

STAT 501 Fall 2025 Course Notes

Chapter 1: Probability Theory

Probability is the foundation upon which all of statistics is built, providing a means for modeling populations, experiments, or almost anything else that could be considered a random phenomenon. Through these models, statisticians are able to draw inferences about populations, inferences based on examination of only a part of the whole. –Casella & Berger, 2002, p. 1

1.1 Set Theory (read on your own)

Just as statistics builds upon the foundation of probability theory, probability theory builds upon set theory. Set theory provides a grammar for probability. ***Read this section in the text on your own! Some review below that is straight from the text...***

- **DEF 1.1** The set, S of all possible outcomes of a particular experiment is called the *sample space* for the “experiment”
- **DEF 1.1.2** An *event* is any collection of possible outcomes of an experiment, that is any subset of S (including S itself)
 - Let $A = \{x \in S : x \in A\}$ be an event in S . A *occurs* if the result of the experiment is in A .
 - **Containment:** $A \subset B \iff x \in A \implies x \in B$
 - **Equality:** $A = B \iff A \subset B$ **and** $B \subset A$
- Given any two sets, A and B in a sample space S , we have the following elementary set operations:
 - **Union:** The union of A and B , denoted $A \cup B$ is the set of elements that belong to either A or B . Write out the set notation to represent the union of two sets, A and B that both exist in the same sample space, S
 - **Intersection:** the intersection of A and B , denoted $A \cap B$, is the set of elements that belong to both A and B . Write out the set notation to represent the union of two sets, A and B that both exist in the same sample space, S

- **Complementation:** The complement of A , written A^c (also as, A' , \overline{A}), is the set of all elements that are not in A . Write out the set notation to represent the union of two sets, A and B that both exist in the same sample space, S

The ability to write out formal set notation for events and elementary set operations on those events will help you prove things like THM 1.1.4. See pg 3. for an example of a formal proof for the distributive law. Try to prove the rest of the Laws on your own before our next class!

THM 1.1.4 For any three events A, B , and C , defined on a sample space S ,

- | | |
|----------------------|--|
| a. Commutativity | $A \cup B = B \cup A$
$A \cap B = B \cap A$ |
| b. Associativity | $A \cup (B \cup C) = (A \cup B) \cup C$
$A \cap (B \cap C) = (A \cap B) \cap C$ |
| c. Distributive laws | $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$
$A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$ |
| d. DeMorgan's Laws | $(A \cup B)^c = A^c \cap B^c$
$(A \cap B)^c = A^c \cup B^c$ |

- **DEF 1.1.5** Two events A and B are *disjoint* (or *mutually exclusive*) if $A \cap B = \emptyset$. The events A_1, A_2, \dots are *pairwise disjoint* (or *mutually exclusive*) if $A_i \cap A_j = \emptyset$ for all $i \neq j$.
- **DEF 1.1.6** If A_1, A_2, \dots are pairwise disjoint and $\cup_{i=1}^{\infty} A_i = S$, then the collection A_1, A_2, \dots forms a *partition* of S .

1.2 Basics of Probability Theory

Recall, a **sample space**, denoted by S is the set of all possible outcomes of a particular experiment. An **event** is any collection of possible outcomes of an experiment, i.e., an event is any subset of S .

Goal: Associate with each event $A \subset S$ a number between 0 and 1 (inclusive) that is called the probability of A , denoted by $P(A)$ (or $Pr(A)$).

- $P(\cdot)$
- $P(A)$ can always be defined for all sets A in a *Borel field*, \mathcal{B} , of subsets of S

DEF 1.2.1 (Borel Field) A collection of subsets of S is called a *sigma algebra* (σ -algebra) or *Borel field*, denoted by \mathcal{B} if it satisfies the following three properties:

Based on these properties, we also know the following:

- $S \in \mathcal{B}$
- $\cap_{i=1}^{\infty} A_i \in \mathcal{B}$

DEF 1.2.4 (Probability Function) Given a sample space S and an associated *sigma algebra* \mathcal{B} , a probability function is a function P with domain \mathcal{B} that satisfies,

- 1.
- 2.
- 3.

Note: Properties 1-3 are called the **Axioms of Probability**.

Example: Suppose we have a three-sided die with numbers 1, 2, 3.

- Sample Space:
- Borel field:
- Function, P is defined as:

Is P a probability function?

Note: In all subsequent material, we assume P is a probability function and all relevant sets are in \mathcal{B} .

The Calculus of Probabilities

Many other properties of the probability function can be derived from the Axioms of Probability.

THM 1.2.8 If P is a probability function and A is any set in \mathcal{B} , then

a.

b.

c.

Proof: On your own - see textbook.

THM 1.2.9 If P is a probability function and A and B are any sets in \mathcal{B} , then

a.

b.

c.

Proof:

a. $P(B \cap A^c) = P(B) - P(A \cap B)$

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

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

Results for Collections of Events

The first part of the next theorem is the basis for the Law of Total Probability, which we'll see in Section 1.3.

THM 1.2.11 If P is a probability function, then

- a. $P(A) = \sum_{i=1}^{\infty} P(A \cap C_i)$ for any partition C_1, C_2, \dots
- b. Boole's Inequality $P(\cup_{i=1}^{\infty} A_i) \leq \sum_{i=1}^{\infty} P(A_i)$ for any sets A_1, A_2, \dots ,

Proof:

- a. $P(A) = \sum_{i=1}^{\infty} P(A \cap C_i)$ for any partition C_1, C_2, \dots

b. $P(\cup_{i=1}^{\infty} A_i) \leq \sum_{i=1}^{\infty} P(A_i)$ for any sets A_1, A_2, \dots ,

Bonferroni's Inequality

If P is a probability function and $A_1, A_2, \dots, A_n \in \mathcal{B}$, then $P(\cap_{i=1}^n A_i) \geq \sum_{i=1}^n P(A_i) - (n - 1)$

Proof:

Examples

1. Consider two sets, $A, B \in \mathcal{B}$, and suppose $P(A) = P(B) = 0.95$. What is the minimum possible probability of their intersection?
2. Create an example with two sets $A, B \in \mathcal{B}$ in which Bonferroni's Inequality isn't all that useful.

Inclusion-Exclusion Identity

The result in Theorem 1.2.9 for a probability of a union of two events can be generalized to find the probability of a finite union of n events. This is called the **inclusion-exclusion identity** (see Miscellanea Section 1.8.1): If P is a probability function and $A_1, A_2, \dots, A_n \in \mathcal{B}$, then

$$P(\cup_{i=1}^n A_i) = \sum_{i=1}^n P(A_i) - \sum_{i < j} P(A_i \cap A_j) + \sum_{i < j < k} P(A_i \cap A_j \cap A_k) + \dots + (-1)^{n-1} P(\cap_{i=1}^n A_i).$$

1.2.3-1.2.4 Combinatorics (Counting)

When we can define a sample space as a set of equally likely outcomes, methods of counting (i.e., combinatorics) can be used in order to construct probability assignments.

Fundamental Theorem of Counting

THM 1.2.14 If a job of k separate tasks, the i^{th} of which can be done in n_i ways, $i = 1, \dots, k$, then the entire job can be done in

Proof by tree diagram:

Examples

1. Dr. Hancock's six-year-old daughter, Adeline, has five summer shirts, three pairs of shorts, and two pairs of sandals. How many possible summer outfits can Adeline wear?

2. Consider a finite sample space $S = \{a, b, c, d\}$. How many possible events (subsets of S) can we create from this sample space?

Possible methods of counting (Table 1.2.1 p. 16)

The number of possible arrangements of size r from n distinct objects, $n \geq r$:

	Without replacement	With replacement
Ordered (Permutation)		
Unordered (Combination)		

Example: Flip a fair coin ten times. In how many ways can we observe 3 heads and 7 tails?
What is the probability of observing 3 heads and 7 tails?

Example: In the Powerball lottery, a player selects five numbers between 1 and 53 and a Powerball number between 1 and 42. For each drawing, five balls are drawn at random from a hopper with 53 white balls numbered 1 to 53. A sixth ball is drawn from a second hopper with 42 red balls numbered 1 to 42.

- Ordered, without replacement
- Ordered, with replacement
- Unordered, without replacement

- Unordered, with replacement

Poker Probabilities

Consider choosing a five-card poker hand from a standard deck of 52 playing cards (four suits: hearts, diamonds, clubs, spades — each suit has ranks 2, 3, \dots , 10, J, Q, K, A).

Find the probability of:

- Four aces (Ans: $48/2,598,960$)
- Four of a kind (a hand that contains four cards of one rank and one card of another rank) (Ans: $624/2,598,960$)
- Royal flush (a hand that contains ranks 10, J, Q, K, A, all of the same suit) (Ans: $4/2,598,960$)
- Straight flush (a hand that contains five cards of sequential rank, all of the same suit) (Ans: $36/2,598,960$)
- Flush (a hand that contains five cards all of the same suit, not all of sequential rank) (Ans: $5,108/2,598,960$)
- Full house (a hand that contains three cards of one rank and two cards of another rank) (Ans: $3,744/2,598,960$)

Calculating probabilities when sampling with replacement or with indistinguishable items

- When sampling r items from n distinct objects *without replacement*, you can use either the ordered or unordered sample space to calculate probabilities. Each outcome in the unordered sample space corresponds to _____ outcomes in the ordered sample space.
- When sampling *with replacement*, each outcome in the unordered sample space corresponds to _____ outcomes in the ordered sample space.

Example 1.2.19: Write out the outcomes in the unordered and ordered sample spaces and find each outcome's probability when sampling $r = 2$ items from $n = 3$ distinct objects *with replacement*.

Example: How many words can be formed from the letters MISSIPPI?

Multinomial Coefficient

If there are r places, and we have m different numbers repeated k_1, k_2, \dots, k_m times, then the number of ordered samples of size r is:

Example: There are eight Statistics faculty in the Department of Mathematical Sciences at MSU — three men and five women. These faculty need to be divided among three committees, with 3 assigned to committee A, 3 assigned to Committee B, and 2 assigned to Committee C. In how many ways can the committee assignments be made?

What if, additionally, each committee was required to have at least one female Statistics faculty?

1.3 Conditional Probability and Independence

Often we have partial information about a certain phenomenon and we want to know how this knowledge affects the probabilities of outcomes of interest, if at all. For example, we might want to know the probability of smoke in the air tomorrow, given that there's smoke in the air today.

Example: An urn contains 10 white, 5 yellow, and 10 black marbles. A marble is chosen at random from the urn, and it is noted that the marble is not black. What is the probability it is yellow? What is your reasoning?

Conditional Probability

DEF 1.3.2 If A and B are events in S and $P(B) > 0$, then the *conditional probability* of A given B is,

Example Revisited

Properties of Conditional Probability

- For any B where $P(B) > 0$, the probability function $P(\cdot|B)$ satisfies the Axioms of Probability:
 1. $P(A|B) \geq 0$ for all $A, B \in \mathcal{B}$
 2. $P(S|B) = 1$
 3. if $A_1, A_2, \dots \in \mathcal{B}$ and $A_i \cap A_j = \emptyset$ for all $i \neq j$, then $P(\cup_{i=1}^{\infty} A_i|B) = \sum_{i=1}^{\infty} P(A_i|B)$
- Other properties of conditional probability:
 1. $P(\emptyset|B) = 0$
 2. $P(A|B) \leq 1$
 3. $P(A^c|B) = 1 - P(A|B)$
 4. $P(A_1 \cup A_2|B) = P(A_1|B) + P(A_2|B) - P(A_1 \cap A_2|B)$

Bayes' Theorem

Derivation

Law of Total Probability If A_1, A_2, \dots , form a partition of S and $P(A_i) > 0 \forall i = 1, 2, \dots$, then for any event B in \mathcal{B}

$$P(B) =$$

Note: This also applies for a finite collection of A_i s

Proof:

Bayes' Theorem (General Version)

THM 1.3.5 Let A_1, A_2, \dots , be a partition of S and $P(A_i) > 0 \forall i = 1, 2, \dots$, and let B be any set in \mathcal{B} . Then for each $i = 1, 2, \dots$,

$$P(A_i|B) =$$

Examples

1. An insurance company believes that people can be divided into two classes: accident-prone and nonaccident-prone. Their records show that an accident-prone person will have an accident at some time within a fixed 1-year period with probability 0.4, whereas this probability is 0.2 for a nonaccident-prone person. If we assume that 30% of the population is accident-prone, what is the probability that a new policy holder will have an accident within a year of purchasing a policy?

Suppose a new policy holder has an accident within a year of purchasing the policy. What is the probability that the person is accident-prone?

2. Suppose the probability a diagnostic test correctly detects the presence of a certain disease is 0.95 and that the proportion of women in a given population suffering from this disease is 0.001. If a woman in the target population does not have the disease, the test will report that she does not have it with probability .8. A woman 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 given that the test indicates she has it?

3. At a psychiatric clinic, the social workers are so busy that, on the average, only 60% of potential new patients that telephone are able to talk immediately with a social worker when they call. The other 40% are asked to leave their phone number. About 75% of the time a social worker is able to return the call on the same day, and the other 25% of the time the caller is contacted the following day. Experience at the clinic indicates that the probability a caller will actually visit the clinic for consultation is 0.8 if the call was immediately able to speak with a social worker, whereas it is 0.6 and 0.4, respectively, if the patient's call was returned the same day or the following day.
- What percentage of people that telephone visit the clinic for consultation?
 - What percentage of patients that visit the clinic did not have to have their telephone calls answered immediately?

Independence Motivation:

If two events A and B are unrelated, then intuitively it should be the case that $P(A|B) = P(A)$ and $P(B|A) = P(B)$.

Also, it follows that $P(A \cap B) = P(A)P(B)$.

Statistical Independence

DEF 1.3.7 Two events, A and B are *statistically independent* if and only if

Note: Independence and disjointness are related properties.

- If A and B are disjoint, then A and B are dependent as long as $P(A) > 0$ and $P(B) > 0$.
- If A and B are independent, then A and B are not disjoint as long as $P(A) > 0$ and $P(B) > 0$.

More Independence

THM 1.3.9 If A and B are independent events, then the following pairs are also independent:

1. A and B^c
2. A^c and B
3. A^c and B^c

Proof:

1. If A and B are independent events, then A and B^c are independent.
2. If A and B are independent events, then A^c and B are independent. (on your own - similar to above)
3. If A and B are independent events, then A^c and B^c are independent.

Mutual Independence

DEF 1.3.12 A collection of events A_1, A_2, \dots, A_n are mutually independent if for any subcollection $A_{i_1}, A_{i_2}, \dots, A_{i_k}$, we have

$$P\left(\cap_{j=1}^k A_{i_j}\right) = \prod_{j=1}^k P(A_{i_j}).$$

Note: Pairwise independence does not imply mutual independence!

Example: Why is independence so wonderful?

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 maneuver successfully through them are 0.6, 0.4, and 0.2, respectively, what is the probability that the mouse...

a. will be able to escape?

b. will not be able to escape?

1.4 Random Variables

The outcome of an experiment is completely described by sample space S with probability function $P(\cdot)$ on its events. By observing a $s \in S$ selected according to $P(\cdot)$, one can address questions concerning the experiment. However, we usually only need to observe some function of $s \in S$ in order to address the questions of interest.

Example: Suppose we roll two dice. There are 36 possible outcomes, and with a slight abuse of notation they can be written concisely as,

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

In the game SKUNK, we don't necessarily care about the specific outcome of a roll. For example, we don't care to know that, specifically a (6,2) was rolled. Instead, we are interested in knowing...

In this example (and many others), we only need to know some function of $s \in S$ in order to address the research question of interest!

Random Variable

DEF 1.4.1 A random variable is a function from a sample space, S into the real numbers.

General Idea:

- We have sample space $S = \{s_1, \dots, s_n\}$ with probability function P and define a random variable X with range $\mathcal{X} = \{x_1, \dots, x_m\}$.
- We use P to define a probability function P_X on $\mathcal{X} : P_X(X = x_i) = P(\{s_j \in S : X(s_j) = x_i\})$.

Example Revisited: Define the random variable of interest and its corresponding probability function for the game of SKUNK

In general, we want $P_X(X \in A) = P(\{s_j \in S : X(s_j) \in A\})$, for any A that is a subset of a certain Borel field of subsets of \mathcal{X} .

1.5 Distribution Functions

With every random variable X , we associate a function called the cumulative distribution function of X .

Cumulative Distribution Function (DEF 1.5.1) The *cumulative distribution function* or *cdf* of a random variable X , denoted by $F_X(x)$ is defined by,

Example Revisited: Find and graph the cdf of $X =$ number of ones rolled on two fair dice.

Properties of CDFs

THM 1.5.3 The function $F_X(x)$ is a cdf if and only if the following three conditions hold:

- 1.
- 2.
- 3.

Example (Discrete) A certain basketball player has a free-throw percentage of $p \times 100$. Assume independence from shot to shot. Let X be the number of shots required to make a basket. Find the cdf of X and verify that it is indeed a cdf.

Example (Continuous) Suppose the time to failure (in hours) of an external laptop monitor has cdf,

$$F_Y(y) = \begin{cases} 0 & y < 0 \\ 1 - \exp\left\{\frac{-y}{2000}\right\} & y \geq 0 \end{cases}$$

Verify that this is a cdf.

Example (Mixture) Suppose we are interested in studying the delay a motorist encounters at a one-way stop sign. Let X denote the amount of delay (in seconds) a randomly selected motorist encounters after making the stop, and let p be the probability there is a delay. Verify that for constant $\lambda > 0$, the following distribution of X is indeed a valid cdf.

$$F_X(x) = \begin{cases} 0 & x < 0 \\ 1 - pe^{-\lambda x} & x \geq 0 \end{cases}$$

Summary What are the major differences between cdfs for discrete and continuous RVs?

Continuous / Discrete RVs

DEF 1.5.7

- A random variable X is continuous if $F_X(x)$ is a continuous function of x .
- A random variable X is discrete if $F_X(x)$ is a step function of x .

Identically Distributed

DEF 1.5.8 The random variables X and Y are identically distributed if, for every set $A \in \mathcal{B}$, $P_X(X \in A) = P_Y(Y \in A)$.

THM 1.5.10 The following two statements are equivalent:

- The random variables X and Y are identically distributed.
- $F_X(t) = F_Y(t)$ for all real t .

Note: X and Y are identically distributed \nRightarrow X and Y are equal.

1.6 Density and Mass Functions

Often, we are interested in “point probabilities” of random variables (RV).

To get these, we consider another function, called either the probability density function (pdf) or probability mass function (pmf), that is often associated with a random variable X and its cdf $F_X(x)$. The terms *pmf* and *pdf* refer, respectively to the discrete and continuous cases.

Probability Mass Function (*pmf*)

DEF 1.6.1 The *probability mass function (pmf)* of a discrete random variable X , denoted $f_X(x)$, is given by

For a discrete random variable X , let $\mathcal{X} = \{x_1, x_2, \dots\}$ be the set of possible discrete values that the random variable can assume with non-zero probability, i.e., $f_X(x_j) = P(X = x_j) = p_j > 0 \ \forall x_j \in \mathcal{X}$. *Note: \mathcal{X} is called the **support** of X .* The *pmf* of X is represented by,

Furthermore, for discrete random variables, there is a special relationship between the CDFs and the pmfs:

Examples Revisited

- A certain basketball player has a free-throw percentage of p . Assume independence from shot to shot. Let X be the number of shots required to make a basket. Find the *pmf* of X ? How did we find the cdf of X ?

- Let X = number of ones rolled on two fair dice. Find the pmf of X using $F_X(x)$.

$$F_X(x) = \begin{cases} 0 & x < 0 \\ 25/36 & 0 \leq x < 1 \\ 35/36 & 1 \leq x < 2 \\ 1 & x \geq 2 \end{cases}$$

Probability density function (pdf)

DEF 1.6.3 The *probability density function (pdf)* of an absolutely continuous random variable X , denoted $f_X(x)$, is the function that satisfies

Furthermore, by the Fundamental Theorem of Calculus,

Notes:

- If X is a continuous random variable, then $P(X = x) = 0$ and $P(a < X < b) = P(a \leq X < b) = P(a < X \leq b) = P(a \leq X \leq b)$.

- If X is a continuous random variable, then $P(a \leq X \leq b) = \int_a^b f_X(x)dx$

- If X is a continuous random variable, then $f_X(x) = \frac{dF_X(x)}{dx}$ needs to exist and be continuous except for, at most, a finite number of points in any finite interval.

Properties of a pdf (or pmf)

THM 1.6.5 a function $f_X(x)$ is a pdf (or pmf) of a random variable X iff

1.

2.

Example: A machine produces copper wire for use in electronics and power cords, and occasionally there is a flaw at some point along the wire. The length of wire (in meters) produced between successive flaws can be represented by the continuous random variable X with pdf of the form,

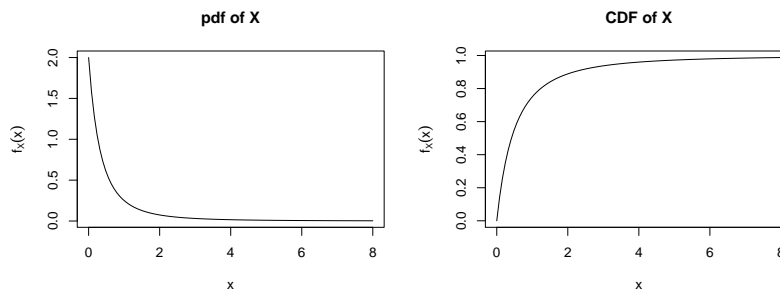
$$f_X(x) = \begin{cases} 0 & x \leq 0 \\ c(1+x)^{-3} & x > 0 \end{cases}$$

- Find c
- Find the cdf of X

Example continued

```
par(mfrow = c(1,2))
#pdf
curve(2*(1 + x)^-3, from = 0, to = 8,
      ylab = expression(f[X](x)), main = "pdf of X")

#cdf
curve(1 - (x + 1)^-2, from = 0, to = 8,
      ylab = expression(f[X](x)), main = "CDF of X")
```



- c. Suppose Dr. Hancock is interested in rewiring her garage door switch because it appears to have lost power. She's getting wire straight off the machine (i.e., beginning is good wire just after a flaw) and she needs 3 meters of wire. What is the probability that a flaw occurs in her wire between 0.40 and 0.45 meters? Find this probability using both the pdf of X and the cdf of X , and represent this probability on the graphs above. **Be sure to write clear and correct mathematical notation to represent the probability you desire!**