y}(x, y) = x + y$ c) $f_{X,Y}(x, y) = \frac{3}{2}(x^2 + y^2)$

Read Ch. 10.

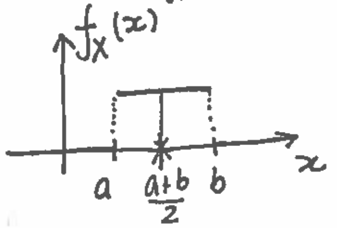
Expected Value of Continuous Random Variables

Def $\mathbb{E}X = \int_{-\infty}^{\infty} x \cdot f_X(x)\,dx$

(same properties as in the discrete case
$\mathbb{E}X = \sum_{R_X} x\, \mathbb{P}_X(x)$)

Ex. $X \sim Unif(a,b)$

$$\mathbb{E}X = \int_a^b \frac{x}{b-a}\,dx = \frac{1}{b-a}\frac{x^2}{2}\Big|_a^b = \frac{1}{2(b-a)}(b^2-a^2) = \boxed{\frac{a+b}{2}}$$



Ex. $X \sim Exp(\lambda)$

$$\mathbb{E}X = \int_0^{\infty} \underbrace{x}_{u}\, \underbrace{\lambda e^{-\lambda x}\,dx}_{dv}$$

integration by parts
$u = x$
$du = dx$
$dv = \lambda e^{-\lambda x}\,dx$
$\Rightarrow v = \int \lambda e^{-\lambda x}\,dx = -e^{-\lambda x}$

$$= x\cdot(-e^{-\lambda x})\Big|_0^{\infty} - \int_0^{\infty}(-e^{-\lambda x})\,dx =$$

$$= (0-0) - \left(\frac{1}{\lambda}e^{-\lambda x}\Big|_0^{\infty}\right) = 0 - \left(0 - \frac{1}{\lambda}\cdot 1\right) = \boxed{\frac{1}{\lambda}}$$

Ex. X = time to breakdown

$X \sim Exp(\lambda)$

let $\mathbb{E}X = 6$ = expected time to breakdown is 6 ho

$\Rightarrow \lambda = \frac{1}{\mathbb{E}X} = \frac{1}{6}$ = hourly rate of breakdowns

Properties of the Expectation:

Let $c = \text{const}$

1) $\mathbb{E}c = c$

2) $\mathbb{E}[cX] = c\,\mathbb{E}X$

3) $\mathbb{E}[X+Y] = \mathbb{E}X + \mathbb{E}Y$

skip

Ex. $T \sim \text{Gamma}(r, \lambda)$ sum of exponentials

r ↓ integer ⇒ Erlang distribution

We saw that $T = X_1 + X_2 + \dots + X_r$, X_i iid $\text{Exp}(\lambda)$

$$\Rightarrow \mathbb{E}T = \frac{r}{\lambda}$$

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

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

$$\left(\mathbb{E}X = \int_0^{\infty} \mathbb{P}(X>x)\,dx\right.$$ ← using tail probability (Proof p. 577)

Ex. $X \sim \text{Exp}(\lambda)$

$$\mathbb{E}X = \int_0^{\infty} \mathbb{P}(X>x)\,dx = \int_0^{\infty} e^{-\lambda x}\,dx = \frac{1}{\lambda}$$

Table of Means (10.1) p. 568

Family	Parameters	Expected value
Uniform	a, b	$\frac{(a+b)}{2}$
derived Exponential	λ	$\frac{1}{\lambda}$
Normal	μ, σ^2	μ
Gamma	α, λ	$\frac{\alpha}{\lambda}$
Erlang	r, λ	$\frac{r}{\lambda}$
Chi-squared	$\nu = df$	ν

Variance

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

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

$$= \int_{-\infty}^{\infty} x^2 f_X(x)\,dx - \left(\int_{-\infty}^{\infty} x f_X(x)\,dx \right)^2$$

$\text{Var}\, X = \sigma_X^2$

Standard deviation:

$\sigma_X = \sqrt{\text{Var}\, X} = \sqrt{\sigma_X^2}$

Properties:

Let $a = \text{const}$

1) $\text{Var}(aX + b) = a^2\, \text{Var}\, X$

2) $\text{Var}(X+Y) = \text{Var}\, X + 2\underbrace{\text{Cov}(X,Y)}_{\text{Covariance}} + \text{Var}\, Y$

Def $\text{Cov}(X,Y) = \mathbb{E}[(X - \mathbb{E}X)(Y - \mathbb{E}Y)]$

$$= \mathbb{E}XY - (\mathbb{E}X)(\mathbb{E}Y)$$

$$= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy\, f_{XY}(x,y)\,dx\,dy - \left(\int_{-\infty}^{\infty} x f_X(x)\,dx\right)\left(\int_{-\infty}^{\infty} y f_Y(\right.$$

Correlation coefficient

Def $\rho = \dfrac{\text{Cov}(X,Y)}{\sigma_X \cdot \sigma_Y}$ $\qquad -1 \leq \rho \leq 1$

- If X and Y are indep. $\Rightarrow$ $\rho = 0$, $\text{Cov}(X,Y) = 0$, and $\text{Var}(X+Y) = \text{Var}\, X + \text{Var}\, Y$

Conditional Expectation

Def $\mathbb{E}[Y \mid X = x] = \int_{-\infty}^{\infty} y \cdot \underbrace{f_{Y|X}(y|x)}_{\text{density of } Y \text{ when we know } X = x}\, dy$

Law of total expectation

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

Conditional Variance

Def $\quad \mathrm{Var}(Y \mid X = x) = \mathbb{E}[(Y - \mathbb{E}[Y \mid X = x])^2 \mid X = x] =$

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

Law of Total Variance

$$\mathrm{Var}(Y) = \mathbb{E}[\mathrm{Var}(Y \mid X)] + \mathrm{Var}(\mathbb{E}[Y \mid X])$$

10.81 Let X and Y be independent random variables having PDFs and finite variances.

a) Without doing any computations, obtain $\mathcal{E}(Y \mid X = x)$ for each possible value x of X. Explain your reasoning.
b) Use Definition 10.2 on page 596 to obtain $\mathcal{E}(Y \mid X = x)$ for each possible value x of X.
c) Without doing any computations, obtain $\mathrm{Var}(Y \mid X = x)$ for each possible value x of X. Explain your reasoning.
d) Use Definition 10.3 on page 596 to obtain $\mathrm{Var}(Y \mid X = x)$ for each possible value x of X.
e) Use part (b) to show that $\mathcal{E}(\mathcal{E}(Y \mid X)) = \mathcal{E}(Y)$, thus verifying the law of total expectation in this case.
f) Use part (d) to show that $\mathcal{E}(\mathrm{Var}(Y \mid X)) + \mathrm{Var}(\mathcal{E}(Y \mid X)) = \mathrm{Var}(Y)$, thus verifying the law of total variance in this case.

10.82 Let X have the standard normal distribution and let $-1 < \rho < 1$. Suppose that, for each $x \in \mathcal{R}$, the conditional distribution of Y given $X = x$ is a normal distribution with parameters ρx and $1 - \rho^2$.

a) Without doing any computations, obtain $\mathcal{E}(Y \mid X)$ and $\mathrm{Var}(Y \mid X)$.
b) Use your results from part (a) and the laws of total expectation and total variance to determine $\mathcal{E}(Y)$ and $\mathrm{Var}(Y)$.
c) Show that your results from part (b) are consistent with those that would be obtained by using the PDF of Y found in Example 9.13(b) on page 517.

10.83 A bus is scheduled to arrive at a bus stop within the next hour. Starting from now, the amount of time, in hours, until the bus actually arrives has a beta distribution with parameters 2 and 1. The number of passengers who have arrived at the bus stop t hours from now has a Poisson distribution with parameter λt, independent of when the bus arrives. Determine the mean and variance of the number of passengers at the bus stop when the bus arrives.

10.84 Suppose that X and Y are independent random variables.

a) What do you think is the best predictor of Y, based on observing X?
b) Use the prediction theorem to determine the best predictor of Y, based on observing X.

10.85 Suppose that Y is some function of X—say, $Y = g(X)$.

a) What do you think is the best predictor of Y, based on observing X?
b) Use the prediction theorem to determine the best predictor of Y, based on observing X.

10.86 Show that the minimum mean square error for prediction equals $\mathcal{E}(\mathrm{Var}(Y \mid X))$.

10.87 Fix a number $c \in (0, 1)$ and divide the interval $(0, 1)$ randomly into two subintervals.

a) What is the expected length of the subinterval that contains c?
b) What value of c maximizes the expected length of the subinterval that contains c?

10.88 In a textile mill, the number of defects per yard of fabric has a Poisson distribution with mean λ. However, λ varies from loom to loom and can be thought of as a random variable Λ whose probability distribution is concentrated on the interval $(0, 3)$. Determine the mean and variance of the number of defects per yard when

a) $\Lambda \sim \mathcal{U}(0, 3)$.
b) $\Lambda \sim \mathcal{T}(0, 3)$.

10.89 Two types of devices—say, Type 1 and Type 2—are used at a factory. The lifetime of a Type k device has mean μ_k and variance σ_k^2. A bin contains $100p\%$ Type 1 devices and the rest Type 2 devices. If one device is selected at random from the bin, what are the mean and variance of its lifetime?

Read Ch 11.

Moment Generating Function (MGF)

Def The MGF of a RV X is given by

$$\boxed{M_X(t) = \mathbb{E}[e^{tX}]}, \quad t \in \mathbb{R}$$

Remarks:

1. • If X is continuous, $M_X(t) = \int_{-\infty}^{\infty} e^{tx} f_X(x)\,dx$

   • If X is discrete, $M_X(t) = \sum_{R_X} e^{tx} \cdot \mathbb{P}_X(x)$.

2. MGF is $\boxed{\text{not}}$ random; for each t we get a number.

3. The MGF is defined for all t, such that $\mathbb{E}e^{tX} < \infty$.

4. The MGF needs to be defined at least for all t in an interval around the origin.

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

$$\boxed{M_X(t)} = \mathbb{E}\, e^{tX} = \sum_{x=0}^{n} e^{tx} \underbrace{\binom{n}{x} p^x (1-p)^{n-x}}_{\mathbb{P}_X(x)} =$$

$$= \sum_{x=0}^{n} \binom{n}{x} (pe^t)^x (1-p)^{n-x} =$$

binomial theory

$$= \boxed{(pe^t + 1 - p)^n}$$

Ex. $X \sim Exp(\lambda)$

$$\boxed{M_X(t)} = \mathbb{E}e^{tX} = \int_0^{\infty} e^{tx} \cdot \underbrace{\lambda e^{-\lambda x}}_{PDF}\,dx = \lambda \int_0^{\infty} e^{(t-\lambda)x}\,dx = \frac{\lambda}{t-\lambda} e^{(t-\lambda)x}\Big|_0^{\infty}$$

• since $\lambda > 0$

assume $t < \lambda$

$$= \frac{-\lambda}{t-\lambda} = \boxed{\frac{\lambda}{\lambda - t}, \quad t < \lambda}$$

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

$$M_X(t) = e^{\mu t + \frac{1}{2}\sigma^2 t^2} \quad , \ t \in \mathbb{R}$$

- if $X \sim N(0,1)$: $\quad M_X(t) = e^{\frac{1}{2}t^2} = e^{\frac{t^2}{2}}$

What is the MGF good for?

1. Generates all the moments: $\mathbb{E}X, \mathbb{E}X^2, \mathbb{E}X^3, \dots$
2. Identifies the distribution
   * like a signature
   * cannot be the same for two distributions
3. Used to prove the Central Limit Theorem

Getting the moments: $\mathbb{E}X^r$, $r = 1, 2, \dots$

$$(M_X(t))' = \frac{dM_X(t)}{dt} = \frac{d\,\mathbb{E}e^{tX}}{dt} = \mathbb{E}\left[\frac{de^{tX}}{dt}\right] = \mathbb{E}[X \cdot e^{tX}]$$

setting $t=0$ (after taking the derivative)

$$M_X'(0) = \mathbb{E}X.$$

$$(M_X(t))'' = \frac{d\,M_X'(t)}{dt} = \frac{d\,\mathbb{E}[X \cdot e^{tX}]}{dt} = \mathbb{E}\left[\frac{d(X \cdot e^{tX})}{dt}\right] = \mathbb{E}[X^2 \cdot e^{tX}]$$

$M_X''(0) = \mathbb{E}[X^2]$ (second moment)

$$\Rightarrow \mathbb{E}X^r = \left.\frac{d^r M_X(t)}{dt^r}\right|_{t=0}$$

Ex. $X \sim \mathcal{E}xp(\lambda)$

$M_X(t) = \frac{\lambda}{\lambda - t}$ $\underbrace{\text{for } t<\lambda}_{\text{need to take derivative on a neighborhood}}$

$M'_X(t) = \lambda \frac{1}{(\lambda-t)^2} \Rightarrow M'_X(0) = \boxed{\frac{1}{\lambda} = \mathbb{E}X}$ first moment (mean)

$M''_X(t) = 2\lambda \frac{1}{(\lambda-t)^3} \Rightarrow M''_X(0) = \boxed{\frac{2}{\lambda^2} = \mathbb{E}X^2}$ second moment

we are interested in variance:

$\operatorname{Var} X = \mathbb{E}X^2 - (\mathbb{E}X)^2 = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}$

Multiplication property of the MGF

If $X_1, X_2, \dots, X_n$ are indep. RVs,

then $M_{\underbrace{X_1+\dots+X_n}_{\text{sum}}}(t) = \underbrace{M_{X_1}(t)\dots M_{X_n}(t)}_{\text{product}}$

Proof ($n=2$):

$\boxed{M_{X_1+X_2}(t)} = \mathbb{E}[e^{(X_1+X_2)t}] = \mathbb{E}[e^{X_1 t} \cdot e^{X_2 t}] \overset{\text{indep.}}{=} \mathbb{E}e^{X_1 t}\,\mathbb{E}e^{X_2 t} =$

$= \boxed{M_{X_1}(t) \cdot M_{X_2}(t)}$

Ex. Sum of independent normal

$X_1 \sim N(\mu_1, \sigma_1^2)$
$X_2 \sim N(\mu_2, \sigma_2^2)$ indep.

$M_{X_1+X_2}(t) = \left(e^{\mu_1 t + \frac{1}{2}\sigma_1^2 t}\right)\left(e^{\mu_2 t + \frac{1}{2}\sigma_2^2 t}\right) = e^{(\mu_1+\mu_2)t + \frac{1}{2}(\sigma_1^2+\sigma_2^2)t}$

Recall, MGF characterizes the distribution uniquely

$\Rightarrow (X_1+X_2) \sim N(\mu_1+\mu_2, \sigma_1^2+\sigma_2^2)$.

<u>Ex.</u> $X \sim \text{Binomial}(n, p)$

$$\mathbb{E}e^{tX} = (e^t p + 1 - p)^n$$

$\boxed{n=1}$ $\mathbb{E}e^{tX} = (e^t p + 1 - p) \sim \text{Bernoulli}(p)$

$$\boxed{\text{If } X_1, \dots, X_n \text{ are iid} \Rightarrow M_{X_1 + \cdots + X_n}(t) = (M_X(t))^n}$$

skip

<u>Ex.</u> MGF of $\text{Exp}(\lambda)$

$$M_X(t) = \frac{\lambda}{\lambda - t}, \quad t < \lambda$$

$\Rightarrow$ MGF of $\underbrace{\text{Gamma}(r, \lambda)}_{\text{sum of iid Exponentials}}$

$$\Rightarrow M_X(t) = \left(\frac{\lambda}{\lambda - t}\right)^r, \quad t < \lambda$$

$$X \sim \text{Gamma}(\alpha, \lambda) \Rightarrow M_X(t) = \left(\frac{\lambda}{\lambda - t}\right)^\alpha, \quad t < \lambda.$$

c) Suppose that X and Y are independent random variables with $X \sim \mathcal{P}(\lambda)$ and $Y \sim \mathcal{P}(\mu)$. Use moment generating functions to show that $X + Y \sim \mathcal{P}(\lambda + \mu)$.

d) Extend the result of part (c) for m independent Poisson random variables.

11.3 Let X have the geometric distribution with parameter p.

a) Determine the MGF of X, including where it's defined.

b) Use the result of part (a) to obtain the mean and variance of X.

11.4 Let X have the uniform distribution on the interval (a, b).

a) Determine the MGF of X.

b) Use the result of part (a) to obtain the mean and variance of X.

11.5 A company insures homes in three cities. Because sufficient distance separates the cities, it's reasonable to assume that the losses occurring among the three cities are independent random variables. The moment generating functions for the three loss distributions are $(1-2t)^{-3}$, $(1-2t)^{-2.5}$, and $(1-2t)^{-4.5}$. Determine the third moment of the combined losses from the three cities.

11.6 Let X be a Cauchy random variable. Does there exist an open interval containing 0 for which M_X is defined? Explain your answer.

11.7 In Example 9.20 on page 536, we showed that, if X and Y are independent random variables with $X \sim \Gamma(\alpha, \lambda)$ and $Y \sim \Gamma(\beta, \lambda)$, then $X + Y \sim \Gamma(\alpha + \beta, \lambda)$.

a) Use MGFs to establish the result referred to and compare your work with that required in Example 9.20.

b) Generalize the result of part (a) to m independent gamma random variables with the same second parameter, thus providing a simple proof of Proposition 9.13 on page 537.

11.8 Suppose that Y has the lognormal distribution with parameters μ and σ^2, which means that $\ln Y \sim \mathcal{N}(\mu, \sigma^2)$. Determine a formula for the nth moment of Y by

a) using a PDF of Y.

b) using the FEF and a PDF of a normal random variable.

c) recalling the formula for the MGF of a normal random variable.

d) For what values of t is the MGF of Y defined? Justify your answer.

e) As we mentioned on page 634, if the MGF of a random variable is defined (exists) in some open interval containing 0, the random variable has moments of all orders. Is the converse of this statement true? Explain your answer.

11.9 For each $n \in \mathcal{N}$, let X_n have the discrete uniform distribution on $\{-1 + 1/n, 1 - 1/n\}$.

a) Heuristically, what is the probability distribution of a random variable X to which $\{X_n\}_{n=1}^{\infty}$ converges in distribution?

b) Mathematically verify your heuristics in part (a) by showing that $M_{X_n}(t) \to M_X(t)$ as $n \to \infty$ for all $t \in \mathcal{R}$.

c) Mathematically verify your heuristics in part (a) by showing that $F_{X_n}(x) \to F_X(x)$ as $n \to \infty$ for all $x \in \mathcal{R}$ at which F_X is continuous.

11.10 Let $X_1, X_2, \ldots$ be independent random variables, each having a $\mathcal{U}(0, 1)$ distribution. For each $n \in \mathcal{N}$, let $U_n = \min\{X_1, \ldots, X_n\}$ and let $V_n = \max\{X_1, \ldots, X_n\}$.

a) Heuristically, what happens to the probability distribution of U_n as $n \to \infty$? Explain your reasoning.

b) Show that $\{U_n\}_{n=1}^{\infty}$ converges in distribution and identify its limiting distribution.

c) Repeat parts (a) and (b) for $\{V_n\}_{n=1}^{\infty}$.

11.11 Let $X_1, X_2, \ldots$ be independent and identically distributed random variables, each having mean μ, variance σ^2, and moment generating function M defined for $-t_0 < t < t_0$.

HW 11

X 1) 8.164 p.475 + extra text

_ 2) my text

✓ 3) 9.26 p.500

✓ 4) 9.45 p.507

✓ 5) 9.64 c) p.519

✓ 6) 9.91 p.528 (hint)

Extra credit:

X 1) 9.77 p.521

8.159 Rayleigh random variable: A random variable R with PDF given by

$$f_R(r) = \frac{r}{\sigma^2}\, e^{-r^2/2\sigma^2}, \qquad r > 0,$$

and $f_R(r) = 0$ otherwise, is called a *Rayleigh random variable* and is said to have the *Rayleigh distribution with parameter* σ^2. Obtain and identify a PDF of the random variable R^2.

8.160 Let $Y = \sigma\sqrt{X}$, where X has the chi-square distribution with n degrees of freedom and σ is a positive real number.

a) Determine a PDF of the random variable Y.

b) Show that, if $n = 1$, then Y has the distribution of the absolute value of a $\mathcal{N}(0, \sigma^2)$ random variable.

c) Show that, if $n = 2$, then Y has the Rayleigh distribution with parameter σ^2, as defined in Exercise 8.159.

d) Maxwell random variable: A random variable S with PDF given by

$$f_S(s) = \frac{\sqrt{2/\pi}}{\sigma^3} s^2 e^{-s^2/2\sigma^2}, \qquad s > 0,$$

and $f_S(s) = 0$ otherwise, is called a *Maxwell random variable* and is said to have the *Maxwell distribution with parameter* σ^2. Show that, if $n = 3$, then Y has the Maxwell distribution with parameter σ^2.

8.161 Suppose that X has the standard Cauchy distribution. Determine and identify the probability distribution of the random variable $1/(1 + X^2)$.

8.162 Let U be uniformly distributed on the interval $(0, 1)$.

a) Without doing any computations, make an educated guess at the probability distribution of the random variable $1 - U$. Explain your reasoning.

b) Determine and identify the probability distribution of $1 - U$.

8.163 Let $X \sim \mathcal{U}(a, b)$ and let $U \sim \mathcal{U}(0, 1)$. Use Proposition 8.16 on page 471 to show that

a) $(X - a)/(b - a) \sim \mathcal{U}(0, 1)$. **b)** $a + (b - a)U \sim \mathcal{U}(a, b)$.

X **8.164 Simulation:** This exercise requires access to a computer or statistical software.

a) Explain how to use Exercise 8.163(b) to simulate 5000 observations of a uniform random variable on the interval (a, b) by using only a basic random number generator.

b) Conduct the simulation described in part (a) when $a = -0.5$ and $b = 0.5$.

c) Provide a graph of your results in part (b). Is the graph what you would expect? Explain.

X **8.165** Let $X \sim \mathcal{U}(0, 1)$ and let m and n be integers with $m < n$. Determine and identify the probability distribution of the random variable $\lfloor m + (n - m + 1)X \rfloor$, where $\lfloor x \rfloor$ denotes the *floor function,* the greatest integer smaller than or equal to x.

8.166 Simulation: This exercise requires access to a computer or statistical software.

a) Let m and n be integers with $m < n$. Explain how to simulate 10,000 observations of a discrete uniform random variable on the set $\{m, m + 1, \ldots, n\}$ by using only a basic random number generator. *Note:* Refer to Exercise 8.165.

b) Conduct the simulation described in part (a) when $m = 1$ and $n = 8$.

c) Provide a graph of your results in part (b). Is the graph what you would expect? Explain.

Theory Exercises

8.167 Show that Proposition 8.10 (page 441) is a special case of Proposition 8.15 (page 468).

8.168 Prove the univariate transformation theorem, Proposition 8.14 on page 466, in the case where g is strictly increasing on the range of X.

HW 10

1) 8.164 p. 475
You can search on the web for a random number generator.
For part c), provide a graph of the empirical CDF.

2) You want to simulate three values of a standard normal RV.
Suppose you have generated the values
0.9838, 0.5, 0.0616 of Unif (0, 1).
What are the three corresponding values of the standard normal random variables?
(Hint: use CDF table on p. A-39.)
Explain your method.

3) 9.26 p. 500

4) 9.45 p. 507

5) 9.64 c) p. 519

*6) 9.77 p. 521 extra credit

7) 9.91 p. 528 ("at random" = uniform)

✓8) 9.150 p. 551

✓9) 9.154 p. 551

✓10) 9.122 p. 543

HW 12 summer 1, 2015
day session

1) 9.150 p.551
2) 9.154 p.551, 552
3) 9.122 p.543
4) 10.82 a), b) p.605

HW 11

1) 10.82 a), b) p. 605
2) 11.2 p. 639
3) 11.3 p. 639
4) 11.4 p. 639 ← part a) only
5) 11.5 p. 639

Extra credit:
1) 11.6 p. 639

Moments of a distribution

Def Let $r = 1, 2, 3, \dots$ integer

The r^{th} moment of a RV X is given by

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

· special cases :

$r=1$: $\mathbb{E}[X] = \mu$ = mean = first moment

$r=2$: $\mathbb{E}[X^2]$ = second moment

$$\Rightarrow \operatorname{Var} X = \underbrace{\mathbb{E}[X^2]}_{\substack{\text{second}\\ \text{moment}}} - \underbrace{(\mathbb{E}X)^2}_{\substack{\text{first moment}\\ \text{squared}}}$$

Note: $\operatorname{Var} X = \underbrace{\mathbb{E}[(X - \mathbb{E}X)^2]}_{\text{a.k.a. second central moment}}$

Ex. $X \sim \text{Unif}(0,1)$

$$\mathbb{E}X^r = \int_0^1 x^r \cdot \frac{1}{1-0}\, dx = \frac{x^{r+1}}{r+1}\Big|_0^1 = \frac{1}{r+1} .$$

Expected Value

Notation: $\mathbb{E}[X] = \mathbb{E}X$ = expected value of the RV X

= mean of X

Def $$\mathbb{E}X = \sum_{x \in R_X} \underbrace{x \cdot p(x)}_{\text{weight the value by its probability}} = \sum_{x \in R_X} \underbrace{x}_{\text{values}} \cdot \underbrace{\mathbb{P}(X=x)}_{\text{weights}}$$

Note: $\mathbb{E}X$ is <u>not</u> random (you can compute it).

Ex. X = number showing on a fair die

x	1	2	3	4	5	6
$\mathbb{P}(X=x)$	1/6	1/6	1/6	1/6	1/6	1/6

$$\mathbb{E}X = 1\cdot\frac{1}{6} + 2\cdot\frac{1}{6} + \cdots + 6\cdot\frac{1}{6} = \frac{1}{6}\cdot(1+6)\cdot\frac{6}{2} = 3.5$$

may not belong to

Ex. $X \sim \text{Bern}(p)$

$$X = \begin{cases} 1, & \text{with prob. } p \\ 0, & \text{with prob. } 1-p \end{cases} \qquad \mathbb{E}X = 1\cdot p + 0\cdot(1-p) = p$$

Ex. Indicator function of a set A

Def $$\mathbb{1}_A = \begin{cases} 1, & \text{if } A \text{ occurs} \\ 0, & \text{if } A \text{ does not occur} \end{cases}$$

$\Rightarrow \mathbb{1}_A \sim$ Bernoulli with prob. of success $p = \mathbb{P}(\mathbb{1}_A = 1) = \mathbb{P}(A)$

$$\mathbb{E}\mathbb{1}_A = 1\cdot\mathbb{P}(A) + 0\cdot(1-\mathbb{P}(A)) = \mathbb{P}(A)$$

Ex. 5 $\quad X \sim \text{Poiss}(\lambda)$

$$\mathbb{P}_X(x) = \frac{e^{-\lambda} \cdot \lambda^x}{x!}, \qquad \begin{array}{l} \lambda > 0 \\ x = 0, 1, 2, \dots \end{array}$$

$$\mathbb{E}X = \sum_{x=0}^{\infty} x \cdot e^{-\lambda} \cdot \frac{\lambda^x}{x!} = \sum_{x=1}^{\infty} x e^{-\lambda} \cdot \frac{\lambda^x}{x!} =$$

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

$$= e^{-\lambda} \cdot \lambda \sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^{-\lambda} \cdot \lambda \cdot e^{\lambda} = \underline{\lambda}$$

skip

Table of Means p. 332

Distr.	Param.	Mean
Bernoulli	p	p – derived
Binomial	n, p	np – derived on later ~~next~~ pg.
~~Hypergeometric~~	~~N, n, p~~	np ~~X~~ – skip
Poisson	λ	λ – derived
Geometric	p	$1/p$ – later with tail prob.
~~Pascal~~	~~p, r~~	r/p ~~X~~ – skip

Fundamental Expected Value Formula

(Recall: $\mathbb{E}X = \sum_{x \in R_X} x \cdot \mathbb{P}(X=$

Ex. $X = \begin{cases} -1 & \text{with prob. } 1/3 \\ 0 & \text{with prob. } 1/3 \\ 1 & \text{with prob. } 1/3 \end{cases}$

$Y = X^2 = \begin{cases} 0 & \text{with prob. } 1/3 \\ 1 & \text{with prob. } 2/3 \end{cases}$

$\mathbb{E}Y = ?$

Method 1: use PMF of Y

$$\mathbb{E}Y = \sum_{y=0}^{1} y \cdot \mathbb{P}(Y=y) = 0 \cdot \frac{1}{3} + 1 \cdot \frac{2}{3} = \boxed{\frac{2}{3}}$$

Method 2: use PMF of X

$$\mathbb{E}Y = \mathbb{E}X^2 = \sum_{x=-1}^{1} x^2 \cdot \mathbb{P}(X=x) = (-1)^2 \cdot \frac{1}{3} + 0^2 \cdot \frac{1}{3} + 1^2 \cdot \frac{1}{3} = \boxed{\frac{2}{3}}$$

In general, suppose we want to compute $\mathbb{E}g(X)$:

$$\mathbb{E}g(X) = \sum_{x \in R_X} g(x) \cdot \mathbb{P}(X=x) \leftarrow \text{fundamental expected value formula.}$$

$$\mathbb{E}g(X_1, X_2) = \sum_{x_1} \sum_{x_2} g(x_1, x_2) \cdot \mathbb{P}_{X_1 X_2}(x_1, x_2)$$

Ex. n questions in exam

Drunk $\rightarrow$ guess

$X = \#$ of correct answers $\sim \text{Bin}(n, p)$

Let $g(X) = e^X$

$$\mathbb{E}e^X = \sum_{x=0}^{n} e^x \binom{n}{x} p^x (1-p)^{n-x} = \sum_{x=0}^{n} (ep)^x \binom{n}{x} (1-p)^{n-x} =$$

let $a = ep$, $b = 1-p$ } use Binomial theory

$$= (ep + (1-p))^n$$

Ex. $p = \frac{1}{2}$, $n = 10$ $\Rightarrow \mathbb{E}e^X = \left(e\frac{1}{2} + 1 - \frac{1}{2}\right)^{10} = \left(\frac{e+1}{2}\right)^{10}$

Properties of the Expectation:

1. $c = \text{const}$ $\quad \mathbb{E}c = c$

$$\mathbb{E}c = \sum_x c \cdot \mathbb{P}(X=x) = c \overbrace{\sum_x \mathbb{P}(X=x)}^{1} = c \cdot 1 = c.$$

2. $\mathbb{E}[cX] = c\,\mathbb{E}X$

3. $\mathbb{E}[X+Y] = \mathbb{E}X + \mathbb{E}Y$

Ex. $\mathbb{E}[3X - 4Y + 6Z - 2] = 3\mathbb{E}X - 4\mathbb{E}Y + 6\mathbb{E}Z - 2$

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

$\mathbb{E}X = np \leftarrow$ we will derive it using the following interpretation:

$Y_i \sim \text{Bernoulli}(p)$, $i = 1, \dots, n$ independent

$$Y_i = \begin{cases} 1 & \text{with prob. } p \\ 0 & \text{with prob. } 1-p \end{cases}$$

$X = Y_1 + Y_2 + \dots + Y_n$ (sum of the 1's = # of successes)

$$\mathbb{E}X = \mathbb{E}[Y_1 + \dots + Y_n] = \underset{\sim}{\mathbb{E}Y_1} + \dots + \mathbb{E}Y_n = np$$

$$\mathbb{E}Y_1 = 1p + 0(1-p) = p$$

Ex. $X \sim$ Hypergeometric
(sampling without replacement)

deep

Population size N

$M = \#$ of people with red hair $= Np \Rightarrow p = \frac{M}{N}$

$X = \#$ of people with red hair in the sample

Sample size $n < N$

$X = \sum_{i=1}^{n} Y_i$, where $Y_i = \begin{cases} 1 & \text{if } i^{th} \text{ person has red hair} \\ 0 & \text{if he doesn't} \end{cases}$

$Y_i \sim$ Bernoulli(p), but <u>not</u> independent

$\mathbb{E} X = \mathbb{E} Y_1 + \cdots + \mathbb{E} Y_n = np \Rightarrow$ with or without replacement we will get the same mean value.

<u>Ex.</u> Estimation of the unknown mean of a population

Take a sample of size n from the population with mean μ.

$\overbrace{X_1, X_2, \dots, X_n}$ with replacement.

$\mathbb{E} X_1 = \mathbb{E} X_2 = \cdots = \mathbb{E} X_n = \mu$

$\overline{X}_n = \frac{X_1 + \cdots + X_n}{n}$

$\hookrightarrow$ average (random variable)

mean of the average

$\mathbb{E} \underbrace{\overline{X}_n}_{\text{random}} = \mathbb{E}\left(\frac{1}{n}(X_1 + \cdots + X_n)\right) = \frac{1}{n} \mathbb{E}(X_1 + \cdots + X_n) = \frac{1}{n} \cdot (\mathbb{E} X_1 + \cdots + \mathbb{E} X_n) =$

not random $= \frac{1}{n} \cdot n\mu = \mu$

$\overline{X}_n$ is said to be an <u>unbiased</u> estimator for μ.

- $\mathbb{E}[X_1 + X_2] = \mathbb{E} X_1 + \mathbb{E} X_2$ (the expectation of the sum is the sum of the expectations)
- $\mathbb{E}[X_1 \cdot X_2] = ?$ If X_1 and X_2 are <u>indep.</u> then $\mathbb{E}[X_1 X_2] = \mathbb{E}[X_1] \cdot \mathbb{E}[X_2]$.

Ex. Let $X_1, \dots, X_n$ be $\underbrace{\text{iid}}$

$\sim \text{Bern}(p)$ independent and identically distributed

Let $Z = X_1 X_2 \dots X_n$

lethod 1: $\mathbb{E}Z = (\mathbb{E}X_1).(\mathbb{E}X_2)\dots(\mathbb{E}X_n) = p^n$

Method 2: Distribution of Z?

$$Z = \begin{cases} 1 & \text{with prob. } p^n \\ 0 & \text{with prob. } (1-p^n) \end{cases}$$

$\Rightarrow Z \sim \text{Bernoulli}(p^n) \Rightarrow \mathbb{E}Z = p^n.$

Using tail probability to compute Expectation

Let X be a random variable with range R_X:

$R_X = \{0, 1, 2, 3, \dots\}$

Then $$\mathbb{E}X = \sum_{n=0}^{\infty} \underbrace{\mathbb{P}(X > n)}_{\text{tail probability}}$$

Let $p_n = \mathbb{P}(X=n)$

$$\mathbb{E}X = \sum_{n=0}^{\infty} n.\mathbb{P}(X=n) = \sum_{n=1}^{\infty} n.\mathbb{P}(X=n) = \sum_{n=1}^{\infty} n.p_n$$

$$\begin{aligned} \mathbb{E}X = p_1 + p_2 + p_3 &+ \dots \\ + p_2 + p_3 &+ \dots \\ + p_3 &+ \dots \\ &+ \dots \end{aligned}$$

$\boxed{\text{sum} \downarrow \quad \sum_{n=1}^{\infty} n.p_n}$ $\boxed{\text{sum} \rightarrow \quad \sum_{n=0}^{\infty} \mathbb{P}(X>n)}$

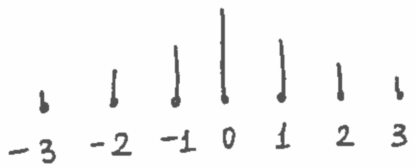
Ex. $X \sim \text{Geom}(p)$

$R_X = \{1, 2, \dots\}$ use the tail probability to compute $\mathbb{E}X$

$$\mathbb{E}X = \sum_{n=0}^{\infty} \mathbb{P}(X > n) = \sum_{n=0}^{\infty} (1-p)^n = \frac{1}{1-(1-p)} = \boxed{\frac{1}{p}}$$

• Variance

X  $\mathbb{E}X = 0$

Y  $\mathbb{E}Y = 0$

same mean, but different shape of the distr., the spread of the data is different.

Measures of dispersion

- $\mathbb{E}|X - \mathbb{E}X|$
- $\sqrt{\mathbb{E}(X - \mathbb{E}X)^2}$ standard deviation

Definition of the Variance

$$\mathrm{Var}(X) = \underset{\uparrow \atop \text{mean}}{\mathbb{E}}\Big[\underbrace{(X - \underset{\text{mean}}{\mathbb{E}X})^2}_{\text{squared deviation from the mean}}\Big]$$

Practical Formula: $\mathrm{Var}(X) = \mathbb{E}[X^2] - (\mathbb{E}X)^2$

Proof: $\mathrm{Var}(X) = \mathbb{E}(X - \mathbb{E}X)^2 = \mathbb{E}[X^2 - 2X\mathbb{E}X + \underset{\text{const}}{(\mathbb{E}X)^2}] =$

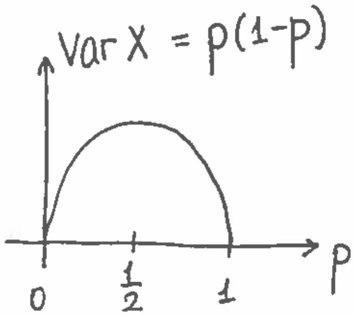
$= \mathbb{E}[X^2] - 2(\mathbb{E}X)^2 + (\mathbb{E}X)^2 = \mathbb{E}[X^2] - (\mathbb{E}X)^2$

Ex. $X \sim \text{Bernoulli}(p)$

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

$\mathbb{E}X^2 = 1^2 \cdot p + 0^2(1-p) = p$

$\mathrm{Var}(X) = p - p^2 = p(1-p)$

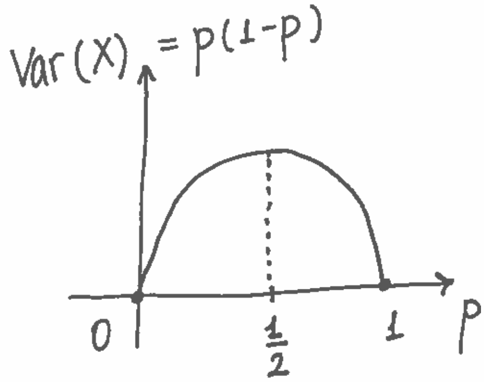


Ex. $X \sim \text{Bernoulli}(p)$

$$\mathbb{E}X = 1 \cdot p + 0 \cdot (1-p) = p$$

$$\mathbb{E}[X^2] = 1^2 \cdot p + 0^2 \cdot (1-p) = p$$

$$\Rightarrow \text{Var}(X) = p - p^2 = p(1-p)$$



Properties of the Variance

$c = \text{const}$

1. $\text{Var}(c) = 0$
2. $\text{Var}(cX) = c^2 \cdot \text{Var}(X)$

$$\begin{aligned}\text{Var}(cX) &= \mathbb{E}[(cX - \mathbb{E}[cX])^2] \\ &= \mathbb{E}[(cX - c \cdot \mathbb{E}X)^2] \\ &= \mathbb{E}[c^2 \cdot (X - \mathbb{E}X)^2] \\ &= c^2 \cdot \mathbb{E}[(X - \mathbb{E}X)^2]\end{aligned}$$

3. $\text{Var}(X + c) = \text{Var}(X)$

Ex. $\text{Var}(3X - 6) = 9\,\text{Var}(X)$

Def Standard Deviation of X

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

Ex. $X \sim \text{Bernoulli}(p)$

$\text{Var}(X) = p(1-p)$

$\sigma_X = \sqrt{p(1-p)}$

e.g. $p = \frac{1}{2} \Rightarrow \text{Var}(X) = \frac{1}{2}(1 - \frac{1}{2}) = \frac{1}{4} \Rightarrow \sigma_X = \frac{1}{2}$

Notation: $\text{Var}(X) = \sigma_X^2$.

Standardized Random Variable

Let X be a RV with mean μ_X and stdev σ_X

$\mu_X = \mathbb{E}X \longrightarrow$ mean

$\sigma_X = \sqrt{Var(X)} \longrightarrow$ standard deviation

Def The standardized RV corresponding to X is the new RV X^*, such that

$$\boxed{X^* = \frac{X - \mu_X}{\sigma_X}}.$$

Then $\mathbb{E}X^* = 0$ and $\sigma_{X^*} = 1$.

Proof: $\mathbb{E}[X^*] = \mathbb{E}\left[\frac{X-\mu_X}{\sigma_X}\right] = \frac{1}{\sigma_X}\big(\underbrace{\mathbb{E}X}_{\mu_X} - \mu_X\big) = 0.$

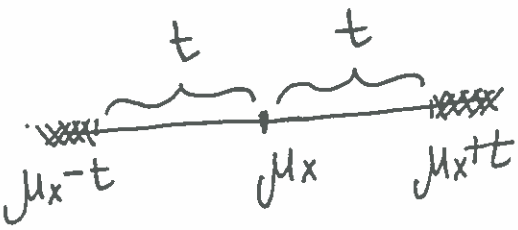
$$Var(X^*) = Var\left(\frac{X-\mu_X}{\sigma_X}\right) = \left(\frac{1}{\sigma_X}\right)^2 \cdot Var(X-\mu_X) =$$

$$= \frac{1}{\sigma_X^2} \cdot \underbrace{Var(X)}_{\sigma_X^2} = 1.$$

skip

Chebyshev's Inequality

$$\mathbb{P}\big(|X - \mu_X| \geq t\big) \leq \frac{Var X}{t^2}$$



Proof: $Var(X) = \sum_{x \in R_X} (x-\mu_X)^2 \cdot \mathbb{P}(X=x)$

$$\geq \sum_{\{|x-\mu_X| \geq t\}} (x-\mu_X)^2 \cdot \mathbb{P}(X=x)$$

this set is no bigger than R_X

$$\geq \sum_{|x-\mu_X| \geq t} t^2 \cdot \mathbb{P}(X=x) = t^2 \cdot \overbrace{\sum_{|x-\mu_X| \geq t} \mathbb{P}(X=x)}^{\mathbb{P}(|X-\mu_X| \geq t)}$$

this is no bigger than $(x-\mu_X)^2$

$$= t^2 \cdot \mathbb{P}(|X-\mu_X| \geq t).$$

Covariance of X and Y

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

$$= \mathbb{E}[XY] - (\mathbb{E}X).(\mathbb{E}Y)$$

Variance $\mathrm{Var}(X) = \mathbb{E}[(X-\mathbb{E}X)^2] = \mathbb{E}X^2 - (\mathbb{E}X)^2$

$$\sigma_X^2 = \mathbb{E}X^2 - \mu_X^2$$

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

$$= \mathbb{E}XY - \mu_X \cdot \mu_Y$$

Properties of the Covariance:

1. $\mathrm{Cov}(X,X) = \mathrm{Var}(X)$
2. $\mathrm{Cov}(X,Y) = \mathrm{Cov}(Y,X)$ $\qquad c = \text{const}$
3. $\mathrm{Cov}(cX,Y) = c.\mathrm{Cov}(X,Y)$
4. $\mathrm{Cov}(X+c,Y) = \mathrm{Cov}(X,Y)$
5. $\mathrm{Cov}(X,Y+Z) = \mathrm{Cov}(X,Y) + \mathrm{Cov}(X,Z)$

$$\Rightarrow \mathrm{Cov}\left(\sum_{j=1}^{n} a_j X_j, \sum_{k=1}^{m} b_k \cdot Y_k\right) = \sum_{j=1}^{n}\sum_{k=1}^{m} a_j b_k \cdot \mathrm{Cov}(X_j, Y_k)$$

EX.
- $\mathrm{Cov}(3X+6, -5Y-3) = \mathrm{Cov}(3X,-5Y) = -15.\mathrm{Cov}(X,Y)$
- $\mathrm{Var}(-5Y-3) = \mathrm{Var}(-5Y) = 25.\mathrm{Var}\,Y$

  take out & square it

- X and Y are Independent

  $\mathrm{Cov}(X,Y) = \mathbb{E}XY - \mathbb{E}X.\mathbb{E}Y = \mathbb{E}X.\mathbb{E}Y - \mathbb{E}X.\mathbb{E}Y = 0$

Note: $\mathrm{Cov} = 0 \not\Rightarrow$ indep. **Ex.** $X = \begin{cases} 2, & 1/4 \\ 1, & 1/4 \\ -1, & 1/4 \\ 2, & 1/4 \end{cases}$ $X^2 = \begin{cases} 1, & 1/2 \\ 4, & 1/2 \end{cases}$ $X^3 = \begin{cases} 8, & 1/4 \\ 1, & 1/4 \\ -1, & 1/4 \\ -8, & 1/4 \end{cases}$

$\mathrm{Cov}(X,X^2) = \underbrace{\mathbb{E}X^3}_{0} - \underbrace{\mathbb{E}X}_{0}.\underbrace{\mathbb{E}X^2}_{\neq 0} = 0$, but X and X^2 are <u>not</u> Indep.

Application of the Covariance

Q: $\mathbb{E}[X+Y] = ?$ **A**: $\mathbb{E}[X+Y] = \mathbb{E}[X] + \mathbb{E}[Y]$

Q: $\mathrm{Var}(X+Y) = ?$

$$\boxed{\mathrm{Var}(X+Y) = \mathrm{Var}(X) + 2\,\mathrm{Cov}(X,Y) + \mathrm{Var}(Y)}$$

special case: if X and Y indep. $\Rightarrow \mathrm{Cov}(X,Y) = 0$

$$\Rightarrow \mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$$

Proof: $\mathrm{Var}(X+Y) = \mathbb{E}[(X+Y-\underbrace{\mu_{X+Y}}_{\underbrace{\mathbb{E}(X+Y)}_{\mathbb{E}X+\mathbb{E}Y}})^2] =$

$$= \mathbb{E}[(X+Y-\mathbb{E}X-\mathbb{E}Y)^2] =$$

$$= \mathbb{E}[((X-\mathbb{E}X)+(Y-\mathbb{E}Y))^2] =$$

$$= \mathbb{E}[(X-\mathbb{E}X)^2 + 2(X-\mathbb{E}X)(Y-\mathbb{E}Y) + (Y-\mathbb{E}Y)^2] =$$

$$= \mathbb{E}[(X-\mathbb{E}X)^2] + 2\mathbb{E}[(X-\mathbb{E}X)(Y-\mathbb{E}Y)] + \mathbb{E}[(Y-\mathbb{E}Y)^2] =$$

$$= \mathrm{Var}(X) + 2\,\mathrm{Cov}(X,Y) + \mathrm{Var}(Y).$$

More generally:

$$\mathrm{Var}\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} \mathrm{Var}(X_i) + 2\sum_{i<j}\sum \mathrm{Cov}(X_i, X_j)$$

$$\mathrm{Var}(X+Y+Z) = \mathrm{Var}(X) + \mathrm{Var}(Y) + \mathrm{Var}(Z) + 2\,\mathrm{Cov}(X,Y) + 2\,\mathrm{Cov}(X,Z) + 2\,\mathrm{Cov}(Y,Z).$$

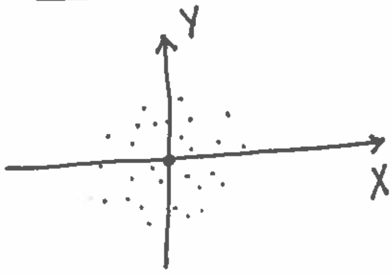
Graphical Representation of Cov(X, Y)

Random Sample $(X_1, Y_1), \dots, (X_n, Y_n)$

Suppose $\mu_X = 0,\ \mu_Y = 0$

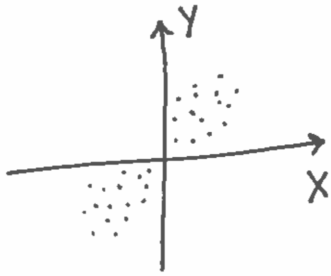
1) X, Y indep.

$\boxed{Cov(X,Y) = 0}$



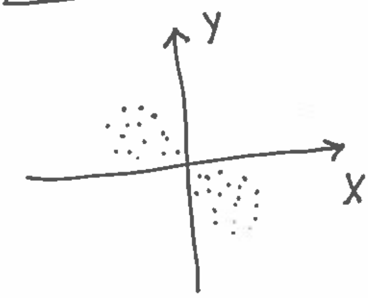
2) positive

$\boxed{Cov(X,Y) > 0}$



3) negative

$\boxed{Cov(X,Y) < 0}$



Correlation between X and Y $\quad \rho(X,Y)$

Suppose $Cov(X,Y) = 2560$ positively correlated

Best is to standardize: $X^* = \dfrac{X - \mu_X}{\sigma_X}$, $Y^* = \dfrac{Y - \mu_Y}{\sigma_Y}$

$(\mathbb{E}X^* = 0,\ Var(X^*) = 1,\ \mathbb{E}Y^* = 0,\ Var(Y^*) = 1)$

Def The correlation coefficient between X and Y is

$$\boxed{\rho(X,Y) = Cov(X^*, Y^*) = \mathbb{E}\left[\left(\frac{X - \mu_X}{\sigma_X}\right)\left(\frac{Y - \mu_Y}{\sigma_Y}\right)\right] = \frac{Cov(X,Y)}{\sigma_X \cdot \sigma_Y}}$$

Example 7.19 p. 373

Given $Cov(X,Y) = 0.224$, $Var(X) = 0.3376$, $Var(Y) = 0.56$

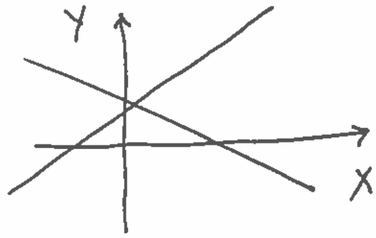
$$\Rightarrow \rho(X,Y) = \frac{0.224}{\sqrt{0.3376} \cdot \sqrt{0.56}} = \underline{\underline{0.515}}$$ bigger than 0.224

Fact: $-1 \leq \rho \leq 1$

Fact: if $Y = aX + b$ (Y is a linear function of X)

$\rho = 1$ if $a > 0$

$\rho = -1$ if $a < 0$



EXAMPLE 7.4 ***Expected Value***

The Lotto An Arizona state lottery, called *Lotto*, is played as follows: A player specifies six numbers of his or her choice from the numbers 1–42; these six numbers constitute the player's "ticket" for which the player pays \$1. In the lottery drawing, six winning numbers are chosen at random (without replacement) from the numbers 1–42. To win a prize, a *Lotto* ticket must contain three or more of the winning numbers. Suppose that the player buys one *Lotto* ticket per week.

a) Determine the expected number of weeks until the player wins a prize.
b) Determine the expected number of prizes that the player wins in a year.

Solution In Example 5.15(b) on page 212, we showed that, if a player buys one ticket, the probability she wins a prize is 0.029 (to three decimal places). Let a success correspond to the player winning a prize during any given week. As lottery results from one week to the next are independent, we have Bernoulli trials with success probability $p = 0.029$.

a) Let X denote the number of weeks until the player wins a prize. Then X has the geometric distribution with parameter $p = 0.029$. Referring to Table 7.4, we see that $\mathcal{E}(X) = 1/p = 1/0.029 = 34.483$. The expected number of weeks until the player wins a prize is roughly 34.5.
b) Let Y denote the number of prizes that the player wins in a year. Then Y has the binomial distribution with parameters $n = 52$ and $p = 0.029$. Again, from Table 7.4, $\mathcal{E}(Y) = np = 52 \cdot 0.029 = 1.508$. The expected number of prizes that the player wins in a year is roughly 1.5. ■

EXERCISES 7.1 Basic Exercises

7.1 Suppose that two balanced dice are rolled. Let X be the sum of the two faces showing. Determine the expected value of the random variable X. *Note:* Exercise 5.24 on page 193 asks for the PMF of X. 7

7.2 An archer shoots an arrow into a square target 6 feet on a side whose center we call the origin. Assume that the archer will actually hit the target and is equally likely to hit any portion of the target. The archer scores 10 points if she hits the bull's eye—a disk of radius 1 foot centered at the origin (region A); she scores 5 points if she hits the ring with inner radius 1 foot and outer radius 2 feet centered at the origin (region B); and she scores 0 points otherwise (region C). Determine the expected score by the archer. *Note:* Exercise 5.27 on page 194 asks for the PMF of the score.

7.3 Urn I contains four red balls and one white ball; Urn II contains one red ball and four white balls. A coin with probability p of a head is tossed. If the result is a head, Urn I is chosen; otherwise, Urn II is chosen. Two balls are then drawn at random without replacement from the chosen urn. Let X be the number of red balls drawn. Obtain the mean of X in case
a) $p = 0.5$ (balanced coin). **b)** $p = 0.4$ (biased coin).
Note: Exercise 5.28 on page 194 asks for the PMF of X in each case.

7.4 Of six men and five women applying for a job at Alpha, Inc., three are selected at random for interviews. Find the expected number of women in the interview pool by using
a) the PMF obtained in Exercise 5.29(a) on page 194.
b) an appropriate expected-value formula obtained in this section.

EXERCISES 7.3 Basic Exercises

Note: Several exercises in this section are continuations of those in Section 7.1.

V 7.54 Suppose that two balanced dice are rolled. Let X be the sum of the two faces showing. Determine the variance of the random variable X by using the
a) defining formula. b) computing formula.

Note: Exercise 7.1 on page 333 asks for the mean of X and also provides a reference for the PMF of X.

7.55 An archer shoots an arrow into a square target 6 feet on a side whose center we call the origin. Assume that the archer will actually hit the target and is equally likely to hit any portion of the target. The archer scores 10 points if she hits the bull's eye—a disk of radius 1 foot centered at the origin (region A); she scores 5 points if she hits the ring with inner radius 1 foot and outer radius 2 feet centered at the origin (region B); and she scores 0 points otherwise (region C). Determine the variance of the score by the archer. *Note:* Exercise 7.2 on page 333 asks for the mean of the score and also provides a reference for its PMF.

7.56 Urn I contains four red balls and one white ball; Urn II contains one red ball and four white balls. A coin with probability p of a head is tossed. If the result is a head, Urn I is chosen; otherwise, Urn II is chosen. Two balls are then drawn at random without replacement from the chosen urn. Let X be the number of red balls drawn. Obtain the variance of X in case
a) $p = 0.5$ (balanced coin). b) $p = 0.4$ (biased coin).

Note: Exercise 7.3 on page 333 asks for the mean of X and provides a reference for the PMF of X in each case.

7.57 Of six men and five women applying for a job at Alpha, Inc., three are selected at random for interviews. Determine the variance of the number of women in the interview pool by using
a) the PMF obtained in Exercise 5.29(a) on page 194.
b) an appropriate variance formula obtained in this section.

7.58 A variable of a finite population has standard deviation 3.2. Let X denote the value of the variable for a randomly selected member of the population. Find the variance of X.

7.59 As part of a screening exam for prospective insured, a physician conducts tests for acute problems. On average, the first positive test occurs with the 25th person tested by the physician. Determine the variance of the number of people tested until
a) the first positive test. b) the fifth positive test.

7.60 Pinworm infestation, commonly found in children, can be treated with the drug pyrantel pamoate. According to the *Merck Manual*, the treatment is effective in 90% of cases. If 20 children with pinworm infestation are treated with pyrantel pamoate, what is the variance of the number cured?

7.61 Refer to Exercise 7.11 on page 334, where sickle cell anemia is discussed. Two people with sickle cell trait have five children. What is the standard deviation of the number of those five children who have sickle cell anemia?

7.62 An upper-level probability class has six undergraduate students and four graduate students. A random sample of three students is taken from the class. Determine the variance of the number of undergraduate students selected if the sampling is
a) without replacement. b) with replacement.
c) Why are your results in parts (a) and (b) different?

EXAMPLE 7.20 ***Correlation of Events***

Suppose that A and B are events with positive probability. Show that I_A and I_B are positively correlated, negatively correlated, or uncorrelated depending on whether $P(B \mid A)$ is greater than, less than, or equal to $P(B)$.

Solution Recall that, for each event E, we have $\mathcal{E}(I_E) = P(E)$. Noting that $I_A I_B = I_{A \cap B}$ and applying the general multiplication rule, we conclude that

$$\begin{aligned}\operatorname{Cov}(I_A, I_B) &= \mathcal{E}(I_A I_B) - \mathcal{E}(I_A)\,\mathcal{E}(I_B) = \mathcal{E}(I_{A\cap B}) - \mathcal{E}(I_A)\,\mathcal{E}(I_B)\\ &= P(A \cap B) - P(A)P(B) = P(A)P(B \mid A) - P(A)P(B)\\ &= P(A)\big(P(B \mid A) - P(B)\big).\end{aligned}$$

Thus I_A and I_B are positively correlated, negatively correlated, or uncorrelated depending on whether $P(B \mid A)$ is greater than, less than, or equal to $P(B)$. *Note:* Referring to the definitions given in Exercise 4.17 on page 136, we see that I_A and I_B are positively correlated, negatively correlated, or uncorrelated depending on whether events A and B are positively correlated, negatively correlated, or independent. ■

EXERCISES 7.4 Basic Exercises

Note: Several exercises in this section are continuations of those in Sections 7.1–7.3.

✓ **7.86** Two balanced dice are rolled, one black and one gray. Let Y and Z be the faces showing on the black and gray dice, respectively, and let X be the sum of the two faces showing.

a) Determine the variances of Y and Z.

b) Express X in terms of Y and Z, and then use properties of variance and the result of part (a) to obtain the variance of X.

c) Compare your answer in part (b) to that obtained in Exercise 7.54 on page 362. Which method for obtaining $\operatorname{Var}(X)$ is easier? Explain your answer.

7.87 An archer shoots an arrow into a square target 6 feet on a side whose center we call the origin. Assume that the archer will actually hit the target and is equally likely to hit any portion of the target. The archer scores 10 points if she hits the bull's eye—a disk of radius 1 foot centered at the origin (region A); she scores 5 points if she hits the ring with inner radius 1 foot and outer radius 2 feet centered at the origin (region B); and she scores 0 points otherwise (region C). Suppose that the archer shoots independently at the target four times. Let X, Y, and Z denote the numbers of times that the archer hits regions A, B, and C, respectively. Also, let T denote the total score in the four shots.

a) Determine and identify the joint PMF of the random variables X, Y, and Z.

b) Express T in terms of X, Y, and Z and then use the FEF and the result of part (a) to obtain $\mathcal{E}(T^2)$.

c) Exercise 7.34 on page 349 asks for $\mathcal{E}(T)$. Use that result and part (b) to find $\operatorname{Var}(T)$.

d) Exercise 7.55 on page 362 asks for the variance of the score for one shot at the target. Use that result and properties of variance to find $\operatorname{Var}(T)$.

e) Identify yet another way to find $\operatorname{Var}(T)$.

7.88 Refer to Example 7.15 on page 367.

a) Use the definition of the variance and the PMF of $X + Y$, which was found in Example 6.23 (see Table 6.7 on page 308), to determine $\operatorname{Var}(X + Y)$.

Let X, Y, and Z denote the numbers of times that she hits regions A, B, and C, respectively. Also, let T denote the total score in the four shots.

a) Determine and identify the joint PMF of the random variables X, Y, and Z.
b) Express T in terms of X, Y, and Z, and then use the FEF and the result of part (a) to obtain $\mathcal{E}(T)$.
c) Determine and identify the univariate marginal PMFs of X, Y, and Z and, from them, find the expected values of X, Y, and Z.
d) Express T in terms of X, Y, and Z and then use the linearity property of expected value and the result of part (c) to obtain $\mathcal{E}(T)$.
e) Exercise 7.2 on page 333 asks for the expected score for one shot at the target. Use that result and the linearity property of expected value to obtain $\mathcal{E}(T)$.
f) Identify yet another way to obtain $\mathcal{E}(T)$.

maybe

7.35 N married couples are randomly seated at a rectangular table, the women on one side and the men on the other. What is the expected number of men who sit across from their wives? *Hint:* One way to solve this problem is to first determine the PMF of the number of men who sit across from their wives and then apply the definition of expected value. However, there is an easier way.

X

•7.36 A random sample of size n is taken from a very large lot of items in which $100p_1\%$ have exactly one defect and $100p_2\%$ have two or more defects, where $0 < p_1 + p_2 < 1$. An item with exactly one defect costs \$1 to repair, whereas an item with two or more defects costs \$3 to repair. Determine the expected cost of repairing the defective items in the sample.

7.37 Determine the expected number of times a balanced die must be thrown to get all six possible numbers. *Hint:* For $1 \le k \le 6$, let X_k denote the number of throws from the appearance of the $(k-1)$th distinct number until the appearance of the kth distinct number.

7.38 Expected utility: One method for deciding among various investments involves the concept of *expected utility*. Economists describe the importance of various levels of wealth by using *utility functions*. For instance, in most cases, a single dollar is more important (has greater utility) for someone with little wealth than for someone with great wealth. Consider two investments—say, Investment A and Investment B. Measured in thousands of dollars, suppose that Investment A yields 0, 1, and 4 with probabilities 0.1, 0.5, and 0.4, respectively, and that Investment B yields 0, 1, and 16 with probabilities 0.5, 0.3, and 0.2, respectively. Let Y denote the yield of an investment. For the two investments, determine and compare

a) $\mathcal{E}(Y)$, the expected yield.
b) $\mathcal{E}(\sqrt{Y})$, the expected utility, using the utility function $u(y) = \sqrt{y}$. Interpret the utility function u.
c) $\mathcal{E}(Y^{3/2})$, the expected utility, using the utility function $v(y) = y^{3/2}$. Interpret the utility function v.

7.39 A lot contains 17 items, each of which is subject to inspection by two quality assurance engineers. Each engineer randomly and independently selects 4 items from the lot. Determine the expected number of items selected by

a) both engineers. b) neither engineer. c) exactly one engineer.
d) Without doing any computations, obtain the sum of the three expected values found in parts (a), (b), and (c). Explain your reasoning.

7.40 Consider two electrical components whose lifetimes observed at discrete time units (e.g., every hour) are independent geometric random variables with parameter p. Use tail probabilities to determine the expected time until

a) the first component to fail. b) the second component to fail.

7.98 An actuary is modeling the sizes of surgical claims and their associated hospital claims, measured in thousands of dollars. The first and second moments of a surgical claim are 5 and 27.4, respectively, the first and second moments of a hospital claim are 7 and 51.4, respectively, and the variance of the total of the surgical and hospital claims is 8. Let X and Y denote the sizes of a combined surgical and hospital claim before and after the application of a 20% surcharge on the hospital portion of the claim, respectively. Determine and interpret the covariance and correlation coefficient of X and Y.

7.99 Let X and Y be discrete random variables defined on the same sample space and having finite variances.

a) Prove that XY has finite mean. *Hint:* First show that $|xy| \le (x^2 + y^2)/2$.

b) Prove that $X + Y$ has finite variance. *Hint:* First show that $(x + y)^2 \le 2(x^2 + y^2)$.

7.100 Verify the computing formula for the covariance: $\operatorname{Cov}(X, Y) = \mathcal{E}(XY) - \mathcal{E}(X)\,\mathcal{E}(Y)$.

7.101 Consider a finite population of size N and a variable defined thereon. Let σ^2 denote the population variance—that is, the arithmetic mean of all possible square deviations of the variable from the population mean for the entire population. Suppose that a random sample of size n is taken with replacement and let $\bar{X}_n$ denote the sample mean. Using results that we have already obtained, explain why $\operatorname{Var}(\bar{X}_n) = \sigma^2/n$. *Note:* No calculations are required.

7.102 Use properties of covariance to show $\operatorname{Cov}(X^*, Y^*) = \operatorname{Cov}(X, Y)/\sqrt{\operatorname{Var}(X) \cdot \operatorname{Var}(Y)}$ and explain why that quantity is unitless.

7.103 Let X have the discrete uniform distribution on $\{-1, 0, 1\}$ and let $Y = X^2$.

a) Show that $\rho(X, Y) = 0$, although X and Y are functionally related and hence associated.

b) Why doesn't the result of part (a) conflict with the correlation coefficient's role as a measure of association?

7.104 In Exercise 7.75 on page 364, we discussed how to best estimate a random variable Y by a constant, c, where "best" means choosing c to minimize the mean square error, $\mathcal{E}\big((Y - c)^2\big)$. Now determine how to best estimate Y by a linear function of a random variable X, again using minimum mean square error as the criterion for "best."

a) Obtain the values of a and b that minimize the mean square error $\mathcal{E}\big([Y - (a + bX)]^2\big)$. *Hint:* Use a calculus technique for finding extrema of functions of two variables.

b) Find the minimum mean square error.

Use the result of part (b) to

c) deduce part (a) of Proposition 7.16 on page 372.

d) deduce parts (c) and (d) of Proposition 7.16.

e) explain why the correlation coefficient is a measure of linear association.

Theory Exercises

7.105 Establish the properties of covariance presented in Proposition 7.12 on page 366.

7.106 Prove the bilinearity property of covariance, Equation (7.40) on page 366.

7.107 Use properties of expected value to establish the computing formula for the covariance, Equation (7.41) on page 366.

'08 Complete the proof of Proposition 7.13 by verifying Equation (7.43) on page 366.

ced Exercises

le variance: Let X be a random variable with finite variance. Suppose that you variance of X and want to estimate it. You take a random sample, $X_1, \ldots, X_n$,

Day 3 4/2

- ## Conditional Expectation

Recall $\mathbb{E}Y = \sum_{y \in R_Y} y \cdot \mathbb{P}(Y=y)$

Suppose we observed that $X = x$ (e.g. Y = height, X = weight of a person)

Interested in mean height of people in the room.

Distribution to use for Y?

Use $\mathbb{P}[(Y=y) \mid (X=x)]$

Def The conditional expectation of Y given $X = x$ is

$$\mathbb{E}[Y \mid X = x] = \sum_{y \in R_Y} y \cdot \mathbb{P}(Y=y \mid X=x)$$

Law of total expectation:

$$\boxed{\mathbb{E}Y = \sum_{x \in R_X} \mathbb{E}[Y \mid X=x] \cdot \mathbb{P}(X=x)}$$

Proof: $\mathbb{E}Y = \sum_{x \in R_X} \sum_{y \in R_Y} y \cdot \underbrace{\mathbb{P}_{XY}(x,y)}_{\mathbb{P}_{Y|X}(y|x) \cdot \mathbb{P}_X(x)}$

$$= \sum_{x \in R_X} \Big(\mathbb{P}_X(x) \cdot \underbrace{\sum_{y \in R_Y} y \cdot \mathbb{P}_{Y|X}(y|x)}_{\mathbb{E}(Y \mid X=x)} \Big)$$

Neat way to write: $\boxed{\mathbb{E}Y} = \sum_{x \in R_X} \mathbb{E}(Y \mid X=x) \cdot \mathbb{P}(X=x)$

$$= \boxed{\mathbb{E}[\mathbb{E}(Y|X)]}.$$

HW read Example 7.23 p. 380

skip

EX. 1 X = SAT score

Y = college grade

A student is picked at random

$\underbrace{\mathbb{E}(Y|X)}$
mean college grade given SAT score X.

EX. 2 Items in a shipment can be defective with $p = 0.1$

Let N = number of items in a shipment, N is a RV

Y = # of defective items in the shipment

Q: Suppose $N = 10$, what is the distribution of Y?

$\mathbb{P}_{Y|N}(y|10) \sim \text{Binomial}(n=10, p=0.1)$

$\mathbb{E}[Y|N=10] = np = 10(0.1) = 1$

In general $\mathbb{E}[Y|N=n] = np = (0.1)n$

$\Rightarrow \mathbb{E}[Y|N] = Np$ (a function of N)

$\mathbb{E}Y = \mathbb{E}(Np) = (\mathbb{E}N) \cdot p$

Suppose we know $\mathbb{P}(N=n)$:

$n =$	10,	11,	12,	13,	14,	15
	5%	10%	10%	20%	35%	20%

$\Rightarrow \mathbb{E}Y = (10 \cdot 5\% + 11 \cdot 10\% + \dots + 15 \cdot 20\%) \cdot p = 13.3p = 1.33$

Conditional Variance

$Var(Y \mid X=x)$ use $P_{Y|X}(y|x)$ to compute it.

Def The conditional variance of Y given $X=x$ is:

$$Var(Y \mid X=x) = \mathbb{E}[(Y - \mathbb{E}[Y \mid X=x])^2 \mid X=x]$$

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

Law of total variance

$$Var(Y) = \mathbb{E}[Var(Y \mid X)] + Var(\mathbb{E}[Y \mid X])$$

read Proof: p. 386

Ex. Sum of Random Number of Random Variables

$S_N = \sum_{i=1}^{N} X_i$, where X_i are iid RVs independent of N

Suppose ... s during the day

... ssengers in bus i

... ssengers during the day

$\Rightarrow$ ... (S_N)

Compute $\mathbb{E}[$ ...) mean and variance do not depend on

Let $\mu = \mathbb{E}X$ ...

$\mathbb{E}[S_N \mid N=n$ ... $+ X_n] = n\mu$

$\Rightarrow \mathbb{E}[S_N \mid N] =$ ...

$\mathbb{E}[S_N] = \mathbb{E}$ ... of buses

Day 3 2/12

$$\mathrm{Var}(S_N) = ?$$

$$\mathrm{Var}(S_N \mid N = n) = \mathrm{Var}(X_1 + X_2 + \dots + X_n) = \mathrm{Var}X_1 + \dots + \mathrm{Var}X_n = n\sigma$$

$$\Rightarrow \mathrm{Var}(S_N \mid N) = N\sigma^2$$

$$\mathrm{Var}(S_N) = \mathbb{E}[\mathrm{Var}(S_N \mid N)] + \mathrm{Var}(\mathbb{E}[S_N \mid N]) =$$

$$= \mathbb{E}[N\sigma^2] + \mathrm{Var}(N\mu) =$$

$$= \sigma^2 \cdot \mathbb{E}[N] + \mu^2 \cdot \mathrm{Var}(N).$$

a) Find the conditional expectation of Y given $X = x$. *Note:* Exercise 6.46(a) on page 288 asks for the conditional PMF of Y given $X = x$.
b) Find the conditional variance of Y given $X = x$.

7.118 Refer to Exercise 7.117 and note that Exercise 6.3 on page 269 asks for the joint PMF of X and Y and the marginal PMF of Y.
a) Use the law of total expectation to determine $\mathcal{E}(Y)$.
b) Determine $\mathcal{E}(Y)$ by using the definition of expected value. Compare your answer to the one obtained in part (a).
c) Use the law of total variance to determine $\operatorname{Var}(Y)$.
d) Determine $\operatorname{Var}(Y)$ by using the definition of variance. Compare your answer to the one obtained in part (c).

7.119 An automobile insurance company divides its policyholders into three groups, which constitute 45.8%, 32.6%, and 21.6% of the policyholders. The average claim amounts for the three groups are \$1457, \$2234, and \$2516, respectively. Obtain the average claim amount among all policyholders.

V **7.120** The hourly number of customers entering a bank for the purpose of making a deposit has a Poisson distribution with parameter 25.8. Each such customer deposits an average of \$574 with a standard deviation of \$3167. For a 1-hour period,
a) find the expected total deposits by all entering customers.
b) find the standard deviation of deposit amounts by all entering customers.

7.121 Part of a homeowner's insurance policy covers one miscellaneous loss per year, which is known to have a 10% chance of occurring. If there is a miscellaneous loss, the probability is c/x that the loss amount is \$$100x$, for $x = 1, 2, \ldots, 5$, where c is a constant. These are the only loss amounts possible. If the deductible for a miscellaneous loss is \$200, determine the net premium for this part of the policy—that is, the amount that the insurance company must charge to break even.

7.122 An automobile insurance company classifies its policyholders as either good or bad drivers, of which there are 75% and 25%, respectively. The mean claim amounts for good and bad drivers are \$2000 and \$3500, respectively, and the standard deviations of those claim amounts are \$200 and \$400, respectively. Determine the mean and standard deviation of the claim amounts for all policyholders of this company.

7.123 Let $X \sim \mathcal{G}(p)$. Assume as known that $\mathcal{E}(X) = 1/p$, as found in Example 7.25 on page 382. Determine $\operatorname{Var}(X)$ by using each of the following methods.
a) Condition on the outcome of the first trial to obtain the second moment of X.
b) Use the law of total variance.

7.124 Let $X \sim \mathcal{NB}(r, p)$. Assume as known that the mean of a geometric random variable with parameter p is $1/p$, as found in Example 7.25 on page 382. Use mathematical induction on r to show that $\mathcal{E}(X) = r/p$ by conditioning on the outcome of the first trial.

7.125 Let $X \sim \mathcal{NB}(r, p)$. Assume as known that $\mathcal{E}(X) = r/p$, as found in Exercise 7.124, and that the variance of a geometric random variable with parameter p is $(1-p)/p^2$, as found in Exercise 7.123. Show that $\operatorname{Var}(X) = r(1-p)/p^2$ by using mathematical induction on r with each of the following methods.
a) Condition on the outcome of the first trial to obtain the second moment of X.
b) Use the law of total variance.

Read Ch 8

Cumulative Distribution Function (CDF)

Def The CDF of a random variable X is a function

$$\boxed{\mathbb{F}_X(x) = \mathbb{P}(X \leq x), \qquad -\infty < x < \infty}$$

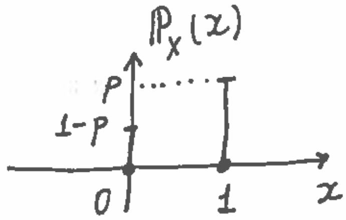
* it accumulates the probability up to x
* needs to be specified for all $x \in (-\infty, \infty)$

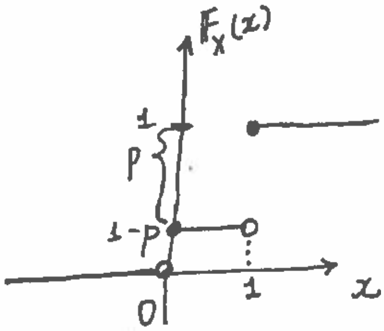
If X is a discrete RV with PMF $\mathbb{P}_X(x) = \mathbb{P}(X = x)$

then $$\mathbb{F}_X(x) = \sum_{u \leq x} \mathbb{P}_X(u).$$

Ex. $X \sim \text{Bern}(p)$

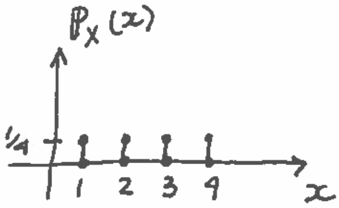
$$X = \begin{cases} 0 & 1-p \\ 1 & p \end{cases}$$

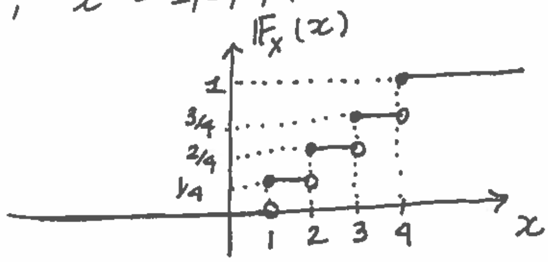




Ex. $X \sim$ Discrete Uniform (4)

PMF $\mathbb{P}(X = x) = \frac{1}{4}$, $x = 1, 2, 3, 4$





specify only for R_X

must specify for all $x \in \mathbb{R}$

Properties of the CDF

1. $\mathbb{F}_X(x)$ is non-decreasing
2. $\mathbb{F}_X(x)$ is right-continuous
3. $\mathbb{F}_X(-\infty) = \lim\limits_{x \to -\infty} \mathbb{F}_X(x) = 0$
4. $\mathbb{F}_X(\infty) = \lim\limits_{x \to \infty} \mathbb{F}_X(x) = 1$

These four properties characterize the CDF.

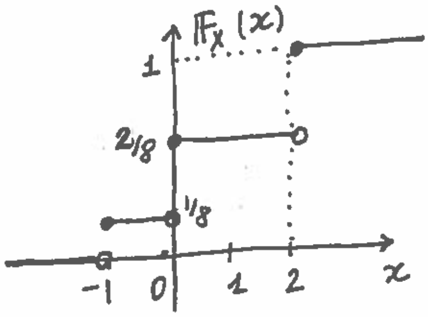
<u>Ex.</u> X is a discrete RV

$\mathbb{P}(1 < X < 30) = \mathbb{P}(X \leqq 29) - \mathbb{P}(X \leqq 1) = \mathbb{F}_X(29) - \mathbb{F}_X(1)$

$\mathbb{P}(X = 30) = \mathbb{P}(X \leqq 30) - \mathbb{P}(X \leqq 29) = \mathbb{F}_X(30) - \mathbb{F}_X(29)$

skip <u>Ex.</u>

$$\mathbb{F}_X(x) = \begin{cases} 0 & \text{for } x < -1 \\ \frac{1}{8} & \text{for } -1 \leqq x < 0 \\ \frac{2}{8} & \text{for } 0 \leqq x < 2 \\ 1 & \text{for } x \geqq 2 \end{cases}$$



satisfies the four properties and hence it is a CDF of a RV X

Find: $\mathbb{P}(X < -2) = 0$

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

$\mathbb{P}(X = -1) = \mathbb{P}(X \leqq -1) - \mathbb{P}(X < -1) = \frac{1}{8} - 0 = \frac{1}{8}$

$\mathbb{P}(X \geqq -1) = 1 - \mathbb{P}(X < -1) = 1$

$\mathbb{P}(1 < X < 2) = \mathbb{P}(X < 2) - \mathbb{P}(X \leqq 1) = \frac{2}{8} - \frac{2}{8} = 0$

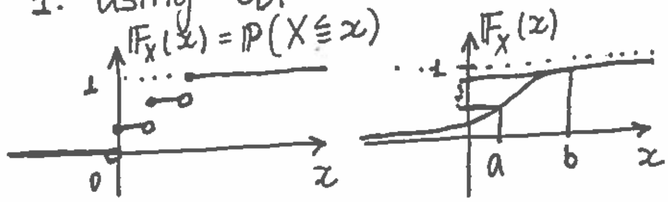
$\mathbb{F}_X(2^-) = \mathbb{P}(X \leqq 2^-) = \frac{2}{8}$

$\mathbb{F}_X(\infty) = 1$

Going from Discrete to Continuous RV

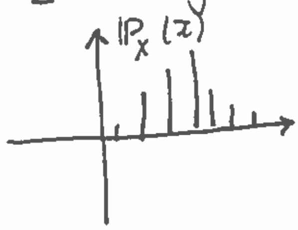
Two perspectives:

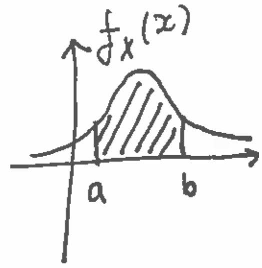
1. Using CDF



satisfies the 4 properties

2. using PMF



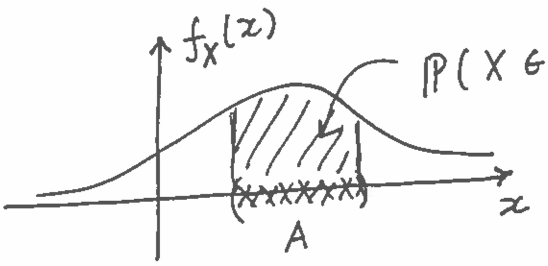


Def X is said to be a continuous RV if there is a function $f_X(x)$ called <u>probability density function</u> PDF such that

Notation: $X \sim f_X(x)$

- $f_X(x) \geq 0$
- $\int_{-\infty}^{\infty} f_X(x)\,dx = 1$
- $\mathbb{P}(a < X \leq b) = \int_a^b f_X(x)\,dx$

<u>Note:</u> $\mathbb{P}(X=a) = 0$ if X is cont. RV $\Rightarrow \mathbb{P}(a \leq X \leq b) = \mathbb{P}(a < X < b)$

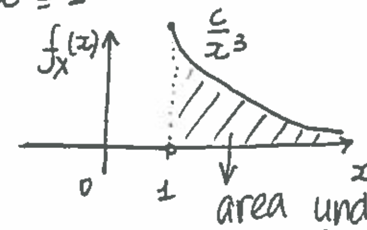


$\mathbb{P}(X \in A) = \int_A f_X(x)\,dx$, where A is any subset of the real line ($A \subset \mathbb{R}$)

$$\underbrace{\mathbb{P}(a \leq X \leq b)}_{\text{unitless}} = \int_a^b \underbrace{f_X(x)}_{\frac{1}{\text{inches}}} \underbrace{dx}_{\text{inches}}$$

Ex. $f_X(x) = \begin{cases} 0 & , x < 1 \\ \dfrac{c}{x^3} & , x \geqq 1 \end{cases}$

1. Find c
2. Find $\mathbb{P}(X \geqq 3)$



area under the curve = 1

1. c = normalization constant → ensures that $\int_{-\infty}^{\infty} f_X(x) = 1$

$$1 = \int_{-\infty}^{\infty} f_X(x)\,dx = \int_{1}^{\infty} \frac{c}{x^3}\,dx = c\int_{1}^{\infty} x^{-3}\,dx = -\frac{1}{2}c\,x^{-2}\Big|_1^{\infty} = 0 + \frac{1}{2}c \cdot 1$$

$$\Rightarrow \boxed{c = 2}$$

$$\Rightarrow f_X(x) = \begin{cases} 0 & , x < 1 \\ \dfrac{2}{x^3} & , x \geqq 1 \end{cases}$$

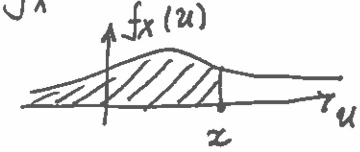
2. $$\mathbb{P}(X \geqq 3) = \int_{3}^{\infty} \frac{2}{x^3}\,dx = -\frac{1}{x^2}\Big|_3^{\infty} = +\frac{1}{3^2} = \frac{1}{9}$$

In the previous example specify $f_X(x)$ for all x:

- $f_X(x) = \begin{cases} 0 & , x < 1 \\ \dfrac{2}{x^3} & , x \geqq 1 \end{cases}$ Best way
- $f_X(x) = \dfrac{2}{x^3}$, $x \geqq 1$ not ideal, but tolerable
- $\cancel{f_X(x) = \dfrac{2}{x^3}}$ wrong e.g. $\mathbb{P}(X \geqq 0) \neq \int_0^{\infty} \frac{2}{x^3}\,dx$
- $f_X(x) = \begin{cases} 0 & , x \leqq 1 \\ \dfrac{2}{x^3} & , x > 1 \end{cases}$ OK to write as because $\mathbb{P}(X = 1) = 0$

<u>CDF</u> $\mathbb{F}_X(x) = \mathbb{P}(X \leq x)$, $-\infty < x < \infty$

⊛ Suppose we are given PDF $f_X(x)$. Find CDF



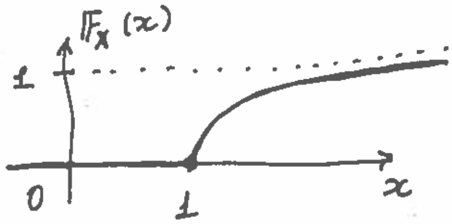
$$\mathbb{F}_X(x) = \mathbb{P}(X \leq x) = \int_{-\infty}^{x} f_X(u)\,du$$

just a dummy variable

<u>EX</u>. Find CDF for previous example

$$f_X(x) = \begin{cases} 0, & x < 1 \\ 2/x^3, & x \geq 1 \end{cases}$$

$$\mathbb{F}_X(x) = \int_{-\infty}^{1} 0\,dx + \int_{1}^{x} \frac{2}{u^3}\,du = -\frac{1}{u^2}\Big|_1^x = -\frac{1}{x^2} + 1$$



$$\mathbb{F}_X(x) = \begin{cases} 0, & x < 1 \\ -\frac{1}{x^2} + 1, & x \geq 1 \end{cases}$$

⊛ How to go from CDF to PDF : take the derivative

$$\mathbb{F}_X(x) = \begin{cases} 0, & x < 1 \\ 1 - \frac{1}{x^2}, & x \geq 1 \end{cases}$$

$$f_X(x) = \frac{d\,\mathbb{F}_X(x)}{dx}$$

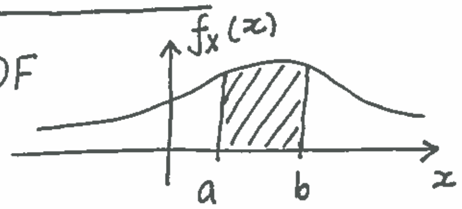
$\Rightarrow f_X(x) = 0, \quad x < 1$

$f_X(x) = \frac{2}{x^3}, \quad x \geq 1$

Continuous RV

skip

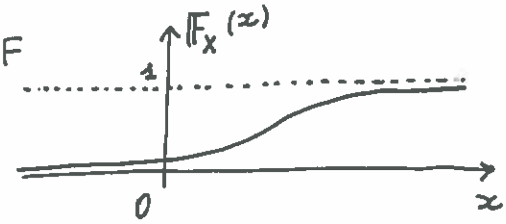
PDF



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

$$\mathbb{P}(X=a) = 0 \Rightarrow \mathbb{P}(a \leq X \leq b) = \mathbb{P}(a<X<b)$$

CDF



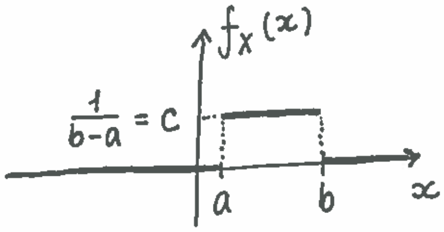
$$\mathbb{F}_X(x) = \mathbb{P}(X \leq x) = \int_{-\infty}^{x} f_X(u)\,du$$

$$f_X(x) = \frac{d\,\mathbb{F}_X(x)}{dx}$$

- ## Probability models

1. Uniform

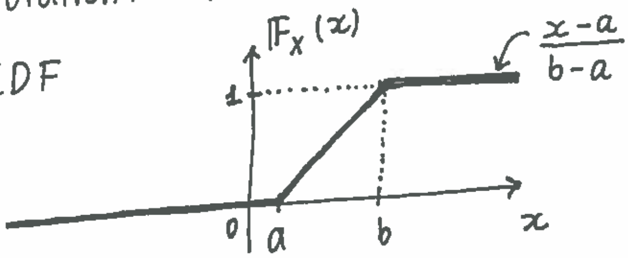
PDF $f_X(x) = \begin{cases} c & \text{, if } a<x<b \\ 0 & \text{, otherwise} \end{cases}$, $-\infty < a < b < \infty$



$c = \frac{1}{b-a}$ area under the curve = 1

Notation: $X \sim Unif(a,b)$

CDF



if $x<a$: $\mathbb{F}_X(x) = 0$

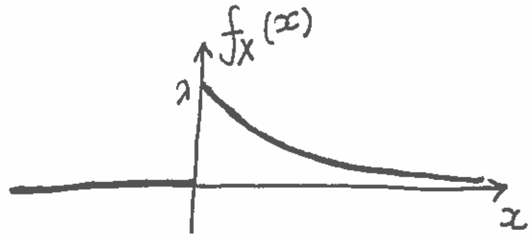
if $x>b$: $\mathbb{F}_X(x) = 1$

if $a \leq x \leq b$:

$$\mathbb{F}_X(x) = \int_a^x \frac{1}{b-a}\,du = \frac{1}{b-a}\int_a^x du = \frac{x}{b}$$

$$\Rightarrow \mathbb{F}_X(x) = \begin{cases} 0 & \text{, if } x<a \\ \frac{x-a}{b-a} & \text{, if } a \leq x \leq b \\ 1 & \text{, if } x>b \end{cases}$$

2. Exponential Distribution



PDF $f_X(x) = \begin{cases} \lambda e^{-\lambda x}, & x \geqq 0 \\ 0, & \text{otherwise} \end{cases}$

Notation: $X \sim Exp(\lambda)$ parameter, $\lambda > 0$

$c = \lambda$ check

$$1 = \int_{-\infty}^{\infty} f_X(x)\,dx = \int_0^{\infty} \lambda e^{-\lambda x} dx = \lambda\left(-\frac{1}{\lambda}\right)e^{-\lambda x}\Big|_0^{\infty} = 0 + 1 = 1.$$

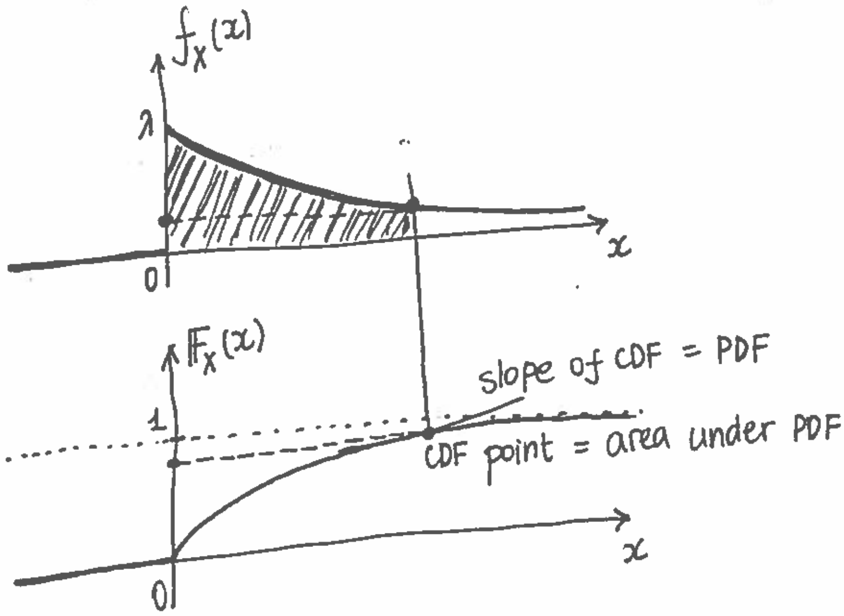
CDF $\mathbb{F}_X(x) = \int_{-\infty}^{x} f_X(u)\,du$

if $x < 0$: $\mathbb{F}_X(x) = 0$

if $x \geqq 0$: $\mathbb{F}_X(x) = \int_0^x \lambda e^{-\lambda u} du = \lambda\left(-\frac{1}{\lambda}\right)e^{-\lambda u}\Big|_0^x =$

$$= -e^{-\lambda u}\Big|_0^x = -e^{-\lambda x} + 1 = 1 - e^{-\lambda x}$$

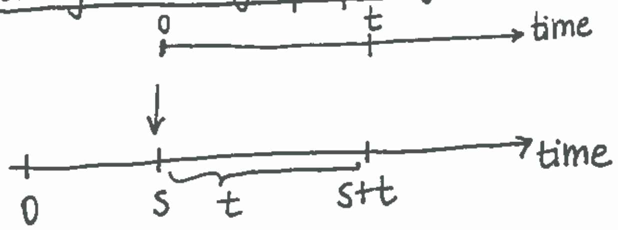
$\Rightarrow \mathbb{F}_X(x) = \begin{cases} 0, & \text{if } x < 0 \\ 1 - e^{-\lambda x}, & \text{if } x \geqq 0 \end{cases}$ $\left[\begin{array}{l} \text{or } x \leqq 0 \\ \quad x > 0 \end{array}\right.$, doesn't matter where the equality is, since it is c



Tail Probability

$$\mathbb{P}(X > x) = 1 - \mathbb{P}(X \leqq x) = 1 - \mathbb{F}_X(x) = \begin{cases} 1 & , \text{ if } x < 0 \\ e^{-\lambda x} & , \text{ if } x \geqq 0 \end{cases}$$

Lack of memory property (no aging)



X = life time of device

Given $X > s$

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

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

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

Cont.	Discrete
Exp	Geom

$Y = \frac{X}{n}$, $X \sim \text{Geom}(p)$

let $\underset{n \to \infty}{p \to 0}$ $np \to \lambda$

$\Rightarrow Y \sim \text{Exp}(\lambda)$.

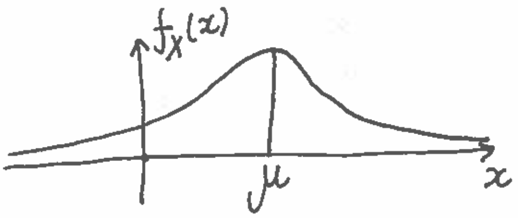
4. Normal Distribution (Gaussian)

Gauss $1+\dots+100$

two parameters : $\mu \in \mathbb{R}$, $\sigma > 0$

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

<u>PDF</u> $f_X(x) = \dfrac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$, $-\infty < x < \infty$

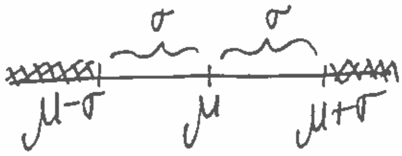


bell curve $\mathbb{E}X = \mu$

$\mathrm{Var}\,X = \sigma^2$

Empirical Rule

$\mathbb{P}(|X-\mu| > \sigma) \approx 0.32$



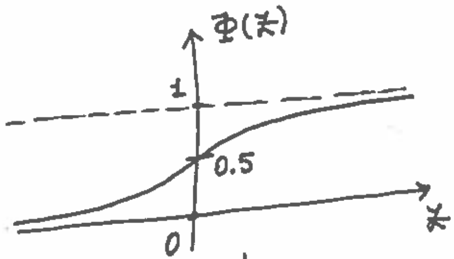
$\mathbb{P}(|X-\mu| > 2\sigma) \approx 0.05$

$\mathbb{P}(|X-\mu| > 3\sigma) \approx 3‰ = 0.003$
per thousand

Standard Normal Distribution

$$Z = \frac{X-\mu}{\sigma} \Rightarrow Z \sim N(0,1)$$

<u>CDF</u> $\Phi(z) = \mathbb{F}_Z(z) = \mathbb{P}(Z \leqq z)$



tabulated

<u>EX.</u> $X \sim N(2,9) \Rightarrow \sigma = 3$

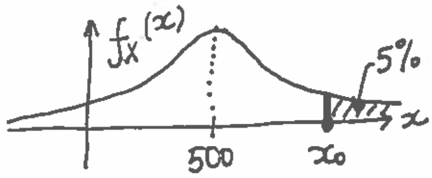
Find $\mathbb{P}(3 \leqq X \leqq 4) = ?$ $= \mathbb{P}\left(\frac{1}{3} \leqq Z \leqq \frac{2}{3}\right) = \Phi\left(\frac{2}{3}\right) - \Phi\left(\frac{1}{3}\right) = 0.7454 - 0.6293 = 0.1$

$= \mathbb{P}\left(\frac{3-2}{3} \leqq Z \leqq \frac{4-2}{3}\right)$

<u>EX.</u> Let X = amount of radiation that can be absorbed by a person before death ensues

$X \sim N(500, (150)^2)$

Above which dosage will only 5% exposed survive?



$0.05 = \mathbb{P}(X \geqq x) = \mathbb{P}\left(Z \geqq \frac{x-500}{150}\right)$

$0.95 = \mathbb{P}\left(Z \leqq \frac{x-500}{150}\right)$

$\frac{x_0 - 500}{150} \approx 1.645$ from table : $\mathbb{P}(Z \leqq 1.64) = 0.9495$, $\mathbb{P}(Z \leqq 1.65) = 0.9505$ $\Rightarrow \mathbb{P}(Z \leqq 1.645) \approx 0.95$

$x_0 \approx 746.75$ röentgens

3. Gamma (α, λ), $\alpha > 0$, $\lambda > 0$

skip

- case where $\alpha = r \geq 1$ integer : $X \sim \text{Gamma}(r, \lambda)$

$$X = T_1 + T_2 + \cdots + T_r \quad , \text{ where } T_i \text{ iid } \text{Exp}(\lambda)$$

↖ time of the r^{th} failure

If r is integer, the Gamma is also called Erlang distr.

$$\mathbb{F}_X(x) = 1 - e^{-\lambda x} \sum_{j=0}^{r} \frac{(\lambda x)^j}{j!} \quad , \text{ for } x \geq 0$$

skip

Continuous	Discrete
Exp	Geometric
Gamma	Pascal

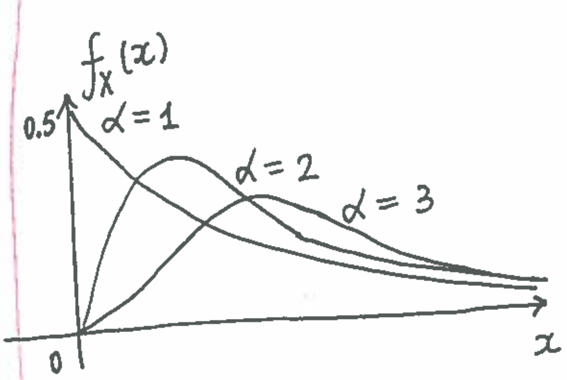
use

skip

- in general α does not have to be integer

$$f_X(x) = \begin{cases} 0 \ , & \text{if } x < 0 \\ \dfrac{\lambda^\alpha}{\Gamma(\alpha)} x^{\alpha-1} \cdot e^{-\lambda x} \ , & \text{if } x \geq 0 \end{cases} ,$$

where $\underbrace{\Gamma(\alpha)}_{\text{Gamma function}} = \int_0^\infty e^{\alpha-1} e^{-x} dx$ cont. extension of the factorial



Property

$\Gamma(\alpha + 1) = \alpha \cdot \Gamma(\alpha)$

if α is integer $(\alpha = n)$

$$\Gamma(n+1) = n\Gamma(n) =$$
$$= n(n-1)\Gamma(n-1) =$$
$$= n(n-1)(n-2)\ldots\Gamma(1)$$

$$\Gamma(1) = \int_0^\infty e^{-x} dx =$$

So $\Gamma(n+1) = n!$

HW6

1) 6.133 p.320 multivariate (hypergeometric)
2) 7.1 p.333
3) 7.3 p.333
4) 7.11 p.334
5) 7.36 p.350

Extra credit:
1) 7.35 p.350

HW6

1) 6.136 p.320 (multinomial)
2) 6.133 p.320 (multivariate hypergeometric)
3) 7.1 p.333 (X+Y pmf)
4) 7.3 p.333 (urns)
5) 7.11 p.334 (binom mean)
6) 7.36 p.350

Extra credit
1) 7.35 p.350

7.5 Expected value as center of gravity: Let X be a random variable with finite range, say, $\{x_1, \ldots, x_m\}$. Set $p_k = p_X(x_k)$ for $k = 1, 2, \ldots, m$. Think of the x-axis as a seesaw and each p_k as a mass placed at point x_k on the seesaw. The *center of gravity* of these masses is defined to be the point $\bar{x}$ on the x-axis at which a fulcrum could be placed to balance the seesaw. Relative to the center of gravity, the torque acting on the seesaw by the mass p_k is proportional to the product of that mass and the signed distance of the point x_k from $\bar{x}$. Show that the center of gravity equals the expected value of X; that is, $\bar{x} = \mathcal{E}(X)$. *Hint:* To balance, the total torque acting on the seesaw must be 0.

7.6 Consider a random experiment with a finite number of possible outcomes and let Ω be the sample space.

a) For a random variable X defined on Ω, show that $\mathcal{E}(X) = \sum_{\omega \in \Omega} X(\omega) P(\{\omega\})$.
b) Show that, if the outcomes are equally likely, the expression in part (a) reduces to the arithmetic average of the possible observations of the random variable X.

7.7 A variable of a finite population has mean 82.4.

a) Let X denote the value of the variable for a randomly selected member of the population. Find the expected value of X.
b) If a random sample of size n is taken from the population with replacement and the value of the variable is observed each time, roughly what will be the sum of the observed values?

7.8 As part of a screening exam for prospective insured, a physician conducts tests for acute problems. On average, the first positive test occurs with the 25th person tested by the physician.

a) Determine the probability that the first positive test occurs by the sixth person tested.
b) On average, how many people must the physician test until obtaining the fifth positive test?

7.9 Pinworm infestation, commonly found in children, can be treated with the drug pyrantel pamoate. According to the *Merck Manual*, the treatment is effective in 90% of cases. If 20 children with pinworm infestation are treated with pyrantel pamoate, what is the expected number cured?

7.10 According to the *Daily Racing Form*, the probability is about 0.67 that the favorite in a horse race will finish in the money (first, second, or third place). Determine the smallest number of races required so that the expected number of times that the favorite finishes in the money is at least 10.

X •**7.11** Sickle cell anemia is an inherited blood disease that occurs primarily in blacks. In the United States, about 15 of every 10,000 black children have sickle cell anemia. The red blood cells of an affected person are abnormal; the result is severe chronic anemia, which causes headaches, shortness of breath, jaundice, increased risk of pneumococcal pneumonia and gallstones, and other severe problems. Sickle cell anemia occurs in children who inherit an abnormal type of hemoglobin, called hemoglobin S, from both parents. If hemoglobin S is inherited from only one parent, the person is said to have sickle cell trait and is generally free from symptoms. There is a 50% chance that a person who has sickle cell trait will pass hemoglobin S to an offspring. If two people who have sickle cell trait have five children, how many children should they expect to have sickle cell anemia?

7.12 An upper-level probability class has six undergraduate students and four graduate students. A random sample of three students is taken from the class. Determine the expected number of undergraduate students selected if the sampling is

a) without replacement. b) with replacement.
c) Why are your results in parts (a) and (b) the same?

HW 7*

✓ 1) p.362 7.54 (var)

✓ 2) p.374 7.86 (var)

✓ 3) p.376 7.101 (Var($\bar{X}_n$))

Extra credit:

X 1) p.375 7.91

HW 7

1) 7.86 · p. 374
2) 7.92 p. 375
3) 7.102 p. 376
4) 7.103 p. 376
5) 7.120 p. 389

Extra credit:

1) 7.91 · p. 375
2) 7.109 p. 376

HW8

(summer 1 '15)
day class

1) 7.92 p. 375
2) 7.102 p. 376
3) 7.103 p. 376
4) 7.120 p. 389

Extra credit:

~~1) 7.~~109 p. 376

HW 8

✓ 1) 8.25 p. 414

✓ 2) 8.27 p. 414

X 3) 8.30 p. 415

8.18 Let X be the number of siblings of a randomly selected student from one of Professor Weiss's classes, as discussed in Example 5.1 on page 176. *Note:* Table 5.5 on page 186 gives the PMF of the random variable X.

8.19 Let Y be the indicator random variable of an event E.

8.20 According to *Vital Statistics of the United States,* published by the U.S. National Center for Health Statistics, chances are 80% that a person aged 20 will be alive at age 65. Of three people aged 20 selected at random, let X denote the number who live to be at least age 65.

8.21 Six men and five women apply for a job at Alpha, Inc. Three of the applicants are selected for interviews. Let X denote the number of women in the interview pool.

8.22 Let X be a discrete random variable.
a) Express F_X in terms of p_X. **b)** Express p_X in terms of F_X.

8.23 Let X have the discrete uniform distribution on the set of the first N positive integers.
a) Without using the PMF of X, obtain and graph the CDF of X.
b) Use part (a) and Equation (8.8) on page 413 to obtain the PMF of X.

8.24 Let X have the geometric distribution with parameter p.
a) Use the FPF to obtain the CDF of X.
b) Use tail probabilities to obtain the CDF of X.
c) Graph the CDF of X.
d) Use the CDF of X and Equation (8.8) on page 413 to obtain the PMF of X.

8.25 Refer to Example 8.3 on page 407. Use the CDF obtained in that example to determine $P(a < W \leq b)$, $P(a < W < b)$, $P(a \leq W < b)$, and $P(a \leq W \leq b)$ for the specified values of a and b.
a) $a = 1, b = 2$ **b)** $a = 0.5, b = 2$ **c)** $a = 1, b = 2.75$ **d)** $a = 0.5, b = 2.75$
e) Repeat parts (a)–(d) by using the FPF and the PMF of W (Table 8.1 on page 407).

8.26 Refer to Example 8.4 on page 408. Use the CDF obtained in that example to determine $P(a < X \leq b)$, $P(a < X < b)$, $P(a \leq X < b)$, and $P(a \leq X \leq b)$ for the specified values of a and b.
a) $a = 0.2, b = 0.8$ **b)** $a = 0, b = 0.8$ **c)** $a = 0.2, b = 1.5$
d) $a = -1, b = 1.5$ **e)** $a = -2, b = -1$ **f)** $a = 1, b = 2$
g) For each of parts (a)–(f), why are the four probabilities identical?

8.27 Refer to Example 8.5 on page 409. Use the CDF obtained in that example to determine $P(a < Z \leq b)$, $P(a < Z < b)$, $P(a \leq Z < b)$, and $P(a \leq Z \leq b)$ for the specified values of a and b.
a) $a = 0.2, b = 0.8$ **b)** $a = 0, b = 0.8$ **c)** $a = 0.2, b = 1.5$
d) $a = -1, b = 1.5$ **e)** $a = -2, b = -1$ **f)** $a = 1, b = 2$
g) For each of parts (a)–(f), why are the four probabilities identical?

8.28 Refer to Example 8.6 on page 410. Use the CDF obtained in that example to determine $P(a < Y \leq b)$, $P(a < Y < b)$, $P(a \leq Y < b)$, and $P(a \leq Y \leq b)$ for the specified values of a and b.
a) $a = 0.2, b = 0.8$ **b)** $a = 0.2, b = 0.75$ **c)** $a = -1, b = 1.5$
d) $a = -1, b = 0.75$ **e)** $a = -2, b = -1$ **f)** $a = 0.75, b = 2$

8.29 Let f be a nonnegative real-valued function such that $\int_{-\infty}^{\infty} f(x)\,dx = 1$. Suppose that X is a random variable that satisfies $P(X \in A) = \int_A f(x)\,dx$ for all subsets A of $\mathcal{R}$.
a) Obtain F_X in terms of f.
b) What is the relationship between F'_X and f?

c) Let a and b be real numbers with $a < b$. Determine $P(a < X \le b)$, $P(a < X < b)$, $P(a \le X < b)$, and $P(a \le X \le b)$ in terms of f.
d) Why are all four answers in part (c) the same?

8.30 For each function f, let X be as in Exercise 8.29. Graph f and determine and graph F_X.
a) $f(x) = \lambda e^{-\lambda x}$ if $x > 0$ and $f(x) = 0$ otherwise, where λ is a positive real number.
b) $f(x) = 1/(b - a)$ if $a < x < b$ and $f(x) = 0$ otherwise, where a and b are real numbers with $a < b$.
c) $f(x) = b^{-1}(1 - |x|/b)$ if $-b < x < b$ and $f(x) = 0$ otherwise, where b is a positive real number.

8.31 Decide whether each function F is the CDF of a random variable by checking properties (a)–(d) of Proposition 8.1 on page 411. For each F that is the CDF of a random variable, classify the random variable as discrete, continuous, or mixed.
a) $F(x) = 0$ if $x < 0$ and $F(x) = 1$ if $x \ge 0$.
b) $F(x) = 0$ if $x < 0$, $F(x) = 1 - p$ if $0 \le x < 1$, and $F(x) = 1$ if $x \ge 1$. Here p is a real number with $0 < p < 1$.
c) $F(x) = 0$ if $x < 0$ and $F(x) = \lfloor x \rfloor$ if $x \ge 0$.
d) $F(x) = 0$ if $x < 0$ and $F(x) = \sum_{n=0}^{\lfloor x \rfloor} a_n$ for $x \ge 0$, where $\{a_n\}_{n=0}^{\infty}$ is a sequence of nonnegative real numbers whose sum is 1.
e) $F(x) = 0$ if $x < 0$ and $F(x) = 1 - e^{-\lambda x}\sum_{j=0}^{r-1}(\lambda x)^j/j!$ if $x \ge 0$. Here λ is a positive real number and r is a positive integer.
f) $F(x) = 0$ if $x < 0$ and $F(x) = x$ if $x \ge 0$.
g) $F(x) = 0$ if $x < -1$, $F(x) = \frac{1}{2} + \frac{3}{8}x$ if $-1 \le x < 1$, and $F(x) = 1$ if $x \ge 1$.

8.32 Let X be a random variable and let m be a real number. Determine the CDF of each of the following random variables in terms of the CDF of X.
a) $Y = \max\{X, m\}$ b) $Z = \min\{X, m\}$

8.33 Let $X_1, \ldots, X_m$ be independent random variables, each having the same probability distribution as a random variable X. Determine the CDF of each of the following random variables in terms of the CDF of X.
a) $Y = \max\{X_1, \ldots, X_m\}$ b) $Z = \min\{X_1, \ldots, X_m\}$

8.34 A function F can be the CDF of a continuous random variable X and still have "flat spots." What is the probabilistic meaning of $F(x) = c$ for all $x \in [a, b]$, where c is a constant and a and b are real numbers with $a < b$?

Theory Exercises

8.35 Prove Equation (8.7) on page 412: If X is a random variable, then $F_X(x-) = P(X < x)$ for all $x \in \mathcal{R}$.

8.36 Prove parts (a), (b), and (d) of Proposition 8.2 on page 412.

Advanced Exercises

8.37 Let X be a random variable. Prove that the CDF of X has a countable number of discontinuities. *Hint:* For each $n \in \mathcal{N}$, consider the set $D_n = \{x \in \mathcal{R} : F_X(x) - F_X(x-) \ge 1/n\}$.

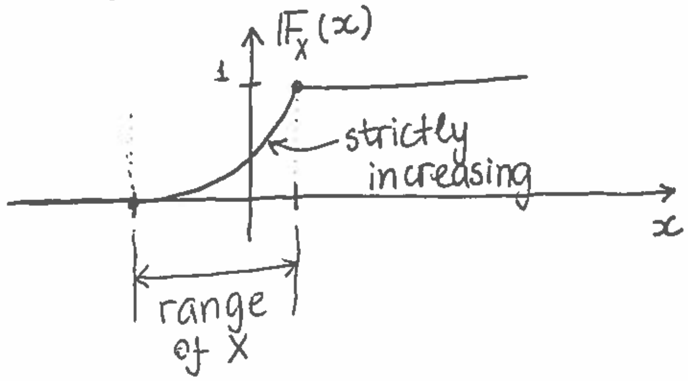
8.38 For each $n \in \mathcal{N}$, let the random variable X_n have the discrete uniform distribution on the set $\{0, 1/n, \ldots, (n-1)/n\}$.
a) Determine the CDF of X_n.

Random Number Generation

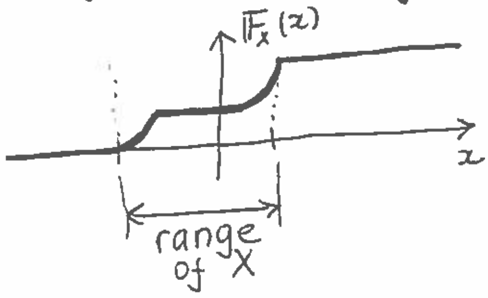
Computer can simulate $U \sim Unif(0,1)$.
How to simulate another distribution using this feature?

Simulation of continuous random variables

Let $X \sim \mathbb{F}_X(x)$ which is strictly increasing on R_X



we are excluding



We will consider

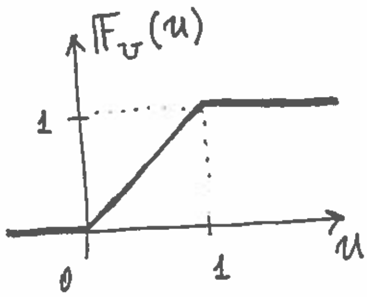
1) $\mathbb{F}_X(x)$: CDF of X
2) $\mathbb{F}_X(\mathbf{X})$: it is a RV
3) $\mathbb{F}_X^{-1}(x)$: inverse function
4) $\mathbb{F}_X^{-1}(U)$: it is a RV
   ↳ $U \sim Unif(0,1)$

Proposition (8.16, p.471)

Let $U \sim Unif(0,1)$.

Then $\boxed{\begin{array}{l} \text{a) } \mathbb{F}_X(X) \sim Unif(0,1) \\ \text{b) } \mathbb{F}_X^{-1}(U) \text{ has same distr. as } X \end{array}}$

Proof: CDF of U



$$\mathbb{F}_U(u) = \begin{cases} 0 & , u < 0 \\ u & , 0 \leq u \leq 1 \\ 1 & , u > 1 \end{cases}$$

a) Let $V = \mathbb{F}_X(X)$.

It is a RV.

Find its CDF.

Take $0 < v < 1$

$$\mathbb{F}_V(v) = \mathbb{P}(V \leqq v) = \mathbb{P}(\mathbb{F}_X(X) \leqq v) \overset{\mathbb{F}_X \nearrow}{=\!=} \mathbb{P}(X \leqq \mathbb{F}_X^{-1}(v))$$

$$= \mathbb{F}_X(\mathbb{F}_X^{-1}(v)) = v \quad \text{for } 0 < v < 1$$

$\Rightarrow V = \mathbb{F}_X(X)$ is $\text{Unif}(0,1)$.

b) Let $W = \mathbb{F}_X^{-1}(U)$.

It is a RV.

Find its CDF.

Take w in the range of $\mathbb{F}_X^{-1}(U)$

$$\mathbb{F}_W(w) = \mathbb{P}(W \leqq w) = \mathbb{P}(\mathbb{F}_X^{-1}(U) \leqq w) =$$

$$= \mathbb{P}(U \leqq \mathbb{F}_X(w)) =$$

$$= \mathbb{F}_X(w)$$

So W and X have same CDF (same distr.).

Use this for simulation:

$$\text{Use } \underbrace{\mathbb{F}_X^{-1}(U)}_{\text{simulate}} = X \quad \text{to simulate}$$

Ex. $X \sim \mathcal{E}xp(\lambda)$ ← simulate this

First, find CDF $\mathbb{F}_X$:

$$\mathbb{F}_X(x) = 1 - e^{-\lambda x} \quad (x > 0 : \text{range of } X)$$

To get the inverse function:

Set $u = 1 - e^{-\lambda x}$ and solve for x:

$$e^{-\lambda x} = 1 - u$$

$$-\lambda x = \ln(|1-u|)$$

$$x = -\lambda^{-1} \cdot \ln(|1-u|)$$

Then, find inverse $\mathbb{F}_X^{-1}$

$$\mathbb{F}_X^{-1}(U) = -\lambda^{-1} \cdot \ln(1-U)$$

← plug-in a value from the computer

Ex. $U = 0.3$

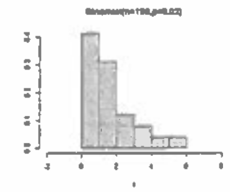
$$\Rightarrow X = -\lambda^{-1} \cdot \ln(1 - 0.3)$$

CLT

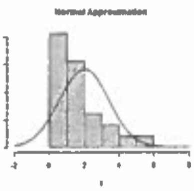
Central Limit Theorem: As $n \to \infty$

$$\frac{\bar{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}} \to N(0,1)$$

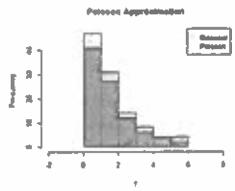
Example 1 $np = 3 < 5$



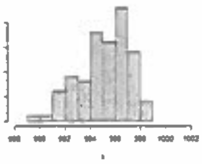
Since $np < 5$ the Normal Approximation fails



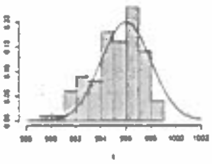
Here n is large, p is small, $np = \lambda = 3$



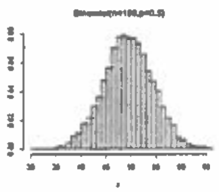
Example 2 $n(1-p) = 4 < 5$



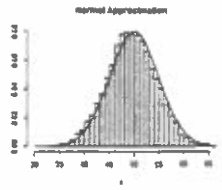
Since $n(1-p) < 5$ the Normal Approximation fails



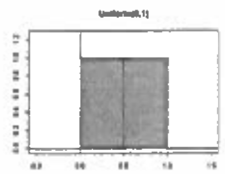
Example 3 $np = 50 > 5$, $n(1-p) = 50 > 5$



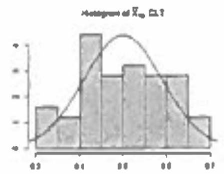
Normal Approximation works well



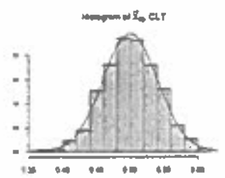
Example 4 Mean of Unif(0,1) is 0.5



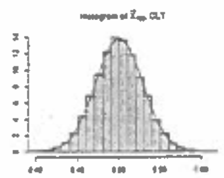
Random sample of size $n = 10 < 30$
Normal Approximation fails



Random sample of size $n = 50 > 30$
Normal Approximation is doing well



Random sample of size $n = 50 > 30$
Normal Approximation is doing better



Knowledge

You want to simulate three values of a standard normal RV. Suppose you have generated the values

0.9838, 0.5, 0.0616 of Unif (0,1).

What are the three corresponding values of the standard normal random variables?

Explain your method.

(Hint: use a standard normal distribution CDF table).

simulation

+

statistics ;)

EXERCISES 9.7 Basic Exercises

9.146 Suppose that, in Example 9.25 on page 547, the times it takes the first and second engineers to inspect the item have $\Gamma(\alpha, \lambda)$ and $\Gamma(\beta, \lambda)$ distributions, respectively.

a) Use the bivariate transformation theorem to obtain a joint PDF of S and T.
b) Use the result of part (a) to obtain and identify a marginal PDF of S.
c) Use the result of part (a) to obtain and identify a marginal PDF of T.
d) Are S and T independent random variables? Justify your answer.

9.147 Suppose that, as in Example 9.25 on page 547, the times it takes the first and second engineers to inspect the item are both exponentially distributed but that the parameters for the exponential distributions differ; say that they are λ and μ, respectively, where $\lambda \neq \mu$.

a) Use the bivariate transformation theorem to obtain a joint PDF of S and T.
b) Use the result of part (a) to obtain a marginal PDF of S.
c) Use the result of part (a) to obtain a marginal PDF of T.
d) Are S and T independent random variables? Justify your answer.

9.148 Let X and Y be continuous random variables with a joint PDF. Apply the bivariate transformation theorem to obtain a PDF of the random variable Y/X. Compare your result with Equation (9.48) on page 540.

9.149 In the petri-dish illustration of Example 9.8 on page 511, let R and Θ denote the polar coordinates of the center of the first spot (visible bacterial colony) to appear.

a) Obtain a joint PDF of R and Θ.
b) Obtain and identify marginal PDFs of R and Θ.
c) Decide whether R and Θ are independent random variables.

9.150 Let U and V be independent $\mathcal{U}(0, 1)$ random variables and let a, b, c, and d be real numbers with $a < b$ and $c < d$.

a) Obtain and identify a joint PDF of $X = a + (b - a)U$ and $Y = c + (d - c)V$.
b) Apply your result from part (a) to explain how to simulate the random (uniform) selection of a point from the rectangle $(a, b) \times (c, d)$ by using a basic random number generator.

9.151 Range and midrange of a random sample: Let $X_1, \ldots, X_n$ be a random sample from a continuous distribution with CDF F and PDF f. The *range* and *midrange* of the random sample are defined to be $R = Y - X$ and $M = (X + Y)/2$, respectively, where $X = \min\{X_1, \ldots, X_n\}$ and $Y = \max\{X_1, \ldots, X_n\}$.

a) Apply the bivariate transformation theorem to obtain a joint PDF of R and M. *Note:* Equation (9.20) on page 498 provides a joint PDF of X and Y.
b) Use the result of part (a) to obtain a marginal PDF of R. Compare your answer to that found in Example 9.6 on page 504.
c) Use the result of part (a) to obtain a marginal PDF of M. Compare your answer to that presented in Exercise 9.116 on page 542.
d) Are R and M independent random variables? Justify your answer.

9.152 Let X and Y be independent $\mathcal{N}(0, \sigma^2)$ random variables. Show that the random variables $X^2 + Y^2$ and Y/X are independent.

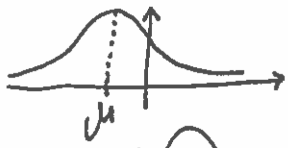
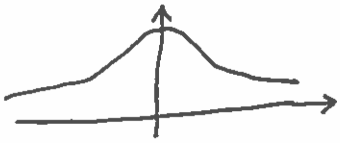
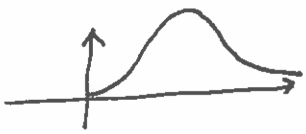
9.153 Let X and Y be independent random variables with $X \sim \Gamma(\alpha, \lambda)$ and $Y \sim \Gamma(\beta, \lambda)$. Are $X + Y$ and Y/X independent? Justify your answer.

9.154 Suppose that X and Y are continuous random variables with a joint PDF.

a) Use the bivariate transformation theorem to obtain a joint PDF of the random variables $X + Y$ and $X - Y$.

Statistics

Sampling Distributions useful in Statistics

Distr.	Parameters	PDF
Normal	μ, σ^2	
Chi-squared	$\nu \geqq 1$ (degrees of freedom)	
Student t	$\nu \geqq 1$	
F	$\nu_1, \nu_2 \geqq 1$	

Chi-squared

Def $K_\nu = \sum_{i=1}^{\nu} Z_i^2$, where $Z_i \sim N(0,1)$ iid

ν = degrees of freedom

Notation: $K_\nu \sim \chi^2_\nu$

Student t

Def $$T_\nu = \frac{\overset{\sim N(0,1)}{Z}}{\sqrt{\frac{\underbrace{K_\nu}_{\sim \chi^2_\nu}}{\nu}}}$$ indep.

$T_\nu \sim t_\nu$

F

Def $$F_{\nu_1,\nu_2} = \frac{\frac{K_{\nu_1}}{\nu_1}}{\frac{K_{\nu_2}}{\nu_2}}$$ indep.

Applications

Population has mean μ and variance σ^2

To estimate (unknown)	Assumption	Use distribution
μ	σ^2 is known	Normal
μ	σ^2 unknown	t-distribution
σ^2	μ known or not	χ^2
compare σ_1^2 and σ_2^2	μ known or not	F
compare μ_1 and μ_2	$\sigma_1^2 = \sigma_2^2$ unknown	t-distribution

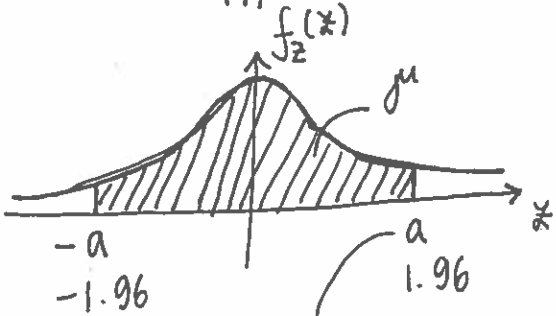
<u>Ex.</u> Estimate μ (σ^2 known) and get a Confidence Interval (CI)

<u>Hypothesis</u>: Population is $N(\mu, \sigma^2)$

μ unknown
σ^2 known

Sample of size $n \Rightarrow$ <u>estimate μ by $\overline{X}_n$</u>

Let $Z = \dfrac{\overline{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}} \sim N(0,1)$



a is some number corresponding to γ

$P(-a < Z < a) = \gamma$

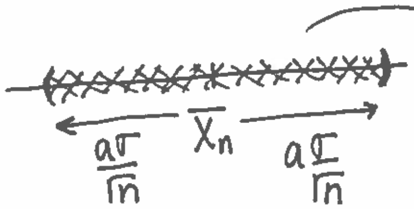
$\gamma = 0.95$
$\Rightarrow a = 1.96$

Magic computation

$$\gamma = \mathbb{P}(-a < Z < a) =$$

$$= \mathbb{P}\left(-a < \frac{\overline{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}} < a\right) =$$

$$= \mathbb{P}\left(\overline{X}_n - a\frac{\sigma}{\sqrt{n}} < \mu < \overline{X}_n + a\frac{\sigma}{\sqrt{n}}\right)$$



gives you confidence interval

- If $\gamma = 95\%$, then 95% of the random intervals (CIs) will cover μ.

$$\boxed{\begin{array}{l}\gamma\text{ - confidence interval for } \mu \\ = \left(\overline{X}_n - a\frac{\sigma}{\sqrt{n}},\ \overline{X}_n + a\frac{\sigma}{\sqrt{n}}\right)\end{array}}$$

Remarks:

1. We started with $\frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \sim N(0,1)$ (a parameter-free distr.)
2. Q: What is better - a large or a short CI?

   A: short
3. Q: How to reduce the length?

   A:
   - σ small
   - n large
   - γ small

Ex. Signal has value X received $\sim N(\mu, 4)$ $\rightarrow \sigma^2 = 4$

$\sigma = 2$

μ unknown

To reduce error
it is sent $n = 9$ times.

Find a 95% CI for μ.

Solution: $\left[\overline{X}_n - a\frac{\sigma}{\sqrt{n}}, \overline{X}_n + a\frac{\sigma}{\sqrt{n}}\right] =$

$$= \left[\overline{X}_9 - (1.96)\cdot\frac{2}{3}, \overline{X}_9 + (1.96)\cdot\frac{2}{3}\right] =$$

$$= \left[\overline{X}_9 - 1.31, \overline{X}_9 + 1.31\right]$$

- 95% of received signals will be there
- the CI will cover μ 95% of the time.

Suppose $\overline{X}_9 = 15$, then the 95% CI becomes $[13.69, 16.31]$.

cannot say
there is a 95% chance
μ will fall there,

[we are 95% confident this interval
contains μ.

$X_1, X_2, \dots, X_n$ random sample from a population X

$\left.\begin{array}{l}\mathbb{E}X = \mu \\ \mathrm{Var}\,X = \sigma^2\end{array}\right\rangle$ population parameters

<u>Estimator for μ:</u>

$$\overline{X}_n = \frac{X_1 + \dots + X_n}{n}$$

Properties: • $\mathbb{E}\overline{X}_n = \mu$ unbiased

• $\left[\begin{array}{l}\overline{X}_n \to \mu \text{ as } n \to \infty \\ \mathrm{Var}\,\overline{X}_n = \frac{\sigma^2}{n} \to 0 \text{ as } n \to \infty\end{array}\right.$ Law of large number

<u>Estimator for σ^2</u>

<u>case 1)</u> μ is known

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

$$V_n^2 \to \sigma^2 \text{ as } n \to \infty.$$

<u>case 2)</u> μ is unknown

$$S_n^2 = \frac{1}{n-1}\sum_{n=1}^{n}(X_i - \overline{X}_n)^2$$

<u>Properties:</u> • $S_n^2 \to \sigma^2$ as $n \to \infty$

• $\mathbb{E}S_n^2 = \mathbb{E}\left[\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X}_n)^2\right] = \cancel{\frac{1}{n-1}\sum_{i=1}^{n}\mathbb{E}\left[\left((X_i-\mu) - (\overline{X}_n-\mu)\right)^2\right]}$

$\cancel{= \frac{1}{n-1}\sum_{i=1}^{n}\left[\mathbb{E}(X_i-\mu)^2 + \mathbb{E}(\overline{X}_n-\mu)^2 - 2\mathbb{E}(X_i-\mu)(\overline{X}_n-\mu)\right] =}$

see next page

$\cancel{= \frac{1}{n-1}\sum_{i=1}^{n}\left[\sigma^2 + \right.}$

Proof that Sample Variance is Unbiased Plus Lots of Other Cool Stuff

Scott D. Anderson

http://www.spelman.edu/~anderson/teaching/437/unbiased/unbiased.html

Fall 1999

Expected Value of S^2

The following is a proof that the formula for the sample variance, S^2, is unbiased. Recall that it seemed like we should divide by n, but instead we divide by n-1. Here's why.

First, recall the formula for the sample variance:

$$\operatorname{var}(x) = S^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}$$

Now, we want to compute the expected value of this:

$$E[S^2] = E\left[\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}\right]$$

$$E[S^2] = \frac{1}{n-1}E\left[\sum_{i=1}^{n}(x_i - \bar{x})^2\right]$$

Now, let's multiply both sides of the equation by n-1, just so we don't have to keep carrying that around, and square out the right side, just like we did with that shortcut formula for SSX, above.

$$(n-1)E[S^2] = E\left[\sum_{i=1}^{n} x_i^2 - 2\bar{x}x_i + \bar{x}^2\right]$$

$$E[Y^2] = E[\bar{x}^2] = \frac{1}{n^2}\sum \mathrm{var}[x_i] + \mu^2$$

$$E[Y^2] = E[\bar{x}^2] = \frac{1}{n^2}\sum \sigma^2 + \mu^2$$

$$E[Y^2] = E[\bar{x}^2] = \frac{1}{n^2}(n\sigma^2) + \mu^2$$

$$E[Y^2] = E[\bar{x}^2] = \frac{1}{n}\sigma^2 + \mu^2$$

We can substitute this stuff for the second term on the RHS of equation 1. Also, note that the first term on the RHS of eqation 1 is the second moment of X, so that can also be re-written. Doing both substitutions gives us:

$$\frac{n-1}{n}E[S^2] = [\sigma^2 + \mu^2] - \left[\frac{1}{n}\sigma^2 + \mu\right]$$

$$\frac{n-1}{n}E[S^2] = \sigma^2 - \frac{1}{n}\sigma^2$$

$$E[S^2] = \sigma^2$$

Whew! That was hard, but solvable. This is why S^2 with the n-1 denominator is an unbiased estimator.

CI for μ when σ is unknown

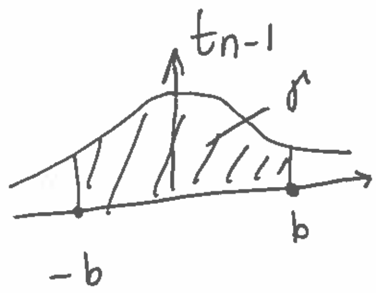
$\downarrow$ use S_n because σ is not known

recall $\chi_n^2 = \sum_{i=1}^{n} Z_i^2$ ← iid $N(0,1)$

$$\frac{(n-1)}{\sigma^2} \cdot S_n^2 = \sum_{i=1}^{n} \left(\frac{X_i - \overline{X}_n}{\sigma} \right)^2 \sim \chi^2_{n-1}$$

Look at $T_n = \dfrac{\overline{X}_n - \mu}{S_n/\sqrt{n}} = \dfrac{\dfrac{\overline{X}_n - \mu}{\sigma/\sqrt{n}}}{\sqrt{\dfrac{1}{\sigma^2} \dfrac{(n-1)S_n^2}{(n-1)}}} =$

$$= \frac{\underbrace{\dfrac{\overline{X}_n - \mu}{\sigma/\sqrt{n}}}_{\sim N(0,1)}}{\sqrt{\dfrac{\underbrace{\frac{(n-1)S_n^2}{\sigma^2}}_{\sim \chi^2_{n-1}}}{(n-1)}}} \sim t_{n-1}$$



$\gamma = \mathbb{P}(-b < t_{n-1} < b)$

parameter - free

if $\gamma = 95\%$, then $b = 2.306$

$= \mathbb{P}\left(-b \leq \dfrac{X_n - \mu}{\frac{S_n}{\sqrt{n}}} \leq b \right) = \mathbb{P}\left(\overline{X}_n - b\dfrac{S_n}{\sqrt{n}} \leq \mu \leq \overline{X}_n + b\dfrac{S_n}{\sqrt{n}} \right)$

HW 12

1) 11.59 p. 672

2) 11.61 p. 673

3) 11.91 p. 679

4) 11.95 p. 680

5) 11.98 p. 680

Hint a) leave it as a sum

b) no need for continuity correction because $\underline{\underline{n}}$ is very large

6) 11.63 p. 673

Extra credit:

1) p. 673 11.66

Hint: use ln.

HW 12

Due date: Aug 04 2014 (Monday)

Midterm II: July 31

Chapters 8, 9, 10

(no Ch 11!)

08/04 Statistics

08/05 Markov chain

08/07 Final

Other Central Limit Theorems

In this section, we examined the central limit theorem in its classical form. Many other versions of the central limit theorem have now been established and can be found in the extensive literature on the subject. These versions weaken, in one way or another, the independence and/or identically-distributed assumptions of the classical central limit theorem, thus providing normal approximations in contexts where the random variables under consideration may not be independent or may not be identically distributed.

EXERCISES 11.4 Basic Exercises

11.55 As reported by a spokesperson for Southwest Airlines, the no-show rate for reservations is 16%—that is, the probability is 0.16 that a person making a reservation will not take the flight. For a certain flight, 42 people have reservations. For each part, determine and compare the exact probability by using the appropriate binomial PMF and an approximate probability by using the integral De Moivre–Laplace theorem (in the form of Proposition 11.11 on page 661). The probability that the number of people who don't take the flight is
a) exactly 5. **b)** between 9 and 12, inclusive. **c)** at least 1. **d)** at most 2.

11.56 In Exercise 8.108 on page 449, you were asked to determine and compare the exact probability of each event in Exercise 11.55 by using the appropriate binomial PMF and an approximate probability by using the local De Moivre–Laplace theorem (in the form of Proposition 8.12 on page 445).
a) If you didn't previously do Exercise 8.108, do it now.
b) Compare the results for all three methods of obtaining the probabilities by constructing a table similar to Table 11.2 on page 663.
c) Discuss your results from part (b) in light of the rule of thumb for using the normal approximation to the binomial distribution.

11.57 Refer to the roulette illustration of Example 11.20 on page 670. Use the integral De Moivre–Laplace theorem to estimate the probability that the gambler will be ahead after
a) 100 bets. **b)** 1000 bets. **c)** 5000 bets.

11.58 Some boys want to play football at a park that doesn't have a formal football field. In an attempt to approximate the length of a football field, one of the boys tries to step-off 100 yards. In actuality, his steps are independent and identically distributed random variables with mean 0.95 yard and standard deviation 0.08 yard. Determine the approximate probability that the distance stepped-off by the boy is within
a) 4 yards of the length of a football field.
b) 6 yards of the length of a football field.

X **11.59** A brand of flashlight battery has normally distributed lifetimes with a mean of 30 hours and a standard deviation of 5 hours. A supermarket purchases 500 of these batteries from the manufacturer. What is the probability that at least 80% of them will last longer than 25 hours? Use the integral De Moivre–Laplace theorem.

11.60 The checkout times at the local food market are independent random variables having mean 3.5 minutes and standard deviation 1.5 minutes.
a) Determine the probability that it will take at least 6 hours to check out 100 customers.
b) Find the probability that the mean checkout time for 100 customers is less than 3.4 minutes.

11.61 In a large city, annual household incomes have a mean of \$35,216 and a standard deviation of \$3,134.

a) What is the probability that the average income of 160 randomly chosen households is below \$34,600?
b) Strictly speaking, in part (a), are the assumptions for the central limit theorem satisfied? Explain your answer.
c) Why is it permissible to use the central limit theorem to solve part (a)?

11.62 The claim amount for a health insurance policy follows a distribution with density function $f(x) = (1/1000)e^{-x/1000}$ for $x > 0$, and $f(x) = 0$ otherwise. The premium for the policy is set at 100 over the expected claim amount. Suppose that 100 policies are sold and that claim amounts are independent of one another.

a) Identify the exact probability distribution of the total claim amount for the 100 policies.
b) Use your answer from part (a) to obtain an expression that gives the exact probability that the insurance company will have claims exceeding the premiums collected.
c) Use the central limit theorem to obtain the approximate probability that the insurance company will have claims exceeding the premiums collected.
d) If you have access to statistical software, use it and your answer from part (a) to obtain the probability that the insurance company will have claims exceeding the premiums collected. Compare this probability to that found in part (c).

11.63 Let X and Y denote the number of hours that a randomly selected person watches movies and sporting events, respectively, during a 3-month period. Assume that $\mathcal{E}(X) = 50$, $\mathcal{E}(Y) = 20$, $\text{Var}(X) = 50$, $\text{Var}(Y) = 30$, and $\text{Cov}(X, Y) = 10$. If 100 people are randomly selected and observed for 3 months, what is the probability that the total number of hours that they watch movies or sporting events is at most 7100?

11.64 A hardware manufacturer knows from experience that 95% of the screws produced by his company are within tolerance specifications. Each shipment of 10,000 screws comes with a warranty that promises a complete refund if more than r screws aren't within tolerance specifications. How small can r be chosen so that no more than 1% of shipments will require complete refunds? Solve this problem by

a) using Chebyshev's inequality.
b) using the central limit theorem.
c) If you have access to statistical software, use it to obtain the exact value of r and compare your answer to those found in parts (a) and (b).

11.65 An air-conditioning contractor plans to offer service contracts on the brand of compressor used in all of the units her company installs. First she must estimate how long those compressors last, on average. To that end, the contractor consults records on the lifetimes of 250 previously used compressors. She plans to use the sample mean lifetime of those compressors as her estimate for the mean lifetime of all such compressors. If the lifetimes of this brand of compressor have a standard deviation of 40 months, what is the probability that the contractor's estimate will be within 5 months of the true mean?

11.66 A particle of initial size s is subjected to repeated impacts. After impact j, a proportion X_j of the particle remains; that is, if Y_j denotes the size of the particle after impact j, then $Y_j = X_j Y_{j-1}$. The random variables $X_1, X_2, \ldots$ are independent, all having the same probability distribution as a random variable X.

a) Use the central limit theorem to identify the approximate probability distribution of Y_n for large n. State explicitly any assumptions that you make.
b) Specialize the results obtained in part (a) to the cases where $X \sim \mathcal{U}(0, 1)$, $X \sim \text{Beta}(2, 1)$, $X \sim \text{Beta}(1, 2)$, and $X \sim \text{Beta}(2, 2)$.

e) Does $X + Y$ have a binomial distribution? Justify your answer.
f) Use MGFs to decide whether X and Y are independent random variables.

11.88 Simulation: This exercise requires access to a computer or graphing calculator with statistical software.
a) Use a Poisson random number generator—that is, a random number generator that simulates observations of a Poisson random variable—to obtain 10,000 observations of a $\mathcal{P}(3)$ random variable.
b) Without doing further computations, roughly what should be the average value of the 10,000 numbers obtained in part (a)? Explain your reasoning.
c) Determine the average value of the 10,000 numbers obtained in part (a) and compare your result to your answer in part (b).

11.89 Monte Carlo integration: In this exercise, you are to extend the technique of Monte Carlo integration considered in Exercises 11.43 and 11.44 on page 656. Suppose that you want to determine the value of $\int_a^b g(x)\,dx$, where g is a Riemann integrable function on the interval $[a, b]$. Further assume that no simple formula exists for the antiderivative of g. Let f be a PDF such that $f(x) > 0$ for $x \in [a, b]$, and $f(x) = 0$ for $x \notin [a, b]$.
a) Let X be a random variable with probability density function f. Show that the random variable $g(X)/f(X)$ has finite expectation.
b) Let $X_1, X_2, \ldots$ be independent and identically distributed random variables with common PDF f. Set

$$\hat{I}_n(f, g) = \frac{1}{n}\sum_{k=1}^{n}\frac{g(X_k)}{f(X_k)}.$$

Show that $\lim_{n\to\infty} \hat{I}_n(f, g) = \int_a^b g(x)\,dx$ with probability 1.
c) Determine when $\hat{I}_n(f, g)$ has finite variance and, in such a case, obtain the variance.
d) Show that, if $a = 0$, $b = 1$, and f is the PDF of a $\mathcal{U}(0, 1)$ random variable, the result of part (b) reduces to the Monte Carlo technique examined in Exercise 11.43. Obtain the variance of $\hat{I}_n(f, g)$ in this case.

11.90 A machine produces a large number of items per day. The quality control engineer samples 100 items from the daily output and declares the machine "out of control" and in need of service if 12 or more of the sampled items are defective. If the machine is actually producing 8% defective items, what is the probability that it will be declared "out of control"?

X **11.91** The second leading genetic cause of mental retardation is Fragile X Syndrome, named for the fragile appearance of the tip of the X chromosome in affected individuals. Worldwide, 1 in every 1500 males is affected, with no ethnic bias. For a sample of 10,000 males, use the integral De Moivre–Laplace theorem (in the form of Proposition 11.11 on page 661) to determine the probability that the number who have Fragile X Syndrome
a) exceeds 7. b) is at most 10.

11.92 The probabilities in parts (a) and (b) of Exercise 11.91 were obtained in Exercise 8.111 on page 450 by using the local De Moivre–Laplace theorem (in the form of Proposition 8.12 on page 445) and in Exercise 5.80 on page 226 by using the Poisson approximation to the binomial distribution.
a) If you didn't previously do Exercises 5.80 and 8.111, do them now.
b) Compare your probability estimates for the three methods (local and integral De Moivre–Laplace approximations and the Poisson approximation).

11.93 Acute rotavirus diarrhea is the leading cause of death among children under age 5, killing an estimated 4.5 million annually in developing countries. Scientists from Finland and

Belgium claim that a new oral vaccine is 80% effective against rotavirus diarrhea. Assuming that the claim is correct, use the integral De Moivre–Laplace theorem to find the probability that, of 1500 cases, the vaccine will be effective in
a) exactly 1225 cases. b) at least 1175 cases.
c) between 1150 and 1250 cases, inclusive.

11.94 As we stated on page 667, a rule of thumb for using the normal approximation is that $n \geq 30$. However, sometimes the approximation works well for quite small n, as we noted in Example 11.19(c) on page 667. Let X_1, X_2, and X_3 be independent and identically distributed random variables, all having the discrete uniform distribution on the set $\{1, 2, 3\}$.
a) By using either a direct listing or counting, determine the PMF of the random variable $X_1 + X_2 + X_3$.
b) Use the result of part (a) to obtain the exact value of $P(X_1 + X_2 + X_3 \leq 7)$.
c) Construct a probability histogram for the random variable $X_1 + X_2 + X_3$ and comment on its degree of resemblance to a normal curve.
d) Obtain an approximate value of the probability required in part (b) by using the normal approximation with a continuity correction. Compare this value to that obtained in part (b).

11.95 Experience has shown that scores on a certain aptitude test have a mean of 73 points and a standard deviation of 5 points. If the aptitude test is given to 140 people, what is the probability that their mean score will exceed 74?

11.96 In Exercise 11.70 on page 674, you were asked to consider the following problem: "A light is continually kept on. Each time a bulb burns out, it's immediately replaced by another bulb. There is a supply of 20 bulbs and the lifetime of each bulb is exponentially distributed with mean 300 hours, independent of the lifetimes of the other bulbs. What is the probability that all bulbs have burned out by 7000 hours?" Answer that same question if replacement time isn't immediate but takes a mean of 0.5 hour with a standard deviation of 0.1 hour, and successive replacement times are independent of one another and of the lifetimes of the bulbs.

11.97 In an analysis of healthcare data, based on a random sample of 48 people, ages have been rounded to the nearest multiple of 5 years. What is the approximate probability that the mean of the rounded ages is within 0.25 year of the mean of the true ages? State explicitly any assumptions that you make.

11.98 An insurance company issues 1250 vision care insurance policies. The number of claims filed by a policyholder under a vision care insurance policy during a 1-year period is a Poisson random variable with mean 2. Assume that the numbers of claims filed by distinct policyholders are independent of one another.
a) Provide an expression for the exact probability that the total number of claims filed during a 1-year period is between 2450 and 2600, inclusive.
b) Use the central limit theorem to approximate the probability that the total number of claims filed during a 1-year period is between 2450 and 2600, inclusive.

11.99 Suppose that $X_1, X_2, \ldots$ are independent and identically distributed positive random variables. Further suppose that $\ln X_j$ has finite nonzero variance. Show that, for large n, the random variable $\prod_{j=1}^{n} X_j$ has approximately a lognormal distribution. *Note:* Refer to Exercise 8.154 on page 474.

11.100 According to Scarborough Research, more than 85% of working adults commute by car. Of all U.S. cities, Washington, D.C., and New York City have the longest mean commute times. Suppose that 30 randomly obtained commute times in Washington, D.C., have a mean of 27.97 minutes. Assuming that the standard deviation of all such commute times is 10.04 minutes, determine a 90% confidence interval for the mean commute time in Washington, D.C. Interpret your result in words.

The law of large numbers

Let $X_1, X_2, \dots, X_n$ iid with mean $\mathbb{E}X$ and variance $Var(X)$

Average $\underbrace{\overline{X_n}}_{RV} = \frac{X_1 + X_2 + \dots + X_n}{n}$

We saw $\mathbb{E}[\overline{X_n}] = \mathbb{E}X$

no cov because indep

$$Var(\overline{X_n}) = Var\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\cdot\sum_{i=1}^{n} Var(X_i) =$$

$$= \frac{1}{n^2}\cdot n\,Var(X) = \frac{Var X}{n} \xrightarrow[n\to\infty]{} 0$$

- performing the same experiment a large number of times
- the result obtained from a large number of trials should be close to the expected value, and will tend to become closer as more trials are performe

Central Limit Theorem and Normal Distribution

Let $X_1, X_2, \ldots$ be iid RVs

$$\mu = \mathbb{E} X_j$$
$$\sigma^2 = \operatorname{Var} X_j$$

We do not know the distr. of X_j, only the mean and variance.

Let $S_n = X_1 + X_2 + \cdots + X_n$

Then $\mathbb{E} S_n = n\mu$

$\operatorname{Var} S_n = n\sigma^2$

Standardized Sum

$$Y_n = \frac{S_n - \mathbb{E} S_n}{\sqrt{\operatorname{Var} S_n}} = \frac{S_n - n\mu}{\sigma\sqrt{n}}$$

$\mathbb{E} Y_n = 0$

$\operatorname{Var} Y_n = 1$ Distr. of $Y_n \not\equiv$ Distr. of Z

Let $Z \sim N(0,1)$

BUT

$\boxed{\text{CLT: As } n \to \infty, \text{ the distribution of the standardized sum } Y_n \text{ tends to the normal } (0,1) \text{ distribution, that is } \lim_{n\to\infty} \mathbb{P}(Y_n \leq x) = \mathbb{P}(Z \leq x)}$

$$\underbrace{\mathbb{F}_{Y_n}(x)}_{\text{CDF of } Y_n} \xrightarrow[n\to\infty]{} \underbrace{\Phi(x)}_{}$$

CDF of Y_n converges to CDF of standard normal as $n \to \infty$.

Normal approximation

For large n, $\mathbb{P}(Y_n \leq x) \approx \mathbb{P}(Z \leq x)$

To apply the CLT, S_n must be the sum of iid RV's, and Y_n is the standardized S_n.

Ex. Nicotine content in a single cigarette has mean $\mu = 0.8$ mg and variance $\sigma^2 = (0.1)^2 \text{mg}^2$.

John smokes 5 packs a week = 100 cigarettes /week

Find $\mathbb{P}$(John consumes at least 82 mg Nicotine in a week).

Solution: X_j = amount of Nicotine John gets in cigarette j

$j = 1, \dots, 100$

Then $S_n = S_{100} = X_1 + X_2 + \cdots + X_{100}$

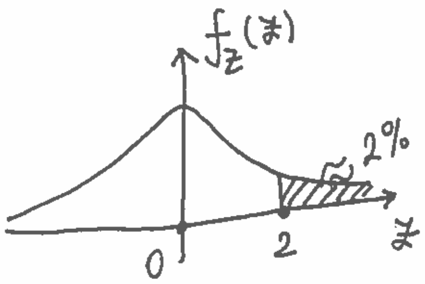
$\mathbb{E} S_n = n\mu = (100).(0.8) = 80$ mg

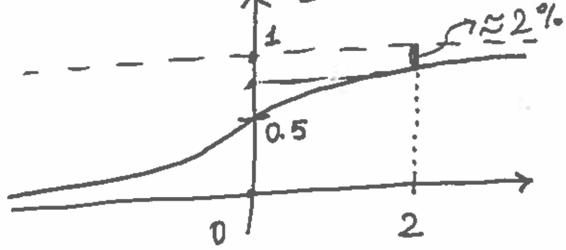
$\sqrt{\operatorname{Var} S_n} = \sigma\sqrt{n} = (0.1).(10) = 1$ mg

$$\mathbb{P}(S_n \geqq 82) = \mathbb{P}\left(\frac{S_n - \mathbb{E}S_n}{\sqrt{\operatorname{Var} S_n}} \geqq \frac{82-80}{1}\right) = \mathbb{P}(Y_n \geqq 2)$$

$$\approx \mathbb{P}(Z \geqq 2) = 0.0228 \approx 2\%$$

$\mathbb{F}_Z(x) = \Phi(x) = \mathbb{P}(Z \leq x)$

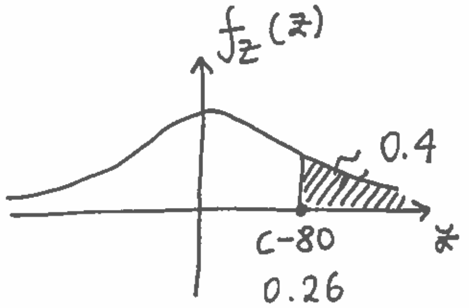




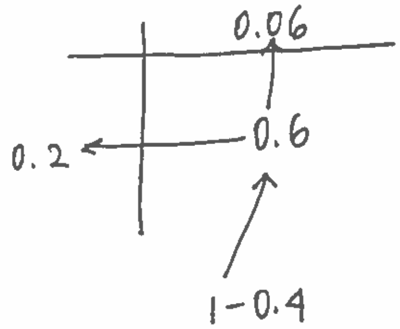
Ex. In previous example, find $\underset{\sim}{c}$ such that

$\mathbb{P}$(John consumes more than c mg Nicotine in a week) = 0.4

$$0.4 = \mathbb{P}(S_n \geqq c) = \mathbb{P}\left(Y_n \geqq \frac{c-80}{1}\right) \approx \mathbb{P}(Z \geqq c-80)$$



Table



$\Rightarrow 0.26 = c - 80$

$\Rightarrow \boxed{c = 80.26 \text{ mg}}$

How to generate a standard normal RV

Ex. Let $X_j \sim \text{Unif}(0,1)$

$n = 12$

$S_{12} = X_1 + \cdots + X_{12}$

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

$$\mathbb{E}X^2 = \int_0^1 x^2 dx = \frac{1}{3} \Rightarrow \sigma^2 = \frac{1}{3} - \frac{1}{4} = \frac{1}{12} \Rightarrow \sigma = \frac{1}{\sqrt{12}}$$

$$Y_{12} = \frac{S_{12} - \mathbb{E}S_{12}}{\sqrt{\text{Var}\, S_{12}}} = \frac{S_{12} - 6}{\cancel{\sqrt{12}} \cdot \frac{1}{\cancel{\sqrt{12}}}} = \boxed{S_{12} - 6 \approx Z \sim N(0,1)}$$

So, to generate standard normal RV, generate 12 Uniforms on [
add them together, and subtract 6.

Approximation of the Binomial

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

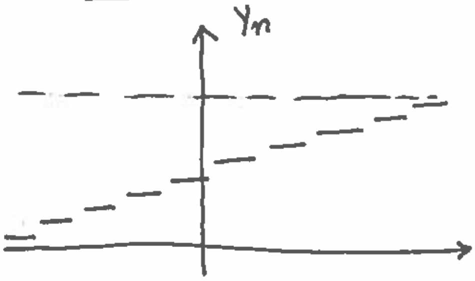
↳ sum of iid RV

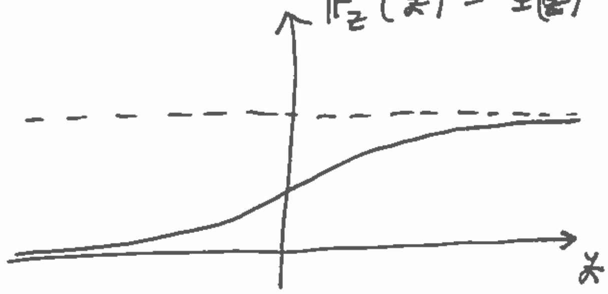
↳ Bernoulli(p) $\quad \mu = p$

$\sigma = \sqrt{p(1-p)}$

$$\Rightarrow Y_n = \frac{S_n - np}{\sqrt{np(1-p)}} \approx Z \sim N(0,1) \text{ if } n \text{ is large}$$

Difficulty: plot CDF





Use continuity correction!

Theorem (De Moivre - Laplace) Proposition 11.11 p. 661

Let $U \sim \text{Binomial}(n, p)$

$\mathbb{P}(a \leq U \leq b)$, a and b are integers $0 \leq a < b \leq n$,

$$\approx \mathbb{P}\left(Z \leq \frac{b + \frac{1}{2} - np}{\sqrt{np(1-p)}}\right) - \mathbb{P}\left(Z \leq \frac{a - \frac{1}{2} - np}{\sqrt{np(1-p)}}\right)$$

Ex. Let $U \sim \text{Binomial}(n = 49, p = \frac{1}{2})$

a. compute $\mathbb{P}(U = 28)$ exactly;

b. compute $\mathbb{P}(U = 28)$ using the normal approximation without continuity correction;

c. same as b. but with continuity correction.

Solution:

a. $\binom{49}{28}\left(\frac{1}{2}\right)^{28}\left(1-\frac{1}{2}\right)^{49-28} = \binom{49}{28}\left(\frac{1}{2}\right)^{49} = 0.069366$

$\approx \boxed{0.0694}$

b. $\mathbb{P}(U=28) = \mathbb{P}\left(\frac{U-\mathbb{E}U}{\sqrt{\mathrm{Var}U}} = \frac{28-24.5}{3.5}\right) \approx \mathbb{P}(Z=1) = \boxed{0}$

c. $\mathbb{P}(U=28) = \mathbb{P}(27.5 \leqq U \leqq 28.5) =$

$$= \mathbb{P}\left(\frac{27.5-24.5}{3.5} \leqq \frac{U-\mathbb{E}U}{\sqrt{\mathrm{Var}U}} \leqq \frac{28.5-24.5}{3.5}\right) \approx$$

$$\approx \mathbb{P}\left(\underbrace{\frac{3}{3.5}}_{0.858} \leqq Z \leqq \underbrace{\frac{4}{3.5}}_{1.143}\right) =$$

$$= \Phi(1.143) - \Phi(0.858) =$$

$$= 0.87351 - 0.804317 = 0.069134$$

$\approx \boxed{0.0691}$

How large should n be?

For Binomial require $np > 5$ AND $n(1-p) > 5$.

In general, for cont. distributions n can be as small as 6.

Remarks:

Q: When to use continuity correction?

A: When S_n is discrete.

Q: When not to use it?

A: When S_n is continuous or we don't know the distr. of S_n

Q: How to use it?

A: Let a, b = integers

RV U = discrete

Then
- $\mathbb{P}(a \leq U \leq b) = \mathbb{P}(a - \frac{1}{2} \leq U \leq b + \frac{1}{2})$
- $\mathbb{P}(U = a) = \mathbb{P}(a - \frac{1}{2} \leq U \leq a + \frac{1}{2})$
- $\mathbb{P}(a < U < b) = \mathbb{P}(a+1 \leq U \leq b-1) =$

$$= \mathbb{P}(a+1-\tfrac{1}{2} \leq U \leq b-1+\tfrac{1}{2}) =$$

$$= \mathbb{P}(a+\tfrac{1}{2} \leq U \leq b-\tfrac{1}{2})$$

Different ways of expressing the normal approximation:

1. For large n, $Y_n = \dfrac{S_n - \mathbb{E}S_n}{\sqrt{\operatorname{Var} S_n}} \approx N(0,1)$.

2. For large n, $S_n \approx N(\mathbb{E}S_n, \operatorname{Var} S_n)$.

3. Use $\overline{X}_n = \dfrac{X_1 + \dots + X_n}{n} = \dfrac{S_n}{n}$

$$\boxed{\text{CLT: } \frac{\overline{X}_n - \mu}{\sigma/\sqrt{n}} \xrightarrow[n\to\infty]{\text{distr.}} N(0,1)}$$

$\mathbb{E}\overline{X}_n = \mathbb{E}X = \mu$

$\operatorname{Var}\overline{X}_n = \dfrac{\operatorname{Var} X}{n} = \dfrac{\sigma^2}{n}$

4. standardized: $\dfrac{\overline{X}_n - \mathbb{E}\overline{X}_n}{\sqrt{\operatorname{Var}\overline{X}_n}} = \dfrac{\overline{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}} = \boxed{\dfrac{\sqrt{n}(\overline{X}_n - \mu)}{\sigma} \approx N(0,1)}$.

Ex. Impurity in a batch of chemicals has mean $\mu = 4.0$ g and variance $\sigma^2 = (1.5)^2$ g. Prepare $n = 50$ batches.

Let $\overline{X}_n$ = average amount of impurity in a batch.

We know $\underbrace{\overline{X}_n \longrightarrow \mu}_{\text{law of large numbers}}$ as $n \to \infty$ (converges to its mean)

but for finite n, $\overline{X}_n$ is random.

Compute $\mathbb{P}(3.5 \leq \overline{X}_n \leq 3.8)$

$$= \mathbb{P}\left(\frac{3.5-4}{\frac{1.5}{\sqrt{50}}} \leq \frac{\overline{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}} \leq \frac{3.8-4}{\frac{1.5}{\sqrt{50}}} \right) \approx$$

$$\approx \mathbb{P}(-3.26 \leq Z \leq -0.94) = \Phi(-0.94) - \Phi(-3.26) =$$

$$= 0.1729 - 0.0091 = 0.1638$$

Q: How to get $\Phi(z) = \mathbb{P}(Z \leq z)$ for $z < 0$ from table?

Ex. $\Phi(-2) = 1 - \Phi(2)$

$\mathbb{P}(Z \leq -2) = \mathbb{P}(Z \geq 2)$

Ex. 11.69 p. 674

Skip

11.69 George and Julia both work for the campus coffee shop. Sales are slow and management is considering letting one of them go. To decide which employee to keep, each is timed on 40 independent trials of making a double decaf skim milk latte. If the sample mean times differ by more than 3 seconds, the person with the larger sample mean will be let go; otherwise, both will be kept. The standard deviation of the time it takes each person to make a double decaf skim milk latte is 4 seconds. If the mean times for both George and Julia are actually the same, what is the probability that George will be let go?

X, Y indep.

$\mu_X = ?$ $\mu_Y = ?$ $\sigma_X = \sigma_Y = 4$

$n = 40$

let $D = X - Y$ be the difference

Test the hypothesis $H_0: \mu_X = \mu_Y$

Reject if $D = \overline{X}_n - \overline{Y}_n \geqq 3$

(If $D = \overline{X}_n - \overline{Y}_n < 3$ we will say hypothesis holds

Compute $\underbrace{\mathbb{P}(D > 3 \mid H_0 \text{ is true})}_{\text{probability of false positive}}$

we would conclude the test is positive (Reject), but that is false

Under H_0: $\mu_X = \mu_Y \Rightarrow \mathbb{E}D = \mathbb{E}\overline{X}_n - \mathbb{E}\overline{Y}_n = \mu_X - \mu_Y = 0$

$\operatorname{Var} D = \operatorname{Var} \overline{X}_n + \operatorname{Var} \overline{Y}_n = \frac{\sigma_X^2}{n} + \frac{\sigma_Y^2}{n} = \frac{16}{40} + \frac{16}{40} = 0.$

$\Rightarrow \mathbb{P}(D > 3 \mid \text{under } H_0) =$

$= \mathbb{P}(\overline{X}_n - \overline{Y}_n > 3) \approx \mathbb{P}\left(Z > \frac{3-0}{\sqrt{0.8}}\right) = \mathbb{P}(Z > 3.3542) =$

$= 1 - \mathbb{P}(Z \leqq 3.35) \approx 0.000398$

$\Rightarrow$ $\mathbb{P}$(false positive is very small).

Normal approximation for Gamma(n, λ)

n ← integer

Why? To apply the normal approximation, we must have a sum of iid random variables, and Gamma(n, λ) is the sum of n iid Exp(λ)

$\Rightarrow$ $S_n \sim$ Gamma(n, λ)

$S_n = X_1 + \cdots + X_n$, $X_i \sim$ Exp(λ), $i = 1, \ldots, n$ indep.

$$\mathbb{E}X_i = \mu = \frac{1}{\lambda}$$

$$\text{Var}\, X_i = \sigma^2 = \frac{1}{\lambda^2}$$

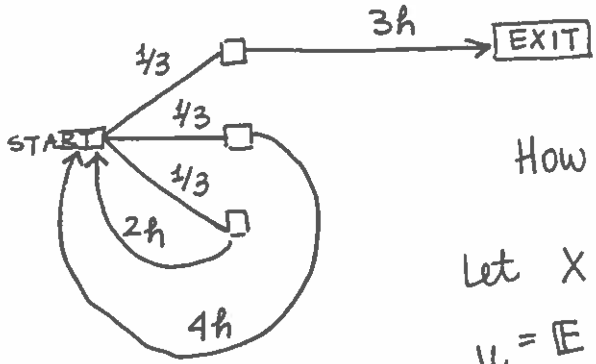
$$\mathbb{E}S_n = n\mu = \frac{n}{\lambda}$$

$$\text{Var}\, S_n = n\sigma^2 = \frac{n}{\lambda^2}$$

$\Rightarrow$ For large n, Gamma$(n, \lambda) \approx N\left(\frac{n}{\lambda}, \frac{n}{\lambda^2}\right)$.

Markov Chains

Ex.



How long on average will it take to e

Let X = time it takes you to exit

$\mu = \mathbb{E}[X] = ?$

$$\mu = \frac{1}{3}\cdot 3 + \frac{1}{3}(4+\mu) + \frac{1}{3}(2+\mu) =$$

Condition on the first choice

$$= 1 + \frac{4}{3} + \frac{\mu}{3} + \frac{2}{3} + \frac{\mu}{3} =$$

$$= 3 + \frac{2}{3}\mu \Rightarrow \mu = 3 + \frac{2}{3}\mu$$

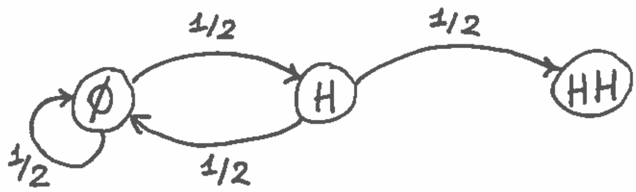
$$\frac{1}{3}\mu = 3$$

$$\boxed{\mu = 9\,h}$$

Ex. Fair coin

How many throws will it take on average until you get two H in a row?

Build a Markov process by:



Want to find μ_ϕ = mean time to go from ϕ to HH.

Let μ_H = mean time to go from H to HH.

Condition on next throw

$$\mu_\phi = \frac{1}{2}(1+\mu_\phi) + \frac{1}{2}(1+\mu_H)$$

$$\mu_H = \frac{1}{2}(1+\mu_\phi) + \frac{1}{2}\cdot 1$$

$$\boxed{\mu_\phi = 6}$$

$$\mu_H = 4$$

Stochastic process

$X_0, X_1, X_2, \dots \qquad (X_n)_{n=0,1,2,\dots}$

n represents "time" (discrete)

The random variable X_n can take values on a set I of "sta

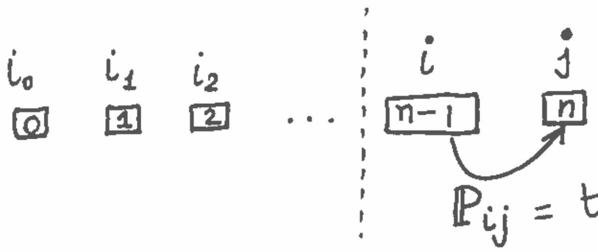
Usually $I = \{0, 1, \dots, N\}$ finite

$I = \{0, 1, 2, \dots\}$ countable infinite

Discrete-time Markov chain

is a discrete-time stochastic process, such that

$$\mathbb{P}(X_n = j \mid \underbrace{X_{n-1} = i, X_{n-2} = i_{n-2}, \dots, X_0 = i_0}_{\text{history until time } n}) = \mathbb{P}(X_n = j \mid \underbrace{X_{n-1} = i}_{\substack{\text{only depends on the state}\\ \text{of the system}\\ \text{on the previous time step}}}) \overset{\text{denote}}{=} \underbrace{\mathbb{P}_{ij}}_{\text{transiti probab}}$$



$\mathbb{P}_{ij}$ = transition probability to move to state j given that you were at state i at the previous time step.

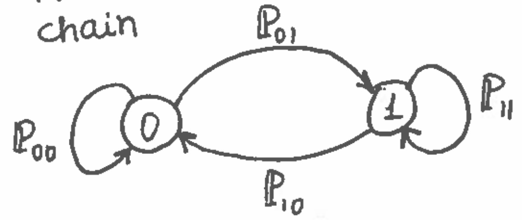
- $\mathbb{P}_{ij} \geq 0$
- $\sum_{j \in I} \mathbb{P}_{ij} = 1 \rightarrow \boxed{n-1}$ i, $\boxed{n}$ $\underbrace{0, 1, \dots}_{I}$ (we have to move to some state from I at time n.)

- We also define

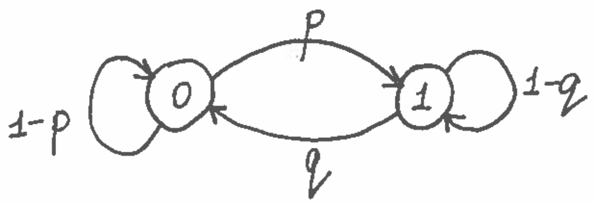
transition matrix

$$\mathbb{P} = \begin{array}{c|c} & \begin{array}{cccc} 0 & 1 & 2 & \cdots \end{array} \\ \begin{array}{c} 0 \\ 1 \\ 2 \\ \vdots \end{array} & \begin{pmatrix} \mathbb{P}_{00} & \mathbb{P}_{01} & \mathbb{P}_{02} & \cdots \\ \mathbb{P}_{10} & \mathbb{P}_{11} & \mathbb{P}_{12} & \cdots \\ \mathbb{P}_{20} & \mathbb{P}_{21} & \mathbb{P}_{22} & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{pmatrix} \end{array} \rightarrow \text{sum of each row is } 1$$

- Alternative way to describe chain

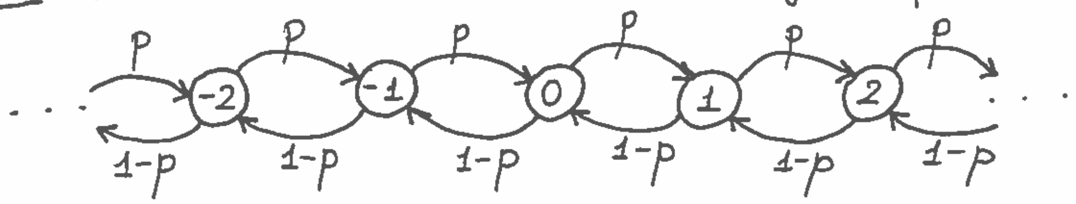


Ex.



$$\mathbb{P} = \begin{pmatrix} 1-p & p \\ q & 1-q \end{pmatrix}$$

Ex. Random walk: $\mathbb{P}(\text{walk to the right}) = p$, $\mathbb{P}(\text{walk left}) = 1-p$



$$\mathbb{P} = \begin{pmatrix} 0 & p & 0 & 0 & 0 & 0 & \cdots \\ 1-p & 0 & p & 0 & 0 & 0 & \cdots \\ 0 & 1-p & 0 & p & 0 & 0 & \cdots \\ 0 & 0 & 1-p & 0 & p & 0 & \cdots \\ \vdots & \vdots & 0 & & & \ddots & \end{pmatrix}$$

P_{ij} = one-step transition probabilities for $n = 1, 2, \dots$

note: $\left(\mathbb{P}_{ij}^{(1)} = \mathbb{P}_{ij}\right)$

Define $P_{ij}^{(m)} = \mathbb{P}(X_m = j \mid X_0 = i)$

given that you start in state i, the prob. to be in state j after m steps

Chapman - Kolmogorov (C.K.) equations:

for $m = 1, 2, \dots$
$n = 1, 2, \dots$

then $$\boxed{\mathbb{P}_{ij}^{(m+n)} = \sum_{k \in I} \mathbb{P}_{ik}^{(m)} \cdot \mathbb{P}_{kj}^{(n)}}$$

Proof: $P_{ij}^{(m+n)} = \mathbb{P}(X_{m+n} = j \mid X_0 = i) = \sum_{k \in I} \mathbb{P}(X_{m+n} = j, X_m = k \mid X_0 = i) =$

$$= \sum_{k \in I} \mathbb{P}(X_{m+n} = j \mid X_m = k, X_0 = i) \cdot \mathbb{P}(X_m = k \mid X_0 = i) =$$

$$\overset{\text{Markovian property}}{=} \sum_{k \in I} \mathbb{P}(X_{m+n} = j \mid X_m = k) \cdot \mathbb{P}(X_m = k \mid X_0 = i) =$$

$$= \sum_{k \in I} \mathbb{P}_{kj}^{(n)} \cdot \mathbb{P}_{ik}^{(m)}$$

For $n = 1, 2, \dots$ define $\mathbb{P}^{(n)}$ - matrix of n-step transition probabilities (entry (ij) is $P_{ij}^{(n)}$)

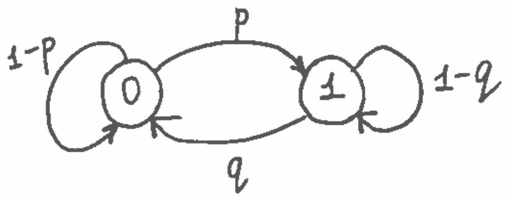
Then, the C.K. equations become $\boxed{\mathbb{P}^{(m+n)} = \mathbb{P}^{(m)} \cdot \mathbb{P}^{(n)}}$

- for $m = n = 1$: $\mathbb{P}^{(2)} = \mathbb{P}^{(1)} \cdot \mathbb{P}^{(1)} = \mathbb{P} \cdot \mathbb{P} = \mathbb{P}^2$
- for $m = 2, n = 1$: $\mathbb{P}^{(3)} = \mathbb{P}^{(2)} \cdot \mathbb{P}^{(1)} = \mathbb{P}^2 \cdot \mathbb{P} = \mathbb{P}^3$

...

$\Rightarrow \mathbb{P}^{(n)} = \underbrace{\mathbb{P}^n}_{\text{matrix multiplication}}$

Ex.



$$\mathbb{P} = \begin{pmatrix} 1-p & p \\ q & 1-q \end{pmatrix}$$

$$\mathbb{P}^2 = \begin{pmatrix} (1-p)^2 + pq & (1-p)p + p(1-q) \\ q(1-p) + q(1-q) & pq + (1-q)^2 \end{pmatrix}$$

Remark: for large n, matrix multiplication can be tedious, so instead we use $\mathbb{P} = R^{-1} D R$

$$D = \begin{pmatrix} 1 & 0 \\ 0 & 1-(p+q) \end{pmatrix}, \quad R = \begin{pmatrix} q & p \\ -1 & 1 \end{pmatrix}, \quad R^{-1} = \frac{1}{q+p} \begin{pmatrix} 1 & -p \\ 1 & q \end{pmatrix}$$

$$\Rightarrow \mathbb{P}^n = (R^{-1} D R)^n = R^{-1} D \underbrace{R R^{-1}}_{I} D \underbrace{R R^{-1}}_{I} D R \dots$$

$$= R^{-1} D^n \cdot R$$

$$D^n = \begin{pmatrix} 1^n & 0 \\ 0 & (1-(p+q))^n \end{pmatrix}$$

$$\Rightarrow \mathbb{P}^{(n)} = \mathbb{P}^n = \frac{1}{q+p}\begin{pmatrix} 1 & -p \\ 1 & q \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & (1-(p+q))^n \end{pmatrix}\begin{pmatrix} q & p \\ -1 & 1 \end{pmatrix}$$

$$-1 < 1-(p+q) < 1$$

$$\Rightarrow |1-(p+q)| < 1$$

$$\Rightarrow (1-(p+q))^n \xrightarrow[n\to\infty]{} 0$$

$$\Rightarrow \text{as } n \to \infty \quad D^n \to \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$$

$$\Rightarrow \lim_{n\to\infty} \mathbb{P}^n = \frac{1}{q+p}\begin{pmatrix} 1 & -p \\ 1 & q \end{pmatrix}\begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}\begin{pmatrix} q & p \\ -1 & 1 \end{pmatrix} =$$

$$= \frac{1}{q+p}\begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} q & p \\ -1 & 1 \end{pmatrix} =$$

$$= \frac{1}{q+p}\begin{pmatrix} q & p \\ q & p \end{pmatrix} = \begin{pmatrix} \frac{q}{q+p} & \frac{p}{p+q} \\ \frac{q}{q+p} & \frac{p}{q+p} \end{pmatrix}$$

$$\Rightarrow \lim_{n\to\infty} \mathbb{P}_{00}^{(n)} = \lim_{n\to\infty} \mathbb{P}_{10}^{(n)} = \frac{q}{q+p}$$

proportion of time you will be ~~lnd~~ in state 0

same, no matter which state you start from

$$\lim_{n\to\infty} \mathbb{P}_{01}^{(n)} = \lim_{n\to\infty} \mathbb{P}_{11}^{(n)} = \frac{p}{p+q}$$

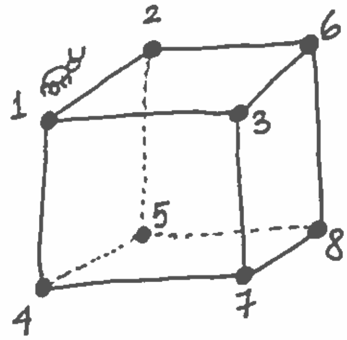
proportion of time you will be in state 1

limiting distribution

Let μ_i = mean time to go from ⓘ to ⑧ , $i = 1, \dots, 8$

want to find μ_1.

$$\left|\begin{aligned}
\mu_1 &= 1 + \tfrac{1}{3}\mu_2 + \tfrac{1}{3}\mu_3 + \tfrac{1}{3}\mu_4 \\
\mu_2 &= 1 + \tfrac{1}{3}\mu_1 + \tfrac{1}{3}\mu_5 + \tfrac{1}{3}\mu_6 \\
\mu_3 &= 1 + \tfrac{1}{3}\mu_1 + \tfrac{1}{3}\mu_6 + \tfrac{1}{3}\mu_7 \\
\mu_4 &= 1 + \tfrac{1}{3}\mu_1 + \tfrac{1}{3}\mu_5 + \tfrac{1}{3}\mu_7 \\
\mu_5 &= 1 + \tfrac{1}{3}\mu_2 + \tfrac{1}{3}\mu_4 + \tfrac{1}{3}\mu_8 \\
\mu_6 &= 1 + \tfrac{1}{3}\mu_2 + \tfrac{1}{3}\mu_3 + \tfrac{1}{3}\mu_8 \\
\mu_7 &= 1 + \tfrac{1}{3}\mu_3 + \tfrac{1}{3}\mu_4 + \tfrac{1}{3}\mu_8 \\
\mu_8 &= 0
\end{aligned}\right.$$



An ant is performing a rando walk on the edge of a cube, where it can select any the three adjoining vertices with equal probability. What is t expected number of s the ant needs till it rea the diagonally opposite ver

By symmetry: $\mu_2 = \mu_3 = \mu_4$

and $\mu_5 = \mu_6 = \mu_7$

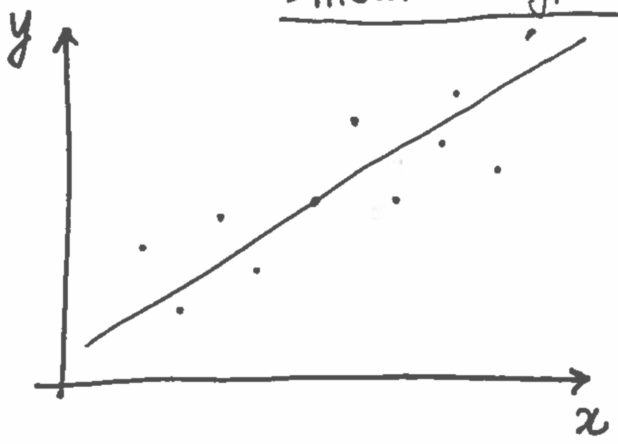
$$\Rightarrow \left|\begin{aligned}
\mu_1 &= 1 + \mu_2 \\
\mu_2 &= 1 + \tfrac{1}{3}(1 + \mu_2) + \tfrac{2}{3}\mu_5 \Rightarrow \tfrac{2}{3}\mu_2 = \tfrac{4}{3} + \tfrac{2}{3}\mu_5 \\
& \Rightarrow \mu_2 = 2 + \mu_5 \\
\mu_5 &= 1 + \tfrac{2}{3}\mu_2
\end{aligned}\right.$$

$$\Rightarrow \mu_5 = 1 + \tfrac{2}{3}(2 + \mu_5) \Rightarrow \tfrac{1}{3}\mu_5 = \tfrac{7}{3}$$

$\Rightarrow \mu_5 = \mu_6 = \mu_7 = 7$; $\mu_2 = \mu_3 = \mu_4 = 9$

$\Rightarrow \boxed{\mu_1 = 10}$

Linear Regression



X, Y - cont. RV

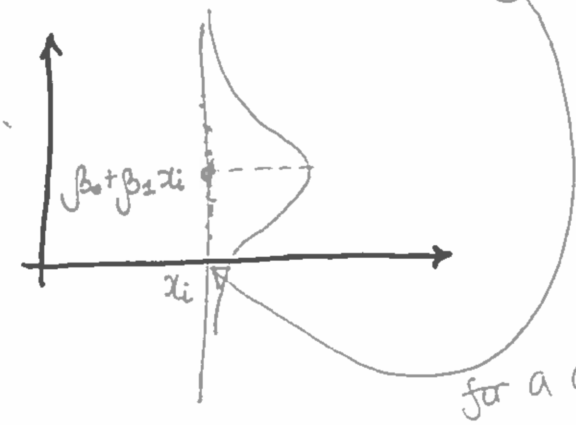
Assume there is a linear relationship between Y and X, Sample $(X_i, Y_i)_{i=1}^{n}$.

$$Y_i = \beta_0 + \beta_1 X_i + \underbrace{\varepsilon_i}_{\text{noise ; error}}$$

where $\varepsilon_i | X \overset{iid}{\sim} N(0, \sigma^2)$

mean 0 — constant variance

$\mathbb{E}[\varepsilon_i | X] = 0$ — $Var(\varepsilon_i | X) = \sigma^2 = c$



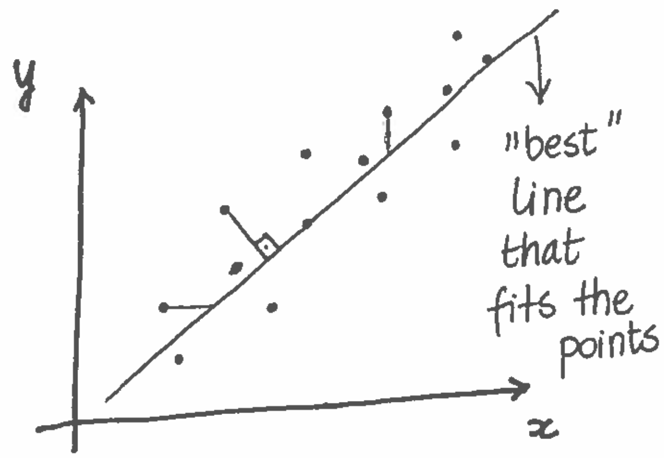
for a given value of $X = x_i$,

$\varepsilon_i \sim N(0, \sigma^2)$

$\Rightarrow (Y_i | X = x_i) = \underbrace{\beta_0 + \beta_1 x_i}_{\text{const}} + \varepsilon_i \sim N$

$\Rightarrow \mathbb{E}[Y_i | X = x_i] = \beta_0 + \beta_1 x_i$

$(Y_i | X = x_i) \sim N(\beta_0 + \beta_1 x_i, \sigma^2)$



X, Y - continuous RV

$(x_i, y_i)_{i=1}^{n}$ pairs of obs.

- used for prediction of Y for a future value of X.

"best"?
- minimize horizontal distances squared
- minimize Euclidean distances sq between points and line
- → minimize vertical distances squared

<u>line</u>: $y = \underbrace{\beta_0 + \beta_1}_{\text{parameters}} x$

$$\min \sum_{i=1}^{n} \left(y_i - (\beta_0 + \beta_1 x_i) \right)^2 = \min_{\beta_0, \beta_1} f(\beta_0, \beta_1)$$

To solve: $\frac{\partial f}{\partial \beta_0} \stackrel{set}{=} 0$ $\qquad \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$

$\frac{\partial f}{\partial \beta_1} \stackrel{set}{=} 0$ $\qquad \hat{\beta}_1 = \frac{\Sigma (x_i - \bar{x})(y_i - \bar{y})}{\Sigma (x_i - \bar{x})^2}$

- interpretation of coefficients

The model is trying to explain the variability in Y, i.e. the line is describing how Y and X are co-varying together.

total variability in Y: $\underbrace{\sum_{i=1}^{n} (y_i - \bar{y})^2}_{SST}$ (how far is y from its mean)^2

variability in Y explained by the regression model: $\underbrace{\sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2}_{SSR}$ (how far from the mean are the fitted values)^2

variability in Y not explained by the model: $SST - SSR = SSE = \sum_{i=1}^{n} (\underbrace{y_i - \hat{y}_i}_{\text{residuals}}$

proportion of variability in Y explained by the model: $\frac{SSR}{SSTO} = R^2$

$0 \leqq R^2 \leqq 1$

Multiple Linear Regression

$$Y_i = \beta_0 + \beta_1 X_{i,1} + \beta_2 X_{i,2} + \dots + \beta_{p-1} X_{i,p-1} + \varepsilon_i$$

$i = 1, \dots, n$

Assume $\varepsilon \mid \mathbb{X} \overset{iid}{\sim} N(0, \sigma^2)$

- matrix notation
- $\hat{\beta} = (\mathbb{X}^T \mathbb{X})^{-1} \mathbb{X}^T \cdot y$

- interpretation of ~~coefficients~~

ANOVA

Source	SS	df	MS	F-ratio
Regression	SSR	$p-1$	$\frac{SSR}{p-1}$	$F = \frac{MSR}{MSE} \sim F_{(p-1, n-p)}$
Error	SSE	$n-p$	$\frac{SSE}{n-p}$	
Total	SST	$n-1$		

$$F = \frac{\frac{\chi^2_R}{p-1}}{\frac{\chi^2_E}{n-p}}$$

- Test overall adequacy of the model

  H_0: $\beta_1 = \beta_2 = \dots = \beta_{p-1} = 0$ (model is not useful in explaining Y)

  H_1: at least one β_j is not zero

  ~~model~~ (at least one of the variables X is useful)

- individual t-tests

- increasing R^2 $\qquad R^2_{adj} = 1 - \frac{n-1}{n-p}(1-R^2)$
- multicolinearity
- Variable selection
  - all possible subsets and use $\overset{\text{some criteria, such as}}{\downarrow}$ R^2_{adj}
  - forward
  - backward

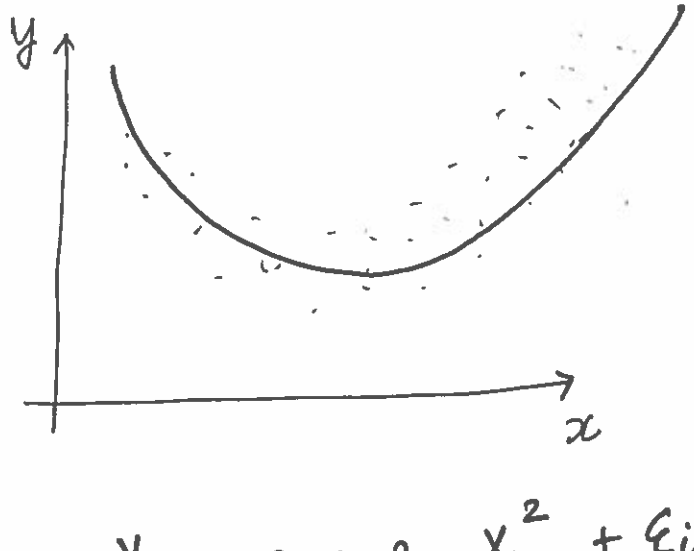
  $\}$ use some criteria such as

  AIC, BIC ...

  $\downarrow$

  $-2\ln(L) + 2p$

Transformations



$$Y_i = \beta_0 + \beta_1 \cdot \underbrace{X_i^2}_{f(X_i)} + \varepsilon_i$$

Y - continuous

X - categorical (k levels)

example: X = weather

{ sunny, cloudy, rainy }

Y = temperature

- indicators ... $Y \sim$ indicators

  - interpretation

ANOVA

$k-1$

$n-k$

$n-1$

F - test

$H_0: \mu_1 = \mu_2 = \cdots = \mu_k$

H_a

$Y \sim$. multiple categorical

- treatments ; ANOVA
- interactions, main effects

A	$k_A - 1$
B	$k_B - 1$
AB	$(k_A - 1)(k_B - 1)$
R	$k_A k_B - 1$
E	$n - k_A k_B$
Total	$n - 1$

Name: ______________________________

MSSP Bootcamp, Summer 2015
Probability Post-Assessment
Friday, August 21, 2015

Justify your answers. Write all of your work on the following blank pages.

Good luck!

Problem 1 Is it true that $\mathbb{P}(A \cup B \cup C) = 1 - \mathbb{P}(A^c | B^c \cap C^c)\mathbb{P}(B^c | C^c)\mathbb{P}(C^c)$? Justify your answer.

$$1 - \mathbb{P}(A^c | B^c \cap C^c) \cdot \mathbb{P}(B^c | C^c) \cdot \mathbb{P}(C^c) =$$

$$= 1 - \frac{\mathbb{P}(A^c \cap B^c \cap C^c)}{\cancel{\mathbb{P}(B^c \cap C^c)}} \cdot \frac{\cancel{\mathbb{P}(B^c \cap C^c)}}{\cancel{\mathbb{P}(C^c)}} \cdot \cancel{\mathbb{P}(C^c)} =$$

$$= \mathbb{P}((A^c \cap B^c \cap C^c)^c) \overset{\text{De Morgan}}{=} \mathbb{P}(A \cup B \cup C).$$

True.

1. 20
2. 20
3. 20
4. 20
5. 20

Total 100

Problem 2 We are given two urns: Urn I contains 5 white and 4 black balls, and Urn II contains 3 white and 4 black balls. An urn is chosen at random, and then two balls are drawn from it without replacement. We define the following events:

A = at least one of the balls is white $= \{wb \text{ or } bw \text{ or } ww\}$

B = at least one of the balls is black $= \{wb \text{ or } bw \text{ or } bb\}$

C = the balls are of different color. $= \{wb \text{ or } bw\}$

a) Are the events A and C independent?

b) Are the events C and $A \cup B$ independent?

c) Compute $\mathbb{P}(A|C)$ and $\mathbb{P}(C|A)$.

d) If two white balls were drawn, what is the probability they were drawn from Urn I?

$\underbrace{5w, 4b}_{I}$ $\quad$ $\underbrace{3w, 4b}_{II}$

a) $\mathbb{P}(A \cap C) \stackrel{?}{=} \mathbb{P}(A) \cdot \mathbb{P}(C)$

$\quad \mathbb{P}(wb \cup bw) = \mathbb{P}(C) \neq \mathbb{P}(A) \cdot \mathbb{P}(C)$, since $\mathbb{P}(A) \neq 1$.

$\Rightarrow$ A and C are not indep.

b) $\mathbb{P}(A \cup B) = \mathbb{P}(\Omega) = 1$

$\mathbb{P}((A \cup B) \cap C) \stackrel{?}{=} \mathbb{P}(A \cup B) \cdot \mathbb{P}(C)$

$\mathbb{P}(\Omega \cap C) = \mathbb{P}(C) = 1 \cdot \mathbb{P}(C)$ ✓ $\Rightarrow$ C and $A \cup B$ are indep.

c) $\mathbb{P}(A|C) = \dfrac{\mathbb{P}(A \cap C)}{\mathbb{P}(C)} = \dfrac{\mathbb{P}(C)}{\mathbb{P}(C)} = 1$

$\frac{5}{18} + \frac{2}{7} = \frac{35+36}{18 \cdot 7} = \frac{71}{2 \cdot 7 \cdot 9}$

$\mathbb{P}(C|A) = \dfrac{\mathbb{P}(A \cap C)}{\mathbb{P}(A)} = \dfrac{\mathbb{P}(C)}{\mathbb{P}(A)} = \dfrac{\frac{1}{2}\left(\frac{5}{9} \cdot \frac{4}{8} \cdot 2 + \frac{3}{7} \cdot \frac{4}{6} \cdot 2\right)}{\frac{1}{2}\left(\frac{5}{9} \cdot \frac{4}{8} \cdot 2 + \frac{5}{9} \cdot \frac{4}{8} + \frac{3}{7} \cdot \frac{4}{6} \cdot 2 + \frac{3}{7} \cdot \frac{2}{6} \cdot\right.}$ $= \dfrac{2 \cdot 71}{3 \cdot 5 \cdot 1}$

d) $\mathbb{P}(I | ww) = \dfrac{\mathbb{P}(ww | I) \cdot \mathbb{P}(I)}{\mathbb{P}(ww | I) \cdot \mathbb{P}(I) + \mathbb{P}(ww | II) \cdot \mathbb{P}(II)} =$

$$= \frac{\frac{5}{9} \cdot \frac{4}{8} \cdot \frac{1}{2}}{\frac{5}{9} \cdot \frac{4}{8} \cdot \frac{1}{2} + \frac{3}{7} \cdot \frac{2}{6} \cdot \frac{1}{2}} = \frac{35}{53}$$

Problem 3 Let $U \sim Unif(0,1)$, and let $X = 3U + 5$. What is the distribution of X?

Circle your answer:

a) $Unif(3,5)$

b) $Unif(3,8)$

c) $Unif(5,8)$

Use the CDF method to derive and identify the distribution of $X = 3U + 5$.

$$\mathbb{F}_X(x) = \mathbb{P}(X \leqq x) = \mathbb{P}(3U + 5 \leqq x) = \mathbb{P}\left(U \leqq \frac{x-5}{3}\right) =$$

$$= \begin{cases} 0, & \frac{x-5}{3} < 0 \\ \frac{x-5}{3}, & \frac{x-5}{3} \in [0,1] \\ 1, & \frac{x-5}{3} > 1 \end{cases} = \begin{cases} 0, & x < 5 \\ \frac{1}{3}x - \frac{5}{3}, & x \in [5,8] \\ 1, & x > 8 \end{cases}$$

$$\Rightarrow X \sim Unif(5,8).$$

Problem 4 Let X_1, X_2, X_3 be independent random variables and $X_1 \sim N(\mu_1, \sigma_1^2)$, $X_2 \sim N(\mu_2, \sigma_2^2)$, $X_3 \sim N(\mu_3, \sigma_3^2)$.

What is the distribution of $X = X_1 + X_2 + X_3$? Circle your answer:

a) $N(\mu_1 + \mu_2 + \mu_3, (\sigma_1 + \sigma_2 + \sigma_3)^2)$

b) $N(\mu_1 + \mu_2 + \mu_3, \sigma_1^2 + \sigma_2^2 + \sigma_3^2)$

Recall the MGF of $Y \sim N(\mu_Y, \sigma_Y^2)$ is given by $M_Y(t) = e^{\mu_Y t + \sigma_Y^2 t^2}$, $t \in \mathbb{R}$.

Use this fact and the multiplication property of the MGF to derive and identify the distribution of $X = X_1 + X_2 + X_3$.

$$M_X(t) = M_{X_1}(t) \cdot M_{X_2}(t) \cdot M_{X_3}(t) =$$

$$= e^{\mu_1 t + \sigma_1^2 t^2} \cdot e^{\mu_2 t + \sigma_2^2 t^2} \cdot e^{\mu_3 t + \sigma_3^2 t^2} =$$

$$= e^{(\mu_1 + \mu_2 + \mu_3)t + (\sigma_1^2 + \sigma_2^2 + \sigma_3^2) \cdot t^2}$$

$$\Rightarrow X \sim N(\mu_1 + \mu_2 + \mu_3, \sigma_1^2 + \sigma_2^2 + \sigma_3^2).$$

Problem 5 Let $X \sim Geom(p)$, $0 < p < 1$. Show that the MGF of

$$M_X(t) = \frac{pe^t}{1-(1-p)e^t},\ t < -ln(1-p)$$

and use the MGF to find the mean of X.

instead ask a problem about conditional expectation

$$\mathbb{P}_X(x) = \sum_{x=1}^{\infty} (1-p)^{x-1} \cdot p$$

$$\Rightarrow M_X(t) = \mathbb{E}[e^{tX}] = \sum_{x=1}^{\infty} e^{tx} \cdot (1-p)^{x-1} \cdot p =$$

$$= \frac{1 \cdot p}{1-p} \cdot \sum_{x=1}^{\infty} (e^t(1-p))^x =$$

$$= \frac{p}{\cancel{(1-p)}} \cdot \frac{e^t \cancel{(1-p)}}{(1-(1-p)e^t)} = \frac{pe^t}{1-(1-p)e^t}$$

$$|\underbrace{e^t}_{>0}\underbrace{(1-p)}_{>0}| < 1 \quad \Rightarrow \quad e^t(1-p) < 1$$

$$e^t < (1-p)^{-1}$$

$$t < -\ln(1-p)$$

$$\mathbb{E}X = M_X'(t)\Big|_{t=0} = \frac{pe^t(1-\cancel{(1-p)e^t}) + \cancel{pe^t(1-p)e^t}}{(1-(1-p)e^t)^2}\Bigg|_{t=0} =$$

$$= \frac{pe^t}{(1-(1-p)e^t)^2}\Bigg|_{t=0} = \frac{p}{(1-(1-p))^2} = \frac{p}{p^2} = \frac{1}{p}.$$

Name: Solutions & Grading Policy

MSSP Bootcamp, Summer 2015
Probability Post-Assessment
Friday, August 21, 2015

Justify your answers. Write all of your work on the following blank pages.

Good luck!

Problem 1 Is it true that $\mathbb{P}(A \cup B \cup C) = 1 - \mathbb{P}(A^c | B^c \cap C^c)\mathbb{P}(B^c | C^c)\mathbb{P}(C^c)$? Justify your answer.

$$1 - \mathbb{P}(A^c | B^c \cap C^c) . \mathbb{P}(B^c | C^c) . \mathbb{P}(C^c) =$$

$$= 1 - \frac{\mathbb{P}(A^c \cap B^c \cap C^c)}{\cancel{\mathbb{P}(B^c \cap C^c)}} \cdot \frac{\cancel{\mathbb{P}(B^c \cap C^c)}}{\cancel{\mathbb{P}(C^c)}} \cdot \cancel{\mathbb{P}(C^c)} =$$

$$= 1 - \mathbb{P}(A^c \cap B^c \cap C^c) = \mathbb{P}((A^c \cap B^c \cap C^c)^c) =$$

De Morgan

$$= \mathbb{P}(A \cup B \cup C). \qquad \text{True.}$$

Problem 2 We are given two urns: Urn I contains 5 white and 4 black balls, and Urn II contains 3 white and 4 black balls. An urn is chosen at random, and then two balls are drawn from it without replacement. We define the following events:

A = at least one of the balls is white
B = at least one of the balls is black
C = the balls are of different color.

a) Are the events A and C independent?

b) Are the events C and $A \cup B$ independent?

c) Compute $\mathbb{P}(A|C)$ and $\mathbb{P}(C|A)$.

d) If two white balls were drawn, what is the probability they were drawn from Urn I?

a) $\mathbb{P}(A \cap C) = \mathbb{P}(wb, bw) = \mathbb{P}(C) \neq \mathbb{P}(A)$. not indep.

b) $\mathbb{P}(A \cup B) = \mathbb{P}(\Omega) = 1$
$\Rightarrow \mathbb{P}((A \cup B) \cap C) = \mathbb{P}(\Omega \cap C) = \mathbb{P}(C) =$
$= \mathbb{P}(C) \cdot 1 = \mathbb{P}(C) \cdot \mathbb{P}(A \cup B) \Rightarrow$ indep.

5w, 4b — I 3w, 4b — II

$A = \{wb, bw, ww\}$

$B = \{bw, wb, bb\}$

$C = \{bw, wb\}$

$$\mathbb{P}(A) = \mathbb{P}(A|I) \cdot \mathbb{P}(I) + \mathbb{P}(A|II) \cdot \mathbb{P}(II) =$$
$$= \frac{1}{2}\left[\frac{5}{9} \cdot \frac{4}{8} \cdot 2 + \frac{5}{9} \cdot \frac{4}{8} + \frac{3}{7} \cdot \frac{4}{6} \cdot 2 + \frac{3}{7}\right]$$
$$= \frac{1}{2}\left[\frac{5 \cdot \cancel{4}}{3 \cdot 2} + \frac{4}{7} + \frac{1}{7}\right] =$$
$$= \frac{1}{2} \cdot \left[\frac{5}{6} + \frac{5}{7}\right] = \frac{5}{2} \cdot \frac{13}{6 \cdot 7}$$

$$\mathbb{P}(C) = \frac{1}{2}\left[\frac{5}{9} \cdot \frac{4}{8} \cdot 2 + \frac{3}{7} \cdot \frac{4}{6} \cdot 2\right] =$$
$$= \frac{5}{2 \cdot 9} + \frac{2}{7} = \frac{35+36}{2 \cdot 7 \cdot 9} = \frac{71}{2 \cdot 7 \cdot 9}$$

c) $$\mathbb{P}(A|C) = \frac{\mathbb{P}(C)}{\mathbb{P}(C)} = 1$$

$$\mathbb{P}(C|A) = \frac{\mathbb{P}(C)}{\mathbb{P}(A)} = \frac{71 \cdot \cancel{2}^{2} \cdot \cancel{6} \cdot \cancel{7}}{\cancel{2} \cdot \cancel{7} \cdot \cancel{9}_{3} \cdot 5 \cdot 13}$$
$$= \frac{2 \cdot 71}{3 \cdot 5 \cdot 13}$$

$$\mathbb{P}(A \cap C) = \mathbb{P}(C)$$

d) $$\mathbb{P}(I|ww) = \frac{\mathbb{P}(ww|I) \cdot \mathbb{P}(I)}{\mathbb{P}(ww|II) \cdot \mathbb{P}(II) + \mathbb{P}(ww|I) \cdot \mathbb{P}(I)} =$$
$$= \frac{\frac{5}{9} \cdot \frac{4}{8} \cdot \frac{1}{2}}{\frac{3}{7} \cdot \frac{2}{6} \cdot \frac{1}{2} + \frac{5}{9} \cdot \frac{4}{8} \cdot \frac{1}{2}} = \frac{\frac{5}{18}}{\frac{1}{7} + \frac{5}{18}} = \cancel{\frac{5/(18+35)}{18/(7 \cdot 18)}} = \cancel{\frac{18}{53}}$$
$$= \frac{5 \cdot 7 \cdot \cancel{18}}{\cancel{18} \cdot 53} = \frac{35}{53}.$$

Problem 3 Let $U \sim Unif(0,1)$, and let $X = 3U + 5$. What is the distribution of X?

Circle your answer:

$X \sim Unif(5,8)$

a) $Unif(3,5)$

b) $Unif(3,8)$

c) $Unif(5,8)$

Use the CDF method to derive and identify the distribution of $X = 3U + 5$.

$$\mathbb{P}(X \leq x) = \mathbb{P}(3U+5 \leq x) = \begin{cases} 0, & \frac{x-5}{3} < 0 \\ \mathbb{P}\left(U \leq \frac{x-5}{3}\right), & \frac{x-5}{3} \in [\\ 1, & \frac{x-5}{3} > 1 \end{cases}$$

$$= \begin{cases} 0, & x < 5 \\ \frac{1}{3}x - \frac{5}{3}, & x \in [5,8] \\ 1, & x > 8 \end{cases}$$

$\Rightarrow X \sim Unif(5,8).$

Problem 4 Let X_1, X_2, X_3 be independent random variables and $X_1 \sim N(\mu_1, \sigma_1^2)$, $X_2 \sim N(\mu_2, \sigma_2^2)$, $X_3 \sim N(\mu_3, \sigma_3^2)$.
What is the distribution of $X = X_1 + X_2 + X_3$? Circle your answer:

a) $N(\mu_1 + \mu_2 + \mu_3, (\sigma_1 + \sigma_2 + \sigma_3)^2)$

b) $N(\mu_1 + \mu_2 + \mu_3, \sigma_1^2 + \sigma_2^2 + \sigma_3^2)$

Recall the MGF of $Y \sim N(\mu_Y, \sigma_Y^2)$ is given by $M_Y(t) = e^{\mu_Y t + \sigma_Y^2 t^2}$, $t \in \mathbb{R}$.
Use this fact and the multiplication property of the MGF to derive and identify the distribution of $X = X_1 + X_2 + X_3$.

$$M_X(t) \overset{\text{indep.}}{=} M_{X_1}(t) . M_{X_2}(t) . M_{X_3}(t) =$$

$$= e^{(\mu_1 t + \sigma_1^2 t^2 + \mu_2 t + \sigma_2^2 t^2 + \mu_3 t + \sigma_3^2 t^2)} =$$

$$= e^{(\mu_1 + \mu_2 + \mu_3).t + (\sigma_1^2 + \sigma_2^2 + \sigma_3^2).t^2}$$

$$\Rightarrow X \sim N(\mu_1 + \mu_2 + \mu_3, \ \sigma_1^2 + \sigma_2^2 + \sigma_3^2)$$

Problem 5 Let $X \sim Geom(p)$, $0 < p < 1$. Show that the MGF of X is given by

$$M_X(t) = \frac{pe^t}{1-(1-p)e^t},\ t < -ln(1-p)$$

and use the MGF to find the mean of X.

$$\cancel{M_X(t) = \mathbb{E}[e^{tX}]} = \sum_{x=0}^{\infty} e^{tx} \cdot (1-p)^{\cancel{x-1}} \cdot p =$$

$$= \cancel{\frac{1}{(1-p)} \cdot p \sum_{x=0}^{\infty} (e^t(1-p))^{x}} =$$

$$= \cancel{\frac{p}{(1-p)} \cdot \frac{1 \cdot e^t}{(1-(1-p)e^t)}}$$

$$M_X(t) = \sum_{x=1}^{\infty} e^{tx} \cdot (1-p)^{x-1} \cdot p =$$

$$= \frac{1}{1-p} \cdot p \sum_{x=1}^{\infty} (e^t(1-p))^x =$$

$$= \frac{p}{\cancel{1-p}} \cdot \frac{e^t\cancel{(1-p)}}{1-e^t(1-p)} = \frac{p \cdot e^t}{1-(1-p)e^t}$$

$$\underbrace{e^t}_{>0}\underbrace{(1-p)}_{>0}$$

$$|e^t(1-p)| < 1$$

$$e^t(1-p) < 1$$

$$e^t < (1-p)^{-}$$

$$t < ln(1$$

$$t < -ln(1$$

$$M_X'(t) = \cancel{\frac{pe^t \cdot (1-(1-p)e^t) - pe^t(-(1-p)e^t)}{(1-(1-p)e^t)^2}}$$

$$= \frac{pe^t}{(1-(1-p)e^t)^2}$$

$$M_X'(t)\Big|_{t=0} = \frac{p}{(1-(1-p))^2} = \frac{p}{p^2} = \frac{1}{p}$$

$$\Rightarrow \mathbb{E}X = \frac{1}{p}.$$

Pre-assessment

1) Let A and B be events such that
$\mathbb{P}(A) = \frac{1}{4}$, $\mathbb{P}(B) = \frac{1}{3}$, $\mathbb{P}(A \cup B) = \frac{1}{2}$.

a) Are events A and B mutually exclusive? Explain;

b) Determine $\mathbb{P}(A \cap B)$.

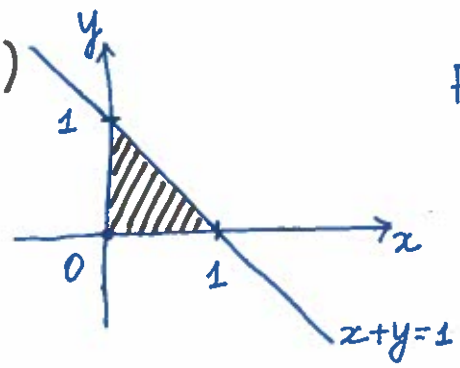
2) How many license plates are there consisting of three digits and two letters if there is no restriction on where the digits and letters are place

3) If $|r| < 1$, determine $\sum_{i=0}^{\infty} a \cdot r^i$.

4) Let $f(x, y) = x^2 + y^3$.

Determine $\frac{\partial f(x,y)}{\partial x}$.

5)



Find the shaded area in two ways:

a) Using double definite integral;

b) Using the formula for the area of a right triangle.

6) Compute $\int_0^{\infty} e^x \cdot 2e^{-2x} \, dx$.

Math Bootcamp for New MSSP Students

Haviland Wright
Aleksandrina Goeva

August 1, 2015

Schedule

Dates	August 17 – 28, M – F
Times	Morning: 9 – 12 || Afternoon: 2 – 5
Materials	On Blackboard

Description

Many graduate programs begin the academic year with a short preparation course in which essential technical topics are reviewed. The idea is to provide a common starting point for students arriving from diverse undergraduate programs. The MSSP Math Bootcamp is in this tradition.

The MSSP Math Bootcamp is intensive – 10 full days of lectures and hands-on mathematics in two weeks. Although some material may be new to you, the bootcamp sessions are designed as a review, so you should recognize most of the topics. Our intention is reinforce what you already know, fill in a few gaps, establish strong working relationships within the new MSSP class, and send you into the first semester of the program with momentum.

The Math Bootcamp is not graded and the only homework that is assigned in advance is preparation for the second week linear algebra sessions+. If, however, you would like to do some reading in advance, we recommend these online open textbooks:

DBR David M Diez, Christopher D Barr, Mine Çetinkaya-Rundel. *OpenIntro Statistics, 3rd edition.* OpenIntro, Inc., 2015. Chapter 2.

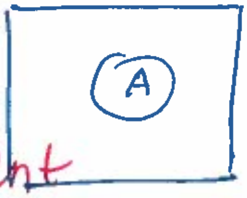
CDW Cherney, David, Tom Denton, and Andrew Waldron. *Linear Algebra.* California: University of California, 2013.

1 Course level Learning Objectives

- **Orientation** Students have had their first educational experience at BU. They have met their classmates and key faculty and staff in the MSSP program. Students are able to assess classroom requirements and organize their work better than before the bootcamp. Students have experience working in groups. They can use blackboard. Students are prepared to hit the ground running.
- **Mathematical Foundations** Students can respond to the mathematical requirements of the MSSP program. They have experience using the basic mathematics they will encounter in the first weeks of the semester. They have practised with statistical problems.
- **Participation Model** Students can work in groups to solve classroom problems and work on a simple project.
- **Computational Experience** Students can access BU computing facilities and have experience using R for simple calculations.

1) Shade the complement of the set A (denoted A^c) on the figure below :

skip



Pre-assessment

1) Let A and B be events such that

$$\mathbb{P}(A) = \frac{1}{4}, \quad \mathbb{P}(B) = \frac{1}{3}, \quad \mathbb{P}(A \cup B) = \frac{1}{2}.$$

a) Are events A and B mutually exclusive? Explain.

b) Determine $\mathbb{P}(A \cap B)$.

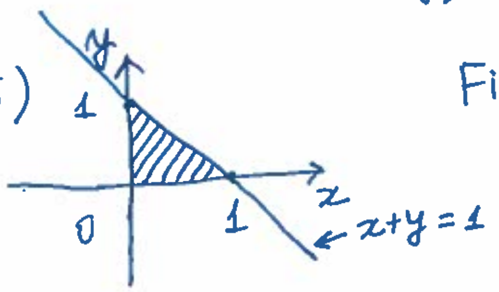
2) How many license plates are there consisting of three digits and two letters if there is no restriction on where the digits and letters are placed?

3) $$\sum_{i=0}^{\infty} a r^i = ?$$

4) Let $f(x,y) = x^2 + y^2$.

$$\frac{\partial f(x,y)}{\partial x} = ?$$

5)



Find the shaded area in two ways:

a) Using a double definite integral.

b) Formula for the area of a right Δ

6) Compute the following integral

$$\int_0^{\infty} e^{x} \cdot 2e^{-2x}\, dx = ?$$

$\frac{\lambda}{t-\lambda}$, $t=1$, $\lambda = 2$

Pre - assessment

② · integration by parts

$$\int_0^\infty x \cdot e^{-x}\, dx = -\int_0^\infty x\, de^{-x} = -x \cdot e^{-x}\Big|_0^\infty + \int_0^\infty e^{-}$$

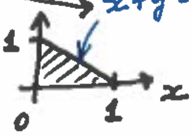
$$= -e^{-x}\Big|_0^\infty = 1$$

③~⑥ · series (geometric)

③ · eq. of a line → $x+y=1$

④ · integration → area

⑤ · check using area of Δ



$$\int_0^1 \int_0^{1-x} dy\, dx = \int_0^1 (1-x)\, dx =$$

$$= \left(x - \frac{1}{2}x^2\right)\Big|_0^1 = 1 - \frac{1}{2} = \frac{1}{2}$$

Area of Δ $= 1 \cdot 1 \cdot \frac{1}{2} = \frac{1}{2}$ ✓

① · sets - complement

- relations

② · combinatorics

- how many license plates
- choose

shade the complement of the event A, denoted A^c.

let A and B be events such that

$\mathbb{P}(A) = 1/4$

$\mathbb{P}(B) = 1/3$

$\mathbb{P}(A \cup B) = 1/2$

a) Are events A and B mutually exclusive? Explain

b) Determine $\mathbb{P}(A \cap B)$.

set difference

p.107 post 3.35 How many license plates are there consisting of three digits and three letters if there is no restriction on where the digits and letters are placed?

All of Statistics - How he does probability 20.07.2015
Ch1-5 (do Ch 1-3)

John Rice - Mathematical statistics book

:
:
expectation
moments (MGF)

- probabilistic modeling
spin of the problems
(connect to statistics)

---

flip your classroom

All of Statistics   Larry Wasserman

Ch 1 Probability

1.1 Introduction

1.2 Sample Spaces and Events

1.3 Probability

1.4 Probability on Finite Sample Spaces

1.5 Independent Events

1.6 Conditional Probability

1.7 Bayes' Theorem

1.8 – ··· – 1.10 Exercises

Ch2 Random Variables

2.1 Introduction

2.2 Distribution Functions and Probability Functions

2.3 Some Important Discrete Random Variables

2.4 Some Important Continuous Random Variables

2.5 Bivariate Distributions

2.6 Marginal Distributions

2.7 Independent Random Variables

2.8 Continuous Distributions

2.9 Multivariate Distributions and iid Samples

2.10 Two Important Multivariate Distributions

2.11 Transformations of Random Variables

2.12 Transformations of Several Random Variables

2.14 Exercises

Ch 3 Expectation

3.1 Expectation of a Random Variable

3.2 Properties of Expectations

3.3 Variance and Covariance

3.4 Expectation and Variance of Important Random Variable

3.5 Conditional Expectation

3.6 Moment Generating Functions

3.8 Exercises

Math Bootcamp for New MSSP Students

Haviland Wright
Aleksandrina Goeva

August 1, 2015

Schedule

Dates	August 17 – 28, M – F
Times	Morning: 9 – 12 || Afternoon: 2 – 5
Materials	On Blackboard

Description

Many graduate programs begin the academic year with a short preparation course in which essential technical topics are reviewed. The idea is to provide a common starting point for students arriving from diverse undergraduate programs. The MSSP Math Bootcamp is in this tradition.

The MSSP Math Bootcamp is intensive – 10 full days of lectures and hands-on mathematics in two weeks. Although some material may be new to you, the bootcamp sessions are designed as a review, so you should recognize most of the topics. Our intention is reinforce what you already know, fill in a few gaps, establish strong working relationships within the new MSSP class, and send you into the first semester of the program with momentum.

The Math Bootcamp is not graded and the only homework that is assigned in advance is preparation for the second week linear algebra sessions+. If, however, you would like to do some reading in advance, we recommend these online open textbooks:

DBR David M Diez, Christopher D Barr, Mine Çetinkaya-Rundel. *OpenIntro Statistics, 3rd edition.* OpenIntro, Inc., 2015. Chapter 2.

CDW Cherney, David, Tom Denton, and Andrew Waldron. *Linear Algebra.* California: University of California, 2013.

1 Course level Learning Objectives

- **Orientation** Students have had their first educational experience at BU. They have met their classmates and key faculty and staff in the MSSP program. Students are able to assess classroom requirements and organize their work better than before the bootcamp. Students have experience working in groups. They can use blackboard. Students are prepared to hit the ground running.
- **Mathematical Foundations** Students can respond to the mathematical requirements of the MSSP program. They have experience using the basic mathematics they will encounter in the first weeks of the semester. They have practised with statistical problems.
- **Participation Model** Students can work in groups to solve classroom problems and work on a simple project.
- **Computational Experience** Students can access BU computing facilities and have experience using R for simple calculations.

Week 1 - Probability

- **Foundation** Students should be able to explain the set theoretic foundations of probability beginning with the concept of a sample space and illustrate their explanation with classic examples – dice, coins, balls in urns. Students should be able to solve simple probability problems with combinatoric methods and demontrate that they have accounted for the entire probability space.
- **Extend the model** Students should be able to continue their explanation of the model to describe conditional probability in set theoretic terms up to and including the Law of Total Probability and Bayes' rule. Students should be able to solve problems and illustrate their solutions with diagrams and algebraic equations.
- **Random Variables** Students should be able to define what a random variable is and explain random variables with respect to sample spaces. Students should be able to illustrate their explanation with standard discrete and continuous distributions – binomial, Poisson, uniform, Normal, chi-square.
- **Cumulative Distribution Function** Students should be able to extend their description of random variables to a definition of the CDF and explain how the CDF of specific random variables can be used to compute probabilities and in the inverse CDF can be used to calculate quantiles Students should be able to show how the CDF is related to the probability mass function of discrete random variables and probability density function (PDF) of continuous random variables. Students should be able to calculate probabilities as sums and integrals of these functions.
- **Bivariate Distributions** Students should be able to extend their description of univariate distributions to bivariate distributions. Student should be able to apply the principles of probability to calculate probabilities for discrete and continuous bivariate distributions. Students should be able to derive marginal and conditional probability distributions from a given bivariate distribution. Students should be able to extend the concept of independent events to distributions and should be able to evaluate whether or not jointly distributed random variables are independent.
- **Transformations of Random Variables** Students should be able to calculate the distributions of transformed random variables under transformations such as arithmetic combinations of two random variables.
- **Expectation** Students should be able to describe the expected value of a random variable and calculate the expectation of discrete and continuous random variables.
- **Variance and Covariance** Students should be able to extend the definition of expected value to describe the calculation of Variance and Covariance. Students should be able to calculate variance and covariance.
- **Moment Generating Functions** Students should be able to define what a MGF is and describe how it is used to calculate moments.

Week 1 Topics

Day 1

Random Experiment
Set Relations
Kolmogorov Axioms of Probability
Law of Partition
Counting Methods
Conditional Probability
Independent Events
Law of Total Probability
Bayes' Rule

Day 2

Random Variable (Def)
Discrete Models (Discrete Uniform, Bernoulli, Binomial, Geometric, Negative Binomial (Pascal), Hypergeometric, Poisson)
Jointly Discrete Random Variables (Joint PMF, Marginal, Conditional Distribution)
Binomial Revisited
Multinomial Distribution
Functions of Discrete Random Variables
Convolution

Day 3

Expected Value (Def, Properties)
Expected Value of a Function of Discrete Random Variables
Variance (Def, Practical Formula, Properties)
Chebyshev's Inequality
Standardized Random Variable
Covariance (Def, Properties, Application, Graphical Representation)
Correlation
Conditional Expectation, Law of Total Expectation
Conditional Variance, Law of Total Variance
CDF (Def, Properties)
Going from Discrete to Continuous Random Variables
PDF (Def, Properties)
Continuous Models (Uniform, Exponential (Connection to Geometric), Gamma and Erlang (Connection to Pascal), Normal and Standard Normal)

Day 4

Functions of Continuous Random Variables (CDF method, PDF method - Univariate Transformation Thm)
Jointly Continuous Random Variables
Marginal and Conditional Densities
Bivariate Transformation Theorem
Convolution, Distribution of the Product of Two Independent Random Variables
Expected Value, Variance, Covariance, Correlation Coefficient
Conditional Expectation, Law of Total Expectation
Conditional Variance, Law of Total Variance
Moments of a Distribution
Moment Generating Function

Day 5

Sum of Random Variables
Law of Large Numbers
Central Limit Theorem
Normal Approximation
Random Number Generation
Simulation of Continuous Random Variables
Statistics: Sampling Distributions, CI for μ (σ known), Estimator for σ, CI for μ (σ unknown)

Week 2

In order to move quickly through the basics and get to the material that will be most useful in your statistical work, you should complete the linear algebra weekend homework assignment before our first session on August 24th. The assignment is described below and can be downloaded from Blackboard. If you have recently taken a linear algebra course, the basic material will be easy for you, and you may want to do just enough problems to show yourself that you're ready for the first session. If you need some practice, take the time over the weekend so you get the most out of the linear algebra sessions in the second week

Most of the linear algebra review will use hand calculations. There will be times, however, when it will be handy to have access to faster computation, and we will use R. The weekend homework includes instructions for installing R and some simple exercises to show you a few basic commands so you can minimize the amount of arithmetic you have to do. The homework includes exercises to get you started with R. You will use R extensively in some of your MSSP coursework. For the math bootcamp, however, we will use R as a highly capable desk calculator.

Weekend Homework

1. Do the "Basic Linear Algebra" problem set that is posted on Blackboard.
2. Download R from cran.r-project. Install R.
3. Do the "R for Linear Algebra" problem set that is posted on Blackboard.

Learning Goals

- **Introduction and Overview** Students should be able to discuss the three key concepts for this weeks review – data structures, statistical models, and geometric visualization. Students should be able to recall at least one example of each of these
- **Vectors** Students should be able to define and describe vectors in terms of direction and magnitude and in terms of points in space. Students should be able to demonstrate vector operations graphically and symbolically. They should be able to correctly identify vector spaces by either proving that the proposed space meets the criteria or providing a counter example demonstrating the the proposed space does not meet the criteria.
- **Inner Products** Students should be able to demonstrate inner product calculations and make drawings to explain what is being calculated.
- **Statistical Applications of Vector Operations** Students should be able to used dot products to calculate sums and sums of squares of sample data. They should be able to explain why sampled data is normed by the degrees of freedom instead of the number of sample points. They should be able to calculate the correlation and demonstrate the connection between the correlation coefficient and the cosine of the angle between two vectors.
- **Matrices** Students should be able to formulate systems of linear equations and vector equations as matrix equations and solve the matrix equations using row operations to transform the matrix into echelon form. Student should be able to discriminate between equations with no solutions, a unique solution, and a family of solutions. Students should be able to add vectors. They should be able to describe the difference between pre-multiplication and post-multiplication. They should be able to explain matrix multiplication in terms of vector operations showing that Ax=b can be understood as linear combinations of the column vectors in A and AB can be understood in terms of dot products. They should be able to transpose matrices and algebraic combinations of matrices. They should be able to explain span in set theoretic and geometric terms. They should be able to do matrix calculations and simplify matrix equations using the matrix algebraic rules.

- **Linear Independence and Rank** Students should be able evaluate whether a proposed set of vectors is independent by solving the homogeneous equation. They should be able to identify basic variables from the pivots in a matrix. They should be able to describe the column space of a matrix in matrix operational and geometric form. They should be able to extend this description to the concept of Dimension and Rank. They should be able to determine the null space of a matrix and describe the relationship between the null space and the column space.

- **Linear Transformations** Students should be able to explain the matrix equation Ax=b as a function in which A acts on x to produce b. They should be able to identify the domain, co-domain, and range of this function. They should be able to describe what a linear transformation is and to illustrate this definition with simple linear functions and non-linear functions. They should be able to apply linear functions to product standard geometric results: shear, contraction, expansion, rotation. They should be able to distinguish between transformations that are into, onto, and one-to-one.

- **Inverse** Students should be able describe the inverse and calculate it by hand for 2x2 matrices. They should be able to list invertibility conditions in terms of row equivalence, pivots, the homogeneous equation, linear independence, span, linear transformations.

- **Applying matrix operations to statistical data** starting data matrices, Students should be able to calculate the SSCP, Covariance, and correlation matrices. they should be able to explain the meaning of the elements of the matrices and produce simple graphics to illustrate their understanding.

- **Derivation of the normal equations for a simple regression** Students should be able to state the simple regression model and derive the normal equations to estimate the parameters by minimizing the sum of squares. They should be able to derive the formula for R-square and build the ANOVA table. They should be able to analyze the result for goodness of fit and describe how the analysis of residuals is related.

- **Regression in Matrix terms** Students are able to express the regression model in matrix terms. They are able to calculated parameter estimates using matrix operations.

- **Orthogonality and the geometry of regression** Students are able to develop the regression model starting with orthogonal projections and a vector equation in which the object vector is not in the span of the columns of the data matrix.

- **Multiple regression** Students can use the normal equations to estimate parameters for multiple regression models.

- **Higher order models** Students can use the design matrix to enter higher order and non-linear values to be used for estimating linear regression coefficients.

- **Markov Chains** Students can construct a Markov chain model and explain how the model works. Students can represent the model symbolically and graphically. Students can translate problem descriptions into a Markov chain models. They can discriminate between problem descriptions that fit the Markov chain model and those that don't. Students can calculate steady state probability vectors for Markov chains. Students can evaluate the structure of the stocastic matrix to assess if the matrix is irreducible. Students can calculate the mean return times for the states of the model

- **Eigenspace and Diagonalization** Students are able to demonstrate the eigenvalue problem illustrated with simple solutions. Students are able to determine if proposed scalars are eigenvalues and calculate the basis for corresponding eigenvectors. Students are able to identify eigenvalues in triangular matrices. They are able to calculate eigen values from 2x2 matrices using the characteristic polynomial. Students are able to diagonalize matrices using the eignvalues and vectors of the matrix.

- **Principal Component Analysis** Students are able to explain and illustrate the principal components model. Students are able to use factor loadings to reduce the dimensionality of a dataset by performing a principal components analysis and evaluating the result. Students are able to use PCA to explore a dataset for outliers.

Topics

Day 1

Basic definitions and operations on vectors
- Real vector spaces and sub-spaces
- Linear combinations of vectors
- Special vectors: null, unit, zero, binary, sign
- Inner product, Vector geometry

Applying vector operations to statistical data
- Sums and sums of squares
- Mean corrected sums of squares
- Degrees of freedom
- Geometry of simple correlation

Basic definitions and operations on matrices
- Systems of linear equations
- Vector equations
- Span, linear independence, dimension, rank
- Matrix inversion
- Coordinate systems, change of basis
- Linear transformations

Applying matrix operations to statistical data
- SSCP, Covariance, Correlation matrices
- Linear models, examples of linear models

Day 2

Simple linear Regression - least squares
Regression in matrix algebra
Analysis of Variance Table
Variances and covariance of regression parameters
Orthogonality
The geometry of Least Squares

Day 3

Multiple regression
Generalized variance
Higher order models
Generalized variance
Markov chains

Day 4

Eigenspace
Principle component analysis

Day 5

Summary & project review

well-defined learning goals – why

assessmen – how

indiv group
written, oral

- didactic
- indiv work
- group

for each unit

formative
sumative
assessment

pretest2

Haviland Wright

07/27/2015

Problem 1

Solve this system of linear equations.

$$\begin{aligned} 4x_1 + 4x_2 + 2x_3 &= 5 \\ 3x_1 + 5x_2 + 3x_3 &= 2 \\ 6x_1 + 6x_2 + 2x_3 &= 4 \end{aligned}$$

Problem 2

Determine which of the following sets of vectors are linearly independent. Justify your answers.

a. $\begin{bmatrix} 10 \\ -6 \\ 2 \end{bmatrix}, \begin{bmatrix} 5 \\ -3 \\ 1 \end{bmatrix}$ **b.** $\begin{bmatrix} 1 \\ -2 \\ -1 \end{bmatrix}, \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 3 \\ -5 \\ -4 \end{bmatrix}$ **c.** $\begin{bmatrix} 1 \\ 1 \\ 0 \\ -1 \end{bmatrix}, \begin{bmatrix} -2 \\ 1 \\ -3 \\ 5 \end{bmatrix}, \begin{bmatrix} 1 \\ -1 \\ 2 \\ -3 \end{bmatrix}, \begin{bmatrix} 0 \\ 9 \\ -4 \\ 4 \end{bmatrix}$

Problem 3

Let $A = \begin{bmatrix} 1 & 1 \\ -2 & -1 \\ -1 & -3 \end{bmatrix}, \mathbf{y} = \begin{bmatrix} 2 \\ -7 \\ 4 \end{bmatrix}$, and define $T : \mathbb{R}^2 \to \mathbb{R}^3$ by $T(x) = Ax$

a. Determine if $\mathbf{y}$ is in the range of T

b. Does T map $\mathbb{R}^2$ onto $\mathbb{R}^3$? Explain.

Problem 4

Let $A = \begin{bmatrix} 1 & 1 \\ 1 & 3 \\ 1 & -1 \end{bmatrix}, \mathbf{b} = \begin{bmatrix} 7 \\ 0 \\ 8 \end{bmatrix}$.Find the least-squares solution of $A\mathbf{x} = \mathbf{b}$.

problem 5

Let $A = \begin{bmatrix} 8 & 7 & 9 & 8 \\ 7 & 4 & 6 & 4 \\ 6 & 3 & 4 & 0 \end{bmatrix}$, $B = \begin{bmatrix} 1 & 2 \\ 1 & 3 \\ 2 & 1 \\ 1 & 1 \end{bmatrix}$, and $C = 3$.

Calculate $AB, CAB, BAC, (AB)^T$, and $(B^T B)^{-1}$.

Problem 6

Find the orthogonal projection of $\begin{bmatrix} 7 \\ 2 \end{bmatrix}$ onto the line through $\begin{bmatrix} 2 \\ -1 \end{bmatrix}$ and the origin.

Problem 7

Find the eigenvalues of and corresponding eigenvectors of $\begin{bmatrix} -1 & 4 \\ 6 & 9 \end{bmatrix}$.

Problem 8

Given that -3 is an eigenvalue of $A = \begin{bmatrix} -1 & 4 \\ 6 & 9 \end{bmatrix}$ and $\begin{bmatrix} 1 \\ 3 \end{bmatrix}$ is an eigenvector of A, diagonalize A and check your answer.

MSSP meeting

17.07.2015

read papers & work out the details

group projects: important characteristics of working in grou
- 30 min / 60 min presentation of the process of group work
- what could fail
- develope a worksheet

1: 1/3 2/3 1 ⇒ self-assessment & peer-assessment

How to structure the projects:

Goals for bootcamp

- skills
- habits

- experience with
- knowing where to go for info
- develope & plan problem solving approach
- follow the plan
- present & explain the solution

backward-faded scaffolding

Ex. **Random Samples** skip

Population of size N

Sample size $n \leqq N$

$A = \{ \text{Mr. Jones is in the sample} \}$

Q: $\mathbb{P}(A) = ?$

The $\underline{\underline{A}}$ depends on how sampling is done.

a) order counts and with replacement

$$\mathbb{P}(A) = 1 - \mathbb{P}(A^c) = 1 - \frac{(N-1)^n}{N^n} = 1 - \left(1 - \frac{1}{N}\right)^n$$

$=|\Omega|$

b) order counts and without replacement

$$\mathbb{P}(A) = 1 - \mathbb{P}(A^c) = 1 - \frac{(N-1)(N-2)\cdots(N-1-(n-1))}{N(N-1)(N-2)\cdots(N-(n-1))} =$$

$$= 1 - \frac{(N-\cancel{1}-n+\cancel{1})}{N} = 1 - \frac{N-n}{N} = \frac{n}{N}$$

c) order does not count, without replacement

$$\mathbb{P}(A) = \frac{\binom{N-1}{n-1}\binom{1}{1}}{\binom{N}{n}} = \frac{\frac{(N-1)!}{(n-1)!\cancel{(N-n)!}} \cdot 1}{\frac{N!}{n!\cancel{(N-n)!}}} = \frac{n}{N}$$

~~pg 101~~
~~pg. 37~~
~~pg. 75~~

[p. 226 poisson

[p. 269
p. 288 joint
p. 320 discrete

[p. 276 , Example 6.9 → multinom
notes

[p. 302, p. 303 Example 6.22 → notes

---

day 3

[p. 320
p. 333
p. 334
p. 350

[p. 362
p. 374
p. 376

p. 389

---

day 4

p. 414
p. 415

p. 426
p. 436
p. 448

p. 449
p. 474
p. 475 ⊕

my text problems

p. 500 p. 51[illegible]
p. 507 p. 52[illegible]

p. 528 , p. 551 , p. 543

---

day 5

p. 605
p. 639

p. 672
p. 673
p. 679

p. 680
p. 673

Week 1 (Aug 17 - Aug 21)

- Day 1 (Aug 17)
  - Assessment test - Basics
    - Sets
    - combinatorics
    - Taylor series
    - Integrals
  - ✓ • Random Experiment
  - ✓ • Set Relations
  - ✓ • Kolmogorov Axioms of Probability
  - ✓ • Law of Partition
  - ✓ • Counting Methods
    - Permutations
    - Combinations
    - ...
  - ✓ • Conditional Probability ; • Independent Events
  - ✓ • Law of Total Probability
  - ✓ • Bayes' Rule

Monday 10 am (July 13)

postassessment

Friday

- **Day 2** (Aug 18)
  - Random Variable (Def)
  - Discrete Models
    - Discrete Uniform
    - Bernoulli
    - Binomial
    - Geometric
    - Negative Binomial (Pascal)
    - Hypergeometric
    - Poisson

reading:
- Bernoulli Random Experiment
- Binomial Random Experiment
- Multinomial Random Experiment

- Jointly Discrete Random Variables
  - Joint PMF
  - Conditional Distr.
  - Marginal Distr.
- Binomial Revisited
- Multinomial Distribution
- Multivariate Hypergeometric
- Functions of Discrete Random Variables
- Convolution

- **Day 3** (Aug 19)
  - Expected Value (Def) + Properties
  - Expected Value of a function of Discrete Random Variables (Def)
  - Variance (Def + Practical Formula) + Properties
  - Chebyshev's Inequality
  - Standardized Random Variable
  - Covariance (Def) + Properties + Application + Graphical Representation
  - Correlation

- Conditional Expectation
- Law of Total Expectation
- Conditional Variance
- Law of Total Variance

Day 4

(Aug 20)

reading:
differentiation
integration

- CDF (Def + Properties)
- Going from Discrete to Continuous RVs
- PDF (Def + Properties)
- Continuous Probability Models
  - Uniform
  - Exponential (connection to Geometric)
  - Gamma (Erlang) (connection to Pascal
  - Normal Distribution
    - Standard Normal Distr.

- - - - - -

- Functions of Continuous Random Variables
  (Univariate Transformation Theorem & CDF method / PDF method)
- Jointly Continuous Random Variables
- Conditional Densities, Marginal Densities
- ~~Multivariate~~ Transformation Theorem (Bivariate)
- Convolution, Distr. of Product of Two indep. RVs
- Expected Value, Variance, Covariance, Correlation Coefficient, Conditional Exp., Law of total Exp. Cond. Variance, Law of Total Var

Day 5
(Aug 21)

- Sum of Random Variables
  - Law of Large Numbers
- CLT – different formulations
- Normal Approximation
  - cont. correction
  - Slides

day 4

- Moments of a distribution (Def)
- MGF (Def) (+ characterizes the distr. completely)
  - How to get the moments
  - Multiplication property
  - MGF of a linear transformation $Y = a + bX$
- ~~Random Number Generators~~
- Simulation of Continuous Random Variables
  - Random Number Generation
  - $F_X^{-1}(U)$ has same distr. as X
  - Example (Exponential).
  - Generate a standard Normal RV
    $S_{12} = X_1 + \dots + X_{12}$, $X_i \overset{iid}{\sim} \text{Unif}(0,1)$
    $\Rightarrow S_{12} - 6 \approx Z$.

Additional Topics:

day 5

- Statistics: Sampling Distributions (N, χ^2, t, F)
- Markov Chain problems

- CI for μ
  - (σ^2 known)
- estimarot for σ^2
- CI for μ
  (σ^2 unknown)
-

Math Bootcamp for New MSSP Students
Sessions: Aug 17 - 28 Materials available: July 15

Sessions Schedule:	Sessions will be Monday through Friday for the two weeks, beginning on August 17. During the middle weekend, there will be a group project and optional tutorial sessions.
Daily schedule:	Morning Session: 09 - 12 -- Classroom instruction Afternoon Session: 14 - 17 -- problems solved in groups
Nightly homework:	2 hours -- submitted on **Blackboard** **Daily quiz** (to be completed before the morning session)
Materials:	On Blackboard, **available after July 15**.
Assessment:	First day / Last day self-assessment and tests The math bootcamp is not graded.

Description:

Short summer prep courses enjoy a long tradition in a wide variety of graduate programs. Most cover technical topics with the intention of providing a common starting point for students arriving from diverse undergraduate programs. The MSSP Math Bootcamp is in this tradition.

Over the last two weeks of August, we will spend six hours a day reviewing a range of statistics foundation topics to set the stage for the MSSP program. Many students will recognize much of the material covered in these sessions. But, it may have been years since you actually worked problems in these areas. Other topics may be new to you. These sessions give us the opportunity to build a common foundation that would simply be assumed once the regular term begins or included by referring you to standard textbooks. Problem sets we do in the next two weeks is intended to reinforce what you know, fill in some gaps, establish working relationships in your MSSP class, and send you into the first semester of the program with momentum.

These sessions are not graded, but they begin and end with an objective test and a self-assessment. The MSSP is a professional program, and this is your first lesson to prepare you for professional life. Your success as a professional data scientist will depend on your ability to assess your strengths and weaknesses so that you can trade on your strengths and work on your weaknesses.

The reading and problem sets assigned as homework will set the stage for the work we do the next day. There will be a short quiz based on the homework that you should complete before coming to class to make sure that you understand the essentials of the material we will discuss. If you have trouble with the quiz, go back to the homework and try the quiz again. Once you are comfortable with the material and your quiz results, turn in your homework on blackboard. Plan to spend an hour or two on homework everyday during the sessions.

Daily sessions will begin with a review of the homework and then in the morning session build on that foundation to address more technical points. In the afternoon session, we will work on problem sets, often in groups. At the end of the day, solutions to the problems will be presented and discussed.

Once the fall term begins, you will make extensive use of computer applications for data analysis and report preparation. In these sessions, however, we will use the computer simply as a means to reduce tedious calculations.

Topics by day:

Week 1

Day 1 — Introductions / Orientation /
Assessment test

Sets, simple probability through Bayes
[This is intended as a sort of warm-up session.

Homework:
    Counting out the event space: Combinatorics
    Binomial coefficients, Multinomial coefficients
    Occupancy problems

Quiz:
    dice, committees, letters, unmarked urns, numbered urns,
    binary strings, flags

Day 2 — Random variables, distributions,
Use homework to motivate Bernoulli, binomial, geometric poisson, exponential, negative binomial

Day 3 — Sums of Random variables
change of variables, moment generating functions, convolution integrals, Fourier transformation, Taylor series

Day 4

Day 5 — Simple simulation

Weekend Assignment   Linear algebra problem set
R vector and matrix functions

Week 2

Homework 1:

The reading for this assignment and the problem set review basic concepts and operations in linear algebra. If you have recently taken a course in which you used linear algebra, the material will be easy for you and you may want to do just enough problems to show yourself that you're ready for the rest of the review. If you have not used linear algebra in a while or never took a linear algebra course, invest the time to carefully read the material, do the example problems in the text, and do the entire problem set. Make sure that you have done the quiz on Blackboard. I will use the quiz scores as guidance in the review.

linear systems, vector equations - existence and uniqueness of solutions
span, kernel, linear independence
Linear systems

just enough R --

Quiz 1

Day 1
Linear models, consistency
transformations
coordinate systems, Change of basis,

Homework 2

polynomials
determinants
dimension, rank,

Quiz 2
statistical applications
sums of squares
eigen values

Day 2
Sum of squares
least squares

Homework 3

Quiz 3

Day 3	Orthogonality Least Squares QR factorization Linear models

Homework 4

Quiz 4

Day 4	Eigen Values Markov chains

Homework 5 -- review 2 weeks

Day 5	Review Assessment test Test review