Stat 262 Notes

Table of contents

Course Goals for STAT 262

STAT 262 is an introduction to the basic probability theory that allows us to the use the statistical methods seen in courses like STAT 102 and STAT 212. By the end of the course, you will:

- Understand and be able to apply the basic rules of probability, including sample spaces, conditional probability, independence, and Bayes' theorem
- Understand and be able to articulate the definition of random variable for a specific scenario
- Recognize and be able to use the different function associated with random variables: cumulative distribution function, probability density/mass function, moment generating function
- Use density/mass functions to find moments
- Use common discrete and continuous probability distributions to model real-world scenarios
- Understand and be able to use multivariate probability distributions, including marginal and conditional distributions, independence, and covariance between random variables
- Derive and use probability distributions of functions of random variables using the methods of transformations, Jacobians, and moment generating functions
- Understand and be able to recognize random samples, including random samples drawn from the normal distribution
- Understand and be able to recognize the distribution of functions of normal random variables
- Understand elementary convergence concepts, including the law of large numbers and the Central Limit Theorem

1 Probability

When we are uncertain about an outcome's occurrence (e.g., whether a coin will come up heads or tails, the number of dots observed on the roll of a die, whether or not the bus will be late), we typically quantify this uncertainty with a probability. Probability is the foundation upon which all of statistics is built, and it a provides a framework for modeling populations, experiments, and almost anything that could be considered a random phenomenon.

1.1 Basic Probability Definitions and Calculations

A **sample space**, denoted by S, is comprised of all possible outcomes of a random phenomenon.

An **event** is a collection of possible outcomes. Each event A is a subset of S.

We want to formalize the idea of the "chance" that event A occurs. We will do this by defining the **probability** of each A, which we denote $P(A)$.

Probabilities are calculated by defining functions on sets, and should be defined for all possible events. One thing that must be true:

$$0 \leq P(A) \leq 1$$

More formally, a probability function is defined as follows.

Given a sample space S, a **probability function** is a function $\mathrm{P}(\cdot)$ that satisfies

1.

2.

3.

Requirements (1) - (3) are called the

Any function $\mathrm{P}(\cdot)$ that satisfies the Axioms of Probability is called a probability function.

Here's a (hopefully) obvious theorem:

Theorem: Let (S, P) be a sample space and associated probability function, respectively. For any event $A \in S$

1.

2.

3.

And another, maybe less obvious.

Theorem: Let (S, P) be a sample space and associated probability function, respectively. For any events A, $B \in S$

1.

2.

3.

We'll use these theorems often when calculating probabilities. However, we first need to figure out how to assign probabilities to specific events. In some cases, we can do that by figuring out how many possible events there are in a sample space.

In a finite sample space, when all outcomes are equally likely, the number of possible outcomes can be used to make probability assignments.

- If the sample space S consists of n possible outcomes, and these n outcomes are equally likely, the probability of any of these $\{s_1, s_2, \ldots, s_n\}$ outcomes is

- If the set A is some collection of these outcomes, then

Example: Toss a fair, six-sided die

Other Examples:

Often, it is difficult to list all of the outcomes in a sample space, even when it is finite. In such circumstances, other methods must be employed to count the number of outcomes in a sample space.

Fundamental Theorem of Counting

Example:

Why is the number of possible outcomes important? Because when each outcome is equally likely, we use the number of possible outcomes to find the probabilities of various events. The problem is that Fundamental Theorem of Counting doesn't always (obviously) work. To determine the number of ways a task can be completed, we often need to consider whether sampling occurs with or without replacement and whether sampling is ordered or unordered. We're not going to go into a ton of detail here. Some of these combinatorics scenarios really aren't helpful to us, and most situations don't involve outcomes that are all equally likely.

Some quick and dirty examples:

Ordered, with replacement: Number of different 6-character license plates if the first 3 characters must be letters and the final 3 characters must be numbers.

In general:

Ordered, without replacement: Number of different lead-off (i.e., first three batters) batting orders for a baseball team consisting of 9 players.

In general:

Unordered, without replacement: A student is to answer 7 out of 10 questions on an exam. How many choice are there?

In general:

Unordered, with replacement: This one is not terribly intuitive, and not terribly useful anyway. I couldn't think of a realistic scenario. In case you care:

Example: Suppose Alabama (which does not have a state lottery) is considering four possible lottery drawing configurations:

1. Six different cages, each with 40 balls numbered 1-40. Winners' selections must be in the same sequence as the numbers drawn from the cages.

2. One cage with 40 balls numbered 1-40. Winners' selections must be in the same sequence as the six numbers drawn from the cage.

3. One cage with 40 balls numbered 1-40 Winners' selections must match the six numbers drawn from the cage.

4. Six different cages, each with 40 balls numbered 1-40. Winners' selections must match the numbers drawn from the cages.

Bottom line: the total number of ways to pick r items from a total of n distinguishable items depends on whether or not order matters, as well as whether sampling is done with or without replacement.

Table 1.1: Number of Ways to Pick r Items from n Distinguishable Objects

	Without Replacement	With Replacement
Order Matters		
Order Doesn't Matter		

Unordered, without replacement will turn out to be the most case we use most often (and we'll use that one A LOT).

1.2 Conditional Probability

Let's look at the following table:

	Survived	Did Not Survive
First Class	201	123
Second Class	118	166
Third Class	181	528

The counts in the table are the number of Titanic passengers that fell into the each of the categories. From this table, we can calculate some probabilities. Let's consider the outcomes First Class and Survived.

Often, we have partial information about a certain phenomenon and wish to know how this affects the probabilities of outcomes of interest to us, if at all. For example, we might want to know the probability a randomly selected student is a sophomore, given that we know they are enrolled in STAT 262.

Example: Toss a fair die. Let $A = \{1\}$ and $B = \{1, 3, 5\}$. What is the probability of throwing a 1 given that an odd number is thrown?

Definition: Conditional Probability

Example, again:

This definition of conditional probability leads to:

Law of Total Probability

If $A_1, A_2, \ldots$ is a collection of mutually exclusive ($A_i \cap A_j = \emptyset$ for all $i \neq j$) and exhaustive ($P(\cup_{i=1}^{\infty} A_i) = 1$) events, and if $P(A_i) > 0$ for all i, then for any event B,

$P(B) =$

Venn Diagram:

General Form of Bayes' Theorem

Suppose $A_1, A_2, \ldots,$ partition S.

$P(A_i|B) =$

Example: Of travelers arriving at a small airport, 60% fly on major airlines, 30% fly on privately owned planes, and the remainder fly on commercially owned planes not belonging to a major airline. Of those traveling on major airlines, 50% are traveling for business reasons, whereas 60% of those arriving on private planes and 90% of those arriving on other commercially owned planes are traveling for business reasons.

Let's first construct a table.

Suppose we randomly select one person arriving at this airport. What is the probability the person . . .

- is traveling on business?

- is traveling for business on a privately owned plane?

- arrived on a privately owned plane, given the person is traveling on business?

- is traveling on business, given the person is flying on a commercially owned plane not belonging to a major airline?

Example: A diagnostic test for a disease is such that it (correctly) detects the disease in 95% of the individuals who actually have the disease. Also, if a person does not have the disease, the test will report that he or she does not have it with probability 0.9. Only 1% of the population has the disease in question. If a person is chosen at random from the population and the diagnostic test indicates that she has the disease, what is the conditional probability that she does, in fact, have the disease?

Example: A student answers a multiple choice exam question that offers four possible answers. Suppose the probability the student knows the answer to the question is 0.8 and the probability the student will guess is 0.2. If the student guesses, the probability of selecting the correct answer is 0.25. If the student correctly answers a question, what is the probability the student really knew the correct answer?

Consider the following table:

Table 1.3: All flights arriving at an airport on a single day

	Late	On Time
Domestic	12	109
International	6	53

Find the probability that a randomly selected flight on this day was on time.

Find the probability that a randomly selected flight was on time, given that it was a domestic flight.

What do you notice about these two values?

Does this make sense in the context of this scenario? What do you think it means?

1.3 Independence

Sometimes the occurrence of one event, B, will have no effect on the probability of another event, A. If A and B are unrelated, then intuitively it should be the case that

Also, it follows that

> **Definition: Statistical Independence**
>
> Two events, A and B are **statistically independent** if and only if

Extending this to multiple events . . .

> **Definition: Mutual Independence**
>
> A collection of events $A_1, A_2, \dots A_n$ are **mututally independent** if and only if for any subcollection $A_{i_1}, A_{i_2}, \dots, A_{i_k}$

Example: A mouse caught in a maze has to maneuver through three successive escape hatches in order to escape. If the hatches operate independently and the probabilities for the mouse to successfully get through them are 0.6, 0.4, and 0.2, respectively, what are the probabilities that the mouse:

- will be able to escape?

- will not be able to escape?

2 Random Variables and their Associated Functions

Often, we are interested in a numerical measurement of the outcome of a random experiment. For example, we might want to know the number of insects treated with a dose of a new insecticide that are killed. In this case, the outcome is the survival status of each 'dosed' insect, and the numerical measurement in which we are interested is the number that died. However, the observed number varies depending on the actual result of the experiment. This type of variable is called a **random variable**.

2.1 Random Variables

Definition: Random Variable

A **random variable** is a function that associates a real number with each element in the sample space. That is, a random variable is a function from a sample space, S, into the real numbers, $\mathbb{R}$.

Example: Suppose we roll two dice. There are 36 possible outcomes:

$$S = \left\{ \begin{array}{cccccc} 1,1 & 2,1 & 3,1 & 4,1 & 5,1 & 6,1 \\ 1,2 & 2,2 & 3,2 & 4,2 & 5,2 & 6,2 \\ 1,3 & 2,3 & 3,3 & 4,3 & 5,3 & 6,3 \\ 1,4 & 2,4 & 3,4 & 4,4 & 5,4 & 6,4 \\ 1,5 & 2,5 & 3,5 & 4,5 & 5,5 & 6,5 \\ 1,6 & 2,6 & 3,6 & 4,6 & 5,6 & 6,6 \end{array} \right\} \Rightarrow$$

In the game of SKUNK, we don't necessarily care about the specific outcome of a roll. For example, we don't care to know that, specifically, a (2,3) was rolled. Instead, what are we interested in?

The number of ones rolled varies among the 36 points in the sample space, and an observed value of the variable, X, denoted by x, depends on the outcome of a random experiment (a roll of two dice). This variable (the number of ones rolled on two dice) is referred to as a random variable.

The **support** of X, or the set of possible values of X, is:

The corresponding probabilities for each of the possible values of X are:

Identifying the type(s) of random variable(s) we are interested in helps us decide which methods and procedures are most appropriate for certain problems and answering specific questions. In general, there are two types of random variables:

- **Discrete Random Variable**: If the sample space of a random variable is finite or countably infinite, then the random variable is a discrete random variable.

- **Continuous Random Variable**: If the sample space of a random variable is uncountably infinite, then the random variable is a continuous random variable. A continuous random variable can take on any value in an interval

2.2 Distribution Functions

Each random variable, X, is associated with a function called the cumulative distribution function (CDF) of X.

Definition: Cumulative Distribution Function

The **cumulative distribution function** or CDF of a random variable X, denoted by $F_X(x)$, is defined as

Example: Before, we found the probabilities for the possible values of X = number of ones rolled on two dice. What is the CDF of X?

Theorem:

The function $F_X(x)$ is a CDF if and only if the following three conditions hold:

-
-
-

A discrete CDF increases only in jumps.

Example: A teenager has passed the written driving exam, but still needs to pass the road exam. Suppose the probability of passing the road exam is p. Let X be the number of attempts required to pass the road test, and assume attempts are independent.

Continuous CDFs have no discontinuities.

Example: Suppose $X \sim$ Uniform(0,1). For this distribution, the CDF is:

$$F_X(x) = \begin{cases} 0 & \text{if } x < 0 \\ x & \text{if } 0 \le x < 1 \\ 1 & \text{if } x \ge 1 \end{cases}$$

For a continuous CDF, $P(X = a) = 0$. Why?

Example: Suppose the time a person must wait in line (in minutes) at airport security has cdf

$$F_X(x) = \begin{cases} 0 & \text{if } x < 0 \\ 1 - \exp(-x/15) & \text{if } 0 \le x < \infty \end{cases}$$

- What is the probability a randomly selected passenger gets through the security line in less than 5 minutes?

- What is the probability a randomly selected passenger waits in the security line longer than 5 minutes?

- What is the probability a randomly selected passenger waits in line between 5 and 10 minutes?

Definition: Identically Distributed

Two random variables X and Y are **identically distributed** if they have the same distribution function. That is,

2.3 Density and Mass Functions

Associated with any random variable X and its CDF $F_X(x)$ is another function, called either the probability density function (pdf) or probability mass function (pmf). The terms pmf and pdf refer, respectively, to the discrete and continuous cases (though pdf often gets used for both types of RVs). Both pmfs and pdfs are concerned with "point probabilities" of random variables.

Definition: Probability Mass Function

The **probability mass function** or pmf of a discrete random variable X, denoted by $f_X(x)$, is given by

Let p_j be the size of the jump in the CDF at $x = x_j$, $j = 1, 2, ...$, then

Example: In game of SKUNK, we defined X = number of ones rolled on two dice. X has pmf:

Theorem:

A function, $f_X(x)$, is a pmf if and only if

-
-

Example: A car dealer has 30 cars available for immediate sale, of which 10 are classified as compact cars. 3 customers arrive and buy cars. Define the random variable N to be the number of compact cars sold. What is the pmf for N?

- What is the probability exactly 2 compact cars are purchased?

- What is the probabilty less than 2 compact cars are purchased?

This shows the special relationship between CDFs and pmfs for discrete random variables:

Example (cont'd): What is the CDF for N=number of compact cars sold?

Probability density functions are associated with (absolutely) continuous random variables and (absolutely) continuous CDFs. A probability mass function (pmf) gives "point probabilities," and pmfs may be summed to get the CDF of a discrete random variable. For continuous random variables, the probability of any point is zero (i.e., $\mathrm{P}(X = a) = 0$). As a result, we define the probability density function (pdf) for a continuous random variable differently.

Relationship between CDFs and pdfs

The relationship between the CDF and the **probability density function** (pdf) of a continuous random variable X is given by

where $f_X(x)$ is the pdf. By the Fundamental Theorem of Calculus,

Example: Each day, mail arrives in the Statistics department office between 9 and 10 am. The mail is no more likely to arrive at one time than another. Let the random variable X represent the time (in hours) after 9 am that the mail arrives in the department office. Earlier, we determined $X \sim$ Uniform(0,1). For this distribution, the CDF is:

$$F_X(x) = \begin{cases} 0 & \text{if } x < 0 \\ x & \text{if } 0 \leq x < 1 \\ 1 & \text{if } x \geq 1 \end{cases}$$

For this distribution, the pdf is:

What is the probability the mail arrives before 9:15 am?

The pdf is a curve that describes the probability of observing X in some range of values, such as between x_1 and x_2. The probability is defined as:

Example (con't): Suppose that a random variable $Y \sim \text{Uniform}(L, U)$. In the previous example, $L = 0$ and $U = 1$. The pdf of Y is:

$$f_Y(y) = \begin{cases} 0 & \text{if } y < L \\ c & \text{if } L \leq y \leq U \\ 0 & \text{if } y > U \end{cases}$$

What should c be?

> Theorem:
>
> A function, $f_X(x)$, is a pdf if and only if
>
> -
>
> -

Example (con't):

- What should c be?

- Let Y represent the time (in minutes) after 9 am that the mail arrives at the office. Assuming mail arrives between 9 and 10 am, and the mail is no more likely to arrive at one time than another, what is the pdf of Y?

Example: Suppose the time between injury accidents in a nuclear power plant (in days) is a random variable has the pdf:

$$f_X(x) = \begin{cases} \frac{1}{3}e^{-x/3} & x > 0 \\ 0 & x \leq 0 \end{cases}$$

- Verify that this function is a pdf.

- What is the probability the time between accidents is between 2 and 5 days?

- What is the probability the time between 2 accidents is more than 365 days?

- Find the CDF of X.

- How could the CDF be used to find the probability that the time between 2 accidents is more than 365 days? Is this value easier to calculate using the CDF or pdf? Why do you think so?

3 Moments

3.1 Expected Values

Remember that a random variable is a function that associates a real number with each element in the sample space. The probability a RV takes on a certain value, or falls within a range of values, is described by the probability distribution (pdf or pmf) of the RV. Often, we are interested in the expected/"average" value of the RV. The expected value can tell us things like the "average" amount we might expect to win (or lose!) in a game, the typical weight range for 3 month old babies, the price we might expect to pay for a typical house in a new city.

Definition: Expected Value

Let $X \sim f_X(x)$, $F_X(x)$ and let g be a function defined on the support of X. The **expected value** of $g(X)$ is $\text{E}[g(X)]$, where

NOTE: The expected value is a characteristic of the underlying distribution of the random variable–its CDF/pdf. It is NOT calculated from data. The sample mean calculated from data could be used to estimate the true expected value (and this is what we did in STAT 102; μ vs $\bar{X}$).

Example: Consider a game in which a fair die is thrown. The player pays \$5 to play, and wins \$2 for each dot that occurs on the roll. Is it worth the \$5 to enter?

Example: An insurance policyholder's loss, X, follows a distribution with pdf

$$f_X(x) = \begin{cases} \frac{2}{x^3} & x > 1 \\ 0 & \text{else} \end{cases} .$$

- What is the expected loss for a randomly selected policyholder?

- The insurance company reimburses a loss up to a benefit limit of 20. What is the expected value of the benefit paid under the insurance policy?

Example: Let $f_X(x) = \begin{cases} e^{-x} & x \geq 0 \\ 0 & \text{else} \end{cases}$.

- Suppose $g(X) = X$. Find $\mathrm{E}[g(X)]$.

- Suppose $g(X) = e^X$. Find $\mathrm{E}[g(X)]$.

If $\mathrm{E}[|g(X)|] = \infty$, then $\mathrm{E}[g(X)]$ does not exist.

Theorem:

Let X be a random variable and let a, b and c be constants. For any functions $g_1(x)$ and $g_2(x)$ whose expectations exist,

Example: The failure of a circuit board interrupts work that utilizes a computing system until a new board is delivered. The delivery time (in days), X, follows a distribution with pdf $f_X(x) = \begin{cases} \frac{1}{4} & 1 \le x \le 5 \\ 0 & \text{else} \end{cases}$. The cost of a board failure and interruption is given by $C = 500 + 100X^2 - 25X$. Find the expected cost associated with a single failed circuit board.

3.2 Moments

A particular set of expected values are called moments.

Definition: Moment

The n^{th} **moment** of X is

Example:

Definition: Central Moment

The n^{th} **central moment** of X is

Example:

Facts about the Variance:

1.

2.

Example (cont'd): Earlier, we considered a random variable, X, representing the delivery time (in days) of a new circuit board, which had pdf $f_X(x) = \begin{cases} \frac{1}{4} & 1 \leq x \leq 5 \\ 0 & \text{else} \end{cases}$.

- Find the variance of the delivery time (in days) of a new circuit board.

- Suppose that the cost of a single failed circuit board is $K = 500 + 6X^2$ (yeah, I know it's not the same cost we used before). Find the variance of the cost associated with a single failed circuit board.

One identity that will come in VERY handy this semester:

e^x	$1 + x + \frac{x^2}{2!} + \frac{x^3}{3!} + \cdots$
Exponential Function	Exponential Function (Taylor's Version)

This is especially helpful with a common discrete distribution called the Poisson.

Example: Suppose $X \sim \text{Poisson}(\lambda)$. This means that $f_X(x) = P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!}$ for $x = 0, 1, 2, ...$

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

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

3.3 Moment Generating Functions

A special expected value that is quite useful is the moment generating function. Moment generating functions can make our lives a LOT easier. We'll first talk about how to find moment generating functions, and then discuss how to use moment generating functions in a few different ways.

3.3.1 Finding Moment Generating Functions

Definition: Moment Generating Function

Let X be a random variable with pdf/pmf $f_X(x)$. The **moment generating function** (mgf) of X is $M_X(t)$, where

Note: At $t = 0$, $M_X(t) = 1$.

Example: Suppose $X \sim \text{Poisson}(\lambda)$. What is the mgf of X?

3.3.2 Using Moment Generating Functions

The moment generating function (mgf) of a random variable is a tool we use for a variety of purposes. Specifically, we use the mgf to:

- Calculate moments (the mean, variance, other quantities calculated from moments like skewness and kurtosis)
- Identify distributions (of a random variable, or some function of a random variable)

Theorem:

If X has mgf $M_X(t)$, then $\mathrm{E}(X^n) = M_X^{(n)}(0)$, where

$M_X^{(n)}(0) =$

That is, the n^{th} moment of X is equal to the n^{th} partial derivative of $M_X(t)$ with respect to t, evaluated at $t = 0$.

Example (cont'd): Suppose $X \sim \text{Poisson}(\lambda)$. Using the mgf we found earlier, let's find the mean and variance of X.

Example: Let $X \sim \text{Gamma}(\alpha, \beta)$. This means that $f_X(x) = \begin{cases} \left(\frac{1}{\Gamma(\alpha)\beta^\alpha}\right) x^{\alpha-1} e^{-x/\beta} & x \geq 0 \\ 0 & \text{else} \end{cases}$, where $\alpha, \beta > 0$. Also note that $\Gamma(\alpha) = \int_0^\infty y^{\alpha-1} e^{-y} dy$, $\Gamma(\alpha+1) = \alpha\Gamma(\alpha)$, and, if n is an integer, $\Gamma(n+1) = n\Gamma(n) = n!$. What is the mgf of X?

Example (cont'd): Using the mgf, find the mean and variance of X.

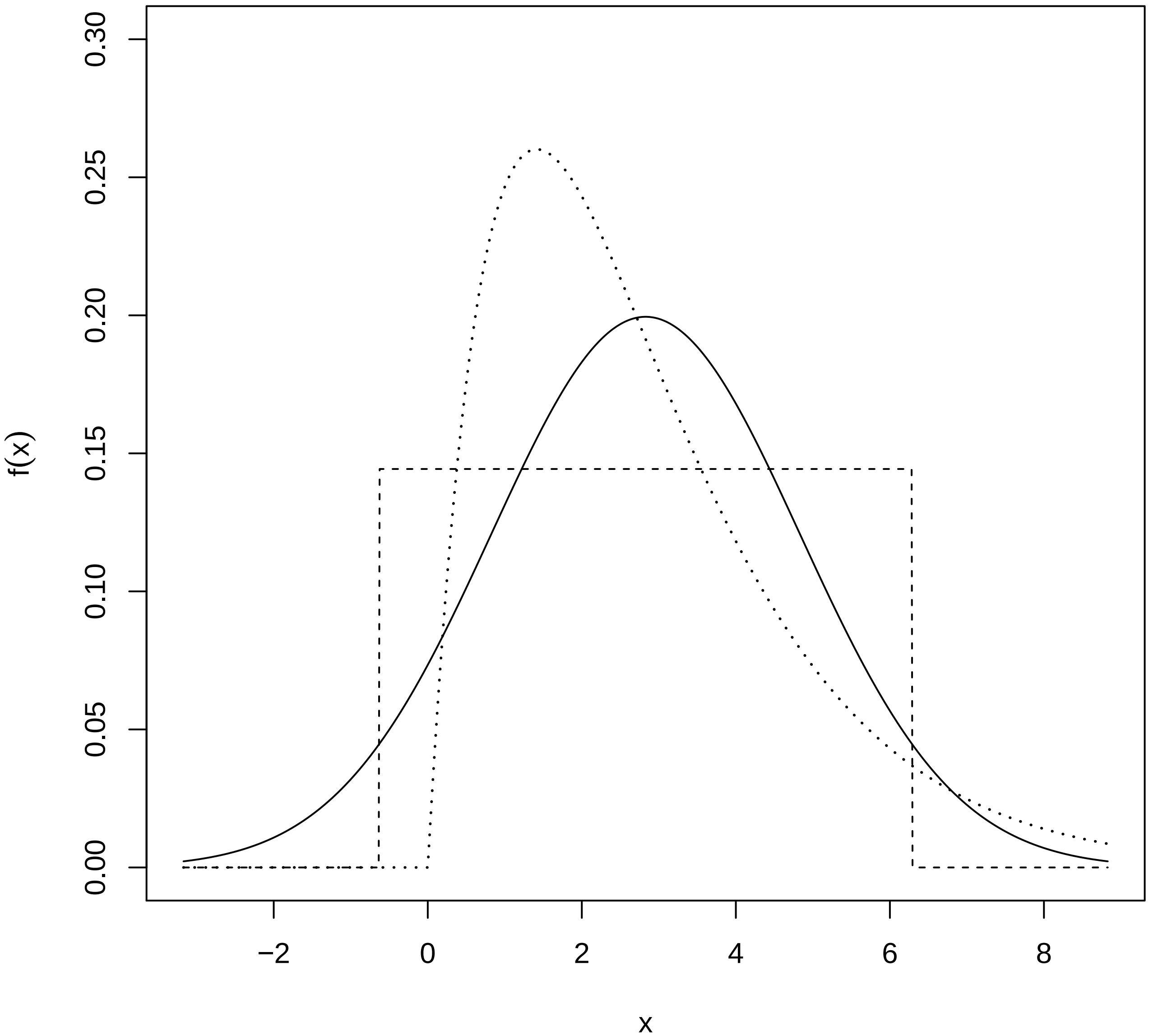
Notice that one of the key steps in finding the mgf of a gamma distribution was recognizing another gamma pdf.

$$\begin{aligned} M_X(t) &= \frac{1}{\Gamma(\alpha)\beta^\alpha} \int_0^\infty \underbrace{x^{\alpha-1} e^{x(t-1/\beta)}}_{\text{looks like } x^{\alpha-1}e^{-x/\beta^*}} \, dx \\ &= \frac{1}{\Gamma(\alpha)\beta^\alpha} \int_0^\infty \underbrace{x^{\alpha-1} e^{-x/(\frac{\beta}{1-\beta t})}}_{\propto \text{gamma}\left(\alpha, \frac{\beta}{1-\beta t}\right) \text{ pdf}} \, dx \end{aligned}$$

The part we recognized is called the kernel of the pdf.

A **kernel** of a function is the main part of the function, the part that remains when constants are ignored.

The mean and variance don't completely describe a distribution. These three distributions all have the same mean and variance.



3.4 Other Quantities Based on Moments

The distribution of a random variable is most often described by its mean and variance. However, as evidence by the previous examples, we might need additional measures. One such quantity is the skewness.

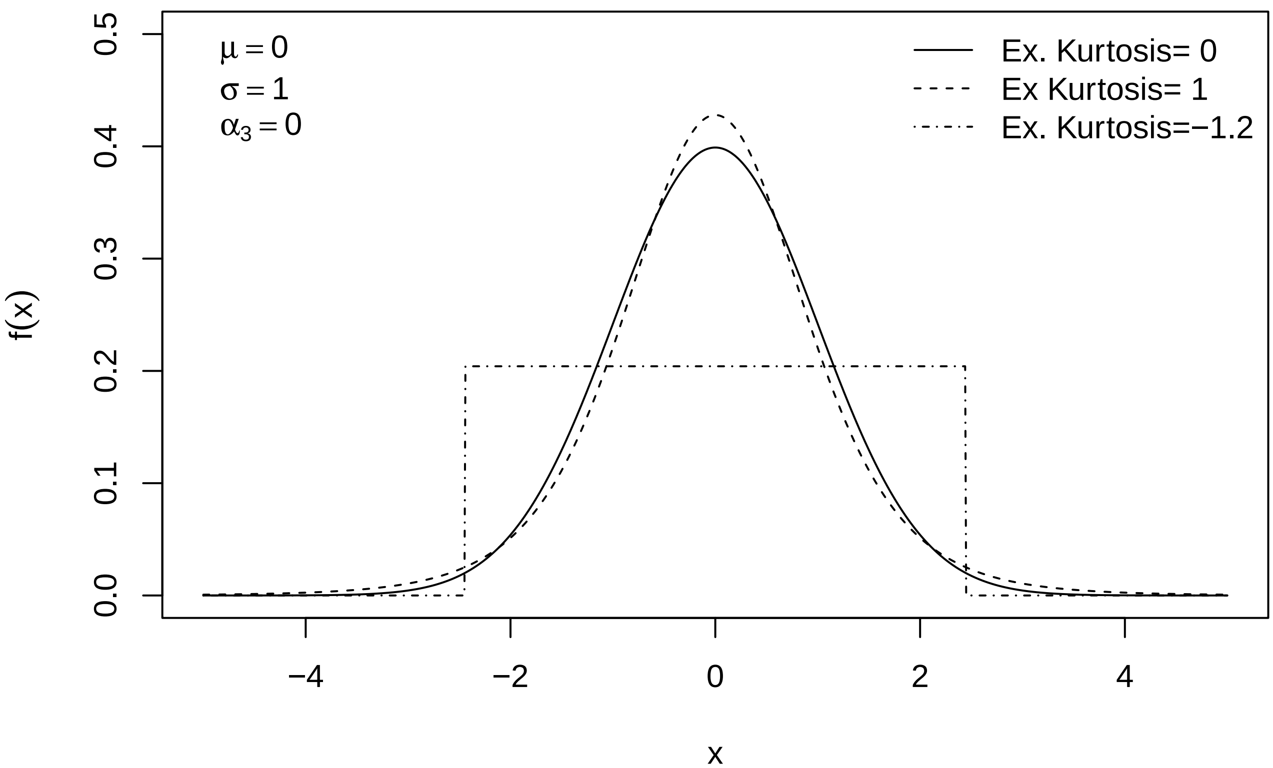
> **Definition: Skewness**
>
> The **skewness** of a random variable X measures the lack of symmetry in the pdf and is defined as

Distributions of random variables can be summarized by a variety of measures, among which are mean, variance/standard deviation and skewness. Random variables with symmetric pdfs/pmfs have a value of ____________ for the skewness. Random variables with right-skewed pdfs/pmfs have ____________ values for the skewness. Random variables with left-skewed pdfs/pmfs have ____________ values for the skewness.

Another such quantity is the kurtosis.

> **Definition: Kurtosis**
>
> The **kurtosis** of a random variable X measures the peakedness or flatness of the pdf and is defined as

The normal distribution has a kurtosis of 3, but often **excess kurtosis** is calculated, rather than kurtosis. Excess kurtosis is $\alpha_4 - 3$, so the normal has excess kurtosis $= 0$.



3.5 Identifying Distributions

Earlier we said that mgfs can also be used to identify distributions, both for RVs and for functions of RVs. How does this work?

> If a moment generating function exists, it is unique

Example: Let X be a random variable with mgf $M_X(t) = e^{3.2(e^t-1)}$. What is the distribution of X?

Example: Let X be a random variable with pdf $f_X(x) = \begin{cases} \frac{1}{\beta} e^{-x/\beta} & x \geq 0 \\ 0 & \text{else} \end{cases}$, where $\beta > 0$.

- What is the mgf of X?

- Where have we seen something similar before?

We can also use mgfs to identify the distributions of functions of random variables (i.e., Transformations–much more on these later).

MGFs of functions of RVs

Assume a random variable X has CDF $F_X(x)$, pdf $f_X(x)$ and mgf $M_X(t)$. If $Y = g(X)$, then the mgf of Y is

$M_Y(t) =$

Example: Consider the linear transformation $Y = aX + b$, where a and b are constants. What is the mgf of Y?

4 Families of Distributions

Statistical distributions are used to model populations. We've already seen this a lot in STAT 102 and STAT 212, when we used the normal distribution to model different populations.

Typically we deal with **families** of distributions. There isn't just one normal distribution–there are an infinite number! What makes one normal distribution different from another are the mean and variance–the parameters of the normal distribution. The name normal refers to a **family** of distributions that change depending on the underlying mean and variance.

Families of distributions are indexed by one (or more!) **parameters**. For the normal, the parameters are μ and σ^2, which represent the underlying mean and variance of the distribution. For other distributions, the parameters may represent other characteristics of the distribution family. For example, we've already seen that the parameter λ represents both the mean and variance of the Poisson distribution. That family of distributions has only parameter.

There are many common distributions; we'll only discuss a few discrete and a few continuous distributions.

4.1 Discrete Distributions

- What makes a distribution discrete?

- Most often:

- We'll discuss: binomial, geometric, negative binomial, hypergeometric, and Poisson
- Summarize: pmf, mean, variance, mgf (if it exists), fun facts

4.1.1 The Binomial Distribution

Binomial Distribution

When it is used? When a random variable is a result of a **Bernoulli process** or a binomial experiment. This requires:

1.

2.

3.

4.

5.

pmf:

$$f_Y(y; n, p) = P(Y = y) = \begin{cases} & y = 0, 1, 2, \ldots, n \text{ and } 0 \leq p \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

Note: The notation used above ($f(y; n, p)$, may also be denoted as $f(y|n, p)$) implies the value of the function is dependent on the values of n and p. A value such as n or p required for the calculation of any probability is called a **distributional parameter**.

mgf: $M_Y(t) =$

Expected Value: $\mathrm{E}(Y) =$

Variance: $\mathrm{Var}(Y) =$

In R:

- To find $P(Y = y)$ use `dbinom(y, size, prob)`
- To find $P(Y \leq y)$ use `pbinom(y, size, prob)`
- To generate n random binomial RVs use `rbinom(n, size, prob)`

4.1.2 The Geometric Distribution

Geometric Distribution

When it is used? When a random variable represents the number of the Bernoulli trial on which the first success occurs. The following conditions are also required:

1.

2.

3.

pmf:

$$f_Y(y;p) = P(Y = y) = \begin{cases} & y = 1, 2, \dots \text{ and } 0 \leq p \leq 1 \\ \\ 0 & \text{otherwise} \end{cases}$$

Note: The notation used above ($f(y;p)$, may also be denoted as $f(y|p)$) implies the value of the function is dependent on the value of p. The value p is the **parameter** of the geometric distribution.

mgf: $M_Y(t) =$

Expected Value: $\mathrm{E}(Y) =$

Variance: $\mathrm{Var}(Y) =$

In R:

- To find $P(Y = y)$ use `dgeom(y-1, prob)`
- To find $P(Y \leq y)$ use `pgeom(y-1, prob)`
- To generate n random binomial RVs use `rgeom(n, prob) + 1`

Key Questions:

- How is a Binomial random variable similar to a Geometric random variable? How are they different?
- Why does the support of the Geometric distribution start at 1, but the support of the Binomial distribution starts at 0?

4.1.3 The Negative Binomial Distribution

Negative Binomial Distribution

When it is used? When a random variable represents the number of the Bernoulli trial on which the r^{th} success occurs ($r = 2, 3, 4$, etc.). The following conditions are also required:

1.

2.

3.

pmf:

$$f_Y(y; r, p) = P(Y = y) = \begin{cases} & y = r, r+1, r+2, \ldots \text{ and } 0 \leq p \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

Note: The notation used above ($f(y; r, p)$, may also be denoted as $f(y|r, p)$) implies the value of the function is dependent on the values of r and p. The values r and p are the **parameters** of the negative binomial distribution.

mgf: $M_Y(t) =$

Expected Value: $\mathrm{E}(Y) =$

Variance: $\mathrm{Var}(Y) =$

In R:

- To find $P(Y = y)$ use `dnbinom(y-r, r, prob)`
- To find $P(Y \leq y)$ use `pnbinom(y-r, r, prob)`
- To generate n random binomial RVs use `rnbinom(n, r, prob) + r`

Key Questions:

- How is a Negative Binomial random variable similar to a Geometric random variable? How are they different?
- Why does the support of the Negative Binomial distribution start at r?

4.1.4 The Hypergeometric Distribution

Hypergeometric Distribution

When it is used? When a random variable represents the number of items with a certain characteristic observed in a sample of size n, drawn from a population with r total items with the characteristic of interest and $N-r$ total items without the characteristic. The following conditions are also required:

1.

2.

pmf:

$$f_Y(y; N, r, n) = P(Y = y) = \begin{cases} & y = 0, 1, 2, \ldots, n \text{ and } \\ & y \leq r \text{ and } n - y \leq N - r \\ 0 & \text{otherwise} \end{cases}$$

Note: The notation used above ($f(y; N, r, n)$, may also be denoted as $f(y|N, r, n)$) implies the value of the function is dependent on the values of N, r and n. The values N, r and n are the **parameters** of the hypergeometric distribution.

Expected Value: $\mathrm{E}(Y) =$

Variance: $\mathrm{Var}(Y) =$

In R:

- To find $P(Y = y)$ use `dhyper(y, r, N-r, n)`
- To find $P(Y \leq y)$ use `phyper(y,r,N-r,n)`
- To generate w hypergeometric RVs use `rhyper(w, r, N-r, n)`

Key Questions:

- How is a Binomial random variable similar to a Hypergeometric random variable? How are they different?

- Why are the bounds $y \leq r$ and $n - y \leq N - r$ required in the support?

4.1.5 The Poisson Distribution

Poisson Distribution

When it is used? When a random variable represents the number of occurrences over a certain amount of time or space.

pmf:

$$f_Y(y;\lambda) = P(Y=y) = \begin{cases} & y = 0,1,2,\dots,n \text{ and } \lambda > 0 \\ \\ 0 & \text{otherwise} \end{cases}$$

Note: The notation used above ($f(y;\lambda)$, may also be denoted as $f(y|\lambda)$) implies the value of the function is dependent on the value of λ. The value λ is the **parameter** of the Poisson distribution.

mgf: $M_Y(t) =$

Expected Value: $\mathrm{E}(Y) =$

Variance: $\mathrm{Var}(Y) =$

In R:

- To find $P(Y=y)$ use `dpois(y,lambda)`
- To find $P(Y \leq y)$ use `ppois(y,lambda)`
- To generate n Poisson RVs use `rpois(n,lambda)`

Useful Facts about the Poisson Distribution:

- Recursive Relationship:
- If Y_1 is Poisson(λ_1) random variable and Y_2 is an independent Poisson(λ_2) random variable, then $X = Y_1 + Y_2$ is a Poisson($\lambda_1 + \lambda_2$) random variable

4.2 Continuous Distributions

- What makes a distribution continuous?
- We'll discuss: uniform, normal, gamma, exponential, chi-square, and beta
- Summarize: pmf, mean, variance, mgf (if it exists), fun facts

4.2.1 The Uniform Distribution

Uniform Distribution

When it is used? When a random variable takes on any value between two limits θ_1 and θ_2 with constant probability.

pdf:

$$f_Y(y;\theta_1,\theta_2) = \begin{cases} & \theta_1 \le y \le \theta_2 \\ & \\ 0 & \text{otherwise} \end{cases}$$

Note: The values θ_1 and θ_2 are the **parameters** of the distribution. The parameter θ_1 is the minimum value the random variable can take on, while θ_2 is the maximum.

mgf: $M_Y(t) =$

Expected Value: $\mathrm{E}(Y) =$

Variance: $\mathrm{Var}(Y) =$

In R:

- To find $P(Y \le y)$ use `punif(y, min= , max= )`
- To generate n random uniform(θ_1, θ_2) RVs use `runif(n, min=theta1, max=theta2)`

Show that if $\theta_1 < \theta_2$ and Y is a random variable uniformly distributed over the interval (θ_1, θ_2), then $\mathrm{E}(Y) = \frac{\theta_1+\theta_2}{2}$.

4.2.2 The Normal Distribution

Normal Distribution

When it is used? All the time! The normal distribution is the most widely used continuous probability distribution, mainly because it is tractable analytically, it follows the familiar bell shape which is consistent with a lot of population models, and the Central Limit Theorem says that, with a large enough sample, the normal distribution can be used to approximate a large variety of other distributions (e.g., Normal approximation to the Binomial). The chi-square, t and F distributions are all by-products of the normal distribution.
pdf:

- **Normal**: $f_Y(y; \mu, \sigma^2) = \begin{cases} & -\infty < y < \infty \\ & \text{where } -\infty < \mu < \infty \\ & \text{and } \sigma^2 > 0 \end{cases}$

- If $Y \sim N(\mu, \sigma^2)$, then $Z = \frac{Y-\mu}{\sigma} \sim N(0, 1)$. The distribution of Z is called the **standard normal distribution**.

  Standard Normal:$f_Z(z) = \begin{cases} & -\infty < z < \infty \end{cases}$

Note: The values μ and σ^2 are the **parameters** of the distribution. The standard normal distribution is a special case of the normal distribution, where $\mu = 0$ and $\sigma^2 = 1$. All normal probabilities are calculated in terms of the standard normal.

mgf: $M_Y(t) =$

Expected Value: $\mathrm{E}(Y) =$
Variance: $\mathrm{Var}(Y) =$

In R:

- To find $P(Y \leq y)$ use `pnorm(y, mean, sd)`
- To generate n random normal(μ, σ^2) RVs use `rnorm(n, mean, sd)`

When using a Normal distribution to approximate a Binomial distribution, what are μ and σ^2? What are suitable conditions needed for this approximation?

4.2.3 The Gamma, Exponential, and Chi-Square Distributions

Gamma, Exponential, and Chi-Square Distributions

When are they used? When a random variable describes the time between events or the time to an event occurring, such as an equipment failure (reliability analysis) or death (survival analysis). They are generally used when the scenario being modeled results in observations with a right-skewed distribution.
pdf:

- **Gamma**: $f_Y(y;\alpha,\beta) = \begin{cases} & 0 < y < \infty \text{ where } \alpha, \beta > 0 \\ 0 & \text{elsewhere} \end{cases}$

  where $\Gamma(\alpha) =$

- **Exponential**: $(\alpha = 1)$ $f_Y(y;\beta) = \begin{cases} & 0 < y < \infty \text{ where } \beta > 0 \\ 0 & \text{elsewhere} \end{cases}$

- **Chi-Square**: $(\alpha = \nu/2 \text{ and } \beta = 2) f_Y(y;\nu) = \begin{cases} & 0 < y < \infty \\ 0 & \text{elsewhere} \end{cases}$

Note: The values α and β are the **parameters** of the distribution. Assuming α is an integer, $\Gamma(\alpha+1) = \alpha!$. The exponential distribution is a special case of the gamma distribution where $\alpha = 1$, and it is often used to model lifetimes. The chi-square distribution is also a special case of the gamma distribution where $\alpha = \nu/2$ and $\beta = 2$, where ν is a positive integer (aka degrees of freedom).

mgf: $M_Y(t) =$

Expected Value: $\mathrm{E}(Y) =$

Variance: $\mathrm{Var}(Y) =$

In R:

- Gamma/Exponential: To find $P(Y \leq y)$ use `pgamma(y, alpha, beta)`

  To generate n random gamma/exp RVs use `rgamma(n, alpha, beta)`

- Chi-Square: To find $P(Y \leq y)$ use `pchisq(y, df)`

  To generate n random chi-square RVs use `rchisq(n, df)`

4.2.4 The Beta Distribution

Beta Distribution

When it is used? When a random variable is defined over the closed interval $0 \le y \le 1$; typically it represents proportions.

pdf:

$$f_Y(y; \alpha, \beta) = \begin{cases} & 0 \le y \le 1 \text{ where } \alpha, \beta > 0 \\ 0 & \text{otherwise} \end{cases}$$

Note: The values α and β are the **parameters** of the distribution.

Expected Value: $\mathrm{E}(Y) =$

Variance: $\mathrm{Var}(Y) =$

In R:

- To find $P(Y \le y)$ use `pbeta(y, alpha, beta)`
- To generate n random beta RVs use `rbeta(n, alpha, beta)`

- How can the Beta distribution be applied to a random variable defined over the interval $a \le y \le b$, where $a \ne 0$ and $b \ne 1$?

- Let $Y \sim \mathrm{Beta}(\alpha, \beta)$. Show that $\mathrm{E}(Y) = \frac{\alpha}{\alpha+\beta}$.