

Stat 262 Notes

Table of contents

Course Goals for STAT 262	3
1 Probability	4
1.1 Basic Probability Definitions and Calculations	4
1.2 Conditional Probability	9
1.3 Independence	16
2 Random Variables and their Associated Functions	17
2.1 Random Variables	17
2.2 Distribution Functions	18
2.3 Density and Mass Functions	22

Course Goals for STAT 262

STAT 262 is an introduction to the basic probability theory that allows us to 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 functions 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 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 $P(\cdot)$ that satisfies

1.

2.

3.

Requirements (1) - (3) are called the

Any function $P(\cdot)$ 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 $\{s_1, s_2, \dots, s_n\}$ 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, \dots 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, \dots , 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 **mutually 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:

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

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 $X = \text{number of ones rolled on two dice}$. 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 $X \sim \text{Uniform}(0,1)$. 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 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

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

- 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 $x = x_j$, $j = 1, 2, \dots$, then

Example: In game of SKUNK, we defined $X = \text{number of ones rolled on two dice}$. 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 probability 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., $P(X = a) = 0$). 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 $X \sim \text{Uniform}(0,1)$. 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?