

MATH 320: Probability

Completed Notes

Colton Gearhart

November 29, 2023

Contents

Test 1	2
Lecture 0 – Course Overview	2
Lecture 1 – Set Theory	5
Lecture 2 – Counting	12
Lecture 3 – Probability	21
Lecture 4 – Conditional Probability	30
Lecture 5 – Independent Events	40
Lecture 6 – Bayes’ Theorem	47
Test 2	53
Lecture 7 – Random Variables	53
Lecture 8 – Distribution Functions	57
Lecture 9 – Summary Measures	71
Test 3	92
Lectures 10 – Discrete Distributions	92
Lectures 11 – Continuous Distributions	127
Lectures 12 – Moment Generating Functions	161
After Test 3	174
Lectures 13 – Functions of Random Variables	174

Test 1

Contents

Lecture 0 – Course Overview	2
Lecture 1 – Set Theory	5
Lecture 2 – Counting	12
Lecture 3 – Probability	21
Lecture 4 – Conditional Probability	30
Lecture 5 – Independent Events	40
Lecture 6 – Bayes’ Theorem	47

Lecture 0 – Course Overview

MATH 320: Probability

Lecture 0: Course Overview

Big picture

Relationship between Probability and Statistics

- In a probability problem, the properties of the population are assumed known, and we use these to infer properties of the sample.

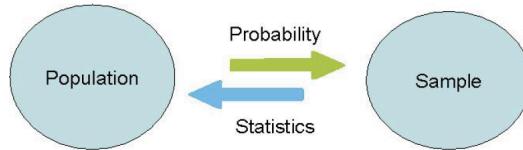


Figure: The reverse actions of Probability and Statistics

$$\underline{p \text{ knows } p} \quad + \quad \underline{s \text{ knows } s}$$

- Whereas statistics is concerned with learning (inferring) population properties from sample information (which is the opposite of probability).
- Example:
 - Suppose we know 75% of batteries last longer than 1500 hours. → **Population info**
We want to know the chance all 10 batteries in a pack last 1500 hours. → **sample info**
 - Suppose that out of 10 batteries, 6 are found to last more than 1500 hours. → **Sample info**
We want to know if that is enough evidence to conclude that the proportion of all batteries that last more than 1500 hours is less than 75%. → **Population info**
- In spite of this difference, statistical inference itself would not be possible without probability because it is based on probability calculations.

Courses you will take

- MATH 320 Probability (Fall)
 - This deals with the building blocks of inferential statistics, which is probability.
 - (1) – Calculating probabilities: Set theory, counting, probabilities of events, random variables
 - (2) – Univariate distributions: Random variables / distributions, probabilities of events, summaries of random variables, applications
 - (3) – Multivariate distributions: Repeat above, now with more than one random variable.
- MATH 321 Mathematical Statistics (Spring)
 - This deals with theory and practice of statistics → We have data, now what do we do with it?
 - Descriptive statistics: Summarizing a whole data set with a single or a few measures (e.g. mean, standard deviation, minimum) and visualizing datasets (e.g. histograms, boxplot).
 - Inferential statistics: Collecting data, analyzing it, and making inferences on parameters using probability concepts.

Exam P syllabus**1. Topic: General Probability (23-30%)****Learning Objectives**

The Candidate will understand basic probability concepts, combinatorics, and discrete mathematics.

2. Topic: Univariate Random Variables (44-50%)**Learning Objectives**

The Candidate will understand key concepts concerning discrete and continuous univariate random variables (including binomial, negative binomial, geometric, hypergeometric, Poisson, uniform, exponential, gamma, normal, lognormal, and beta) and their applications.

3. Topic: Multivariate Random Variables (23-30%)**Learning Objectives**

The Candidate will understand key concepts concerning multivariate discrete random variables, the distribution of order statistics, and linear combinations of independent random variables, along with associated applications.

Lecture 1 – Set Theory

MATH 320: Probability

Lecture 1: Set Theory

Chapter 1: Probability (1.1)

Where does data come from

Definitions

- An **experiment** is the process by which an observation/outcome is made, which cannot be predicted with certainty (outcomes are random).
- An **outcome** of an experiment is any possible observation of that experiment (often called sample points).

Examples: Write the set of all possible outcomes for the following experiments.

1. Sampling students and computing the average number of study hours each day:

$$\Omega = [0, 24] \rightarrow \text{Continuous, uncountable}$$

2. Roll die, record number that appears:

$$\Omega = \{1, 2, 3, 4, 5, 6\} \rightarrow \text{Discrete, countable}$$

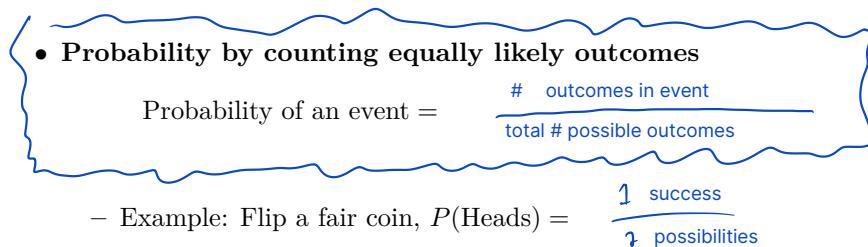
3. Roll die, record first role that a one appears:

$$\Omega = \{1, 2, 3, \dots\} \rightarrow \text{Countable, infinite}$$

What is probability and how to calculate it

Intuitively, probability is the likelihood of something occurring.

One approach (simplest case)



- Events are not always equally likely, so cannot determine probabilities by counting. But there is a simple way to estimate that probability.

Another approach

- **Empirical probability** (based on collected data).

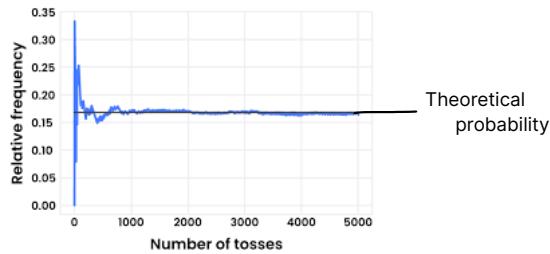
- Relative frequency estimate of the probability of an event

Probability of an event =

$$\frac{\# \text{ times event occurs}}{\# \text{ of trials}}$$

- These are two ways of looking at probability.

- If outcomes are equally likely, relative frequency \approx counting for a very large number of trials (e.g. if we simulated rolling a fair die 10,000 times, $\frac{\# \text{ } 6\text{s}}{10,000} \approx \frac{1}{6}$).



Third approach

- **Subjective probability** is asking a well-informed person for his/her personal estimate of the probability of an event (relying on experience and personal recollections of relative frequencies in the past).
- The rest of this chapter will be building in more precise mathematical framework for probability.
- Counting will play a big role, but keep in mind that in practice many probability numbers actually used in calculations may come from relative frequencies or subjective estimates.

Set Theory

The mathematical basis of probability is sets. Set theory is useful here to provide a precise language for dealing with the outcomes in a probability experiment.

Definitions

- A **set** is a collection of objects (such as the numbers 1, 2, 3, 4, 5, 6).

- The objects are called **elements** of the set. Outcomes of an experiment correspond to elements in a set.

- Writing sets: **Notation is important!** 

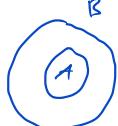
We use capital letters to denote sets such as A, B, C.

Can list the elements in braces if only a few, or use set-builder notation for large or infinite sets.

- Example: All positive numbers can be written as

"A = the set of all x's such that (condition) $x > 0$ "

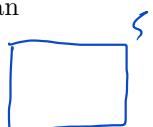
$$A = \{x : x > 0\}$$



- **Subset** $A \subset B$ means that every element in A is also an element of B.

- The **sample space** S (aka outcome space) is the set of all outcomes of an experiment.

- There are different types of sample spaces.
- Countable or uncountable and if numeric: discrete or continuous.

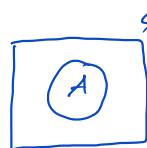


- An **event** is a collection of possible outcomes of an experiment, that is, any subset of S.

1. Examples: Roll die, record the number that appears.

- Define notation: $A = \text{roll even #}$

- Show event: $A = \{2, 4, 6\} \subset S = \{1, 2, 3, 4, 5, 6\}$



An event occurs when at least one element in the event has occurred.

2. Is S an event? yes! $S \subset S$

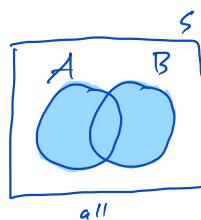
- The **null (empty) set** is the set containing no elements. $\emptyset = \{\}$

Basic operations (algebra of sets)

- **Union:** $A \cup B = \text{"A or B"}$

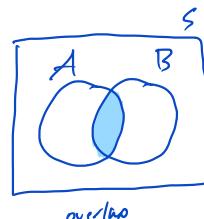
The set of elements that belong to either A or B or both

Compound events



- **Intersection:** $A \cap B = \text{"A and B"}$

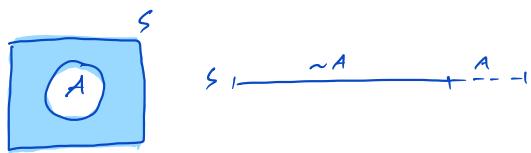
The set of elements that belong to both A and B



- **Complement:** $\sim A$ (or A' , A^c , \bar{A}) = "complement of A", "Not A"

The set of elements (in the sample space) that are _____

So _____ all elements of A from the original sample space S .



- Examples: Let $S = \{1, 2, 3, 4, 5, 6\}$ and $A_1 = \{1\}$, $A_2 = \{2, 3, 4\}$, $A_3 = \{4, 5, 6\}$. Find each of the following events:

$$A_1 \cup A_2 = \{1, 2, 3, 4\}$$

$$A_2 \cup A_3 = \{2, 3, 4, 5, 6\}$$

$$A_1 \cap A_2 = \{3\} = \emptyset$$

$$A_2 \cap A_3 = \{4\}$$

$$\sim A_3 = \{1, 2, 3, 5, 6\}$$

$$\sim(A_1 \cup A_2) = \sim\{1, 2, 3, 4\} = \{5, 6\}$$

Set identities

- These laws help simplify (rewrite) events that are stated verbally or in set notation when solving problems.

- **Commutative Law:** Reordering

Order doesn't matter

$$A \cup B = B \cup A$$

$$a + b = b + a$$

$$A \cap B = B \cap A$$

- **Associative Law:**

Changing location of parentheses

Order of operations with () doesn't matter

$$(A \cup B) \cup C = A \cup (B \cup C)$$

$$(a + b) + c = a + (b + c)$$

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

- **Distributive Law:** Distribute set operation

Distribute to items in ()

$$A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$$

$$a(b+c) = ab + ac$$

$$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$$

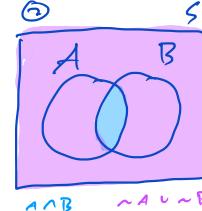
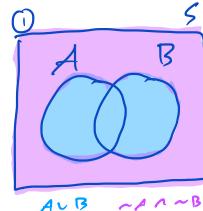


- **De Morgan's Law:** Distributing complement (flip everything)

$$\textcircled{1} \quad \sim(A \cup B) = \sim A \cap \sim B$$

$$\textcircled{2} \quad \sim(A \cap B) = \sim A \cup \sim B$$

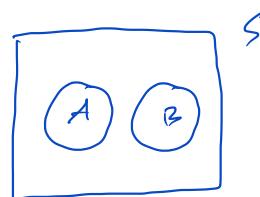
Relationships among sets



- Definitions

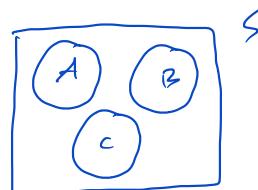
– Two events are **mutually exclusive** (or **disjoint**) if _____.

No overlap $A \cap B = \emptyset$



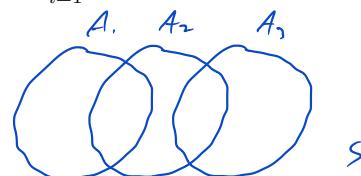
– The events A_1, A_2, \dots are **pairwise mutually exclusive** if _____ every pair is mutually exclusive.

Formally, the condition is stated as: if $A_i \cap A_j = \emptyset$ for all $i \neq j$.



– Events A_1, \dots, A_k are **exhaustive** if when combined, they form _____ entire sample space _____.

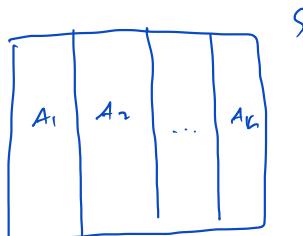
Formally, this is written as: $\bigcup_{i=1}^k A_i = A_1 \cup \dots \cup A_k = S$



– Events A_1, \dots, A_k form a **partition** of S if they are exhaustive _____ and pairwise _____ mutually exclusive.

complete S

no overlap



- Examples: Given the sample space and events below, classify the relationship between events.

$$S = \{1, 2, 3, 4, 5, 6\} \quad \text{and} \quad A_1 = \{1\}, A_2 = \{2, 3, 4\}, A_3 = \{4, 5, 6\}, A_4 = \{5, 6\}.$$

a. A_1 and A_2 are _____

b. A_2 and A_3 are _____

c. A_1, A_2, A_3 are _____

but are _____

d. A_1, A_2, A_4 are _____

$$\text{To check: 1. exhaustive} \rightarrow A_1 \cup A_2 \cup A_4 = \{1\} \cup \{2, 3, 4\} \cup \{5, 6\} = S$$

$$\text{2. pairwise mutually exclusive} \rightarrow A_1 \cap A_2 = A_1 \cap A_4 = A_2 \cap A_4 = \emptyset$$

Lecture 2 – Counting

MATH 320: Probability

Lecture 2: Counting

Chapter 1: Probability (1.2)

Counting basics

Motivation

- There are ways to count the number of outcomes in certain types of random experiments. Thus, we need to develop some counting principles.

This is useful in finding probabilities of events associated with these random experiments.

- Example: Suppose we have a shuffled deck and we deal seven cards. What is the probability that we draw no queens?

$$P(\text{no Queens}) = \frac{\# \text{ successes}}{\# \text{ possibilities}} = \frac{\# \text{ hands with no queens}}{\# \text{ possible hands}}$$

[in Context](#)

Simple counting examples

1. Suppose our class 100 students. 78 students are mathematical science majors and 50 students are actuarial science majors. 41 students are double majors in mathematical science and actuarial science.

- (a) How many students are not mathematical science majors?

$$100 - 78 = 22$$

- (b) How many students major in mathematical sciences or actuarial sciences?

$$78 + 50 - 41 = 87$$

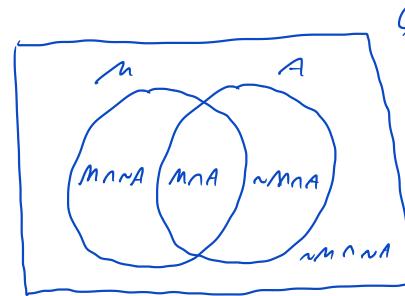
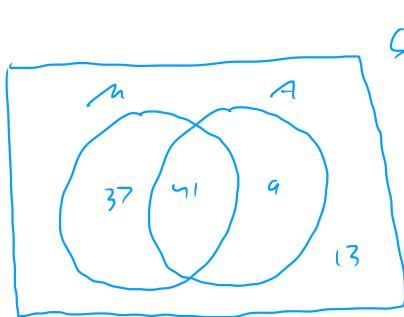
doubles

2. A single card is drawn from a well-shuffled deck. How many cards are hearts or clubs?

$$13 + 13 - 0 = 26$$

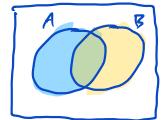
Venn diagrams

- **Venn Diagrams** are helpful for visualizing all of the components of a counting problem and can easily extended to three events.



Basic rules

- **Notation:** The number of elements in the event (set) $A = n(A)$
- **Complements counting rule:** For any finite sample space S and event A
 $\rightarrow n(\sim A) = n(S) - n(A)$
- **General union counting rule:**
 For any two events A and B in any finite sample space $\rightarrow n(A \cup B) = n(A) + n(B) - n(A \cap B)$
- **Special case union counting rule:** If A and B are mutually exclusive $\rightarrow n(A \cup B) = n(A) + n(B)$ $n(A \cap B) = 0$



Counting outcomes of an experiment

- **Tree diagrams** give a simple graphical display of all possible cases (pairs of outcomes) in problem/experiments if the number of outcomes is not unreasonably large.

When drawing tree diagrams, think about the stages of the experiment.

- Example: Suppose we are testing for the presence of a disease. There are two things to consider, if the person has the disease (which is unknown) and the result of the test, positive or negative. Let's define:

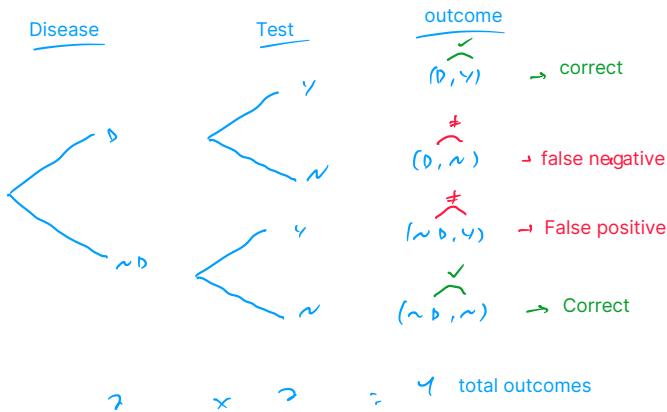
D = the person tested has the disease

$\sim D$ = the person tested does not have the disease

Y = the test is positive

N = the test is negative

Find how many outcomes are possible and what each of them are.



- When experiments get larger, we can use the following idea.

Multiplication principle for counting:

If a job consists of k separate tasks, the i th of which can be done in n_i ways ($i = 1, \dots, k$), then the entire job can be done in $n_1 \times n_2 \times \dots \times n_k$ ways.

Task 1	Task 2	\dots	Task k	Total outcomes
n_1	n_2	\dots	n_k	$n_1 \times n_2 \times \dots \times n_k$

- Example: Sally has 6 pairs of socks, 4 shorts, 5 shirts, and 3 sunglasses. How many ways can she get dressed?

$$\begin{array}{ccccccc} & \overset{6}{\cancel{x}} & & \overset{4}{\cancel{x}} & & \overset{5}{\cancel{x}} & \\ \text{TASK: } & \underset{1}{\cancel{1}} & & \underset{2}{\cancel{2}} & & \underset{3}{\cancel{3}} & \\ & & & & & & \end{array} \quad \approx 360 \quad \text{total ways}$$

- ★ • It is very important to correctly define the sub experiments, then can just use the rule. Each "task / sub experiment" is like a level in our tree diagram.

This is also an important principle because we can use it to develop some more counting techniques.

Permutations, combinations and partitions

Counting number of ways

- After defining tasks in an experiment, often we need to count the number of possible ways to perform each task. In doing so, there are four important criteria to consider:

1. The number of distinguishable items $\rightarrow n$
2. The number of objects we are going to select $\rightarrow r$
3. Order matters or not?
4. With replacement or without?

- Possible methods of counting

	Ordered	Unordered
With replacement	n^r	
Without replacement	$p(n)$	(n)

Ordered, with replacement

- Example: How many four-letter words can the letters A through Z produce?

$$\begin{array}{l} n = 26 \\ r = 4 \end{array} \quad \underline{26} \times \underline{26} \times \underline{26} \times \underline{26} = 26^4$$

- Ordered, with replacement:** Given n distinguishable objects, there are \underline{n} ways to choose with replacement an ordered sample of r objects.



- Strategy: When doing counting problems, think about (and actually draw) the "slots". This will help with what numbers to use AND to determine if order matters.

This illustrates an application of the multiplication principle where each "slot" is a separate task.

ALWAYS
* across

Ordered, without replacement

- Example: How many ways can Bob, Mary and Jane sit in three seats?

This question is really asking how many Permutations of these three are there?

$$\text{Seat } \frac{3}{1} \times \frac{2}{2} \times \frac{1}{3} = 6 = 3!$$

- Ordered without replacement (all n):** The number of permutations of n distinct objects is $\frac{n! = n(n-1)(n-2)\dots(2)(1)}{\text{"n-factorial"}}$. $\rightarrow 0! = 1$

- Example: What is the number of four-letter code words selecting from the 26 letters of the alphabet without replacement?

$$\text{Position } \frac{26}{1} \times \frac{25}{2} \times \frac{24}{3} \times \frac{23}{4} = P(26, 4)$$

- Ordered without replacement ($r \leq n$):** The number of r -permutations of n distinct objects (aka **permutation** of n objects taken r at a time) is $P(r, n)$.

$$\left. \begin{aligned} P(r, n) &= n(n-1)\dots(n-r+1) \\ &\downarrow = \frac{n!}{(n-r)!} \end{aligned} \right\} \begin{aligned} P(r, n) &= \frac{n!}{(n-r)!} = n! \\ P(0, n) &= \frac{n!}{(n-0)!} = 1 \end{aligned}$$

$$\text{Example: } P(3, 10) = \frac{10!}{7!} = \frac{10 \cdot 9 \cdot 8 \cdot 7!}{7!} = 10 \times 9 \times 8 = 720$$

Unordered, without replacement

- Two scenarios: Among 8 students, (a) selecting 1st, 2nd and 3rd place winners (b) selecting 3 committee members among 8 students. What is the difference?

order matters in a + not in b

- **Unordered without replacement ($r \leq n$):** The number of combinations of n objects taken r at a time is $\frac{(n)_r}{r!}$.

$$(n)_r = \frac{n!}{(n-r)! r!} \quad \text{"choose } r\text{"}$$

A **combination** is an unordered group (more formally, an r -element subset of the original n distinct objects), and $\binom{n}{r}$ counts the total number of different groups possible.

- Useful property: $\binom{n}{r} = \binom{n}{n-r}$

If a group of r is made, then a group of $n - r$ is made and vice versa.

Relationship between combinations and permutations

- Both of these can be thought of as two sub experiments involving the other and demonstrates how **order** impacts the counting tool.

Permutation \rightarrow Combination

1. ordered selection of r from n
2. Remove order from groups
 \Rightarrow Divide out duplicates

Example: Committee of 3 from 7

$$\begin{array}{c} A \\ \underline{B} \\ \underline{C} \\ \vdots \\ \vdots \end{array} \quad \left\{ \begin{array}{l} 6 \text{ times } = 3! \\ \text{ALL same group} \end{array} \right.$$

$$\rightarrow \frac{P(n)}{r!} \times \frac{1}{r!} = \frac{n!}{(n-r)!} \times \frac{1}{r!} = \binom{n}{r}$$

$$\rightarrow \binom{n}{r} = \frac{P(n)}{r!}$$

Combination \rightarrow Permutation

1. unordered selection of r from n
2. order all r
 \Rightarrow multiply by possible arrangements

Example: Rank 3 from 7

$$\begin{array}{c} A \\ \underline{A} \\ \underline{B} \\ \vdots \\ \vdots \end{array} \quad \left\{ \begin{array}{l} \text{each can happen} \\ 6 = 3! \text{ ways} \end{array} \right.$$

$$\rightarrow \binom{n}{r} \times P(r) = \frac{n!}{(n-r)!} \times r! = P(r)$$

$$\rightarrow P(r) = \binom{n}{r} r!$$

Examples

1. Determining ordered vs unordered.

Find the number of ways to do each of the following.

- (a) Rank your favorite 4 desserts from the menu of 10 items.

$$\text{order } \checkmark \Rightarrow P(10, 4)$$

- (b) Select which 3 side dishes to serve out of the 15 from your cookbook.

$$\text{order } X \Rightarrow \binom{15}{3}$$

- (c) Determine the jobs ~~for 3~~^{out of 8} members at the dinner party: set the table, serve the food, do the dishes.

$$\begin{array}{l} \xrightarrow{\text{meaning to slots}} \text{order } \checkmark \\ \xrightarrow{\text{table}} \quad \quad \quad \xrightarrow{\text{serve}} \quad \quad \quad \xrightarrow{\text{dishes}} \\ \Rightarrow P(8, 3) \end{array}$$

- In some problems involving ordering, the ordering is not obvious or implied, but rather implicit (like when making an assignment list).

AA BEST way to think about it: If the "slots" have meaning, then order matters.

2. Combined problems

A company has 20 male employees and 30 female employees. They are forming a committee that will have two male members and three female members. In how many ways can this committee be chosen?

$$\begin{array}{ccc} \text{order } X & & \\ \xrightarrow{\text{Choose: 2 from 20}} & \times & \xrightarrow{\text{Choose: 3 from 30}} \\ \binom{20}{2} & & \binom{30}{3} \\ \text{males} & & \text{females} \end{array}$$

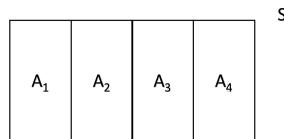
- Many counting problems include combined the use of the multiplication principle, permutations and combinations.

So just separate a scenario into tasks, count each task individually and then multiply each tasks total ways to get the total overall number of ways.

More than two groups

- **Partitioning** refers to the process of breaking a large group into separate smaller groups.

Will learn how to count number of ways to divide all available objects:



The combination problems previously discussed are simple examples of partitioning problems.

- Example: Flip 5 coins. How many observation sequences are there in which there are two heads and three tails?



- The basic idea of a combination divides n distinct objects into two groups: a group of chosen objects and a group of unchosen objects.

$$\text{This is why } \binom{n}{r} = \binom{n}{n-r} \text{ is called a binomial coefficient.}$$

- This can be extended to more than two groups.

Example: There are 10 students. How many ways can we make three groups with sizes $\frac{1}{3}$, 3 and $\frac{1}{3}$.

$$\text{choose: } \underbrace{\binom{10}{3}}_{\text{3 from 10}} \times \underbrace{\binom{7}{3}}_{\text{3 from 7 left}} = \frac{10!}{3!3!} \times \frac{7!}{3!4!} = \frac{10!}{3!3!4!} \rightarrow \text{group sizes}$$

- **Counting partitions:** The number of partitions of n objects into k distinct groups of sizes n_1, n_2, \dots, n_k (where $n_1 + \dots + n_k = n \iff$ splitting up entire group) is given by:

$$\binom{n}{n_1, n_2, \dots, n_k} = \frac{n!}{n_1! n_2! \cdots n_k!}$$

This is called the **multinomial coefficient**.

OR partition of Positions \rightarrow

$$\text{MISSISSIPPI} = \frac{\{1\}}{M} \quad \frac{\{2, 5, 8, 11\}}{I} \quad \frac{\{3, 4, 6, 7\}}{S} \quad \frac{\{9, 10\}}{P}$$

Counting $\rightarrow S = \{1, 2, \dots, 11\}$

$$\text{MISSISSIPPI} = \frac{\{1\}}{M} \quad \frac{\{1, 4, 7, 10\}}{I} \quad \frac{\{2, 3, 5, 6, 8, 9\}}{S} \quad \frac{\{2, 3\}}{P}$$

2-8

- Example: Find the number of ways to rearrange the letters in the word MISSISSIPPI.

$$\frac{\text{total } 11}{\text{dups}} = \frac{11!}{1!4!4!2!} = \binom{11}{1, 4, 4, 2}$$

Ls
 $\begin{array}{l} M \rightarrow 1 \\ I \rightarrow 4 \\ S \rightarrow 4 \\ P \rightarrow 2 \\ \hline 11 \end{array}$

- Counting partitions can also be thought of as the number of ways to arrange n objects where n_1 objects are alike, n_2 objects are alike and so forth.

To account for the repetitions when counting distinct permutations (arrangements), we need to divide out the duplicates.

 It is the groups that matter, not the order within the groups.

Summary: When to use which counting tool (formula)

• Ordered

- With replacement: $\underline{n^r}$ ex) how many 6 digit passwords can the digits 0-9 make?
- Without replacement: $\underline{P(r)}$ ex) 7 possible vacation destinations, rank your top 3.

• Unordered

- Without replacement: $\underline{\binom{n}{r}}$ ex) select 5 out of 12 players to start the basketball game.
- More than two groups: $\underline{(n_1, \dots, n_k)}$ ex) make smaller teams of size 2, 3, and 3 from 8 players.

- With replacement: This exists, but we won't worry about it.

Lecture 3 – Probability

MATH 320: Probability

Lecture 3: Probability

Chapter 1: Probability (1.1)

Probability by counting equally likely outcomes

- Now we can update our original definition of probability using the counting concepts.

- Definition: Let A be an event from a sample space in which all outcomes are equally likely. The **probability of A** , denoted $P(A)$, is defined by:

$$\text{Probability of an event} = \frac{\text{Number of outcomes in the event}}{\text{Total number of possible outcomes}}$$

$$P(A) = \frac{n(A)}{n(S)}$$

- Examples: A standard 52 card deck is shuffled and one card is picked at random. Find the probabilities of the following events:

(a) Club and King: $P(C \cap K) = \frac{1}{52}$

(b) Club or King: $P(C \cup K) = \frac{13 + 4 - 1}{52} = \frac{16}{52}$

(c) Not Club nor King: $P[\sim(C \cup K)] = \frac{52 - 16}{52} = \frac{36}{52}$

(d) Club or Hearts: $P(C \cup H) = \frac{13 + 13}{52} = \frac{26}{52}$

- We will formalize these probability ideas soon.

More counting probability problems

- Now we can use all of the counting tools we've learned and our new probability knowledge to look at more interesting problems.



- COUNTING STRATEGY:** Solve the numerator and denominator separately.
 - Numerator: "IS a condition" (selecting from restricted sample space).
 - Denominator: "NO condition" (selecting from unrestricted sample space).

$$\text{prob} = \frac{\# \text{ successes}}{\# \text{ possibilities}}$$

- Examples:

- A box of jerseys for a pick-up game of basketball contains 8 extra-large jerseys, 7 large jerseys, and 5 medium jerseys. If you are first to the box and grab 3 jerseys, what is the probability that you randomly grab 3 extra-large jerseys.

Counting

(1) numerator

- only selecting from XL jerseys
- 3 from 8

order $X \rightarrow nlr$

$$P(3XL) = \frac{\binom{8}{3}}{\binom{20}{3}} \rightarrow \text{only } n \text{ changes} \rightarrow r \text{ is the same}$$

$$\downarrow = \frac{14}{1285}$$

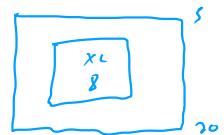
(2) Denominator

- select from ALL jerseys
- 3 from 20

Direct way

$$\text{Selection} \quad \frac{P}{1} \times \frac{7}{2} \times \frac{6}{3} = \frac{14}{1285}$$

↳ w/o rep \Rightarrow drop # each time



- Suppose we have a shuffled deck and deal seven cards. What is the probability that we draw no queens?

- Counting
- select 7 non-queens (48)
 - Select 7 from ALL (52)

$$P(\text{No queens}) = \frac{\binom{48}{7}}{\binom{52}{7}} \approx 0.55$$

Direct

$$\frac{48}{52} \times \frac{47}{51} \times \dots \times \frac{42}{46} \approx 0.55$$

Not queen \longrightarrow



- Suppose we have a shuffled deck and deal three cards. What is the probability that we draw exactly one queen?

- Counting
- Select 1 Q from 4 AND 2 NQ from 48
 - Select 3 from ALL (52)

$$P(1Q, 2NQ) = \frac{\binom{4}{1} \binom{48}{2}}{\binom{52}{3}} \rightarrow \frac{4}{12} = \frac{1}{3}$$

$$\approx 0.204$$

3 sequences
(3)

Direct

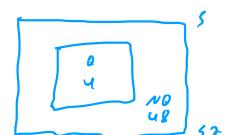
$$\frac{4}{52} \times \frac{48}{51} \times \frac{47}{50} \approx 0.068 \quad \text{Counting}$$

All same prob

$$\begin{array}{ccc} \overline{Q} & \overline{Q} & \overline{Q} \\ \overline{NQ} & \overline{NQ} & \overline{NQ} \\ \overline{NQ} & \overline{NQ} & \overline{Q} \end{array} =$$

seq * P(seq)

$$\Rightarrow 3 * 0.068 \approx 0.204$$



- Remember in order to use counting tools when finding probabilities, all outcomes need to be equally likely.

This means just the smallest possible results of the experiment (e.g. drawing a single card), not events (e.g. king or heart).

Generalizing probability

Motivation

- In real data studies, outcomes are rarely equally likely. So we need methods to work with probabilities in these scenarios as well.
- Example: A research study into the percentage of births which involve more than one child leads to the following probability table:

Number of children	1	2	3
Probability	0.9760	0.0311	0.0019

(1)

Intuitively we can easily find the probability of an event, such as a randomly selected birth involving more than one child, based on this table.

$$\begin{aligned}
 P(\text{children} > 1) &= P(2 \cup 3) \\
 &= P(2) + P(3) \\
 &= 0.0311 + 0.0019 = 0.033
 \end{aligned}$$

Let's break down what we implicitly did.

- Assigned probabilities to each of the individual outcomes in the sample space (i.e. the table).
- Wrote the event of interest in terms of the outcomes of interest.
- Added probabilities of the mutually exclusive outcomes.

- This previous example illustrates a natural method for assigning probabilities to events of certain types of experiments.

Sample point method for calculated probabilities

- Theorem: Let $S = \{O_1, \dots, O_n\}$ be a finite set, where all O_i are individual outcomes each with probability $P(O_i) \geq 0$ and $\sum P(O_i) = 1$. For any $A \in S$,

$$P(A) = \sum_{O_i \in A} P(O_i)$$

- Using this theorem, we have steps / techniques to find probability of any event.

- Example: When players A and B play tennis, the probability that A wins is $2/3$. Suppose that A and B play two matches. What is the probability that A wins at least one match?

Let AB denote the outcome that player A wins the first game and player B wins the second.

- What are the sample space and individual outcomes?

$$\Omega = \{AA, AB, BA, BB\}$$

- Find the probability of each individual outcomes.

$$P(AA) = \frac{2}{3} \cdot \frac{2}{3} = \frac{4}{9}$$

$$P(AB) = \frac{2}{3} \cdot \frac{1}{3} = \frac{2}{9}$$

$$P(BA) = \frac{1}{3} \cdot \frac{2}{3} = \frac{2}{9}$$

$$P(BB) = \frac{1}{3} \cdot \frac{1}{3} = \frac{1}{9}$$

- Find the event of interest and express it as a union of individual outcomes.

$$W = AA \cup AB \cup BA \rightarrow P(W) = P(AA) + P(AB) + P(BA)$$

$$\downarrow = \{AA, AB, BA\} \quad \downarrow = \frac{2}{9}$$

$\star \star$ OR in probability \iff Addition

General definition of probability

- Not all sample spaces are finite or easy to handle. So there are axioms that give general properties that an assignment of probabilities to events must have.

- Axioms of Probability:** If you define a way to assign a probability $P(A)$ to any event A , the following axioms must be true:

$$1. P(A) \geq 0 \text{ for any event } A \rightarrow \text{nonnegative probabilities}$$

$$2. P(S) = 1 \rightarrow \text{Total probability} = 1$$

$$3. \text{Suppose } A_1, \dots, A_n \text{ is a (possibly infinite) sequence of pairwise mutually exclusive events. Then}$$

$$P(A_1 \cup \dots \cup A_n) = P(A_1) + \dots + P(A_n) \rightarrow \begin{aligned} &\text{sample point method} \\ &= \text{add probabilities} \\ &\text{if mutually exclusive} \end{aligned}$$

- Using this new definition, we have two important properties:

$$1. \text{Any event can be expressed as a union of mutually exclusive outcomes.}$$

$$2. \text{The probability of an event is the sum of probabilities of the mutually exclusive outcomes.}$$

Theorems and their proofs

- Theorem: For probability assignment $P(\cdot)$ and any event A in the sample space S ,

$$(a) P(\emptyset) = 0$$

$$(b) P(A) \leq 1$$

$$(c) P(\sim A) = 1 - P(A)$$

\star Complement probability

- Proofs (easier to prove out of order):

(c) $P(\sim A) = 1 - P(A)$

Strategy for proofs: Rewrite events we know and use axioms to simplify their probabilities.



$$\begin{aligned} \rightarrow S &= A \cup \sim A \\ \rightarrow P(S) &= P(A \cup \sim A) \\ 1 &= P(A) + P(\sim A) \\ P(\sim A) &\stackrel{\checkmark}{=} 1 - P(A) \end{aligned}$$

axioms
 $\xleftarrow{2} P(S) = 1 \quad \xrightarrow{3} A \cap \sim A = \emptyset \Rightarrow ME$

(b) $P(A) \leq 1$

HINT: Use part (c) of theorem

$$\begin{aligned} \rightarrow P(\sim A) &= 1 - P(A) \quad , \text{theorem (c)} \\ P(A) &= 1 - P(\sim A) \quad \geq 0 \quad \text{by axiom 1} \\ \Rightarrow P(A) &\stackrel{\checkmark}{\leq} 1 \end{aligned}$$

Another way
 $\rightarrow 0 \leq P(\sim A) \quad \text{some event} \quad , \text{axiom 1} \quad , \text{theorem (c)}$
 $= 1 - P(A) \quad , \text{theorem (c)}$
 $P(A) \stackrel{\checkmark}{\leq} 1$

Combined with Axiom 1, this theorem gives us bounds on the probability of any event A:

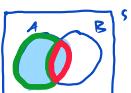
$$\boxed{0 \leq P(A) \leq 1}$$

(a) $P(\emptyset) = 0$

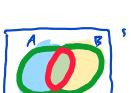
HINT: Use part (c) of theorem

$$\begin{aligned} \rightarrow P(\emptyset) &= 1 - P(\sim \emptyset) \quad , \text{theorem (c)} \\ &= 1 - P(S) \quad , \sim \emptyset = S \\ &= 1 - 1 \quad , \text{axiom 2} \\ &\stackrel{\checkmark}{=} 0 \end{aligned}$$

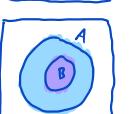
- Theorem: For probability assignment $P(\cdot)$ and any events A and B in the sample space S ,



(a) $P(A \cap \sim B) = P(A) - P(A \cap B)$



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



(c) ~~If $B \subset A$, then $P(B) \leq P(A)$~~

If $B \subset A$, then $P(B) \leq P(A)$

- Proofs:

(a) $P(A \cap \sim B) = P(A) - P(A \cap B)$

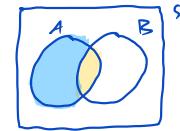
HINT: Start with A and use identity $A = (A \cap B) \cup (A \cap \sim B)$

$$\hookrightarrow A = (A \cap B) \cup (A \cap \sim B)$$

$$\hookrightarrow P(A) = P[(A \cap B) \cup (A \cap \sim B)]$$

$$\downarrow = P(A \cap B) + P(A \cap \sim B)$$

$$P(A \cap \sim B) \stackrel{?}{=} P(A) - P(A \cap B)$$



axiom 3
 $(A \cap B) + (A \cap \sim B)$ are ME

>

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

HINT: Use identity $(B \cup A) = (B \cup A) \cap (\underbrace{B \cup \sim B})$

$\hookrightarrow S$

$$\hookrightarrow B \cup A = (B \cup A) \cap (B \cup \sim B)$$

$$\downarrow = B \cup (A \cap \sim B)$$

by distributive law >

$$\hookrightarrow P(B \cup A) = P[B \cup (A \cap \sim B)]$$

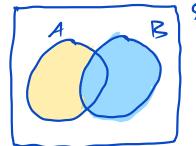
$$\downarrow = P(B) + P(A \cap \sim B)$$

axiom 3, $B + (A \cap \sim B)$ are ME

$$\downarrow = P(B) + [P(A) - P(A \cap B)]$$

ct theorem (a) >

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



(c) If $B \subset A$, then $P(B) \leq P(A)$

HINT: Start with Axiom 1 and $P(A \cap \sim B)$

$$\hookrightarrow 0 \leq P(A \cap \sim B)$$

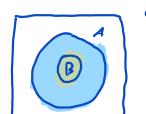
axiom 1 >

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

ct theorem (<) >

$$\downarrow = P(A) - P(B)$$

$B \subset A$ >



More examples and concepts / theorems

1. A fair coin is flipped successively until the same face is observed on successive flips. Find the probability that it will take three or more flips of the coin to observe the same face on two consecutive flips.

$$\rightarrow S = \{\overbrace{(\textcircled{1}, [3, 4, \dots])}^A\}$$

$$\rightarrow A = \{3, 4, \dots\}$$

\star Important strategy

$$\rightarrow P(A) = 1 - P(\sim A)$$

$$\downarrow = 1 - \frac{1}{2} \quad \text{U} \quad \text{2 flips } P(\text{same}) = \frac{2}{4}$$

sample point method

$$\left\{ \text{HH, HT, TH, TT} \right\}$$

each $P(\cdot) = \frac{1}{4}$

2. Find the probability that in a room of 20 people, there are at least two people sharing the same birthday.

$$P(\text{At least 2 shared}) = 1 - P(\text{None shared})$$

$$= 1 - \frac{P(365)}{365^{20}} \quad \text{U} = \frac{0}{\textcircled{1}}$$

success = $\frac{365}{1} \times \frac{364}{2} \times \dots \times \frac{346}{20}$

All different \Rightarrow w/o rep

total = $\frac{365}{1} \times \frac{364}{2} \times \dots \times \frac{346}{20}$

w/ rep

$$\approx 1 - 0.584$$

$$\approx 0.416$$

- Note on order: When solving these types of problems, the numerator and the denominator must ALWAYS MATCH (be consistent) in terms of order.

So should never have ~~nPr~~
 ~~nCr~~

3. A cryptocurrency exchange sells Bitcoin, Litecoin and Ethereum.

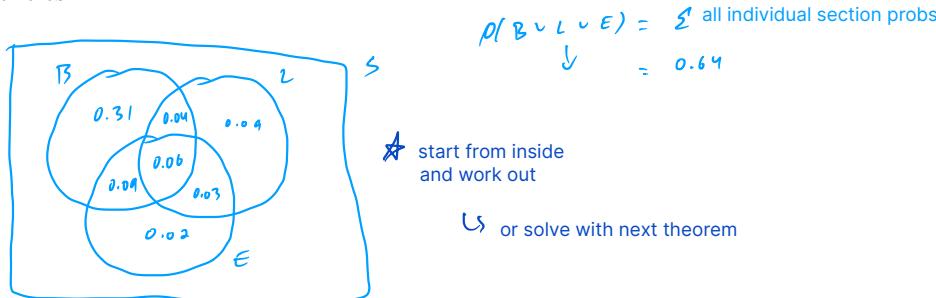
Let $B = \{\text{buys Bitcoin}\}$, $L = \{\text{buys Litecoin}\}$ and $E = \{\text{buys Ethereum}\}$.

Based on past sales the exchange determines that for any new customer

$$P(B) = 0.50, P(L) = 0.22, P(E) = 0.20, P(B \cap L) = 0.10,$$

$$P(B \cap E) = 0.15, P(L \cap E) = 0.09, \text{ and } P(B \cap L \cap E) = 0.06.$$

Find the probability that a new customer purchases at least one of these three currencies.



- Theorem: For any events A , B , and C ,

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

The analogous counting rule for the union of three events is:

$$n(A \cup B \cup C) = n(A) + n(B) + n(C) - n(A \cap B) - n(A \cap C) - n(B \cap C) + n(A \cap B \cap C)$$

Quick review

- Everything we are studying begins from experiments. From these, we obtain outcomes, which together make up the sample space.
- We want to know probabilities of events (subsets of S). To calculate this, we defined how to assign probabilities in general via the axioms.
- Then there are two methods to compute probabilities:
 - sample point method $\rightarrow P(\text{Event}) = \frac{\text{# of outcomes in event}}{\text{# of outcomes}}$
 - Counting tools \rightarrow only when all outcomes are equally likely

Odds

- Odds of an event A are generally written as a ratio of two integers, such as $2:3$ "2 to 3". The odds against A are given by the reverse ratio $3:2$.

Formally, odds are another way to represent probability (but the terms are not interchangeable).

- Definition: The **odds** for an event A are defined as the ratio $P(A)$ to $P(\sim A)$.

$$\text{Odds of } A = \frac{P(A)}{P(\sim A)} = \frac{P(A)}{1 - P(A)} = \frac{P(\text{Event occurring})}{P(\text{Event not occurring})} \quad \left\{ \begin{array}{l} \text{odds: for} \\ \frac{P(w)}{P(l)} \end{array} \right. \quad \left. \begin{array}{l} \text{Against} \\ \frac{P(l)}{P(w)} \end{array} \right.$$

- Converting between odds and probability: $a:b \leftrightarrow P(A) = \frac{a}{a+b}$

Example: Suppose a soccer team is in a playoff game.

- (a) If the probability winning is 0.20, what are the odds of winning?

$$0.2 = \frac{1}{5} \rightarrow \frac{1 \text{ part "w" in}}{4 \text{ "parts" lose}} \Rightarrow 1:4$$

- (b) If the odds of losing are 7:9, find the probability of winning.

$$P(w) = \frac{w}{l+w} = \frac{9}{16}$$

Lecture 4 – Conditional Probability

MATH 320: Probability

Lecture 4: Conditional Probability

Chapter 1: Probability (1.3)

In some probability problems a condition is given which restricts your attention to a subset of the sample space.

Conditional probability by counting

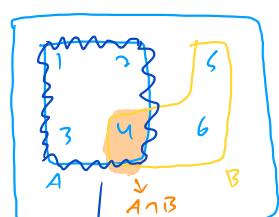
Motivating example

- Example: Roll a die. $A = \{1, 2, 3, 4\}$ and $B = \{4, 5, 6\}$.

We have the following probabilities: $P(A) = \frac{4}{6}$ $P(B) = \frac{3}{6}$

If we know that A already occurred, what is the probability of B ? $P(B|A) = \frac{1}{4}$

- This is known as a conditional probability.



Updated sample spaces new sample space

- The sample space is the set of all possible outcomes.

In other words, only the outcomes in the sample space can be observed.

- When an event occurs, we don't know specifically which outcome was observed.

- 'A occurred' means we know that the outcome is one of the elements in A. That is, there is no chance to get outcomes not in A.

So these numbers are excluded from the sample space.

- Thus, an event which already occurred restricts the sample space and it becomes an updated sample space.

Examples with contingency tables

- Example: Here is attendance and college major of students:

	Statistics	Art	Chemistry	Total
Perfect	100	40	80	220
Good	20	50	70	140
Poor	30	15	30	75
Total	150	105	180	435

- Suppose we are told the selected student has good attendance, what is the probability the student is a chemistry major?
 extra info
 main event

Said equivalently: What is the probability the student is a chemistry major *given that the student has good attendance?*

$$P(\text{Chem} \mid \text{Good}) = \frac{\frac{70}{140}}{n(S)} = \frac{n(\text{Chem} \cap \text{Good})}{n(\text{Good})} = \frac{p(\text{Chem} \cap \text{Good})}{p(\text{Good})}$$

Joint prob
Marginal prob

Another way we could interpret this result is: Of ~~all students~~ ~~good attenders~~ ~~are chemistry majors.~~ ~~50%~~

- Thus the restricted sample space still applies to contingency tables, where now we look only within a single row or column because we have additional (GIVEN) information.
 And we see that these conditional probability problems could also be solved using probabilities.

- Types of probabilities:

- Marginal probabilities: Refer to one event; use column / row totals to find these.
- Joint probabilities: Refer to two (or more) events; use numbers in the middle of the table to find these.

	1	2	3	Total
A	Joint freq			Marginal freq
B				
Total	Marginal freq		Total Total	

- Examples) $P(\text{Stats}) = \frac{150}{435}$



Denominator is the overall total

$P(\text{Perfect} \cap \text{Stats}) = \frac{100}{435}$



Defining conditional probability

- The previous examples showed two natural ways of finding conditional probability. The first was based on counting and the second on probabilities.

- Conditional probability by counting ~~not~~ equally likely outcomes

$$P(A | B) = \frac{n(A \cap B)}{n(B)}, \quad \text{where } n(B) > 0$$

When outcomes are not equally likely, this rule does not apply. Then we need a general definition of conditional probability.

- Definition: The **conditional probability** of an event A given the occurrence of the event B is:

$$P(A | B) = \frac{P(A \cap B)}{P(B)}, \quad \text{provided } P(B) > 0 \quad \left\{ \begin{array}{l} P(A|B) = \frac{\text{Joint prob } A \cap B}{\text{marginal prob of } B} \end{array} \right.$$

- The notation $P(A | B)$ is read "the conditional probability of A given B ".

A is the main event of interest and B is called the conditioning event.

Examples

- 100 cars will be painted in a production line. Of these cars 25 will be painted blue, 75 will be painted red, 12 will get a clear coat, and 9 of the blue cars will get a clear coat at random.

	C	$\sim C$	
B	9	16	25
R	3	72	75
	12	88	100

- What is the probability that a car will be painted blue?

$$P(B) = \frac{25}{100}$$

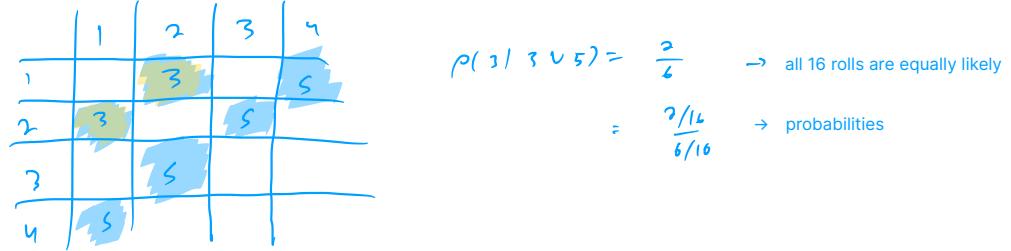
- Given that a car is blue, what is the probability that it got a clear coat?

$$P(C|B) = \frac{9}{25}$$

- What is the probability that the car is red, given that it did not get a clear coat?

$$P(R | \sim C) = \frac{72}{88}$$

2. A pair of fair four sided dice is rolled and the sum is determined. Find the probability that a sum of 3 is rolled given that a sum of 3 or 5 is rolled.



3. Probabilities for the number of auto insurance claims are given in the table below.

Number of claims	0	1	2	3
Probability	0.72	0.22	0.05	0.01

Find the probability that a policyholder files exactly 2 claims, given that the policyholder has filed at least one claim. $\rightarrow C = \geq 1 \text{ claim}$

$$\begin{aligned} P(2|C) &= \frac{P(2 \cap C)}{P(C)} \quad \rightarrow P(C) = \sum \text{probs for } 1, 2, 3 \\ &= \frac{P(2)}{P(C)} \quad \downarrow = 0.28 \\ &= \frac{0.05}{0.28} \quad \circ [1 \boxed{2} 3] \\ &\quad \circ C = \{1, 2, 3\} \Rightarrow 2 \cap C = 2 \end{aligned}$$

This tells us ≈ 0.179 of policyholders who file a claim file exactly 2 claims.

- Note on conditional probability:

$$\begin{aligned} - \text{ ALWAYS } P(B | A) &= \frac{P(A \cap B)}{P(A)} \\ - \text{ ONLY SOMETIMES } P(B | A) &= \frac{P(B)}{P(A)} \quad \rightarrow \text{can simplify numerator} \\ &\quad \text{ONLY IF } B \subset A \end{aligned}$$

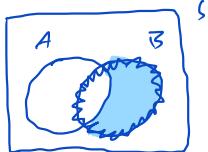


Applying conditional probability

Probability rules for conditional probability

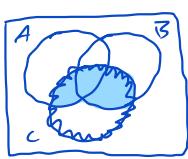
- All of the rules (axioms) for a probability assignment from “Lecture 3 – Probability” apply to a conditional probability function $P(\cdot | B)$ as well.
- In addition, the theorems we proved also hold true for $P(\cdot | B)$.
- Examples and (a few) new theorems:

– We know $P(\sim A) = 1 - P(A)$, using the previous example we can find $P(\sim A | C)$.



$$P(\sim A | C) = \frac{P(C | C)}{P(C)} = \frac{1 - P(A | C)}{1} = 1 - 0.179 = 0.821$$

In general: $P(\sim A | B) = 1 - P(A | B)$



– We know $P(A \cup B) = P(A) + P(B) - P(A \cap B)$. Now applying this to the conditional probability of $A \cup B$ given C ,

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

Multiplication rule for probability

- The definition of conditional probability can be rewritten as a multiplication rule for probabilities, where we are trying to get an expression for the joint probability of A and B .

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

- **Multiplication rule for probability:** Given events A and B ,

$$P(A \cap B) = P(A) P(B|A), \quad P(A) > 0$$

$$P(A \cap B) = P(B) P(A|B), \quad P(B) > 0$$

- Interpretation: $\underbrace{P(A \cap B)}_{\text{Both events occurred}} = \underbrace{P(A) * P(B|A)}_{\text{A occurred, then B occurred later}}$

AND in probability \Rightarrow multiplication

- Example: At a country fair game there are 25 balloons on a board of which 10 balloons are yellow, 8 are red and 7 are green. A player throws darts at balloons to win a prize and randomly hits one of them. If a player throws two darts in a row what is the probability that both balloons hit are yellow?

$$\text{Direct} \rightarrow P(YY) = \frac{\frac{10}{25}}{P(Y_1)} \times \frac{\frac{9}{24}}{P(Y_2|Y_1)}$$

$$\text{Counting} \rightarrow P(YY) = \frac{\binom{10}{2}}{\binom{25}{2}} = \frac{\frac{10 \cdot 9}{2! \cdot 2!}}{\frac{25 \cdot 24 \cdot 23 \cdot 21}{24!}} = \frac{\frac{10 \cdot 9 \cdot 23 \cdot 21}{2! \cdot 2!}}{\frac{25 \cdot 24 \cdot 23 \cdot 21}{24!}} = \frac{10 \cdot 9}{25 \cdot 24}$$

$\hookrightarrow \text{ncr} = \text{w/o rep}$

The “direct way” of solving counting probability problems is just an application of the multiplication rule for probability.

- Often in a probability experiment, it can be easier to assign $P(A)$ and $P(B | A)$ rather than $P(A \cap B)$. Then we can easily compute $P(A \cap B)$ using these.

Example: Two cards are drawn at random from a standard deck without replacement, as in the previous example. Find the probability that both are kings.

$$P(B_1) * P(K_2 | B_1) = \frac{4}{52} * \frac{3}{51} = P(K_1 \cap K_2)$$

- This multiplication rule can be extended to any number of events as well.

– More generally, given k events A_1, \dots, A_k

$$P(A_1 \cap \dots \cap A_k) = P(A_1) * P(A_2 | A_1) * \dots * P(A_k | A_1 \cap \dots \cap A_{k-1})$$

– Example: For $k = 3$ events, the multiplication rule is

$$P(A_1 \cap A_2 \cap A_3) = P(A_1) * P(A_2 | A_1) * P(A_3 | A_1 \cap A_2)$$

- Example:

- From an ordinary deck of playing cards, cards are to be drawn successively at random and without replacement. What is the probability that the third spade appears on the sixth draw?

Counting

$$\frac{\binom{13}{2} \binom{39}{4}}{\binom{52}{5}} * \frac{11}{47} \approx 0.0064$$

2 spades from 5 1 spade

- Same question as (a), except now with replacement.

direct

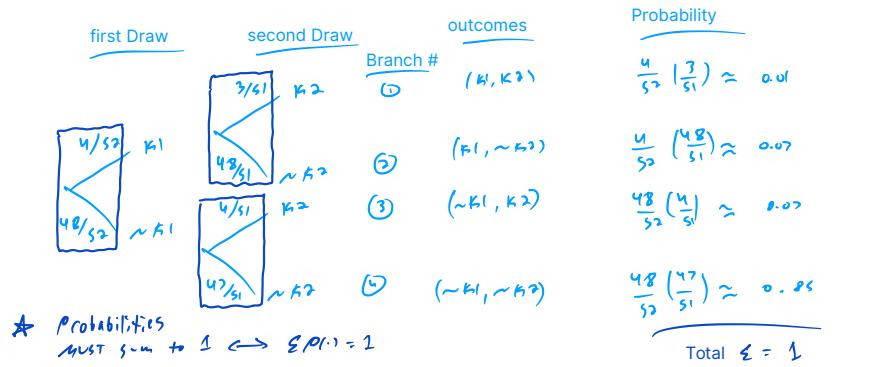
$$\overbrace{s}^{\# \text{ seq}} \quad \overbrace{s}^{\# \text{ seq}} \quad \overbrace{\sim s}^{\text{Prob}} \quad \overbrace{\sim s}^{\text{Prob}} \quad \overbrace{\sim s}^{\text{Prob}} \quad \overbrace{\sim s}^{\text{Prob}}$$

$$\Rightarrow \underbrace{\left(\frac{13}{52}\right)^2}_{\text{2 Spades from 5}} \underbrace{\left(\frac{39}{52}\right)^3}_{\text{1spade}} \left(\frac{13}{52}\right)$$

Using tree diagrams in probability problems

- Experiments involving multiple stages, such as drawing two cards without replacement can be summarized completely using trees!
- Examples:

1. Lets continue the drawing two kings example from before:



(a) What is the probability of exactly one king?

$$\begin{aligned}
 &\rightarrow \text{using tree} \quad \rightarrow P(\text{exactly 1 King}) = (2) + (3) \\
 &\rightarrow \text{Formally writing out} \rightarrow P[(K1 \cap \sim K2) \cup (\sim K1 \cap K2)] = P(K1 \cap \sim K2) + P(\sim K1 \cap K2) \\
 &\qquad\qquad\qquad\downarrow \\
 &\qquad\qquad\qquad = P(K1)P(\sim K2 | K1) + P(\sim K1)P(K2 | \sim K1) \\
 &\qquad\qquad\qquad = \frac{4}{52} \left(\frac{48}{51} \right) + \frac{48}{52} \left(\frac{4}{51} \right)
 \end{aligned}$$

(b) What is the probability of two kings or no kings?

$$\begin{aligned}
 &\curvearrowleft P(2K \text{ OR } \text{OK}) = (0) + (4) = \frac{4}{52} \left(\frac{3}{51} \right) + \frac{48}{52} \left(\frac{47}{51} \right) \\
 &\qquad\qquad\qquad P[(K1, K2) \cup (\sim K1, \sim K2)]
 \end{aligned}$$

(c) What is the probability of at least one king?

$$\begin{aligned}
 P(\text{At least 1 king}) &= (1) + (2) + (3) \\
 &\downarrow \quad = 1 - (4) \quad = 1 - \frac{48}{52} \left(\frac{47}{51} \right)
 \end{aligned}$$

can easily use complements rather than solving for many branches.

(d) Find the probability that the second card is a king (K_2), given that the first card drawn was a king.

$$P(K_2 | K_1) = \frac{3}{51}$$

– This is just knowing parts of the tree. For problems that are just asking for “one level” of the tree or “one stage” of an experiment, it is often easier to find conditional probability by

- (1) Taking into account the information by updating the scenario and then
- (2) Solving like normal.

- Creating tree diagram and summing up the final probabilities of interest simplifies many harder problems!

In most experiments, there will be a natural order of events, which will make setting up the tree intuitive.

- General tree

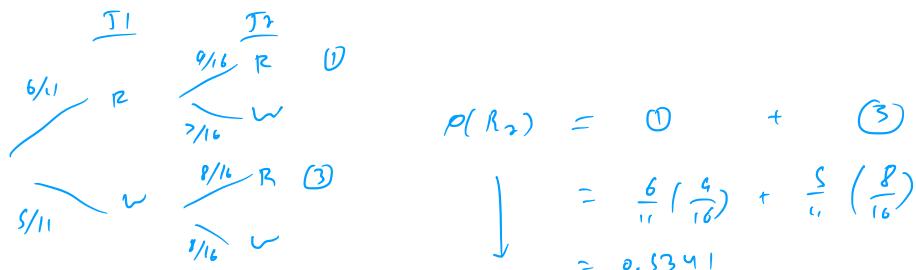
	<u>Stage 1</u>	<u>Stage 2</u>	<u>Probability and Outcome</u>
	A	$P(B A)$ • B	$P(A) \times P(B A) = P(A \cap B)$
	$P(\bar{A})$	$P(\bar{B} A)$ • \bar{B}	$P(A) \times P(\bar{B} A) = P(A \cap \bar{B})$
	\bar{A}	$P(B \bar{A})$ • B	$P(\bar{A}) \times P(B \bar{A}) = P(\bar{A} \cap B)$
		$P(\bar{B} \bar{A})$ • \bar{B}	$P(\bar{A}) \times P(\bar{B} \bar{A}) = P(\bar{A} \cap \bar{B})$

2. There are two jars, Jar 1 and Jar 2. Jar 1 has 6 red and 5 white chips and Jar 2 has 8 red and 7 white chips.

- (a) If you first pick a red chip from Jar 1 and transfer the chip to Jar 2, what is the probability of then picking a red chip from Jar 2?



- (b) If you first pick a chip (don't know which color) from Jar 1 and transfer the chip to Jar 2, what is the probability of then picking a red chip from Jar 2?



3. An insurance company sells several types of insurance policies including auto policies and homeowner policies.

- Let A be those people with an auto policy only and $P(A) = 0.3$
- Let H be those people with a homeowner's policy only and $P(H) = 0.2$
- Let $A+H$ be those people with both an auto and homeowner's policy (but no other policies) and $P(A+H) = 0.2$

Further let R be the event that the person will renew at least one of these policies. From past experience the following conditional probabilities are assigned:

$$P(R | A) = 0.6, P(R | H) = 0.7 \text{ and } P(R | A+H) = 0.8.$$

Given that a person selected at random has an auto or homeowner policy, what is the conditional probability that a person will renew at least one of those policies?

Policy Type	Renewal Status	Probability
A	R	0.3 * 0.6 = 0.18
	not R	0.3 * 0.4 = 0.12
H	R	0.2 * 0.7 = 0.14
	not R	0.2 * 0.3 = 0.06
A+H	R	0.2 * 0.8 = 0.16
	not R	0.2 * 0.2 = 0.04
other	R	0.3 * 0.6 = 0.18
other	not R	0.3 * 0.4 = 0.12

$P(R | A \cup H \cup A+H) =$

$$\begin{aligned}
 & \frac{P(R \cap (A \cup H \cup A+H))}{P(A \cup H \cup A+H)} \\
 &= \frac{(1) + (3) + (5)}{0.3 + 0.2 + 0.2} \\
 &= \frac{0.3(0.6) + 0.2(0.7) + 0.2(0.8)}{0.3 + 0.2 + 0.2} \\
 &\approx 0.686
 \end{aligned}$$

distribute, add, rewrite w/ mult rule
cancel

Lecture 5 – Independent Events

MATH 320: Probability

Lecture 5: Independent Events

Chapter 1: Probability (1.4)

The independence of events

Motivation

- For certain pairs of events, the occurrence of one of them may or may not change the probability of the occurrence of the other.
- Example: Roll a die. $A = \{1, 2, 3, 4\}$, $B = \{1, 3, 5\}$ and $C = \{4, 5, 6\}$.

Compute the probabilities of $P(C | A)$ and $P(B | A)$, then compare them with $P(C)$ and $P(B)$, respectively.

$$1. P(C | A) = \frac{n(C \cap A)}{n(A)} = \frac{1}{4} \quad 2. P(B | A) = \frac{2}{4} = \frac{1}{2} : P(B)$$
$$\neq \frac{3}{6} = P(C)$$

- How can we interpret the results above?

1. The knowledge of the occurrence of A has changed the probability of C .
2. $P(B)$ is not affected by (does not depend on) the occurrence of A .

Definition of independence

- Two events A and B , are **independent** if $P(A \cap B) = P(A) P(B)$ ①

$$\text{If } P(A) > 0 \text{ and } P(B) > 0, \text{ then } A \perp\!\!\!\perp B \iff P(A|B) = P(A) \quad \text{②} \quad P(B|A) = P(B) \quad \text{③}$$
$$\therefore \text{IND} \quad \hookrightarrow \quad P(A \cap B) = \frac{P(A) P(B)}{P(B)}$$

Otherwise, events are said to be dependent.

- ⚠ • To check for independence, we only need to check one of the three conditions. If one is true, then all are true.

Example: If a fair coin is tossed twice, then $S = \{HH, HT, TH, TT\}$. Let $H1$ be the event that the first toss is a head, and $H2$ be the event that the second toss is a head. Check if $H1$ and $H2$ are independent.

$$\rightarrow H1 = \{\text{HH}, \text{HT}\}, H2 = \{\text{HH}, \text{TH}\} \rightarrow P(H1) = P(H2) = \frac{1}{2}$$

$$\rightarrow P(H1 | H2) = \frac{1}{2} \rightarrow \frac{P(H1 \cap H2)}{P(H2)} = \frac{1}{2}$$

$$\frac{1}{2} = \frac{1}{2} \Rightarrow H1 \perp\!\!\! \perp H2$$

- Many experiments are best approached by assuming that successive trials are independent, just like successive tosses of a coin.

There is another common problem in which independence and dependence are intuitively clear.

$$\text{with replacement} \Rightarrow \text{II} \quad \& \text{without replacement} \Rightarrow \cancel{\text{II}} \\ (\text{not independent})$$

Example: Drawing cards, probabilities change if the card drawn is not replaced.

Multiplication rule for independent events

- The general multiplication rule for any two events is

$$P(A \cap B) = P(A) P(B|A) \quad \text{if } A \perp\!\!\! \perp B, \quad P(B|A) = P(B) \Rightarrow \\ \downarrow = P(A) P(B)$$

If II, find Joint by multiplying Marginal probabilities.

- Multiplication rule for independent events**

If A and B are independent events, $P(A \cap B) = P(A) \times P(B)$

- This multiplication rule makes some problems very easy if independence is immediately recognized. However, it may be tricky to check for in practice. So in many problems when it is not intuitively obvious, it is simply given as an assumption.

Example: Suppose the probability of hitting a target is 0.2 and ten shots are fired independently.

- What is the probability the target is hit at least once?

$$P(\text{Hit} \geq 1) = 1 - P(\text{Hit} = 0) \\ = 1 - \binom{10}{0} (0.2)^0 (0.8)^{10} \\ = 1 - 0.8^{10} \quad \hookrightarrow P(\text{Hit}) P(\text{Not Hit}) \dots \\ \text{Instead } \cancel{P(\text{Not Hit})} \dots \text{ bc } \text{II}$$

- What is the conditional probability the target is hit twice, given that it is hit at least once?

$$P(\text{Hit} \geq 2 | \text{Hit} \geq 1) = \frac{P(\text{Hit} \geq 2 \cap \text{Hit} \geq 1)}{P(\text{Hit} \geq 1)} \quad \cancel{\text{II}}$$

$$= \frac{P(\text{Hit} \geq 2)}{P(\text{Hit} \geq 1)} \rightarrow 1 - P(\text{Hit} = 0, 1) \\ = 1 - [0.8^{10} + \binom{10}{1} (0.2)^1 / 0.8^{10}] \\ = \frac{1 - [0.8^{10} + \binom{10}{1} (0.2)^1 / 0.8^{10}]}{1 - (0.8)^{10}}$$

- Summary:

- If A and B are independent, can easily compute $P(A \cap B) = P(A) \times P(B)$.
- 
- If A and B are mutually exclusive, can easily compute $P(A \cup B) = P(A) + P(B)$

Theorems

- If A and B are independent events, then the following pairs of events are also independent:
 - A and $\sim B$
 - $\sim A$ and B
 - $\sim A$ and $\sim B$

- Proof of (a)

Setup: Want to show $P(A \cap \sim B) = P(A) \cdot P(\sim B)$

One way:

$$A = A \cap \sim B \cup (A \cap B)$$



$$P(A) = P[(A \cap \sim B) \cup (A \cap B)]$$

$$\downarrow = P(A \cap \sim B) + P(A \cap B)$$

✓ b/c disjoint \Rightarrow

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

✓ $A \perp\!\!\!\perp B$ by assumption \Rightarrow

$$= P(A) - P(A) \cdot P(B)$$

$$= P(A) [1 - P(B)]$$

$$\checkmark = P(A) \cdot P(\sim B) \Rightarrow A \perp\!\!\!\perp \sim B$$

Another way using conditional probabilities:

$$\begin{aligned} \rightarrow P(A \cap \sim B) &= P(A) \cdot P(\sim B | A) && \text{general multiplication rule } \\ &= P(A) \cdot [1 - P(B | A)] && \text{conditional complement } \\ &= P(A) \cdot [1 - P(B)] && \text{✓ } A \perp\!\!\!\perp B \text{ by assumption } \\ &\checkmark = P(A) \cdot P(\sim B) \Rightarrow A \perp\!\!\!\perp \sim B \end{aligned}$$

- Similar logic can be used to prove (b), and (a) and (b) imply (c).

- How independence relates to the other relationships for events.

- These special cases of independence involve probabilities of zero. To check these, we have to use the most general definition of independence: $P(A \cap B) = P(A) \cdot P(B)$ because it can handle these cases.

$$\begin{aligned} P(B|A) &= \frac{P(B \cap A)}{P(A)} \\ &\downarrow = \text{undefined} \neq P(B) \end{aligned}$$

This is why the definition of independence needed the additional restrictions when checking $P(A|B) = P(A)$ and $P(B|A) = P(B)$. $P(\cdot) > 0$

- (a) If $P(A) = 0$ or $P(B) = 0$, the definition of independence always holds.

- Assume $P(A) = 0$

$$\rightarrow 0 \leq P(A \cap B) \leq P(A) = 0$$

axiom 1 $A \cap B \subseteq A$

$$\rightarrow P(A \cap B) \stackrel{?}{=} P(A) P(B)$$

$$0 \neq 0 \Rightarrow A \perp\!\!\! \perp B \text{ if } P(A) = 0$$

- similar logic for if $P(B) = 0 \Rightarrow A \perp\!\!\! \perp B$

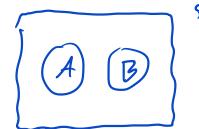
- (b) If A and B are mutually exclusive, show if A and B are independent.

Need $P(A \cap B) = P(A) \cdot P(B)$ for independence:

$$\downarrow$$

If disjoint $\rightarrow 0 = P(A) P(B)$

ONLY TRUE if $P(A) = 0$ OR $P(B) = 0$



If events are mutually exclusive, there is a very dependent relationship (if one event occurs, the other cannot occur).

- (c) If $B \subset A$, show if A and B are independent.

Need $P(A \cap B) = P(A) \cdot P(B)$ for independence.

$$\downarrow$$

If $B \subset A \rightarrow P(B) = P(A) P(B)$



ONLY TRUE if $P(A) = 0, P(B) = 0, \text{ or } P(A) = 1$

$P(A) = 0 \Rightarrow P(B) = 0$ b/c subsets are smaller

Extending the independence definition

- Before extending the definition of independence to more than two events, let's see an example.
- Example: A jar contains four marbles numbered 1, 2, 3, and 4. One marble is to be drawn at random. Let the events A , B , and C be defined by $A = \{1, 2\}$, $B = \{1, 3\}$ and $C = \{1, 4\}$.

a) check if pairs are independent

$$\begin{aligned} & \rightarrow P(A)P(B) = P(A)P(C) = P(B)P(C) = \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{4} \\ & \rightarrow P(A \cap B) = P(A \cap C) = P(B \cap C) = \frac{1}{4} \\ & \Rightarrow A \perp\!\!\!\perp B, A \perp\!\!\!\perp C, B \perp\!\!\!\perp C \end{aligned}$$

b) check if all three are independent

$$P(A \cap B \cap C) = ? P(A)P(B)P(C) \quad \left. \begin{array}{l} P(A) = \frac{1}{2} \\ P(B) = \frac{1}{2} \\ P(C) = \frac{1}{2} \end{array} \right\} \Rightarrow \text{not all } \perp\!\!\!\perp$$

This shows that A , B , and C are pairwise independent, but not mutually independent.

- Definition: Events A , B , and C are **mutually independent** if and only if they are pairwise independent (i.e. (A, B) , (A, C) and (B, C) are independent pairs) and if $P(A \cap B \cap C) = P(A)P(B)P(C)$.

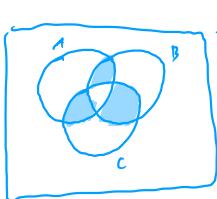
Examples

1. Suppose that A , B and C are mutually independent events and that $P(A) = 0.5$, $P(B) = 0.8$ and $P(C) = 0.9$. Find the following probabilities:

- (a) All three events occur.

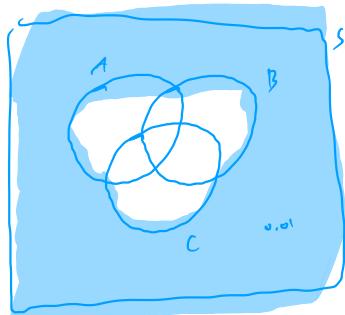
$$\begin{aligned} P(A \cap B \cap C) &= P(A) P(B) P(C) \\ &\downarrow \\ &= 0.5 \cdot 0.8 \cdot 0.9 = 0.36 \end{aligned}$$

- (b) Exactly two of the three events occur



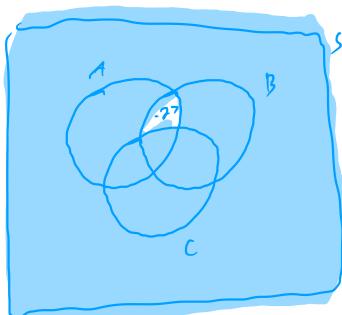
$$\begin{aligned} P(A \cap B \cap \sim C) + P(A \cap \sim B \cap C) + P(\sim A \cap B \cap C) \\ &= 0.5 \cdot 0.8 \cdot 0.1 + 0.5 \cdot 0.2 \cdot 0.9 + 0.5 \cdot 0.8 \cdot 0.1 \\ &= 0.04 + 0.09 + 0.04 \\ &= 0.27 \end{aligned}$$

- (c) None of the events occur



$$\begin{aligned} P(\sim A \cap \sim B \cap \sim C) &= P(\sim A) P(\sim B) P(\sim C) \\ &\downarrow \\ &= 0.5 \cdot 0.2 \cdot 0.1 \\ &= 0.01 \end{aligned}$$

2. Let A , B and C be (mutually) independent events such that $P(A) = 0.5$, $P(B) = 0.6$ and $P(C) = 0.1$. Calculate $P(\sim A \cup \sim B \cup C)$.



$$\begin{aligned} P(\sim A \cup \sim B \cup C) &= 1 - P(A \cap B \cap \sim C) \\ &\downarrow \\ &= 1 - 0.5 \cdot 0.6 \cdot 0.9 \\ &= 1 - 0.27 \\ &= 0.73 \end{aligned}$$

Lecture 6 – Bayes’ Theorem

MATH 320: Probability

Lecture 6: Bayes' Theorem

Chapter 1: Probability (1.5)

Law of total probability

Motivation

- Example: A company has three assembly lines. The first line, the second line and the third produce 30%, 50% and 20% of productions, respectively.

Additionally, 1 of 100 productions is defective in Line 1; 2 of 100 productions are defective in Line 2; 3 of 100 productions are defective in Line 3.

We want to know the probability of defectives produced in the company.

• STRATEGY:

- Define all (unconditional) events given in the problem and find their probabilities.

If conditional events are given, define them using unconditional events.

$$\begin{array}{ll} L_1 = \text{Line 1} \rightarrow P(L_1) = 0.3 & D = \text{Defective} \rightarrow P(D|L_1) = 0.01 \\ L_2 = \text{Line 2} \rightarrow P(L_2) = 0.5 & P(D|L_2) = 0.02 \\ L_3 = \text{Line 3} \rightarrow P(L_3) = 0.2 & P(D|L_3) = 0.03 \end{array}$$

- Find the event of interest. Try to express the event of interest as a composition (union) of the given events.

Often it is desirable to form compositions mutually exclusive or independent events.

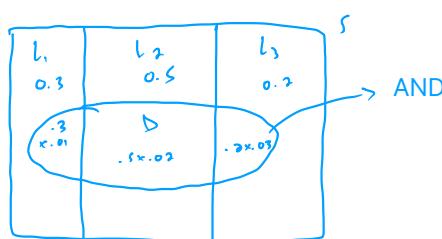
$$\begin{aligned} P(D) &= P[(D \cap L_1) \cup (D \cap L_2) \cup (D \cap L_3)] \\ &\downarrow = P(D \cap L_1) + P(D \cap L_2) + P(D \cap L_3) \quad \downarrow \text{skip to here} \end{aligned}$$

- Use the general multiplication rule to find the probability for the event of interest.

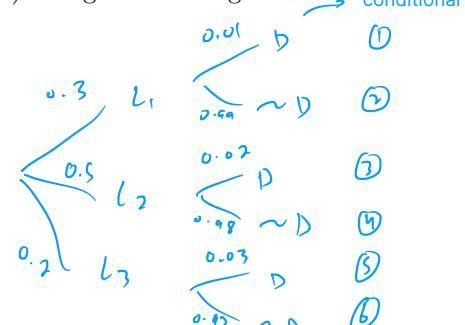
$$\begin{aligned} P(D) &= \stackrel{(1)}{P(L_1)} \cdot \stackrel{(2)}{P(D|L_1)} + \stackrel{(3)}{P(L_2)} \cdot \stackrel{(4)}{P(D|L_2)} + \stackrel{(5)}{P(L_3)} \cdot \stackrel{(6)}{P(D|L_3)} \\ &\downarrow = 0.3 \cdot 0.01 + 0.5 \cdot 0.02 + 0.2 \cdot 0.03 \\ &= 0.019 \end{aligned}$$

- Visualizing scenario:

- Using a Venn Diagram:

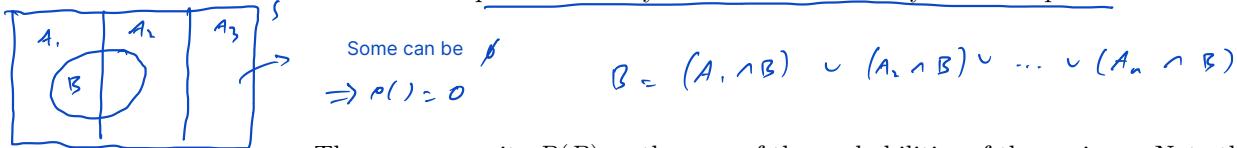


- Using a Tree Diagram:



Law of total probability

- The **law of total probability** says that a partition (A_1, \dots, A_n) of the sample space will lead to a partition of any event B into mutually exclusive pieces.



Then we can write $P(B)$ as the sum of the probabilities of those pieces. Note that an event A and its complement $\sim A$ always partition S .

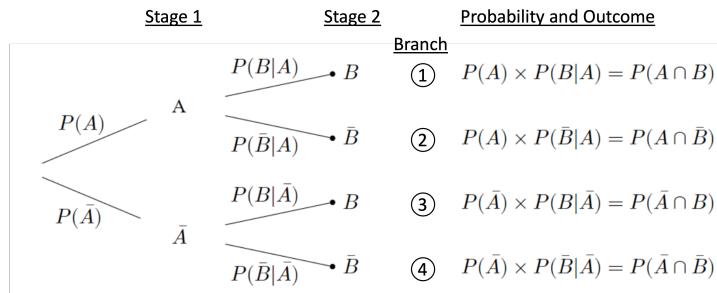
- Definition: **Law of total probability**

Let B be an event. If A_1, \dots, A_n partition the sample space, then

$$\begin{aligned} P(B) &= P\left[\bigcup_{i=1}^n (A_i \cap B)\right] \\ &\stackrel{\text{partition}}{=} P(A_1 \cap B) + \dots + P(A_n \cap B) \\ &\stackrel{\text{disjoint}}{=} P(A_1) P(B|A_1) + \dots + P(A_n) P(B|A_n) \\ &\stackrel{\text{multiplication rule}}{=} \sum_{i=1}^n P(A_i) P(B|A_i) \end{aligned}$$

- Using the general tree diagram below, we can summarize

Law of total probability = $P(\text{Second stage event}) = \sum \text{Branches of interest}$



$$P(B) = \textcircled{1} + \textcircled{3}$$

$$P(A|B) = \frac{\textcircled{1}}{\textcircled{1} + \textcircled{3}}$$

Bayes' Theorem

Bayes' Theorem

- Using the same terminology, we can summarize

Bayes' Theorem = $P(\text{First stage event} | \text{Second stage event}) = \frac{\text{Main branch of interest}}{\sum \text{All branches of interest}}$

- In essence, Bayes' Theorem reverses the natural order of the tree for the conditional probability of interest.

- Continuing example:

Find the probability that a defective product was made in Line 1.

$$\begin{aligned} P(L_1 | D) &= \frac{P(L_1 \cap D)}{P(D)} = \frac{\textcircled{1}}{\textcircled{1} + \textcircled{3} + \textcircled{2}} \\ &= \frac{0.3(0.01)}{0.3(0.01) + 0.5(0.02) + 0.7(0.03)} \\ &\approx 0.158 \end{aligned}$$

• Definition: **Bayes' Theorem**

Let B be an event. If A_1, \dots, A_n partition the sample space, then

$$\begin{aligned} P(A_i | B) &= \frac{P(A_i \cap B)}{P(B)} \\ &= \frac{P(A_i) P(B|A_i)}{\sum_{j=1}^n P(A_j) P(B|A_j)} \quad \text{law of total probability} \end{aligned}$$

- Example: At the beginning of a certain study of a group of persons, 15% were classified as heavy smokers, 30% as light smokers, and 55% as nonsmokers. In the five year study, it was determined that the death rates of the heavy and light smokers were five and three times that of the nonsmokers, respectively.

A randomly selected participant died over the five-year period; calculate the probability that the participant was a nonsmoker.

The tree diagram illustrates the distribution of smokers across three categories: Heavy (0.15), Light (0.3), and Non (0.55). For each category, there are two branches representing death (D) and survival ($\sim D$). The death rates are indicated by circled numbers: Heavy smokers have a rate of 5 (circled 1), Light smokers have a rate of 3 (circled 2), and Non-smokers have a rate of 1 (circled 3).

$$\begin{aligned} P(\text{non} | D) &= \frac{P(\text{non} \cap D)}{P(D)} \\ &= \frac{\textcircled{3}}{\textcircled{1} + \textcircled{2} + \textcircled{3}} \\ &= \frac{0.55}{0.15(5) + 0.3(3) + 0.55(1)} \\ &= 0.25 \end{aligned}$$

Bayes' Theorem from another perspective

- Bayes' Theorem is all about changing probabilities based on new evidence.
- In a previous example, we drew a tree diagram about testing for the presence of a disease and the result of the test. We used the following events:

D = the person tested has the disease

$\sim D$ = the person tested does not have the disease

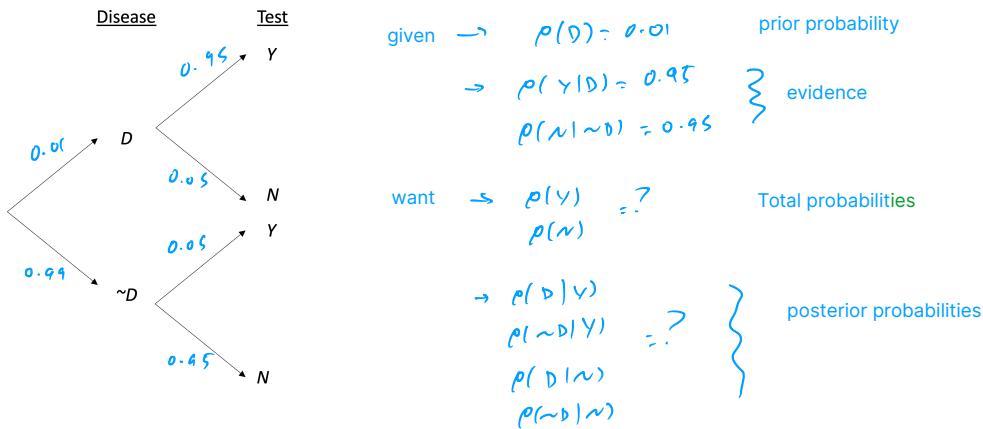
Y = the test is positive

N = the test is negative

Lets now consider a disease test that is "95% accurate", which can be defined as follows:

- If you have the disease, 95% chance of a positive test.
- If you do not have the disease, 95% chance of a negative test.

Further, suppose only 1% of the population actually have this disease (aka prevalence).



- Terminology:
 - Prior probability: Original unconditional probabilities
 - Evidence: Conditional probability given the prior information
 - Posterior probability: Prior probability conditioned on the new evidence
- Some calculations in context: The good and the bad of Bayes' Theorem
 - Find the probability of testing positive (total probability).

$$\begin{aligned}
 P(Y) &= \textcircled{1} + \textcircled{2} \\
 &= P(D, Y) + P(\sim D, Y) \\
 &= 0.01(0.95) + 0.99(0.05) \\
 &= 0.059
 \end{aligned}$$

- Lets' solve for the probability of having the disease *given that you test positive*.
 - For a randomly selected person from the population, we had our original prior probability of having the disease, $P(D) = 0.01$.
(We don't know if they do or don't have the disease, it remains unknown).
 - Then this person got tested, and tested positive; this is our evidence.

Intuitively, this likelihood of the person having the disease should increase; we are adjusting the prior probability upwards based on the new evidence.

(c) Now we can calculate this new posterior probability.

$$P(D|Y) = \frac{\textcircled{1}}{\textcircled{1} + \textcircled{2}} = \frac{P(D,Y)}{P(Y)} = \frac{0.01(0.05)}{0.059} = 0.161$$

3. Suppose you know that someone has tested positive for this disease. What is the probability that the person does not actually have the disease?

$$P(\sim D|Y) = 1 - P(D|Y) = 1 - 0.161 = 0.839$$

$$\Downarrow = \frac{\textcircled{2}}{\textcircled{1} + \textcircled{2}} \quad \frac{P(\sim D,Y)}{P(Y)} = \frac{0.99(0.05)}{0.059} = 0.839$$

- The practical information here is interesting.

- The "95% accurate" test will classify 5.9% of the population as positives, compared to the true prevalence of $P(D) = 0.01$.
- This is the good side of Bayes' Theorem! By updating our prior probability with the new evidence, we drastically increased our information about this person having the disease.

$$P(D) = 0.01 \rightarrow P(D|Y) = 0.161$$

- 83.9% of the individuals who tested positive will actually not have the disease.

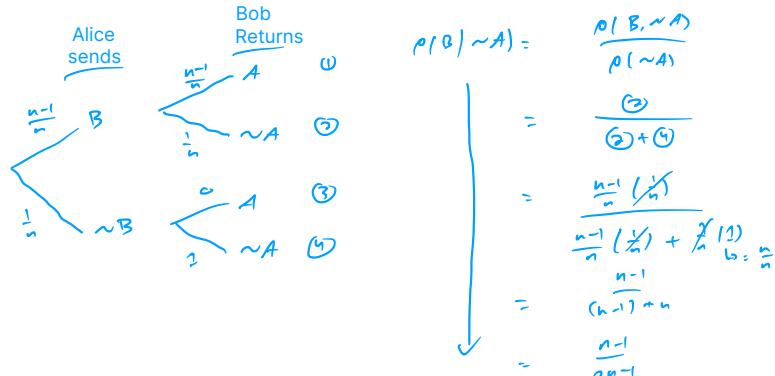
This percentage depends heavily on the prevalence, for example if $P(D) = 0.1 \rightarrow P(\sim D | Y) = \underline{32.1\%}$; and if $P(D) = 0.001 \rightarrow P(\sim D | Y) = \underline{98.1\%}$.

Final example

given

- Alice writes to Bob and does not receive an answer. Assuming that one letter in n is lost in the mail, find the probability that Bob received the letter. It is to be assumed that Bob would have answered the letter if he had received it.

Let A = Alice receives letter from Bob and B = Bob receives letter from Alice.



Test 2

Contents

Lecture 7 – Random Variables	53
Lecture 8 – Distribution Functions	57
Lecture 9 – Summary Measures	71

Lecture 7 – Random Variables

MATH 320: Probability

Lecture 7: Random Variables

Chapters 2 and 3: Distributions (2.1 and 3.1)

Why do we study statistics?

- The main purpose of studying statistics is because we want to study experiments and their outcomes.
- We want to analyze data from experiments numerically. But, outcomes are not always quantitative.
 - So we have to assign numbers to outcomes. Thus, random variables connect outcomes to numbers.
 - The advantage using random variables is that they are easily summarized.
- Intuitive definition: A **random variable** is a numerical quantity whose value depends on chance.

Types of random variables (RVs)

- Examples: Determine if each describes a RV.
i.e. Is the outcome (a) is a number? (b) depends on chance?

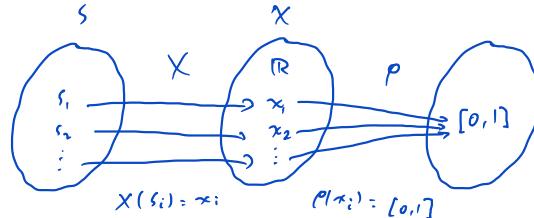
#	chance	Rv
✓	✓	⇒ ✓
✓	✓	⇒ ✗
✗	✓	⇒ ✓
✓	✓	⇒ ✓

Discrete {
1. You are tossing a coin twice and will bet on the number of heads.
2. You go to Las Vegas and begin to put quarters in a slot machine. Let X be the number of quarters you play in order to first win of any amount.
3. You are tossing a coin twice and will bet on specific outcomes such as HT .

Continuous {
4. A resident of Muncie is selected at random, and their height is measured.
- Similar to sample spaces, there are different kinds of random variables.
★ ★ This will be a very important distinction to make at the start of every single problem for the rest of the course.
- Random variables can be discrete (only distinct values are possible) or continuous (measured on a continuous scale).
 - When classifying a random variable as discrete or continuous, we are really just identifying the kind of mathematical model we will use.
 - Calculus-based mathematics is the most efficient way to analyze a random variable such as heights (which we may only measure as discrete to a certain precision).

Definitions and notation

- Functions *map* the input (domain, support) to the output (range).
- Our general definition of probability was a way to assign a probability $P(A)$ to any event A where all the axioms needed to be satisfied. This, more formally, is a function.
- A **random variable** is a function from a sample space S into real numbers.



Random variable

Input: outcome s_i

Output: any real # x_i

Maps: $S \rightarrow \mathbb{R}$

Probability

any real # $p(x_i)$

$[0,1]$

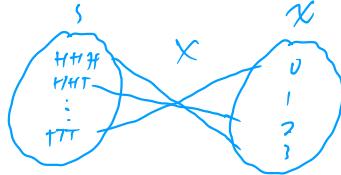
events $\rightarrow [0,1]$

A

- Notation: We will use uppercase letters, such as X, Y, Z, \dots to denote a random variable and lowercase letters, such as x, y, z, \dots to denote a particular value that a random variable may assume.
- Definition: The set of possible values of X is the **range** of X , \mathcal{X} .
- Summary of notation:
 - X = Random variable.
 - x_i = Individual values of X .
 - \mathcal{X} = Range of $X \rightarrow$ set of all $x_i = \{x_1, x_2, \dots\}$ or $[x_a, x_b]$
- It is important to know the distinction between the outcomes in an experiment (sample space) and the range.
- Examples:
 1. Toss three fair coins and observe the results. Let X equal the number of heads obtained.
 - (a) What is the sample space and range of X ?

$$S = \{HHH, HHT, \dots, TTT\} \rightarrow X = \{0, 1, 2, 3\}$$

(b) Show the connection between S and X .



2. Let X be the time to failure for a machine part. Find the range.

$$X: [0, \infty)$$

3. You are waiting for the bus to arrive. If it arrives in under 5 minutes, you will get on the bus. If not, you will walk to your destination.

Let X be the random variable such that $X = 1$ if you get on the bus and $X = 0$ if you walk. Is X a continuous or discrete random variable?



- Types of random variables definitions

X is a **discrete random variable** if the range X is a finite or countable set.

X is a **continuous random variable** if the range X is an interval (or union of intervals) on the real number line.

Connection between random variables and probability

- We would like use random variables to express events, because we can calculate probabilities of events.
- Notation: $\{X = x\}$ is the set of outcomes in the sample space assigned the value x by the random variable X .
 $X = x$ means the random variable X was realized with a specific value x .
So it is an event. As a result, we can compute the probability of $\{X = x\}$.
- Notation: We used to have events like $A \cap B$ or now $\{X = x\}$ in $P(\cdot)$, but we will now use $P(X = x)$ for simplicity.

Example: Continuing the previous three coin toss scenario, find the following events and their probabilities:

$$\{X = 1\} = \{HHT, THT, TTH\} \rightarrow P(X=1) = 3/8$$

$$\{X = 3\} = \{HHH\} \rightarrow P(X=3) = 1/8$$

Lecture 8 – Distribution Functions

MATH 320: Probability

Lecture 8: Distribution Functions

Chapters 2 and 3: Distributions (2.1 and 3.1)

Probability functions

Probabilities for discrete random variables

- Definition: The **probability mass function (pmf)** of a discrete random variable X is given by

$$f_X(x) = P(X = x), \quad \text{for all } x$$

- For a discrete random variable with a small number of outcomes, the pmf can be given in a table. When there is a very large or infinite number of possible outcomes, $f(x)$ can be given in a formula.
- Example: The number of injury claims per month is modeled by a random variable N with

$$P(N = n) = \frac{1}{(n+1)(n+2)}, \quad \text{where } n \geq 0.$$

- a) Determine the probability of two claims.

n	0	1	2	\dots
$P(N=n)$	$\frac{1}{2}$	$\frac{1}{6}$	$\frac{1}{12}$	\dots

$$\begin{aligned} P(N=2) &= \frac{1}{(2+1)(2+2)} \\ &= \frac{1}{3 \cdot 4} \\ &= \frac{1}{12} \end{aligned}$$

- b) Determine the probability of at most two claims.

$$\begin{aligned} P(N \leq 2) &= P(N=0, 1, 2) \\ &= P(N=0) + P(N=1) + P(N=2) \\ &= \frac{1}{2} + \frac{1}{6} + \frac{1}{12} \\ &= \frac{9}{12} \end{aligned}$$

- c) Determine the probability of at least two claims.

$$\begin{aligned} P(N \geq 2) &= 1 - P(N \leq 1) \\ &= 1 - P(N \leq 1) \\ &= 1 - [P(N=0) + P(N=1)] \\ &= 1 - [\frac{1}{2} + \frac{1}{6}] \\ &= \frac{2}{3} \end{aligned}$$



Probabilities for continuous random variables

- For continuous random variables, does the pmf exist? **No!**

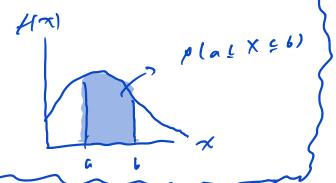
These intervals are not countable, so we cannot use a pmf to assign probabilities.

- Instead, to find the general probability $P(a \leq X \leq b)$, we find the area bounded by $f(x)$ and the x -axis between $x = a$ and $x = b$ using integration.

More specifically, we use a density function and find areas under the density function curve.

- Definition: A **probability density function (pdf)** is a continuous random variable X is a real-valued function that can be used to find probabilities using

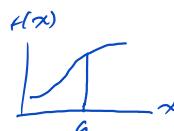
$$P(a \leq X \leq b) = \int_a^b f(x) dx$$



- Notes:

- There is no probability associated with a single point:

$$\text{For } a \in \mathcal{X}, \quad P(X = a) = \int_a^a f(x) dx = 0$$



- As a result: For any interval (a, b) , it doesn't matter if we include or exclude the endpoints in the continuous case (unlike with discrete).

$$\text{For } (a, b) \in \mathcal{X}, \quad P(a < X < b) = P(a \leq X \leq b) = \int_a^b f(x) dx$$

$a < x \leq b$
 $a \leq x < b$ (implied $a \neq b$)

- Example: Let $f(x) = \frac{1}{12}(x^2 + 1)$, for $0 \leq x \leq 3$.

Find: (a) $P(X \leq 1)$

(b) $P(X \geq 1)$

(c) $P(0.5 \leq X \leq 1.5)$.

$$\begin{aligned} & \int_0^1 \frac{1}{12}(x^2 + 1) dx \\ &= \frac{1}{12} \left[\frac{1}{3}x^3 + x \right] \Big|_0^1 = F(1) - F(0) \\ &= \frac{1}{12} \left[\frac{1}{3} + 1 \right] \\ &= \frac{1}{9} \end{aligned}$$

$$\begin{aligned} & \int_1^3 \frac{1}{12}(x^2 + 1) dx \\ &= \left[-\frac{1}{12}(x^2 + 1) \right] \Big|_1^3 = 1 - P(X \leq 1) \\ &= 1 - \frac{1}{9} \\ &\approx 8/9 \end{aligned}$$

$$\begin{aligned} & \int_{1/2}^{3/2} \frac{1}{12}(x^2 + 1) dx \\ &= \frac{1}{12} \left[\frac{1}{3}x^3 + x \right] \Big|_{1/2}^{3/2} = F(3/2) - F(1/2) \leftarrow \begin{array}{l} \text{Show work} \\ \text{at least up to} \\ \text{this stem on} \\ \text{Integrals,} \\ \text{then can calc} \end{array} \\ &= \frac{1}{12} \left[\frac{1}{3} \left(\frac{3}{2}\right)^3 + \frac{3}{2} \right] - \frac{1}{12} \left[\frac{1}{3} \left(\frac{1}{2}\right)^3 + \frac{1}{2} \right] \\ &\approx 0.174 \end{aligned}$$

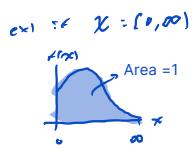


Valid pmfs and pdfs

- There are rules that these new probability functions must follow, similar to the axioms our original probability assignments needed to satisfy.

- Theorem: A function $f_X(x)$ is a pdf (or pmf) of a random variable X if and only if
 - $f_X(x) \geq 0$ for all x . **Non-negative**

$$(b) \sum_x f_X(x) = 1 \text{ (pmf)} \quad \text{or} \quad \int_{-\infty}^{\infty} f_X(x) dx = 1 \text{ (pdf). Total probability}$$

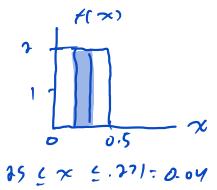


- Important note about $f(x)$ values.

-  – For the discrete case, $f(x)$ values were actually probabilities.
e.g. if $f(5) = 0.23$, there is a 23% probability of $x = 5$.
- For the continuous case, the values of $f(x)$ are NOT probabilities themselves; they define areas which give probabilities.

The values $f(x)$ must be positive, but they can be greater than 1.

Example: Let $f(x) = 2$, $0 \leq x \leq 0.5$.



- Note these are why continuous variables use density functions and not mass functions.
Cannot assign probability to every single point \Leftrightarrow Can't satisfy $\sum_{x \in \mathcal{X}} f(x) = 1$.

infinite range on \mathbb{R} $\mathcal{X} = [0, \infty]$ 

- Examples:

- Verify the pmf for X is valid.

x	1	2	3	4
$f_X(x)$	0.43	0.12	0.3	0.15

① $\sum f(x) = 1$ ✓
② $\sum f(x) = 1$ ✓

- Verify $f_X(x) = 2x^{-3}$, $1 \leq x < \infty$ is a valid pdf.

$$\text{③ } f(x) \geq 0 \text{ over } [1, \infty] \quad \text{④ } \int_1^{\infty} 2x^{-3} dx = \left[-\frac{1}{2}x^{-2} \right]_1^{\infty} = \frac{1}{2} + \frac{1}{2}$$

- Let X have the following pdf:

$$f_X(x) = \begin{cases} cx^2 & 0 \leq x \leq 2 \\ 0 & \text{otherwise} \end{cases}$$

Find the value of c for which $f_X(x)$ is a valid pdf.

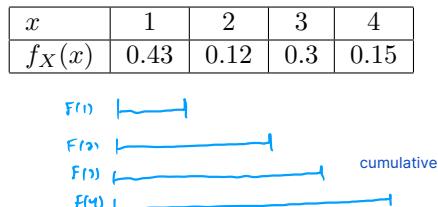
$$\begin{aligned} \Rightarrow 1 &= \int_0^2 cx^2 dx = \left[\frac{c}{3}x^3 \right]_0^2 \\ &\downarrow \\ &= \frac{8}{3}c \\ &\Rightarrow c = \frac{3}{8} \end{aligned}$$

$$\Rightarrow \int_0^2 \frac{3}{8}x^2 dx = 1 \quad \text{+ all } f(x) \geq 0 \text{ over } [0, 2]$$

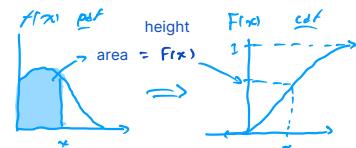
The cumulative distribution function

Concept of a cdf

- Examples: Discrete case



$$f_X(x) = \frac{3}{8}x^2 \quad 0 \leq x \leq 2$$



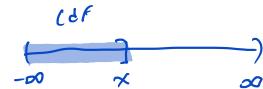
- In these examples, probabilities were obtained by cumulatively adding successive probabilities in the table above or by accumulating more probability as we increased the upper bound.

If we do this throughout the entire range of either case, we obtain the cumulative distribution function $F(x)$. Every random variable X has an associated cumulative distribution function that can be defined as follows.

Defining a cdf

- Definition: The **cumulative distribution function** or **cdf** of a random variable X , denoted $F_X(x)$, is defined by

$$F_X(x) = P_X(X \leq x), \quad \text{for all } x.$$



- Notes about $F(x)$: ALWAYS

- The cdf is defined for $-\infty < x < \infty$ always.
 - The range of every cdf is $0 \leq F(x) \leq 1 \iff$
 - $F_X(x)$ is a non-decreasing function.

- DISCRETE case

Example:

- (a) Using the pmf table for X below, find the cdf $F_X(x)$ as a table.

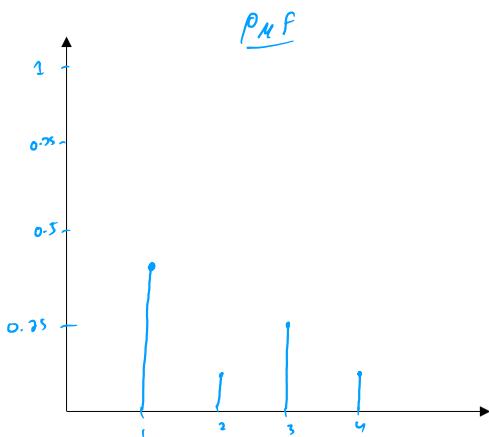
x	1	2	3	4
$f_X(x)$	0.43	0.12	0.3	0.15
$F_X(x)$	0.43	0.55	0.85	1

(b) Write $F_X(x)$ as a piecewise function:

$F_X(x) = \begin{cases} 0 & x < 1 \\ 0.43 & 1 \leq x < 2 \\ 0.55 & 2 \leq x < 3 \\ 0.85 & 3 \leq x \leq 4 \\ 1 & x > 4 \end{cases}$

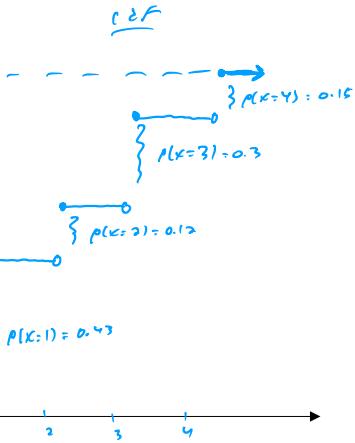
No probability until 12
Spike at 1, constant until 2

(c) Plot the pmf and cdf.



Observations of pmf

1) Sum of the heights = 1



Properties of cdf

2) Positive values (probabilities) only at $x = 1, 2, 3, 4$

\Rightarrow left limit is 0 & right limit is 1

3) No probability between x values

\Rightarrow Constant (horizontal) between x values
(with value at next smallest x)

Properties

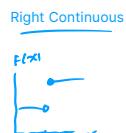
- The last entry in a table for $F(x)$ of a finite discrete random variable will always be 1.
- Even though $f(x)$ is only defined for certain values of x , we can define $F(x)$ for any real number.
- $F_X(x)$ is a right-continuous step-function.

• CONTINUOUS case

Example: Let ~~continuous~~

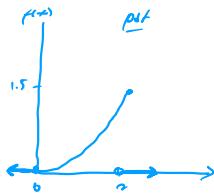
$$(a) \text{ Find the cdf } F(x). \quad f(x) = \begin{cases} \frac{3}{8}x^2 & 0 \leq x \leq 2 \\ 0 & \text{otherwise} \end{cases}$$

$$F_X(x) = \int_0^x \frac{3}{8}t^2 dt = \left[\frac{1}{8}t^3 \right]_0^x = \frac{1}{8}x^3, \quad 0 \leq x \leq 2$$

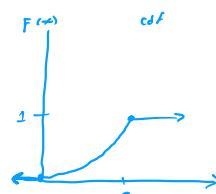


$$F_X(x) = \begin{cases} 0 & x < 0 \\ \frac{1}{8}x^3 & 0 \leq x < 2 \\ 1 & x \geq 2 \end{cases} \rightarrow \text{check } F_X(0) = \frac{1}{8}0^3 = 0 \\ F_X(2) = \frac{1}{8}2^3 = 1$$

(b) Plot the pdf and cdf.



Observations of ~~cdf~~



Properties of cdf

1) Plot starts at 0 and ends at 1 $\Rightarrow 0 \leq F(x) \leq 1$

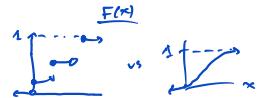
2) Continuous at change points \Rightarrow continuous over \mathbb{R}
(not necessarily differentiable)

Properties

* check

- $F_X(\text{lower limit}) = 0$ and $F_X(\text{upper limit}) = 1$

- $F_X(x)$ is always a continuous function (even though the pdf $f_X(x)$ does not necessarily have to be continuous over \mathbb{R}).



- Types of random variables (another way to define).
 - A random variable X is **discrete** $\Leftrightarrow F_X(x)$ is a step function of x .
 - A random variable X is **continuous** $\Leftrightarrow F_X(x)$ is a continuous function of x .

Relationship between the cdf and pdf

- Since $F(x)$ is defined by integrating $f(x)$, it is clear that the derivative of $F(x)$ is $f(x)$. This simple relationship is very important when the derivative $F'(x)$ exists.

$$\boxed{F'(x) = f(x)}$$

- Said another way: We can define the **pdf** of a continuous random variable X as the function that satisfies

$$F_X(x) = \int_{-\infty}^x f(t) dt \quad \text{for all } x.$$

Then using the Fundamental Theorem of Calculus, if $f_X(x)$ is continuous,

$$\frac{d}{dx} F_X(x) = f_X(x)$$

- This relationship means that the pdf (or pmf) contains the same information as the cdf. So knowing one of these about a random variable is very important and allows researchers to analyze it many, many ways.

know pdf (or pmf) \Leftrightarrow know Cdf

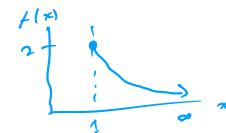
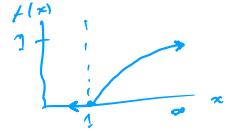
same info

- Example: Let X have the following cdf:

$$F(x) = P(X \leq x) = 1 - \frac{1}{x^2}, \quad 1 \leq x < \infty$$

Find the pdf $f(x)$.

$$f(x) = F'(x) = \frac{d}{dx} (1 - x^{-2}) \\ \downarrow \\ = 2x^{-3}, \quad x \geq 1$$



Finding probabilities using the cdf



- ALWAYS
 - Cdf gives a LEFT probability.



- DISCRETE case
 - Cdf adds all probabilities of points less than or equal to x . Formula version of this for a particular value $x = a$:

$$F(a) = P(X \leq a) = \sum_{x \leq a} f(x)$$



- The complement of this is:

$$1 - F(x) = 1 - P(X \leq x) = P(X > x)$$

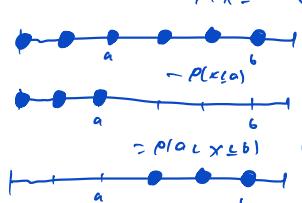


which represents the probability X is greater than x , exclusive.



- Interval probabilities: $P(a < X \leq b) = P(X \leq b) - P(X \leq a) = F(b) - F(a)$

$$P(a < X \leq b) = P(X \leq b) - P(X \leq a) \approx F(b) - F(a)$$

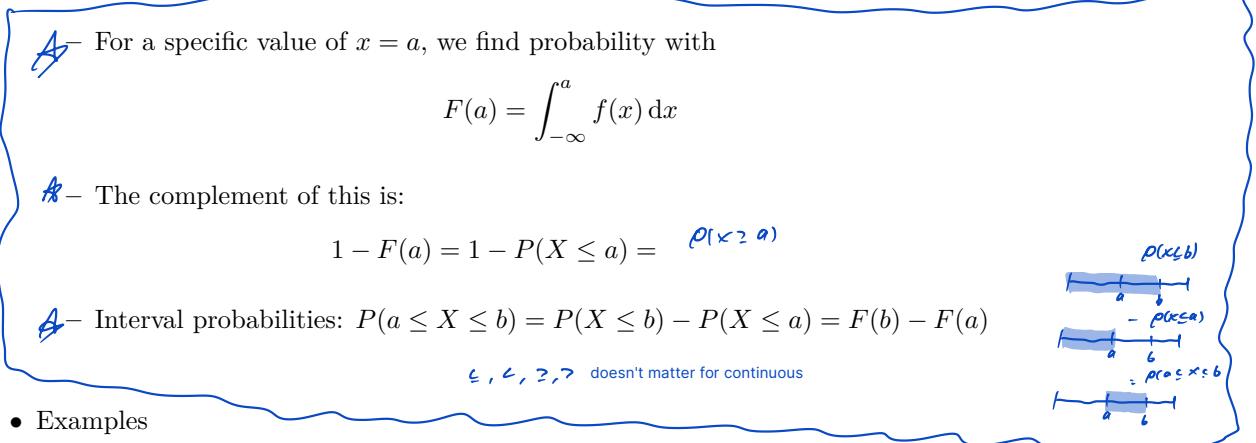


- CONTINUOUS case

- Cdf accumulates all of the probability less than or equal to x , which means we are finding the probability up to x . So this x is the upper bound of the integral.

To not confuse our letters, we change the letter of the variable in the function being integrated (and in the differential dt).

$$F_X(x) = \int_{-\infty}^x f(t) dt$$

 A - For a specific value of $x = a$, we find probability with

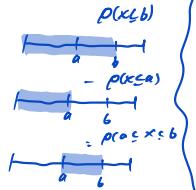
$$F(a) = \int_{-\infty}^a f(x) dx$$

A - The complement of this is:

$$1 - F(a) = 1 - P(X \leq a) = P(X > a)$$

A - Interval probabilities: $P(a \leq X \leq b) = P(X \leq b) - P(X \leq a) = F(b) - F(a)$

$\leftarrow, \leftarrow, \geq, \geq$ doesn't matter for continuous



- Examples

1. Let X have the cdf table below.

x	1	2	3	4	5
$F_X(x)$	0.16	0.63	0.67	0.78	1.00

Find (a) $P(X \leq 2)$ (b) $P(X > 3)$ (c) $P(X \geq 3)$ (d) $P(X < 4)$

$$\therefore F(2) = 0.63$$

$$1 - P(X \leq 3)$$

$$= 1 - F(3)$$

$$= 1 - 0.67$$

$$= 0.33$$

$$1 - P(X \leq 2)$$

$$= 1 - F(2)$$

$$= 1 - 0.63$$

$$= 0.37$$

$$P(X \leq 3)$$

$$= F(3)$$

$$= 0.67$$

(e) $P(2 \leq X \leq 4)$

$$F(4) - F(2)$$

$$= 0.78 - 0.63$$

$$= 0.15$$

(f) $P(1 < X \leq 5)$

$$P(1 \leq X \leq 5)$$

$$= F(5) - F(1)$$

$$= 1 - 0.16$$

$$= 0.84$$

2. Let $F_X(x) = \frac{x^3}{8}$, $0 \leq x \leq 2$.

Find (a) $P(X \leq 1)$ (b) $P(X < 1.5)$ (c) $P(X \geq 1.25)$ (d) $P(X > 0.75)$

$$F(1) = \frac{1}{8}$$

$$F(1.5) = \frac{1.5^3}{8}$$

$$1 - F(1.25)$$

$$= 1 - \frac{1.25^3}{8}$$

$$1 - F(0.75)$$

$$= 1 - \frac{0.75^3}{8}$$

(e) $P(0.25 \leq X \leq 1)$

$$F(1) - F(0.25)$$

$$= \frac{1^3}{8} - \frac{0.25^3}{8}$$

(f) $P(1 < X \leq 1.5)$

$$F(1.5) - F(1)$$

$$= \frac{1.5^3}{8} - \frac{1^3}{8}$$

Examples

• Discrete

1. The cdf for the years until patients are asymptomatic for a certain disease is shown below:

Number of years (x)	1	2	3	4	5
$F(x)$	0.53	0.78	0.9	0.97	1.00

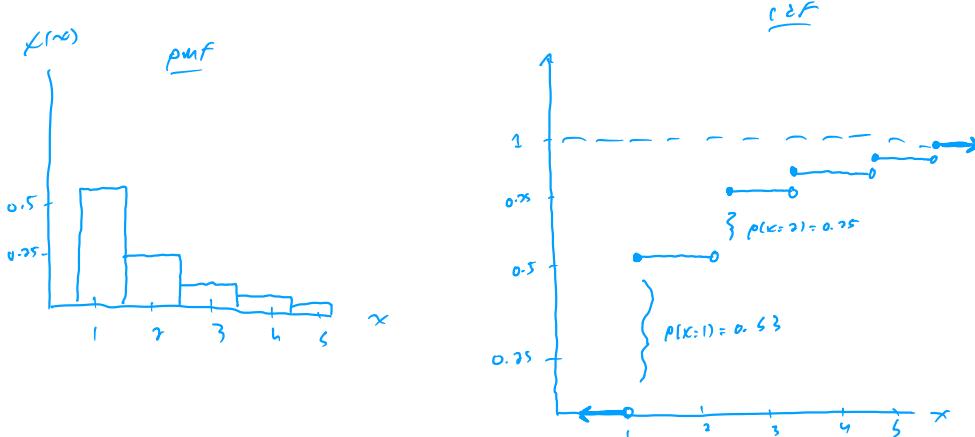
- a) Find $P(X < 4)$ and $P(X \leq 3)$.

$$\begin{aligned} \underline{P(X < 4)} &= F(3) = 0.9 \\ &\quad \nearrow F(3) - F(2) \\ &= 0.9 - 0.78 \\ &= 0.12 \end{aligned}$$

- b) Write the piecewise functions of the pmf and cdf.

$$\begin{aligned} f(x) &= \begin{cases} 0.53 & x=1 \\ 0.25 & x=2 \\ 0.12 & x=3 \\ 0.07 & x=4 \\ 0.03 & x=5 \end{cases} \\ F_X(x) &= \begin{cases} 0 & x < 1 \\ 0.53 & 1 \leq x < 2 \\ 0.78 & 2 \leq x < 3 \\ 0.90 & 3 \leq x < 4 \\ 0.97 & 4 \leq x < 5 \\ 1 & x \geq 5 \end{cases} \end{aligned}$$

- c) Plot a histogram of the pmf and a line graph of the cdf.

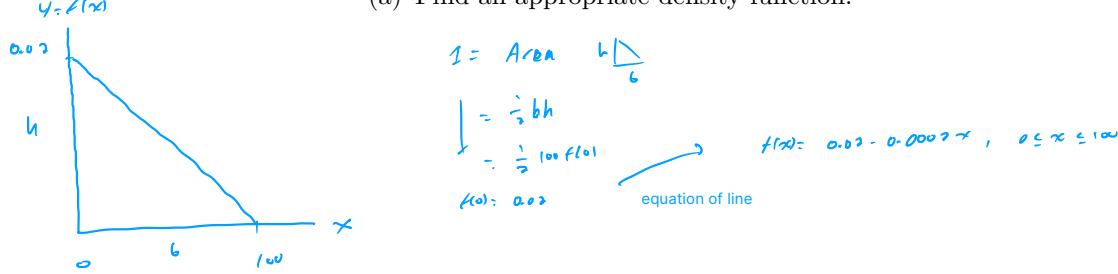


- Continuous

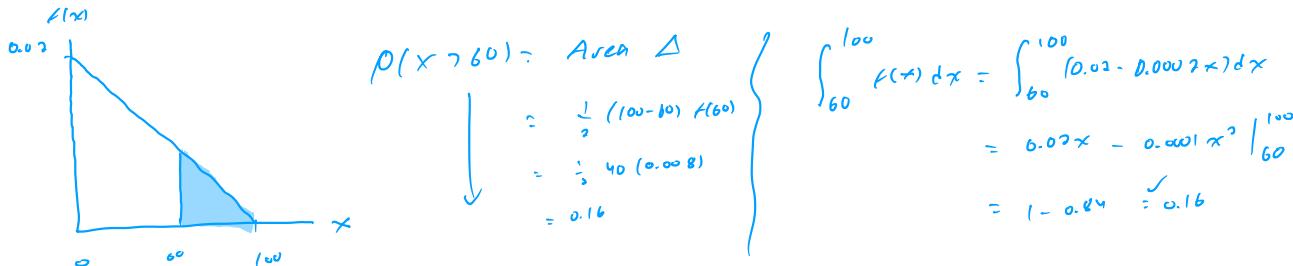
Straight-line densities

2. Suppose we are offering a warranty insurance policy which pays for repairs on a new appliance. We know from experience that repair costs X on a single policy will be on the interval $[0, 100]$, with probabilities highest for the lowest cost (\$0) and decreasing in a straight line fashion until x reaches \$100.

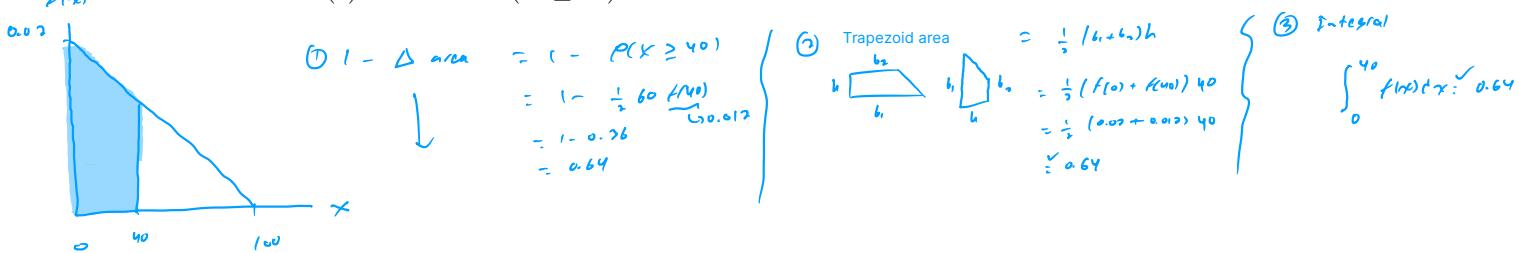
- (a) Find an appropriate density function.



- (b) Calculate $P(X > 60)$.



- (c) Calculate $P(X \leq 40)$.



Conclusion: For straight-line densities, it is usually easier to find probabilities as areas of trapezoids or triangles.

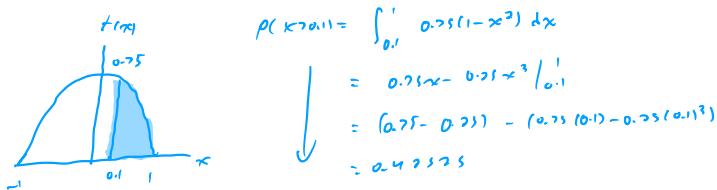
Symmetric densities

3. A risky investment has widely varying possible return percentages for next year.
 Best case: Return on investment (ROI) 100% (doubles money by getting money invested back plus 100% of the amount invested); Worst case: -100% (loses all money invested).

The percentage return is a random variable X with that can be anywhere between the worst and best case, depending on the state of the economy. The pdf is:

$$f(x) = \begin{cases} 0.75(1-x^2) & -1 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

- (a) Find the probability the investor has a ROI greater than 10%.



- (b) Find the cdf $F(x)$; then find $P(X \geq 0.1)$.

$$\begin{aligned} F(x) &= \int_{-1}^x 0.75(1-t^2) dt = 0.75t - 0.25t^3 \Big|_{-1}^x = F(x) - F(-1) \\ &= 0.75x - 0.25x^3 - (0.75(-1) - 0.25(-1)^3) \\ &= 0.25(x^3 + 3x + 2), \quad -1 \leq x \leq 1 \\ &\hookrightarrow \text{check: } \frac{F(-1) = 0}{F(1) = 1} \end{aligned}$$

$$\begin{aligned} P(X \geq 0.1) &= 1 - F(0.1) \\ &= 1 - 0.25(0.1)^3 + 3(0.1)^2 + 2 \\ &= 1 - 0.57575 \\ &\checkmark 0.42425 \end{aligned}$$

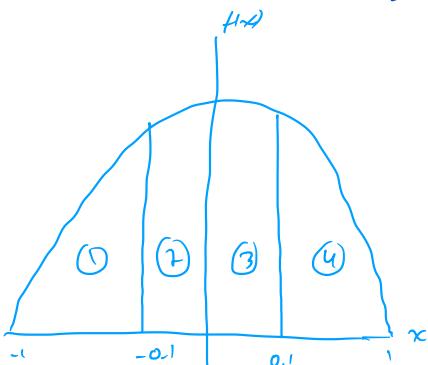
- This is actually a symmetric density, which means we have information about other probabilities as well just from finding the probability in part (b).

$$P(X \geq 0.1) = \textcircled{4} = 0.42425 = \textcircled{1}$$

$$P(0 < X < 0.1) = \textcircled{3} = 0.5 - \textcircled{4} = 0.07575 = \textcircled{2}$$

$$P(-0.1 < X < 0.1) = P(|X| < 0.1) = \textcircled{2} + \textcircled{3} = 2(0.07575) = 0.1495$$

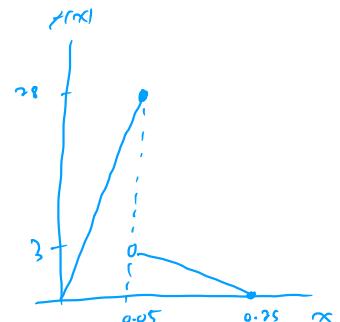
$$P(\{X < -0.1\} \cup \{X > 0.1\}) = P(|X| > 0.1) = \textcircled{2} + \textcircled{3} = 1 - P(|X| < 0.1) = 1 - \textcircled{2} - \textcircled{3} = 0.8505$$



Piecewise densities

4. The density function for a continuous random variable can be defined piecewise and fail to be continuous at some points. Let X have the following pdf:

$$f_X(x) = \begin{cases} 0 & x < 0 \\ 560x & 0 \leq x \leq 0.05 \\ -15x + 3.75 & 0.05 < x \leq 0.25 \\ 0 & x > 0.25 \end{cases}$$



- (a) Show that the total probability is 1.

$$\begin{aligned} 1 &= \int_{-\infty}^{\infty} f(x) dx \\ &= \int_0^{0.05} 560x dx + \int_{0.05}^{0.25} (-15x + 3.75) dx \\ &= \frac{1}{2}(0.05)(28) + \frac{1}{2}(3)(0.2) \\ &\stackrel{\checkmark}{=} 0.7 + 0.3 \end{aligned}$$

↗ Need separate integral for each function

↗ straight line densities ⇒ use areas of shapes

- (b) Find $P(0.03 \leq X \leq 0.07)$.

$$\begin{aligned} &= \int_{0.03}^{0.05} 560x dx + \int_{0.05}^{0.07} (-15x + 3.75) dx \\ &\quad \text{f}(0.03) = 16.8 \quad \text{f}(0.07) = 2.7 \\ &\quad \text{or } (0.7 - \Delta) + \text{or } (0.3 - \Delta) \\ &\quad (0.7 - \frac{1}{2}(0.03)(16.8)) + (0.3 - \frac{1}{2}(0.07)(2.7)) \\ &\quad = 0.448 + 0.057 \\ &\quad = 0.505 \end{aligned}$$

(c) Find the cdf $F_X(x)$.

STRATEGY: Find the cdf in cases.

$$\text{Case 1: } 0 \leq x \leq 0.05 \Rightarrow f(x) = 560x$$

$$\begin{aligned} \rightarrow P(X \leq x) &= \int_0^x 560t dt \\ &= 280t^2 \Big|_0^x \\ &= 280x^2 \end{aligned}$$

$$\text{Case 2: } 0.05 < x \leq 0.25 \Rightarrow f(x) = -15x + 3.75$$

$$\begin{aligned} \rightarrow P(X \leq x) &= F(0.05) + \int_{0.05}^x (-15t + 3.75) dt \\ &= 0.7 + \left[-7.5t^2 + 3.75t \right]_{0.05}^x \\ &\quad + \left[-7.5x^2 + 3.75x - \underbrace{(-7.5(0.05)^2 + 3.75(0.05))}_{= 0.16875} \right] \\ &= -7.5x^2 + 3.75x + 0.53125 \end{aligned}$$

$$F_X(x) = \begin{cases} 0 & x < 0 \\ 280x^2 & 0 \leq x < 0.05 \\ -7.5x^2 + 3.75x + 0.53125 & 0.05 \leq x \leq 0.25 \\ 1 & 0.25 < x \end{cases}$$

Lecture 9 – Summary Measures

MATH 320: Probability

Lecture 9: Summary Measures

Chapters 2 and 3: Distributions (2.2, 2.3, and 3.1)

Expected value

Data reduction

- When we try to interpret numerical information that has a wide range of values, we like to reduce our confusion by looking at a single number which summarizes the information.

Sample of n data points \rightarrow 1 or 2 summary measures.

Motivating example

- When the quizzes are returned, students are interested in the quiz average as well as the distribution of grades.

Lets say we have the following quiz scores: 6, 7, 8, 9

First we are going to calculate the mean like we normally would, then do some rearranging.

$$\text{Mean} = \frac{6 + 7 + 8 + 9}{4} = \frac{1}{4}(6 + 7 + 8 + 9) = 6\left(\frac{1}{4}\right) + 7\left(\frac{1}{4}\right) + 8\left(\frac{1}{4}\right) + 9\left(\frac{1}{4}\right) = 7.5$$

Written out like this, we can think of $\frac{1}{4}$ as a probability and the numbers as x 's, which are particular instances of the random variable X . Then we have a probability function.

Quiz score (x)	6	7	8	9
$f(x)$	1/4	1/4	1/4	1/4

mean = 7.5

Now what if we said that a score of 9 is more likely than the other scores. Our new pmf is:

Quiz score (x)	6	7	8	9
$f(x)$	1/6	1/6	1/6	1/2

The new mean going to increase. Let's calculate it:

$$\text{Mean} = 6\left(\frac{1}{6}\right) + 7\left(\frac{1}{6}\right) + 8\left(\frac{1}{6}\right) + 9\left(\frac{1}{2}\right) = 8 > 7$$

- What we are actually calculating here is called the expected value, which we can think of as a weighted average of the x 's where the probabilities $p(x=x)$ are the weights.

This is how we get mean (aka expected value) of a random variable from its pmf, which is usually what we are given.

Defining expected value

- Definitions

Let X be a discrete random variable. The **expected value** of X is defined by

$$E(X) = \sum_{x \in X} x f(x)$$

Let X be a continuous random variable. The **expected value** of X is defined by

$$E(X) = \int_{-\infty}^{\infty} x f(x) dx$$

The expected value of the random variable X is often denoted by the Greek letter μ .

$$E(X) = \mu$$

- Examples

- Pmf for the random variable X , the number of health insurance claims filed by a policyholder in a year, is given in the table below.

Number of claims (x)	0	1	2	3
$f(x)$	0.28	0.43	0.20	0.09

Find the expected number of claims.

$$\begin{aligned} E(X) &= \sum x f(x) \\ &= 0(0.28) + 1(0.43) + 2(0.20) + 3(0.09) \\ &= 1.1 \end{aligned}$$

- Let X be the loss severity random variable for the warranty policy with density:

$$f(x) = \begin{cases} 0.02 - 0.0002x & 0 \leq x \leq 100 \\ 0 & \text{otherwise} \end{cases}$$

Find the expected loss.

$$\begin{aligned} E(X) &= \int_{-\infty}^{\infty} x f(x) dx = \int_0^{100} x (0.02 - 0.0002x) dx \\ &= \left[0.01x^2 - \frac{0.0002}{3}x^3 \right]_0^{100} F(100) - F(0) \\ &= 0.01(100^2) - \frac{0.0002}{3}(100^3) \\ &= \frac{100}{3} \approx 33.33 \end{aligned}$$

3. Expected value of a piecewise density function example: Let X have the following pdf:

$$f_X(x) = \begin{cases} 560x & 0 \leq x \leq 0.05 \\ -15x + 3.75 & 0.05 < x \leq 0.25 \\ 0 & \text{otherwise} \end{cases}$$

Find the expected value.

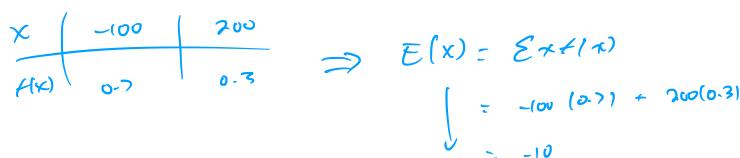
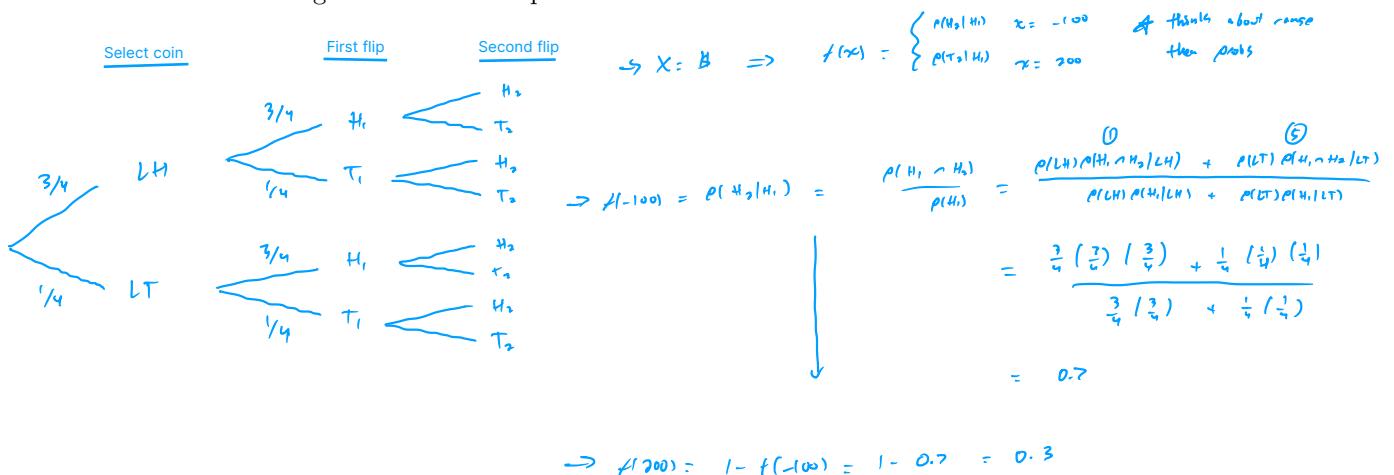
- *STRATEGY:* Integrate in pieces.

$$\begin{aligned} E(X) &= \int_{-\infty}^{\infty} x f(x) dx \\ &= \int_0^{0.05} x (560x) dx + \int_{0.05}^{0.25} x (-15x + 3.75) dx \\ &= 0.0233 + 0.035 \\ &= 0.05833 \end{aligned}$$

4. Smith is offered the following gamble: he is to choose a coin at random from a large collection of coins and toss it randomly.

- $3/4$ of the coins in the collection are loaded toward head (LH) and $1/4$ are loaded towards a tail (LT).
- If a coin is loaded towards a head, then when the coin is tossed randomly, there is a $3/4$ probability that a head will turn up and a $1/4$ probability that a tail will turn up. Similarly, if the coin is loaded towards tails, then there is a $3/4$ chance of tossing a tail on any given toss.
- If Smith tosses a head, he loses \$100 and if he tosses a tail, he wins \$200.
- Smith is allowed to obtain “sample information” about the gamble. When he chooses the coin at random, he is allowed to toss it once before deciding to accept the gamble with that same coin.

Suppose Smith tosses a head on the sample toss. Find Smith's expected gain/loss on the gamble if it is accepted.



Expected value of a function of a random variable

Motivating example

- Suppose X is a random variable, but we are actually interested in a function of the random variable $g(X)$ (which is another random variable).

One common application of this is when $g(X) = aX + b$.

- Now suppose the table from the previous Example 1 is for a type of policy which guarantees a fixed payment of \$100 dollars for each claim.

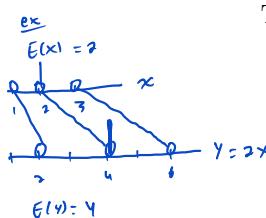
Then the amount paid to a policy holder in a year is just \$100 multiplied by the number of claims. The total claim amount is a new random variable $\text{Total_Claims} \times 100$.

We now have two random variables, X and Y , and each has their own pmf.

Number of claims (x)	0	1	2	3
Total claim amount (y)	0	100	200	300
$f(y) = f(x)$	0.28	0.43	0.20	0.04

(a) Find the expected total claim amount.

$$\begin{aligned}
 E(Y) &= 0(0.28) + 100(0.43) + 200(0.2) + 300(0.09) \\
 &= 100[0(0.28) + 1(0.43) + 2(0.2) + 3(0.09)] \\
 &\quad \downarrow \text{E}(X) \\
 &\approx 100 \underbrace{\frac{E(X)}{\approx 1.1}}_{\approx 110}
 \end{aligned}$$



Theorems

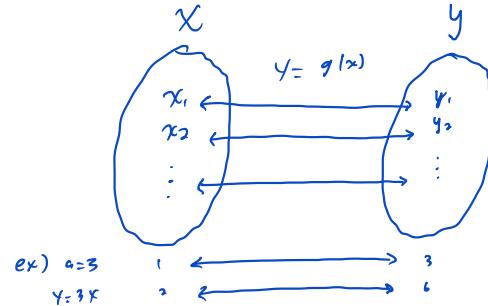
- Theorem: For any constant a and random variable X ,

$$E(aX) = a E(X)$$

We can derive this!

$$\begin{aligned}
 E_Y(Y) &= \sum y f_Y(y) \\
 &= \sum a x f_Y(ax) && \text{(substitute } Y=ax\text{)} \\
 &= \sum a x f_X(x) && \text{if } Y=ax \Rightarrow f_Y(ax) = f_X(x) \\
 &= a \sum x f_X(x) && \text{ex: } Y=3x \Leftrightarrow x=\frac{Y}{3} \\
 &= a E(X) && \text{same as}
 \end{aligned}$$

- Transformation mapping
 - If $Y = g(X)$ is a one-to-one function, then the inverse of $g(X)$ exists. So we can go “backwards” in our mapping.



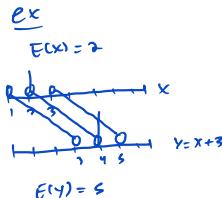
- We can also extend this rule for adding a constant. Continuing example:

- (b) Lets say the insurance company has a yearly fixed cost of \$20 per policyholder for administering the insurance policy. So the total cost in a year for a policy is the sum of the claim payments and the administrative cost.

Write our new random variable for the total cost and find the expected cost per policy per year.

x	$f(x)$	$z = 100x + 20$	$f(z)$
0	0.28	20	0.28
1	0.43	120	0.43
2	0.20	220	0.20
3	0.09	320	0.09

$$\begin{aligned} \text{Total cost} &= E(Z) = E(100X + 20) \\ &= 100E(X) + E(20) \\ &= 100(1.1) + 20 \\ &= 130 \end{aligned}$$



- Theorem: For any constants a and b and random variable X ,

$$E(aX + b) = aE(X) + b$$

Discrete derivation:

< follows same pattern as deriving $E(ax)$ >

$$E(Y) = E[Y f_Y(Y)] = E[(ax+b) \underbrace{f_Y(ax+b)}_{=f_X(x)}] = E[ax f_X(x) + b \underbrace{\sum x f(x)}_{=E(X)}] = a E[X f(X)] + b E[X] = a E(X) + b$$

$\Delta Y = ax + b$ substitute

Continuous derivation:

$$\text{If } Y = ax + b \rightarrow E(Y) = E(ax + b) = \int_{-\infty}^{\infty} (ax + b) f(x) dx$$

$$= a \int_{-\infty}^{\infty} xf(x) dx + b \int_{-\infty}^{\infty} f(x) dx$$

$$= a E(X) + b$$

- Simple example: Continuing previous example 2. Suppose that due to inflation the losses on the warranty policy are expected to increase by 5% ~~from \$100~~ plus fixed cost of \$30.

$$Y = 1.05X + 30 \rightarrow E(Y) = E(1.05X + 30) = 1.05 \underbrace{E(X)}_{\$100} + 30 = 65$$

- Theorem: For some constant random variable $X = a$,

$$E(X) = E(a) = a$$

(Discrete) Derivation:

$$f(x) = \begin{cases} 1 & x=a \\ 0 & \text{elsewhere} \end{cases} \Rightarrow E(x) = \sum x f(x)$$

$$= a \cdot 1 = a$$

Generalizing expected value of a function of a random variable

- Now suppose we are working with functions that are more complex than $Y = g(X) = aX + b$, specifically functions that are not one-to-one function. Let's see how to find the expected value when this is the case.
- Example: Let X be a discrete random variable with the pmf given below. Find the expected value of $Y = g(X) = X^2$.

x	-1	0	1
$f(x)$	0.2	0.6	0.20
$g(x) = x^2$	$(-1)^2$	0^2	1^2
sum $f(x)$			

$$\rightarrow E(g(x)) = E(x^2) = \sum x^2 f(x)$$

$$= 1(0.2) + 0(0.6) + 1(0.2)$$

$$= 0.4$$

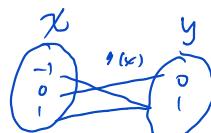
$$\Rightarrow \begin{array}{c|cc|c} y = g(x) & 0 & & 1 \\ \hline f(y) & 0.6 & & 0.4 \end{array}$$

$$\rightarrow E(Y) = \sum y f(y)$$

$$= 0(0.6) + 1(0.4)$$

$$= 0.4$$

More complex because $g(x)$ is not a one-to-one function.



- This example illustrates two major points.

- The distribution table for X can be converted into a preliminary table for $g(X)$ with entries for $g(x)$ and $f(x)$, but some regrouping may be necessary to get the actual distribution table for $Y = g(X)$.
- Even though the tables are not the same, they lead to the same results for the expected value of $Y = g(X)$.

- Theorem:

Let X be a discrete random variable. The **expected value** of $Y = g(X)$ is given by

$$E(Y) = \sum_y y f_y(y) = E(g(x)) = \sum_x g(x) f_x(x)$$

↳ simplest way

Let X be a continuous random variable. The **expected value** of $Y = g(X)$ is given by

$$E(Y) = \int_{-\infty}^{\infty} y f_y(y) dy = E(g(x)) = \int_{-\infty}^{\infty} g(x) f_x(x) dx$$

↳ only way so far

- Examples:

1. Let the random variable X have the pmf $f(x) = \frac{x}{10}$ for $x = 1, 2, 3, 4$.

Find $E(X^2)$ and $E[5 - X]$. $\rightarrow E(X^2) = 3$

$$\Rightarrow E(X^2) = \sum_{x=1,2,3,4} x^2 \left(\frac{x}{10}\right) = \sum_{x=1,2,3,4} \frac{x^3}{10} = 10$$

$$\Rightarrow E[5 - X] = E(5 - X) = 5 - E(X) = 5 - \frac{3}{4} = 4.25$$

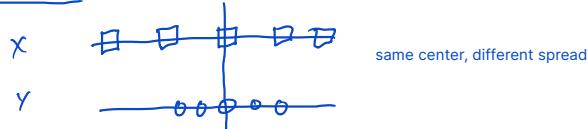
2. Let $f(x) = 3x^{-4}$ for $x > 1$. Find $E(2X^2)$.

$$E(2X^2) = \int_1^{\infty} 2x^2 (3x^{-4}) dx = \int_1^{\infty} 6x^{-2} dx = -6x^{-1} \Big|_1^{\infty} = 6$$

Variance and standard deviation

Measures of spread

- The mean of a random variable gives a nice single summary number to measure central tendency. However two different random variables can have the same mean and still be quite different.



Motivating example

- Below is the pmfs of quiz scores for two different classes.

- Find the mean of each.

Class 1: RV X

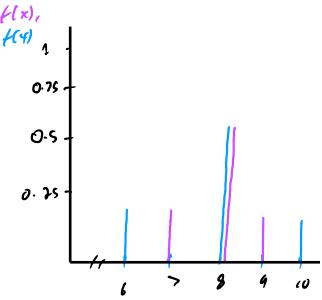
Score (x)	7	8	9
$f(x)$	0.2	0.6	0.2

Class 2: RV Y

Score (y)	6	8	10
$f(y)$	0.2	0.6	0.2

$$\mu_x = E(x) = \sum x f(x) = 8$$

$$\mu_y = E(y) = \sum y f(y) = 8$$



Means are the same, but obviously the two random variables are quite different. There is much more variation or dispersion in Y than X . The question is, how to measure that variation?

- We could look at the distance between each individual value of x or y from the mean of its distribution.

$$x - \mu_x \rightarrow E(x - \mu_x) = E(x) - \mu_x = 0$$

Always true for any random variable

This is always true for any random variable.

- However, if we look at the square of the distance between the mean, this problem does not occur. Now find the expected values of each of these new pmfs.

RV = $(X - \mu_X)^2$

$(x - 8)^2$	$(7 - 8)^2 = 1$	$(8 - 8)^2 = 0$	$(9 - 8)^2 = 1$
$f(x)$	0.2	0.6	0.2

RV = $(Y - \mu_Y)^2$

$(y - 8)^2$	$(6 - 8)^2 = 4$	$(8 - 8)^2 = 0$	$(10 - 8)^2 = 4$
$f(y)$	0.2	0.6	0.2

$$E[(x - \mu_x)^2] = E[(x - 8)^2 f(x)] = 1(0.2) + 0(0.6) + 1(0.2) = 0.4$$

$$E[(y - \mu_y)^2] = E[(y - 8)^2 f(y)] = 4(0.2) + 0(0.6) + 4(0.2) = 1.6$$

- This is the single measure of variation that is most widely used in probability theory.

Defining variance and standard deviation

- Definition: The **variance** of a random variable X is defined to be

$$V(X) = E[(X - \mu_x)^2]$$

- Definition: The **standard deviation** of a random variable is the square root of its variance. It is denoted by the greek letter σ .

$$\sigma = \sqrt{V(X)}$$

- Notes

- The variance is also written as σ^2 .
- $SD(X)$ is in the same units as X and $V(X)$ is in units².
- Variance is just a special expected value with $g(X) = (X - \mu)^2$.

- Definitions:

Discrete

$$V(X) = \sum_{x} (x - \mu_x)^2 f(x)$$

Continuous

$$V(X) = \int_{-\infty}^{\infty} (x - \mu_x)^2 f(x) dx$$

- Examples:

- Let X be the random variable with $f(x) = \frac{3}{8}x^2$ for $0 < x < 2$. Find σ_X^2 .

$$\textcircled{1} \quad E(X) : \int_{-\infty}^{\infty} x f(x) dx = \int_0^2 x \left(\frac{3}{8}x^2\right) dx \\ \downarrow \\ = \frac{3}{32} x^4 \Big|_0^2 [F(2) - F(0)] \\ \approx 1.5$$

$$\textcircled{2} \quad V(X) = E[(X - \mu_x)^2] \stackrel{\mu_x = 1.5}{=} \int_{-\infty}^{\infty} (x - 1.5)^2 f(x) dx \\ \downarrow \\ = \int_0^2 (x - 1.5)^2 \frac{3}{8}x^2 dx \\ = \frac{3}{8} \int_0^2 (x^3 - 3x^2 + \frac{9}{4})x^2 dx \\ = \frac{3}{8} \int_0^2 (x^5 - 3x^4 + \frac{9}{4}x^3) dx \\ = \frac{3}{8} \left(\frac{1}{5}x^5 - \frac{3}{4}x^4 + \frac{9}{16}x^4 \right) \Big|_0^2 \\ = 0.15$$

2. Let $f(x) = \frac{3x+4}{22}$ for $x = -1, 0, 1, 2$.

Find the mean, variance and standard deviation of X .

$$\textcircled{1} \quad E(X) = \sum_{\substack{x=-1,0,1,2}} x f(x) = \sum_{\substack{x=-1,0,1,2}} x \left(\frac{3x+4}{22} \right) = \frac{1}{22} \sum_{\substack{x=-1,0,1,2}} (3x^2 + 4x) = \frac{1}{22} \left[(3(-1)^2 + 4(-1)) + \dots + (3(2)^2 + 4(2)) \right] = \frac{13}{11}$$

$$\textcircled{2} \quad V(X) = \sum_{\substack{x=-1,0,1,2}} (x - E(X))^2 f(x) = \sum_{\substack{x=-1,0,1,2}} (x - \frac{13}{11})^2 \left(\frac{3x+4}{22} \right) = \frac{1}{22} \left[((-1 - \frac{13}{11})^2 (3(-1) + 4)) + \dots + ((2 - \frac{13}{11})^2 (3(2) + 4)) \right] = \frac{95}{121}$$

$$\textcircled{3} \quad SD(X) = \sqrt{V(X)} = \sqrt{95/121}$$

Expectation as a linear operator

- Now let's say we want to find the expected value of $g(X) = (X - 3)^2 = X^2 - 6X + 9$.
- This is a more complicated function of X because it has multiple X 's. To find the expected value of $g(X)$ in this case, we can do what intuitively makes sense.

$$E(g(X)) = E((X-3)^2) = E(X^2 - 6X + 9) = E(X^2) - 6E(X) + 9$$

$\underbrace{\qquad}_{\text{can algebra within } E(\cdot)}$ $\underbrace{\qquad}_{\text{Separate & simplify}}$

- Theorem: The reason this works is because of the following property of expectation:

$$E \left[\sum_{i=1}^k c_i g_i(X) \right] = \sum_{i=1}^k c_i E[g_i(X)]$$

In fancy words, the expected value of a linear combination equals a linear combination of expected values.

- Because of the property, expectation is often called a linear (or distributive) operator.

Integration and derivation are linear operators as well:

$$\int [g(x) + f(x)] dx = \int g(x) dx + \int f(x) dx$$

$$\frac{d}{dx} [g(x) + f(x)] = \frac{d}{dx} g(x) + \frac{d}{dx} f(x)$$

Another way to calculate variance

- Using the property we just showed for expectation, we can arrive at another way to calculate the variance of a random variable.

$$\begin{aligned}
 V(X) &= E[(X-\mu)^2] && \text{(definition)} \\
 &= E[X^2 - 2\mu X + \mu^2] && \text{(expand)} \\
 &= E[X^2] - 2\mu \underbrace{E[X]}_{\mu} + \mu^2 && \text{(distribute & simplify)} \\
 &= E[X^2] - \mu^2
 \end{aligned}$$

Theorem:

- Another way to calculate variance:

$$V(X) = E[X^2] - \mu^2$$

mean

- Examples:

- Below is a pmf table for X .

x	0	1	2	3
$f(x)$	0.5	0.3	0.06	0.14

Find $V(X)$ using the new formula.

$$\textcircled{1} \quad E(X) = \sum_x x f(x) = 0(0.5) + \dots + 3(0.14) = 0.84$$

$$\textcircled{2} \quad E(X^2) = \sum_x x^2 f(x) = 0^2(0.5) + \dots + 2^2(0.14) = 1.8$$

$$\textcircled{3} \quad V(X) = E(X^2) - (E(X))^2 = 1.8 - 0.84^2 = 1.0844$$

- Find the variance of the warranty loss random variable using the alternate formula.

$$f(x) = \begin{cases} 0.02 - 0.0002x & 0 \leq x \leq 100 \\ 0 & \text{otherwise} \end{cases}$$

$$\textcircled{1} \quad E(X) = \frac{100}{3}$$

$$\begin{aligned}
 \textcircled{2} \quad E(X^2) &= \int_{-\infty}^{\infty} x^2 f(x) dx = \int_0^{100} x^2 (0.02 - 0.0002x) dx \\
 &= \left[\frac{0.02}{3} x^3 - \frac{0.0002}{4} x^4 \right]_0^{100} F(100) - F(0) \\
 &= 666.67
 \end{aligned}$$

$$\textcircled{3} \quad V(X) = E(X^2) - (E(X))^2 = (666.67 - \frac{100}{3})^2 = 555.56$$

Calculating the variance from the definition would require evaluation of the integral:

$$V(X) = \int_0^{100} \left(x - \frac{100}{3} \right)^2 (0.02 - 0.0002x) dx$$

This would be straightforward, but time consuming relative to the other way if done by hand. Of course with computing, calculation time is not an issue.

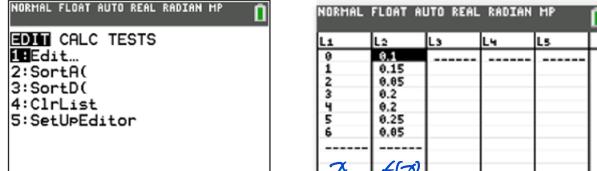
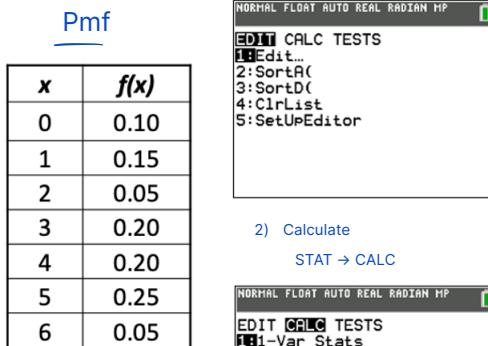
-  • Important practical note: The calculation of variance can be done more easily by hand using this alternate form rather than the definition.

But using the definition is more efficient for computers when large values of X are present (problems with overflow due to the magnitude of X^2).

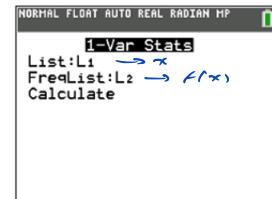
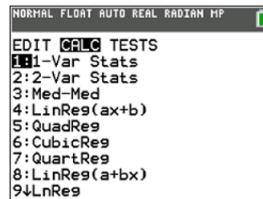
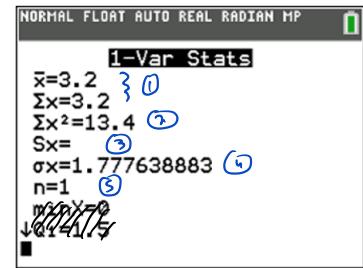
Technology session

- Using TI-84 (and TI-30XS MultiView) to calculate $E(X)$ and $SD(X)$.

- 1) Enter data
STAT



Results



① $\bar{x} = E(x) = \sum x f(x)$
" $E(x)$ "

② $E(x^2) = E(x^2)$

③ $S_x = \sqrt{\sum x^2 - \bar{x}^2}$

If we typed in probabilities correctly

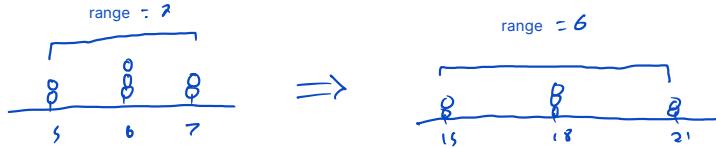
④ $S_x = \text{BLANK or Error}$
↳ data standard deviation

⑤ $\sigma x = \sqrt{\sum x^2 - \bar{x}^2}$
↳ Check ✓

Variance and standard deviation of a function of a random variable

Variance and standard deviation of $Y = aX + b$

- Previously, we saw how multiplying by a coefficient a and adding a constant b affected the expected value of our new random variable $Y = aX + b$. Now let's study the affect they have on the variance and standard deviation.
- First we will look at only the affect of the coefficient.



Derivation of $V(aX)$:

$$\begin{aligned} \text{If } Y = aX \rightarrow V(Y) &= E[(Y - \mu_Y)^2] \\ \hookrightarrow \mu_Y &= a\mu_X \quad \left| \begin{array}{l} = E[(aX - a\mu_X)^2] \\ = E[a^2(X - \mu_X)^2] \\ = a^2 E[(X - \mu)^2] \\ \leq a^2 V(X) \end{array} \right. \end{aligned}$$

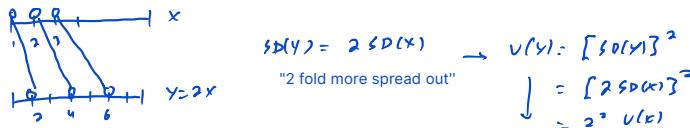
- Theorem: For any constant a and random variable X , (alternate notation)

$$V(aX) = a^2 V(X) \quad \sigma_{aX}^2 = a^2 \sigma_X^2$$

The standard deviation of aX can now be obtained by taking the square root.

$$SD(aX) = \sqrt{a^2 V(X)} = |a| SD(X) \quad \sigma_{aX} = |a| \sigma_X$$

- Intuitively, here's why the a is squared for $V(aX)$:



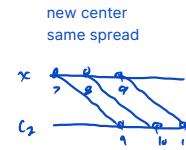
- Example: Using the previous quiz scores from class 1, let's investigate two different types of curves the instructor could use and their impact on the expected value and variance.

– Curve 1: Each individual score is raised by 25%. $C_1 = 1.25 \times$

$$\begin{aligned} E(C_1) &= E(1.25X) \quad \left| \begin{array}{l} V(C_1) = V(1.25X) \\ = 1.25 E(X) \quad \downarrow \\ = 10 \end{array} \right. \\ &\quad \left| \begin{array}{l} \downarrow \\ = 1.25^2 V(X) \\ = 0.625 \end{array} \right. \end{aligned}$$

- Curve 2: Every score gets an additional 2 points. $c_2 = x + 2$

Original score (x)	7	8	9
Curved score (c_2)	9	10	11
$f(c_2) = f(x)$	0.2	0.6	0.2



using calculator (showing work) $\cancel{\text{not}}$

$$\rightarrow \text{Entered data} \rightarrow l_1 = c_2 \\ l_2 = f(c_2)$$

$$\rightarrow \text{TI 84} \rightarrow \text{1 var stats } \{ \text{list} = l_1, \text{freqlist} = l_2 \} \rightarrow (\sigma_{c_2} = 0.632)^2 \Rightarrow V(c_2) \approx 0.4 \\ \text{TI 30xs} \rightarrow \text{1 var stats } \{ \text{data} = l_1, \text{freq} = l_2 \} \downarrow = V(x)$$

- This example shows the intuitive idea that if all values are shifted by exactly b units, the mean changes but the dispersion around the new mean is exactly the same as before.

- Theorem: For any constants a and b and random variable X ,

$$V(aX + b) = a^2 V(X)$$

- Example: Let's say your Test 1 grades X had the pmf given below (grades were out of 100 points).

The test corrections gave you $1/2$ of the missed points back. Find the expected value and standard deviation of the grades after test corrections. Compare this to the original expected value and standard deviation.

- Define a new random variable $Y = \text{grade after test corrections}$.

Grade (x)	$f(x)$	$y = 0.5x + 50$	$f(y) = f(x)$
60	0.02	80	.02
65	0.04	82.5	.04
70	0.17	85	.17
75	0.23	87.5	.23
80	0.21	90	.21
85	0.15	92.5	.15
90	0.12	95	.12
95	0.05	97.5	.05
100	0.01	100	.01

$$\begin{aligned} Y &= x + (100-x)0.5 \\ &= x + 50 - 0.5x \\ &= 0.5x + 50 \end{aligned}$$

(b) Calculate $E(Y)$ and $SD(Y)$.

Short way $\rightarrow E(Y) + V(Y)$ Indirectly

① Find $E(X) + V(X)$ first.

Entered data $\rightarrow L_1 = x, L_2 = f(x)$
2-var stats ($L_1 = L_1, \text{Freq}L_1 = L_2$) (TI 84)

$$\Rightarrow E(X) = 79.05$$

\bar{x} in calc

$$S^2(X) = 8.357 \Rightarrow V(X) = 8.357^2$$

σ_x in calc

$$\downarrow = 69.034$$

② Use formulas to find $E(Y) + V(Y) \Rightarrow SD(Y)$

$$\rightarrow E(Y) = E(0.5x + 50) = 0.5 \underbrace{E(X)}_{\bar{x}=79.05} + 50 = 89.525$$

$$\rightarrow V(Y) = V(0.5x + 50) = 0.5^2 V(x) = 0.5^2 \cdot 69.034 \Rightarrow SD(Y) = \sqrt{V(Y)}$$

$$\downarrow = 4.179$$

Long way $\sim E(Y) + SD(Y)$ directly

① finding raw pmf of Y

② Then find $E(Y) + SD(Y)$

→ Entered data $\rightarrow L_1 = x, L_2 = f(x), L_3 = 0.5L_1 + 50$ (TI 84)

→ 2-var stats (Data: L_1 , Freq: L_2) (TI 84)

$$\Rightarrow \boxed{E(Y) = 89.525}$$

$\frac{\text{check}}{n=1}$

$$\boxed{SD(Y) = 4.179}$$

\downarrow

Final comparison: Test corrections have higher mean and less variability.

- Note that we have learned and practiced how to find the expected value of non-linear functions of X such as $Y = X^2 - 7$, but not necessarily for variance. Here's why:

$$\rightarrow E(Y) = E(X^2 - 7) = E(X^2) - 7 = \left\{ \int_{-\infty}^{\infty} x^2 f(x) dx \right\} - 7 \quad \text{fairly easy}$$

$$\rightarrow V(Y) = V(X^2 - 7) = V(X^2)$$

→ options → 1) using definition: $V(X^2) = E[(X^2 - E(X^2))^2]$

$$\downarrow = \int_{-\infty}^{\infty} (x^2 - E(x^2))^2 f(x) dx$$

2) using alternate formula

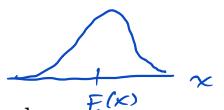
$$V(X^2) = E[X^2] - [E(X^2)]^2$$

$$\downarrow = E(X^4) - [E(X^2)]^2$$

possible, just more involved

3) Find $f_Y(y)$ & find $V(Y)$ like normal \rightarrow Haven't learned this yet!

Evaluating expected value (as a predictor of X)



- Expected value is a measure of center (i.e. where a distribution is located).

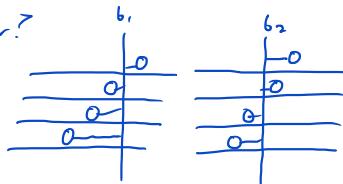
In other words, we can interpret $E(X)$ as a good guess at a value of X . Here's why this interpretation makes sense:

- Suppose that we measure the distance between a random variable X and a constant b by $(X - b)^2$. The closer b is to X , the smaller this quantity is.

$(X - b)^2$ is a random variable, which is not a good measure because X changes every time.

$E[(X - b)^2]$ is constant, which makes it a good measure because it never changes.

b_1 , or b_2 "closer" to center?
 \Rightarrow minimize collective distance to every point?



Summary Measures

9-16

- Find the value of b that minimizes $E[(X - b)^2]$ and, hence will provide us with a good predictor of X (i.e. the optimal value of b).

Steps:

(a) Rearrange $g(b) = E[(X - b)^2] = E(x^2 - 2xb + b^2)$

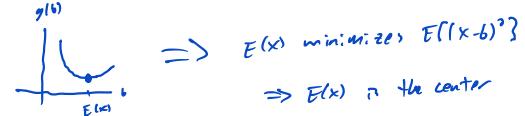
$$\downarrow \quad = E(x^2) - 2E(x)b + b^2$$

- (b) Goal: Find b to minimize step result of step (a).

$$\begin{aligned} \Rightarrow g'(b) &= \frac{d}{db} [E(x^2) - 2E(x)b + b^2] \\ \downarrow &= -2E(x) + 2b \\ \Rightarrow 0 &= -2E(x) + 2b \\ \Rightarrow b &= E(x) \end{aligned}$$

- (c) Confirm minimum with second derivative test:

$$\begin{aligned} g''(b) &= \frac{d}{db} [-2E(x) + 2b] \\ \downarrow &= 2 > 0 \Rightarrow \text{minimum } \checkmark \end{aligned}$$



Other measures of center

Introduction

- The mean of a random variable is the most widely used single measure of central tendency.

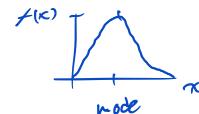
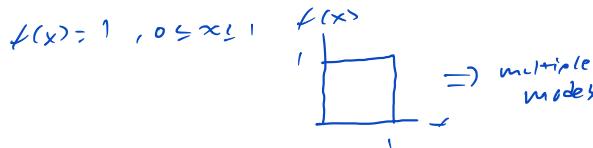
There are other measures which are also informative. Each of these has their own interpretation, advantages and shortcomings.

Mode

- Definition: The **mode** is the x value(s) which maximizes the distribution function $f(x)$.

For discrete random variables, the mode is the x with the highest probability (we can think of this as the most likely); there can be multiple modes.

For continuous random variables it is the x value where $f(x)$ is the highest.



Median

- The **median** m of a continuous random variable X is the solution of the equation

$$F(m) = P(X \leq m) = 0.5$$

Thus, we can think of the median m as the “equal areas point”, which means the x value dividing the lower 50% probability and the upper 50%.

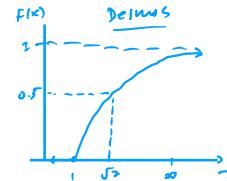
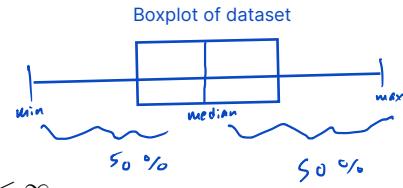
- Examples

- Let X have the following cdf:

$$F(x) = P(X \leq x) = 1 - \frac{1}{x^2}, \quad 1 \leq x < \infty$$

Find the median m .

$$\begin{aligned} 0.5 &= F(m) \\ \downarrow &= 1 - \frac{1}{m^2} \\ 0.5 &= \frac{1}{m^2} \\ m^2 &= 2 \\ m &= \pm\sqrt{2} \rightarrow \text{in range } 1 < x < \infty \\ m &= \sqrt{2} \end{aligned}$$



- The straight-line density example about losses on a warranty insurance policy had the following pdf and cdf (which we now know how to find):

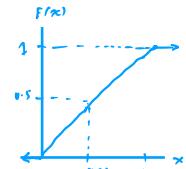
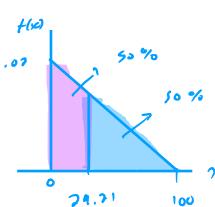
$$f(x) = \begin{cases} 0.02 - 0.0002x & 0 \leq x \leq 100 \\ 0 & \text{otherwise} \end{cases}$$

$$F(x) = \int_0^x 0.02 - 0.0002t dt = 0.02x - 0.0001x^2, \quad 0 \leq x \leq 100$$

Find the median m .

$$\begin{aligned} \rightarrow 0.5 &= F(m) \\ \downarrow &= 0.02m - 0.0001m^2 \\ 0 &= \underbrace{-0.0001m^2}_{a} + \underbrace{0.02m}_{b} - \underbrace{0.5}_{c} \\ \rightarrow m &= \frac{-0.02 \pm \sqrt{0.02^2 - 4(-0.0001)(-0.5)}}{2(-0.0001)} \\ \Rightarrow m &= 29.21 \quad \text{or} \quad 70.71 \\ m &\in X \quad m \notin X \end{aligned}$$

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$



This has a nice intuitive interpretation. Half of the losses will be less than $\$79.71$ and the other half will be greater.

3. Symmetric densities: If the density function is symmetric, the median can be found without calculation.

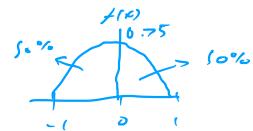
* Result: For symmetric densities, the median m is equal to the point of symmetry.

Example: Find the expected value and the median for the ROI example, where

$$f(x) = \begin{cases} 0.75(1-x^2) & -1 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

$$\begin{aligned} E(x) &= \int_{-\infty}^{\infty} x f(x) dx = \int_{-1}^1 x (0.75(1-x^2)) dx \\ &= \frac{3}{8}x^2 - \frac{3}{16}x^4 \Big|_{-1}^1 F(1) - F(-1) \\ &= \left(\frac{3}{8} - \frac{3}{16}\right) - \left(\frac{3}{8} - \frac{3}{16}\right) \\ &= 0 \end{aligned}$$

- median
① find cdf $F(x)$
② set $0.5 = F(m)$
③ solve for m



Result: In general, the median and mean are not equivalent.

But when the density function is symmetric, these two measures of center are equal.

Percentiles

- For the previous example we worked out about warranty losses, the median could be interpreted as separating the top 50% of losses from the bottom 50% of losses. For this reason, the median is called the 50th percentile.

Other percentiles can be defined using similar reasoning. For example, the 90th percentile separates the top 10% and the bottom 90%. Below is a general definition of percentile.

- Definition: Let X be a continuous random variable and $0 \leq p \leq 1$. The **100pth percentile** of X is the number x_p defined by

$$F(x_p) = p$$

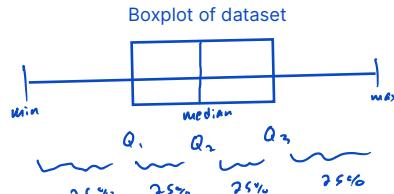
- Special percentiles:

$x_{0.25} = 25^{\text{th}}$ percentile → First, Q_1

$x_{0.50} = 50^{\text{th}}$ percentile → Second, $Q_2 = \text{median}$

$x_{0.75} = 75^{\text{th}}$ percentile → Third, Q_3

Quartiles (split area under density curve into quarters)



- Another measure of spread (variation):

- Generally in probability theory, we only talk about standard deviation and variance.
- In practice (and elementary statistics courses), another measure is often used: inter quartile range (IQR).

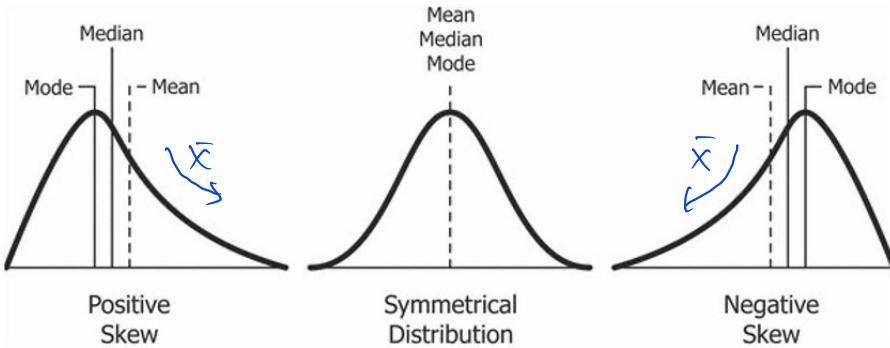
- Definition: The **inter quartile range (IQR)** is equal to the difference of the third and first quartiles:

$$IQR = Q_3 - Q_1$$

In probability theory, the IQR is measuring the range of the middle 50% of probability. For datasets, it measures how far data is spread out around the median.

- In studies, The IQR is typically reported in favor of the standard deviation when the distribution of data is skewed.

This is because the SD gets inflated due to skewness and outliers, whereas the IQR does not (i.e. it is a **resistant measure**).



- Examples:

- Find the IQR of X using the following cdf:

$$F(x) = P(X \leq x) = 1 - \frac{1}{x^2}, \quad 1 \leq x < \infty.$$

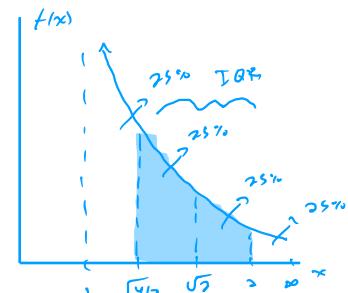
Q₁

$$\begin{aligned} 0.25 &= \frac{1}{4} = F(x_{0.25}) \\ \downarrow &= 1 - \frac{1}{x^2} \\ \frac{1}{4} &= x^2 \\ x_{0.25} &= \pm \sqrt{4/3} \end{aligned}$$

Q₃

$$\begin{aligned} 0.75 &= \frac{3}{4} = F(x_{0.75}) \\ \downarrow &= 1 - \frac{1}{x^2} \\ 4 &= x^2 \\ x_{0.75} &= 2 \end{aligned}$$

$$\begin{aligned} \rightarrow IQR &= Q_3 - Q_1 = x_{0.75} - x_{0.25} \\ &= 2 - \sqrt{4/3} \\ &\approx 0.85 \end{aligned}$$



5. The time X in months until failure of a certain product has the pdf

$$f(x) = (1/2)x e^{-(x/2)^2} \quad \text{for } 0 < x < \infty.$$

Find the first and second quartiles of X .

$$\begin{aligned} \textcircled{1} \quad F(x) &= \int_0^x \frac{1}{2}t e^{-\left(\frac{t}{2}\right)^2} dt \\ &= \int_0^x \frac{1}{2}t e^{-\frac{t^2}{4}} dt \quad \rightarrow \text{let } u = \frac{t^2}{4} \Rightarrow du = \frac{1}{2}t dt \\ &= \int_0^{x^2/4} e^{-u} du \\ &= -e^{-u} \Big|_0^{x^2/4} \\ &= e^{-x^2/4} - (-e^0) \\ &= 1 - e^{-x^2/4}, \quad 0 < x < \infty \end{aligned}$$

$$\begin{aligned} \textcircled{2} \quad Q_1 \rightarrow 0.25 &= F(x_{0.25}) & Q_2 \rightarrow 0.5 &= F(m) \\ \downarrow &= 1 - e^{-x^2/4} & \downarrow &= 1 - e^{-m^2/4} \\ \ln(e^{-x^2/4}) &= \ln(0.25) & \vdots & \vdots \\ -x^2/4 &= \ln(3/4) & m &= \sqrt{-4 \ln(0.5)} \approx 1.6651 \\ x &= \sqrt{-4 \ln(3/4)} \approx 1.0727 \end{aligned}$$

Test 3

Contents

Lectures 10 – Discrete Distributions	92
Lectures 11 – Continuous Distributions	127
Lectures 12 – Moment Generating Functions	161

Lectures 10 – Discrete Distributions

MATH 320: Probability

Lecture 10: Discrete Distributions

Chapters 2: Distributions (2.1, 2.3, 2.4, 2.5, 2.6)
2.7

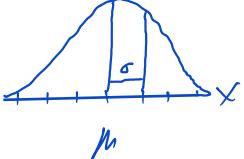
Introduction

- Statistical distributions are used to model populations.

We usually deal with a **family** of distributions rather than a single distribution (family = type of distribution).

- This family is indexed by one or more parameters, which allows us to vary certain characteristics of the distribution while staying with one functional form.

The functional form determines the unique features of the distribution.



- For example, if the distribution of a population is symmetric and bell-shaped, then a **Normal** distribution is a reasonable choice. Then we will specify the parameters μ & σ^2 .

- In the previous sections, we saw a number of examples of discrete probability distributions.

Recall a random variable X is said to have a discrete distribution if X , range of X is countable.

- In the next sections we will study some special distributions that are extremely useful and widely applied, some of which we have already seen before.

Discrete uniform distribution

Definition

- Scenario: If a finite number of values are equally likely to be observed, then a discrete uniform distribution is used to model.

- Definition: A random variable X has a **discrete uniform** (N_0, N_1) distribution if

$$P(X = x | N_0, N_1) = \frac{1}{N_1 - N_0 + 1} \quad x = N_0, \dots, N_1,$$

of numbers included in X

where N_0 and N_1 are specified integers ($N_0 \leq N_1$).

- If X follows a discrete uniform (DU) distribution with parameters N_0 and N_1 , we can summarize this using the following notation:

 $\text{RV} \sim \text{Distribution}(\text{parameters}) \quad X \sim \text{Discrete uniform}(N_0, N_1)$

“The random variable X **follows** (is distributed as / assumed to have) a discrete uniform distribution with parameters N_0 and N_1 ”.

Can also use this more generally with $X \sim F_X(x)$ or $X \sim f_X(x)$.

- Example: Rolling a fair 6-sided die.

$$P(X = x | N_0, N_1) = \frac{1}{N_1 - N_0 + 1} = \frac{1}{6-1+1} = \frac{1}{6}, \quad x = 1, 2, \dots, 6$$

$f(x|N_0, N_1)$

- A note on notation: When we are dealing with parametric distributions, the distribution is dependent on the values of the parameters. In order to emphasize this fact and keep track of the parameters, write them in the pmf preceded by: | (given).

In general, we have $P(X = x | \theta)$, where θ could be a vector of parameters.

 ex) $DU \rightarrow \theta = (N_0, N_1)$

Mean and variance

- One of the advantages of having families of distributions is that the pmfs have the same functional form (i.e. follow a certain pattern). This is also true for their expected value and variances.

- If a random variable X has a discrete uniform distribution (N_0, N_1) ,

$$E(X) = \frac{N_0 + N_1}{2} \quad V(X) = \frac{(N_1 - N_0 + 1)^2 - 1}{12} \Rightarrow SD(X) = \sqrt{V(X)}$$

- The mean and variance are functions of the parameters.

- Continuing example: Find the mean and variance of rolling a fair 6-sided die.

Use formulas:

$$E(X) = \frac{N_0 + N_1}{2} = \frac{1+6}{2} = 3.5$$

$$V(X) = \frac{(N_1 - N_0 + 1)^2 - 1}{12} = \frac{(6-1+1)^2 - 1}{12} = \frac{35}{12}$$

Confirm with definition:

$$E(X) = \sum x f(x) = \frac{1}{6} \sum_{x=1}^6 x = 3.5$$

Summary

- If we know (or assume) the family of a distribution and the values of the parameters
 \Leftrightarrow know pmf \Leftrightarrow cdf

We also know the expected value and variance as well.

- This statement works backwards too.
- **IMPORTANT STRATEGY** when solving problems:

A A A If we recognize that we have a pmf where the range of the random variable and the probabilities match the scenario of a specific distribution, then that random variable must follow that specific distribution.

- Examples: How can we model the following populations?

- Suppose $f_X(x) = \frac{1}{4}$, $x = 7, 8, 9, 10$

\hookrightarrow equally likely \hookrightarrow Discrete range in a sequence

$$\Rightarrow Y \sim DV(N_0 = 7, N_1 = 11)$$

$$\text{Find } E(X^2). \quad V(X) = E(X^2) - (E(X))^2 \Rightarrow E(X^2) = \frac{V(X)}{N_0} + \frac{(E(X))^2}{N_1}$$

$$= \frac{(10 - 7)^2 - 1}{12} + \left(\frac{7+10}{2}\right)^2$$

$$= \frac{16}{12} + (8.5)^2$$

- Suppose the values 3, 6, 9, 12, 15 are equally likely.

$$X \sim DV(N_0 = 1, N_1 = 5) \rightarrow \text{transform } Y = 3X$$

- Suppose the values 5, 9, 13, 17 are equally likely.

$$X \sim DV(N_0 = 1, N_1 = 4) \rightarrow \text{transform } Y = 4X + 1$$

Bernoulli distribution

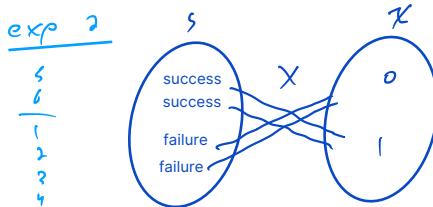
Motivation

- A**
- The following four distributions are constructed based on Bernoulli experiments: Binomial, Geometric, Hypergeometric, Negative Binomial.
 - Example experiments:
 - 1. Produce a product and see if the product is defective. $\begin{matrix} \text{Success} \\ \text{defective} \end{matrix}$
 - 2. Roll a die and see if a number greater than 4 appears. $\# > 4$
 - 3. Choose a student in this class at random and see if a female student is chosen. $\begin{matrix} \text{Female} \\ \approx \text{True / false} \end{matrix}$
 - What is the common characteristic of these experiments? $\approx \text{True / false}$
 - The main event of interest is labeled Success and the other is labeled failure.

Definition

- Scenario: The random variable for which 0 and 1 are chosen to describe the two possible values is called a **Bernoulli random variable**.
- The outcomes in Success and Failure are assigned 1 and 0 by the Bernoulli random variable, respectively.

$$\mathcal{X} = \{0, 1\}$$



- Definition: A random variable X has a **Bernoulli** (p) distribution if

$$P(X = x | p) = \begin{cases} 1 - p = q & x = 0 \\ p & x = 1 \\ 0 & \text{otherwise} \end{cases}$$

where the parameter p represents the probability of success and $0 < p < 1$.

Equivalently, $f(x | p) = p^x (1 - p)^{1-x}$, $x = 0, 1$.

- Notation: $X \sim \text{Bernoulli}(p)$
(Ber)
- Example: A basketball player shoots a free throw with a 75% probability of success. Let X denote the number of points scored.
 $\hookrightarrow \mathcal{X} = \{0, 1\}$

$$X \sim \text{Ber}(p=0.75)$$

- (b) Find the pmf of X .

$$f_X(x | p=0.75) = 0.75^x (1-0.75)^{1-x}, x=0,1$$

- (c) Find the expected value of X (by hand using the definition).

$$\begin{aligned} E(X) &= \sum_{x=0}^1 x f(x) \\ &= 0(0.75) + 1(0.75) \\ &= 0.75 \\ &\boxed{= p} \end{aligned}$$

- (d) Find the variance of X (by hand using the definition).

$$\begin{aligned} V(X) &= E[(X - \mu_x)^2] = \sum_{x=0}^1 (x - 0.75)^2 f(x) \\ &= (0 - 0.75)^2 (0.75) + (1 - 0.75)^2 (0.75) \\ &= 0.1875 \\ &= 0.75(0.25) \\ &\boxed{= pq} \end{aligned}$$

Mean and variance

- If $X \sim \text{Bernoulli}(p)$

$$E(X) = p$$

$$V(X) = p(1-p) = pq \Rightarrow SD(X) = \sqrt{pq}$$

- Derive $E(X)$. HINT: Generalize the calculations from the previous examples.

$$\begin{aligned} E(X) &= \sum_{x=0}^1 x f(x) \\ &= 0(1-p) + 1(p) \\ &\stackrel{\checkmark}{=} p \end{aligned}$$

- Derive $V(X)$.

$$\begin{aligned} V(X) &= E[(X - \mu_X)^2] \\ &= \sum_{x=0}^1 (x - p)^2 f(x) \\ &= (0 - p)^2 q + (1 - p)^2 p \\ &= p^2 q + (1-p)^2 p \\ &= pq \underbrace{[p + (1-p)]}_{\geq 1} \\ &\stackrel{\checkmark}{=} pq \end{aligned}$$

** We can always try to derive the mean and variance of distributions by going back to the definitions.

Binomial distribution

Motivation

- Example experiments:

1. A basketball player gets 10 3-point shots and has a probability of success (making a basket) of 1/5. Let X be the number of makes out of the 10 attempts.
2. A factory produces 100 products with the probability of defect 2/100. Let Y be the number of defective products out of 100 products.
3. Roll a fair die 3 times. Let Z be the number of 6s out of the 3 rolls.

- What is the common characteristic of these experiments?

success / failure multiple times

Binomial experiments and binomial random variables

- Definition: An experiment is called a **binomial experiment** if all of the following hold:

These are the conditions needed in order to have a binomial experiment.

1. The experiment consists of a fixed number, n , of identical trials.
(Note that **identical** = from the same family with the same parameter values.)
2. Each trial results in one of two events: success or failure.
3. There is a constant probability of success, p , for each trial.
4. The trials are independent.
5. The outcome is a sequence of successes and failures. e.g. $\{0, 1, 1, 0, 1\}$

Said another way, a binomial experiment is a sequence of n identical Bernoulli trials.

- Definition: If X is the number of successes in a binomial experiment, X is called a **binomial random variable.**

- Examples:

- (a) Check the conditions to determine if example experiment 1 is indeed a binomial experiment.

1. $n=10$ trials (shots) ✓
2. Each trial is success (make) or failure (miss) ✓
3. constant probability of success ($p=0.2$) ✓
4. Each trial is independent ✓
5. outcome is a sequence of makes or misses ✓ $S F S F F S \dots$
 | o

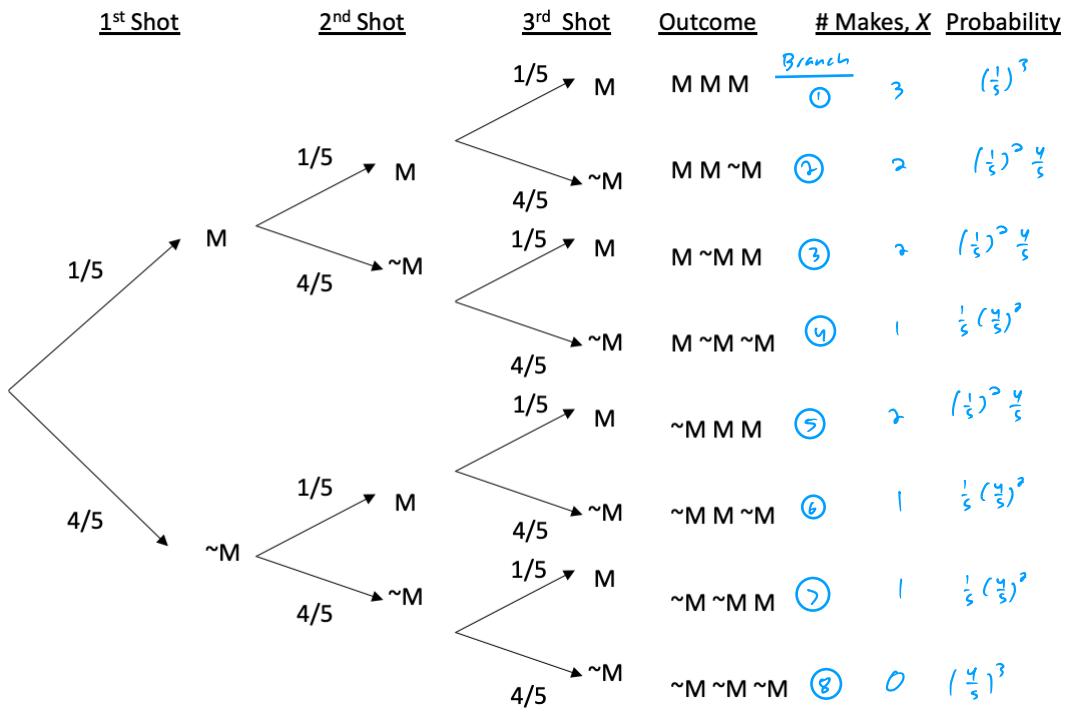
$$\Rightarrow X \sim \text{Binomial}(n=10, p=0.2)$$

- (b) Are the conditions met for experiments 2 and 3? If so, identify n and p .

$$\begin{aligned} \text{Exp 2, } Y & \quad \left. \right\} \rightarrow \text{All conditions are met} \quad \checkmark \\ \text{Exp 3, } Z & \\ Y & \sim \text{Bin}(n=100, p=2/100) \\ Z & \sim \text{Bin}(n=3, p=1/6) \end{aligned}$$

Binomial probabilities

- Probabilities for binomial distribution are just an application of the multiplication rule for independent events.
- Example: Let's say our basketball player has only 3 shot attempts now. Let's visualize the tree diagram illustrating the different outcomes of this binomial experiment ($M = \text{make}$, $\sim M = \text{miss}$).



(a) Find $P(X = 3)$. $= \textcircled{1} (\frac{1}{5})^3$

\downarrow $=$ (1 branch) (probability of branch)

(b) Find $P(X = 2)$. $= \frac{\textcircled{2} + \textcircled{3} + \textcircled{5}}{3} [(\frac{1}{5})^2 \frac{4}{5}]$

$=$ (# branches with exactly 2 makes) \neq (Probability of a single branch with exactly 2 makes)

Another way to think about the number of branches:

3 sequences that include exactly 2 makes.

This could also be found using _____.

$$\begin{array}{r} 1 \\ 1 \\ 0 \\ 0 \\ \hline 1 \\ 0 \\ 1 \\ 1 \end{array}$$

$$\begin{array}{ccccccccc} S & S & F & S & F & S & F & F \\ \underline{S} & \underline{S} & \underline{F} & \underline{S} & \underline{F} & \underline{S} & \underline{F} & \underline{F} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{array}$$

(c) Now suppose $n = 8$. Find $P(X = 4)$.

$\binom{8}{4} = 70$ sequences = 70 ways to choose 4 success out of the $n = 8$ trials.

Now just need the probability of 4 success and 4 failures:

$$P(X=4 | n=8, p) = \underbrace{\binom{8}{4}}_{\text{# of sequences}} \underbrace{p^4 q^4}_{\text{prob}}$$

- We can use these patterns to obtain a general formula for the $P(X = x)$ for a binomial random variable with (n, p) .

Definition

- Definition: A random variable X has a **binomial** distribution based on n trials with success probability p if the pmf of X has the form

$$P(X = x | n, p) = \binom{n}{x} p^x (1-p)^{n-x}, \quad x = 0, 1, \dots, n$$

↗ # sequences ↗ probability of each sequence
 ↗ $x! (n-x)!$

where $0 < p < 1$ and n is an integer such that $n \geq 1$.

- Notation: $X \sim \text{Binomial}(n, p)$

range $X = \{0, 1, 2, \dots, n\}$

- Example: Suppose a student is taking a multiple choice exam with 15 questions (4 choices each). They will be randomly guessing on each question. Let X be the number of questions out of 15 for which the student guesses correctly.

- a) What is the distribution of X ? conditions met ✓

$$X \sim \text{Binomial}(n=15, p=0.25)$$

- b) Find the pmf of X .

$$f_X(x | p, n) = \binom{15}{x} 0.25^x 0.75^{15-x}, \quad x = 0, 1, \dots, 15$$

Always need to include range

- c) Find $P(X = 7)$.

$$P(X=7) = \binom{15}{7} 0.25^7 0.75^8$$

Note: on Hw, okay to leave answers in this form

- d) Find the probability of at least one correct.

$$\begin{aligned}
 &\text{at least 1 none} \\
 P(X \geq 1) &= 1 - P(X=0) \\
 &\downarrow \\
 &= 1 - \binom{15}{0} 0.25^0 0.75^{15} \\
 &= 1 - 0.75^{15}
 \end{aligned}$$

e) Find $P(X < 6)$.

$$P(X < 6) = P(X \leq 5) = \sum_{x=0}^5 \binom{15}{x} 0.25^x 0.75^{15-x}$$

True because discrete

f) Find $P(X \geq 6)$.

$$P(X \geq 6) = 1 - P(X \leq 5) = \sum_{x=6}^{15} \binom{15}{x} 0.25^x 0.75^{15-x}$$

Right left $x=6$

- Binomial cdf: $P(X \leq x) = \sum_{t=0}^x \binom{n}{t} p^t q^{n-t}$

Right sided probability: $P(X > x) = 1 - P(X \leq x)$

be careful with $\geq x$ vs $> x$

Mean and variance

- If $X \sim \text{Binomial}(n, p)$

$$E(X) = np$$

$$V(X) = np(1-p) = npq$$

$$\Rightarrow SD(X) = \sqrt{npq}$$

- Continuing example:

(f) Find the expected value and variance of the number of questions guessed correctly.

$$\begin{aligned} E(X) &= np \\ &\downarrow \\ &= 15(0.25) \\ &\downarrow \\ &= 3.75 \text{ correct questions} \end{aligned}$$

$$\begin{aligned} V(X) &= npq \\ &\downarrow \\ &= 15(0.25)(0.75) \\ &\downarrow \\ &= 2.8125 \Rightarrow SD(X) = 1.68 \text{ correct questions} \end{aligned}$$

(g) Suppose each question is worth 4 points. Find the expected value and variance of the number of points missed on the exam if it is worth a total of 60 points.

* X = number of question correct:

* Y = number of points earned: $Y = 4X$

* Z = number of points missed: $Z = 60 - Y = 60 - 4X$

$$\rightarrow E(Z) = E(60 - 4X) = 60 - 4E(X) = 60 - 4 \cdot 3.75 = 45$$

$$\rightarrow V(Z) = V(60 - 4X) = (-4)^2 V(X) = 4^2 \cdot 2.8125 = 45$$

More examples

1. A coin is weighted so that the probability of flipping heads is 0.65. The coin is flipped 10 times and each flip is independent of every other flip. Let X be the number of heads in the 10 flips.

(a) Give the pmf of X , μ_X and σ_x^2 . $X \sim \text{Bin}(n=10, p=0.65)$

$$f(x) = \left(\frac{10}{x}\right) 0.65^x 0.35^{10-x}, \quad x=0, 1, \dots, 10 \quad \rightarrow \quad E(X) = 10(0.65) = 6.5 \\ V(X) = \sqrt{10(0.65)(0.35)} \approx 1.508$$

(b) Find: $P(X = 3)$

$$P(X < 3)$$

$$P(X \geq 3).$$

$$P(X=3) = \text{Binompdf} \left(\begin{array}{rcl} \text{trials} & = & 10 \\ p & = & 0.65 \\ x & \text{ value} & = 3 \end{array} \right)$$

↓

$$\approx 0.0217$$

$$= \int p(x \leq 2) = \text{Binomcdf} \left(\begin{array}{rrr} \text{trials} & = 10 \\ p & = 0.65 \\ x & \text{value} \\ & = 2 \end{array} \right)$$

↓

≈ 0.0048

$$\begin{aligned} &= 1 - P(X \leq 2) \\ &= 1 - \text{Binomcdf}(10, 0.65, 2) \\ &\approx 0.4952 \end{aligned}$$



(c) If $Y = 10 - X$ give the distribution of Y . Find $P(Y \leq 1)$.

failures # successes

original c

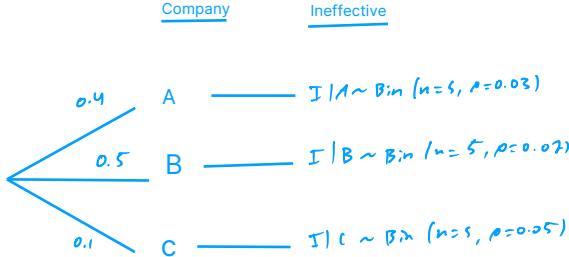
$$Y \sim \text{Binomial}(n=10, p=0.35)$$

$$\rightarrow P(Y \leq 1) = \text{Binomcdf}(10, 0.35, 1) = 0.8695 \iff P(X \geq 1) = 1 - \text{Binomcdf}(10, 0.35, 0)$$

2. A hospital obtains 40% of its flu vaccine shipments from Company A, 50% from Company B, and 10% from Company C.

From manufacturing specifications, it is known that 3% of the vials from A are ineffective, 2% from B are ineffective, and 5% from C are ineffective.

The hospital tests five vials from each large shipment from a company. If at least one of the five is ineffective, find the conditional probability of that shipment having come from C.



$$\begin{aligned}
 P(C | I \geq 1) &= \frac{P(C \cap I \geq 1)}{P(I \geq 1)} \\
 &\quad \swarrow \quad \downarrow \quad \searrow \\
 &= \frac{P(C) P(I \geq 1 | C)}{P(A) P(I \geq 1 | A) + P(B) P(I \geq 1 | B) + P(C) P(I \geq 1 | C)} \\
 &\quad \swarrow \quad \downarrow \quad \searrow \\
 P(I \geq 1 | \cdot) &= 1 - P(I = 0 | \cdot) \\
 &\quad \downarrow \quad : \quad 1 - \rho^5 \\
 &= \frac{\alpha(1 - (0.95)^5)}{0.4(1 - (0.97)^5) + \alpha(1 - (0.98)^5) + \alpha(1 - (0.95)^5)} \\
 &\approx 0.1779
 \end{aligned}$$

Relationship between Bernoulli and binomial

- We can think of the result of a binomial experiment as a sequence of 0s and 1s with length n , where each individual number is the result of a Bernoulli experiment.
- X is the number of successes. Take $n = 5$ for example. We could have:

$$\begin{array}{ll} (0, 0, 1, 0, 0) & x=1 \\ (1, 0, 1, 1, 0) & x=3 \\ (1, 1, 1, 1, 1) & x=5 \end{array} \quad \left. \begin{array}{l} x=\text{Sum of the vectors} \\ \downarrow \text{iid} \dots n \end{array} \right\}$$

- Suppose that $X \sim \text{Bin}(n, p)$ and $Y_1, \dots, Y_n \stackrel{\text{iid}}{\sim} \text{Ber}(p) \iff Y_i \stackrel{\text{iid}}{\sim} \text{Ber}(p)$ where $iid = \text{independent and identically distributed}$.

$$X = \sum_{i=1}^n Y_i$$

Bernoulli distribution is a special case of Binomial when $n=1 \rightarrow \text{Ber}(p) = \text{Binomial}(1, p)$

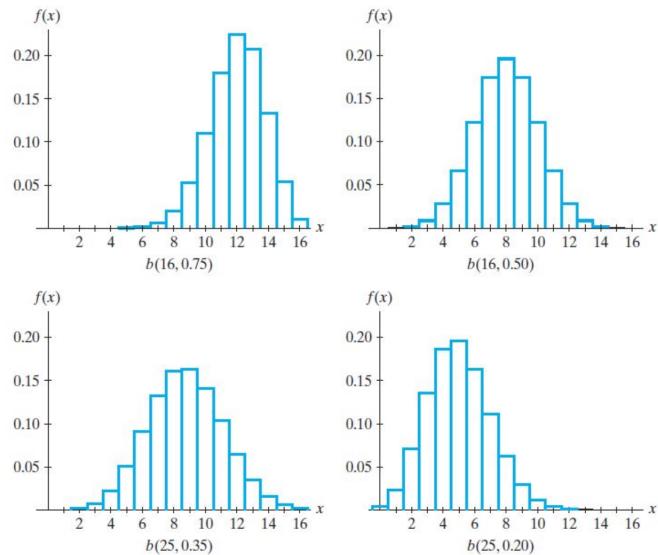
This is intuitively obvious and can be verified in the distribution formula as well.

- Compare the expected values and variances.

Note that the expected value of a sum is the sum of expected values.

$$\begin{aligned} E(X) &= E\left[\sum_{i=1}^n Y_i\right] & V(X) &= V\left[\sum_{i=1}^n Y_i\right] \\ np &= \sum_{i=1}^n E(Y_i) & npq &= \sum_{i=1}^n V(Y_i) \\ \downarrow &= np & \downarrow &= npq \end{aligned}$$

Visualizing binomial distributions



Geometric distribution

Motivation

- Example experiments:
 1. A basketball player shoots three pointers (with success probability of 1/5) until they make the first one. Let X be the number of shots it took to make the first one.
 2. An oil prospector will drill a succession of holes in a given area to find a productive well. He stops when he finds a productive well. Let Y be the number wells drilled before the first productive well.
 3. Roll a fair die until the number 6 appears. Let Z be the number of rolls to get the first 6.
- What is the common characteristic of these experiments? **success/ failure, go until first success**

Geometric experiments and geometric random variables

Same as Binomial

- The general setting for a geometric distribution problem has many features in common with a binomial distribution problem.
- Characteristics of a **geometric experiment**.
 - 1. Each trial is a Bernoulli experiment.
 - 2. The trials are identical and independent.
 - 3. The sample space is: $S = \{S, FS, FFS, FFFF, \dots\}$
- Geometric is a waiting time random variable, with the number of success fixed at 1.
- There are TWO forms of a **geometric RV**. We will be using the FIRST.
 - These are just two different ways to interpret outcomes in the sample space (i.e. define the range of the random variable).
 - ★ 1. The random variable of interest X is the number of trials. \rightarrow like exp 1
 - 2. The random variable of interest Y is the number of failures before the first success. \rightarrow like exp 2
- Relationship between the two formulations.
 - The first way X includes the trial on which the success occurs, whereas the second way Y does not. $\Rightarrow Y = X - 1$
 - This obviously will result in the pmf, expected value, etc. being different between X and Y .

- Example: Check the conditions to determine if example experiment 1 is indeed a geometric experiment.

Each shot follows Independent Bernoulli and want number of shots needed for first make (\times)

$$\text{e.g. } P(X=1) = 0.8^3 (1-0.2)$$

\hookrightarrow counting trials \star

Definition

- Informal derivation of the pmf: We can use the multiplication rule of independent events to find the probability of x trials to get the first success.

1. Success on first trial: $P(X=1) = q^0 p$

In terms of X

$$= q^{1-1} p$$

2. Success on second trial: $P(X=2) = q^1 p$

$$= q^{2-1} p$$

\downarrow success on last trial
 \downarrow Failure on previous $x-1$ trials

3. Success on third trial: $P(X=3) = q^2 p$

$$= q^{3-1} p$$

- If $X \sim \text{Geometric}(p)$,
(Geo)

$$P(X=x | p) = q^{x-1} p, \quad x = 1, 2, 3, \dots$$

Alternate form: $Y = \# \text{ failures before first success}$

\star Note different ranges

$$\hookrightarrow P(Y=y | p) = q^y p, \quad y = 0, 1, 2, \dots$$

Mean and variance

- If $X \sim \text{Geometric}(p)$,

$$E(X) = \frac{1}{p},$$

$$V(X) = \frac{1-p}{p^2} = \frac{q}{p^2}$$

Alternate form: $Y \rightarrow E(Y) = \frac{q}{p} \neq E(X)$

$$V(Y) = \frac{q}{p^2} \stackrel{Y=X-1}{=} V(X)$$

- Continuing example: Let $X \sim \text{Geometric}(p = 1/5)$. How many trials do you expect to see the first success (make)?

$$E(X) = \frac{1}{p} = \frac{1}{1/5} = 5 \rightarrow \text{this makes sense } \checkmark$$

Geometric probabilities

- A geometric random variable is a geometric series:

$$\begin{aligned} p(X=1) &= pe^0 \\ p(X=2) &= pe^1 \\ p(X=3) &= pe^2 \\ &\vdots \end{aligned}$$

- Sums of geometric series:** We will use these properties to derive formulas for geometric probabilities.

Let q be a real number such that $|q| < 1$, and let m be any positive integer $m \geq 1$.

- Infinite geometric series:

$$(a) \sum_{i=0}^{\infty} q^i = q^0 + q^1 + q^2 + \cdots = \frac{1}{1-q} \quad (b) \sum_{i=1}^{\infty} q^i = \frac{q}{1-q}$$

PATTERN: Numerator = first term in the series and denominator = $1 - q$ ALWAYS when infinite sum.

- Another sum:

$$(c) \sum_{i=0}^m q^i = \underbrace{\sum_{i=0}^{\infty} q^i}_{(a)} - \underbrace{\sum_{i=m+1}^{\infty} q^i}_{(b)} = \frac{1}{1-q} + \frac{q^{m+1}}{1-q} = \frac{1 - q^{m+1}}{1-q}$$

- By these sums of a geometric series, we can compute the following probabilities. For positive integers x, a , and b such that $a < b$.

- Total probability: $P(X < \infty) = 1$.

- Derive this:

$$P(X < \infty) = \sum_{x=1}^{\infty} q^{x-1} p$$

Doesn't match (a) by looks, but if we write it out...

$$\begin{aligned} &= P \left(\sum_{x=1}^{\infty} q^{x-1} \right) \\ &= q^0 + q^1 + q^2 + \dots \\ &= \frac{1}{1-q} \\ &= \frac{p}{1-p} \\ &= 1 \end{aligned}$$

(a) \Rightarrow can use result

- Cdf: $P(X \leq x) = 1 - e^{-x}$

$$\begin{aligned} F(x) = P(X \leq x) &= \sum_{t=1}^x t^{-1} p = P \sum_{t=1}^x t^{-1} \\ &\stackrel{\text{Transform}}{=} \sum_{j=0}^{x-1} e^j \quad \text{let } t = j+1 \Rightarrow \text{Bounds} \\ &\quad t=1 \Rightarrow j=0 \\ &= P \left[\frac{1 - e^{-(x-1)}}{1 - e^{-1}} \right] \\ &= 1 - e^{-x} \end{aligned}$$

Intuitive derivation of cdf:

$$\begin{aligned}
 F(x) &= P(X \leq x) \\
 &= 1 - \underbrace{P(X > x)}_{\hookrightarrow \text{need more than } x \text{ trials}} \\
 &= 1 - \underbrace{q^x}_{\Rightarrow \text{failed first } x \text{ times}} \\
 &\quad \text{doesn't matter what happens next}
 \end{aligned}$$

– Right probability (exclusive): $P(X > x) = \sum_{i=x+1}^{\infty} q^{i-1} p$ Directly

$$\begin{aligned}
 &= 1 - P(X \leq x) = 1 - F(x) \\
 &= 1 - (1 - q^x) \\
 &= q^x
 \end{aligned}$$



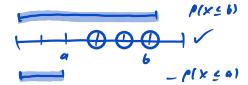
– Right probability (inclusive): $P(X \geq x) = P(X > x-1) = q^{x-1}$



$$\begin{aligned}
 \text{OR} \quad &= 1 - (1 - P(X \leq x-1)) \\
 &= 1 - (1 - q^{x-1}) \\
 &= q^{x-1}
 \end{aligned}$$

– Interval probability: $P(a < X \leq b) = P(X \leq b) - P(X \leq a)$

$$\begin{aligned}
 &= 1 - q^b - (1 - q^a) \\
 &= q^a - q^b
 \end{aligned}$$



– Interval probability (both inclusive): $P(a \leq X \leq b) = P(a-1 < X \leq b)$

$$\begin{aligned}
 &= q^{a-1} - q^b
 \end{aligned}$$

- Example: Suppose that the probability of engine malfunction during any one-hour period is $p = 0.02$. Let X be the number of one-hour intervals until the first malfunction.

- a) What is the distribution of X ? Conditions met ✓

$$X \sim \text{Geometric}(p=0.02)$$

- b) Find the pmf of X .

$$f_X(x) = 0.98^{x-1} (0.02), \quad x=1, 2, 3, \dots$$

- c) Find the probability that an engine will survive 3 hours.

$$\begin{aligned} P(X > 3) &= 0.98^3 \approx 0.94 \\ &\downarrow = 1 - \text{geometcdf}(p=0.02, k=3) \approx 0.94 \end{aligned}$$

we know the first 3 trials (hours) are failure (surviving)
Then we don't care when success (malfunction) happens

- d) Find the probability that an engine will survive 3 hours, but die before or during the 6th hour.

$$\begin{aligned} P(3 < X \leq 6) &= 0.98^3 - 0.98^6 \approx 0.055 \\ &\downarrow = \text{geometcdf}(p=0.02, k=6) - \text{geometcdf}(p=0.02, k=3) \\ &\Rightarrow \text{geompdf}(p=0.02, x=k) \end{aligned}$$

- e) Suppose that an engine survives 3 hours. Find the probability that the engine will survive 6 hours.

conditional

$$\begin{aligned} P(X > 6 | X > 3) &= \frac{P(X > 6)}{P(X > 3)} \quad \{X > 6\} \cap \{X > 3\} \Rightarrow \frac{P(X > 6)}{P(X > 3)} = \frac{P(X > 3)}{P(X > 3)} \\ &\downarrow = \frac{0.98^6}{0.98^3} \\ &= 0.98^3 \quad \Leftrightarrow P(X > 3) \rightarrow \text{memoryless property} \end{aligned}$$

Memoryless property

same calculation as

- The geometric distribution has an interesting property, known as the “memoryless” property. For integers $s > t$, it is the case that

$$P(X > s | X > t) = P(X > s - t)$$

- To help us understand what this means, let's interpret some events for $s > t$.

- $\{X > s\} \rightarrow$ First s trials are failures

- $\{X > t\} \rightarrow$ First t trials are failures

- $\{X > s\} \mid \{X > t\} \rightarrow$ Next $s-t$ trials are failures after the first t trials are failures

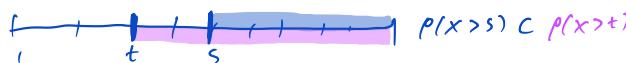
- For example, let's say I flip a coin three times and don't get a tails. Do these past failures help to increase the probability of getting a tails next?

No!

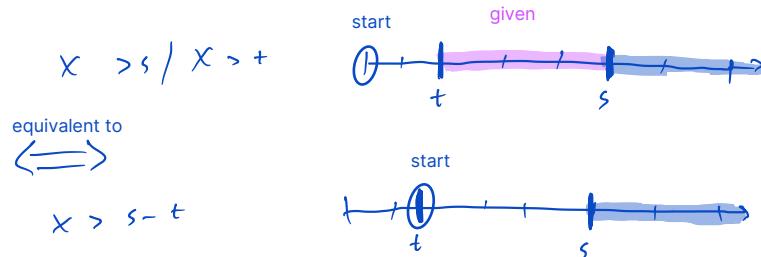
FFF FFF
FFF

- We want to know how (if at all) this past information influences the future event. Recall $P(X > a) = q^a$.

$$P(X > s | X > t) = \underbrace{\frac{P(\{X > s\} \cap \{X > t\})}{P(X > t)}}_{a} = \frac{P(X > s)}{P(X > t)} = \frac{q^s}{q^t} = \boxed{q^{s-t}} = P(X > s-t)$$

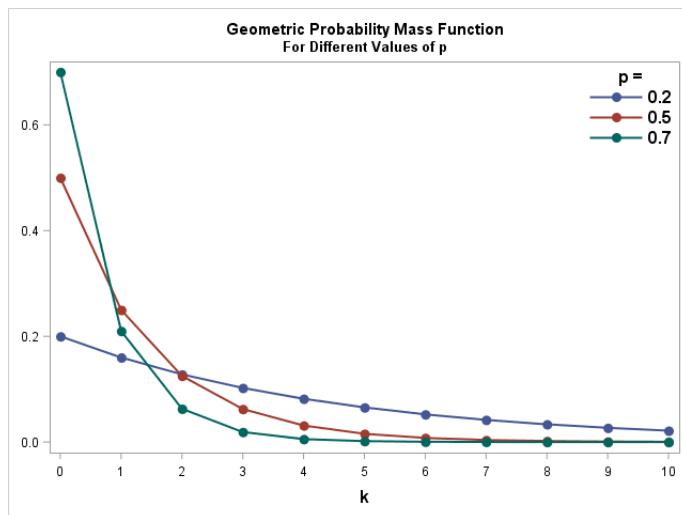


- The probability of getting an additional failures $s-t$, having already observed t failures is the same probability of observing $s-t$ failures at the start of the sequence.



- That's why this property is called the memoryless property. Even though we have some failures before, a geometric experiment is sort of restarting again. Thus, the geometric distribution "forgets" what has occurred.

Visualizing geometric distributions



Negative binomial distribution

Motivation

- Example experiments:
 1. A basketball player shoots three pointers (with success probability of 1/5) until they make the 4th shots. Let X be the number of shots it took to make the 4th one.
 2. An oil prospector will drill a succession of holes in a given area to find a productive well. He stops when he finds three productive wells. Let Y be the number wells drilled before the 3rd productive well.
- What is the common characteristic of these experiments?

success/ failure
go until succeed multiple times

Negative binomial experiments and negative binomial random variables

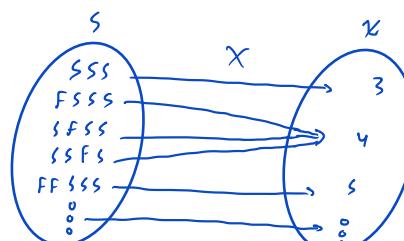
- The negative binomial experiment is a generalized version of the geometric experiment.
- Characteristics of a **negative binomial experiment**.
 - 1. Each trial is a Bernoulli experiment.
 - 2. The trials are identical and independent.
 - 3. The number of successes to stop a negative binomial experiment is denoted by r .
 - This is a parameter of the distribution.
- **Negative binomial random variable.**

(Following the logic from the geometric distribution, there is also two ways to define a negative binomial random variable. We will only discuss the one corresponding to the version of the geometric we are using.)

 - The random variable of interest X is the number of trials until r successes.
 - Simple demonstration of the random variable mapping: When $r = 3$, the sample space and range are:

– In general, the range of X is

$$\star \quad X = \{r, r+1, \dots\}$$



- Example: Check the conditions to determine if example experiment 1 is indeed a negative binomial experiment.

→ Geometric experiment with $r=4$ successes ✓
 → want total number of shots (x)

Definition

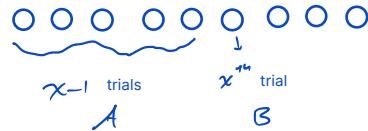
- Informal derivation of the pmf: We can again use the multiplication rule of independent events to find the probability of x trials to get r successes.

1. The outcomes in the event $\{X = x\}$ must follow these two rules:

- A) The first $x - 1$ trials result in $r - 1$ successes and $\underbrace{x - r}_{\text{total number of failures}}$ failures.
- B) The last x^{th} trial has to result in success.

2. Let's visualize these rules (and think of them as events) to get some probabilities.

- Consider a sequence of Bernoulli trials with probability of success p , $0 < p < 1$. Let $x = r, r + 1, \dots$



A) $A \sim \text{Binomial}(n = x-1, p) \rightarrow$ want $P(A = r-1)$
 $P(A) = \binom{x-1}{r-1} p^{r-1} q^{x-r} \rightarrow$ $\underbrace{x-1 - (r-1)}_{\text{A} \perp\!\!\!\perp \text{B}}$

B) $B \sim \text{Bernoulli}(p) \rightarrow$ want $P(B = 1)$
 $P(B) = p$

3. Putting these together:

$$f_X(x | r, p) = P(A \cap B) = P(A) P(B) = \left[\binom{x-1}{r-1} p^{r-1} q^{x-r} \right] p$$

$$\downarrow \quad \stackrel{?}{=} \quad \binom{x-1}{r-1} p^r q^{x-r}$$

- If $X \sim \text{Negative binomial}(r, p)$,

$$P(X = x | r, p) = \binom{x-1}{r-1} p^r q^{x-r} \quad x = r, r + 1, \dots$$

\downarrow \curvearrowright probability of each individual sequence

of sequences

- Note: This is called the "negative binomial"; distribution because it looks like the binomial pmf, except with minus ones in the combination.

Mean and variance

- If $X \sim \text{Negative binomial}(r, p)$,

$$E(X) = \frac{r}{p}, \quad V(X) = \frac{r(1-p)}{p^2} = \frac{rq}{p^2} \implies SD(X) = \sqrt{\frac{rq}{p^2}}$$

Example

- You are playing the slot machine on which the probability of a win on any individual trial is 0.05. You will play until you win twice. Let X denote the number of plays in order to get your second win.

- a) What is the distribution of X ? Conditions met ✓

$$X \sim NB(r=2, p=0.05)$$

- b) Find the pmf of X .

$$f_{X|2}(x) = \binom{x-1}{2-1} 0.05^2 0.95^{x-2}, \quad x=2, 3, \dots$$

- c) Find the probability that the second win occurs on the 8th play.

$$P(X=8) = \binom{7}{1} 0.05^2 0.95^6 \approx 0.0129$$

- d) Find $E(X)$ and $V(X)$.

$$\begin{aligned} E(X) &= \frac{r}{p} & V(X) &= \frac{rv}{p^2} \\ &= \frac{2}{0.05} & &= \frac{2(0.95)}{0.05^2} \\ &= 40 \text{ plays} & &= 720 \rightarrow SD(X) \approx 27.57 \text{ plays} \end{aligned}$$

- e) Find the probability that the fourth win occurs before the 10th play.

* Be careful with bounds

$$Y \sim NB(r=4, p=0.05) \implies P(Y < 10) = \sum_{y=4}^9 \binom{y-1}{4-1} 0.05^4 0.95^{y-4}$$

Relationship between geometric and negative binomial

- We stated the negative binomial distribution is a generalized version of the geometric distribution. Here's why.
 - The negative binomial is the sum of independent and identical geometric experiments:

Formally, if $X \sim \text{NB}(r, p)$ and $Y_1, \dots, Y_r \stackrel{\text{ iid }}{\sim} \text{Geo}(p)$.

$$X = \sum_{i=1}^r y_i$$

- We can see this for the expected value and variance too.

$$E(X) = r \underbrace{\left(\frac{1}{p}\right)}_{\hookrightarrow E(Y_i)} \text{ geometric mean}$$

$$v(x) = r \underbrace{\frac{q}{p^2}}_{\hookrightarrow = v(y_i)}$$

geometric variance

- Intuitively this makes sense.

- If $r = 2$, Y_1 is the number of trials to get the first success and Y_2 is the number of subsequent trials to get the second success.
 - If we are waiting for the second success, we wait through Y_1 trials for the first success, and then repeat the process as we go through Y_2 subsequent trials for the second success, for a total of $X = Y_1 + Y_2$ trials.

$$\text{e.g.) } V=3 \rightarrow \boxed{\text{FS FFS FS}} \quad \begin{array}{l} \text{one NB exp} \\ \rightarrow x=7 \end{array}$$

$$\boxed{\text{FS FFS FS}} \quad \begin{array}{l} \text{three geo exps} \\ Y_1=2 \quad Y_2=3 \quad Y_3=2 \quad \rightarrow E Y_i = 2+3+2 = 7 \end{array}$$

- Note: Even though Y_1 and Y_2 follow the same kind of geometric distribution, Y_1 and Y_2 can have different values. So $Y_1 + Y_2$ is NOT the same as $2Y_1$.

- The geometric distribution is a special case of the negative binomial distribution when $r = 1$, which counts the number of trials for the first success

This again is intuitively obvious and can be verified with the distribution formula.

$$X \sim NB(r=1, \rho) \iff \text{Geometric}(\rho)$$

Hypergeometric distribution

Motivation

- Polling problem:

– Suppose you live in a large city which has 1,000,000 registered voters. The voters will vote on a tax issue next month and you want to estimate the percent of the voters who favor the issue.

You cannot poll everyone, so you randomly sample 100 voters and ask them if they favor the issue. What are your chances of correctly estimating the true percentage in favor of the issue?

This is a MATH 321 question! We could build a confidence interval to estimate the unknown p .

– For now, we can make a simplifying assumption to answer this question. Suppose the true percentage of the voters in favor of the issue is 65%. We don't actually know this number, it is what we are trying to estimate.

– But with this assumption, when polling voters, we are really doing a binomial experiment.

$$\begin{aligned} \rightarrow X &= \text{number in favor of issue} \\ \downarrow &\sim \text{Binomial}(n=100, p=0.65) \\ \rightarrow p(X=0.65) &=? \end{aligned}$$

- Checking assumptions for this polling problem.

– Each voter (trial) is in favor (success) or not in favor (failure).

Random sampling, so each successive voters are independent.

Sampling $n = 100$ voters, and get a sequence of S/F.

Is there a constant probability of success p ? *No!*

– The usual method of sampling voters is called **sampling without replacement**.

	<u>voter 1</u>	<u>voter 2</u>	
options :	1,000,000	999,999	\Rightarrow probability in favor changes slightly
success probability:	p_1	$\neq p_2$	

– When there is a very large population N and the sample n is very small in comparison, not replacing changes things very little on each trial, and it is still reasonable to model this scenario with the binomial distribution.

In intro stats classes, this condition is met if $\frac{1}{10}N \geq n$.

– The hypergeometric distribution will handle sampling without replacement exactly for any population size.

- Example experiments:

1. Five cards are dealt from a standard deck. Let X be the number of aces in the hand.
2. There are three red chips and four blue chips in a bowl. We randomly select four chips without replacement from the bowl. Let Y be the number of blue chips selected.

- What is the common characteristic of these experiments?

number of successes in some amount of trials, without replacement

Example calculations

- Informal derivation of the pmf via an example: We have actually already been doing hypergeometric problems (sampling without replacement) when we did certain counting problems.

- (Test 1 Q3) A gym teacher is picking students to compete in a pickup basketball game. There are 45 students in the class, which includes 25 boys and 20 girls. The teacher picks 7 students from the 45 at random and without replacement. Find the probability that the team includes exactly 4 girls.

$$P(4 \text{ Girls, } 3 \text{ Boys}) = \frac{\binom{20}{4} \binom{25}{3}}{\binom{45}{7}} = \frac{\left(\begin{array}{l} \# \text{ of ways} \\ \text{choose 4 Girls} \end{array}\right) \left(\begin{array}{l} \# \text{ of ways} \\ \text{choose 3 Boys} \end{array}\right)}{\left(\begin{array}{l} \# \text{ of ways} \\ \text{choose 7 overall} \end{array}\right)}$$

\downarrow
 ≈ 0.246

- We can easily generalize this to find the probability that the selected team includes any number of girls between 0 and 7. Let X be the number of girls selected.

$$P(X=x) = \frac{\binom{20}{x} \binom{25}{7-x}}{\binom{45}{7}}, \quad x=0, 1, 2, \dots, 7$$

- This leads to the following pmf for X :

x	0	1	2	3	4	5	6	7
$f(x)$	0.011	0.078	0.222	0.318	0.246	0.102	0.021	0.002

- Again, successive selections are dependent on previous one. So the probabilities change. \rightarrow Direct way

$$P(G_1) = \frac{20}{45} \approx 0.444 \quad P(G_2 | G_1) = \frac{19}{44} \approx 0.433$$

$\frac{\text{Girl 1}}{\text{Girl 2}}$

- Now we can formalize this example and relate it to how we will talk about hypergeometric experiments and random variables.

In Example

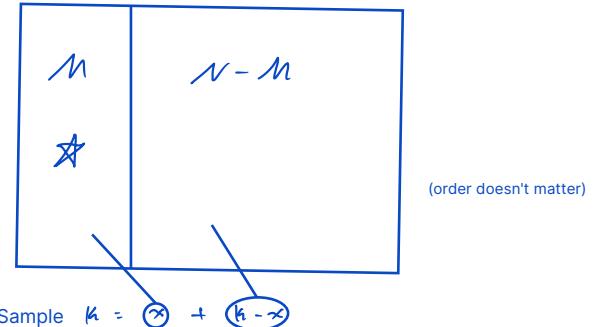
Hypergeometric experiments and hypergeometric random variables

→ $K = 7$ team members $N=45$ total students $M=20$ Girls $\Rightarrow N-M = 25$ Boys• Characteristics of a **hypergeometric experiment**.

1. Each trial is a Bernoulli experiment.
2. Trials are not identical and not independent. → x , number of girls selected for team
3. A sample of size K is being taken from a finite population of size N .
4. The population has a subgroup of size M that is of interest.

• **Hypergeometric random variable**.

- The random variable of interest X is the number of members of the subgroup in the sample taken.

Population N 

Definition

- If $X \sim \text{Hypergeometric}(N, M, K)$,

$$P(X = x | N, M, K) = \frac{\binom{M}{x} \binom{N-M}{K-x}}{\binom{N}{K}}, \quad x = 0, 1, \dots, \min(M, K)$$

↓

$$= \frac{\left(\begin{array}{l} \text{choose } x \text{ successes} \\ \text{from } M \end{array} \right) \left(\begin{array}{l} \text{choose } K-x \text{ failures} \\ \text{from } N-M \end{array} \right)}{\left(\begin{array}{l} \text{choose } x \text{ overall} \\ \text{from } N \end{array} \right)}$$

- Two cases for the range of X .

- ∅ 1. In most applications $M \geq K$, which means the sample size is smaller than the subgroup of interest.

- This implies: $\mathcal{X} = \{0, 1, \dots, \min(M, K)\}$.

$$\text{ex) } M=20 \quad K=7$$

$$x=0, 1, 2, \dots, 7$$

2. But if the sample size K is quite large and we lose this restriction, the formula will still be applicable.

- But with new range: $\mathcal{X} = \{\max(0, K + M - N), \dots, \min(M, K)\}$.

$$\begin{array}{r} 20+10-25 \\ \downarrow \quad \downarrow \\ 10 \quad 20 \\ = 5 \end{array}$$

$$\text{ex) } N=25$$

$$K=20$$

$$M=10$$

$$x=5, 6, \dots, 10$$

$$N-M=15$$

Examples

1. A lot consisting of 200 fuses is inspected by the following procedure: 20 fuses are chosen at random and tested; if at least 19 blow at the correct amperage, the lot is accepted. If a lot contains 10 defective fuses what is the probability of accepting this lot?

Solve this problem two different ways.

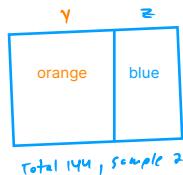
$$\begin{aligned} X &= \text{number of defective fuses} \\ \downarrow &\sim H_6(N=200, M=10, K=20) \\ \Rightarrow x &= 0, 1, 2, \dots, 10 \end{aligned}$$

$$\begin{aligned} P(X \leq 1) &= P(X=0) + P(X=1) \\ &= \frac{\binom{10}{0}/\binom{190}{20} + \binom{10}{1}/\binom{190}{19}}{\binom{200}{20}} \\ &\approx 0.7372 \end{aligned}$$

$$\begin{aligned} Y &= \text{Number of Correct fuses} \\ \downarrow &\sim H_6(N=200, M=190, K=20) \\ \Rightarrow y &= 10, 11, \dots, 20 \end{aligned}$$

$$\begin{aligned} P(Y \geq 19) &\leq P(X=19) + P(X=20) \\ &= \frac{\binom{190}{19}/\binom{190}{20} + \binom{190}{20}/\binom{190}{0}}{\binom{200}{20}} \end{aligned}$$

2. A bag contains 144 ping-pong balls. More than half of the balls are painted orange and the rest are painted blue. Two balls are drawn at random without replacement. The probability of drawing two balls of the same color is the same as the probability of drawing two balls of different colors. How many orange balls are in the bag?



$$\begin{aligned} X &\sim H_6(N=144, M=y, K=2) \\ \Rightarrow 144-y &= z \\ y > z \end{aligned}$$

$$\begin{aligned} P(X=0, z) &= P(X=1) \\ \frac{\binom{y}{0}\binom{z}{2} + \binom{y}{1}\binom{z}{1}}{\binom{144}{2}} &= \frac{\binom{y}{1}\binom{z}{1}}{\binom{144}{2}} \\ \binom{z}{2} + \binom{y}{2} &= yz \end{aligned}$$

$$\begin{aligned} \frac{z!}{2!(z-2)!} + \frac{y!}{2!(y-2)!} &= \\ \frac{z(z-1)}{2} + \frac{y(y-1)}{2} &= \end{aligned}$$

$$2(z-1) + y(y-1) = 2yz \rightarrow z = 144-y$$

$$(144-y)(143-y) + y(y-1) = 2y(144-y)$$

$$y^2 - 287y + 20592 + y^2 - y = -7y^2 + 288y$$

$$4y^2 - 576y + 20592 = 0$$

$$y^2 - 144y + 5198 =$$

$$(y-6)(y-78) =$$

$$y = 78 \text{ or } y = 66$$

y must be $> z$

Mean and variance

- If $X \sim \text{Hypergeometric}(N, M, K)$,

$$E(X) = K \left(\frac{M}{N} \right),$$

$$V(X) = K \left(\frac{M}{N} \right) \left(\frac{N-M}{N} \right) \left(\frac{N-K}{N-1} \right) \implies \text{SD}(X) = \sqrt{V(X)}$$

Interpretation of the mean: We want the sample to have the same proportion of successes as the population.

- Although the mean and variance of the hypergeometric random variable look complicated, it is actually easily understood when we compare it to those of binomial(n, p).

$$E(Y) = np \quad V(Y) = npq$$

$\hookrightarrow Y$

This switch from hypergeometric to binomial means that we are now sampling with replacement.

- For hypergeometric, the population size is N and there are M elements of interest.

The probability of success is $p = \frac{M}{N}$.

And the sample size K is like the number of trials n .

Then the mean and variance of HG(N, M, K) can be re-expressed like this:

In terms of Y

$$E(X) = \frac{np}{\binom{M}{n}}$$

$$V(X) = \frac{n \rho q \left(1 - \frac{n-k}{N-1} \right)}{\binom{M}{n} \left(1 - \frac{M-n}{N-1} \right)} = \frac{n \rho q}{\binom{M}{n}}$$

- Observations when comparing with vs without replacement:

- Means: The expected values are the same regardless of the type of sampling.
- Variances: The variances are somewhat smaller when sampling without replacement.

$$\frac{n-k}{N-1} \leq 1$$

This final term in the hypergeometric variance is often called the **finite population correlation factor**, which is an adjustment because the sampling in the hypergeometric experiment is done without replacement and is therefore not independent.

Examples

1. A standard 52 card deck contains 4 different suits, each with 13 cards. Five cards are dealt from a standard deck. Let X be the number of spades in the hand.

- (a) What is the distribution of X ?

$$X \sim H(52, 13, 5)$$

- (b) Find the pmf of X .

$$f(x) = \frac{\binom{13}{x} \binom{39}{5-x}}{\binom{52}{5}}, \quad x=0, 1, \dots, 5$$

- (c) Find the probability of two or three spades in the hand dealt.

$$P(X=2) + P(X=3) = \sum_{x=2}^3 \frac{\binom{13}{x} \binom{39}{5-x}}{\binom{52}{5}}$$

- (d) Find the probability of at least two spades in the hand dealt.

$$P(X \geq 2) = \sum_{x=2}^5 \frac{\binom{13}{x} \binom{39}{5-x}}{\binom{52}{5}}$$

- (e) Find the probability of at most 4 spades in the hand.

$$\begin{aligned} P(X \leq 4) &= \sum_{x=0}^4 \frac{\binom{13}{x} \binom{39}{5-x}}{\binom{52}{5}} = 1 - P(X=5) \\ &= 1 - \frac{\binom{13}{5} \binom{39}{0}}{\binom{52}{5}} \end{aligned}$$

- (f) Find $E(X)$ and $V(X)$.

$$\begin{aligned} E(X) &\approx np \\ &= 5 \left(\frac{13}{52} \right) \\ \downarrow &\simeq 1.25 \text{ spades} \end{aligned}$$

$$\begin{aligned} V(X) &\approx npq \left(\frac{n-q}{n-1} \right) \\ &= 5 \left(\frac{13}{52} \right) \left(\frac{21}{52} \right) \left(\frac{47}{51} \right) \\ \downarrow &\simeq 0.8639 \end{aligned}$$

2. A machine shop orders 200 bolts from a supplier. To determine whether to accept the shipment of bolts, the manager of the facility randomly selects 30 bolts without replacement. If of the 30 randomly selected bolts 2 or less are found to be defective, he concludes that the shipment is acceptable.

- (a) If 20% of the bolts in the population are defective, what is the probability that the shipment is acceptable?

$$X \sim H(1/N=200, M=0.2(200), K=30) \quad \rightarrow \quad P(X \leq 2) = \sum_{x=0}^2 \frac{\binom{40}{x} \binom{160}{30-x}}{\binom{200}{30}} \downarrow \approx 0.0331$$

- (b) Now assume the 30 bolts are chosen with replacement. If 20% of the bolts in the population are defective, what is the probability that the shipment is acceptable?

$$Y \sim \text{Binomial}(n=30, p=0.2) \quad \rightarrow \quad P(Y \leq 2) = \sum_{y=0}^2 \binom{30}{y} 0.2^y 0.8^{30-y} \downarrow = \text{Binomial}(30, 0.2, 2) \approx 0.0442$$

Relationship between hypergeometric and binomial

- Hypergeometric experiment is without replacement and binomial is with replacement.
- As the population size goes to infinity ($N \rightarrow \infty$), the mean and variance of hypergeometric(N, M, K) converges to those of binomial(n, p).

Here's why:

- We saw the expected values were already equivalent in the finite case. $E(X) = E(Y)$
- The last term in the variance of HG goes to 1, and we are left with what is equivalent to the binomial variance.

$$\lim_{N \rightarrow \infty} \frac{N-K}{N-1} = 1 \Rightarrow \lim_{N \rightarrow \infty} V(X) = V(Y) = npq$$

Just because the means and variances are of hypergeometric converge to those of binomial as we let the sample size go to infinity (i.e. in asymptotics), we cannot say that the random variables are identical.

- This turns out to be a true statement, be we need to show convergence in distribution, which is a topic of graduate level probability theory.

Summary of commonly used discrete distributions so far

- We studied four distributions based on Bernoulli experiments:
 - Binomial, Geometric, Negative Binomial, and Hypergeometric.
- Throughout all of these, there were three important aspects:

1. Number of successes
2. number of trials
3. probability of success

- A • We can organize the four distributions based on what we are interested in (the random variable) and what we are given (as parameters).
- Distributions counting the number of successes.

Binomial & Hypergeometric

Interested in ①

② + ③ are given as parameters.

Only difference is _____.

- Distributions counting the number of trials.

Geometric & Negative Binomial

Interested in ②.

① + ③ are given as parameters.

Only difference is _____.

Poisson Distribution

Motivation

- Example experiments:
 1. The number of accidents at a particular intersection during a time period of one week.
 2. The number of earthquakes in California during a time period of one year.
 3. The number of flaws in 100 feet of wire.
- What is the common characteristic of these experiments?

Counting the number of occurrences during a given period

Poisson experiments and Poisson random variables

- **Poisson experiments:** The Poisson distribution can be used to model several different types of experiments.
 - A scenario in which we are waiting for an occurrence (such as waiting for a bus, or waiting for customers to arrive in a bank).
 - The number of occurrences in a given time interval (such as experiments 1 and 2) or on physical objects (such as experiment 3).
 - Spatial distributions (such as the distribution of fish in a lake).

All of these situations are about an event which is said to occur at an average rate λ per given unit (usually time period, could be area, location, etc.).

- **Poisson random variable.**
 - The random variable of interest X is the number of events in a given unit.
 - The range of X is

$$X = \{0, 1, 2, \dots\}$$

Definition

- First we will look at an example to get an idea of the types of problems that we use the Poisson distribution for.
- Example: An analyst studies data on accidents and an intersection and concludes that accidents occur there at an “average rate of $\lambda = 2$ per month”.

The number of accidents X in a month is a random variable. And the Poisson distribution can be used to find probabilities $P(X = x)$ in terms of x and λ the average rate.

- Definition: A random variable X follows the Poisson distribution with parameter (or average rate) λ if

$$P(X = x | \lambda) = \frac{e^{-\lambda} \lambda^x}{x!}, \quad x = 0, 1, 2, \dots \quad \text{and} \quad \lambda > 0$$

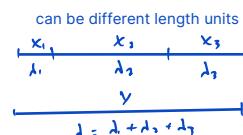
Note: λ is the expected value of the number of occurrences per unit.

- Notation: $X \sim \text{Poisson}(\lambda)$

- Theorem (Combining Poisson random variables): If $X_i \stackrel{\text{IID}}{\sim} \text{Poisson}(\lambda_i)$ (not necessarily identical because λ_i 's can be different), then

$$Y = \sum_{i=1}^n X_i$$

$\downarrow \sim \text{Poisson}(\lambda = \sum_{i=1}^n \lambda_i)$



If X_i 's are identical and therefore *iid*, then the new parameter for Y becomes $n\lambda$ and the number of occurrences in n units is

$$Y \sim \text{Poisson}(n\lambda)$$

Mean and variance

- If $X \sim \text{Poisson}(\lambda)$,

$$E(X) = \lambda,$$

$$V(X) = \lambda$$

- It should be obvious that the expected value is equal to λ , the average rate at which events occur.

Examples

1. Continuing accidents example: The number of accidents in a month at this intersection can be modeled using the Poisson distribution with an average rate of $\lambda = 2$.

$$X \sim \text{Poisson}(\lambda=2) \Rightarrow f(x|\lambda=2) = \frac{e^{-2} 2^x}{x!}, x=0,1,2,\dots$$

Now we can easily find probabilities and the mean and variance.

- (a) Find the probability there are no accidents this month.

$$P(X=0) = \frac{e^{-2} 2^0}{0!} = e^{-2} = \text{poisson pdf } (\lambda=2, x=0) \approx 0.135$$

- (b) Find the probability there is one accident this month.

$$P(X=1) = \frac{e^{-2} 2^1}{1!} = 2e^{-2} = \text{poisson pdf } (\lambda=2, x=1) = 0.2707$$

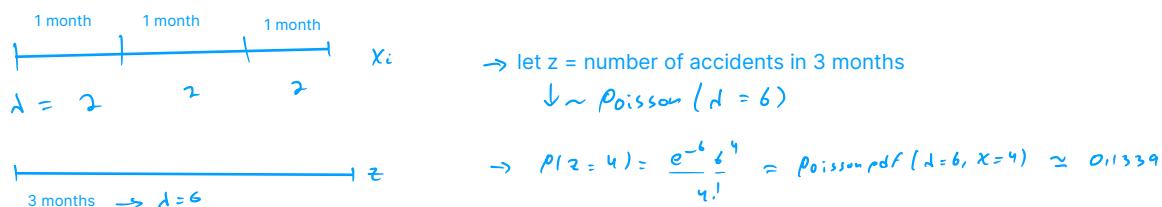
- (c) Find the probability there are more than two accidents this month.

$$\begin{aligned} P(X > 2) &= 1 - P(X \leq 2) \\ &= 1 - \sum_{x=0}^2 \frac{e^{-2} 2^x}{x!} \\ &= 1 - \text{poisson cdf } (\lambda=2, x=2) \approx 0.3233 \end{aligned}$$

- (d) Suppose each accident costs the driver \$350, find the mean and variance for the cost of accidents this month.

$$\begin{aligned} Y = 350X &\Rightarrow E(Y) = E(350X) \\ &= 350 \underbrace{E(X)}_{\text{Incorrect}} \\ &\downarrow = \$700 \quad \cancel{\Rightarrow} \quad V(k) = 700 \quad \text{because} \quad Y = 350X, y = \{0, 350, 700, \dots\} \\ &\quad \downarrow \not\sim \text{Poisson} \\ V(Y) = V(350X) &= 350^2 \underbrace{V(X)}_{\text{Correct}} \\ &\downarrow = 245,000 \text{ dollars-squared} \quad \quad \quad Z = X_1 + X_2 + \dots + X_{350} \\ &\quad \downarrow \sim \text{Poisson } (\lambda=700), z = \{0, 1, 2, \dots\} \end{aligned}$$

- (e) Find the probability there are four accidents in a three month span.



2. Assume the number of hits, X , per baseball game has a Poisson distribution. If the probability of a no-hit game is $1/10000$, find the probability of having 4 or more hits in a particular game.

$$\rightarrow X \sim \text{Poisson}(\lambda)$$

$\hookrightarrow ??$ solve for λ using known info

$$\rightarrow P(X=0) = \frac{1}{10000}$$

$$\frac{e^{-\lambda} \lambda^0}{0!} = \downarrow$$

$$\ln(e^{-\lambda}) = \ln\left(\frac{1}{10000}\right)$$

$$-\lambda = \ln\left(\frac{1}{10000}\right)$$

$$\Rightarrow \lambda = -\ln(10000)$$

$\rightarrow P(X \geq 4) = 1 - P(X \leq 3)$

$= 1 - \text{Poisson CDF } (\lambda = -\ln(10000), x=3)$
 $= 1 - \sum_{x=0}^3 \frac{e^{-\ln(10000)} (-\ln(10000))^x}{x!}$
 ≈ 0.9817

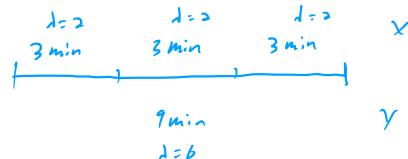
3. In a large city, telephone calls to 911 come on the average of two every 3 minutes. If the city assumes an approximate Poisson process, what is the probability of five or more calls arriving in a 9-minute period?

$$\rightarrow X \sim \text{Poisson } (\lambda = 2)$$

$\hookrightarrow 3 \text{ min}$

$$\rightarrow Y \sim \text{Poisson } (\lambda = 3 \cdot 2 = 6)$$

$\hookrightarrow 9 \text{ min}$



$$\rightarrow P(Y \geq 5) = 1 - P(Y \leq 4)$$

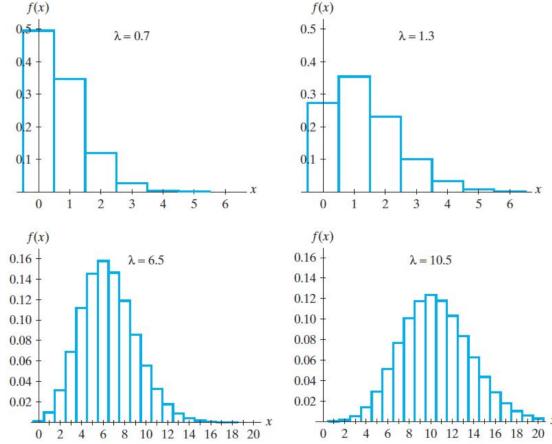
$$\downarrow$$

$$= 1 - \text{Poisson CDF } (\lambda=6, x=4)$$

$$= 1 - \sum_{y=0}^4 \frac{e^{-6} 6^y}{y!}$$

$$\approx 0.7144$$

Visualizing Poisson distributions



Poisson approximation to the binomial distribution

- In grad school :), you will learn that Poisson is the limiting distribution of binomial. Said slightly more formally:

$$\lim_{n \rightarrow \infty} \text{Binomial}(n, p = \frac{\lambda}{n}) = \text{Poisson}(\lambda = np)$$

- This means that for large n and small p , the Poisson distribution can be used to approximate the binomial distribution.

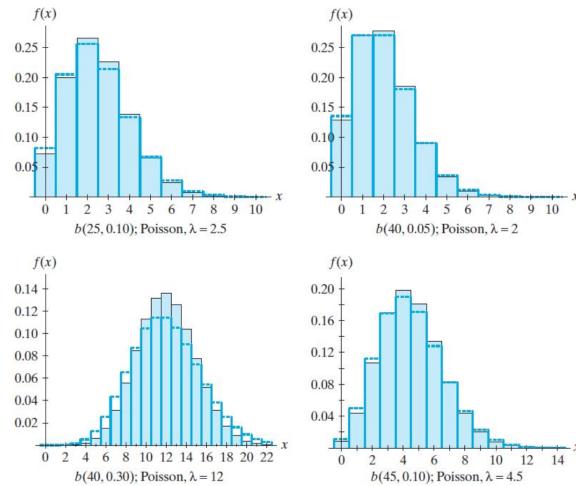


Figure 2.6-2 Binomial (shaded) and Poisson probability histograms

Lectures 11 – Continuous Distributions

MATH 320: Probability

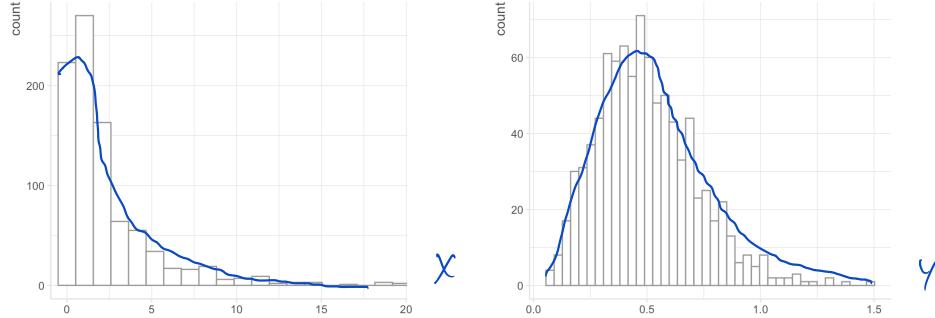
Lecture 11: Continuous Distributions

Chapters 3: Distributions (3.1, 3.2, 3.3)

Introduction

- Recall that statistical distributions are used to model populations; this is the goal!
- Suppose an insurance company has the following data from two samples of $n = 900$ policyholder losses for two different policies.

We can think of these as observations from loss random variables X and Y .



- The question is: How can we model the population that these losses are coming from?
Said another way: In theory there is some distribution that these losses follow, and we want to figure out which family it is and what the parameter values are \Rightarrow Find equation for smooth curve.
- Once the researcher knows this, they can figure out lots of things that are useful in analyzing how losses will occur in the future. For example (suppose X is in thousands of dollars):
 - Can find the expected claim amount $E(X)$. Then use this to set a premium amount to ensure at least breakeven.
 - Can figure out the probability of “major” losses, say $P(X > 15 = \$15,000)$.
 - Or if there is a deductible of say \$2,000, can find the percentage of claims that the insurance company is not responsible to pay for: $P(X < 2 = \$2,000)$.

- If we already have data, we could simply approximate these values with empirical formulas:

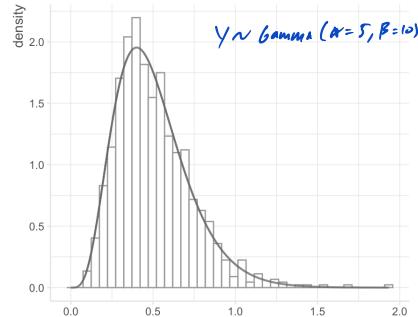
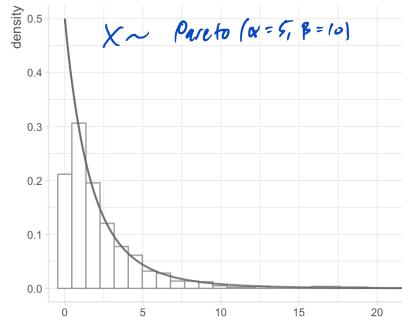
$$- E(X) \approx \frac{1}{n} \sum_{i=1}^n x_i = \frac{1}{900} \sum_{i=1}^{900} x_i$$

$$- P(X > 15 = \$15,000) \approx \frac{n(X > 15)}{900}$$

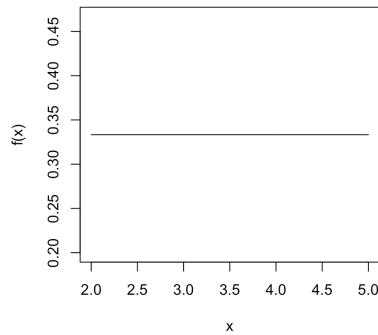
$$- P(X < 2 = \$2,000) \approx \frac{n(X < 2)}{900}$$

- But these are just approximations and are based on a single sample of data. Having population information is much better and much more generalizable!

Observation: When we overlay these density curves, they match the histograms excellently.

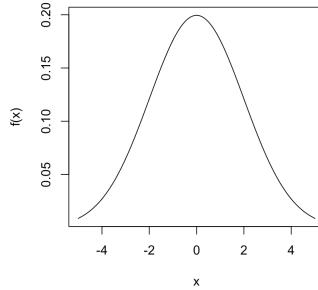


- Here is an overview of different population shapes and the distributions families that can be used to model them.
- If a population is distributed evenly between two numbers like



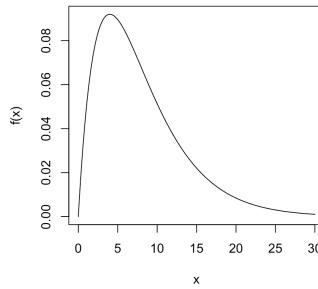
uniform distribution must be used.

- If a population is bell-shaped and symmetric like



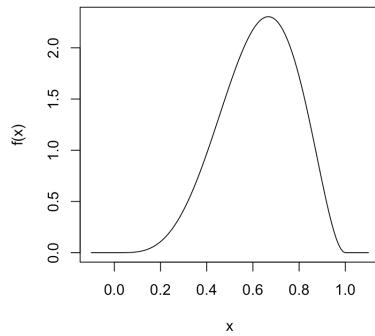
Normal or t distributions can be used, among others.

- If a population is skewed like



Gamma, lognormal or χ^2 distributions can be used, among others.

- If a population has bounded range (support) between two points and is not evenly distributed



Beta distribution can be used. This is useful when modeling probabilities $0 \leq p \leq 1$.

Very cool tangent!

- We saw in the previous example that we could model the data with a gamma distribution, specifically:

$$Y \sim \text{Gamma}(\alpha = 5, \beta = 10)$$

- In practice, once the researcher selects the family (gamma), they would then have to estimate the parameter values using techniques such as maximum likelihood estimation (will learn this next semester!).
 - The goal would be to have: $\hat{\alpha} \approx 5$ and $\hat{\beta} \approx 10$.
 - This strategy would be in the context of Classical statistics, where parameter values are fixed constants; there is another branch of statistics called Bayesian statistics.
- In Bayes, parameters are considered random variables and can have their own probability distributions. For example: $\alpha \sim \text{Exponential}(3)$ and $\beta \sim \text{Uniform}(1, 5)$.

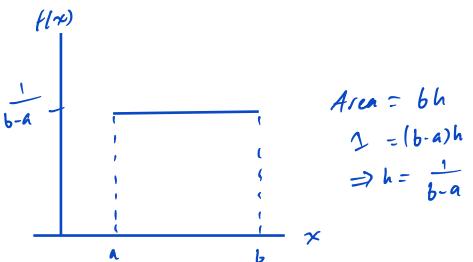
Uniform distribution

Definition

- The continuous uniform distribution is defined by spreading probability uniformly over an interval $[a, b]$ (X can also be thought of as the outcome when a point is randomly selected from the interval $[a, b]$).
- If $X \sim \text{Uniform}(a, b)$

$$f(x | a, b) = \begin{cases} \frac{1}{b-a} & a \leq x \leq b \\ 0 & \text{otherwise} \end{cases}$$

where the parameters a and b are real numbers.



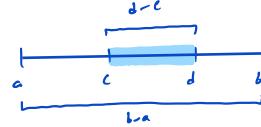
- Characteristics of a uniform distribution.
 - Constant probability over the entire interval.
 - Bounded support. $a, b \neq \infty$
 - Symmetric. \Rightarrow mean = median

Probabilities

- The probability of any subinterval in the range (support) is proportional only to the length of the subinterval.

• For $a \leq c \leq d \leq b$

$$P(c \leq X \leq d) = \frac{d-c}{b-a}$$



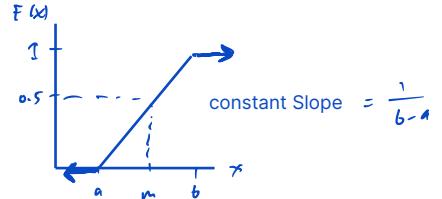
- We can generalize this to find $P(X \leq x)$ for values of x in the interval $[a, b]$.

$$P(X \leq x) = P(a \leq X \leq x) = \frac{x-a}{b-a}$$



- Then we can define the cdf $F(x)$ for a uniform random variable X on $[a, b]$ by:

$$F(x | a, b) = \begin{cases} 0 & x < a \\ \frac{x-a}{b-a} & a \leq x \leq b \\ 1 & x > b \end{cases}$$



Lifetime random variables and survival functions

- In many applied problems, the random variable interest is a time variable T . This time could represent:
 - Time until death of a person (a standard insurance application).
 - Time until the machine part fails.
 - Time until a disease ends.
 - Time it takes to serve a customer in a store.
- The uniform distribution doesn't give a very realistic model of human lifetimes, but is often used as an illustration of a lifetime model because of its simplicity.
- Example: Let T be the time from birth until death of a randomly selected member of a population. Assume that T has a uniform distribution on $[0, 100]$. Then

$$f(t) = \begin{cases} 1/100 & 0 \leq t \leq 1 \\ 0 & \text{otherwise} \end{cases} \quad \text{and} \quad F(t) = \begin{cases} 0 & t < 0 \\ t/100 & 0 \leq t \leq 1 \\ 1 & t > 100 \end{cases}$$

- The function $F(t)$ gives us the probability that the person dies by age t .
- For example: The probability of death by age 57 is

$$P(T \leq 57) = F(57) = \frac{57}{100} = 0.57$$

- Most of us are interested in the probability that we will survive past a certain age. In this example, we might wish to find the probability that we survive beyond age 57. This is simply the probability that we do *not* die by age 57.

$$P(T > 57) = 1 - F(57) = 1 - \frac{57}{100} = 0.43$$

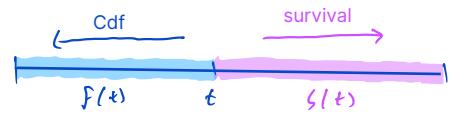
- The probability of surviving from birth past a given age t is called a survival probability and denoted by $S(t)$.

- Definition: The **survival function** is

$$S(t) = P(T > t) = 1 - F(t)$$

\approx Complement of cdf

Right probability = 1 - left probability



- The survival function for a uniform random variable is

$$\begin{aligned} S(t) &= 1 - F(t) \\ &= 1 - \frac{t-a}{b-a} \\ &= \frac{b-a - (t-a)}{b-a} \\ &= \frac{b-t}{b-a}, \quad a \leq t \leq b \end{aligned}$$

Mean and variance

- If $X \sim \text{Uniform}(a, b)$

$$E(X) = \frac{a+b}{2}$$

$$V(X) = \frac{(b-a)^2}{12}$$

$$\Rightarrow \text{SD}(X) = \sqrt{\frac{(b-a)^2}{12}}$$

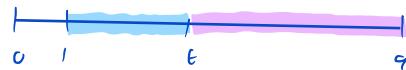
= median m (midpoint of interval)

Summarizing example

- Let T be the time in months from initial use of a machine part until failure, where $T \sim \text{Uniform}(1, 9)$.

- Find the pdf, cdf and survival function of T .

$$f(t) = \begin{cases} \frac{1}{8} & 1 \leq t \leq 9 \\ 0 & \text{otherwise} \end{cases}$$



$$f(t) = \frac{t-1}{8}, \quad 1 \leq t \leq 9$$

$$S(t) = \frac{9-t}{8}, \quad 1 \leq t \leq 9$$

b) Find the probability the machine part fails between 5 and 7 months.

$$P(5 \leq T \leq 7) = \frac{7-5}{8} = \frac{2}{8}$$



c) Find the probability the machine part lasts longer than 4 months.

$$P(T \geq 4) = S(4) = \frac{9-4}{8} = \frac{5}{8}$$



d) Find the expected value and standard deviation for the lifetime of the machine part.

$$E(T) = \frac{1+9}{2} = 5$$

$$V(T) = \frac{(9-1)^2}{12} = \frac{64}{12} \Rightarrow SD(T) = \sqrt{\frac{64}{12}}$$

Exponential distribution

(Brief) Motivation

- Recall when observing a Poisson process, we counted the number of occurrences in a given interval. This number was a discrete random variable with a Poisson distribution.



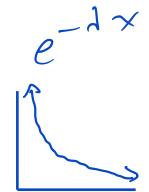
- But not only is the number of occurrences a random variable, the waiting times between successive occurrences are also random variables. These are continuous and follow an exponential distribution.
- Formal statement: It can be shown that the exponential distribution gives the probability for the waiting time between successive Poisson events.

Derivation of density

- The main part of the density function is an exponential decay function with parameter λ , which represents average number of events occurring (i.e. average rate of events) per unit of time in a Poisson process.
- To get this function to be a valid pdf, when integrated over the appropriate range $[0, \infty)$ it needs to equal 1.

Happens immediately \swarrow Never occurs \searrow

$$\begin{aligned} \int_0^\infty e^{-\lambda t} dt &= -\frac{1}{\lambda} e^{-\lambda t} \Big|_0^\infty = f(\infty) - f(0) \\ &= 0 - \left(-\frac{1}{\lambda}\right) \\ \lambda \times [] &= \left[\frac{1}{\lambda}\right] \times 1 \quad \text{need to multiply through by } \lambda \\ \checkmark 1 &\implies \int_0^\infty \lambda e^{-\lambda t} dt = 1 \end{aligned}$$

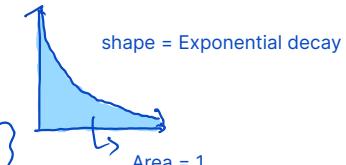


This process is the same as finding c such that:

$$\int_0^\infty c e^{-\lambda t} dt = 1$$

$$\downarrow = \lambda$$

- The value c is called a **normalizing constant**.
 - All of the distributions we will study from now on have a normalizing constant, some of which are quite complex.
 - The purpose is the same though, functions of x determine the shape of a function (e.g. $e^{-\lambda x}$), then c converts it to a valid pdf.



Definition

If $T \sim \text{Exponential}(\lambda)$

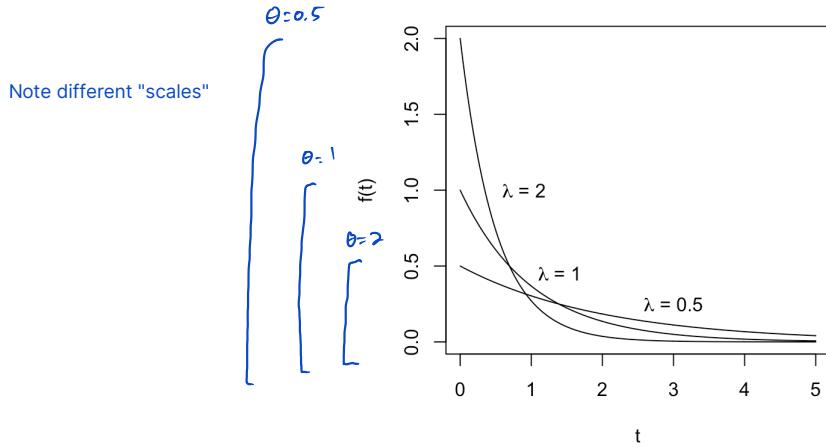
$$f(t | \lambda) = \lambda e^{-\lambda t}, \quad t \geq 0 \quad \text{and} \quad \lambda > 0$$

- Characteristics of an exponential distribution.
 - Gives the probability for the waiting time between successive Poisson events.
 - Right-skewed density function.
 - Unbounded support. $0 \leq t < \infty$
 - The exponential random variable is a continuous version of the geometric random variable, which waits for the first Success in a discrete sample space.
 - Also has the memoryless property: $P(T > a + b | T > a) = P(T > b)$.
- Exponential and geometric are the only distributions with the memoryless property.

() 1 1
F F F S

Parameter

- As stated, λ is the rate at which events occur in a Poisson process. Here is how it affects the exponential density function.
- As λ increases, events happen more often in a time interval, which consequently means we are waiting less and less for the next event to occur.



We see that for larger values of λ , there is more probability close to zero and less in the tail.

- Just like with some of the discrete distributions, there are different versions of the exponential distribution.
 - All versions have a random variable that is giving probabilities for waiting times. So they are not different in the way that we had two different versions of the geometric random variable (counting number of trials vs number of failures).
 - Rather they differ in what the parameters represent.
 - The definition given above uses the **rate parameterization** of the exponential distribution. $\text{rate} = \lambda$
 - There is also a **scale parameterization**, where the scale parameter θ is defined by

$$\text{scale } \theta = \frac{1}{\text{rate}} = \frac{1}{\lambda} \quad \Rightarrow \quad T \sim \text{Exp}(\theta = \frac{1}{\lambda}) \quad \text{and} \quad f(t) = \frac{1}{\lambda} e^{-t/\lambda}$$
- This obviously will impact the formulas for the cdf, mean, variance, etc.

Probabilities

- Probabilities for the exponential distribution are easy to solve by hand (unlike the next distributions).
- Example: Accidents at a busy intersection occur at an average rate of $\lambda = 2$ per month according to a Poisson process. Let T be the random variable for the time between accidents.

a) Find $f(t)$. $f(t) = 2e^{-2t}$, $t \geq 0$

- b) Cdf $P(T \leq t)$: Find the probability that the waiting time for the next accident is less than 2.

$$P(T \leq 2) = \int_0^2 2e^{-2t} dt = -e^{-2t} \Big|_0^2 = e^{-2(0)} - e^{-2(2)} = 1 - e^{-4}$$

- c) Survival $P(T > t)$: Find the probability that the waiting time for the next accident is longer than 1 month.

$$P(T > 1) = \int_1^\infty 2e^{-2t} dt = -e^{-2t} \Big|_1^\infty = e^{-2(1)} - e^{-2(\infty)} = e^{-2} = e^{-2}$$

- d) Interval $P(a \leq T \leq b)$: Find the probability that the waiting time for the next accident is between 0.5 and 1.5 months

$$P(0.5 \leq T \leq 1.5) = \int_{0.5}^{1.5} 2e^{-2t} dt = -e^{-2t} \Big|_{0.5}^{1.5} = e^{-2(0.5)} - e^{-2(1.5)} = e^{-1} - e^{-3}$$

- Cdf and survival function:

If $T \sim \text{Exponential}(\lambda)$

$$F(t) = 1 - e^{-\lambda t}$$

$$S(t) = 1 - F(t)$$

$$\begin{aligned} &= 1 - (1 - e^{-\lambda t}) \\ &= e^{-\lambda t} \end{aligned}$$

Note that we could derive these formally:

$$F(t) = \int_0^t f(u) du \quad \text{and} \quad S(t) = \int_t^\infty f(u) du$$

- Using the cdf: Recall once the cdf $F(x)$ is known for a random variable X , it can be used to find the probability that X lies in any interval since

$$P(a \leq X \leq b) = P(X \leq b) - P(X \leq a) = F(b) - F(a)$$

For the exponential distribution, we have:

$$P(a \leq T \leq b) = F(b) - F(a) = (1 - e^{-\lambda b}) - (1 - e^{-\lambda a}) = e^{-\lambda a} - e^{-\lambda b}$$

Mean and variance

scale parameterization
 $E(T) = \theta$
 $V(T) = \theta^2$

- If $T \sim \text{Exponential}(\lambda)$

$$E(T) = \frac{1}{\lambda}$$

$$V(T) = \frac{1}{\lambda^2}$$

$$\Rightarrow SD(T) = \sqrt{\frac{1}{\lambda^2}} = \frac{1}{\lambda} = E(T)$$

- Example: Find the mean and variance for the previous accident example: $T \sim \text{Exp}(\lambda=2)$

$$E(T) = \frac{1}{2} \text{ months}$$

$$V(T) = \frac{1}{2^2} = \frac{1}{4} \text{ months}^2$$

Memoryless property

- Just like the geometric distribution, the exponential distribution has the **memoryless property**.

- Theorem: If $T \sim \text{Exponential}(\lambda)$

$$P(T > a + b | T > a) = P(T > b)$$

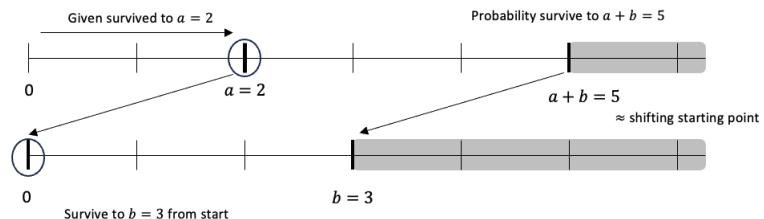
- (Using a concrete example to help conceptualize) The probability that a bulb will operate for more than time $a + b$ time units, given that has already operated for at least a time units, is the same as the probability that a new bulb will operate for at least b time units.

In practical terms, this means that the length of time a light bulb has already operated does not affect its chances of operating for additional time units.

Example: $a = 2, b = 3, a + b = 5$

$$P(T > 2 + 3 | T > 2)$$

equivalent to
 \Leftrightarrow
 $P(T > 3)$



- Proof:

$$P(T > a+b | T > a) = \frac{P(T > a+b)}{P(T > a)} = \frac{S(a+b)}{S(a)} = \frac{e^{-\lambda(a+b)}}{\cancel{e^{-\lambda a}}} = e^{-\lambda b} = S(b) \stackrel{\checkmark}{=} P(T > b)$$

- Example: Let T be the time to failure of a machine part, where $T \sim \text{Exponential}(\lambda = 0.001)$.

- (a) Given that the part has operated 100 hours, find the probability it will operate for 150 hours.

$$P(T > 150 | T > 100) = P(T > 50) = e^{-0.001 \cdot 50}$$

- (b) Given that the part has operated 100 hours, find the probability it will operate for $100 + x$ hours ($0 \leq x < \infty$).

$$P(T > 100+x | T > 100) = P(T > x) = e^{-0.001 \cdot x} = S(x)$$

- Note that survival functions are unique, just like cdfs and pdfs.

And the final expression in part (b) is the survival function $S(x)$ for a random variable which is exponentially distributed on $[0, \infty)$ with $\lambda = 0.001$; so it must have this distribution. This has a nice intuitive interpretation:

- Lifetime of a new part $\sim \text{Exp}(\lambda = 0.001)$ on $[0, \infty)$
- Remaining lifetime of a 100 hr old part $\sim \text{Exp}(\lambda = 0.001)$ on $[0, \infty)$

Waiting time for a Poisson process

- In the intro, we stated that the exponential distribution gives the waiting time between Poisson events. Here's why.
- If the number of events in a time period of length 1 is a Poisson random variable with parameter λ , then the number of events in a time period of length t is a Poisson random variable with parameter $\underline{\lambda t}$. Let this be the random variable X .

This is a reasonable assumption (think about the accidents in 1 month vs 3 months example).

- We can find the probability of no accidents in an interval of length t easily:

$$P(X=0) = \frac{e^{-\lambda t} (\cancel{\lambda t})^0}{\cancel{t!}} = e^{-\lambda t}$$

However, no accidents in an interval of length t is equivalent to saying the waiting time T for the next accident is greater than t . Thus

$$P(X=0) = P(T > t) = S(t) = e^{-\lambda t} \implies T \sim \text{Exp}(\lambda)$$

- Example: Customers arrive in a certain shop according to an approximate Poisson process at a mean rate of 40 per 2 hours. What is the probability that the shopkeeper will have to wait more than 5 minutes for the arrival of the first customer?

$$\rightarrow \lambda = \frac{40 \text{ customers}}{2 \text{ hours}} = \frac{20 \text{ customers}}{1 \text{ hour}} * \frac{1 \text{ hour}}{60 \text{ min}} = \frac{1}{3} \Rightarrow T \sim \text{Exp}(\lambda = \frac{1}{3})$$

$$\rightarrow P(T > 5) = S(5) = e^{-\frac{1}{3}(5)} \approx 0.1889$$

Last example

- The lifetime of a machine has an exponential distribution with a mean of 3 years. The manufacturer is considering offering a warranty and considers two types of warranties.

- ω_1 – Warranty 1 pays 3 if the machine fails in the first year, 2 if the machine fails in the second year, and 1 if the machine fails in the third year, with no payment if the machine fails after 3 years.
- ω_2 – Warranty 2 pays $3e^{-T}$ at time T years.

Find the expected warranty payment under each of the two warranties.

$$\rightarrow E(T) = \frac{1}{\lambda} = 3 \rightarrow T \sim \text{Exp}(\lambda = \frac{1}{3})$$

$$\Rightarrow \lambda = \frac{1}{3}$$

$$\rightarrow \omega_1 = \begin{cases} \rho(T \leq 1) & w_1 = 3 \\ \rho(1 \leq T \leq 2) & w_1 = 2 \\ \rho(2 \leq T \leq 3) & w_1 = 1 \end{cases} \Rightarrow E(\omega_1) = \sum w_i \rho(w_i)$$

$$= 3 \rho(T \leq 1) + 2 \rho(1 \leq T \leq 2) + 1 \rho(2 \leq T \leq 3)$$

$$= \downarrow + 2[\rho(T \leq 2) - \rho(T \leq 1)] + [\rho(T \leq 3) - \rho(T \leq 2)]$$

$$= \rho(T \leq 1) + \rho(T \leq 2) + \rho(T \leq 3)$$

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

$$\downarrow \simeq 1.4022$$

$$\rightarrow \omega_2 = 3e^{-T} \Rightarrow E(\omega_2) = \int_0^\infty \underbrace{3e^{-t} f(t)}_{w_2} dt$$

$$\downarrow$$

$$= \int_0^\infty 3e^{-t} \frac{1}{3} e^{-\frac{1}{3}t} dt$$

$$= \int_0^\infty e^{-\frac{4}{3}t} dt$$

$$= -\frac{3}{4} e^{-\frac{4}{3}t} \Big|_0^\infty F(\infty) - F(0)$$

$$= -\frac{3}{4} (0 - 1)$$

$$= 3/4$$

Continuous Distributions

Individual waiting times

$$X_i \sim \text{Exp}(\beta)$$

11-14

Gamma distribution

(Brief) Motivation

$$X = X_1 + X_2 + X_3 \\ \downarrow \sim \text{Gamma}(\alpha=3, \beta)$$

- Relationship between the exponential distribution and the gamma distribution:

Recall

Geometric = number of trials for 1st success

analogous
same relationship

Exponential = waiting time for first event

negative Binomial = number of trials for r successes

for a poisson process

$$\sum_{i=1}^r \text{Geo}(p) = NB(r, p)$$

Gamma = waiting time for α^{th} event

$$\sum_{i=1}^{\alpha} \text{Exp}(\beta) = \text{Gamma}(\alpha, \beta)$$

II events

- The gamma distribution can also be applied in other problems where the exponential distribution is useful like analysis of failure time of a machine part or survival time for a disease.

Definition

Scale parameterization

$$f(x) = \frac{x^{\alpha-1} e^{-\beta x}}{\beta^\alpha \Gamma(\alpha)}$$

- If $X \sim \text{Gamma}(\alpha, \beta)$

Rate parameterization

$$f(x | \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x}, \quad x \geq 0 \quad \text{and} \quad \alpha, \beta > 0$$

Normalizing constant main function

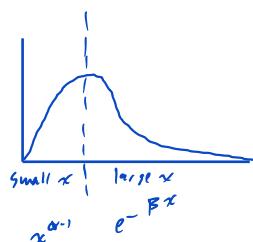
- Intuitive idea behind pdf: \rightarrow setting shape

– The main part of $f(x)$ is $x^{\alpha-1} e^{-\beta x}$.

As $x \rightarrow \infty$, $x^{\alpha-1} \rightarrow \infty$ for $\alpha > 1$.

As $x \rightarrow \infty$, $e^{-\beta x} \rightarrow 0$ for $\beta > 0$.

When x is small, $x^{\alpha-1} e^{-\beta x}$ is dominated by $x^{\alpha-1}$; when x is large, it is dominated by $e^{-\beta x}$. Thus as $x \rightarrow \infty$, $x^{\alpha-1} e^{-\beta x} \rightarrow 0$.



- Once we model the shape of the density function, we need to find the constant c such that

$$\int_0^\infty c x^{\alpha-1} e^{-\beta x} dx = 1$$

By some calculus, we can show that $c = \frac{\Gamma(\alpha)}{\beta^\alpha}$, where

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1} e^{-x} dx, \quad \alpha > 0$$

- The quantity $\Gamma(\alpha)$ is known as the **gamma function**. Here is a summary of the $\Gamma(\alpha)$:
- Gives a value for any $\alpha > 0$.
- We can think of the gamma function as an extension of the factorial definition from the positive integers to all positive real numbers.

Integration by parts can show

$$\Gamma(t) = (t-1) \Gamma(t-1)$$

ex) $\Gamma(2.5) = 1.5 \Gamma(1.5) = 1.5 \cdot 0.5 \Gamma(0.5) \approx 1.329$

- Whenever $\alpha = n$, where n is a positive integer, repeated integration by parts shows that

$$\begin{aligned} \Gamma(n) &= (n-1) \Gamma(n-1) = (n-1)(n-2) \cdots (2)(1) \Gamma(1) \\ &\downarrow \\ &\int_0^\infty e^{-x} dx = 1 \\ &\hookrightarrow \sim \text{Exp}(A=1) \end{aligned}$$

Thus, when n is a positive integer, we have

$$\Gamma(n) = (n-1)!, \quad n = 1, 2, \dots$$

- The pdf for the gamma distribution is essentially the gamma function as defined above, except it introduces another parameter in the exponential term, β .

- Characteristics of a gamma distribution:

- Gives the probability for the waiting time until the k^{th} occurrence in a Poisson process.
- Right-skewed density function.
- Unbounded support. $0 \leq x < \infty$

- Two important special cases of the gamma distribution:

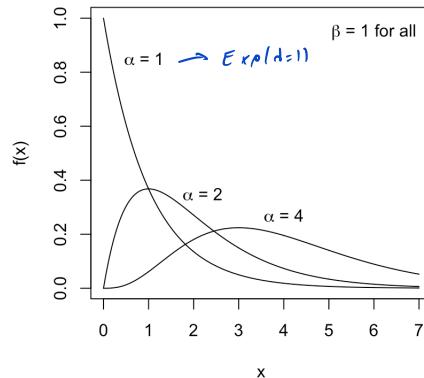
1. When $\alpha = 1 \rightarrow X \sim \text{Exponential}(\beta)$. Thus gamma can be considered a generalized version of the exponential random variable.

$$f(x) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} = \beta e^{-\beta x} \underset{\text{rate}}{\sim} \text{Exp}(\lambda=\beta)$$

2. When $\alpha = r/2, \beta = 2 \rightarrow X \sim \chi^2(r)$ (read “Chi-squared”). This distribution is important in statistical inference, especially in analysis of variance (aka ANOVA), which is the basis of experimental design methods.

Parameters

- The parameters of the distributions we have studied so far have a direct interpretation, but here α and β do not. But we can still give some meaning to them. Here is how they affect the gamma density function.
- $\alpha = \text{shape}$ parameter.

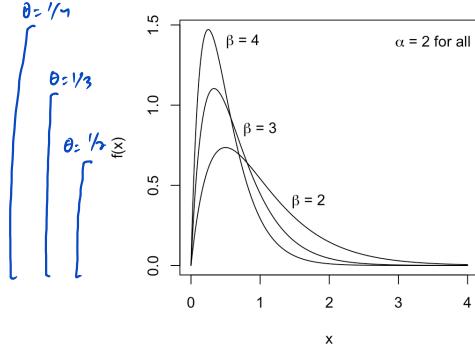


We see that for different values of α (and constant β), there are different Shapes of the gamma densities. Hence “shape” parameter.

Additionally, for larger values of α , the density increases longer out from zero because the polynomial part of the density function is more powerful for small x .

Notice that when $\alpha = 1$ we have the familiar negative exponential curve. All exponential densities have this shape because of this fixed α .

- $\beta = \text{rate}$ parameter (or $\theta = 1/\beta = \text{scale}$ parameter).



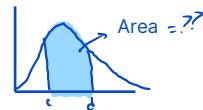
We see that for different values of β and constant α , the general Shapes of the gamma densities look similar, but the Scales are different. Hence "scale" parameter, which is equal to the inverse rate.

- Remember that the rate vs scale parameterizations will cause slight differences in the formulas for the pdf, mean, variance, etc. in other resources.

Probabilities

- Probabilities for the gamma distribution are difficult. If $X \sim \text{Gamma}(\alpha, \beta)$:
 - Can only be solved by hand if α is an integer.
 $\alpha = 1 \rightarrow \text{Exponential. } \checkmark$
 $\alpha = 2 \rightarrow \text{Requires integration by parts once. } \checkmark \text{ reasonable}$
 $\alpha > 2 \rightarrow \text{Requires repeated integration by parts } \alpha - 1 \text{ times. } \text{Possible, but not practical}$
 - If α is not an integer and $0 < c < d < \infty$, cannot calculate $P(c < X < d)$ by integration because it is impossible to give a closed-form expression for

$$\int_c^d \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} dx$$



Thus we must use statistical software to compute the probability that a gamma random variable falls in an interval.

- Software do implement a general gamma cdf, but not anything one needs to know.

Mean and variance

Scale parameterization
 $E(X) = \alpha\theta$
 $V(X) = \alpha\theta^2$

- If $X \sim \text{Gamma}(\alpha, \beta) \quad \alpha, \beta > 0$

$$E(X) = \frac{\alpha}{\beta}$$

$$V(X) = \frac{\alpha}{\beta^2} \quad \Rightarrow \quad \text{sd}(X) = \sqrt{\frac{\alpha}{\beta^2}}$$

- If $\alpha = 1$, $E(X)$ and $V(X)$ for exponential follow.

$$= \frac{1}{\beta} \quad = \frac{1}{\beta^2}$$

Summarizing examples:

1. Let X be random variable for the waiting time (in months) from the start of observation until the second accident at an intersection. Assume $X \sim \text{Gamma}(\alpha = 2, \beta = 3)$.

- (a) Find the pdf of X .

$$f(x) = \frac{3^2}{\Gamma(2)} x^{2-1} e^{-3x} = 2x e^{-3x}, \quad 0 \leq x < \infty$$

- (b) Find $E(X)$ and $V(X)$. $E(X) = \frac{\alpha}{\beta} = \frac{2}{3}$ months

$$V(X) = \frac{\alpha}{\beta^2} = \frac{2}{9} \text{ months}^2$$

- (c) Find the probability the total waiting time for the second accident is between 1 and 2 months.

$$\begin{aligned} P(1 \leq X \leq 2) &= \int_1^2 f(x) dx \\ &= \int_1^2 2x e^{-3x} dx \\ &= \left[2x \left(-\frac{1}{3} e^{-3x} \right) - \int -\frac{1}{3} e^{-3x} (2) dx \right]_1^2 \\ &= \left[-3x e^{-3x} - e^{-3x} \right]_1^2 = F(2) - F(1) \\ &= \left(-3(2)e^{-3(2)} - e^{-3(2)} \right) - \left(-3(1)e^{-3(1)} - e^{-3(1)} \right) \\ &= -7e^{-6} - 4e^{-3} \\ &\approx 0.1818 \end{aligned}$$

Integration by parts
 $u = 2x \quad dv = e^{-3x}$
 $du = 2dx \quad v = -\frac{1}{3}e^{-3x}$
 $\left[uv - \int v du \right]_a^b$

2. Biologists investigating a stretch of desert find certain fossils according to a Poisson process at a mean rate of 500 per kilometer $1000m$.

What is the probability that the biologists will have to investigate more than 5 meters in order to find the first four fossils?

$$\rightarrow X \sim \text{Gamma}(\alpha = 4, \beta = \frac{500}{1000} \text{ meters})$$

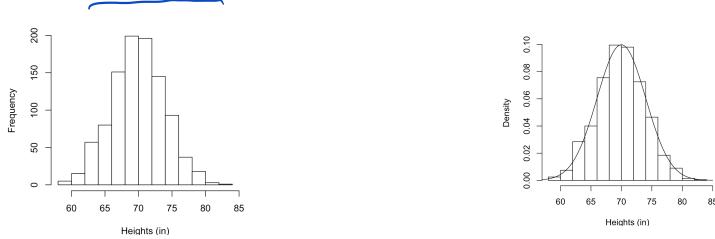
$$\rightarrow P(X > 5) = \int_5^\infty f(x) dx = 1 - \text{gammacdf}(\alpha = 4, \beta = 0.5, x = 5) \approx 0.2576$$

* If there was a calculator function

Normal distribution

Applications

- The **normal distribution** is the most widely-used of all the distributions we have and will discuss. It can be used to model a wide range of natural phenomena that follow the “bell-shaped” pattern in their relative frequency distribution (histogram).



- Examples include variables such as test scores, physical measurements (height, weight, length) of organisms, and repeated measurements of the same quantity on different occasions or by different observers (measurement error), stock portfolio returns, insurance portfolio losses, etc.
- Every normal density curve has this shape, and the normal density model is used to find probabilities for all of the natural phenomena who histograms display this pattern.
- Random variables whose histograms are well-approximated by a normal density curve are called approximately normal.

Definition

If $X \sim \text{Normal}(\mu, \sigma^2)$

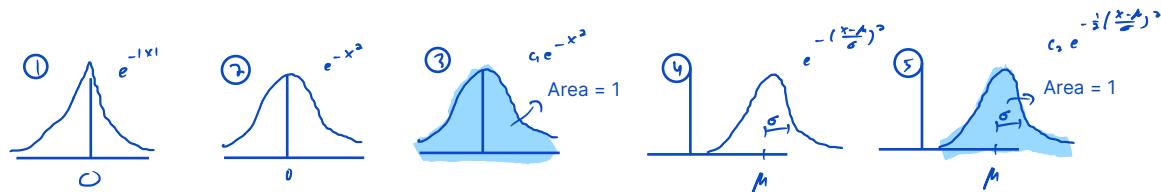
$$f(x | \mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right], \quad -\infty < x < \infty \quad \text{and} \quad -\infty < \mu < \infty, \quad \sigma > 0$$

Notation

$$\exp[a] = e^a$$

$$= \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right]$$

- Intuitive idea behind pdf:
 - Set the shape → Bell-shaped and symmetric (exponential decay from center in both directions)
 - Change to smooth curve
 - Find constant to make valid pdf.
 - Add location and scale parameters to shift and scale the density, respectively.
 - Adjust constant to take into account new parameters.



- Characteristics of a normal distribution.

– Density function is bell-shaped and symmetric.

– Unbounded support (range). $-\infty < x < \infty$

But the density decreases exponentially as x runs away from the center. Thus, the normal distribution is also used for bounded data in practice (e.g. test scores).

$\rightarrow ACT = [0, 16]$

- The density function for the normal distribution is difficult to integrate as we will see. For example, to show that the normal distribution is a valid pdf, we need to use polar coordinates.

$$\int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right] dx \stackrel{?}{=} 1$$

Parameters, expected value and variance

- The normal distribution is somewhat special in the sense that it's two parameters provide us with complete information about the exact shape (the spread) and location (where it's centered) of the distribution.

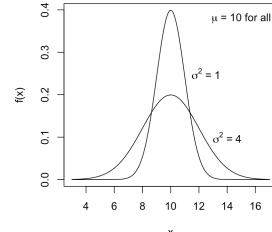
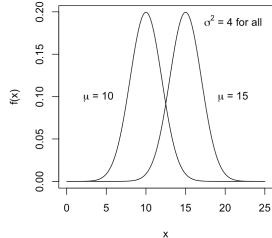
- If $X \sim \text{Normal}(\mu, \sigma^2)$

$$E(X) = \mu$$

$$V(X) = \sigma^2$$

$$SD(X) = \sigma$$

- These are really easy to derive using ~~MGF~~. Moment generating functions (will learn later).
- Here is how they affect the normal density function.



- In practice, often the standard deviation σ is used rather than the variance σ^2 . So it is common to see $X \sim \text{Normal}(\mu, \sigma)$.

Probabilities

- Suppose we are looking at a national exam whose scores X are approximately normal with $\mu = 500$ and $\sigma = 100$. To find the probability a score is between 600 and 750, we have to evaluate the integral:

$$\begin{aligned} &\rightarrow X \sim \text{normal}(\mu = 500, \sigma = 100) \\ &\rightarrow P(600 < X < 750) = \int_{600}^{750} \frac{1}{100\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-500}{100}\right)^2\right] dx \stackrel{?}{=} ? \end{aligned}$$

- This cannot be done in closed-form (just like the gamma distribution), but it can be approximated using numerical methods using software, such as TI-84s as shown next.

Continuing ex) $\rightarrow X \sim \text{normal} (\mu = 500, \sigma = 100)$

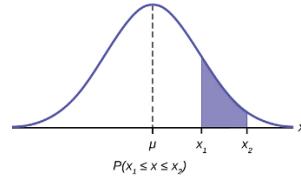
$$\rightarrow P(600 < X < 750) = \text{NormalCdf} \left\{ \begin{array}{l} \text{lower} = 600 \\ \text{upper} = 750 \\ \mu = 500 \\ \sigma = 100 \end{array} \right\} \approx 0.1524$$

★ show work using calculator

11-21

Finding Probabilities

- Choose correct Dist: 2ND \rightarrow VARS
 - ALWAYS want **normalcdf()**
 - Enter in information (endpoints and parameters)
 - lower = lower boundary
 - upper = upper boundary
 - μ = Mean
 - σ = SD
- If you have TI-83, you would type `normalcdf(lower, upper, mean, sd)`



```
NORMAL FLOAT AUTO REAL RADIAN HP
1:normalpdf
2:normalcdf
3:invNorm
4:invT
5:tndf
6:tcdf
7:zndf
8:zcdf
9:randInt
```

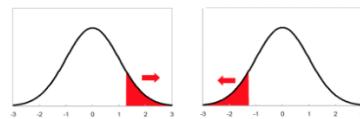
Ex) If $X \sim \text{Normal}(\mu = 40, \sigma = 5)$,
 $P(26 < X < 39)$

```
NORMAL FLOAT AUTO REAL RADIAN HP
normalcdf
lower:26
upper:39
μ:40
σ:5
Paste
```

```
NORMAL FLOAT AUTO REAL RADIAN HP
normalcdf(26,39,40,5)
0.4181851216
```

*** Special Cases

If finding a left tailed probability, enter lower = -10000 (some really big negative number)
 If finding a right tailed probability, enter upper = 10000 or 10^6 (really big positive number)



- In the olden days before such software was readily available, another way of finding normal probabilities involving tables of areas for a standard normal distribution was developed.

NOTE: Exam P formula sheet only gives LEFT probabilities for POSITIVE Z values
 \Rightarrow We need to use properties of the normal curve to find probabilities for $-Z$. So you have to know this way.

Entries represent the area under the standardized normal distribution from $-\infty$ to z , $\Pr(Z < z)$

The value of z to the first decimal is given in the left column. The second decimal place is given in the top row.

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879



$$\text{Left of } -z = \text{Right of } +z$$

$$\begin{aligned} P(Z < -0.12) &= P(Z > 0.12) \\ &= 1 - P(Z < 0.12) \\ &= 1 - 0.5478 \\ &= 0.4522 \end{aligned}$$

- We will now cover this method, and the basic properties of the normal distribution which are behind it, in a series of steps.

- Step 1: Linear transformation of normal random variables

– Theorem: If $X \sim \text{Normal}(\mu, \sigma^2)$ and $Y = aX + b$. Then

$$Y \sim \text{Normal}(a\mu + b, a^2\sigma^2)$$

$$E(Y) = E(aX + b) = aE(X) + b$$

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

– It is easy to show that the mean and variance of Y using properties we have learned before.

★ – But the crucial statement is that Y is also normally distributed. ★

This is actually really easy to show with moment generating functions (~~(M.G.F.)~~) and can also be shown using the pdf and a transformation of X .

– Example: If $X \sim \text{Normal}(\mu = 10, \sigma^2 = 4)$, find the distribution of $Y = 0.5X - 3$.

$$\begin{aligned} E(Y) &= E(0.5X - 3) = 0.5 \underbrace{E(X)}_{\mu=10} - 3 = 2 \\ \text{Var}(Y) &= \text{Var}(0.5X - 3) = 0.5^2 \underbrace{\text{Var}(X)}_{\sigma^2=4} = 1 \end{aligned} \quad \Rightarrow Y \sim \text{Normal}(\mu = 2, \sigma^2 = 1)$$

- Step 2: Transformation to a standard normal

– Using the linear transformation property of normal random variables, we can transform any normal random variable X with μ and standard deviation σ into a standard normal random variable with mean 0 and standard deviation 1.

– The transformation used to do this is:

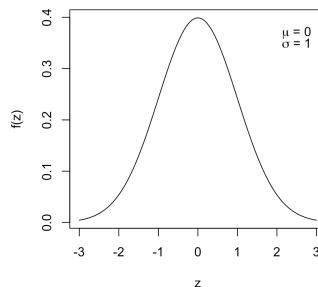
$$Z = \frac{X - \mu}{\sigma} = \frac{1}{\sigma} X - \frac{\mu}{\sigma}$$

– Using the previous theorem, we know $Z \sim \text{Normal}$ and can easily confirm the mean and standard deviation.

$$E(Z) = E\left(\frac{1}{\sigma}X - \frac{\mu}{\sigma}\right) = \frac{1}{\sigma} \underbrace{E(X)}_{\mu} - \frac{\mu}{\sigma} = 0 \quad \left| \quad \text{Var}(Z) = \text{Var}\left(\frac{1}{\sigma}X - \frac{\mu}{\sigma}\right) = \frac{1}{\sigma^2} \underbrace{\text{Var}(X)}_{\sigma^2} = 1 \right.$$

– Thus, if we have the standard normal random variable $Z \sim \text{Normal}(0, 1)$

$$f(z) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}z^2\right], \quad -\infty < z < \infty$$

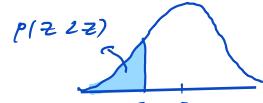


- This density function equation looks simpler, but still requires numerical integration.
- Forming $(X - \mu)/\sigma$ is known as **standardizing** X . Note that we can standardize any random variable, not just normals.
- Example: Suppose $X \sim \text{Exponential}(\lambda)$. $\Rightarrow E(X) = \frac{\lambda}{\lambda} = \lambda$ $V(X) = \frac{\lambda^2}{\lambda} = \lambda^2$

$$Z = \frac{X - \lambda}{\sigma} = \frac{X - \lambda}{\sqrt{\lambda}} = \lambda X - 1 \rightarrow E(Z) = \frac{\lambda}{\sqrt{\lambda}} = \sqrt{\lambda}$$

However Z doesn't necessarily have the same distribution of X . $Z \sim ?$

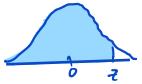
• Step 3: Using z-tables



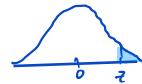
- Tables of areas under the density curve for the distribution of Z have been constructed for use in probability calculations.



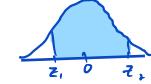
- The table gives values for the cdf of Z , $F_Z(z) = P(Z \leq z)$. **LEFT probabilities**
- Can find any probability we would like using the cdf:



$$\text{Cdf (left)} \quad P(Z \leq z) = \text{Table}$$

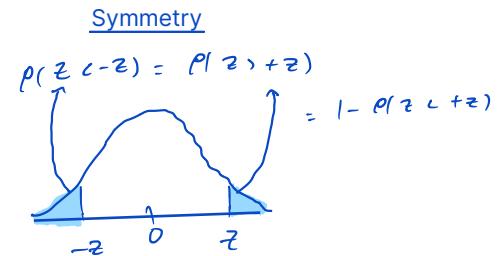


$$\text{Survival (right)} \quad P(Z > z) = 1 - P(Z \leq z)$$

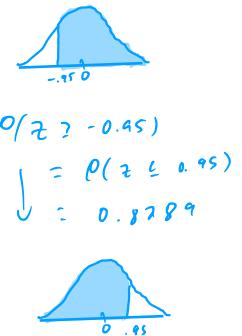


$$\text{Interval} \quad P(z_1 \leq Z \leq z_2) = F_z(z_2) - F_z(z_1)$$

$$\downarrow \quad = P(z \leq z_2) - P(z \leq z_1)$$



$$\begin{aligned} \text{- Examples: } P(-1.22 \leq Z \leq 2.1) &= P(z \leq 2.1) - P(z \leq -1.22) \\ &= 0.9821 - (1 - P(z \leq 1.22)) \\ &= 0.9821 - 0.0178 \quad \uparrow = 0.8898 \\ &= 0.8709 \end{aligned}$$



• Step 4: Finding probabilities for any normal X

- Once we know how to find probabilities for Z , we can use the transformation in step 1 to find probabilities for any normal random variable X with mean μ and standard deviation σ using the identity:

$$P(x_1 \leq X \leq x_2) = P\left(\frac{x_1 - \mu}{\sigma} \leq \frac{X - \mu}{\sigma} \leq \frac{x_2 - \mu}{\sigma}\right) = P(z_1 \leq Z \leq z_2)$$

$$\text{where } z_1 = \frac{x_1 - \mu}{\sigma} \text{ and } z_2 = \frac{x_2 - \mu}{\sigma}.$$

Z-scores = distance from mean in standard deviation units

- National exam examples: If $X \sim \text{Normal}(\mu = 500, \sigma = 100)$. Find the following probabilities:

$$P(X \leq 800) = P\left(\frac{X - 500}{100} \leq \frac{800 - 500}{100}\right) = P(Z \leq 3) = 0.9987$$



$$\begin{aligned} P(600 \leq X \leq 750) &= P(1 \leq Z \leq 2.5) = P(Z \leq 2.5) - P(Z \leq 1) \\ z_1 &= \frac{600 - 500}{100} = 1 \\ z_2 &= \frac{750 - 500}{100} = 2.5 \\ &\downarrow \\ &= 0.9938 - 0.8413 \\ &= 0.1525 \\ (\text{calc answer } &\approx 0.1524) \end{aligned}$$

Percentiles

- We can also find percentiles of the standard normal distribution from the table.

- Recall for $0 \leq p \leq 1$ the **$100p^{th}$ percentile** of X is the number x_p defined by

$$F(x_p) = p \rightarrow F_Z(z_p) = p$$

- If X is a normal random variable with mean μ and standard deviation σ , then we can easily find x_p , using z_p and the basic relationship of X and Z .

$$z_p = \frac{x_p - \mu}{\sigma} \Rightarrow x_p = z_p \sigma + \mu$$

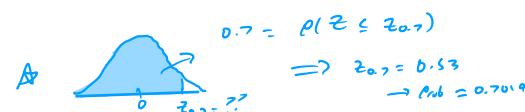
- Examples: If you scored in the 70^{th} percentile of test scores, what was your score?

$$\rightarrow X \sim \text{Normal}(\mu = 500, \sigma = 100)$$

$$\rightarrow F(z_{0.7}) = 0.7$$

↳ 0.53 (find 0.7 probability on z-table
(middle values))

$$x_{0.7} = \frac{100 z_{0.7} + 500}{100} = 553$$



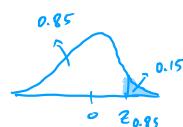
★ Closest Z to get probability at least 0.7
 $P(Z \leq z_{0.7}) \geq 0.7$

Find the test score corresponding to the top 15% of scores.

$$\rightarrow \text{Top 15\%} = z_{0.85} = z_{0.85}$$

$$\rightarrow F(z_{0.85}) = 0.85 \rightarrow x_{0.85} = \frac{100 z_{0.85} + 500}{100} = 604$$

↳ 1.04
gives prob = 0.8509208



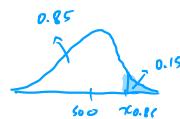
$$ex) \rightarrow P(X \leq x_{0.7}) = 0.7$$

$$\rightarrow x_{0.7} = \text{invNorm} \left(\begin{array}{l} \text{area} = 0.7 \\ \mu = 500 \\ \sigma = 100 \end{array} \right) \approx 552.44$$



$$\rightarrow P(V \leq x_{0.85}) = 0.85$$

$$\Rightarrow x_{0.85} = 607.64$$

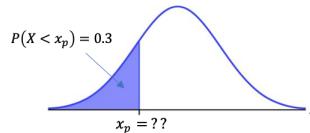


11-25

- This can also be done using software such as TI-84s.

Finding Percentiles

- Choose $\text{invNorm}()$: $2^{\text{ND}} \rightarrow \text{VARS}$
- Enter in information (area and parameters)
 - area = Probability (LEFT, percentile!)
 - μ = Mean
 - σ = SD



Ex) If $X \sim \text{Normal}(\mu = 10, \sigma = 1.5)$,
 $x_{0.30} \approx 9.21$



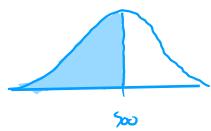
If you have TI-83, you would type $\text{invNorm}(\text{area}, \text{mean}, \text{sd})$

*** If you have a right tail (upper probability), you need to rewrite as probability to the left!

Harder examples

- Let $X \sim N(\mu, \sigma^2)$, $P(X < 500) = 0.5$ and $P(X > 650) = 0.0228$.

Find μ and σ^2 .



$$\rightarrow \text{symmetric} \Rightarrow \text{mean} = \text{median} = 500 = \mu$$

$$\rightarrow P(X > 650) = 0.0228$$

$$\Rightarrow P(X < 650) = 1 - 0.0228$$

$$P\left(\frac{X-500}{\sigma} < \frac{650-500}{\sigma}\right) = 0.9772$$

$$P(Z < 2) = 0.9772$$

$$\Rightarrow \frac{650-500}{\sigma} = 2 \Rightarrow \sigma = 75 \text{ and } \sigma^2 = 5625$$

- Let $X \sim N(\mu = 25, \sigma^2 = 36)$. Find c such that $P(|X - 25| \leq c) = 0.9544$

$$\begin{aligned} P(|X - 25| \leq c) &= P\left(-\frac{c}{6} \leq \frac{X-25}{6} \leq \frac{c}{6}\right) = P\left(-\frac{c}{6} \leq Z \leq \frac{c}{6}\right) \\ &= P(Z \leq \frac{c}{6}) - P(Z \leq -\frac{c}{6}) \\ &= \downarrow - (1 - P(Z \leq \frac{c}{6})) \\ &= 2P(Z \leq \frac{c}{6}) - 1 = 0.9544 \\ &\Rightarrow P(Z \leq \frac{c}{6}) = \frac{0.9544}{2} \\ &P(Z \leq z) = 0.9772 \\ &\Rightarrow \frac{c}{6} = 2 \Rightarrow c = 12 \end{aligned}$$

- If $X \sim N(\mu = 1, \sigma^2 = 16)$, find $P(X^2 - 4X < 21)$.

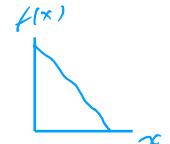
$$\begin{aligned} P(X^2 - 4X < 21) &= P(X^2 - 4X - 21 < 0) \\ &= P((X-2)(X+3) < 0) \\ &\quad \text{true when } \begin{cases} (+) (+) \text{ or } (-)(+) \\ x > 2 \Rightarrow X \\ x > -3 \Rightarrow X \\ x < 2 \Rightarrow X \end{cases} \\ &= P(-3 < X < 2) \\ &= P\left(-\frac{3-1}{4} \leq \frac{X-1}{4} \leq \frac{2-1}{4}\right) \xrightarrow{\text{NormalCDF}} \begin{pmatrix} \text{lower} & = -1 & = 1 \\ \text{upper} & = 2 & = 0 \\ \mu & = 1 & = 1 \\ \sigma & = 4 & = 1 \end{pmatrix} \approx 0.7754 \\ &= P(-1 \leq Z \leq 1.5) \end{aligned}$$

Central Limit Theorem

Sums of independent, identically distribution random variables

- We will now demonstrate one reason why the normal distribution is so useful in applications.
- Motivating example: Recall the previous straight-line density example, where the random variable X represented the loss on a single warranty insurance policy. It was not normally distributed.

We found that $E(X) = \frac{100}{3}$ and $V(X) = \frac{5,000}{9}$ and were able to find probabilities for X . However, this information applies only to a single policy.



The company selling insurance has more than one policy and must look at its total book of business by adding up all of the losses on all policies.

- Generalizing this scenario: Suppose the loss on a single insurance policy follows a non-normal density X . Companies will have say 1,000 policies and be interested in the total loss S , which is the sum of the losses on all individual policies X_i :

$$S = X_1 + X_2 + \cdots + X_{1000}$$

- If we assume that all of the policies are *iid* (independent and follow the same distribution), then the **Central Limit Theorem (CLT)** shows the sum is approximately normal (even though the individual policies X_i are not) \Rightarrow which means we can use normal probability methods to find probabilities for the total loss.

↳ Normalcdf() and z-tables

- **Central Limit Theorem** Let X_1, \dots, X_n be independent random variables, all of which have the same probability distribution and thus the same mean μ and variance σ^2 . If n is large, the sum

$$S = X_1 + X_2 + \cdots + X_n$$

will be approximately normal with mean $n\mu$ and variance $n\sigma^2$.

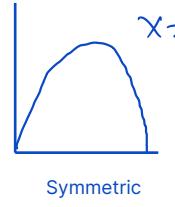
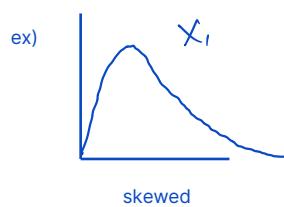
Written succinctly:

If $X_i \stackrel{iid}{\sim} f(x)$ with mean μ + variance σ^2
then $S = \sum_{i=1}^n X_i \sim \text{Normal}(\mu n, \sigma^2 n)$ if n is large

- Notes about CLT:

- This is a super powerful theorem!
- X_i do not have to be normally distributed. When this is the case, the CLT results in an approximately normal distribution.
- How large must n be? This depends on how close the original distribution is to the normal.





$n_1 \gg n_2$ for accurate CLT approximations

11-27

Some elementary statistics books define $n > 30$ as "large". This will not always be the case. For example, skewed distributions will require larger values of n compared to symmetric distributions in order for the results of the CLT to be decent.

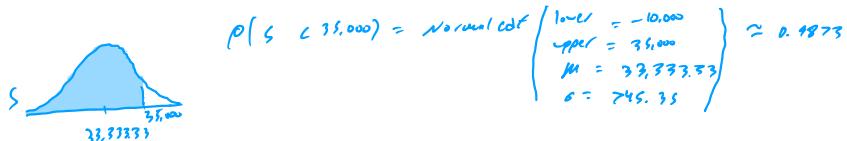
- In general, as n increases, approximations based on the CLT get better and better.
- If $X_i \stackrel{iid}{\sim} \text{Normal}(\mu, \sigma^2)$, then the CLT results in an exactly normal distribution for all n .

In theory
As $n \rightarrow \infty$
 \rightarrow Normal
exactly

- Return to motivating example: Find the distribution of the sum of losses S . $E(X) = \frac{100}{9}$ $V(X) = \frac{5000}{81}$

$$S = \sum_{i=1}^{1000} X_i \rightarrow S \sim \text{Normal} \left(\mu = 100 \left(\frac{100}{9} \right), \sigma^2 = \sqrt{100 \left(\frac{5000}{81} \right)} \right)$$

Find the probability total losses are less than \$35,000.



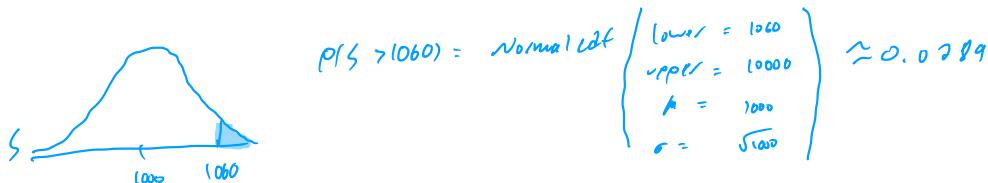
This shows the company that it is not likely to need more than \$35,000 to pay claims, which is helpful in planning.

More examples

1. Another example: Suppose the number of claims filed on for a particular policy follow a Poisson distribution with a mean of 2 claims per year and the company has a portfolio of 500 active policies this year, which are assumed to be independent.
 - (a) Find the distribution of the total number of filed claims for the entire portfolio.

$$\begin{aligned} \rightarrow X_i &\stackrel{iid}{\sim} \text{Poisson}(\lambda = ?) \Rightarrow S = \sum_{i=1}^{500} X_i \\ \Rightarrow E(X) &= V(X) = \lambda = 2 \qquad \qquad \downarrow \sim \text{Normal} \left(\mu = 500(2), \sigma^2 = \frac{500(2)}{500} = 1000 \right) \\ \star \text{By CLT} \star & \end{aligned}$$

- (b) Find the probability there will be more than 1060 claims this year.



2. Two instruments are used to measure the height, h , of a tower. The error made by the less accurate instrument is normally distributed with mean 0 and standard deviation $0.0056h$. The error made by the more accurate instrument is normally distributed with mean 0 and standard deviation $0.0044h$. Assuming the two measurements are independent random variables, what is the probability that their average value is within $.005h$ of 0?

$$\begin{aligned} & \rightarrow X_1 \sim \text{Normal}(\mu = 0, \sigma = 0.0056h) \quad \left. \begin{array}{l} \\ \end{array} \right\} X_1 \perp\!\!\!\perp X_2 \\ & X_2 \sim \text{Normal}(\mu = 0, \sigma = 0.0044h) \\ & \rightarrow P(-0.005h \leq \frac{X_1 + X_2}{2} \leq 0.005h) = P(-0.01h \leq \underbrace{X_1 + X_2}_{\hookrightarrow X_1 + X_2 \sim \text{Normal}(\mu = \mu_1 + \mu_2, \sigma^2 = \sigma_1^2 + \sigma_2^2)} \leq 0.01h) \\ & \qquad \qquad \qquad \downarrow = 0 + 0 \quad \downarrow = (0.0056h)^2 + (0.0044h)^2 \\ & \qquad \qquad \qquad = 0 \\ & = \text{Normal cdf} \left\{ \begin{array}{l} \text{lower} = -0.01h \\ \text{upper} = 0.01h \\ \mu = 0 \\ \sigma = \sqrt{(0.0056h)^2 + (0.0044h)^2} \\ h \text{ is cancel} \end{array} \right\} \\ & \qquad \qquad \qquad \approx 0.8347 \end{aligned}$$

Summary

- In general, normal distribution is quite valuable because it applies in so many situations where independent and identical components are being added.
 - Even though normal is continuous, the CLT works with discrete distributions as we saw above.
 - And the CLT showed us why so many random variables are approximately normally distributed.
- This occurs because many useful random variables are themselves sums of other independent random variables.

→ Contextually → Total claims = Sum of individual claims

Total losses = Sum of individual losses

$$\begin{aligned} \rightarrow \text{Distributions} \rightarrow & \sum_{i=1}^n \text{Bernoulli} \sim \text{Binomial}(n, p) \\ & \sum_{i=1}^n \text{Geometric} \sim \text{Negative Binomial}(r, p) \\ & \sum_{i=1}^n \text{Poisson}(\lambda_i) \sim \text{Poisson}(\lambda) \\ & \sum_{i=1}^n \text{Exponential} \sim \text{Gamma}(n, \beta) \end{aligned}$$

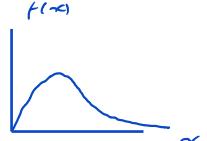
for large summations, we can approximate the ending distributions with normal

Lognormal distribution

(Brief) Motivation

- An alternative to the gamma distribution for asymmetric data is the lognormal distribution.

This distribution is more widely used than gamma to model skewed populations because we can take advantage of the properties of the normal distribution.



- Examples include insurance claim severity and investment returns.

Definition

- A random variable is called **lognormal** if its natural logarithm is normally distributed.

If $Y \sim \text{Lognormal} \iff \ln(Y) \sim \text{Normal}(\mu, \sigma^2)$

$$\ln(Y) = \text{Normal}(\mu, \sigma^2)$$

↓ = normal

- Stated another way: A random variable Y is lognormal if $Y = e^X$ for some normal random variable X with mean μ and variance σ^2 .

If $X \sim \text{Normal}(\mu, \sigma^2)$ and $Y = e^X \iff Y \sim \text{Lognormal}$

$$f(y | \mu, \sigma^2) = \frac{1}{y\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(\ln(y) - \mu)^2}{2\sigma^2}\right], \quad y \geq 0 \quad \text{and} \quad -\infty < \mu < \infty, \quad \sigma > 0$$

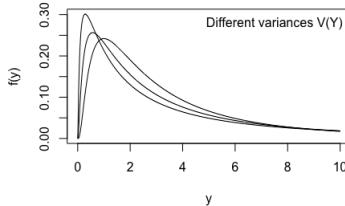
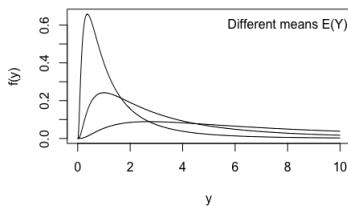
very similar to normal pdf except

- Characteristics of a lognormal distribution.

- Right-skewed density function.
- Unbounded support. $0 \leq y \leq \infty$

Parameters, expected value and variance

- Note that the parameters μ and σ^2 represent the mean and variance of the normal random variable X which appears in the exponent.
- Here is how they affect the lognormal density function.



- The mean and variance of the actual lognormal distribution Y are:

If $X \sim \text{Normal}(\mu, \sigma^2)$ and $Y = e^X$

$$E(Y) = e^{\mu + \frac{\sigma^2}{2}}$$

$$V(Y) = e^{2\mu + \sigma^2} (e^{\sigma^2} - 1) \Rightarrow \text{SD}(Y) = \sqrt{V(Y)}$$

Probabilities

- Just like with the normal, we cannot integrate the pdf, but we do not need to. The cdf can be found directly from the cdf for the normally distributed exponent.
- $F_Y(y) = P(Y \leq y) = P(e^X \leq y) = P[\ln(e^X) \leq \ln(y)] = P(X \leq \ln(y))$
- This means we just need to do algebra in the probability statement, then use normal-cdf() or z-tables like normal on the resulting number.

Example

- Let the claim severity $X \sim \text{Normal}(\mu = 7, \sigma^2 = 0.25)$ and $Y = e^X$.

- (a) Find $E(Y)$ and $V(Y)$.

$$E(Y) = e^{7 + \frac{0.25}{2}} = 1,242.65 \quad V(Y) = e^{2(7) + 0.25} (e^{0.25} - 1) = 438,524.80$$

- (b) Find the probability a claim is less than or equal to 1300.

$$\begin{aligned} P(Y \leq 1300) &= P(e^X \leq 1300) = P(X \leq \ln(1300)) = P(Z \leq \frac{\ln(1300) - 7}{0.5}) \\ &= \text{Normal CDF} \left(\begin{array}{l} \text{lower} = -10.00 \\ \text{upper} = \ln(1300) \\ \mu = 7 \\ \sigma = 0.5 \end{array} \right) \approx 0.6737 \end{aligned}$$

- (c) Find the probability a claim is between 900 and 1200.

$$\begin{aligned} P(900 \leq Y \leq 1200) &= P(900 \leq e^X \leq 1200) \\ &= P(\ln(900) \leq X \leq \ln(1200)) \\ &= \text{Normal CDF} \left(\begin{array}{l} \text{lower} = \ln(900) \\ \text{upper} = \ln(1200) \\ \mu = 7 \\ \sigma = 0.5 \end{array} \right) \\ &\approx 0.225 \end{aligned}$$

Beta distribution

(Brief) Motivation

- The beta distribution is defined on the interval $[0, 1]$.
- Thus it is often used as a model for probabilities such as the proportion of impurities in a chemical product or the proportion of time that a machine is under repair

Definition

- If $X \sim \text{Beta}(\alpha, \beta)$

$$f(x | \alpha, \beta) = \frac{1}{B(\alpha, \beta)} x^{\alpha-1} (1-x)^{\beta-1}, \quad 0 \leq x \leq 1 \quad \text{and} \quad \alpha, \beta > 0,$$

normalizing constant
main function

where

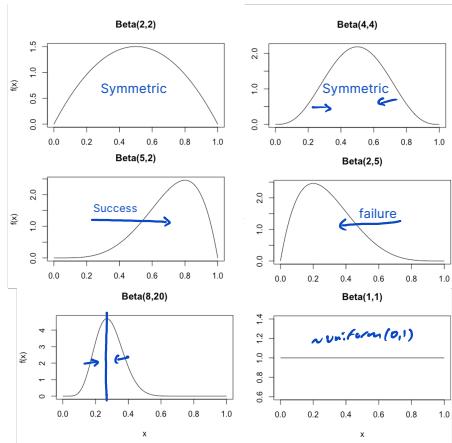
$$B(\alpha, \beta) = \int_0^1 x^{\alpha-1} (1-x)^{\beta-1} dx = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$$

- Intuitive idea behind pdf: Like the gamma distribution, we construct the formula as the function of x and then find the constant out from to make the density valid.
- Characteristics of a beta distribution.
 - In general, asymmetric density function.
 - Bounded support. $0 \leq x \leq 1$

Parameters, expected value and variance

- ~~The beta random variable X is often used as a model for probability and proportion. Thus, x represents the probability of success and $1 - x$ represents the probability of failure.~~

~~So, α and β sort of represent the chances (or weight) of success and failure, respectively.~~
- For example, when $\alpha = 5$ and $\beta = 2$, the chance of success is stronger than the chance of failure. So the mode and mean lean towards 1. In addition, as the parameters increase, the distributions tighten around the expectations.



- If $X \sim \text{Beta}(\alpha, \beta)$

$$E(X) = \frac{\alpha}{\alpha + \beta}$$

$$V(X) = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)} \Rightarrow SD(X) = \sqrt{V(X)}$$

Probabilities

- When α and β are integers greater than 1, the cdf can be found by integrating a polynomial. Else we need to use software.

Example

- A management firm handles investment accounts for a large number of clients. The percent of clients who contact the firm for information or services in a given month is a beta random variable with $\alpha = 4$ and $\beta = 3$.

- (a) Find the pdf $f(x)$ and the cdf $F(x)$.

$$\begin{aligned} \rightarrow f(x) &= \frac{1}{B(4,3)} x^{4-1} (1-x)^{3-1} = 60x^3(1-x)^2 = 60x^3(1 - 2x + x^2) = 60x^3 - 120x^4 + 60x^5, \quad 0 \leq x \leq 1 \\ &\text{b} = \frac{\Gamma(4)\Gamma(3)}{\Gamma(4+3)} = \frac{3!2!}{6!} \\ \rightarrow F(x) &= \int_0^x f(t) dt = \int_0^x [60t^3 - 120t^4 + 60t^5] dt = 15t^4 - 24t^5 + 10t^6 \Big|_0^x = 15x^4 - 24x^5 + 10x^6, \quad 0 \leq x \leq 1 \end{aligned}$$

- (b) Find the probability the percent of clients contacting the firm in a month is less than 40%.

$$F(0.4) = 15(0.4)^4 - 24(0.4)^5 + 10(0.4)^6 = 0.1792$$

- (c) Find the expected value and variance for the percent of clients contacting the firm.

$$\rightarrow E(X) = \frac{\alpha}{\alpha + \beta} = \frac{4}{4+3} = \frac{4}{7}$$

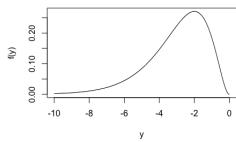
$$\rightarrow V(X) = \frac{\alpha\beta}{(\alpha + \beta)^2(\alpha + \beta + 1)} = \frac{\frac{4}{7} \cdot \frac{3}{7}}{(\frac{7}{7})^2(\frac{7}{7} + 1)} \approx 0.0306$$

Transformations

- We have covered the standard shapes that can be modeled by these distributions. To be more flexible, we can use functions of these random variables to model variations of the standard shapes.
- Examples:

- If a population is left-skewed and the range is $-\infty < Y \leq 0$. How can we model the population?

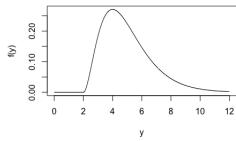
$$\rightarrow X \sim \text{Gamma} \quad \text{OR} \quad X \sim \text{Lognormal}$$



$$\rightarrow Y = -X$$

- If a population is right-skewed, but the range is $a \leq Y < \infty$, $a \neq 0$. How can we model the population?

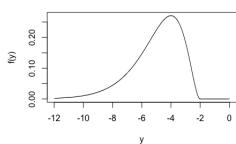
$$\rightarrow X \sim \text{Gamma} \quad \text{OR} \quad X \sim \text{Lognormal}$$



$$\rightarrow Y = X + a$$

- If a population is left-skewed and the range is $-\infty < Y \leq a$, $a \neq 0$. How can we model the population?

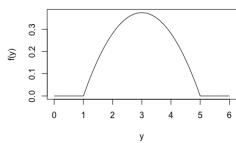
$$\rightarrow X \sim \text{Gamma} \quad \text{OR} \quad X \sim \text{Lognormal}$$



$$\rightarrow Y = -(X+a)$$

- If a population is bounded between $a \leq Y \leq b$. How can we model the population?

$$\rightarrow X \sim \text{Beta}$$



$$\rightarrow Y \rightarrow X : \begin{array}{l} a \leq y \leq b \\ [a, b] \end{array} \quad \begin{array}{l} 0 \leq y-a \leq b-a \\ [0, 1] \end{array}$$

$$\rightarrow X = \frac{y-a}{b-a} \quad 0 \leq X \leq 1$$

$$\Rightarrow Y = (b-a)X + a$$

Lectures 12 – Moment Generating Functions

MATH 320: Probability

Lecture 12: Moment Generating Functions

Chapters 2 and 3: Distributions (2.3, 2.4, 2.6, 2.7, 3.1, 3.2, 3.3)

Moments

Moments

- Definition: The n^{th} moment of X is $E(X^n)$.

Technically, these moments are “about zero (the origin)”: $E[(X - 0)^n]$.

The first and second moments are simply $E(x) + E(x^2) \Rightarrow v(x) = \text{Second moment} - \text{First moment}^2$

- Example: Calculate the third moment of $X \sim \text{Binomial}(n = 3, p = 0.2)$ using the pmf table below:

x	0	1	2	3
$f(x)$	0.512	0.384	0.096	0.008

$$E(x^3) = \sum x^3 f(x) = \begin{aligned} & 0^3(0.512) + 1^3(0.384) \\ & + 2^3(0.096) + 3^3(0.008) \\ & = 1.368 \end{aligned}$$

- There are other “types” of moments as well.

– Given random variable X and constant b ,

$E(X - b)$ is the **first moment of X about b** .

$E[(X - b)^n]$ is the **n^{th} moment of X about b** .

– If we let $b = \mu = E(X)$, then we get what are called **central moments**:

$E[(X - \mu)^n] \rightarrow$ “about the center $E(X)$ ”.

- Definition: The variance of a random variable X is its second central moment.

$$\text{if } n=2 \rightarrow E[(X-\mu)^2] = v(x)$$

Note *

$E(x^3) \neq (E(x))^3$
 \hookrightarrow new RV

Moment generating functions

Defining moment generating function and its properties

- The moment generating function (mgf) of random variable X (or the distribution of X), denoted $M_X(t)$, is defined by

$$M_X(t) = E(e^{tX}) = \sum_{x \in X} e^{tx} f(x) = \int_{-\infty}^{\infty} e^{tx} f(x) dx$$

- Notes:

$$E[g(x)] \text{ where } g(x) = e^{tx}$$

* e^{tX} is a function of t and X and it is a random variable.

* $E(e^{tX})$ is the expectation with respect to X , which takes away the randomness of $X \implies$ Result is only a function of t (variable, not constant).

- The mgf has a number of useful properties:

- Finding moments: The derivatives of $M_X(t)$ can be used to find the moments of the random variable X (i.e. take the derivative and evaluate at $t = 0$).

$$M'_X(0) = E(X), \quad M''_X(0) = E(X^2), \quad \dots, \quad M_X^{(n)}(0) = E(X^n)$$

- Distribution of a function of a random variable: The mgf of $aX + b$ can be found easily if the mgf of X is known.

$$M_{aX+b}(t) = e^{tb} M_X(at)$$

- Uniqueness: If a random variable X has the mgf of a known distribution, then X has that distribution.

- Formalizing this new notation from (1):

$$E(X^n) = M_X^{(n)}(0) = \left. \frac{d}{dt^n} M_X(t) \right|_{t=0}$$

 In words this means: The n^{th} moment of X is equal to the n^{th} derivative of $M_X(t)$ evaluated at $t = 0$.

Derivation / proof / intuition

- Finding moments:

Now we can show exactly how / why mgfs generate moments (hence the name).

- We will start by looking at the infinite series representation of $e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!}$ and then substituting the random variable tX for x in this series, we obtain:

$$e^{tX} = 1 + tX + \frac{t^2 X^2}{2!} + \frac{t^3 X^3}{3!} + \dots$$

(b) Now we take the expected value of each side. (with respect to x)

$$\left\{ \begin{array}{l} E(e^{tx}) = 1 + tE(x) + \frac{t^2}{2!} E(x^2) + \frac{t^3}{3!} E(x^3) + \dots \\ \hookrightarrow M_x(t) \end{array} \right.$$

Using the distributive property
of expectation and derivation

(c) Then we take the derivative. (with respect to t)

$$M_x'(t) = 0 + E(x) + tE(x^2) + \frac{t^2}{2} E(x^3) + \dots$$

(d) Then we evaluate at $t = 0$.

$$M_x'(0) = 0 + E(x) + 0E(x^2) + \frac{0^2}{2} E(x^3) + \dots$$

\downarrow $\checkmark E(x) \rightarrow$ first moment all have "t" term \Rightarrow cancel out

(e) Now take the second derivative and evaluate at $t = 0$ again.

$$\begin{aligned} \rightarrow M_x''(t) &= \frac{d}{dt} M_x'(t) \\ &\downarrow = 0 + 0 + E(x^2) + tE(x^3) + \frac{t^2}{2!} E(x^4) + \dots \\ \rightarrow M_x''(0) &= 0 + 0 + E(x^2) + 0E(x^3) + \frac{0^2}{2!} E(x^4) + \dots \\ &\downarrow = E(x^2) \rightarrow \text{Second moment} \end{aligned}$$

cancel out

(2) Distribution of a function of a random variable:

- Maybe the most important use of mgfs is to find out the distribution of a function of a random variable, specifically $Y = aX + b$.
- Earlier theorems only gave us only $E(aX + b)$ and $V(aX + b)$, but mgfs allow us to find out the distribution, and thus the pmf / cdf.

- Proof of theorem: $Y = aX + b$

$$\begin{aligned} \rightarrow M_Y(t) = E(e^{ty}) &\Leftrightarrow M_{aX+b}(t) = E[e^{(ax+b)t}] \\ &= E[e^{atx + bt}] \\ &= E_x[e^{atx} e^{bt}] \quad \text{constant value with respect to } X \text{ (for the expectation)} \\ &= e^{bt} E[e^{atx}] \\ &= e^{bt} M_x(at) \end{aligned}$$

(3) Uniqueness:

- When we introduced pmfs / pdfs and cdfs of the commonly used distributions, we stated a useful property:

If we recognize that we have a pmf / pdf where the range of the random variable and the probabilities / density function match the scenario of a specific distribution, then that random variable must follow that specific distribution.

- This is true for mgfs as well.

 Mgfs are unique. This means that if a random variable X has the moment generating function of a known random variable, it must be that kind of variable.

- We will not prove this.

Example: Using mgfs

x	0	1	2	3
e^{tx}	$e^{t \cdot 0} = 1$	e^t	e^{2t}	e^{3t}
$f(x)$	0.512	0.384	0.096	0.008

$\sim \text{Bin}(n=3, p=0.2)$

- (a) Find the moment generating function of X , $M_X(t)$.

$$\begin{aligned} M_X(t) &= E(e^{tx}) = \sum e^{tx} f(x) \\ &\rightarrow M_X(t) = 0.512(1) + 0.384e^t + 0.096e^{2t} + 0.008e^{3t} \end{aligned}$$

- (b) Find $E(X)$ using the mgf found in part (a).

Step 1: Take the derivative of $M_X(t)$ with respect to t .

$$\begin{aligned} \rightarrow M'_X(t) &= \frac{d}{dt} M_X(t) \\ &\downarrow = 0 + 0.384e^t + 0.096(2)e^{2t} + 0.008(3)e^{3t} \end{aligned}$$

Step 2: Now evaluate the derivative at $t = 0$ (i.e. plug in $t = 0$ and simplify).

$$\begin{aligned} \rightarrow M'_X(0) &= M'_X(t) \Big|_{t=0} \\ &\downarrow = 0 + 0.384e^0 + 0.096(2)e^{2(0)} + 0.008(3)e^{3(0)} \\ &= 0.6 = E(X) \end{aligned}$$

Confirm
Binomial mean = $np = 3(0.2) = 0.6$

Calc review e^x

$$\begin{aligned} \rightarrow \frac{d}{dx} e^x &= e^x \\ \rightarrow \frac{d}{dt} e^{f(x)} &= f'(x) e^{f(x)} \\ \rightarrow \frac{d}{dx} e^{ax} &= ae^{ax} \end{aligned}$$

- (c) Find $V(X)$ using the mgf.

Note that the higher derivatives can be used in the same way \Rightarrow Take the second derivative of $M_X(t)$ with respect to t and then evaluate at $t = 0$.

$$\begin{aligned} \rightarrow M''_X(t) &= \frac{d}{dt} [M'_X(t)] \\ &\downarrow = 0 + 0.384e^t + 0.096(2^2)e^{2t} + 0.008(3^2)e^{3t} \end{aligned}$$

$$\begin{aligned} \rightarrow M''_X(0) &= M''_X(t) \Big|_{t=0} \\ &\downarrow = 0 + 0.384e^0 + 0.096(2^2)e^{2(0)} + 0.008(3^2)e^{3(0)} \\ &= 0.84 = E(X^2) \end{aligned}$$

For Binomial

$$\begin{aligned} \text{variance} \rightarrow V(X) &= E(X^2) - (E(X))^2 \Rightarrow E(X^2) = V(X) + (E(X))^2 = npq + (np)^2 \\ &= 3(0.2)(0.8) + (3(0.2))^2 \leq 0.84 \end{aligned}$$

Final points

- Expanding on property (1) of mgfs (for discrete random variables, replace summation with integration for continuous).

$$M_X(t) = E(e^{tX}) = \sum e^{tx} f(x)$$

Middle step: $M'_X(t) = \sum x e^{tx} f(x)$

$$M''_X(t) = \sum x^2 e^{tx} f(x)$$

Property: $M'_X(0) = \sum x f(x) = E(X)$

$$M''_X(0) = \sum x^2 f(x) = E(X^2)$$

- The general form is the following:

$$M_X^{(n)}(t) = \sum x^n e^{tx} f(x)$$

$$M_X^{(n)}(0) = \sum x^n f(x) = E(X^n)$$

- Variance: Using these new ideas and notations, we have a new way to write the variance using mgfs:

$$\text{Var}(X) = M''_X(0) - [M'_X(0)]^2 = M''_X(t)|_{t=0} - [M'_X(t)|_{t=0}]^2$$

- Mgf's for random variables don't always exist. There is technically more to the definition $M_X(t) = E(e^{tX})$.

– Condition: It must be the case that the expectation exists for t in some neighborhood of 0. That is, there is an $h > 0$ such that, for all t in $-h < t < h$, $E(e^{tX})$ exists. If the expectation does not exist in the neighborhood of 0, we say that the mgf does not exist.

 – Said another way: Derivations of $M_X(t)$ of all orders exist at $t = 0 \implies M_X(t)$ is continuous at $t = 0$.

- Many standard probability distributions have moment generating functions which can be found fairly easily. This gives us another way to derive the mean and variance formulas stated previously.

Mgfs of commonly used discrete distributions

Discrete uniform random variable mgf

- The mgf for a discrete uniform random variable is straightforward to find.
- Let $X \sim \text{Discrete uniform } (N_0, N_1)$

$$\begin{aligned} M_X(t) &= E(e^{tx}) = \sum_{x=N_0}^{N_1} e^{tx} f(x) = \frac{1}{N_1 - N_0 + 1} \\ &\downarrow \\ &= \frac{1}{N_1 - N_0 + 1} \sum_{x=N_0}^{N_1} e^{tx} \end{aligned}$$

- If $X \sim \text{Discrete uniform } (N_0, N_1)$

$$M_X(t) = \frac{1}{N_1 - N_0 + 1} \sum_{x=N_0}^{N_1} e^{tx}$$

Bernoulli random variable mgf

- The mgf for a bernoulli random variable is straightforward to find.
- Let $X \sim \text{Bernoulli}(p)$

x	0	1
e^{tx}	$e^{t0} = 1$	e^t
$f(x)$	$1 - p = q$	p

$$M_X(t) = 1q + pe^t$$

- If $X \sim \text{Bernoulli}(p)$

$$M_X(t) = (1 - p) + pe^t = q + pe^t$$

Binomial random variable mgf

- The mgf for a binomial random variable is easy to find in the simplest cases. Then we can generalize the pattern without proof.

- We just saw the mgf when $X \sim \text{Binomial}(n=1, p) \sim \text{Bernoulli}(p)$

$$M_X(t) = (q + pe^t)^{\textcircled{1}}$$

- Now let $X \sim \text{Binomial}(n=2, p)$. $M_X(t) = q^2 + 2pqe^t + p^2e^{2t}$

x	0	1	2
e^{tx}	1	e^t	e^{2t}
$f(x)$	q^2	$2pq$	p^2

$$\begin{aligned} M_X(t) &= q^2 + 2pqe^t + p^2e^{2t} \\ &= \frac{1}{\cancel{q}} + \frac{2pq}{\cancel{q}} \cancel{e^t} + \frac{p^2}{\cancel{q}} \cancel{e^{2t}} \\ &= (1 + pe^t)^{\textcircled{2}} \end{aligned}$$

→ Now extend pattern for n trials

- If $X \sim \text{Binomial}(n, p)$

$$M_X(t) = (q + pe^t)^n$$

- To derive the mean and variance of the binomial distribution using the mgf, we would simply need to do the following:

- $E(X)$: Take the derivative of $M_X(t)$ and evaluate at $t = 0$.
- $V(X)$: Take the second derivative of $M_X(t)$ and evaluate at $t = 0$.

Then use the result of $E(X)$ and the alternate variance form
 $V(X) = E(X^2) - [E(X)]^2$.

- Warning: This requires careful attention when taking the derivatives in order to correctly keep track of everything.

- Example: You are working with a random variable X , and find that its mgf is:

$$M_X(t) = (0.2 - 0.8e^t)^7 \underset{\text{uniqueness}}{\implies} X \sim \text{Binomial}(n=7, p=0.2)$$

Geometric random variable mgf

- The derivation of the mgf for a geometric random variable is not bad either. It relies on the sum of an infinite geometric series.

- Let $X \sim \text{Geometric}(p) \rightarrow X = \{1, 2, 3, \dots\}$

$$\begin{aligned} \rightarrow M_X(t) &= E(e^{tX}) = \sum_{x=1}^{\infty} e^{tx} f(x) \\ &= \frac{p}{q} \sum_{x=1}^{\infty} (qe^t)^x \quad \xrightarrow{\text{Geometric series}} \frac{\infty a^x}{\sum_{x=1}^{\infty} a^x} = \frac{a}{1-a} \\ &= \frac{p}{q} \left[\frac{qe^t}{1-qe^t} \right] \\ &\checkmark = \frac{pe^t}{1-qe^t} \end{aligned}$$

Restriction is to ensure mgf is not undefined
(i.e. divide by zero)

- If $X \sim \text{Geometric}(p)$

$$M_X(t) = \frac{pe^t}{1-qe^t} \quad t < -\ln(q)$$

- Just like with the binomial, need to be careful when deriving the mean and variance using the mgf.
- Example: Let $X \sim \text{Geometric}(p)$. Suppose $Y = X - 1$. Find the mgf of Y , $M_Y(t)$.

Transformation

$$\begin{aligned} M_Y(t) &= M_{X-1}(t) = e^{-t} M_X(t) \\ &= e^{-t} \frac{pe^t}{1-qe^t} \\ &= \frac{p}{1-qe^t} \end{aligned}$$

Negative binomial random variable mgf

- We will not derive this. But we can make use of the pattern / relationship of Bernoulli and binomial.

$$\left. \begin{array}{l} \sum_{i=1}^{\infty} \underbrace{\text{Bernoulli}}_{\downarrow} = \text{Binomial} \\ M_X(t) \rightarrow q + pe^t \rightarrow (q + pe^t)^{\oplus} \end{array} \right\} \begin{array}{l} \sum_{i=1}^{\oplus} \underbrace{\text{Geometric}}_{\downarrow} = \text{Negative Binomial} \\ \frac{pe^t}{1-qe^t} \rightarrow \left[\frac{pe^t}{1-qe^t} \right]^{\oplus} \end{array}$$

- If $X \sim \text{Negative binomial}(r, p)$

$$M_X(t) = \left[\frac{pe^t}{1-qe^t} \right]^r \quad t < -\ln(q)$$

Poisson random variable mgf

- The derivation of the mgf for a Poisson random variable is quite short as well, but will be left for grad school. It makes use of the series for e^x .

- If $X \sim \text{Poisson}(\lambda)$

$$M_X(t) = e^{\lambda(e^t - 1)}$$

- Derivation: Use the mgf to derive the mean and variance of a Poisson(λ) random variable.

$$\textcircled{1} \quad E(X) = M_X'(0)$$

$$\rightarrow M_X'(t) = \frac{d}{dt} M_X(t) = \frac{d}{dt} [e^{\lambda(e^t - 1)}]$$

$$\downarrow \quad = \lambda e^t e^{\lambda(e^t - 1)}$$

$$\rightarrow M_X'(0) = \cancel{\lambda e^0 e^{\lambda(e^0 - 1)}} = \cancel{\lambda}$$

$$\downarrow \quad = \cancel{\lambda}$$

$$\textcircled{2} \rightarrow E(X^2) = M_X''(0)$$

$$\rightarrow M_X''(t) = \frac{d}{dt} M_X'(t) = \frac{d}{dt} [\underbrace{\lambda e^t}_u \underbrace{e^{\lambda(e^t - 1)}}_v]$$

$$\downarrow \quad = \lambda e^t e^{\lambda(e^t - 1)} + (\lambda e^t)^2 e^{\lambda(e^t - 1)}$$

Calc Review: Product rule
 Let $u = f(x)$ & $v = g(x)$
 $(uv)' = u'v + uv'$

$$\rightarrow M_X''(0) = \cancel{\lambda e^0 e^{\lambda(e^0 - 1)}} + \cancel{(\lambda e^0)^2 e^{\lambda(e^0 - 1)}}$$

$$\downarrow \quad = \lambda + \lambda^2$$

$$\textcircled{3} \rightarrow V(X) = E(X^2) - [E(X)]^2$$

$$\downarrow \quad = \lambda + \lambda^2 - (\lambda)^2$$

$$\downarrow \quad = \lambda$$

- Example: Let $X \sim \text{Poisson } (\lambda = 2)$. Suppose $Y = 3X + 5$. Find the mgf of Y , $M_Y(t)$.

Transformation

$$M_Y(t) = M_{3x+5}(t) = e^{5t} M_X(3t)$$

$$\downarrow \quad = e^{5t} [e^{3t}(e^{3t} - 1)]$$

Hypergeometric random variable mgf

- Apparently it exists, but we will not discuss it.

Moment generating functions for continuous random variables

Introduction

- Some continuous random variables have useful mgfs, which can be written in closed form and easily applied, and others do not.

We will discuss the mgf of the uniform, gamma and normal distributions. The beta and lognormal distributions do not have useful mgfs, and the Pareto mgf does not exist.

- Note that we could find mgfs of any of the non-named distribution examples we have seen before. They would just require application of the definition and probably lots of integration by parts (as most $f(x)$'s were non-simple functions of X , e.g. ROI example $f(x) = 0.75(1 - x^2)$, $-1 \leq x \leq 1$).

But the mgfs we will discuss here are much more interesting and common.

Uniform mgf

- The derivation of the uniform mgf is just a direct application of the definition.

$$\begin{aligned}
 M_X(t) &= E(e^{tx}) = \int_a^b e^{tx} \frac{1}{b-a} dx \\
 &\downarrow = \frac{1}{t(b-a)} e^{tx} \Big|_{x=a}^{x=b} \\
 &= \frac{e^{tb} - e^{ta}}{t(b-a)}
 \end{aligned}$$

- If $X \sim \text{Uniform}(a, b)$,

$$M_X(t) = \frac{e^{tb} - e^{ta}}{t(b-a)}, \quad t \neq 0$$

Gamma mgf

- The gamma mgf can be easily derived. It requires the identity which was shown earlier when introducing the gamma function $\Gamma(\alpha)$:

$$\int_0^\infty x^{\alpha-1} e^{-\beta x} dx = \frac{\Gamma(\alpha)}{\beta^\alpha}, \quad x > 0 \quad \text{and} \quad \alpha, \beta > 0.$$

- To derive $M_X(t)$ for $X \sim \text{Gamma}(\alpha, \beta)$ we just need to go from the definition. This will also be left for grad school.

- If $X \sim \text{Gamma}(\alpha, \beta)$,

$$M_X(t) = \left(\frac{\beta}{\beta - t} \right)^\alpha$$

- If we wanted to, we could then easily go through the process of taking the derivatives of $M_X(t)$ and evaluating at $t = 0$ to show that

$$E(X) = \frac{\alpha}{\beta}, \quad V(X) = E(X^2) - [E(X)]^2 = \frac{\alpha}{\beta^2}$$

Exponential mgf

- Since the exponential distribution is a special case of the gamma distribution,

$$X \sim \text{Gamma}(\alpha=1, \beta) = \text{Exp}(\beta)$$

also know
we ~~can find~~ the mgf of the exponential distribution.

- If $X \sim \text{Exponential}(\beta)$,

$$M_X(t) = \frac{\beta}{\beta - t}, \quad t < \beta$$

Normal mgf

- The normal mgf can actually be derived without fancy calculus skills, just fancy algebra skills.

- If $X \sim \text{Normal}(\mu, \sigma^2)$,

$$M_X(t) = e^{\mu t + \frac{\sigma^2 t^2}{2}}$$

- Lets derive the mean of the normal distribution using the mgf:

$$\rightarrow M_X'(t) = \frac{d}{dt} \left[e^{\mu t + \frac{\sigma^2 t^2}{2}} \right]$$

$$\downarrow = \mu + \sigma^2 t \left[e^{\mu t + \frac{\sigma^2 t^2}{2}} \right]$$

$$\rightarrow M_X'(0) = (\mu + \sigma^2 \cdot 0) e^{\mu(0) + \frac{\sigma^2(0^2)}{2}} = 1$$

$$\downarrow = \mu = \mathbb{E}(x)$$

- Derivation of the variance is straightforward, just requires careful application of the product rule.

Question: What must $E(X^2)$ equal?

$$\begin{aligned} E(x^2) &= V(x) + (E(x))^2 \\ \downarrow &= \sigma^2 + \mu^2 \end{aligned}$$

- We can now prove the following theorem that was shown earlier without proof:

Linear transformation of normal random variables: If $X \sim \text{Normal}(\mu, \sigma^2)$ and $Y = aX + b$. Then

$$Y \sim \text{Normal}(a\mu + b, a^2\sigma^2)$$

$$\begin{aligned} M_{aX+b}(t) &= e^{tb} M_X(at) \\ &= e^{tb} e^{\mu(at)} + \frac{\sigma^2(at)^2}{2} \\ &= e^{(a\mu+b)t} + \frac{(a^2\sigma^2)t^2}{2} \\ &\stackrel{M}{\sim} \text{Normal}(a\mu + b, a^2\sigma^2) \text{ MGF} \end{aligned}$$

- Now let's apply the same strategy to the specific linear transformation of standardizing and show the mgf of the standard normal distribution Z .

$$\begin{aligned} \rightarrow Z &= \frac{X-\mu}{\sigma} = \frac{1}{\sigma} X - \frac{\mu}{\sigma} \\ \rightarrow M_Z(t) &: M_{\frac{1}{\sigma}X - \frac{\mu}{\sigma}}(t) = e^{-\frac{\mu}{\sigma}t} M_X(\frac{1}{\sigma}t) = e^{-\frac{\mu}{\sigma}t} e^{\mu(\frac{1}{\sigma}t) + \frac{\sigma^2(\frac{1}{\sigma}t)^2}{2}} \\ &\quad \downarrow \\ &= e^{\sigma t + \frac{t^2}{2}} \\ &\sim \text{Normal}(\mu=0, \sigma^2=1) \end{aligned}$$

- Lets try this strategy for a distribution other than normal.

In the previous section, we had an example about standardizing an exponential distribution $X \sim \text{Exponential}(\lambda)$. We showed the mean and variance of the standardized exponential:

$$Z = \frac{X - \mu}{\sigma} = \frac{X - 1/\lambda}{1/\lambda} = \cancel{X} - 1 \implies E(Z) = 0, \quad V(Z) = 1$$

But did not have a conclusion about the distribution of Z .

$$\begin{aligned} M_Z(t) &= M_{\lambda X-1}(t) = e^{-t} M_X(\lambda t) \\ &= e^{-t} \left[\frac{\lambda}{\lambda - \lambda t} \right] \\ &= e^{-t} \left[\frac{\lambda}{\lambda(1-t)} \right] \\ &= \frac{1}{e^{t(1-t)}} \neq \text{Exponential Mgf} \\ \Rightarrow Z &\not\sim \text{Exp} \end{aligned}$$

After Test 3

Contents

Lectures 13 – Functions of Random Variables	174
---	-----

Lectures 13 – Functions of Random Variables

MATH 320: Probability

Lecture 13: Functions of Random Variables

Chapter 5: Distributions of Functions of Random Variables (5.1)

Deductibles and caps: Expected value of a function of a random variable

Expected value of a loss or claim

- These examples are in insurance applications, but are just expected value of a function of a random variable problems.
- Insurance loss.
 - Example: (a) The amount of a single loss X for an insurance policy is exponential, with density function

$$f(x) = 0.002e^{-0.002x}, \quad x \geq 0 \quad \Rightarrow \quad X \sim \text{Exp}(\lambda = 0.0002)$$

So the (base) expected value of a single loss is: $E(X) = \frac{1}{\lambda} = 500$

- Insurance with a deductible.
 - Continuing example: (b) Suppose now the insurance policy has a deductible of \$100 for each loss. Find the expected value of a single claim.
 - ** Now loss amount \neq claim amount
 - **STRATEGY:** We need to write a new function $g(X)$ that represents the new claim amount taking into account the deductible.
 - $g(X)$ will be a piecewise function. So think about the values $g(X)$ takes in cases based on the range of X .

NOTE: We are thinking about the values of the claim from the insurance company's perspective.

claim amount

$$g_1(x) = \begin{cases} 0 & 0 < x < 100 \\ x-100 & x > 100 \end{cases}$$

↓
policyholder pays first 100

$E[g_1(x)] = \int_0^\infty g_1(x) f(x) dx$
 $\hookrightarrow E_{\text{Exp}}(\lambda = 0.0002)$
 $= \int_0^\infty 0 f(x) dx + \int_{100}^\infty (x-100) f(x) e^{-0.0002x} dx$
 $= 0 + \text{Integration by parts} >$
 $= 409.37 < 500 = E(X)$

- Insurance with a deductible and a cap.
 - Continuing example: (c) Now suppose the insurance policy has a deductible of \$100 per claim AND a restriction that the largest amount paid on any claim will be \$700.
 - Strategy: Use the same strategy as before for the first case, then just need to take into account the cap.

claim amount

$$g_2(x) = \begin{cases} 0 & 0 < x \leq 100 \\ x-100 & 100 < x \leq 800 \\ 700 & x > 800 \end{cases}$$

→ payments are capped at 700

⇒ loss > 800 receives a payment of $800 - 100 = 700$

$$\left\{ \begin{aligned} E[g_2(x)] &= \int_0^\infty g_2(x) f(x) dx \\ &\quad \hookrightarrow \text{Exp}(d=0.002) \\ &= \int_0^{100} 0 f(x) dx + \int_{100}^{800} (x-100) f(x) dx + \int_{800}^\infty 700 f(x) dx \\ &= 0 + \text{< Integration by parts>} + 700 \int_x(800) \\ &= 167.09 + 143.33 \\ &= 708.42 < 409.37 = E[g_1(x)] \end{aligned} \right.$$

- Another example: The amount of a single loss X for an insurance policy has the density function $f(x)$ for $x \geq 0$ with deductible of \$150 and cap of \$900.

- (a) Find a function $g(X)$ for the amount paid (claim amount) for a loss x .
- (b) Write the integral to solve for the expected claim amount.

a)

$$g(x) = \begin{cases} 0 & 0 < x \leq 150 \\ x-150 & 150 < x \leq 1050 \\ 900 & x > 1050 \end{cases}$$

b) $E[g(x)] = \int_0^{150} 0 f(x) dx + \int_{150}^{1050} (x-150) f(x) dx + \int_{1050}^\infty 900 f(x) dx$

- In general, if loss x with deductible d and cap c , we have

$$g(x) = \begin{cases} 0 & x \leq d \\ x-d & d < x \leq d+c \\ c & x > d+c \end{cases}$$

The distribution $Y = g(X)$

Transformations so far

- We have already seen simple methods for finding $E[g(X)]$ and $V[g(X)]$ for any type of variable.
- Example: The monthly maintenance cost for a machine $X \sim \text{Exponential}(\lambda = 0.01)$. Next year costs will be increased 5% due to inflation. Thus next year's monthly cost is $Y = g(X) = 1.05X$.

Find $E(Y)$.

$$E(x) = \frac{1}{\lambda} = 100 \rightarrow E(y) = E(1.05x) = 1.05E(x) = 105$$

- Note we did not need to know the distribution of Y for this calculation.

However, the mean and variance alone are not sufficient to enable us to calculate probabilities for $Y = g(X)$, we need the actual distribution function $f(x)$.

- Discrete example: Same X with a new (discrete) model and inflation costs $Y = g(X) = 1.05X$:
- (a) Find the distribution of $Y = g(X)$.
- (b) Find $P(Y < 100)$.

x	$f(x)$	$y = 1.05x$	$f(y)$
0	0.28	0	0.28
50	0.43	52.5	0.43
100	0.20	105	0.20
150	0.09	157.5	0.09

$$\begin{aligned} P(Y < 100) &= P(Y = 0) + P(Y = 52.5) \\ &= 0.28 + 0.43 \\ &= 0.71 \end{aligned}$$

- For the original continuous model, it is not as simple to find the new distribution.

Continuous transformations example

- Continuing example: Using the original $X \sim \text{Exponential}(\lambda = 0.01)$ model...
- Find $P(Y \leq 100)$.

GOAL: Get the probability statement to be with respect to X .

STRATEGY: "Indirectly" find the probability for Y based on the known cdf of X and using some simple algebra. Note that this is the same strategy we used to find lognormal probabilities based on the normal cdf.

$$\begin{aligned} &\rightarrow \text{know } F_x(x) = 1 - e^{-0.01x}, \quad x \geq 0 \\ &\Rightarrow P(Y \leq 100) = P(1.05x \leq 100) && \text{< substituting >} \\ &\quad \downarrow && \text{< rearrange to get } x \text{ by itself >} \\ &\quad = P(x \leq \frac{100}{1.05}) \\ &\quad = F_x(\frac{100}{1.05}) = 1 - e^{-0.01(\frac{100}{1.05})} \\ &\quad \approx 0.614 \end{aligned}$$

- Find the cdf $F_Y(y)$.

STRATEGY: Use the same reasoning as above, just for a general y :
 $P(Y \leq 100) = F_Y(100) \rightarrow P(Y \leq y) = F_Y(y)$ for any value $y \geq 0$.

$$\begin{aligned} F_Y(y) &= P_Y(Y \leq y) \\ &= P_{X'}(1.05X \leq y) \\ &= P_X(X \leq \frac{y}{1.05}) \\ &\downarrow \quad = F_X(\frac{y}{1.05}) = 1 - e^{-0.01(\frac{y}{1.05})}, \quad y \geq 0 \end{aligned}$$

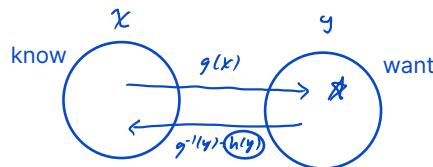
- Note that the range of X is the interval $[0, \infty)$. The range for $Y = 1.05X$ is the same interval. This will not always be the case for transformations $g(X)$.

 **STRATEGY:** Apply $g(x)$ to all pieces, ALWAYS check both sides.

$$\begin{aligned} \rightarrow x: \quad 0 &\leq x < \infty \\ \rightarrow y = g(x) = 1.05x: \quad 1.05(0) &\leq 1.05x < 1.05(\infty) \\ &\quad 0 \leq y < \infty \end{aligned}$$

Inverses

- Finding the distribution of $Y = g(X)$ like we did above is much simpler when the transformation function $g(X)$ has an inverse.
- Recall that the function $g(X)$ defines a mapping from the original range of X to a range of Y . That is,



** We do not know stuff (pdf, cdf, etc.); so we have to use the inverse function to go backwards. Y is completely determined by X .

- When do inverse functions exist?

If the function $g(x)$ is strictly **monotone** \Rightarrow one-to-one \iff inverse exists.

$u > v \Rightarrow g(u) > g(v) \rightarrow$ strictly increasing

$u > v \Rightarrow g(u) < g(v) \rightarrow$ strictly decreasing

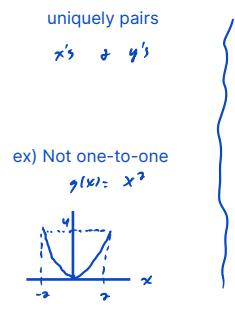
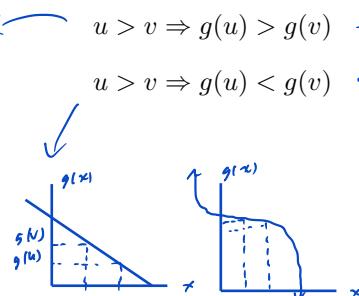
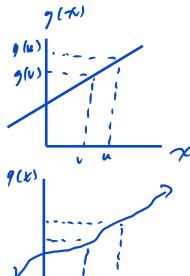
uniquely pairs

$x's \leftrightarrow y's$

$$g^{-1}(y) = h(y) \checkmark$$

ex) Not one-to-one

$$g(x) = x^2$$



- Summary and results:

For a function $g(x)$ that strictly increasing or strictly decreasing on the range of X , we can find an inverse function $h(y)$ defined on the range of Y . Thus we have:

$$h[g(x)] = x \quad + \quad g[h(y)] = y$$

** Strategy when problem solving:

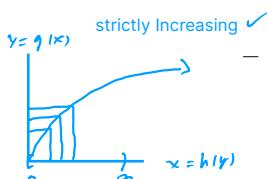
1. Draw a figure of the transformation.
- If transformation is strictly increasing or strictly decreasing over \mathcal{X} , then use the methods described next.
2. Check range of Y (i.e. ALSO transform range of X to range of Y).

Using $F_X(x)$ to find $F_Y(y)$ for $Y = g(X)$

- We will only generalize the methods for when $g(X)$ has an inverse. If this is true, then there are two cases.
- Case 1: $g(x)$ is strictly increasing on the range of X**

- Let $h(y)$ be the inverse function of $g(x)$. The function $h(y)$ will also be strictly increasing. In this case, we can find $F_Y(y)$ as follows:

$$\begin{aligned} F_Y(y) &= P(Y \leq y) \\ &= P(g(X) \leq y) \\ &= P(h[g(x)] \leq h(y)) \\ &= P(x \leq h(y)) \\ &= F_X(h(y)) \end{aligned}$$



- Example: Let $X \sim \text{Exponential}(\lambda = 3)$. Find the cdf of $Y = \sqrt{X}$. $\rightarrow F_X(x) = 1 - e^{-3x}, x \geq 0$

There are two ways that we can solve this.

Long way

$$\begin{aligned} F_Y(y) &= P(Y \leq y) \\ &= P(\sqrt{X} \leq y) \\ &= P(X \leq y^2) \\ &= F_X(y^2) = 1 - e^{-3y^2} \end{aligned}$$

Short way

$$\begin{aligned} &\rightarrow y = \sqrt{x} \quad \xrightarrow{\text{solve for } x} \quad x = h(y) = y^2 \\ &\rightarrow F_Y(y) = F_X(h(y)) \\ &\downarrow = F_X(y^2) = 1 - e^{-3y^2} \end{aligned}$$

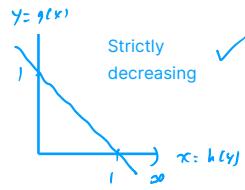
Range:

$$\begin{array}{ll} x: & 0 \leq x \leq \infty \\ y = g(x) = \sqrt{x}: & \sqrt{0} \leq \sqrt{x} \leq \sqrt{\infty} \\ & 0 \leq y \leq \infty \end{array}$$

- Case 2: $g(x)$ is strictly decreasing on the range of X

- Let $h(y)$ be the inverse function of $g(x)$. The function $h(y)$ will also be strictly decreasing. In this case, we can find $F_Y(y)$ as follows:

$$\begin{aligned}
 \rightarrow F_Y(y) &= P(Y \leq y) && \text{have to flip inequality because decreasing function} \quad \cancel{\star} \\
 &= P(g(X) \leq y) \\
 &= P(h[g(x)] \geq h(y)) \\
 &= P(X \geq h(y)) \quad \rightarrow \text{survival function} \\
 &= 1 - F_X(h(y))
 \end{aligned}$$



- Example: Let $X \sim \text{Exponential}(\lambda = 3)$. Find the cdf of $Y = 1 - X$.

Again, we can do the long ("derivation") way or short way (skip to end result).

Long way

$$\begin{aligned}
 F_Y(y) &= P(Y \leq y) \\
 &= P(1-X \leq y) \\
 &= P(X \geq 1-y) \\
 &= 1 - F_X(1-y) = 1 - [1 - e^{-3(1-y)}] \\
 &= e^{-3(1-y)}, \quad y \leq 1
 \end{aligned}$$

Short way

$$\begin{aligned}
 \rightarrow y &= 1-x \implies x = h(y) = 1-y \\
 \rightarrow F_Y(y) &= 1 - F_X(1-y) \\
 &\quad \downarrow \quad = 1 - [1 - e^{-3(1-y)}] \\
 &\quad \downarrow \quad = e^{-3(1-y)}, \quad y \leq 1
 \end{aligned}$$

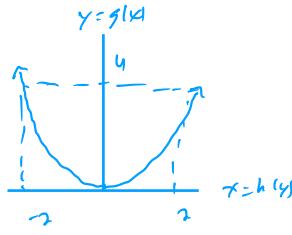
Range:

$$x : 0 \leq x < \infty$$

$$y = g(x) = 1-x : 1-\infty \leq 1-x < 1-\infty$$

$$1 \leq y < \infty$$

$$\Rightarrow \boxed{-\infty < y \leq 1} \quad \cancel{\star} \text{ bounds flip sides}$$



Not monotone (one-to-one)
 \Rightarrow no $h(y)$ over x

- If $g(x)$ does NOT have an inverse

– Example: Let $X \sim \text{Uniform}(a = -2, b = 2)$. Find the cdf of $Y = X^2$.

13-7



$$f_x(x) = \frac{1}{4}$$

$$= \frac{x+2}{4}$$

$$\rightarrow F_Y(y) = P(Y \leq y)$$

$$= P(X^2 \leq y)$$

$$= P(|X| \leq \sqrt{y})$$

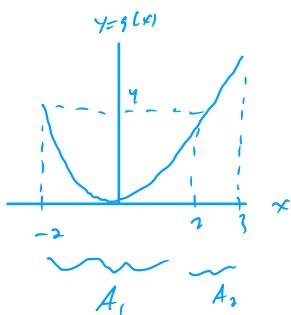
$$= P(-\sqrt{y} \leq X \leq \sqrt{y})$$

$$= F_X(\sqrt{y}) - F_X(-\sqrt{y}) = \frac{\sqrt{y} + 2}{4} - \frac{-\sqrt{y} + 2}{4}$$

$$= \frac{\sqrt{y}}{2}, \quad 0 \leq y \leq 4$$

– It can be even more complicated if there isn't a "balanced" range of Y .

Example: Let $X \sim \text{Uniform}(-2, 3)$. find the cdf of $Y = X^2$.



\Rightarrow Solve each A_i individually + piecewise

$$F_Y(y) = \begin{cases} \dots & x \in A_1 \\ \dots & x \in A_2 \end{cases}$$

– Both of these cases will be left for grad school :)

Finding the density function $f_Y(y)$ for $Y = g(X)$

- Finding $F_Y(y)$ gives us all the information that is needed to calculate probabilities for Y , as shown below:

$$P(Y \leq y) = F_Y(y) \quad P(Y \geq y) = 1 - F_Y(y) \quad P(a \leq Y \leq b) = F_Y(b) - F_Y(a)$$

Thus there is no real need to find the density function $f_Y(y)$. If the density function is required, it can be found by differentiating the cdf:

$$f_Y(y) = \frac{d}{dy} F_Y(y)$$

- If X is continuous, it is usually easier to find the cdf of Y and then the pdf of Y (rather than skipping straight to the pdf). But we will learn both methods, which we shall name:

1. Cdf method

$$\text{pdf } X \rightarrow \text{cdf } X \rightarrow \text{cdf } Y \rightarrow \text{pdf } Y$$

2. Pdf method

$$\text{pdf } X \longrightarrow \text{pdf } Y$$

(aka change of variable technique)

- Again when working in situations when $g(x)$ has an inverse, there are two cases:

- Case 1: $g(x)$ is strictly increasing on the range of X**

- Setup: $h(y)$ is the inverse of $g(x)$ and $h(y)$ is strictly increasing.
- Previous results: $F_Y(y) = F_X(h(y))$
- We can find the pdf $f_Y(y)$ as follows:

$$f_Y(y) = \frac{d}{dy} F_Y(y) = \frac{d}{dy} F_X(h(y)) = F'_X(h(y)) h'(y) = f_X(h(y)) h'(y)$$

Substitute chain rule simplify

- Case 2: $g(x)$ is strictly decreasing on the range of X**

- Setup: $h(y)$ is the inverse of $g(x)$ and $h(y)$ is strictly decreasing.
- Previous results: $F_Y(y) = 1 - F_X(h(y))$
- We can find the pdf $f_Y(y)$ as follows:

$$(Same\ steps) \quad f_Y(y) = \frac{d}{dy} F_Y(y) = \frac{d}{dy} [1 - F_X(h(y))] = -F'_X(h(y)) h'(y) = -f_X(h(y)) h'(y)$$

- Since $h(y)$ is decreasing, its derivative is negative. Thus the final expression above is actually positive.

rewrite as $\rightarrow f_X(h(y)) \underbrace{(-h'(y))}_{<0} = \underbrace{(f_X(h(y)) / h'(y))}_{(negatives\ cancel)} = \text{case 1 } f_Y(y) \text{ because } h'(y) > 0 \text{ when increasing}$

Just simplify with absolute value

\nearrow holds in both cases



- Theorem: Let X have cdf $F_X(x)$ with range \mathcal{X} , $Y = g(X)$ with range \mathcal{Y} and inverse $h(y)$.

- If $g(x)$ is strictly increasing on $\mathcal{X} \rightarrow F_Y(y) = F_X(h(y))$ for $y \in \mathcal{Y}$.
- If $g(x)$ is strictly decreasing on $\mathcal{X} \rightarrow F_Y(y) = 1 - F_X(h(y))$ for $y \in \mathcal{Y}$.
- If $g(x)$ is strictly increasing or strictly decreasing on \mathcal{X} , then

$$f_Y(y) = f_X(h(y)) \cdot |h'(y)| \quad \text{for } y \in \mathcal{Y}.$$

- Return to previous examples: Let $X \sim \text{Exponential}(\lambda = 3)$.

- (a) Find the pdf of $Y = \sqrt{X}$.

<u>Cdf method</u> ① previously showed $F_Y(y) = 1 - e^{-3y^2}$ ② $f_Y(y) = \frac{d}{dy} F_Y(y) = \frac{d}{dy} [1 - e^{-3y^2}]$ \downarrow $= -e^{-3y^2} (-3(2y))$ $= 6ye^{-3y^2}, \quad y \geq 0$	<u>Pdf method</u> $\rightarrow x = h(y) = y^2 \rightarrow h'(y) = 2y$ $\rightarrow f_Y(y) = f_X(h(y)) h'(y) $ \downarrow $= 3e^{-3(y^2)} 2y $ \downarrow $= 6ye^{-3y^2}, \quad y \geq 0$
--	---

(ranges shown previously)

- (b) Find the pdf of $Y = 1 - X$.

<u>Cdf method</u> ① previously showed $F_Y(y) = e^{-3(1-y)}$ ② $f_Y(y) = \frac{d}{dy} F_Y(y) = \frac{d}{dy} [e^{-3(1-y)}]$ \downarrow $= e^{-3(1-y)} (1)$ $= 3e^{-3(1-y)}, \quad y \leq 1$	<u>Pdf method</u> $\rightarrow x = h(y) = 1 - y \rightarrow h'(y) = -1$ $\rightarrow f_Y(y) = f_X(1-y) h'(y) $ \downarrow $= 3e^{-3(1-y)} -1 $ \downarrow $= 3e^{-3(1-y)}, \quad y \leq 1$
---	---

(ranges shown previously)

More examples

1. Let X be the outcome when you roll a fair four sided die. If you get $Y = |X - 2|$ dollars based on your roll, find $f_Y(y)$.

 Discrete Random variable
(don't need cdf/pdf method)

$$\rightarrow X = \{1, 2, 3, 4\} \rightarrow Y = \{0, 1, 2\}$$

$$\rightarrow f_Y(0) = P(X=2) = \frac{1}{4}$$

$$f_Y(1) = P(X=1) + P(X=3) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}$$

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

$$\rightarrow f_Y(y) = \begin{cases} \frac{1}{4} & y=0 \\ \frac{1}{2} & y=1 \\ \frac{1}{4} & y=2 \end{cases}$$

2. Let $X \sim \text{Poisson}(\lambda = 4)$. If $Y = X^2$, find the pdf of Y .

$$\rightarrow Y = g(x) = x^2 \text{ is one-to-one over } X = \{0, 1, 2, \dots\} \checkmark$$

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

$$= f_X(\sqrt{y}) = e^{-4} \frac{4^{\sqrt{y}}}{(\sqrt{y})!}, y = \{0^2, 1^2, 2^2, 3^2, \dots\}$$

$\downarrow Y=x^2$

