# MATH 320: Probability
# Lectures

Colton Gearhart

April 22, 2024

## Contents

# Test 1

## Contents



## 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.

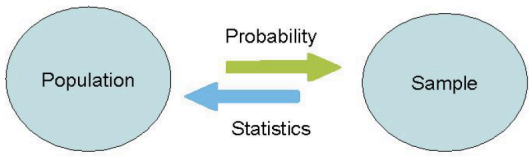



Figure: The reverse actions of Probability and Statistics

- 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. $\rightarrow$

    We want to know the chance all 10 batteries in a pack last 1500 hours. $\rightarrow$

  - Suppose that out of 10 batteries, 6 are found to last more than 1500 hours. $\rightarrow$

    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%. $\rightarrow$

- 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.
  - Calculating probabilities: Set theory, counting, probabilities of events, random variables
  - Univariate distributions: Random variables / distributions, probabilities of events, summaries of random variables, applications
  - 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:

2. Roll die, record number that appears:

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

## What is probability and how to calculate it

Intuitively, probability is the likelihood of something occurring.

One approach (simplest case)

- **Probability by counting equally likely outcomes**

  Probability of an event =

  - Example: Flip a fair coin, $P(\text{Heads}) =$

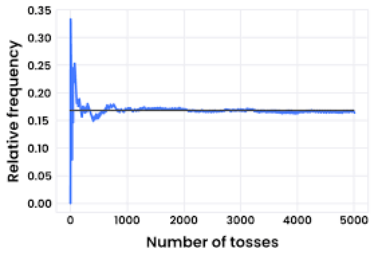
- Events are not always equally likely, so cannot determine probabilities by counting. But there is a simple way to estimate that probability.

Another approach

- **Empirical probability** (based on collected data).
- Relative frequency estimate of the probability of an event

  Probability of an event =

- 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, $\frac{\#\,6s}{10{,}000} \approx \frac{1}{6}$).




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. _____ of an experiment correspond to _____ 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) $x > 0$"

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

- The **sample space** $S$ (aka outcome space) is the set of _____ 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 _____ of $S$.
  1. Examples: Roll die, record the number that appears.
     - Define notation:
     - Show event:

       An event occurs when at least one element in the event has occurred.
  2. Is $S$ an event?
- The **null (empty) set** is the set containing _____.

Basic operations (algebra of sets)

- **Union**:

  The set of elements that belong to _____

- **Intersection**:

  The set of elements that belong to _____

- **Complement**:

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

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

- Example: 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:

  $A_1 \cup A_2 =$ $\qquad$ $A_2 \cup A_3 =$

  $A_1 \cap A_2 =$ $\qquad$ $A_2 \cap A_3 =$

  $\sim A_2 =$ $\qquad$ $\sim(A_1 \cup A_2) =$

Set identities

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

- **Commutative Law**: Reordering $\qquad$ Order doesn't matter

  $A \cup B =$

  $A \cap B =$

- **Associative Law**:

  Changing location of parentheses $\qquad$ Order of operations with ( ) doesn't matter

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

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

- **Distributive Law**: Distribute set operation $\qquad$ Distribute to items in ( )

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

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

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

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

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

Relationships among sets

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

  - The events $A_1, A_2, \ldots$ are **pairwise mutually exclusive** if _____ is mutually exclusive.

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

  - Events $A_1, \ldots, A_k$ are **exhaustive** if when combined, they form _____.

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

  - Events $A_1, \ldots, A_k$ form a **partition** of $S$ if they are _____ and _____ mutually exclusive.

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

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

  (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 _____

  To check:

  1. $A_1 \cup A_2 \cup A_4 = \{1\} \cup \{2, 3, 4\} \cup \{5, 6\} =$

  2. $A_1 \cap A_2 = A_1 \cap A_4 = A_2 \cap A_4 =$

## 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?

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?

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

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

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) $A =$

- **Complements counting rule**: For any finite sample space $S$ and event $A$

- **General union counting rule**:

  For any two events $A$ and $B$ in any finite sample space

- **Special case union counting rule**: If $A$ and $B$ are mutually exclusive

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 two 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 $(i = 1, \ldots, k)$, then the entire job can be done in $n_1 \times n_2 \times \cdots \times n_k$ ways.

| Task 1 | Task 2 | $\cdots$ | Task $k$ | Total outcomes |
|---|---|---|---|---|
| $n_1$ | $n_2$ | $\cdots$ | $n_k$ | $n_1 \times n_2 \times \cdots \times n_k$ |

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

- 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
  2. The number of objects we are going to select
  3. Order matters or not?
  4. With replacement or without?

- Possible methods of counting

| | Ordered | Unordered |
|---|---|---|
| With replacement | | |
| Without replacement | | |

Ordered, with replacement

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

- **Ordered, with replacement**: Given $n$ distinguishable objects, there are _____ 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.

Ordered, without replacement

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

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

- **Ordered without replacement (all $n$)**: The number of permutations of $n$ distinct objects is _____.

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

- **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 _____.

  Example: $P\binom{10}{3} =$

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?

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

  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: $\binom{n}{r} =$

  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 $\longrightarrow$ Combination | Combination $\longrightarrow$ Permutation |
|---|---|
| 1. | 1. |
| 2. | 2. |
| Example: Committee of 3 from 7 | Example: Rank 3 from 7 |

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.

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

   (c) Determine the jobs for three members out of 8 at the dinner party: set the table, serve the food, do the dishes.

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

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

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?

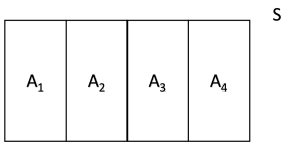
   - Many counting problems include combined the use of the multiplication principle, permutations and combinations.

     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.

  This is why $\binom{n}{r} = \binom{n}{n-r}$ 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 3, 3 and 4.

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

$$\binom{n}{n_1,\, n_2,\, \ldots,\, n_k} =$$

  This is called the **multinomial coefficient**.

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

- 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 _____.

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

Summary: When to use which counting tool (formula)

- **Ordered**
  - **With replacement**: _____ ex) how many 6 digit passwords can the digits 0-9 make?
  - **Without replacement**: _____ ex) 7 possible vacation destinations, rank your top 3.
- **Unordered**
  - **Without replacement**: _____ ex) select 5 out of 12 players to start the basketball game.

    **More than two groups**: _____ 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:

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

$$P(A) =$$

- 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:

  (b) Club or King:

  (c) Not Club nor King:

  (d) Club or Hearts:

- 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).
- Examples:
  1. 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.

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

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

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

  This means the 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 | 0.9670 | 0.0311 | 0.0019 |

  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.

  Let's break down what we implicitly did.

  1. Assigned probabilities to each of the _____ in the sample space (i.e. the table).
  2. Wrote the _____ of interest in terms of the outcomes of interest.
  3. _____ 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 $S = \{O_1, \ldots, O_n\}$ 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.

  1. What are the sample space and individual outcomes?

  2. Find the probability of each individual outcomes.

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

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. $P(A)$ for any event $A$.
  2. $P(S) =$ .
  3. Suppose $A_1, \ldots, A_n$ is a (possibly infinite) sequence of pairwise mutually exclusive events. Then
     $$P(A_1 \cup \cdots \cup A_n) =$$

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

  1. **Any _____ can be expressed as a union of mutually exclusive outcomes**.
  2. **The probability of an event is the _____ 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) $P(\emptyset) =$ (b) $P(A) \leq$ (c) $P(\sim A) =$

- Proofs (easier to prove out of order):

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

*PROOF STRATEGY*: Rewrite events we know and use axioms to simplify their probabilities.

(b) $P(A) \leq 1$ *HINT: Use part (c) of theorem*

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

(a) $P(\emptyset) = 0$ *HINT: Use part (c) of theorem*

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

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

  (b) $P(A \cup B) =$

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

- Proofs:

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

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

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

*HINT: Use identity $(B \cup A) = (B \cup A) \cap (B \cup \sim B)$*

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

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

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.

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

   - 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

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

   $$P(B) = 0.50, P(L) = 0.22, P(E) = 0.20,$$
   $$P(B \cap L) = 0.10, P(B \cap E) = 0.15, P(L \cap E) = 0.09,$$
   $$P(B \cap L \cap E) = 0.06$$

   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 _____.

  From these, we obtain _____, which together make up the _____.
- We want to know _____ of events (subsets of $S$).

  To calculate this, we defined how to assign _____ in general via the _____.
- Then there are two methods to compute _____:

  1.

  2.

Odds

- Odds of an event $A$ are generally written as a _____ of two integers, such as _____. The odds against $A$ are given by the _____.

  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 $P(\sim A)$.

  Odds of $A =$

- Converting between odds and probability:

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

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

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

# 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: $P(A) =$ $P(B) =$

  If we know that $A$ already occurred, what is the probability of $B$?

- This is known as a _____ probability.

Updated sample spaces

- The _____ is the set of all possible outcomes.

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

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

- '$A$ occurred' means we know that the outcome is one of the elements in _____. That is, there is no chance to get outcomes _____.

  So these numbers are _____ from the sample space.

- Thus, an event which already occurred _____ the sample space and it becomes an _____ 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?

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

  Another way we could interpret this result is: Of students with Good attendance, _____ are Chemistry majors.

- 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: $P(\text{Stats}) =$ $P(\text{Perfect} \cap \text{Stats}) =$

**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 for equally likely outcomes**

  $P(A \mid B) =$

  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:

  $P(A \mid B) =$

- The notation $P(A \mid B)$ is read "the conditional probability of $A$ given $B$".
  - $A$ is the main event of interest and $B$ is called the conditioning event.

Examples

1. 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.

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

   (b) Given that a car is blue, what is the probability that it got a clear coat?

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

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 policy holder has filed at least one claim.*

This tells us $\approx$ _____ of policyholders who file a claim file exactly 2 claims.

- Note on conditional probability:
  - ALWAYS $P(B \mid A) =$
  - ONLY SOMETIMES $P(B \mid A) =$

**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 $P(\cdot \mid B)$ as well.
- In addition, the theorems we proved also hold true for $P(\cdot \mid B)$.
- Examples and (a few) new theorems:
  - We know $P(\sim A) = 1 - P(A)$, using the previous example we can find $P(\sim 2 \mid C)$.

    In general: $P(\sim A \mid B) =$
  - 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 $C$,

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

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

  $P(A \cap B) =$

  $P(A \cap B) =$
- Interpretation:

- 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?

  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 $P(B \mid A)$ 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.

- This multiplication rule can be extended to any number of events as well.
  - More generally, given $k$ events $A_1, \ldots, A_k$

    $P(A_1 \cap \cdots \cap A_k) =$

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

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

- Example:

  (a) 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?

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

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?

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

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

Can easily use _____ rather than solving for many branches.

(d) Find the probability that the second card is a king ($K2$), given that the first card drawn was a king.

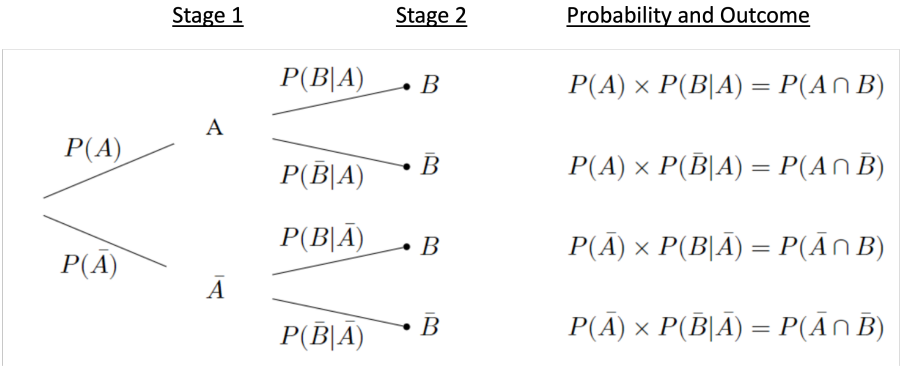
– This is just knowing parts of the tree. For problems that are just asking for "one level" of the tree or "one stage" of an experiment, it is often easier to find conditional probability by

(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




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 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 an homeowner 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:

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

   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?

## 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 $P(C \mid A)$ and $P(B \mid A)$, then compare them with $P(C)$ and $P(B)$, respectively.

  1. $P(C \mid A) =$
  2. $P(B \mid A) =$

- How can we interpret the results above?
  1. The knowledge of the occurrence of $A$ has _____ the probability of $C$.
  2. $P(B)$ is _____ the occurrence of $A$.

Definition of independence

- Two events $A$ and $B$, are **independent** if

  If $P(A) > 0$ and $P(B) > 0$, then $A \perp\!\!\!\perp B \iff$

  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.

- 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.

  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

  If $\perp\!\!\!\perp$, find _____ probability by multiplying _____ probabilities.

- **Multiplication rule for independent events**

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

- 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.

  (a) What is the probability the target is hit at least once?

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

- Summary:
  - If $A$ and $B$ are independent, can easily compute $P(A \cap B) = P(A) \cdot P(B)$
  - If $A$ and $B$ are mutually exclusive, can easily compute $P(A \cup B)$

Theorems

- If $A$ and $B$ are independent events, then the following pairs of events are also independent:
  (a) $A$ and $\sim B$
  (b) $\sim A$ and $B$
  (c) $\sim A$ and $\sim B$
  - Proof of (a)

    Setup: Want to show

    One way:

    Another way using conditional probabilities:

  - 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.

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

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

Assume $P(A) = 0$.

Similar logic for if $P(B) = 0 \implies A \perp\!\!\!\perp B$.

(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:

If events are mutually exclusive, there is a very _____ relationship (if one event occurs, the other _____ 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.

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 the pairs are independent.

  (b) Check if all three are independent.

  This shows that $A$, $B$, and $C$ are _____ **independent**, but not _____ **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.

   (b) Exactly two of the three events occur

   (c) None of the events occur

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 $P(\sim A \cup \sim B \cup C)$.

# 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*:

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

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

  (b) 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.

  (c) Use the general multiplication rule to find the probability for the event of interest.

- Visualizing scenario:

  (a) Using a Venn Diagram:

  (b) Using a Tree Diagram:

Law of total probability

- The **law of total probability** says that a partition $(A_1, \ldots, A_n)$ 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 $S$.

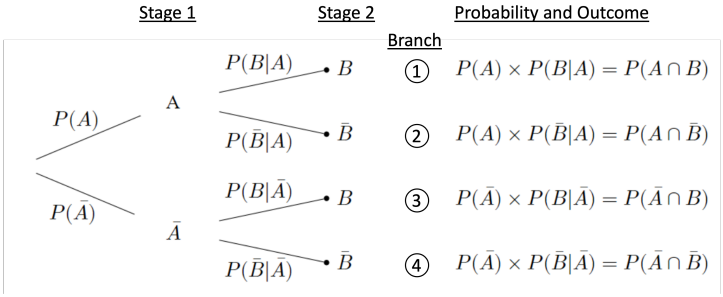
- Definition: **Law of total probability**

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

  $P(B) =$

- Using the general tree diagram below, we can summarize

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




## Bayes' Theorem

Bayes' Theorem

- Using the same terminology, we can summarize

  Bayes' Theorem = $P$(First stage event | 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.

- Definition: **Bayes' Theorem**

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

  $P(A_i \mid B) =$

- 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.

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:

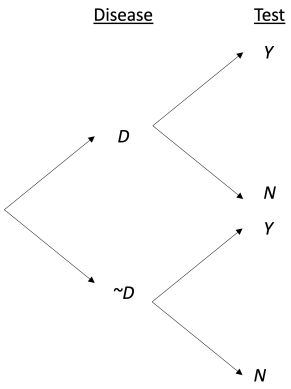
  $D$ = the person tested has the disease
  $\sim D$ = the person tested does not have the disease
  $Y$ = the test is positive
  $N$ = the test is negative

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

a. If you have the disease, 95% chance of a positive test.

b. 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

  1. Find the probability of testing positive (total probability).

  2. Lets' solve for the probability of having the disease *given that you test positive.*

     (a) 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).

     (b) Then this person got tested, and tested positive; this is our evidence.

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

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

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?

- The practical information here is interesting.
  1. The "95% accurate" test will classify _____ of the population as positives, compared to the true prevalence of $P(D) = 0.01$.
  2. By updating our prior probability with the new evidence, we drastically increased our information about this person having the disease. This is the good side of Bayes' Theorem!
  3. _____ of the individuals who tested positive will actually _____ have the disease.

     This percentage depends heavily on the prevalence, for example if $P(D) = 0.1 \rightarrow P(\sim D \mid Y) =$ _____; and if $P(D) = 0.001 \rightarrow P(\sim D \mid Y) =$ _____.

Final example

- 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

**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?

  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 $HT$.
  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 | Probability |
|---|---|---|
| Input: | | |
| Output: | | |
| Maps: | | |

- Notation: We will use uppercase letters, such as $X$, $Y$, $Z$, ... to denote a random variable and lowercase letters, such as $x$, $y$, $z$, ... 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 $x_i = \{x_1, x_2, \ldots\}$ 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 $X$?

(b) Show the connection between $S$ and $X$.

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

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 _____ is a finite or countable set.

  $X$ is a **continuous random variable** if the _____ 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 _____ 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 _____. 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:

  $\{X = 1\} =$

  $\{X = 3\} =$

# 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.

b) Determine the probability of at most two claims.

c) Determine the probability of at least two claims.

Probabilities for continuous random variables

- For continuous random variables, does the pmf exist?

  These intervals are not countable, so we _____ 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

$$P(a \leq X \leq b) = \int_a^b f(x)\,\mathrm{d}x$$

- Notes:
  - There is no probability associated with a single point:

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

  - 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).

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

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

  Find: (a) $P(X \leq 1)$ $\qquad$ (b) $P(X \geq 1)$ $\qquad$ (c) $P(0.5 \leq X \leq 1.5)$.

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

  (a) $f_X(x) \geq$ for all $x$.

  (b) $\sum_x f_X(x) =$ (pmf) or $\int_{-\infty}^{\infty} f_X(x)\,\mathrm{d}x =$ (pdf).

- 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 $f(x) = 2, \quad 0 \leq x \leq 0.5$.

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

- Examples:

  1. Verify the pmf for $X$ is valid.

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

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

  3. 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.

## The cumulative distribution function

Concept of a cdf

- Examples: Discrete case

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

Continuous case

$$f_X(x) = \tfrac{3}{8}x^2 \quad 0 \le x \le 2$$

- 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

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

- Notes about $F(x)$: ALWAYS
  - The cdf is defined for $-\infty < x < \infty$ always.
  - The range of every cdf is $0 \le F(x) \le 1 \iff$ Limits:
  - $F_X(x)$ is a non-decreasing function.

- DISCRETE case

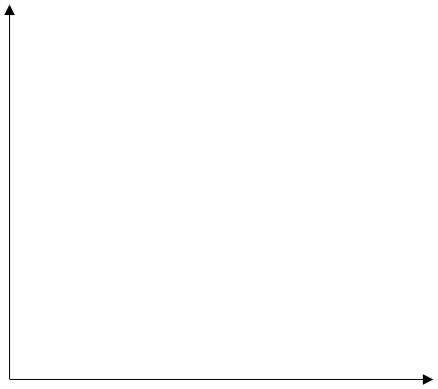
  Example:

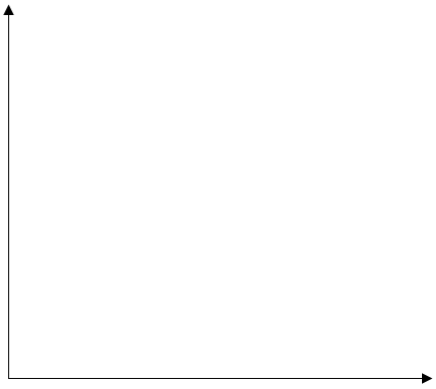
  (a) Using the pmf table for $X$ below, find the cdf $F_X(x)$ as a table.

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

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

(c) Plot the pmf and cdf.





Observations of pmf — Properties of cdf

1) Sum of the heights = _____ $\Longrightarrow$

2) Positive values (probabilities) only at _____ $\Longrightarrow$

3) No probability _____ $x$ values $\Longrightarrow$

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

$$f(x) = \begin{cases} \frac{3}{8}x^2 & 0 \le x \le 2 \\ 0 & \text{otherwise} \end{cases}$$

  (a) Find the cdf $F(x)$.

(b) Plot the pdf and cdf.

| Observations of cdf | | Properties of cdf |
|---|---|---|
| 1) Plot starts at _____ and ends at _____ | $\Longrightarrow$ | |
| 2) _____ at change points | $\Longrightarrow$ | |

Properties

- $F_X(\text{lower limit}) = 0$ 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** $\Longleftrightarrow$ $F_X(x)$ is a step function of $x$.
  - A random variable $X$ is **continuous** $\Longleftrightarrow$ $F_X(x)$ 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.

$$F'(x) = f(x)$$

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

$$F_X(x) = \int_{-\infty}^{x} f(t)\,\mathrm{d}t \qquad \text{for all } x.$$

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

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

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

- Example: Let $X$ have the following cdf:

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

Find the pdf $f(x)$.

Finding probabilities using the cdf

- ALWAYS
  - Cdf gives a _____ 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:

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

    which represents the probability $X$ is greater than $x$, _____.

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

- 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 $x$ 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 $\mathrm{d}t$).

$$F_X(x) = \int_{-\infty}^{x} f(t)\,\mathrm{d}t$$

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

$$F(a) = \int_{-\infty}^{a} f(x)\,\mathrm{d}x$$

- The complement of this is:

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

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

- 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) $P(X < 4)$

(e) $P(2 \leq X \leq 4)$ (f) $P(1 < X \leq 5)$

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

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

(e) $P(0.25 \leq X \leq 1)$ (f) $P(1 < X \leq 1.5)$

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 $P(X = 3)$.

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

  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:

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

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

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

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

$P(X \ge 0.1) =$

$P(0 < X < 0.1) =$

$P(-0.1 < X < 0.1) =$

$P(\{X < -0.1\} \cup \{X > 0.1\}) =$

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.

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

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

*STRATEGY:* Find the cdf in cases.

Case 1: $0 \le x \le 0.05 \qquad \Longrightarrow \qquad f(x) = 560x$

Case 2: $0.05 \le x \le 0.25 \qquad \Longrightarrow \qquad f(x) = -15x + 3.75$

# 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 summaries the information.

  Sample of _____ data points $\longrightarrow$ _____ 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.

  Written out like this, we can think of _____ 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 |

  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 _____. Let's calculate it:

- What we are actually calculating here is called the _____, which we can think of as a _____ of the _____ where the _____ 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

  $$E(X) =$$

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

  $$E(X) =$$

  The expected value of the random variable $X$ is often denoted by the Greek letter $\mu$.

  $$E(X) =$$

- Examples

  1. 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.

  2. Let $X$ be the loss severity random variable for the warranty policy with density:

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

     Find the expected loss.

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.

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 _____.

  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 ($y$) | | | | |
| $f(y)$ | | | | |

(a) Find the expected total claim amount.

Theorems

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

$$E(aX) =$$

  We can derive this!

- 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.

| $x$ | $f(x)$ | $z =$ | $f(z)$ |
|---|---|---|---|
| 0 | 0.28 | | |
| 1 | 0.43 | | |
| 2 | 0.20 | | |
| 3 | 0.09 | | |

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

$$E(aX + b) =$$

Discrete derivation:

Continuous derivation:

- Simple example: Continuing previous example 2. Suppose that due to inflation the losses on the warranty policy are expected to increase by 5% and have an additional fixed cost of \$30 for the following year. Find the expected value for the following year.

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

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

(Discrete) Derivation:

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 |

More complex because $g(X)$ is _____ function.

- This example illustrates two major points.
  1. 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 $Y = g(X)$.
  2. 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

  $$E(Y) =$$

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

  $$E(Y) =$$

- 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 $E[X(5 - X)]$.

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

## Variance and standard deviation

Measures of spread

- The mean of a random variable gives a nice single summary number to measure _____ 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.
  (a) 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 |

Means are the _____, 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?

(b) We could look at the distance between each individual value of $x$ or $y$ from the mean of its distribution.

This is always true for any random variable.

(c) 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 |

- 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

$$V(X) =$$

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

$$\sigma =$$

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

- Definitions:

  Discrete  Continuous

- Examples:
  1. Let $X$ be the random variable with $f(x) = (3/8)x^2 \quad$ for $0 < x < 2$. Find $\sigma_X^2$.

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

   Find the mean, variance and standard deviation of $X$.

Expectation as a linear operator

- Now lets 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.

- 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:

  $\int [g(x) + f(x)]\,\mathrm{d}x =$

  $\frac{\mathrm{d}}{\mathrm{d}x}\left[g(x) + f(x)\right] =$

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.

- Theorem: Another way to calculate variance:

$$V(X) =$$

- Examples:
  1. 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.

  2. Find the variance of the warranty loss random variable using the alternate formula.

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

Calculating the variance from the definition would require evaluation of the integral:

$$V(X) = \int_0^{100} \left(x - \frac{100}{3}\right)^2 (0.02 - 0.0002x)\,\mathrm{d}x$$

This would be straightforward, but time consuming relative the the other way if done by hand. Of course with computing, calculation time is not an issue.

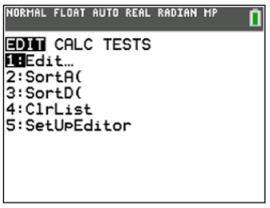
- Important practical note: The calculation of variance can be done more easily by hand using this alternate form rather than the definition.

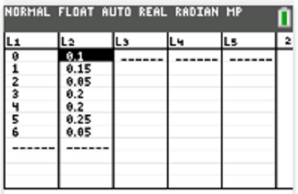
  But using the definition is more efficient for computers when large values of $X$ are present (problems with overflow due to the magnitude of $X^2$).

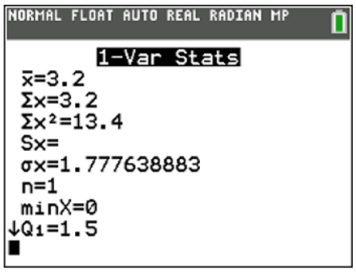
Technology session

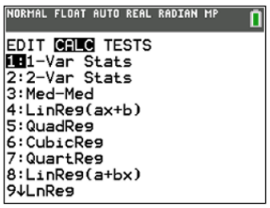
- Using TI-84 (and TI-30XS MultiView) to calculate $E(X)$ and $SD(X)$.

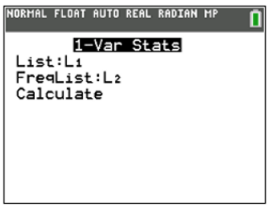
| x | f(x) |
|---|---|
| 0 | 0.10 |
| 1 | 0.15 |
| 2 | 0.05 |
| 3 | 0.20 |
| 4 | 0.20 |
| 5 | 0.25 |
| 6 | 0.05 |



**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 $V(aX)$:

- Theorem: For any constant $a$ and random variable $X$, (alternate notation)

  $$V(aX) = \qquad\qquad \sigma^2_{aX} =$$

  The standard deviation of $aX$ can now be obtained by taking the square root.

  $$SD(aX) = \qquad\qquad \sigma_{aX} =$$

- Intuitively, here's why the $a$ is squared for $V(aX)$:

- Example: Using the previous quiz scores from class 1, lets 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%.

– Curve 2: Every score gets and additional 2 points.

| Original score ($x$) | 7 | 8 | 9 |
|---|---|---|---|
| Curved score ($c_2$) | | | |
| $f(c_2)$ | 0.2 | 0.6 | 0.2 |

- 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 $X$,

$$V(aX + b) =$$

- 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.

  (a) Define a new random variable $Y$ = grade after test corrections.

| Grade ($x$) | $f(x)$ |
|---|---|
| 60 | 0.02 |
| 65 | 0.04 |
| 70 | 0.17 |
| 75 | 0.23 |
| 80 | 0.21 |
| 85 | 0.15 |
| 90 | 0.12 |
| 95 | 0.05 |
| 100 | 0.01 |

(b) Calculate $E(Y)$ and $SD(Y)$.

Short way $\rightarrow$ $E(Y)$ & $V(Y)$ indirectly

1) Find $E(X)$ and $V(X)$ first

2) Use formulas to find $E(Y)$ and $V(Y) \Rightarrow SD(Y)$

Long way $\rightarrow$ $E(Y)$ & $SD(Y)$ directly

1) Find new pmf of $Y$

2) Then find $E(Y)$ and $V(Y)$

Final comparison: Test corrections have _____ mean and _____ 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:

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 _____, which is not a good measure because $X$ changes every time.

  $E[(X - b)^2]$ is _____, which makes it a good measure because it never changes.

- 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 $g(b) = E[(X-b)^2]$:

  (b) Goal: Find $b$ to minimize step result of step (a).

  (c) Confirm minimum with second derivative test:

**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

1. Let $X$ have the following cdf:

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

Find the median $m$.

2. 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):

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

$$F(x) = \int_0^x 0.02 - 0.0002t \, \mathrm{d}t \; = \; 0.02x - 0.0001x^2, \qquad 0 \leq x \leq 100$$

Find the median $m$.

This has a nice intuitive interpretation. Half of the losses will be less than _____ and the other half will be greater.

3. Symmetric densities example: Find the expected value and the median for the ROI example, where

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

Results:

If the density function is symmetric, the median can be found without calculation; specifically, the median $m$ is equal to the point of symmetry.

In general, the median and mean are not equivalent. But when the density function is _____, these two measures of center are _____.

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 $50^{\text{th}}$ percentile.

  Other percentiles can be defined using similar reasoning. For example, the $90^{\text{th}}$ 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 $0 \le p \le 1$. The **$100p^{th}$ percentile** of $X$ is the number $x_p$ defined by

$$F(x_p) = p$$

- Special percentiles: **Quartiles** split area under density curve into quarter.

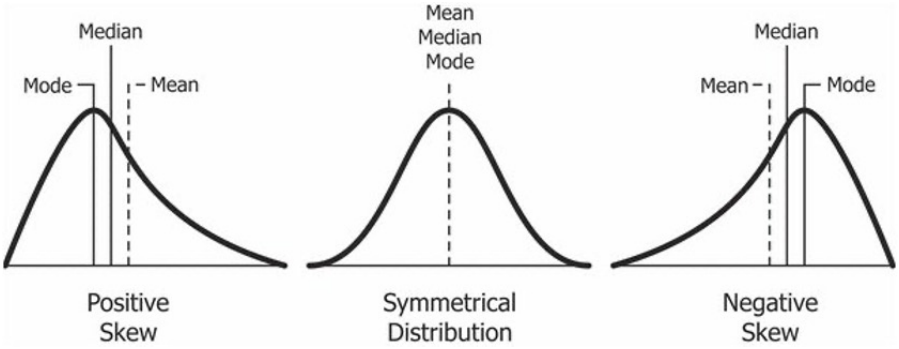
  $x_{0.25} = 25^{th}$ percentile

  $x_{0.50} = 50^{th}$ percentile

  $x_{0.75} = 75^{th}$ percentile

- 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:

4. Find the IQR of $X$ using the following cdf:

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

5. The time $X$ in months until failure of a certain product has the pdf
   $f(x) = (1/2)\, x\, \mathrm{e}^{-(x/2)^2}$ for $0 < x < \infty$.
   Find the first and second quartiles of $X$.

# Test 3

## Contents



## Lecture 10 – Discrete Distributions

**MATH 320: Probability**

# Lecture 10: Discrete Distributions

Chapter 2: Distributions (2.1, 2.3, 2.4, 2.5, 2.6, 2.7)

## Introduction

- Statistical distributions are used to model _____.

  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 _____, 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 of symmetric and bell-shaped, then a _____ distribution is a reasonable choice. Then we will specify the parameters _____.

- 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 _____ 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

$$P(X = x \mid N_0, N_1) = \frac{1}{N_1 - N_0 + 1}, \qquad x = N_0, \ldots, N_1,$$

  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:

  $$\text{RV} \sim \text{Distribution (parameters)} \qquad X \sim \text{Discrete uniform } (N_0, N_1)$$

  "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.

  $P(X = x \mid N_0, N_1) =$

- 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: $\mid$ (given).

  In general, we have $P(X = x \mid \theta)$, where $\theta$ could be a vector of parameters.

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 $(N_0, N_1)$,

  $$E(X) = \frac{N_0 + N_1}{2} \qquad V(X) = \frac{(N_1 - N_0 + 1)^2 - 1}{12}$$

- 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: Confirm with definition:

Summary

- If we know (or assume) the family of a distribution and the values of the parameters

  $\iff$ know _____ $\iff$ _____

  We also know the expected value and variance as well.

- This statement works backwards too.
- *IMPORTANT STRATEGY* when solving problems:

  If we recognize that we have a pmf where the 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?
  1. Suppose $f_X(x) = \frac{1}{4}, \quad x = 7, 8, 9, 10$

     Find $E(X^2)$.

  2. Suppose the values 3, 6, 9, 12, 15 are equally likely.

  3. Suppose the values 5, 9, 13, 17 are equally likely.

**Bernoulli distribution**

Motivation

- 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.
  2. Roll a die and see if a number greater than 4 appears.
  3. Choose a student in this class at random and see if a female student is chosen.
- What is the common characteristic of these experiments?

- The main event of interest is labeled _____ and the other is labeled _____.

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 _____ and _____ by the Bernoulli random variable, respectively.

- Definition: A random variable $X$ has a **Bernoulli** $(p)$ distribution if

$$P(X = x \mid p) = \begin{cases} 1-p & x = 0 \\ p & x = 1 \\ 0 & \text{otherwise} \end{cases}$$

  where the parameter $p$ represents the probability of success and $0 < p < 1$.

  Equivalently, $f(x \mid p) = p^x(1-p)^{1-x}, \quad x = 0, 1.$

- Notation: $X \sim \text{Bernoulli}\,(p)$

- Example: A basketball player shoots a free throw with a 75% probability of success. Let $X$ denote the number of points scored.

  (a) What is the distribution of $X$?

  (b) Find the pmf of $X$.

  (c) Find the expected value of $X$ (by hand using the definition).

  (d) Find the variance of $X$ (by hand using the definition).

Mean and variance

- If $X \sim \text{Bernoulli}\,(p)$

$$E(X) = p \qquad\qquad V(X) = p(1-p) = pq$$

- Derive $E(X)$. *HINT*: Generalize the calculations from the previous examples.

- Derive $V(X)$.

** 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?

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.

  Said another way, a binomial experiment is a sequence of _____.

- Definition: If $X$ is the number of successes in a binomial experiment, $X$ is called a **binomial random variable**.

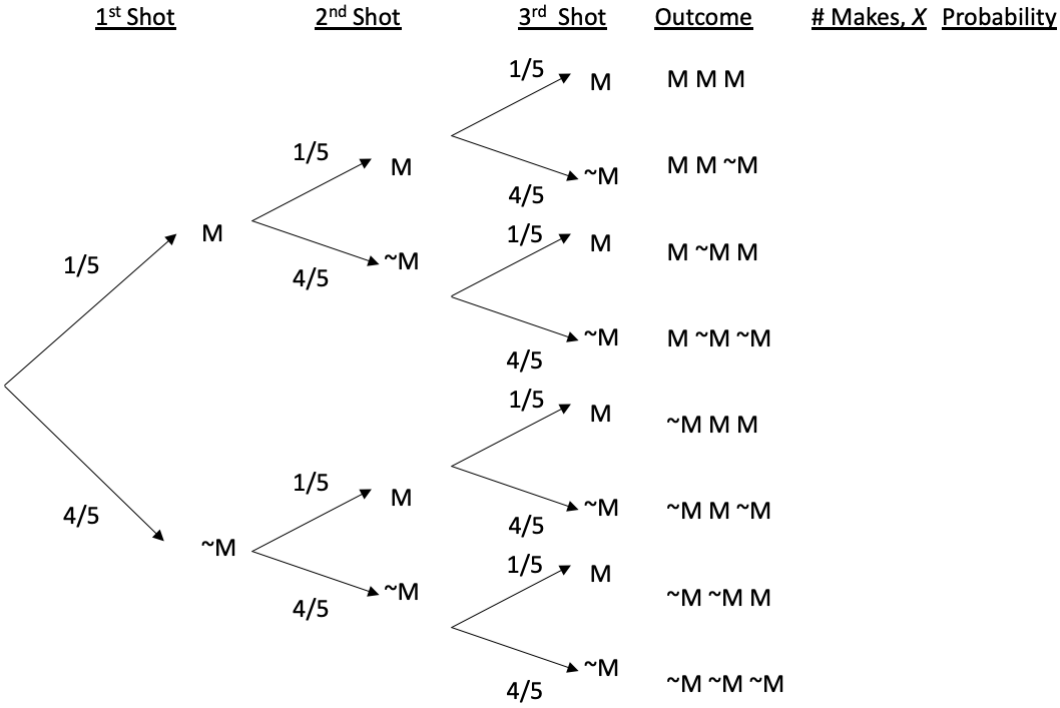
- Examples:

  (a) Check the conditions to determine if example experiment 1 is indeed a binomial experiment.

  (b) Are the conditions met for experiments 2 and 3? If so, identify $n$ and $p$.

Binomial probabilities

- Probabilities for binomial distribution are just an application of the multiplication rule for independent event.
- Example: Lets say our basketball player has only 3 shot attempts now. Lets visualize the tree diagram illustrating the different outcomes of this binomial experiment ($M$ = make, $\sim M$ = miss).




(a) Find $P(X = 3)$.

(b) Find $P(X = 2)$.

Another way to think about the number of branches:

_____ that include exactly 2 makes.

This could also be found using _____.

(c) Now suppose $n = 8$. Find $P(X = 4)$.

$=$ sequences $=$ ways to choose 4 success out of the $n = 8$ trials.

Now just need the probability of 4 success and 4 failures:

- 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

$$P(X = x \mid n, p) = \binom{n}{x} p^x (1-p)^{n-x}, \qquad x = 0, 1, \ldots, n$$

where $0 < p < 1$ and $n$ is an integer such that $n \geq 1$.

- Notation: $X \sim \text{Binomial}\,(n, p)$

- 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 $X$?

b) Find the pmf of $X$.

c) Find $P(X = 7)$.

d) Find the probability of at least one correct.

e) Find $P(X < 6)$.

f) Find $P(X \geq 6)$.

- Binomial cdf:

  Right sided probability:

Mean and variance

- If $X \sim \text{Binomial}\,(n, p)$

$$E(X) = np \qquad\qquad V(X) = np(1-p) = npq$$

- Continuing example:

  (f) Find the expected value and variance of the number of questions guessed correctly.

  (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:
  * $Z$ = number of points missed:

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 the number of heads in the 10 flips.

   (a) Give the pmf of $X$, $\mu_X$ and $\sigma_x^2$.

   (b) Find: $P(X = 3)$ $P(X < 3)$ $P(X \geq 3)$.

   (c) If $Y = 10 - X$ give the distribution of $Y$. Find $P(Y \leq 1)$.

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.

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:

  $(0, 0, 1, 0, 0)$

  $(1, 0, 1, 1, 0)$

  $(1, 1, 1, 1, 1)$

- Suppose that $X \sim \text{Bin}(n, p)$ and $Y_1, \ldots, Y_n \overset{\perp\!\!\!\perp}{\sim} \text{Ber}(p) \iff Y_i \overset{iid}{\sim} \text{Ber}(p)$ where $iid$ = independent and identically distributed.

  $$X =$$

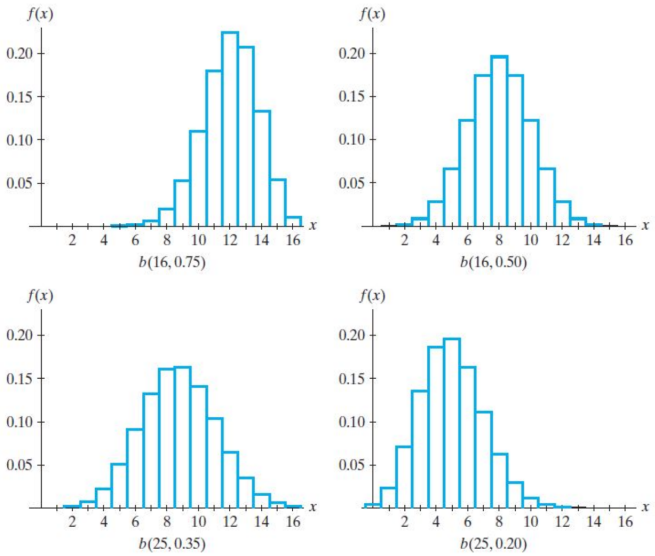
  Bernoulli distribution is a special case of Binomial when

  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 _____.

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?

Geometric experiments and geometric random variables

- 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: $S =$
- Geometric is a waiting time random variable, with the number of successes 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.
  2. The random variable of interest $Y$ is the number of failures before the first success.
- Relationship between the two formulations.
  - The first way $X$ includes the trial on which the success occurs, whereas the second way $Y$ does not.
  - 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.

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:

  2. Success on second trial:

  3. Success on third trial:

- If $X \sim$ Geometric $(p)$,

$$P(X = x \mid p) =$$

  Alternate form: $Y =$

Mean and variance

- If $X \sim$ Geometric $(p)$,

$$E(X) = \frac{1}{p}, \qquad V(X) = \frac{1-p}{p^2} = \frac{q}{p^2}$$

  Alternate form:

- Continuing example: Let $X \sim$ Geometric$(p = 1/5)$. How many trials do you expect to see the first success (make)?

Geometric probabilities

- A geometric random variable is a geometric series:

- **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) $\displaystyle\sum_{i=0}^{\infty} q^i = q^0 + q^1 + q^2 + \cdots = \frac{1}{1-q}$ (b) $\displaystyle\sum_{i=1}^{\infty} q^i = \frac{q}{1-q}$

    *PATTERN*: Numerator = _____ and denominator = _____.

  - Another sum:

    (c) $\displaystyle\sum_{i=0}^{m} q^i =$

- 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: $P(X < \infty) =$ .

    Derive this:

  - Cdf: $P(X \leq x) =$

Intuitive derivation of cdf:

- Right probability (exclusive): $P(X > x) =$

- Right probability (inclusive): $P(X \geq x) =$

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

- Interval probability (both inclusive): $P(a \leq X \leq b) =$

- 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 $X$?

  b) Find the pmf of $X$.

c) Find the probability that an engine will survive 3 hours.

d) Find the probability that an engine will survive 3 hours, but die before or during the $6^{\text{th}}$ hour.

e) Suppose that an engine survives 3 hours. Find the probability that the engine will survive 6 hours.

Memoryless property

- The geometric distribution has an interesting property, known as the "memoryless" property. For integers $s > t$, it is the case that

$$P(X > s \mid X > t) = P(X > s - t)$$

- To help us understand what this means, let's interpret some events for $s > t$.
  - $\{X > s\}$
  - $\{X > t\}$
  - $\{X > s\} \mid \{X > t\}$

- For example, lets 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?

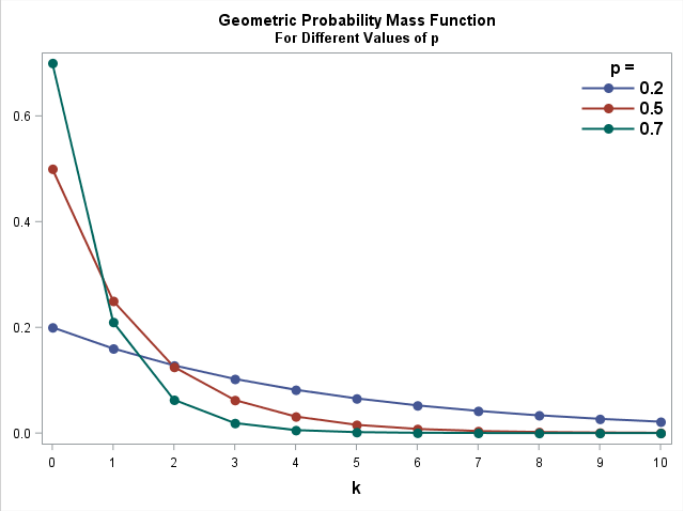
- We want to know how (if it all) this past information influences the future event. Recall $P(X > a) = q^a$.

  $P(X > s \mid X > t) =$

- 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 4 shots. Let $X$ be the number of shots it took to make the $4^{\text{th}}$ 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 $3^{\text{rd}}$ productive well.
- What is the common characteristic of these experiments?

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

- Example: Check the conditions to determine if example experiment 1 is indeed a negative binomial experiment.

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 $x - r$ failures.
     B) The last $x^{\text{th}}$ trial has to result in success.

  2. Lets 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 $x = r, r + 1, \ldots$.

     A) $A \sim$
        $P(A) =$

     B) $B \sim$
        $P(B) =$

  3. Putting these together:

     $f_X(x \mid r, p) = P(A \cap B) =$

- If $X \sim$ Negative binomial $(r, p)$,

$$P(X = x \mid r, p) = \binom{x-1}{r-1} p^r \, q^{x-r} \qquad x = r, r + 1, \ldots$$

- 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 $X \sim$ Negative binomial $(r, p)$,

$$E(X) = \frac{r}{p}, \qquad V(X) = \frac{r(1-p)}{p^2} = \frac{rq}{p^2}$$

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 $X$?

  b) Find the pmf of $X$.

  c) Find the probability that the second win occurs on the $8^{\text{th}}$ play.

  d) Find $E(X)$ and $V(X)$.

  e) Find the probability that the fourth win occurs before the $10^{\text{th}}$ play.

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 $Y_1, \ldots, Y_r \overset{\perp\!\!\!\perp}{\sim} \text{Geo}(p)$.

  $$X =$$

- We can see this for the expected value and variance too.

- 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.

  - 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.

## 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.

- 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 $p$?
  - The usual method of sampling voters is called **sampling without replacement**.

  - 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?

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.
  - 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.
  - 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 | | 0.102 | 0.021 | 0.002 |

  - Again, successive selections are dependent on previous one. So the probabilities change.
  - Now we can formalize this example and relate it to how we will talk about hypergeometric experiments and random variables.

Hypergeometric experiments and hypergeometric random variables

- Characteristics of a **hypergeometric experiment**.
  1. Each trial is a Bernoulli experiment.
  2. Trials are not identical and not independent.
  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.
  - Three parameters:
    (a) $N$ = Population size
    (b) $M$ = Number of objects of interest
    (c) $K$ = Sample size
  - Lets visualize this scenario.

Definition

- If $X \sim \text{Hypergeometric}\,(N, M, K)$,

$$P(X = x \mid N, M, K) = \frac{\binom{M}{x}\binom{N-M}{K-x}}{\binom{N}{K}}, \qquad x = 0, 1, \ldots, \min(M, K)$$

- 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: $\mathcal{X} = \{0, 1, \ldots, \min(M, K)\}$.
  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: $\mathcal{X} = \{\max(0, K + M - N), \ldots, \min(M, K)\}$.

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.

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?

Mean and variance

- If $X \sim \text{Hypergeometric}\,(N, M, K)$,

$$E(X) = K\left(\frac{M}{N}\right), \qquad V(X) = K\left(\frac{M}{N}\right)\left(\frac{N-M}{N}\right)\left(\frac{N-K}{N-1}\right)$$

  Interpretation of the mean: We want the sample to have the same propoertion 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$(n, p)$.

  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 _____.

  And the sample size $K$ is like the number of trials .

  Then the mean and variance of $\text{HG}(N, M, K)$ can be re-expressed like this:

$$E(X) = \qquad\qquad V(X) =$$

- Observations when comparing with vs without replacement:

  1. Means: The expected values are the same regardless of the type of sampling.
  2. Variances: The variance is somewhat smaller when sampling without replacement.

     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 $X$?

   (b) Find the pmf of $X$.

   (c) Find the probability of two or three spades in the hand dealt.

   (d) Find the probability of at least two spades in the hand dealt.

   (e) Find the probability of at most 4 spades in the hand.

   (f) Find $E(X)$ and $V(X)$.

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?

   (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?

Relationship between hypergeometric and binomial

- Hypergeometric experiment is _____ replacement and binomial is _____ replacement.
- As the population size goes to infinity ($N \to \infty$), 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.

  – The last term in the variance of HG goes to 1, and we are left with what is equivalent to the binomial variance.

  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.

  2.

  3.

- 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.

    Interested in .

    are given as parameters.

    Only difference is _____.

  - Distributions counting the number of trials.

    Interested in .

    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?

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 $\lambda$ 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

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 $\lambda$ the average rate.

- Definition: A random variable $X$ follows the Poisson distribution with parameter (or average rate) $\lambda$ if

$$P(X = x \mid \lambda) = \frac{\mathrm{e}^{-\lambda}\lambda^x}{x!}, \qquad x = 0, 1, 2, \ldots \quad \text{and} \quad \lambda > 0$$

  Note: $\lambda$ is the expected value of the number of occurrences per unit.

- Notation: $X \sim \text{Poisson}\,(\lambda)$

- Theorem (Combining Poisson random variables): If $X_i \overset{\perp\!\!\!\perp}{\sim} \text{Poisson}\,(\lambda_i)$ (not necessarily identical because $\lambda_i$s can be different), then

$$Y = \sum_{i=1}^{n} X_i$$

  If $X_i$'s are identical and therefore *iid*, then the new parameter for $Y$ becomes _____ and the number of occurrences in $n$ units is

Mean and variance

- If $X \sim \text{Poisson}\,(\lambda)$,

$$E(X) = \lambda \qquad\qquad V(X) = \lambda$$

- It should be obvious that the expected value is equal to $\lambda$, 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 $\lambda = 2$.

   Now we can easily find probabilities and the mean and variance.

   (a) Find the probability there are no accidents this month.

   (b) Find the probability there is one accident this month.

   (c) Find the probability there are more than two accidents this month.

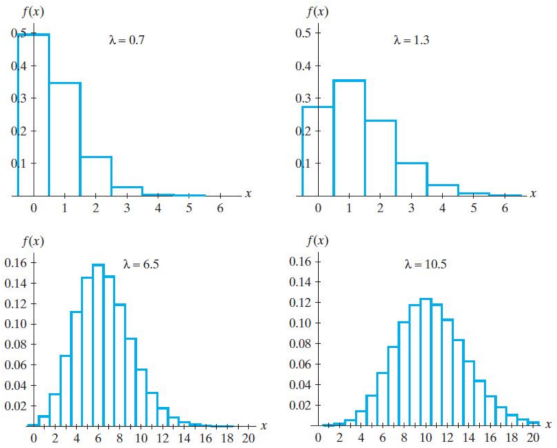
   (d) Suppose each accident costs the driver \$350, find the mean and variance for the cost of accidents this month.

   (e) Find the probability there is 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.

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?

Visualizing Poisson distributions




Poisson approximation to the binomial distribution

- In grad school :), you will learn that Poisson is the is the limiting distribution of binomial. Said slightly more formally:

$$\lim_{n\to\infty} \text{Binomial}(n, p = \frac{\lambda}{n}) = \text{Poisson}(\lambda = np)$$

- This means that for large $n$ and small $p$, the Poisson distribution can be used to approximate the binomial distribution.

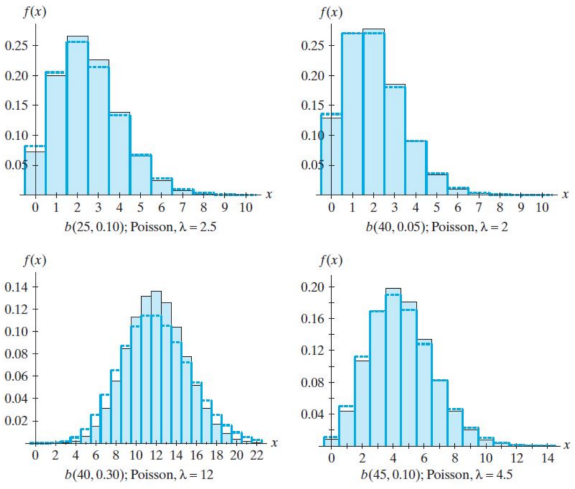



**Figure 2.6-2** Binomial (shaded) and Poisson probability histograms

# Lecture 11 – Continuous Distributions

**MATH 320: Probability**

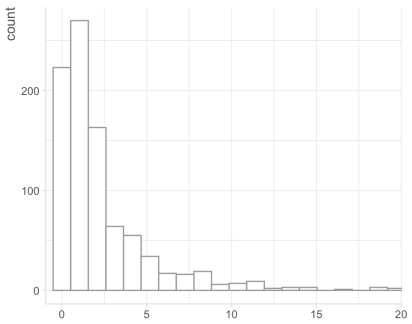
# Lecture 11: Continuous Distributions

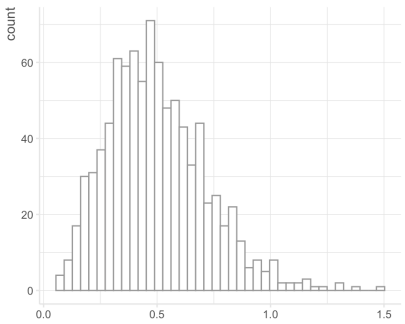
Chapter 3: Distributions (3.1, 3.2, 3.3)

## 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 me model the population that these losses are coming from?

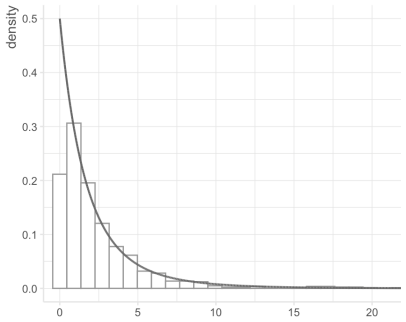
  Said another way: In theory there is some distribution that these losses follow, and we want to figure out which family it is and what the parameter values are $\Longrightarrow$ Find equation for smooth curve.

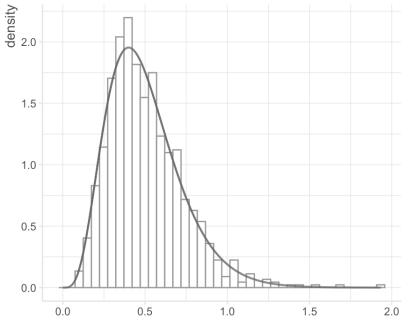
- Once the researcher knows this, they can figure out lots of things that are useful in analyzing how losses will occur in the future. For example (suppose $X$ is in thousands of dollars):
  - Can find the expected claim amount $E(X)$. Then use this to set a premium amount to ensure at least breakeven.
  - 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:
  - $E(X) \approx \dfrac{1}{n}\sum_{i=1}^{n} x_i$
  - $P(X > 15 = \$15,000) \approx$
  - $P(X < 2 = \$2,000) \approx$

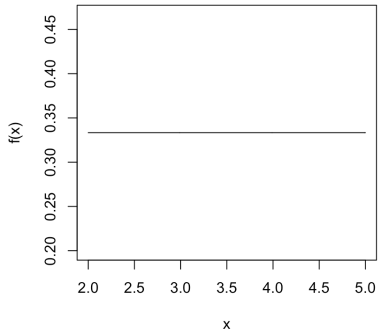
- But these are just approximations and are based on a single sample of data. Having population information is much better and much more generalizable!

  Observation: When we overlay these density curves, they match the histograms excellently.



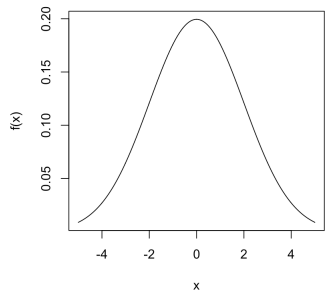




- Here is an overview of different population shapes and the distributions families that can be used to model them.
- If a population is distributed evenly between two numbers like



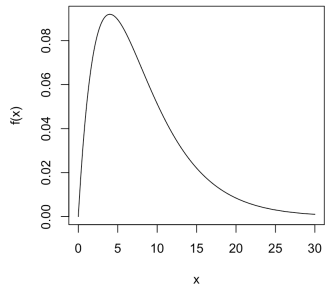

_____ distribution must be used.

- If a population is bell-shaped and symmetric like



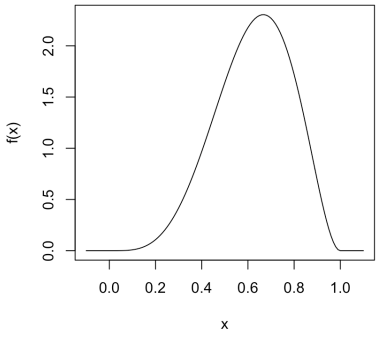

_____ or _____ distributions can be used, among others.

- If a population is skewed like




_____, _____ or _____ distributions can be used, among others.

- If a population has bounded range (support) between two points and is not evenly distributed




_____ distribution can be used. This is useful when modeling _____.

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 $X \sim \text{Uniform}\,(a, b)$
$$f(x \mid a, b) = \begin{cases} \dfrac{1}{b-a} & a \leq x \leq b \\ 0 & \text{otherwise} \end{cases}$$
where the parameters $a$ and $b$ are real numbers.
- Characteristics of a uniform distribution.
  - Constant probability over the entire interval.
  - Bounded support.
  - Symmetric.

Probabilities

- The probability of any subinterval in the range (support) is proportional only to the _____ of the subinterval.

- For $a \le c \le d \le b$
$$P(c \le X \le d) =$$

- We can generalize this to find $P(X \le x)$ for values of $x$ in the interval $[a, b]$.
$P(X \le x) =$

- Then we can define the cdf $F(x)$ for a uniform random variable $X$ on $[a, b]$ by:
$$F(x \mid a, b) = \begin{cases} 0 & x < a \\ & a \le x \le b \\ 1 & x > b \end{cases}$$

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
$$f(t) = \begin{cases} 1/100 & 0 \le t \le 1 \\ 0 & \text{otherwise} \end{cases} \quad \text{and} \quad F(t) = \begin{cases} 0 & t < 0 \\ t/100 & 0 \le t \le 1 \\ 1 & t > 100 \end{cases}$$
  - 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

  - 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.

- 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

$$S(t) = P(T > t) = 1 - F(t)$$

- The survival function for a uniform random variable is

Mean and variance

- If $X \sim \text{Uniform}\,(a, b)$

$$E(X) = \frac{a+b}{2} \qquad\qquad V(X) = \frac{(b-a)^2}{12}$$

Summarizing example

- Let $T$ be the time in months from initial use of a machine part until failure, where $T \sim \text{Uniform}\,(1, 9)$.

  a) Find the pdf, cdf and survival function of $T$.

b) Find the probability the machine part fails between 5 and 7 months.

c) Find the probability the machine part lasts longer than 4 months.

d) Find the expected value and standard deviation for the lifetime of the machine part.

**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 $\lambda$, 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.

This process is the same as finding $c$ such that:

- 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. $\mathrm{e}^{-\lambda x}$), then $c$ converts it to a valid pdf.

Definition

- If $T \sim \text{Exponential}\,(\lambda)$

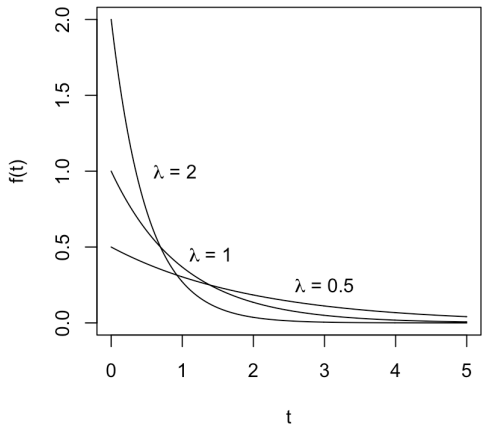
$$f(t \mid \lambda) = \lambda \mathrm{e}^{-\lambda t}, \qquad t \geq 0 \quad \text{and} \quad \lambda > 0$$

- Characteristics of an exponential distribution.
  - Gives the probability for the waiting time between successive Poisson events.
  - Right-skewed density function.
  - Unbounded support.
  - The exponential random variable is a continuous version of the _____ random variable, which waits for the first _____ in a discrete sample space.
  - Also has the _____ property: $P(T > a + b \mid T > a) = P(T > b)$.

    Exponential and geometric are the only distributions with the memoryless property.

Parameter

- As stated, $\lambda$ is the rate at which events occur in a Poisson process. Here is how it affects the exponential density function.
- As $\lambda$ 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 $\lambda$, 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.
    1. The definition given above uses the **rate parameterization** of the exponential distribution.
    2. There is also a **scale parameterization**, where the scale parameter $\theta$ is defined by

       $$\text{scale } \theta = \frac{1}{\text{rate}} = \frac{1}{\lambda} \qquad \Longrightarrow$$

  - 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 $f(t)$.

  b) Cdf $P(T \leq t)$: Find the probability that the waiting time for the next accident is less than 2.

  c) Survival $P(T > t)$: Find the probability that the waiting time for the next accident is longer than 1 month.

  d) Interval $P(a \leq X \leq b)$: Find the probability that the waiting time for the next accident is between 0.5 and 1.5 months

- Cdf and survival function:

  If $T \sim$ Exponential $(\lambda)$

  $$F(t) = \qquad\qquad S(t) =$$

  Note that we could derive these formally:

- 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

$$P(a \leq X \leq b) =$$

For the exponential distribution, we have:

$P(a \leq T \leq b) =$

Mean and variance

- If $T \sim \text{Exponential}\,(\lambda)$

$$E(T) = \frac{1}{\lambda} \qquad\qquad V(T) = \frac{1}{\lambda^2}$$

- Example: Find the mean and variance for the previous accident example:

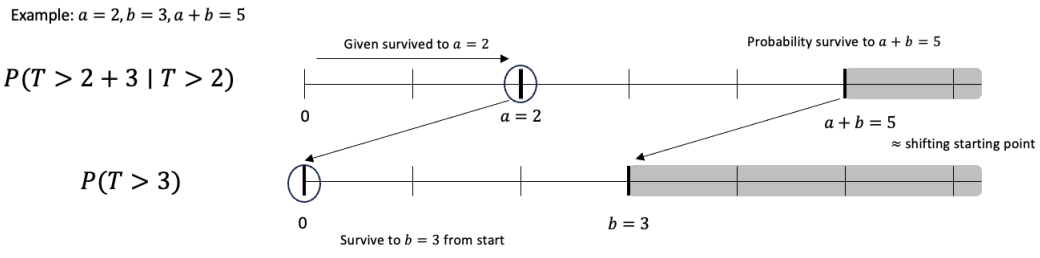
Memoryless property

- Just like the geometric distribution, the exponential distribution has the **memoryless property**.
- Theorem: If $T \sim \text{Exponential}\,(\lambda)$

$$P(T > a + b \mid T > a) = P(T > b)$$

- (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.



- Proof:

- 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.

  (b) Given that the part has operated 100 hours, find the probability it will operate for $100 + x$ hours $(0 \leq x < \infty)$.

- 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 $\sim$

  – Remaining lifetime of a 100 hr old part $\sim$

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 $\lambda$, then the number of events in a time period of length $t$ is a Poisson random variable with parameter _____. 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:

  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

- 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?

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.
  - 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.
  - Warranty 2 pays $3e^{-T}$ at time $T$ years.

  Find the expected warranty payment under each of the two warranties.

**Gamma distribution**

(Brief) Motivation

- Relationship between the exponential distribution and the gamma distribution:

- The gamma distribution can also 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

- If $X \sim \text{Gamma}\,(\alpha, \beta)$

$$f(x \mid \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)}\, x^{\alpha-1}\, \mathrm{e}^{-\beta x}, \qquad x \geq 0 \quad \text{and} \quad \alpha, \beta > 0$$

- Intuitive idea behind pdf:
  - The main part of $f(x)$ is $x^{\alpha-1}\, \mathrm{e}^{-\beta x}$.

    As $x \to \infty$, $x^{\alpha-1} \to$ for $\alpha > 1$.

    As $x \to \infty$, $e^{-\beta x} \to$ for $\beta > 0$.

    When $x$ is small, $x^{\alpha-1}\, e^{-\beta x}$ is dominated by _____; when $x$ is large, it is dominated by _____. Thus as $x \to \infty$, $x^{\alpha-1}\, e^{-\beta x} \to$ .

- Once we model the shape of the density function, we need to find the constant $c$ such that

$$\int_0^\infty c\, x^{\alpha-1}\, e^{-\beta x}\, \mathrm{d}x = 1$$

By some calculus, we can show that $c = \frac{\beta^\alpha}{\Gamma(\alpha)}$, where

$$\Gamma(\alpha) = \int_0^\infty x^{\alpha-1}\, \mathrm{e}^{-x}\, \mathrm{d}x, \qquad \alpha > 0$$

- 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

$$\Gamma(t) = (t-1)\,\Gamma(t-1)$$

  - Whenever $\alpha = n$, where $n$ is a positive integer, repeated integration by parts shows that

$$\Gamma(n) = (n-1)\,\Gamma(n-1) = (n-1)(n-2)\cdots(2)(1)\,\Gamma(1)$$

    Thus, when $n$ is a positive integer, we have

$$\Gamma(n) = (n-1)!, \qquad n = 1, 2, \ldots$$

  - The pdf for the gamma distribution is essentially the gamma function as defined above, except it introduces another parameter in the exponential term, $\beta$.

- Characteristics of a gamma distribution:
  - Gives the probability for the waiting time until the $\alpha^{th}$ occurrence in a Poisson process.
  - Right-skewed density function.
  - Unbounded support.

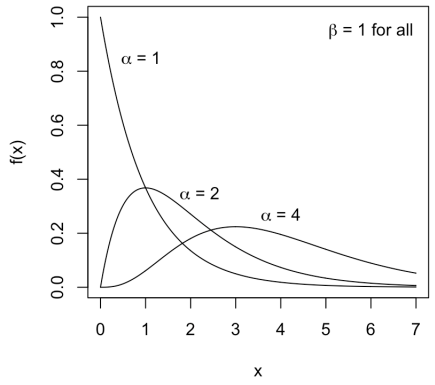
– Two important special cases of the gamma distribution:

1. When $\alpha = 1 \rightarrow X \sim$ Exponential $(\beta)$. Thus gamma can be considered a generalized version of the exponential random variable.

2. When $\alpha = r/2, \quad \beta = 2 \rightarrow X \sim \chi^2(r)$ (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 $\alpha$ and $\beta$ do not. But we can still give some meaning to them. Here is how they affect the gamma density function.
- $\alpha =$ **shape** parameter.

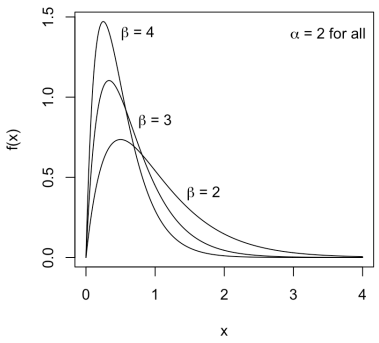



We see that for different values of $\alpha$ (and constant $\beta$), there are different _____ of the gamma densities. Hence "shape" parameter.

Additionally, for larger values of $\alpha$, the density increases longer out from zero because the polynomial part of the density function is more powerful for small $x$.

Notice that when _____ we have the familiar negative exponential curve. All exponential densities have this shape because of this fixed $\alpha$.

- $\beta =$ **rate** parameter (or $\theta = 1/\beta =$ **scale** parameter).




  We see that for different values of $\beta$ and constant $\alpha$, the general _____ of the gamma densities look similar, but the _____ 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 $\alpha$ is an integer.

    $\alpha = 1 \rightarrow$ Exponential.

    $\alpha = 2 \rightarrow$ Requires integration by parts once.

    $\alpha > 2 \rightarrow$ Requires repeated integration by parts $\alpha - 1$ times.
  - If $\alpha$ is not an integer and $0 < c < d < \infty$, we cannot calculate $P(c < X < d)$ by integration because it is impossible to give a closed-form expression for

$$\int_c^d \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} \, \mathrm{e}^{-\beta x} \, \mathrm{d}x$$

    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

- If $X \sim \text{Gamma}\,(\alpha, \beta)$ $\qquad \alpha, \beta > 0$

$$E(X) = \frac{\alpha}{\beta} \qquad\qquad V(X) = \frac{\alpha}{\beta^2}$$

- If $\alpha = 1$, $E(X)$ and $V(X)$ for exponential follow.

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 $X$.

   (b) Find $E(X)$ and $V(X)$.

   (c) Find the probability the total waiting time for the second accident is between 1 and 2 months.

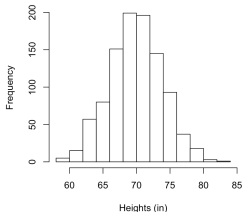
2. Biologists investigating a stretch of desert find certain fossils according to a Poisson process at a mean rate of 500 per kilometer.

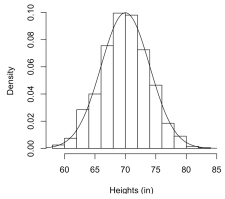
   What is the probability that the biologists will have to investigate more than 5 meters in order to find the first four fossils?

## 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 $X \sim \text{Normal}\,(\mu, \sigma^2)$

$$f(x \mid \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right], \qquad -\infty < x < \infty \quad \text{and} \quad -\infty < \mu < \infty, \quad \sigma > 0$$

- Intuitive idea behind pdf:
  1. Set the shape $\rightarrow$ Bell-shaped and symmetric (exponential decay from center in both directions)
  2. Change to smooth curve
  3. Find constant to make valid pdf.
  4. Add location and scale parameters to shift and scale the density, respectively.
  5. Adjust constant to take into account new parameters.

- Characteristics of a normal distribution.
  - Density function is bell-shaped and symmetric.
  - Unbounded support (range).

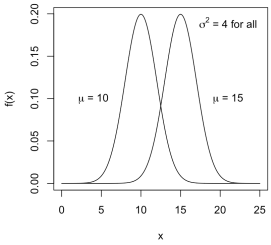
    But the density decreases exponentially as $x$ runs away from the center. Thus, the normal distribution is also used for bounded data in practice (e.g. test scores).
- The density function for the normal distribution is difficult to integrate as we will see. For example, to show that the normal distribution is a valid pdf, we need to use polar coordinates.

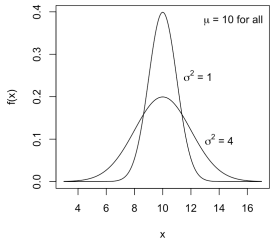
Parameters, expected value and variance

- 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 $X \sim \text{Normal}\,(\mu, \sigma^2)$

$$E(X) = \mu \qquad V(X) = \sigma^2 \qquad SD(X) = \sigma$$

- These are really easy to derive using moment generating functions (will learn later).
- Here is how they affect the normal density function.




- In practice, often the standard deviation $\sigma$ is used rather than the variance $\sigma^2$. So it is common to see $X \sim \text{Normal}\,(\mu, \sigma)$.

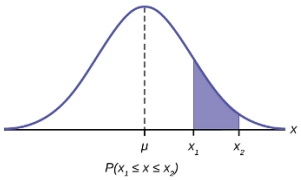
Probabilities

- 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:

- 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.

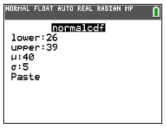
Finding Probabilities

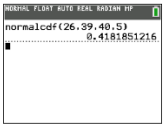
1. Choose correct Dist: 2ND → VARS
   - ALWAYS want **normalcdf()**
2. Enter in information (endpoints and parameters)
   - lower = lower boundary
   - upper = upper boundary
   - μ = Mean
   - **σ** = SD





Ex) If $X \sim Normal(\mu = 40, \sigma = 5)$, $P(26 < X < 39)$



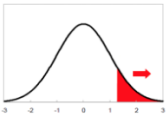



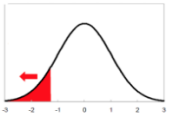

- If you have TI-83, you would type normalcdf(lower, upper, mean, sd)

*** 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 $\Longrightarrow$ 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 -∞ 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 |

- 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

    $$Y \sim \text{Normal}\,(a\mu + b, a^2\sigma^2)$$

  - 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 mgfs 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 $Y = 0.5X - 3$.

- **Step 2: Transformation to a standard normal**

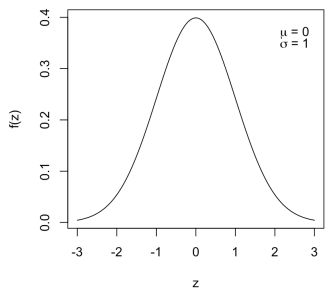
  - Using the linear transformation property of normal random variables, we can transform any normal random variable $X$ with $\mu$ and standard deviation $\sigma$ into a **standard normal** random variable with mean 0 and standard deviation 1.
  - The transformation used to do this is:

    $$Z = \frac{X - \mu}{\sigma} =$$

  - Using the previous theorem, we know $Z \sim$ Normal and can easily confirm the mean and standard deviation.

  - Thus, if we have the standard normal random variable $Z \sim \text{Normal}\,(0, 1)$

    $$f(z) = \frac{1}{\sqrt{2\pi}} \exp\left[-\frac{1}{2}z^2\right], \qquad -\infty < z < \infty$$



- 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 $X \sim$ Exponential $(\lambda)$.

  However $Z$ doesn't necessarily have the same distribution of $X$.

- **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 $Z$, $F_Z(z) = P(Z \leq z)$.
  - Can find any probability we would like using the cdf:

    $P(Z \leq z) =$

    $P(Z > z) =$

    $P(z_1 \leq Z \leq z_2) =$

  - Examples:

- **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 $\mu$ and standard deviation $\sigma$ using the identity:

$$P(x_1 \leq X \leq x_2) = P\Big(\frac{x_1-\mu}{\sigma} \leq \frac{X-\mu}{\sigma} \leq \frac{x_2-\mu}{\sigma}\Big) = P(z_1 \leq Z \leq z_2)$$

  where $z_1 = \dfrac{x_1-\mu}{\sigma}$ and $z_2 = \dfrac{x_2-\mu}{\sigma}$.

- National exam examples: If $X \sim \text{Normal}\,(\mu = 500, \sigma = 100)$. Find the following probabilities:

  $P(X \leq 800) =$

  $P(600 \leq X \leq 750) =$

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

$$F(x_p) = p$$

- If $X$ is a normal random variable with mean $\mu$ and standard deviation $\sigma$, then we can easily find $x_p$, using $z_p$ and the basic relationship of $X$ and $Z$.

$$z_p = \frac{x_p - \mu}{\sigma}$$

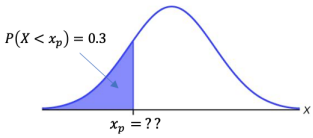
- Examples: If you scored in the $70^{th}$ percentile of test scores, what was your score?

  Find the test score corresponding to the top 15% of scores.

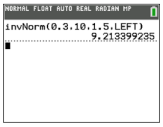
- This can also be done using software such as TI-84s.

Finding Percentiles

1. Choose invNorm(): 2ND → VARS
2. Enter in information (area and parameters)
   - area = Probability (LEFT, percentile!)
   - μ = Mean
   - **σ** = SD




Ex) If $X \sim Normal(\mu = 10, \sigma = 1.5)$, $x_{0.30} \approx 9.21$


If you have TI-83, you would type invNorm(area,mean,sd)

*** If you have a right tail (upper probability), you need to rewrite as probability to the left!

Harder examples

1. Let $X \sim \mathrm{N}(\mu, \sigma^2)$, $P(X < 500) = 0.5$ and $P(X > 650) = 0.0228$.
   Find $\mu$ and $\sigma^2$.

2. Let $X \sim \mathrm{N}(\mu = 25, \sigma^2 = 36)$. Find $c$ such that $P(|X - 25| \leq c) = 0.9544$

3. If $X \sim \mathrm{N}(\mu = 1, \sigma^2 = 16)$, find $P(X^2 - 4X < 21)$.

## 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) $\Longrightarrow$ which means we can use normal probability methods to find probabilities for the total loss.

- **Central Limit Theorem** Let $X_1, \ldots, X_n$ be independent random variables, all of which have the same probability distribution and thus the same mean $\mu$ and variance $\sigma^2$. 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:

- 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.

Some elementary statistics books define $n \geq 30$ 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 $X_i \overset{iid}{\sim} \text{Normal}(\mu, \sigma^2)$, then the CLT results in an exactly normal distribution for all $n$.

- Return to motivating example: Find the distribution of the sum of losses $S$.

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.

   (b) Find the probability there will be more 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?

Summary

- In general, normal distribution is quite valuable because it applies in so many situations were 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.

## 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 $Y \sim \text{Lognormal} \iff \ln(Y) \sim \text{Normal}\,(\mu, \sigma^2)$.

- Stated another way: A random variable $Y$ is lognormal if $Y = \mathrm{e}^X$ for some normal random variable $X$ with mean $\mu$ and variance $\sigma^2$.

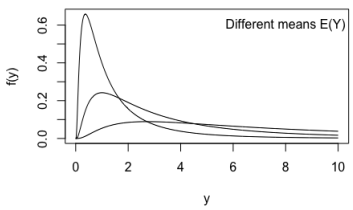
  If $X \sim \text{Normal}\,(\mu, \sigma^2)$ and $Y = \mathrm{e}^X \iff Y \sim \text{Lognormal}$

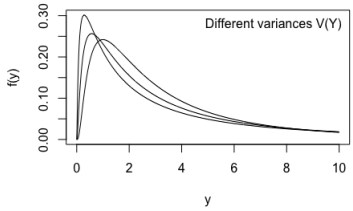
  $$f(y \mid \mu, \sigma^2) = \frac{1}{y\sqrt{2\pi\sigma^2}} \exp\left[-\frac{(\ln(y) - \mu)^2}{2\sigma^2}\right], \qquad y \geq 0 \quad \text{and} \quad -\infty < \mu < \infty, \quad \sigma > 0$$

- Characteristics of a lognormal distribution.
  - Right-skewed density function.
  - Unbounded support.

Parameters, expected value and variance

- Note that the parameters $\mu$ and $\sigma^2$ 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 $Y = \text{e}^X$

$$E(Y) = \text{e}^{\mu + \frac{\sigma^2}{2}} \qquad\qquad V(Y) = \text{e}^{2\mu+\sigma^2}(\text{e}^{\sigma^2} - 1)$$

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.
- $F_Y(y) =$

- This means we just need to do algebra in the probability statement, then use normalcdf() or $z$-tables like normal on the resulting number.

Example

- Let the claim severity $X \sim \text{Normal}\,(\mu = 7, \sigma^2 = 49)$ and $Y = \text{e}^X$.

  (a) Find $E(Y)$ and $V(Y)$.

  (b) Find the probability a claim is less than or equal to 1300.

  (c) Find the probability a claim is between 900 and 1200.

**Beta distribution**

(Brief) Motivation

- The beta distribution is defined on the interval $[0, 1]$.
- Thus it is often used as a model for _____ such as the proportion of impurities in a chemical product or the proportion of time that a machine is under repair

Definition

- If $X \sim \text{Beta}\,(\alpha, \beta)$

$$f(x \mid \alpha, \beta) = \frac{1}{B(\alpha, \beta)}\, x^{\alpha-1}(1-x)^{\beta-1}, \qquad 0 \le x \le 1 \quad \text{and} \quad \alpha, \beta > 0,$$

  where

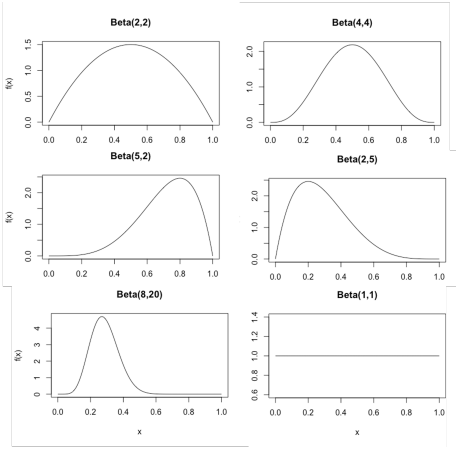
$$B(\alpha, \beta) = \int_0^1 x^{\alpha-1}(1-x)^{\beta-1}\,\mathrm{d}x = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$$

- 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.

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, $\alpha$ and $\beta$ 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 $X \sim \text{Beta}\,(\alpha, \beta)$

$$E(X) = \frac{\alpha}{\alpha + \beta} \qquad\qquad V(X) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$$

Probabilities

- When $\alpha$ and $\beta$ 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 $\beta = 3$.

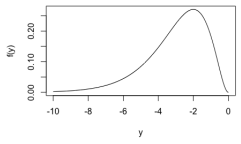
  (a) Find the pdf $f(x)$ and the cdf $F(x)$.

  (b) Find the probability the percent of clients contacting the firm in a month is less than 40%.

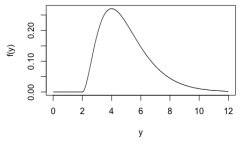
  (c) Find the expected value and variance for the percent of clients contacting the firm.

**Transformations**

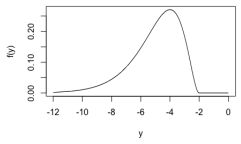
- We have covered the standard shapes that can be modeled by these distributions. To be more flexible, we can use functions of these random variables to model variations of the standard shapes.
- Examples:
  1. If a population is left-skewed and the range is $-\infty < Y \leq 0$. How can we model the population?



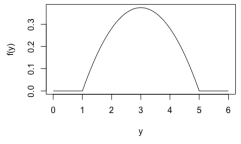

  2. If a population is right-skewed, but the range is $a \leq Y < \infty$, $a \neq 0$. How can we model the population?




  3. If a population is left-skewed and the range is $-\infty < Y \leq a$, $a \neq 0$. How can we model the population?




  4. If a population is bounded between $a \leq Y \leq b$. How can we model the population?



# Lecture 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 $\boldsymbol{n^{th}}$ **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

- 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 |

- There are other "types" of moments as well.
  - Given random variable $X$ and constant $b$,

    $E(X-b)$ is the **first moment of $\boldsymbol{X}$ about $\boldsymbol{b}$**.

    $E[(X-b)^n]$ is the $\boldsymbol{n^{th}}$ **moment of $\boldsymbol{X}$ about $\boldsymbol{b}$**.

  - 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.

## 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
$$M_X(t) = E(\mathrm{e}^{tX})$$
- Notes:
  * $\mathrm{e}^{tX}$ is a function of $t$ and $X$ and it is a random variable.
  * $E(\mathrm{e}^{tX})$ is the expectation with respect to $X$, which takes away the randomness of $X \Longrightarrow$ Result is only a function of $t$ (variable, not constant).
- The mgf has a number of useful properties:

  (1) 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 $t = 0$).
  $$M'_X(0) = E(X), \quad M''_X(0) = E(X^2), \quad \ldots \quad , \quad M_X^{(n)}(0) = E(X^n)$$

  (2) Distribution of a function of a random variable: The mgf of $aX + b$ can be found easily if the mgf of $X$ is known.
  $$M_{aX+b}(t) = \mathrm{e}^{tb} M_X(at)$$

  (3) Uniqueness: If a random variable $X$ has the mgf of a known distribution, then $X$ has that distribution.

- Formalizing this new notation from (1):
$$E(X^n) = M_X^{(n)}(0) = \frac{\mathrm{d}}{\mathrm{d}t^n} M_X(t)\bigg|_{t=0}$$
  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

(1) Finding moments:

Now we can show exactly how / why mgfs generate moments (hence the name).

(a) We will start by looking at the infinite series representation of $\mathrm{e}^x = \sum_{k=0}^{\infty} \frac{x^k}{k!}$ and then substituting the random variable $tX$ for $x$ in this series, we obtain:

(b) Now we take the expected value of each side (with respect to $X$).

(c) Then we take the derivative (with respect to $t$).

(d) Then we evaluate at $t = 0$.

(e) Now take the second derivative and evaluate at $t = 0$ again.

(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:

(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 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

(a) Find the moment generating function of $X$, $M_X(t)$.

| $x$ | 0 | 1 | 2 | 3 |
|---|---|---|---|---|
| $\mathrm{e}^{tx}$ | | | | |
| $f(x)$ | 0.512 | 0.384 | 0.096 | 0.008 |

(b) Find $E(X)$ using the mgf found in part (a).

Step 1: Take the derivative of $M_X(t)$ with respect to $t$.

Step 2: Now evaluate the derivative at $t = 0$ (i.e. plug in $t = 0$ and simplify).

(c) Find $V(X)$ using the mgf.

Note that the higher derivatives can be used in the same way $\Longrightarrow$ Take the second derivative of $M_X(t)$ with respect to $t$ and then evaluate at $t = 0$.

Final points

- Expanding on property (1) of mgfs (for discrete random variables, replace summation with integration for continuous).

$$M_X(t) = E(\mathrm{e}^{tX}) = \sum \mathrm{e}^{tx}\, f(x)$$

Middle step: $M'_X(t) =$ $M''_X(t) =$

Property: $M'_X(0) =$ $M''_X(0) =$

- The general form is the following:

$$M_X^{(n)}(t) = \sum x^n \mathrm{e}^{tx}\, f(x)$$

$$M_X^{(n)}(0) = \sum x^n\, f(x) = E(X^n)$$

- Variance: Using these new ideas and notations, we have a new way to write the variance using mgfs:

$$V(X) = M''_X(0) - \left[M'_X(0)\right]^2 = M''_X(t)\big|_{t=0} - \left[M'_X(t)\big|_{t=0}\right]^2$$

- Mgfs for random variables don't always exist. There is technically more to the definition $M_X(t) = E(\mathrm{e}^{tX})$.
  - 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 $-h < t < h$, $E(\mathrm{e}^{tX})$ 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 $X \sim$ Discrete uniform $(N_0, N_1)$

- If $X \sim$ Discrete uniform $(N_0, N_1)$

$$M_X(t) = \frac{1}{N_1 - N_0 + 1} \sum_{x=N_0}^{N_1} \mathrm{e}^{tx}$$

Bernoulli random variable mgf

- The mgf for a bernoulli random variable is straightforward to find.
- Let $X \sim$ Bernoulli $(p)$

| $x$ | 0 | 1 |
|---|---|---|
| $\mathrm{e}^{tx}$ | $\mathrm{e}^{t0} = 1$ | $\mathrm{e}^{t}$ |
| $f(x)$ | $1 - p = q$ | $p$ |

- If $X \sim$ Bernoulli $(p)$

$$M_X(t) = (1 - p) + p\mathrm{e}^{t} = q + p\mathrm{e}^{t}$$

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.
  1. We just saw the mgf when $X \sim \text{Binomial}\,(n = 1, p)$

     $M_X(t) = q + p\text{e}^t$

  2. Now let $X \sim \text{Binomial}\,(n = 2, p)$.

| $x$ | 0 | 1 | 2 |
|---|---|---|---|
| $\text{e}^{tx}$ | 1 | $\text{e}^t$ | $\text{e}^{2t}$ |
| $f(x)$ | $q^2$ | $2pq$ | $p^2$ |

- If $X \sim \text{Binomial}\,(n, p)$

$$M_X(t) = (q + p\text{e}^t)^n$$

- 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 it's mgf is:

  $M_X(t) = (0.2 + 0.8\text{e}^t)^7 \Longrightarrow$

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 $X \sim \text{Geometric}\,(p)$

- If $X \sim$ Geometric $(p)$

$$M_X(t) = \frac{p\mathrm{e}^t}{1 - q\mathrm{e}^t} \qquad t < -\ln(q)$$

- Just like with the binomial, need to be careful when deriving the mean and variance using the mgf.
- Example: Let $X \sim$ Geometric $(p)$. Suppose $Y = X - 1$. Find the mgf of $Y$, $M_Y(t)$.

  This is the mgf of the alternate form of geometric ($Y$ = number of failures before first success).

Negative binomial random variable mgf

- We will not derive this. But we can make use of the pattern / relationship of Bernoulli and binomial.

- If $X \sim$ Negative binomial $(r, p)$

$$M_X(t) = \left[\frac{p\mathrm{e}^t}{1 - q\mathrm{e}^t}\right]^r \qquad t < -\ln(q)$$

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 $\mathrm{e}^x$.
- If $X \sim$ Poisson $(\lambda)$

$$M_X(t) = \mathrm{e}^{\lambda(\mathrm{e}^t - 1)}$$

- Derivation: Use the mgf to derive the mean and variance of a Poisson($\lambda$) random variable.

- Example: Let $X \sim \text{Poisson}\,(\lambda = 2)$. Suppose $Y = 3X + 5$. Find the mgf of $Y$, $M_Y(t)$.

Hypergeometric random variable mgf

- Apparently it exists, but we will not discuss it.

**Moment generating functions for continuous random variables**

Inroduction

- 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 $f(x) = 0.75(1 - x^2), \quad -1 \leq x \leq 1$).

  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.

- If $X \sim \text{Uniform}\,(a, b)$,

$$M_X(t) =$$

Gamma mgf

- The gamma mgf can be easily derived. It requires the identity which was shown earlier when introducing the gamma function $\Gamma(\alpha)$:

$$\int_0^\infty x^{\alpha-1}\,\mathrm{e}^{-\beta x}\,\mathrm{d}x = \frac{\Gamma(\alpha)}{\beta^\alpha}, \qquad x > 0 \quad \text{and} \quad \alpha, \beta > 0.$$

- 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

$$E(X) = \frac{\alpha}{\beta}, \qquad V(X) = E(X^2) - [E(X)]^2 = \frac{\alpha}{\beta^2}$$

Exponential mgf

- Since the exponential distribution is a special case of the gamma distribution,

  we also know the mgf of the exponential distribution.

- If $X \sim \text{Exponential}\,(\beta)$,

$$M_X(t) =$$

Normal mgf

- The normal mgf can actually be derived without fancy calculus skills, just fancy algebra skills.
- If $X \sim \text{Normal}\,(\mu, \sigma^2)$,

$$M_X(t) = \mathrm{e}^{\mu t + \frac{\sigma^2 t^2}{2}}$$

- Lets derive the mean of the normal distribution using the mgf:

- Derivation of the variance is straightforward, just requires careful application of the product rule.

  Question: What must $E(X^2)$ equal?

- 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

  $$Y \sim \text{Normal}\,(a\mu + b, a^2\sigma^2)$$

- Now lets apply the same strategy to the specific linear transformation of standardizing and show the mgf of the standard normal distribution $Z$.

- 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:

  $$Z = \frac{X - \mu}{\sigma} = \frac{X - 1/\lambda}{1/\lambda} = \lambda X - 1 \qquad \Longrightarrow \qquad E(Z) = 0, \qquad V(Z) = 1$$

  But did not have a conclusion about the distribution of $Z$.

# After Test 3

## Contents



## Lecture 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 an 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

$$f(x) = 0.002\mathrm{e}^{-0.002x}, \quad x \geq 0 \qquad \Longrightarrow \qquad X \sim \mathrm{Exp}\,(\lambda = 0.0002)$$

    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  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.

- 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.

- 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.

- In general, if loss $x$ with deductible $d$ and cap $c$, we have

## 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 $E(Y)$.

- Note we did not need to 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 $f(y)$.
- 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 | | |
| 50 | 0.43 | | |
| 100 | 0.20 | | |
| 150 | 0.09 | | |

- 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 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.

- Find the cdf $F_Y(y)$.

  *STRATEGY*: Use the same reasoning as above, just for a general $y$:
  $P(Y \leq 100) = F_Y(100) \longrightarrow P(Y \leq y) = F_Y(y)$ for any value $y \geq 0$.

- 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*: How to check range $\rightarrow$ Apply $g(x)$ to all pieces, ALWAYS need to check both sides.

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 _____ to a _____. That is,

  ** We do not know stuff (pdf, cdf, etc.); so we have to use the inverse function to go backwards. $\mathcal{Y}$ is completely determined by $\mathcal{X}$.

- When do inverse functions exist?

  If the function $g(x)$ is strictly **monotone** $\implies$ one-to-one $\iff$ inverse exists.

  $u > v \Rightarrow g(u) > g(v)$

  $u > v \Rightarrow g(u) < g(v)$

- 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:

** *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:

  - Example: Let $X \sim \text{Exponential}\,(\lambda = 3)$. Find the cdf of $Y = \sqrt{X}$.

    There are two ways that we can solve this.

    Long way  Short way

- **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:

  - 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 Short way

- **If $g(x)$ does NOT have an inverse**
  - Example: Let $X \sim \text{Uniform}\,(a = -2, b = 2)$. Find the cdf of $Y = X^2$.

  - It can be even more complicate if there isn't a "balanced" range of $Y$.
    Example: Let $X \sim \text{Uniform}\,(a = -2, b = 3)$. Find the cdf of $Y = X^2$.

  - 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:

  $P(Y \leq y) =$ $P(Y \geq y) =$ $P(a \leq Y \leq b) =$

  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:

  $$f_Y(y) = \frac{\mathrm{d}}{\mathrm{d}y} F_Y(y)$$

- 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
  2. Pdf method

     (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:

- **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:

  - Since $h(y)$ is decreasing, its derivative is negative. Thus the final expression above is actually positive.

- Theorem: Let $X$ have cdf $F_X(x)$ with range $\mathcal{X}$, $Y = g(X)$ with and range $\mathcal{Y}$ and inverse $h(y)$.

  - If $g(x)$ is strictly increasing on $\mathcal{X} \longrightarrow F_Y(y) = F_X(h(y))$ for $y \in \mathcal{Y}$.
  - If $g(x)$ is strictly decreasing on $\mathcal{X} \longrightarrow F_Y(y) = 1 - F_X(h(y))$ for $y \in \mathcal{Y}$.

  - If $g(x)$ is strictly increasing or strictly decreasing on $\mathcal{X}$, then

$$f_Y(y) = f_X(h(y)) \cdot |h'(y)| \qquad \text{for } y \in \mathcal{Y}.$$

- Return to previous examples: Let $X \sim \text{Exponential}\,(\lambda = 3)$.

  (a) Find the pdf of $Y = \sqrt{X}$.

  Cdf method  Pdf method

  (b) Find the pdf of $Y = 1 - X$.

  Cdf method  Pdf method

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 $f_Y(y)$.

2. Let $X \sim \text{Poisson}\,(\lambda = 4)$. If $Y = X^2$, find the pmf of $Y$.