STATS50 Lecture Notes

Zed Chance

Fall 20

Contents

1 What is statistics?

The practice of science of collecting and analyzing numerical data in large quantities, especially for the purpose of inferring proportions in a whole from those in a representative sample.

Two major types of statistics:

1. Descriptive statistics
   (a) Consists of methods for organizing and summarizing information
   (b) Graphs, charts, tables, calculating averages, measures of variation, and percentiles
2. Inferential statistics
   (a) Consists of methods for drawing and measuring the reliability of conclusions about a population based on information from a sample of the population

Probability theory is the science of uncertainty. It enables us evaluate and control the likelihood that a statistical inference is correct. In general, probability theory provides the mathematical basis for inferential statistics.

An **experiment** is a process that results in an outcome that cannot be predicted in advance with certainty.

Example 1. Tossing a coin, weighing the contents of a box of cereal, rolling a die ...

The set of all possible outcomes of an experiment is called the **sample space** for the experiment (often notated by the capital letter S). For example:

1. Consider the experiment of tossing a fair coin. There are two possible outcomes: heads (H) or tails (T). The sample space is $S = \{H, T\}$, this is a finite sample space.
2. There are six possible outcomes when a six-sided die is rolled: $1, 2, 3, 4, 5, 6$, so the sample space is $S = \{1, 2, 3, 4, 5, 6\}$, this is a finite sample space.
3. Imagine a punch with a diameter 10mm that punches holes in sheet metal. Because of variations in the angle of the punch and slight movements in the sheet metal, the diameters of the holes vary between 10.0mm and 10.2mm. What might be a reasonable sample space for the experiment of punching a hole? (Say we let x represent the diameter, in mm, of a punched out hole.) The sample space would be $S = \{x \,|\, 10.0 < x < 10.2\}$, this is an infinite sample space.

A subset of a sample space is called an **event**:

1. Can be represented by a Venn diagram.
2. The sample space when a six-sided die is rolled is $S = \{1, 2, 3, 4, 5, 6\}$. The event the die comes up even is $E = \{2, 4, 6\}$.
3. For the previous hole punch example, the event a hole has a diameter less than 10.1mm is the subset $\{x : 10.1 < x < 10.1\}$

Note 1. The empty set: $\emptyset$, is a set with no outcomes or elements in it. For example $\emptyset = \{\}$. The empty set is an event as is the entire sample space. Sometimes referred to as an "impossible event". The sample space is sometimes referred to as a "certain event".

We say an event has occurred if the outcome of the experiment is one of the outcomes in the event.

Example 2. Refer to the previous example of rolling a die: If the die comes up 2 when it is rolled, then we can say that the event $E = \{2, 4, 6\}$, the event an even number is rolled, has occurred.

2 Combining events

The **union** of two events, A and B, denoted by $A \cup B$, is the set of outcomes that belong either to A or B, or to both. In words, $A \cup B$ means "A or B". Thus, the event $A \cup B$ occurs whenever either A or B (or both) occurs.

The **intersection** of two events A and B, denoted $A \cap B$, is the set of outcomes that belong to both A and to B, in words this means "A and B". Thus the event $A \cap B$ occurs whenever both A and B occur.

The **complement** of an event A, denoted A^C is the set of outcomes that do not belong to A. In words, A^C means "not A". Thus, the event A^C occurs whenever A does *not* occur.

Example 3. When a die is rolled, the sample space is:

$$S = \{1, 2, 3, 4, 5, 6\}$$

Consider the following events:

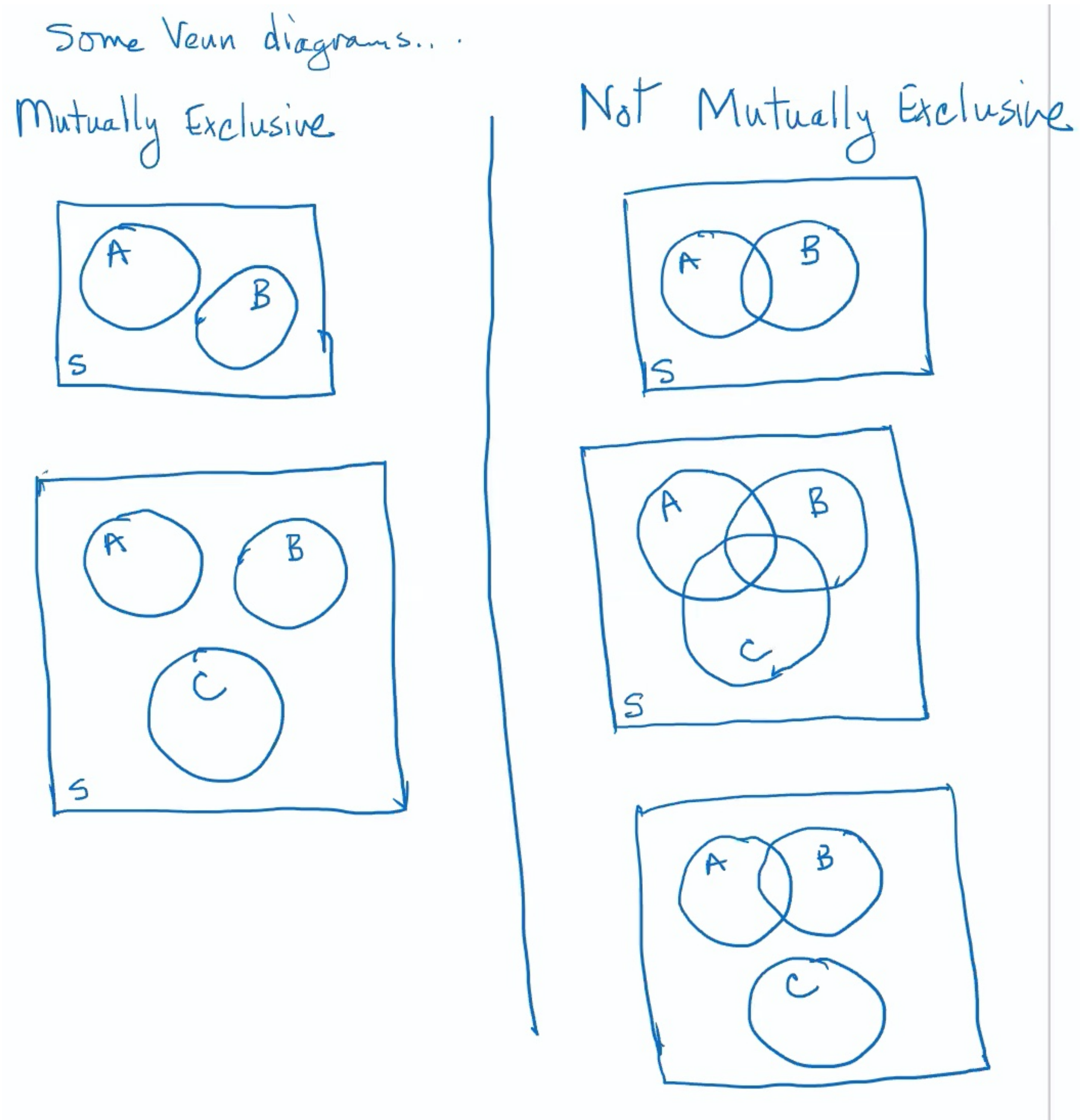
- $A =$ "the event the die comes up even" $= \{2, 4, 6\}$
- $B =$ "the event the die comes up 4 or more" $= \{4, 5, 6\}$
- $C =$ "the event the die comes up at most 2" $= \{1, 2\}$
- $D =$ "the event the die comes up 3" $= \{3\}$

So,

$$\begin{aligned}
A \cap B &= \{4, 6\} \\
A \cup B &= \{2, 4, 5, 6\} \\
A \cup B \cup C &= \{1, 2, 4, 5, 6\} \\
A \cap B \cap C &= \emptyset \\
(A \cap B) \cup C &= \{4, 6\} \cup \{1, 2\} = \{1, 2, 4, 6\} \\
A^C &= \{1, 3, 5\} \\
B \cap A^C &= \{5\} \\
B^C \cup C^C &= \{1, 2, 3, 4, 5, 6\} \\
B^C \cap C^C &= \{3\}
\end{aligned}$$

2.1 Mutually exclusive events

The events A and B are **mutually exclusive** if they have no outcomes in common. More generally, a collection of events $A_1, A_2, \ldots, A_n$ is said to be mutually exclusive if no two of them have any outcomes in common.



Example 4. Consider the previous example's events.

Decide whether the following collection of events are mutually exclusive or not:

1. A, B are not mutually exclusive, they have a 4 in common.
2. B, C are mutually exclusive, no common outcomes.
3. A, B, C are not mutually exclusive (see 1)
4. B, C, D are mutually exclusive, no common outcomes.
5. A, B, C, D are not mutually exclusive (see 1,3)

3 Probability

Given any experiment and any event A:

- The expression $P(A)$ denotes the probability that the event A occurs.
- $P(A)$ is the proportion of times that event A would occur in the long run, if the experiment were to be repeated over and over again.

Example 5. Suppose the probability of a weighted coin comes up heads is 0.75, so:

$$P(\text{heads}) = 0.75$$

I would expect the weighted coin to come up heads 75% of the time.

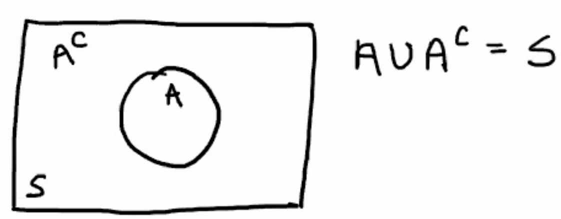
3.1 The axioms of probability

Let S be a sample space, then $P(S) = 1$. For any event A, $0 \leq P(A) \leq 1$. The long run relative frequency is always between 0% and 100%. If A and B are mutually exclusive events, then $P(A \cup B) = P(A) + P(B)$.

More generally: if $A_1, A_2, \cdots$ are mutually exclusive events, then: $P(A_1 \cup A_2 \cup \cdots) = P(A_1) + P(A_2) + \cdots$.

3.2 Complementation rule

For any event A, $P(A^C) = 1 - P(A)$.



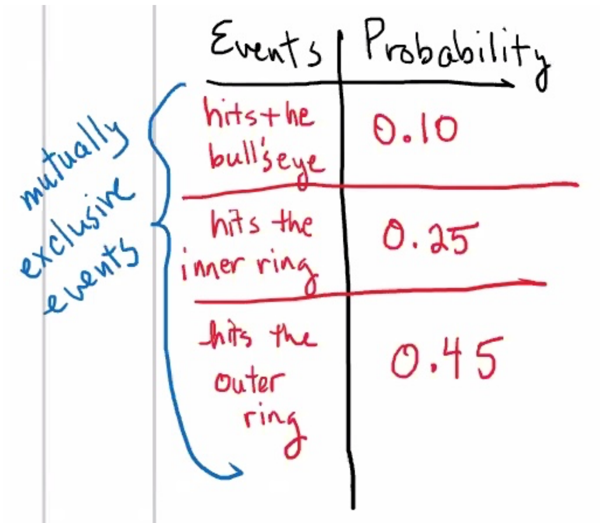
Events A and A^C are mutually exclusive. So:

$$P(A \cup A^C) = P(A) + P(A^C) = P(S) = 1$$

Remember: $P(\emptyset) = 0$.

Example 6. A target on a test firing range consists of a bull's eye with two concentric rings around it. A projectile is fired at the target. The probability it hits the bull's eye is 0.1, the probability that it hits the inner ring is 0.25, and the probability it hits the outer ring is 0.45:

What is the probability that the projectile hits the target?



$$P(\text{target}) = P(\text{bullseye} \cup \text{inner} \cup \text{outer})$$

Since they are mutually exclusive, they can be summed.

$$\begin{aligned} P(\text{hits target}) &= P(\text{hits bullseye OR hits inner ring OR hits outer ring}) \\ &= P(\text{hits bullseye}) + P(\text{hits inner ring}) + P(\text{hits outer ring}) \\ &= 0.10 + 0.25 + 0.45 \\ &= 0.80 \end{aligned}$$

This means there is a 80% chance of hitting the target.

What is the probability that it misses the target?

$$\begin{aligned} P(\text{miss}) &= 1 - P(\text{target}) \\ &= 1 - .8 \\ &= .2 \end{aligned}$$

Example 7. If A is an event containing outcomes $O_1, \ldots, O_n$, that is if $A = \{O_1, \ldots, O_n\}$ then

$$P(A) = P(O_1) + P(O_2) + \cdots + P(O_n)$$

A six sided die is rolled, $S = \{1, 2, 3, 4, 5, 6\}$

Let $E = \{2, 4, 6\}$

Then:

$$P(E) = P(2) + P(4) + P(6)$$

If the die was known to be fair and that each outcome was equally likely, then each individual outcome would have the probability of happening $\frac{1}{6}$ of the time.

So:

$$P(E) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{1}{2}$$

4 Sample spaces with equally likely outcomes

A population from which an item is sampled at random can be thought of as a sample space with equally likely outcomes.

If a sample space has N has equally likely outcomes, the probability of each outcome is $\frac{1}{N}$.

Example 8. A card is randomly selected from a deck of 52 cards. Find the probability the selected card is:

1. A spade

$$P(\text{spade}) = \frac{13}{52} = .25$$

2. A face card

$$P(\text{face card}) = \frac{12}{52} = .23$$

3. A spade or a face card

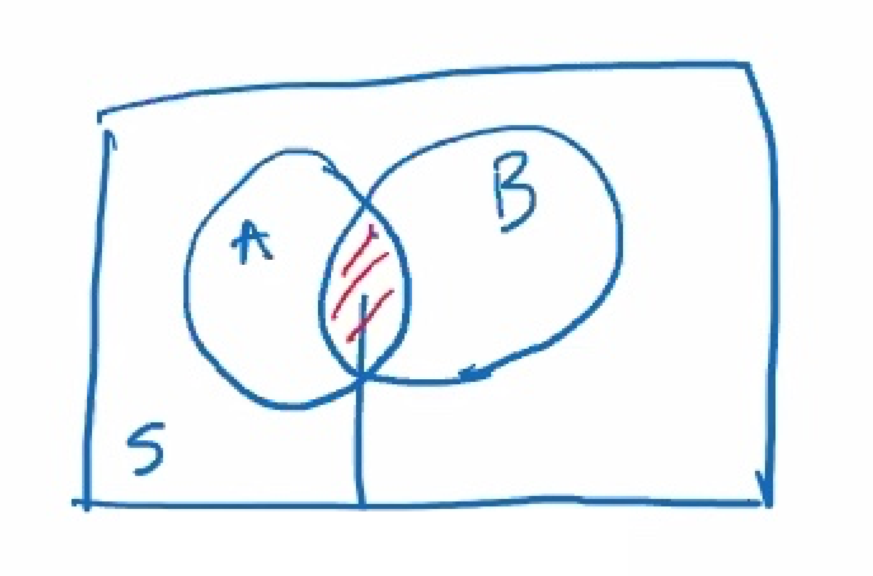
$$\begin{aligned} P(\text{spade or face}) &= \frac{13 + (12 - 3)}{52} && (1) \\ &= \frac{22}{52} && (2) \\ &= .42 && (3) \end{aligned}$$

Can't simply add the previous probability because they are not mutually exclusive events!

4.1 General addition rule

Let A and B be any events, then:

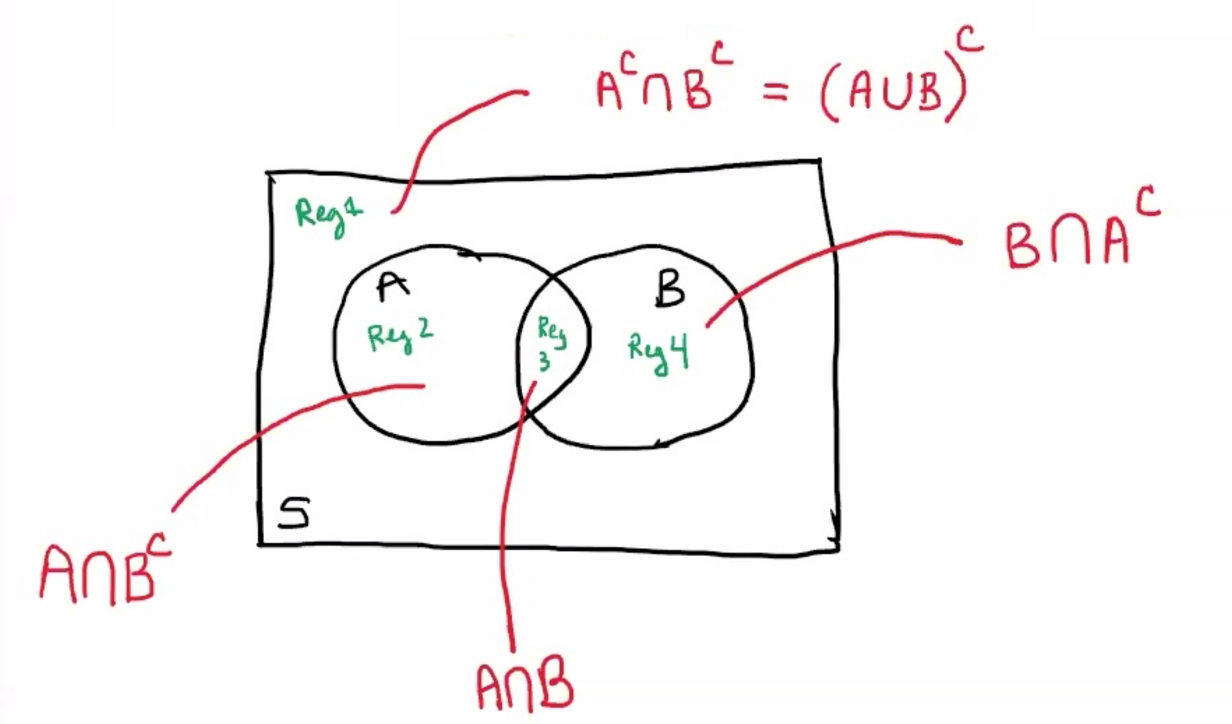
$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$



We subtract $P(A \cap B)$, so we don't double count the overlap.

This can be used when events are not mutually exclusive.

Mini-proof:



$$A \cup B = (A \cap B^C) \cup (A \cap B) \cup (B \cap A^C)$$
$$P(A \cup B) = \underbrace{P(A \cap B^C)}_{P(a)} + \underbrace{P(A \cap B) + P(B \cap A^C)}_{P(B)-P(A \cap B)}$$

A special takeaway from this:

$$A = \underbrace{(A \cap B) \cup (A \cap B^C)}_{\text{mutually exclusive}}$$

Which means

$$P(A) = P(A \cap B) + P(A \cap B^C)$$

and this fact is often used.

Example 9. In a process that manufactures aluminum cans, the probability that a can has a flaw on its side is 0.02, the probability that a can has a flaw on the top is 0.03, and the probability that a can has a flaw on both the top and the side is 0.01

$$P(\text{flaw on side}) = 0.02 \qquad P(\text{flaw on top}) = 0.03 \qquad P(\text{flaw on top AND flaw on side}) = 0.01$$

(a) What is the probability that a randomly chosen can has a flaw?

$$\begin{aligned} P(\text{flaw}) &= P(\text{flaw on top AND flaw on side}) \\ &= P(\text{flaw on side}) + P(\text{flaw on top}) - P(\text{flaw on top AND flaw on side}) \\ &= 0.02 + 0.03 - 0.01 \\ &= 0.04 \end{aligned}$$

(b) What is the probability that is has no flaw?

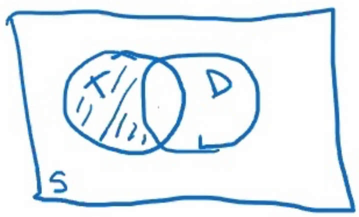
$$\begin{aligned} P(\text{no flaw}) = 1 - P(\text{flaw}) \\ &= 1 - 0.4 \\ &= 0.96 \end{aligned}$$

(c) What is the probability that a can has a flaw on the top but not on the side?

Aside:

T = flaw on top

D = flaw on side

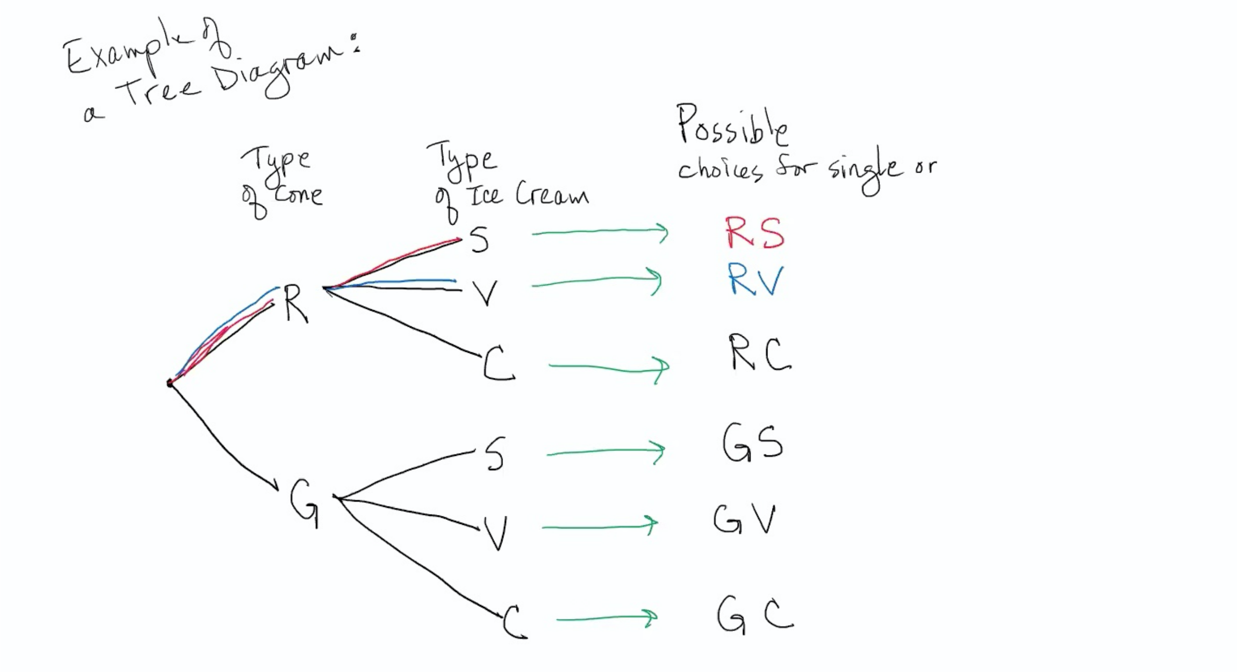


$$P(T \cap D^c) = P(T) - P(T \cap D)$$

$$\begin{aligned}
P(\text{flaw on top AND no flaw on side}) &= P(\text{flaw on top}) - P(\text{flaw on top AND flaw on side}) \\
&= 0.03 - 0.01 \\
&= 0.02
\end{aligned}$$

5 Counting methods

A small local ice cream parlor offers 2 different choices of cones (regular and sugar cone), and 3 different flavors of ice cream (strawberry, vanilla, and chocolate). How many choices are there for a single (cone and one scoop of ice cream) order?



$(2)\ (3) = 6$ possible choices

example of a product table

choices of ice cream

choices of cones	strawberry (B)	vanilla (V)	chocolate (C)
regular (R)	RB	RV	RC
sugar cone (G)	GB	GV	GC

There are $(2)(3) = 6$ choices for a single order.

What if there are also 4 choices of toppings?

$$2 \cdot 3 \cdot 4 = 24$$

5.1 Fundamental principle of counting

Definition 1. If an operation can be performed in n_1 ways, and if for each of these ways a second operation can be performed in n_2 ways, then the total number of ways to perform the two operations is $n_1 \cdot n_2$.

In general:

Assume that k operations are to be performed. If there are n_1 ways to perform the second operation, and if for each of these ways there are n_2 ways to perform the second operation, and if for each choice of ways to perform the first two operations there are n_3 ways to perform the third operation, and so on, then the total number of ways to perform the sequence of k operations is:

$$n_1 \cdot n_2 \cdot n_3 \cdot \dots \cdot n_k$$

Example 10. When ordering a certain type of computer, there are 3 choices of hard drives, 4 choices of amount of memory, 2 choices of video card, and 3 choices of monitor. How many different computer configurations can be ordered?

$$3 \cdot 4 \cdot 2 \cdot 3 = 72$$

Example 11. A club M has 5 members. Suppose club $M = \{$Arthur, Bo, Caitlyn, David, Eva$\}$ (abbreviated $M = \{A, B, C, D, E\}$)

In how many ways can club M elect a president and a secretary, if no one may hold more than one office, and the secretary must be a man.

Version 1

Operations: ① Select a President ② Select a secretary

Number of Ways: $\boxed{} \times \boxed{} =$

Scratch work

A, B, C, D, E A, B, D

$\boxed{} \times \boxed{} =$

A, B, C, D, E A, B, D

$\boxed{} \times \boxed{} =$

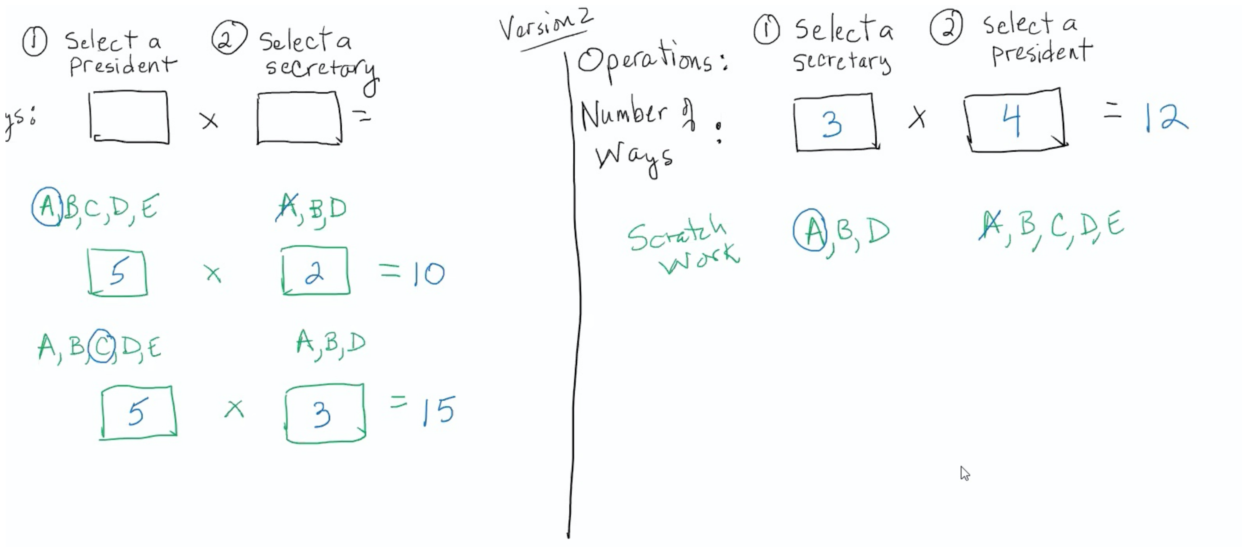
Version 2

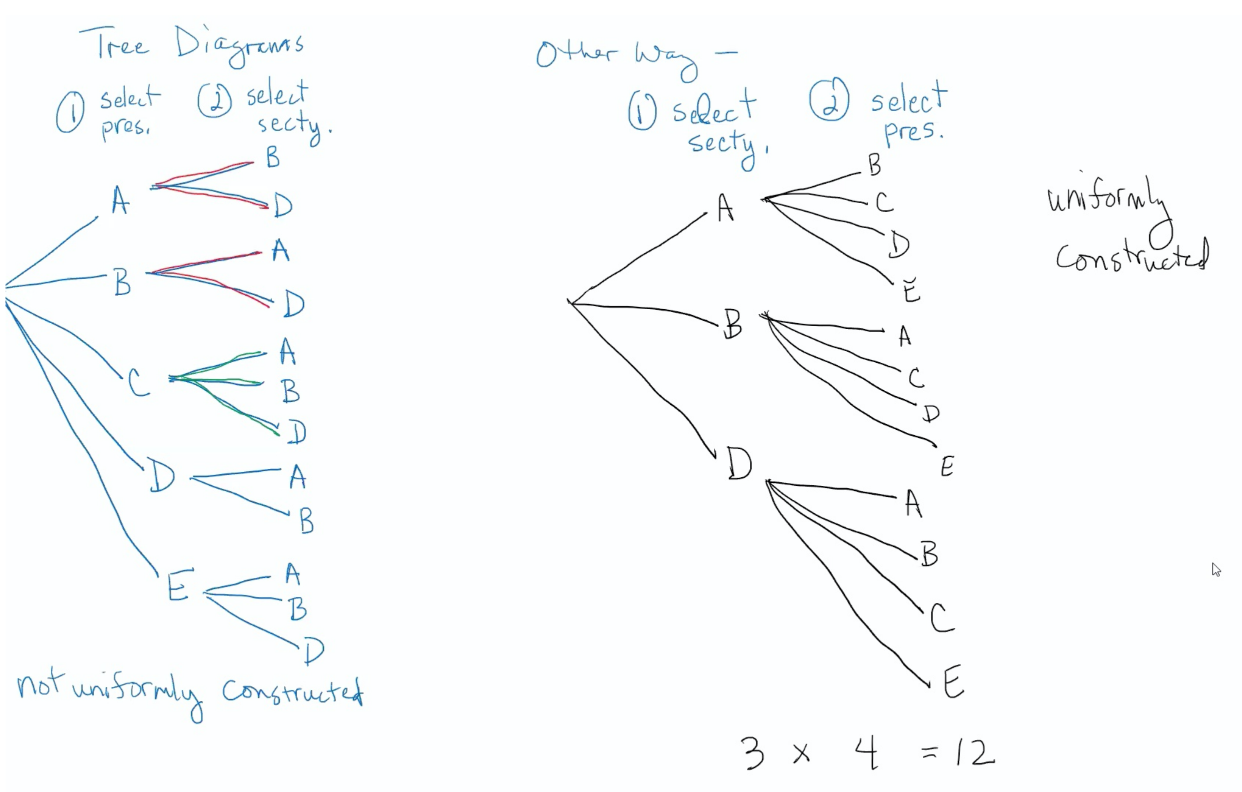
Operations: ① Select a secretary ② Select a president

Number of Ways: $\boxed{} \times \boxed{} =$

Scratch work

A, B, D A, B, C, D, E





As a rule of thumb: count the most restricted parts first. (Start with the subset first).

Example 12. In some states, auto license plates have contained three letters followed by 3 digits. How many such license plates are possible?

$$26^3 \cdot 10^3 = 17,576,000$$

How many such license plates end in 555?

$$26^3 \cdot 1^3 = 17,576$$

5.2 Permutations

A permutation is an ordering of a collection of objects.

Example 13. There are six permutations of the letters A, B, C:

ABC, ACB, BAC, BCA, CAB, CBA

The definition of $n!$ "n factorial": The product of the first n positive integers and is denoted by $n!$. That is for any positive integer $n! = n(n-1)(n-2)\cdots(3)(2)(1)$ and $0! = 1$

Theorem 1. The number of permutations of n objects is $n!$.

Example 14. Let club $M = \{A, B, C, D, E\}$

How many ways can all of the club members arrange themselves in a row for a photo?

$$5! = 120$$

How many ways can the club elect a president, a secretary, and a treasurer if no one can hold more than one office?

$$\frac{5!}{2!} = 60$$

5.2.1 Factorial Formula for Permutations

Definition 2. The number of **permutations**, or *ordered selections (arrangements)*, of r objects chosen from a set of n object is given by:

$$_nP_r = \frac{n!}{(n-r)!}$$

Permutations can be used anytime we need to know the number of ordered selections of r objects that can be selected from a collection of n objects.

We use permutations only in cases that satisfy these conditions:

1. Repetitions are not allowed.
2. Order is important.

$$_5P_3 = \frac{5!}{(5-3)!} = \frac{5!}{2!} = 60$$

Example 15. An ATM requires a four digit PIN using the digits 0-9 (the first digit may be 0). How many such PINs have no repeated digits?

$$_{10}P_4 = \frac{10!}{6!} = 5040$$

5.3 Combinations

A combination is an unordered selection (subset).

Example 16. Let $M = \{A, B, C, D, E\}$

A committee of 3 members would be an unordered selection (a combination), for example: $\{A, B, C\}$. Order does not matter!

5.3.1 Factorial Formula for Combinations

Definition 3. The number of **combinations**, or *unordered selections (subsets)*, of r objects chosen from a set of n objects is given by:

$$_nC_r = \frac{_nP_r}{r!} = \frac{n!}{r!(n-r)!}$$

Another commonly used notation for combinations is:

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

Example 17. Consider club M = Arthur, Bo, Caitlyn, David, Eva = $\{A, B, C, D, E\}$. A committee of three members would be an example of a combination, or unordered selection. For example one such comittee might consist of $\{A, B, C\}$. Here, order does not matter as it is just a group of 3 people in a comittee.

Use the combinations formula to determine the number of different size 3 committees (subsets) that could possibly be made from club M.

$$\binom{5}{3} = \frac{5!}{3!(5-3)!} = 10$$

Each group of 3 letters corresponds to $3! = 3 \cdot 2 \cdot 1 = 6$ permutations

and so on

ABC	ABD	ABE	ACD	ACE	ADE	BCD	BCE	BDE	CDE
ACB	ADB	AEB	ADC	AEC	AED	BDC	BEC	BED	CED
BAC	BAD	BAE	CAD	CAE	DAE	CBD	CBE	DBE	DCE
BCA	BDA	BEA	CDA	CEA	DEA	CDB	CEB	DEB	DEC
CAB	DAB	EAB	DAC	EAC	EAD	DBC	EBC	EBD	ECD
CBA	DBA	EBA	DCA	ECA	EDA	DCB	ECB	EDB	EDC

FIGURE 2.4 The 60 permutations of three objects chosen from five.

Total number of perms of 3 objects from as set of 5 is $_5P_3 = 60$

(60 total permutations) $\div$ (6 permutations per group of 3)

$$\frac{60 \text{ perms}}{6 \text{ perms/group of 3}} = 10 \text{ groups of } 3$$

Example 18. How many different ways are there to select a jury of 12 people from a pool of 20?

$$\binom{20}{12} = 125,970$$

Example 19. Assume that in a class of 12 students, a project is assigned in which the students will work in groups. 3 groups are formed, consisting of 5, 4, and 3 students. Find the number of ways in which the groups can be formed.

$$\binom{12}{5}\binom{7}{4}\binom{3}{3} = 27,720$$

6 Conditional Probability

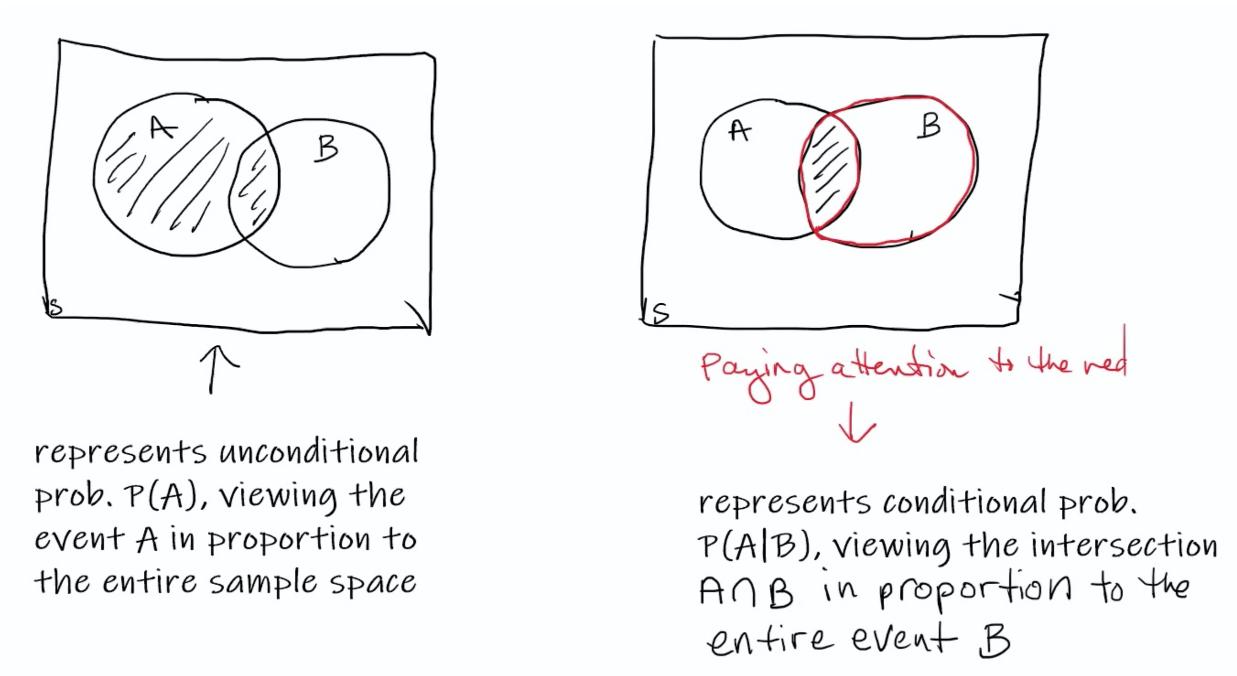
The probability that event B occurs given that event A occurs is called **conditional probability**. This is denoted:

$$P(B \mid A)$$

A conditional probability is a probability that is based on a **part** of the sample space. (An unconditional probability would be based on the entire sample space.)

Let A and B be events with $P(B) \neq 0$. The conditional probability of A given B is:

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



Example 20. In a genetics experiment, the researcher mated two *Drosophila* fruit flies and observed the traits of 300 offspring. The results are shown in the table below.

One of these offspring is randomly selected and observed for the two genetic traits. Determine the follow probabilities. Round your answer to three decimal places.

1. What is the probability that a fly has both normal eye color and normal wing size?

$$P(\text{normal eye and normal wing}) = \frac{140}{300} = .467$$

2. What is the probability that the fly has vermillion eyes?

$$P(\text{vermillion eyes}) = \frac{154}{300} = .513$$

3. What is the probability that the fly has normal eye color, given the fly has miniature wings?

$$P(\text{normal eye — mini wings}) = \frac{6}{157} = .038$$

4. What is the probability that the fly has either vermillion wings or miniature wings (or both)?

$$P(\text{vermillion eyes or mini wings}) = \frac{151 + 3 + 6}{300} = .533$$

Example 21. In a process that manufactures aluminum cans, the probability that a can has a flaw on its side is 0.02, the probability that a can has a flaw on top is 0.03, and the probability that a can has a flaw on both the side and the top is 0.01.

1. What is the probability that a can will have a flaw on the side, given that it has a flaw on top?

$$\begin{aligned} P(\text{flaw on side} \mid \text{flaw on top}) &= \frac{P(\text{flaw on side and flaw on top})}{P(\text{flaw on top})} \\ &= \frac{.01}{.03} \approx .333 \end{aligned}$$

2. What is the probability that a can will have a flaw on top, given that it has a flaw on the side?

$$P(\text{flaw on top} \mid \text{flaw on side}) = \frac{.01}{.02} = .5$$

6.1 Independent Events

Definition 4. Two events A and B are **independent** if the probability of each event remains the same whether or not the other occurs.

If $P(A) \neq 0$ and $P(B) \neq 0$ then A and B are independent if:

$$P(B \mid A) = P(B) \qquad P(A \mid B) = P(A)$$

If either $P(A) = 0$, or $P(B) = 0$, then A and B are independent.

Note 2. If A and B are independent, then the follow pairs of events are also independent: A and B^C, A^C and B, and A^C and B^C.

The concept of independence can be extended to more than two events:

Events $A_1, A_2, \ldots, A_n$ are independent if the probability of each remains the same no matter which of the others occur.

$$P(A_i \mid A_{i1} \cap \cdots \cap A_{in}) = P(A_i)$$

Example 22. Consider the experiment of randomly selecting one card from a deck of 52 playing cards. Let:

$$
\begin{aligned}
F &= \text{event a face card is selected} \\
K &= \text{event a king is selected} \\
H &= \text{event a heart is selected}
\end{aligned}
$$

$$
\begin{aligned}
P(K) &= \frac{4}{52} \approx 0.077 \\
P(K \mid F) &= \frac{4}{12} \approx 0.333 \\
P(K \mid F) &\neq P(K)
\end{aligned}
$$

So, K is not independent of F. The percentage of Kings among face cards (33.3%) is not the same as the percentage of Kings among all cards (7.7%).

$$P(K \mid H) = \frac{1}{13} \approx 0.077$$

Note 3.

$$P(K \mid H) = P(K)$$

So, K is independent of the event H. The percentage of Kings among the hearts is the same percentage of Kings among all cards.

6.2 Multiplication Rule

Definition 5. If A and B are two events with $P(B) \neq 0$, then $P(A \cap B) = P(B) \cdot P(A \mid B)$.

If A and B are two events with $P(A) \neq 0$, then $P(A \cap B) = P(A) \cdot P(B \mid A)$.

If $P(A) \neq 0$ and $P(B \neq 0)$, then both above equations hold.

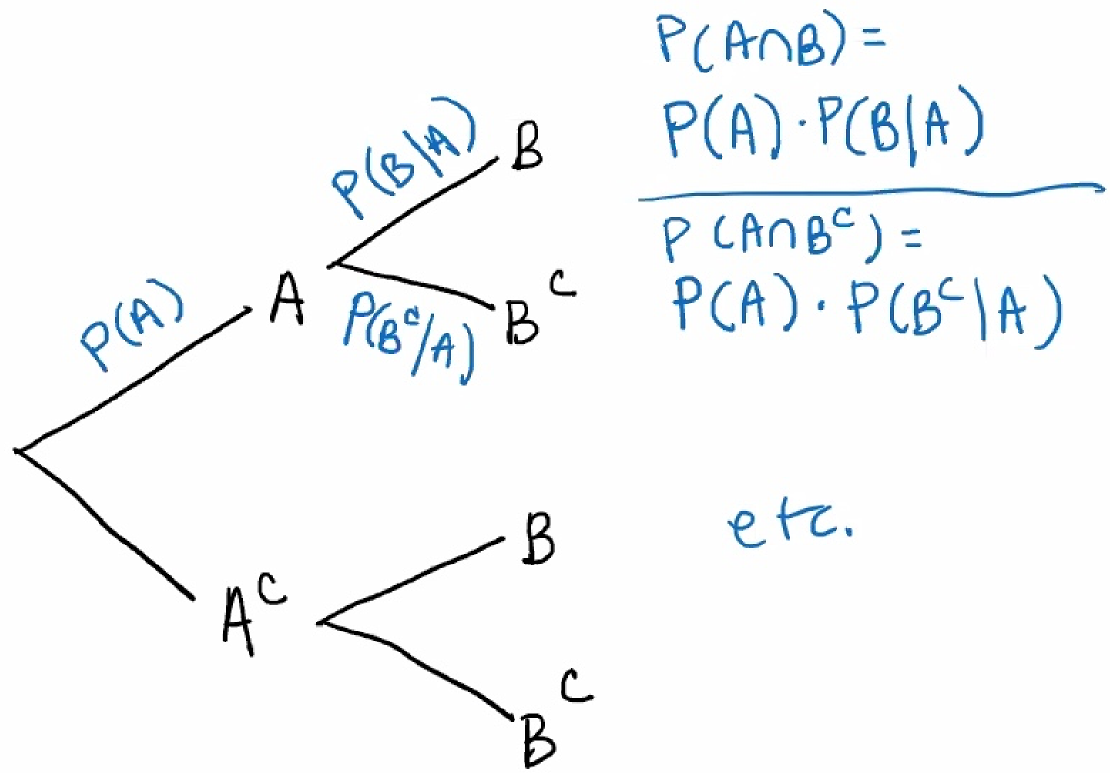
When two events are independent, then $P(A \mid B) = P(A)$ and $P(B \mid A) = P(B)$ and we have the following:

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

This result can be extended to any number of events. If $A_1, A_2, \ldots, A_n$ are independent events, then for each collection $A_{i1}, \ldots, A_{in}$ of events:

$$P(A_{i1} \cap A_{i2} \cap \cdots \cap A_{in}) = P(A_{i1})P(A_{i2}) \cdots P(A_{in})$$

Remember, $A \cap B = B \cap A$ and $A \cup B = B \cup A$.



$$
\begin{aligned}
P(B^C \mid A) &= \frac{P(A \cap B^C)}{P(A)} \\
&= 1 - P(B \mid A)
\end{aligned}
$$

Mutually exclusive events are not usually independent, we are usually working with events that have non-zero probabilities.

Example 23. Imagine an urn with 7 balls, 3 yellow and 4 green. 2 balls are selected in succession, without replacement.

Find the probability of selecting a yellow ball first, and a green ball second.

$$
\begin{aligned}
P(Y_1 \cap G_2) &= P(Y_1) \cdot P(G_2 \mid Y_1) \\
&= \frac{3}{7} \cdot \frac{4}{6} \\
&= \frac{12}{42}
\end{aligned}
$$

Find the probability of selecting a yellow ball first, a green ball second, and a green ball third.

$$\begin{aligned} P(Y_1 \cap G_2 \cap G_3) &= P(Y_1) \cdot P(G_2 \mid Y_1) \cdot P(G_3 \mid Y_1 \cap G_2) \\ &= \frac{3}{7} \cdot \frac{4}{6} \cdot \frac{3}{5} \\ &= \frac{36}{210} \end{aligned}$$

Now suppose three balls are selected in succession, with replacement. Find the probability of selecting a yellow ball first, a green ball second, and a green ball third.

$$\begin{aligned} P(Y_1 \cap G_2 \cap G_3) &= P(Y) \cdot P(G) \cdot P(G) \\ &= \frac{3}{7} \cdot \frac{4}{7} \cdot \frac{4}{7} \end{aligned}$$

Example 24. A vehicle contains two engines, a main engine and a backup. The engine component fails only if both engines fail. The probabilty that the main engine fails is 0.05, and the probability that the backup engine fails is 0.10. Assume that the main and backup engines function independently. What is the probability that the engine component fails?

$$\begin{aligned} P(\text{failure}) &= P(\text{main failure}) \cdot P(\text{backup failure}) \\ &= 0.05 \cdot 0.10 \\ &= 0.005 \end{aligned}$$

Example 25. A system contains two components, A and B. Both components must function for the system to work. The probability that component A fails is 0.08, and the probability that component B fails is 0.05. Assume the two components function independently. What is the probabilty that the system functions?

$$\begin{aligned} P(\text{functions}) &= P(\text{A functions}) \cdot P(\text{B functions}) \\ &= [1 - P(\text{A functions}^C)] \cdot [1 - P(\text{B functions}^C)] \\ &= (1 - .08) \cdot (1 - .05) \\ &= .874 \end{aligned}$$

Example 26. Tests are performed in which structures consisting of concrete columns welded to steel beams are loaded until failure. 80% of failurs occur in the beam, while 20% occur in the weld. Five structures are tested.

What is the probability that all five failures occur in the beam?

$$0.8^5 = 0.327$$

What is the probability that at least one failure occurs in the beam?

$$1 - 0.2^5 = .99968$$

6.3 Law of Total Probability

Definition 6. If $A_1, A_2, \ldots, A_n$ are mutually exclusive and exhaustive events, and B is any event, then

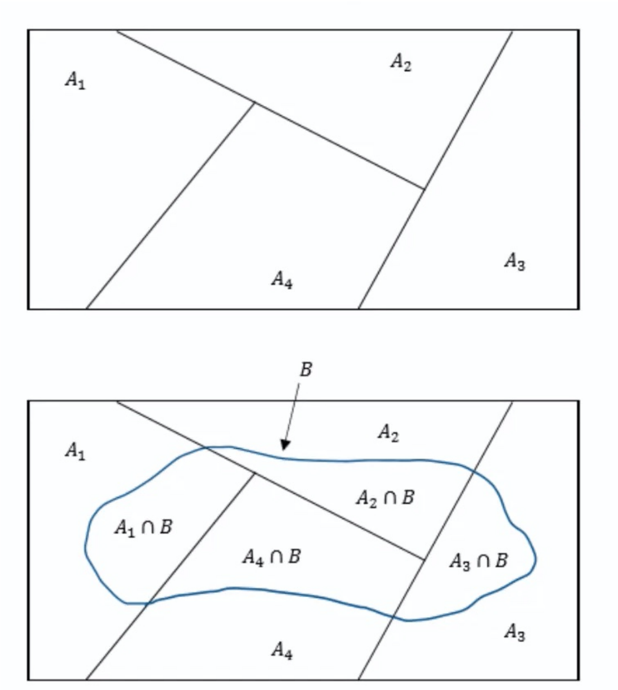
$$P(B) = P(A_1 \cap B) + P(A_2 \cap B) + \cdots + P(A_n \cap B)$$

Equivalently, if $P(A_i) \neq 0$, for each A_i,

$$P(B) = P(B \mid A_1)P(A_1) + P(B \mid A_2)P(A_2) + \cdots + P(B \mid A_n)$$

Exhaustive means that the union of all of the events covers the entire sample space.

In the figure below, the mutually exhaustive events A_1, A_2, A_3, A_4 divide event B into mutually exclusive subsets.



6.4 Bayes' Rule

Definition 7. **Special case**: Let A and B be events with $P(A) \neq 0$, $P(A^C) \neq 0$, and $P(B) \neq 0$, then

$$P(A \mid B) = \frac{P(B \mid A)P(A)}{P(B \mid A)P(A) + P(B \mid A^C)P(A^C)}$$

Bayes' rule provides a formula that allows us to calculate one of the conditional probabilities, if we know the other one.

General case: Let $A_1, A_2, \ldots, A_n$ be mutually exclusive events and exhaustive events with $P(A_i) \neq 0$ for each A_i. Let B be any event with $P(B) \neq 0$.

$$\begin{aligned} P(A_k \mid B) &= \frac{P(B \mid A_k)P(A_k)}{P(B \mid A_1)P(A_1) + \cdots + P(B \mid A_n)P(A_n)} \\ &= \frac{P(B \mid A_k)P(A_k)}{\sum_{i=1}^{n} P(B \mid A_i)P(A_i)} \end{aligned}$$

Note 4. In general,

$$P(A \mid B) \neq P(B \mid A)$$

Example 27. During frequent trips to a certain city, a traveling salesman stays at hotel A 50% of the time, hotel B 30% of the time, and hotel C 20% of the time. When checking in, there is some problem with the reservation 3% of the time at hotel A, 6% of the time at hotel B, and 10% of the time at hotel C. Suppose the salesperson travels to this city.

Let

$$\begin{aligned} A &= \text{stays at hotel A} \\ B &= \text{stays at hotel B} \\ C &= \text{stays at hotel C} \\ R &= \text{problem with res} \end{aligned}$$

and

$$\begin{aligned} P(A) &= 0.5 \\ P(B) &= 0.3 \\ P(C) &= 0.2 \\ P(R \mid A) &= 0.03 \\ P(R \mid B) &= 0.06 \\ P(R \mid C) &= 0.1 \end{aligned}$$

Find the probability that the salesperson stays at hotel A and has a problem.

$$\begin{aligned} P(A \cap R) &= P(R \mid A) \cdot P(A) \\ &= 0.03 \cdot 0.5 \\ &= 0.015 \end{aligned}$$

Find the probability that the salesperson has a problem with the reservation.

$$\begin{aligned} P(R) &= (0.03 \cdot 0.5) + (0.06 \cdot 0.3) + (0.1 \cdot 0.2) \\ &= 0.053 \end{aligned}$$

Suppose the salesperson has a problem with the reservation, what is the probability that the salesperson is staying at hotel A?

We can use the values from the last 2 problems:

$$\begin{aligned} P(A \mid R) &= \frac{P(A \cap R)}{P(R)} \\ &= \frac{0.015}{0.053} \\ &= 0.283 \end{aligned}$$

7 Random Variables

A **random variable** assigns a numerical value to each outcome in a sample space (typically denoted by capital letters, like X, Y, Z). We can think of a random variable X as a function.

Example 28. When a balanced coin is tossed three times, 8 equally likely outcomes are possible: HHH HHT HTH HTT THH THT TTH TTT.

Let X denote the total number of heads obtained in 3 tosses, then X is a random variable where

$$X = \{0, 1, 2, 3\}$$

Two important types of random variables:

- Discrete
  - One whose possible values form a discrete set; that is values can be ordered and there are gaps between adjacent values.
  - Set of integers.
  - Set of whole numbers.
  - Usually invole a 'count' of something.
- Continuous
  - Possible values always contain an interval; that is all the points between some two numbers.
  - Usually involves a 'measure' of something.

7.1 Discrete Random Variables

Definition 8. The **probability mass function** (pmf) of a discrete random variable X is the function

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

The probability mass function is sometimes called the probability distribution.

Note 5. If the values of the probability mass function are added over all the possible values of X, the sum equals 1. That is,

$$\sum_x p(x) = \sum_x P(X = x) = 1$$

where the sum is over all the possible values of X.

Example 29. The number of flaws in a 1-inch length of copper wire manufactured by a certain process varies from wire to wire. Overall, 48% of the wires produced have no flaws, 39% have 1 flaw, 12% have 2 flaws, and 1% have 3 flaws. Let X be the number of flaws in a randomly selected piece of wire.

The possible values of X are:

$$X = \{0, 1, 2, 3\}$$

The pmf is a table showing the possible values of X and their corresponding probabilities.

Definition 9. The **cumulative distribution function** (cdf) of a random variable X is the function

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

In general, for any discrete random variable X, the cumulative distribution function $F(x)$ can be computed by summing the probabilities of all the possible values of X that are less than or equal to x. That is,

$$F(x) = \sum_{t \leq x} p(t) = \sum_{t \leq x} p(X = t)$$

Note 6. $F(x)$ is ddefined for any number x, not just for possible values of X. For a discrete randomy variable X, the graph of F will be a "step function".

Example 30. The given values can be turned into a cdf.

$$\begin{aligned}
F(0) &= P(X \leq 0) = p(0) = .48 \\
F(1) &= P(X \leq 1) = p(0) + p(1) = .87 \\
F(2) &= P(X \leq 2) = .48 + .39 + .12 = .99 \\
F(3) &= P(X \leq 3) = 1
\end{aligned}$$

If you want to calculate numbers in between the given:

$$\begin{aligned}
F(-1.4) &= P(X \leq -1.4) = 0 \\
F(0.5) &= P(X \leq 0.5) = .48 \\
F(.99) &= P(X \leq .99) = .48 \\
F(1.3) &= P(X \leq 1.3) = .87 \\
F(4) &= P(X \leq 4) = 1
\end{aligned}$$

There is only a discrete amount of values for X so you must check what the probability X is less than what youre checking.

The cdf is a piecewise defined function:

$$F(x) = \begin{cases} 0 & x < 0 \\ 0.48 & 0 \leq x \leq 1 \\ 0.87 & 1 < x \leq 2 \\ 0.99 & 2 < x \leq 3 \\ 1 & x \geq 3 \end{cases}$$

7.1.1 Mean of a for Discrete Random Variables

Definition 10. Let X be a discrete random variable with probability mass function $p(x) = P(X = x)$. The **mean** of X is given by

$$\mu_X = \sum_X x \cdot P(X = x)$$

where the sum is over all possible values of X.

The mean of X is sometimes called the expectation, or expected value, of X and may also be denoted by $E(X)$. Sometimes we leave off the subscript and write the sum as

$$\mu = \sum_X xp(x)$$

Mean: for a small finite data set, the mean would simply be an arithmetic average. The mean of a discrete random variable is a generalization of this idea and represents a **weighted** average.

Example 31. Consider a population of 8 students with the following ages in years: 19, 20, 20, 19, 21, 27, 20, 21

Say:

$$X = \text{age of randomly selected student}$$

Arithmetic average would be:

$$\begin{aligned} &= \frac{19 + 20 + 20 + 19 + 21 + 27 + 20 + 21}{8} \\ &= 20.875 \end{aligned}$$

But in a weighted average:

$$\begin{aligned} &= \frac{19(2) + 20(3) + 21(2) + 27(1)}{8} \\ &= 19(.25) + 20(.375) + 21(.25) + 27(.125) \\ &= \sum_x xp(x) = \mu \\ \mu &= \sum_x xp(x) \\ &= 19(.25) + 20(.375) + 21(.25) + 27(.125) \\ &= 20.875 \end{aligned}$$

7.1.2 Variance and Standard Deviation of a Discrete Random Variable

Definition 11. Let X be a discrete random variable with a probability mass function $p(x) = P(X = x)$.

- The **variance** of X is given by

$$\sigma_X^2 = \sum_X (x-\mu)^2 P(X = x)$$

- An alternate formula for the variance is given by

$$\sigma_X^2 = \sum_X x^2 P(X = x) - \mu_X^2$$

- The variance of X may also be denoted by $V(x)$ or by σ^2.
- The **standard deviation** is the square root of the variance

$$\sigma_X = \sqrt{\sigma_X^2}$$

The **variance** can also be written as

$$\sigma^2 = \sum_x \left[(x-\mu)^2 \cdot p(x)\right]$$

The alternate formula for the variance is known as the **computing formula**.

The **mean** is a measure of center. The **variance** and **standard deviation** are measures that indicate how far, on average, an observed value of X is from the mean.

Example 32. Consider a population of 8 students with the following ages in years: 19, 20, 20, 19, 21, 27, 20, 21

Say:

$$X = \text{age of randomly selected student}$$

$x - \mu$ is the **deviation from mean**. $(x-\mu)^2$ is the **squared deviation**.

$$\begin{aligned}
\sigma^2 &= \sum_x (x-\mu)^2 \cdot p(x) \\
&= (-1.875)^2(.25) + (-.875)^2(.375) + (.125)^2(.25) + (6.125)^2(.125) \\
&= 5.859375 \\
&\approx 5.859 \text{ years}^2
\end{aligned}$$

The variance is the "mean of the squared deviations".

$$\begin{aligned}
\sigma &= \sqrt{\sigma^2} \\
&= \sqrt{5.859375} \\
&\approx 2.421\,\text{years}
\end{aligned}$$

In words: Observations are on average roughtly 2.421 years from the mean.

Example 33. Using the previous example, lets use the **computing formula**.

$$\sigma_x^2 = \sum_x \left[x^2 p(x)\right] - \mu_x^2$$

Recall

$$\mu = 20.875 \text{ years}$$

So using the computing formula:

$$V(x) = \sigma^2 = \sum_x x^2 p(x) - \mu^2$$

$$\begin{aligned}
\sigma^2 &= \sum_x x^2 p(x) - \mu^2 \\
&= 19^2(.25) + 20^2(.375) + 21^2(.25) + 27^2(.125) - \mu^2 \\
&= 441.625 - (20.875)^2 \\
&= 5.859375 \\
&\approx 5.859 \\
\sigma &\approx 2.421
\end{aligned}$$

7.2 Probability Density Functions

Definition 12. A random variable is **continuous** if its probabilities are given by areas under a curve. This curve is called a **probability density function** (pmf) for the random variable. The probability density function is sometimes called the probability distribution.

Let X be a continuous random variable with probability density function $f(x)$. Let a and b be any two numbers, with $a < b$.

The proportion of the population whose values of X lie between a and b are given by $\int_a^b f(x)\,\mathrm{d}x$, the area under the probability density function between a and b. This is the probability that the random variable X takes on a value between a and b. Note that the area under the curve does not depend on whether the endpoints a and b are included in the interval. So, probabilities involving X do not depend on whether endpoints are included. That is,

$$\begin{aligned}
P(a \leq X \leq B) &= P(a \leq X < b) \\
&= P(a < X \leq b) \\
&= P(a < X < b) \\
&= \int_a^b f(x)\,\mathrm{d}x
\end{aligned}$$

In additions,

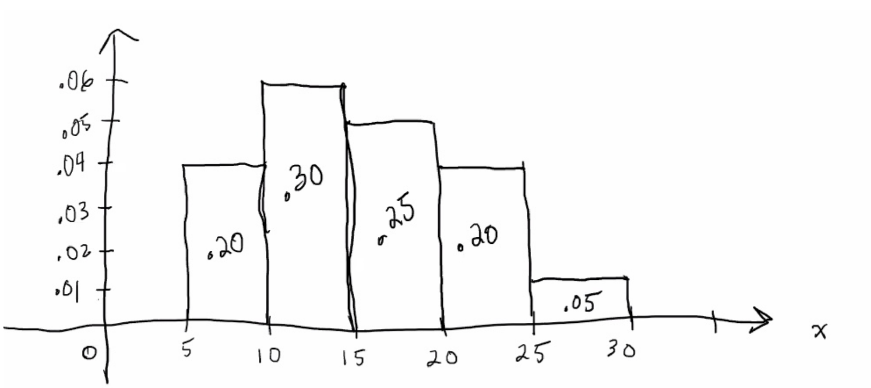
$$
\begin{aligned}
P(X \leq b) &= P(X < b) \\
&= \int_{-\infty}^{b} f(x)\,\mathrm{d}x \\
P(X \leq a) &= P(X > a) \\
&= \int_{a}^{\infty} f(x)\,\mathrm{d}x
\end{aligned}
$$

Always find the probability that X is in a range, never for a single value.

Note also, for $f(x)$ to be a legitimate pdf, it must satisfy the following conditions:

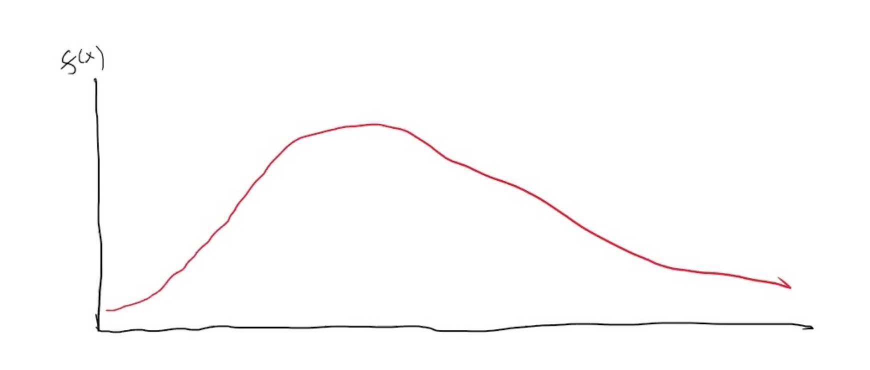
1. $f(x) \geq 0$ for all x.
2. $\int_{-\infty}^{\infty} f(x)\,\mathrm{d}x = 1$. The area under the curve is equal to 1.

Some background: In a probability histogram, areas of rectangles correspond to probabilities. For example with a sample of continuous data, data can be organized into categories, or bins.



The areas of these rectangles correspond to probabilities and let us know the proportion of values of x that fall into a certain category. The sum of the areas of the rectangles must be 1. Note: the vertical axis is not the probability, the area is.

The probability histogram for a large continuous population could be drawn with extremely narrow rectangles and might look like this curve:



Example 34. A hole is drilled in a sheet-metal component, and then a shaft is inserted through the hole. The shaft clearance is equal to the different between the radius of the hole and radius of the shaft. Let the random variable X denote the clearance, in millimeters. The probability density function of X

is

$$f(x) = \begin{cases} 1.25(1 - x^4) & 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

Components with clearances larger than 0.8 mm must be scrapped. What proportion of components are scrapped?

$$\begin{aligned} P(X > 0.8) &= \int_a^\infty f(x)\,\mathrm{d}x \\ &= \lim_{t\to\infty} \int_a^t f(x)\,\mathrm{d}x \\ &= \int_{0.8}^1 1.25(1 - x^4)\,\mathrm{d}x + \lim_{t\to\infty} \int_a^t 0\,\mathrm{d}x \\ &= 1.25 \int_{0.8}^1 1 - x^4\,\mathrm{d}x \\ &= 1.25 \left[x - \frac{1}{5}x^5\right]_{0.8}^1 \\ &= 1.25 \left[(1 - \frac{1}{5}) - (0.8 - \frac{.8^5}{5})\right] \\ &\approx 0.0819 \end{aligned}$$

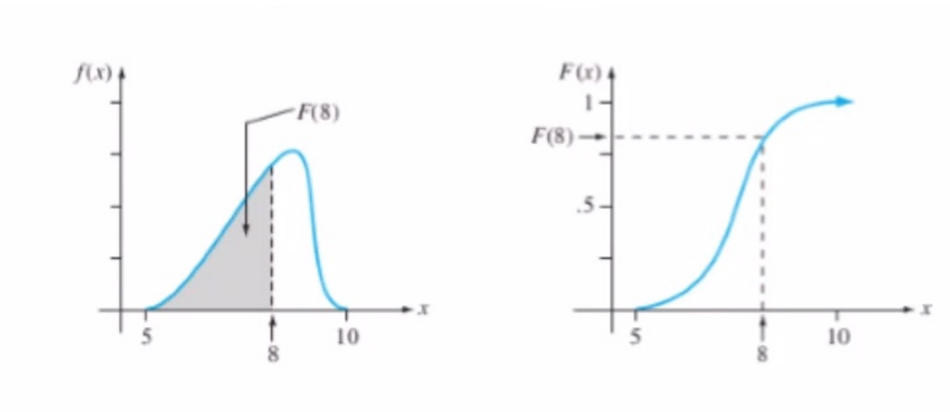
7.3 Cumulative Distribution Functions of a Continuous Random Variable

Definition 13. Let X be a continuous randome variable with probability density function $f(x)$. The **cumulative distribution function** (cdf) of X is the function

$$F(x) = P(X \leq x) = \int_{-\infty}^{\infty} f(t)\,\mathrm{d}t$$

For example, for each x, $F(x)$ is the area under the density curve to the left of x.

Here is an illustration of a pdf and its associated cdf:



Example 35. A hole is drilled in a sheet-metal component, and then a shaft is inserted through the hole. The shaft clearance is equal to the different between the radius of the hole and radius of the shaft. Let the random variable X denote the clearance, in millimeters. The probability density function of X is

$$f(x) = \begin{cases} 1.25(1-x^4) & 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

Find the cumulative distribution function of $F(x)$.

We know:

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)\,\mathrm{d}t$$

Case 1: $x \leq 0$

$$\begin{aligned} F(x) &= \int_{-\infty}^{x} 0\,\mathrm{d}t \\ &= 0 \end{aligned}$$

Case 2: $0 < x < 1$

$$\begin{aligned} F(x) &= \int_{-\infty}^{x} f(t)\,\mathrm{d}t \\ &= \int_{-\infty}^{0} f(t)\,\mathrm{d}t + \int_{0}^{x} f(t)\,\mathrm{d}t \\ &= 0 + \int_{0}^{x} 1.25(1-t^4)\,\mathrm{d}t \\ &= 1.25\int_{0}^{x} 1-t^4\,\mathrm{d}t \\ &= 1.25\left[t - \frac{1}{5}t^5\right]_0^x \\ &= 1.25\left[(x - \frac{1}{5}x^5) - (0-0)\right] \\ &= 1.25\left(x - \frac{1}{5}x^5\right) \end{aligned}$$

Case 3: $x \geq 1$

$$
\begin{aligned}
F(x) &= \int_{-\infty}^{x} f(t)\,\mathrm{d}t \\
&= \int_{-\infty}^{0} f(t)\,\mathrm{d}t + \int_{0}^{1} f(t)\,\mathrm{d}t + \int_{1}^{x} f(t)\,\mathrm{d}t \\
&= 0 + \int_{0}^{1} 1.25(1-t^4)\,\mathrm{d}t + 0 \\
&= 1
\end{aligned}
$$

The area under the entire cdf should be 1 by definition.

We can define our cdf as:

$$
F(x) = \begin{cases} 0 & x \leq 0 \\ 1.25(x - \frac{1}{5}x^5) & 0 < x < 1 \\ 1 & x \geq 1 \end{cases}
$$

7.4 Mean and Variance for Continuous Random Variables

Definition 14. Let X be a continuous random variable with probability density function $f(x)$. Then the mean of X is given by

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

The mean of X is sometimes called the expectation, or expected value, of X and may also be denoted by $E(X)$ or μ.

Definition 15. Let X be a continuous random variable with probability density function $f(x)$. Then

- The variance of X is given by

$$\sigma_x^2 = \int_{-\infty}^{\infty} (x0\mu_X)^2 f(x)\, \mathrm{d}x$$

- An alternate formula for the variance is given by

$$\sigma_X^2 = \int_{-\infty}^{\infty} x^2 f(x)\, \mathrm{d}x - \mu_X^2$$

- The variance of X may also be denoted by $V(x)$ or by σ^2.
- The standard deviation is the square root of the variance

$$\sigma_X = \sqrt{\sigma_X^2}$$

Example 36. A hole is drilled in a sheet-metal component, and then a shaft is inserted through the hole. The shaft clearance is equal to the different between the radius of the hole and radius of the shaft. Let the random variable X denote the clearance, in millimeters. The probability density function of X is

$$f(x) = \begin{cases} 1.25(1 - x^4) & 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

Find the mean clearance and the variance of the clearance.

Recall:

$$E(X) = \mu = \int_{-\infty}^{\infty} x f(x)\, \mathrm{d}x$$

So:

$$
\begin{aligned}
\mu &= \int_0^1 x f(x)\,\mathrm{d}x \\
&= 1.25 \int_0^1 x(1-x^4)\,\mathrm{d}x \\
&= 1.25 \int_0^1 x - x^5\,\mathrm{d}x \\
&= 1.25 \left[\frac{1}{2}x^2 - \frac{1}{6}x^6\right]_0^1 \\
&= 1.25 \left[\frac{1}{2} - \frac{1}{6}\right] \\
&\approx 0.4167
\end{aligned}
$$

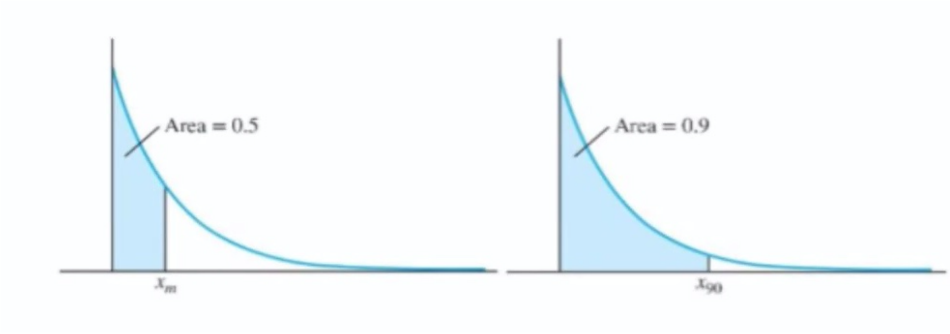
Lets find the variance:

$$
\begin{aligned}
V(x) = \sigma^2 &= \int_{-\infty}^{\infty} x^2 f(x)\,\mathrm{d}x - \mu^2 \\
&= \int_0^1 x^2 f(x)\,\mathrm{d}x - (.4167)^2 \\
&= \int_0^1 x^2 \left[1.25(1-x^4)\right]\,\mathrm{d}x - (.4167)^2 \\
&= 1.25 \int_0^1 x^2 - x^6\,\mathrm{d}x - (.4167)^2 \\
&= 1.25 \left[\frac{1}{3}x^3 - \frac{1}{7}x^7\right]_0^1 - (.4167)^2 \\
&= 1.25 \left(\frac{1}{3} - \frac{1}{7}\right) - (.4167)^2 \\
&\approx .0645\,\mathrm{mm}^2
\end{aligned}
$$

7.5 Median and Percentiles

The **median** of a data set divides the data into two equal parts: the bottom 50% and the top 50%. In terms of probability density function, the median (denoted by x_m) is the point at which half the area under the curve is to the left, and half the area is to the right.

More generall, for p between 0 and 100, the p**th percentile** (denoted by x_p) is the number that divides the bottom $p\%$ of the data from the top $(100-p)\%$. For example, the 90th percentile (denoted by x_{90}) is the number that divides the bottom 90% of the data from the top 10%. In terms of a probability density function, the 90th percentile is the point x_{90} at which the area under the curve to the left of x_{90} is 0.9, and the area under the curve to the right of x_{90} is 0.1.



Definition 16. Let X be a continuous random variable with a probability function $f(x)$ and cumulative distribution function $F(x)$.

- The median of X is the point x_m that solves the equation

$$F(x_m) = P(X \leq x_m) = \int_{-\infty}^{x_m} f(x)\,\mathrm{d}x = 0.5$$

- If p is any number between 0 and 100, the pth percentile is the point x_p that solves the equation

$$F(x_p) = P(X \leq x_p) = \int_{-\infty}^{x_p} f(x)\,\mathrm{d}x = \frac{p}{100}$$

- The median is the 50th percentile.

Example 37. A certain radioactive mass emits alpha particles from time to time. The time between emissions, in seconds, is random with probability density function

$$f(x) = \begin{cases} 0.1e^{-0.1x} & x > 0 \\ 0 & x \leq 0 \end{cases}$$

1. Find the median time between emissions.

   We are looking for x_m such that $F(x_m) = 0.5$.

$$
\begin{aligned}
F(x_m) = P(X \le x_m) &= \int_{-\infty}^{x_m} f(x)\,\mathrm{d}x \\
&= \int_{0}^{x_m} 0.1e^{-0.1x}\,\mathrm{d}x \\
&= 0.1\int_{0}^{x_m} e^{-0.1x}\,\mathrm{d}x \\
&= 0.1\left[\frac{1}{-0.1}e^{-0.1x}\right]_0^{x_m} \\
&= -e^{-0.1x}\,\Big|_0^{x_m} \\
&= -e^{-0.1x_m} - (-e^0) \\
&= -e^{-0.1x_m} + 1
\end{aligned}
$$

Set it equal to our 0.5 because we are looking for the median:

$$
\begin{aligned}
-e^{-0.1x_m} + 1 &= 0.5 \\
e^{-0.1x_m} &= 0.5 \\
\ln(e^{-0.1x_m}) &= \ln(0.5) \\
-0.1x_m &= \ln(0.5) \\
x_m &= \frac{\ln(0.5)}{-0.1} \\
x_m &\approx 6.931\,\text{seconds}
\end{aligned}
$$

Interpretation: half of the times between emissions are less than 6.931 seconds and half are more.

2. Find the 60th percentile of the times.

   To find the 60 percentile we can set it equal to 0.6:

$$
\begin{aligned}
F(x_{60}) = -e^{-0.1x_{60}} + 1 &= 0.6 \\
-e^{-0.1x_{60}} &= -0.4 \\
e^{-0.1x_{60}} &= 0.4 \\
-0.1x_{60} &= \ln(0.4) \\
x_{60} &= \frac{\ln(0.4)}{-0.1} \\
x_{60} &\approx 9.163\,\text{seconds}
\end{aligned}
$$

Interpretation: 60% of the times between emissions are less than 9.163 seconds and 40% are greater.

7.6 Linear Functions of Random Variables

Definition 17. Adding a constant

If X is a random variable and b is a constant, then

$$\mu_{X+b} = \mu_X + b \qquad \sigma^2_{X+b} = \sigma^2_X$$

Using the alternate notation

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

In general, when a constant is added to a random variable, the mean is shifted by that constant, and the variance is unchanged.

Example 38. Assume that steel rods produced by a certain machine have a mean length of 5.0 in and a variance of $\sigma^2 = 0.003$ in^2.

1. What is the mean length of the assembly?

   Each rod is attached to a base that is 1.0 in long. So the mean length of the assembly will be:

   $$5.0 + 1.0 = 6.0\,\text{inches}$$

2. Can you tell what the variance of the length of the assembly might be?

   Since each length is increased by the same amount, the spread in the lengths doesn't change. So the variance remains the same:

   $$\sigma^2 = 0.003\,\text{in}^2$$

3. Let X represent the length of a randomly chosen rod and let Y represent the length of the assembly. Write Y in terms of the variable X.

   $$\begin{aligned} y &= \text{length of rod} \\ x &= \text{added length} \\ y &= x + 1 \end{aligned}$$

4. Use statistical notation to show how to compute μ_Y.

   $$\begin{aligned} \mu_y &= \mu_{x+1} \\ &= \mu_x + 1 \\ &= 5 + 1 \\ &= 6 \end{aligned}$$

5. Use statistical notation to show how to compute σ_Y^2.

$$\begin{aligned}\sigma_y^2 &= \sigma_{x+1}^2 \\ &= \sigma_x^2 \\ &= 0.003\,\text{in}^2\end{aligned}$$

Note that the standard deviation of Y would be:

$$\begin{aligned}\sigma_y &= \sqrt{0.003} \\ &\approx 0.055\,\text{in}\end{aligned}$$

Definition 18. Multiplying by a constant

If X is a random variable and a is a constant, then

$$\mu_{aX} = a\mu_X \qquad \sigma_{aX}^2 = a^2\sigma_X^2$$

Using alternative notation

$$E(aX) = aE(X) \qquad V(aX) = a^2V(X)$$

In general, when a random variable is multiplied by a constant, its mean is multiplied by the same constant, but its variance is multiplied by the square of the constant.

Example 39. Continuing the last example:

1. If we measure the lengths of the rods described above in centimeters rather than inches, what will the mean length be? (Use 2.54 cm = 1 in.)

$$5 \cdot 2.54 = 12.7\,\text{cm}$$

2. When the length X of a rod is measured in inches, the variance σ_X^2 must have units of in^2. If we measure the lengths of the rods in centimeters, what must the units of the variance be?

   To convert our variance to cm^2 we can multiply:

$$\begin{aligned}\sigma_x^2 &= 0.003 \cdot 2.54^2 \\ &= 0.019\,\text{cm}^2\end{aligned}$$

3. Let Z represent the length of a rod, in cm. Write Z in terms of the random variable X.

$$Z = 2.54X$$

4. Use statistical notation to show how to compute μ_Z.

   The mean of the rod length in cm:

$$\begin{aligned}\mu_Z &= \mu_{2.54X} \\ &= 2.54\mu_X \\ &= 2.54(5.0) \\ &= 12.7\,\text{cm}\end{aligned}$$

5. Use statistical notation to show how to compute σ_Z^2.

$$\begin{aligned}\sigma_Z^2 &= \sigma_{2.54X}^2 \\ &= 2.54^2 \cdot \sigma_X^2 \\ &= 2.54^2(0.003) \\ &\approx 0.019\,\text{cm}^2\end{aligned}$$

   Note that the standard deviation of this is:

$$\begin{aligned}\sigma_Z &= \sqrt{2.54^2 \cdot 0.003} \\ &\approx 0.139\,\text{cm}\end{aligned}$$

7.7 Linear Combinations of Random Variables, Means and Variances

Definition 19. If $X_1, \ldots, X_n$ are random variables, then the mean of the sum $X_1 + \cdots + X_n$ is given by

$$\mu_{X_1+\cdots+X_n} = \mu_{X_1} + \cdots + \mu_{X_n}$$

"The mean of the sum is the sum of the means."

The sum $X_1 + \cdots + X_n$ is a special case of a **linear combination**.

Definition 20. If $X_1, \ldots, X_n$ are random variables and $c_1, \ldots, c_n$ are constants, then the random variable

$$c_1X_1 + \cdots + c_nX_n$$

is called a **linear combination** of $X_1, \ldots, X_n$.

Definition 21. If $X_1, \ldots, X_n$ are random variables and $c_1, \ldots, c_n$ are constants, then the mean of the linear combination $c_1X_1 + \cdots + c_nX_n$ is given by

$$\mu_{c_1X_1+\cdots+c_nX_n} = c_1\mu_{X_1} + \cdots + c_n\mu_{X_n}$$

Definition 22. If $X_1, \ldots, X_n$ are *independent* random variables then the variance of the sum $X_1 + \cdots + X_n$ is given by

$$\sigma^2_{X_1+\cdots+X_n} = \sigma^2_{X_1} + \cdots + \sigma^2_{X_n}$$

"The variance of the sum is the sum of the variances."

In general, if $X_1, \ldots, X_n$ are *independent* random variables, and $c_1, \ldots, c_n$ are constants, then the variance of the linear combination $c_1X_1 + \cdots + c_nX_n$ is given by

$$\sigma^2_{c_1X_1+\cdots+c_nX_n} = c_1^2\sigma^2_{X_1} + \cdots + c_n^2\sigma^2_{X_n}$$

The variance of the sum is the sum of the variances, WHEN the X_i's are independent.

Note 7. Two frequently encountered linear combinations are the sum and the different of two random variables. It is interesting to note that, when the random variables are independent, the variance of the sum is the same as the variance of the different.

For example, if X and Y are *independent* random variables with variance σ^2_X and σ^2_Y, then the variance of the sum $X + Y$ is

$$\sigma^2_{X+Y} = \sigma^2_X + \sigma^2_Y$$

The variance of the different $X - Y$ is

$$\sigma^2_{X-Y} = \sigma^2_X + \sigma^2_Y$$

Example 40. A piston is placed inside a cylinder. The clearance is the distance between the edge of the piston and the wall of the cylinder, and is equal to one-half the different between the cylinder diameter and the piston diameter. Assume the piston diameter has a mean of 80.85 cm with a standard deviation of 0.02 cm. Assume the cylinder diameter has a mean of 80.95 cm with a standard deviation of 0.03 cm.

1. Find the mean clearance.

   Lets define our values

$$
\begin{aligned}
C &= \text{clearance} \\
X &= \text{cylinder diameter} \\
Y &= \text{piston diameter} \\
C &= \frac{1}{2}(X - Y)
\end{aligned}
$$

To find the mean clearance:

$$
\begin{aligned}
\mu_C &= \mu_{\frac{1}{2}X - \frac{1}{2}Y} \\
&= \frac{1}{2}\mu_X - \frac{1}{2}\mu_Y \\
&= \frac{1}{2}(80.95) - \frac{1}{2}(80.85) \\
&\approx 0.050\,\text{cm}
\end{aligned}
$$

2. Assuming that the piston and cylinder are chosen independently, find the standard deviation of the clearance.

   We must first find the variance:

$$
\begin{aligned}
\sigma_C^2 &= \sigma^2_{\frac{1}{2}X - \frac{1}{2}Y} \\
&= \sigma^2_{\frac{1}{2}X} + \sigma^2_{-\frac{1}{2}Y} \\
&= (.5)^2\sigma_Y^2 + (-.5)^2\sigma_Y^2 \\
&= (.5)^2(.03)^2 + (-.5)^2(.02)^2 \\
&= 3.25 \cdot 10^{-4}\,\text{cm}^2
\end{aligned}
$$

   The standard deviation of the clearance is the square root of the variance:

$$
\begin{aligned}
\sigma_C &= \sqrt{\sigma_C^2} \\
&= \sqrt{(.5)^2(.03)^2 + (-.5)^2(.02)^2} \\
&\approx 0.018\,\text{cm}
\end{aligned}
$$

Example 41. Let X, Y, Z be independent random variables.

$$\begin{aligned}\mu_X &= 30\\ \mu_Y &= -18\\ \mu_Z &= 25\\ \sigma_X &= 5\\ \sigma_Y &= 3\\ \sigma_Z &= 2.5\end{aligned}$$

1. Find the mean and standard deviation of
   (a) $2X + 4Y + 10$

$$\begin{aligned}\mu_{2X+4Y+10} &= \mu_{2X+4Y} + 10\\ &= 2\mu_X + 4\mu_Y + 10\\ &= 2(30) + 4(-18) + 10\\ &= -2\end{aligned}$$

$$\begin{aligned}\sigma^2_{2X+4Y+10} &= 2^2\sigma_X^2 + 4^2\sigma_Y\\ &= 4(5^2) + 16(3^2)\\ &= 244\\ \sigma &= \sqrt{244}\\ &\approx 15.6\end{aligned}$$

   (b) $3X + 2Y - 4Z$

$$\begin{aligned}\mu_{3X+2Y-4Z} &= 3\mu_X + 2\mu_Y - 4\mu_Z\\ &= -46\end{aligned}$$

$$\begin{aligned}\sigma^2_{3X+2Y-4Z} &= 3^2\sigma_X^2 + 2^2\sigma_Y^2 + (-4)^2\sigma_Z^2\\ &= 9(5^2) + 4(3^2) + 16(2.5^2)\\ &= 361\\ \sigma &= \sqrt{361}\\ &= 19\end{aligned}$$

7.8 Independent and Simple Random Samples, Mean and Variance of a Sample Mean

Definition 23. A **population** is the entire collection of objects or outcomes about which informatino is sought.

A **sample** is a subset of a population, containing the objects or outcomes that are actually observed.

A **simple random sample** (SRS) of size n is a sample chosen by a method in which each collection of n population items is equally likely to make up the sample, just as in a lottery.

When a simple random sample of numerical values is drawn from a population, each item in the sample can be thought of as a random variable. The items in a simple random sample may be treated as indepent, except when the sample is a large proportion (more than 5%) of a finite population.

For our work from here on, unless explicitly stated to the contrary, we will assume this exception has not occured, so that the values in a simple random sample may be treated as independent random variables.

In general:

If $X_1, \ldots, X_n$ is a simple random sample, then $X_1, \ldots, X_n$ may be treated as independent random variables, all with the same distribution.

When $X_1, \ldots, X_n$ are independent random variables, all with the same distribution, it is sometimes said that $X_1, \ldots, X_n$ are **independent and identically distributed** (i.i.d.).

Items in a **sample** are independent if knowing the values of some items does not help predict the values of the others.

A "pretend" example:

Define a population as: all students at sac state. Define X as the age of a student at sac state.

Consider a simple random sample of size 5. Let:

$$X_1, X_2, X_3, X_4, X_5$$

be a simple random sample. For $i = 1, \ldots, 5$ each X_i is the age of a student at sac state. One particular sample may be:

$$X_1 = 32 X_2 = 20 X_3 = 18 X_4 = 21 X_5 = 25$$

Definition 24. Sample Mean

If $X_1, \ldots, X_n$ is a simple random sample from a population with mean μ and variance σ^2, then the **sample mean** is denoted by $\bar{X}$ and is given by

$$\bar{X} = \frac{X_1 + \cdots + X_n}{n}$$

Rewriting the above expression, we see that the sample mean $\bar{X}$ is a linear combination

$$\bar{X} = \frac{1}{n}X_1 + \cdots + \frac{1}{n}X_n$$

From this fact, we can think about how to derive the mean and variance of $\bar{X}$:

Suppose $\bar{X}$ is the population variable with mean μ and variance σ^2. For the simple random sample $X_1, \ldots, X_n$ from this population, $X_1, \ldots, X_n$ all have the same distribution as the population variable X. It follows that

$$\begin{aligned}
\mu_{\bar{X}} = \mu_{\frac{1}{n}X_1 + \cdots + \frac{1}{n}X_n} &= \frac{1}{n}\mu_{X_1} + \cdots + \frac{1}{n}\mu_{X_n} \\
&= \underbrace{\frac{1}{n}\mu + \cdots + \frac{1}{n}\mu}_{\text{there are } n \text{ of these}} \\
&= (n)\left(\frac{1}{n}\right)\mu \\
&= \mu
\end{aligned}$$

and, since the items in a simple random sample may be treated as indepent random variables

$$\begin{aligned}
\sigma^2_{\bar{X}} = \sigma^2_{\frac{1}{n}X_1 + \cdots + \frac{1}{n}X_n} &= \frac{1}{n^2}\sigma^2_{X_1} + \cdots + \frac{1}{n^2}\sigma^2_{X_n} \\
&= \underbrace{\frac{1}{n^2} + \cdots + \frac{1}{n^2}\sigma^2}_{\text{there are } n \text{ of these}} \\
&= (n)\left(\frac{1}{n^2}\right)\sigma^2 \\
&= \frac{\sigma^2}{n}
\end{aligned}$$

A sample mean is an arithmetic average.

So considering our last example, with our sample size of 5:

$$n = 5$$

So the sample mean can be computed:

$$\bar{X} = \frac{X_1 + X_2 + X_3 + X_4 + X_5}{5}$$

$\bar{X}$ is a variable. One particular value of a sample mean, using set above would me:

$$\begin{aligned}\bar{X} &= \frac{32 + 20 + 18 + 21 + 25}{5} \\ &= 23.2\,\text{years}\end{aligned}$$

The value of $\bar{X}$ will change depending on which sample we are looking at.

Example 42. A process that fills plastic bottles with a beverage has a mean fill volume of 2.013 L and a standard deviation of 0.005 L. A case contains 24 bottles. Assuming that the bottles in a case are a simple random sample of bottles filled by this method, find the mean and standard deviation of the average volume per bottle in a case.

Let $V_1, \ldots, V_{24}$ be a simple random sample. For $i = 1, \ldots, 24$, each V_i is the volume of a bottle.

Average volume per bottle of a case of 24 bottles is the sample mean:

$$\bar{V} = \frac{V_1 + \cdots + V_{24}}{24}$$

Mean of the average volume per bottle in a case:

$$\begin{aligned}\mu_{\bar{V}} &= \mu \\ &= 2.013\,\text{L}\end{aligned}$$

The standard deviation:

$$\begin{aligned}\sigma_{\bar{V}} &= \frac{\sigma}{\sqrt{n}} \\ &= \frac{0.005}{\sqrt{24}} \\ &\approx 0.001\,\text{L}\end{aligned}$$

8 Commonly Used Distributions

8.1 Bernoulli Distribution

Many applications of probability and statistics concern the repetition of an experiment, where each repetition is called a *trial*. A **Bernoulli trial** is an experiment with exactly two possible outcomes, labaled "success" and "failure". The probability of success is denoted by p and the probability of failured is denoted as $1 - p$.

Definition 25. For any Bernoulli trial, we define a random variable X as follows

If the experiment results in success, then $X = 1$, otherwise $X = 0$.

It follows that X is a discrete random variable with probabiliity mass function $p(x)$ defined by

$$p(0) = P(X = 0) = 1 - p$$
$$P(1) = P(X = 1) = p$$

The random variable X is said to have a **Bernoulli distribution** with parameter p.

The notation is $X \sim Bernoulli(p)$.

The mean and variance of a Bernoulli Random Variable is

$$\mu_X = p \qquad \sigma_X^2 = p(1 - p)$$

Example 43. For example, the tossing of a coin is a simple Bernoulli trial. There are 2 possible outcomes, heads or tails. We can define a heads as "success". With a fair coin, the probability of success is $p = 0.5$.

Let $X = 1$ if the coin comes up heads. Let $X = 0$ otherwise. Then X has a Bernoulli distribution with $p = 0.5$. So:

$$X\ Bernoulli(0.5)$$

Example 44. Ten percent of the components manufactured by a certain process are defective. A component is chosen at random. Let $X = 1$ if the component is defective and $X = 0$ otherwise. What is the distribution of X?

The success probability is:

$$p(1) = P(X = 1) = 0.1$$

X has a Bernoulli distribution with parameter $p = 0.1$. That is,

$$X\ Bernoulli(0.1)$$

What is the mean and variance?

Recall that the mean is:

$$
\begin{aligned}
\mu_x &= \sum_x xp(x) \\
&= 0(1-p) + 1(p) \\
&= p
\end{aligned}
$$

Recall that the variance is:

$$
\begin{aligned}
\sigma_x^2 &= \sum_x (x-\mu)^2 \cdot p(x) \\
&= (0-p)^2(1-p) + (1-p)^2(p) \\
&= p^2(1-p) + (1-p)^2(p) \\
&= p(1-p)\left[p + (1-p)\right] \\
&= p(1-p)
\end{aligned}
$$

So

$$
\begin{aligned}
\mu_x &= 0.1 \\
\sigma_x^2 &= 0.1(0.9) \\
&= .09
\end{aligned}
$$

8.2 Binomial Distribution

Definition 26. If a total of n Bernoulli trials are conducted and

- The trials are independent.
- Each trial has the same success probability p.
- X is the number of successes in the n trials

then X has the **binomial distribution** with parameters n and p, denoted $X \sim Bin(n, p)$.

If $X \sim Bin(n, p)$, the probability mass function of X is

$$p(x) = P(X = x) = \begin{cases} \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x} & x = 0, 1, 2, \ldots, n \\ 0 & \text{otherwise} \end{cases}$$

Recall

$$\binom{n}{x} = \frac{n!}{x!(n-x)!} =_n C_x$$

The mean and variance of a binomial random variable is

$$\mu_X = np \qquad \sigma_X^2 = np(1-p)$$

Note 8. Assume that a finite population contains items of twp types, successes and failures, and that a simple random sample is drawn from the population. Then if the sample size is no more than 5% of the population, the binomial distribution may be used to model the number of successes. For example, a sampling "without replacement" experiment can be treated as binomial as long as the sample size (number of trials) n is at most 5% of the population size.

Example 45. Suppose a biased coin has a probability of 0.6 of coming up heads. The coin is tossed three times. Let X = number of heads in 3 tosses. Lets make sure this is a binomail distribution:

- $n = 3$ Bernoulli trialgs (coin tossed 3 times 2 possible outcomes H or T)
- The trials are independent.
- $p = 0.6$, probability of success is probability coin comes up heads and this stays the same from trial to trial.

So X has a binomial distribution with parameters $n = 3$ and $p = 0$. That is,

$$X \ Bin(3, 0.6)$$

Lets find the probability that exactly 2 of the 3 tosses come up heads:

$$P(X = x) = \begin{cases} \frac{n!}{3!(3-x)!}(0.6)^x(0.4)^{3-x} & x = 0, \dots, n \\ 0 & \text{otherwise} \end{cases}$$

So the probability of exactly 2 heads:

$$\begin{aligned} P(X = 2) &= \frac{3!}{2!1!}(0.6)^2(0.4)^1 \\ &= 0.432 \end{aligned}$$

$$\begin{aligned} P(X = 2) &= 3(0.6)^2(0.4)^1 \\ &= \underbrace{\binom{3}{2}}_{\text{how many outcomes have exactly 2 heads}} \cdot \overbrace{(0.6)^2(0.4)^1}^{\text{prob. of an individual outcome with 2 heads}} \end{aligned}$$

A variation of the Bernoulli random variable.

Mini-proof:

A binomial random variable can be written as a sum of Bernoulli random variables. Assume n independent Bernoulli trials are conducted, each with a success probability p. Consider $Y_1, \dots, Y_n$ defined as follows:

For $i = 0, \dots, n$:

$$Y_i = 1$$

If the ith trial results in success, otherwise:

$$Y_i = 0$$

Let X = number of success among n trials. Then:

$$X = \underbrace{Y_1 + \cdots + Y_n}_{\text{sum gives the number of } Y_i \text{ that equal 1}}$$

So the mean of X is the mean of that sum:

$$\begin{aligned} \mu_X &= \mu_{Y_1} + \cdots + \mu_{Y_n} \\ &= \underbrace{p + \cdots + p}_{n \text{ of these}} \\ &= np \end{aligned}$$

The variance:

$$\begin{aligned}\sigma_X^2 &= \sigma_{Y_1}^2 + \cdots + \sigma_{Y_n}^2 \\ &= \underbrace{p(1-p) + \cdots + p(1-p)}_{n \text{ of these}} \\ &= np(1-p)\end{aligned}$$

Example 46. Approximately 40% of all pizza orders are carry-out. Suppose 20 pizza orders are randomly selected

1. Find the probability that at most 8 are carry-out orders.

   2 outcomes: carry-out order, or not carry-out order. We'll say success is a carry-out order, so $p = 0.4$. 20 pizza orders are randomly selected, so $n = 20$.

$$X \ Bin(20, 0.4)$$

2. Find the probability that exactly 10 are carry out orders.

   We can use our calculator with $P(X = 10)$

$$= .117142$$

3. Find the probability that at least 7 are carry out orders.

   We want to find $P(X \geq 7)$:

$$= .749989$$

4. Find the probability that between 5 and 11 (inclusive) are carry-out orders.

   We are looking for $P(5 \leq X \leq 11)$

$$= .892522$$

5. Find the mean, variance, and standard deviation of the number of pizza orders that are carry out among 20 pizza orders.

$$\begin{aligned}\mu_X &= np \\ &= 20(0.4) \\ &= 8\end{aligned}$$

   "We can expect (on average) 8 out of every 20 pizzas to be carry-out."

$$\begin{aligned}
\sigma_X^2 &= np(1-p) \\
&= 20(0.4)(0.6) \\
&= 4.8 \\
\sigma_X &= \sqrt{4.8} \\
&\approx 2.19
\end{aligned}$$

8.2.1 Sample proportion

Definition 27. Sometimes the success probability p associated with a certain Bernoulli trial is unknown.

Suppose n independent trials are conducted and the number of X successes is counted. To estimate the success probability p, we compute the **sample proportion** $\hat{p}$ given by

$$\hat{p} = \frac{\text{number of successes}}{\text{number of trials}} = \frac{X}{n}$$

Note 9. The sample proportion $\hat{p}$ is an estimate of the success probability p, and, in general, is not equal to p.

8.2.2 Bias and Uncertainty

In general, bias is the difference between the mean of the estimator and the true mean. When bias equals 0, the mean of the estimator is equal to the true mean (the expected value of the estimator is equal to the true value of the parameter) and we say the estimator is unbiased.

Uncertainty is the amount of error expected in an estimate of the true mean and is given by the standard deviation of the estimator.

Definition 28. Let n denote the sample size, and let X denote the number of successes, where $X \sim Bin(n, p)$.

- The mean of the sample proportion $\hat{p}$ is

$$\begin{aligned}\mu_{\hat{p}} &= \mu_{\frac{X}{n}} \\ &= \frac{1}{n}\mu_X \\ &= \frac{1}{n}np \\ &= p\end{aligned}$$

  For example, the expected value of the sample proportion $\hat{p}$ is equal to the population parameter p, that is $E(\hat{p}) = p$.

  So $\hat{p}$ is an unbiased estimator of p.

- The uncertainty is the standard deviation of the sample proportion $\hat{p}$. The variance of $\hat{p}$ is

$$\begin{aligned}\sigma^2_{\hat{p}} &= \sigma^2_{\frac{X}{n}} \\ &= \frac{1}{n^2}\sigma^2_X \\ &= \frac{1}{n^2}np(1-p) \\ &= \frac{p(1-p)}{n}\end{aligned}$$

  So the standard deviation of the sample proportion $\hat{p}$ is

$$\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$$

Example 47. The safety commissioner in a large city wants to estimate the proportion of buildings in the city that are in violation of fire codes. A random sample of 40 buildings is chosen for inspection, and 4 of them are found to have fire code violations.

1. Estimate the proportion of buildings in the city that have fire code violations, and find the uncertainty in that estimate.

   We have a sample size of $n = 40$. Let X = number of buildings with fire code violations among 40. In general:

$$X \; Bin(40, p)$$

   p is unknown.

   Lets first compute the sample proportion:

$$\begin{aligned}\hat{p} &= \frac{X}{n}\\ &= \frac{4}{40}\\ &= 0.1\end{aligned}$$

Next we can calculate the uncertainty:

$$\begin{aligned}\sigma_{\hat{p}} &= \sqrt{\frac{p(1-p)}{n}}\\ &\approx \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\\ &\approx \sqrt{\frac{0.1(1-0.1)}{40}}\\ &\approx 0.047\end{aligned}$$

2. Approximately how many additional buildings must be inspected so that the uncertainty in the sample proportion of buildings in violation is only 0.02?

   Find the sample size n needed so that $\sigma_{\hat{p}} = 0.02$. We don't have p so we must use our value of $\hat{p} = 0.1$.

$$\begin{aligned}0.02 &= \sqrt{\frac{p(1-p)}{n}}\\ &= \sqrt{\frac{0.1(1-0.1)}{n}}\\ n &= \frac{(0.1)(0.9)}{(0.002)^2}\\ &= 225\end{aligned}$$

   So we need to subtract the 225 from the 40 already inspected. Therefore:

$$\begin{aligned}n &= 225 - 40\\ &= 185\end{aligned}$$

   So we'd need to inspect 185 more buildings to get our uncertainty down to 2%. If our additional inspections happened to be a decimal, then we round up to the nearest whole number to guarantee we have the right uncertainty.

8.3 Hypergeometric Distribution

Definition 29. Assume a finite population contains N items, of which R are classified as successes and $N-R$ are classified as failures. Assume that n items are sampled (without replacement) from this population, let X represent the number of successes in the sample. Then X has the **hypergeometric distribution** with parameters N, R, and n, which can be denoted $X \sim H(N, R, n)$.

The probability mass function of X is

$$p(x) = P(X = x) = \begin{cases} \frac{\binom{R}{x}\binom{N-R}{n-x}}{\binom{N}{n}} & \max(0, R+n-N) \leq x \leq \min(n, R) \\ 0 & \text{otherwise} \end{cases}$$

The mean and variance of X are

$$\mu_X = \frac{nR}{N} \qquad \sigma_X^2 = n\frac{R}{N}\left(1 - \frac{R}{N}\right)\left(\frac{N-N}{N-1}\right)$$

8.4 Geometric Distribution

Definition 30. Let X represent the number of trials up to and including the first success in a sequence of independent Bernoulli trials with a constant success probability p. Then X has a **geometric** distribution with parameter p, written $X \sim Geom(p)$.

The probability mass function X is

$$p(x) = P(X = x) = \begin{cases} p(1-p)^{x-1} & x = 1, 2, \cdots \\ 0 & \text{otherwise} \end{cases}$$

The mean and variance of X are

$$\mu_x = \frac{1}{p} \qquad \sigma_x^2 = \frac{1-p}{p^2}$$

Some background: Think of an experiment like phoning a friend until you get through. The number of calls necessary until the first success (getting through) is realized is the value of a geometric random variable.

Let $X =$ the number of trials (calls) up to and including the first success (getting through).

Let $p =$ probability of success, and let $s =$ success, $f =$ failure. So,

$$P(s) = p \qquad P(f) = 1 - p$$

possible outcomes	Probabilities
s	$P(X=1) = p(s) = p$
f s	$P(X=2) = p(f) \cdot p(s) = (1-p)p$
f f s	$P(X=3) = (1-p)^2 \cdot p$
f f f s	$P(X=4) = (1-p)^3 \cdot p$
f f f f s	$P(X=5) = (1-p)^5 \cdot p$

Example 48. A test of weld strength involves loading welding joints until a fracture occurs. For a certain type of weld, 80% of the fractures occur in the weld itself, while the other 20% occur in the beam. A number of welds are tested. Let X be the number of tests up to and including the first test that results in a beam fracture.

1. What is the distribution of X?

   Each test is a Bernoulli Trial. Success = "a beam fracture".

$$p(\text{success}) = 0.2$$
$$X \sim Geom(0.2) = (0.2)(0.8)^{x-1}$$

2. Find $P(X = 3)$

   In other words, find the probability it takes for 3 tests to realize the beam fracture.

$$P(X = 3) = (0.2)(0.8)^2 = 0.128$$

3. Find the mean and variance of X.

$$\begin{aligned}\mu_x &= \frac{1}{p} \\ &= \frac{1}{0.2} \\ &= 5\end{aligned}$$

   On average it takes 5 tests to realize the 1st beam fracture.

$$\begin{aligned}\sigma_x^2 &= \frac{1-p}{p^2}\\ &= \frac{1-0.2}{0.2^2}\\ &= 20\end{aligned}$$

8.5 Negative Binomial Distribution

Definition 31. Let X represent the number of trials up to and incliuding the rth success in a sequence of independent Bernoulli trials with a constant success probability p. Then X has a **negative binomial** distribution with paramter p, written $X \sim NB(r, p)$.

The probability mass function X is

$$p(x) = P(X = x) = \begin{cases} \binom{x-1}{r-1} p^r (1-p)^{x-r} & x = r, r+1, \cdots \\ 0 & \text{otherwise} \end{cases}$$

The mean and variance of X are

$$\mu_x = \frac{r}{p} \qquad\qquad \sigma_x^2 = \frac{r(1-p)}{p^2}$$

Aside 1.

$$? = \underbrace{\binom{x-1}{r-1} p^{r-1} (1-p)^{x-r}}_{\text{a binomial probability}} \cdot p$$

Aside:

$r = 4$

4th success

f S f f S f S f | S

f f S f S f f S | S $n = 9$ trials

probability p

Prob. of getting 3 successes in 8 trials

$$\underbrace{\binom{8}{3} p^3 \cdot (1-p)^5 \cdot p}_{\binom{8}{3} p^4 (1-p)^5}$$

Example 49. In a test of weld strength, 80% of the tests result in a fracture in the weld, while the other 20% result in a fracture in the beam. Let X denote the number of tests up to and including the third beam fracture.

1. What is the distribution of X?
   Success = beam fracture. X has the negative binomial distribution with $r = 3$ and $p = 0.2$, that is
   $$X \sim NB(3, 0.2)$$

2. Find $P(X = 8)$
   This is the probability it takes exactly 8 tests to realize the 3rd beam fracture.
   $$\begin{aligned} P(X=8) &= \binom{8-2}{2}(0.2)^3(0.8)^{8-3} \\ &= \binom{7}{2}(0.2)^3(0.8)^5 \\ &\approx 0.0551 \end{aligned}$$
   Another way to compute it with your calculator:
   $$\binom{7}{2}(0.2)^2(0.8)^5 \cdot (0.2)$$
   $$\text{binompdf}(7, .2, 2) \cdot 0.2$$

3. Find the mean and variance of X.

   Aside 2. If $X \sim NB(r,p)$ then $X = Y_1 + \cdots + Y_r$ where $Y_1, \ldots, Y_r$ are independent variables, each with the $Geom(p)$ distribution.

   informal example using 8 trials with 3 successes $(r = 3)$

   $\underbrace{FFS}_{Y_1=3}\ \underbrace{FS}_{Y_2=2}\ \underbrace{FFS}_{Y_3=3}$ $\qquad Y_1, Y_2, Y_3$

   $Y_1 + Y_2 + Y_3 = 3 + 2 + 3 = 8$ trials

   The mean of the variable X is the sum of the means of the Y_is. The variance of X is the sum of the variances of the Y_is.

   So,
   $$\begin{aligned} \mu_x &= \frac{r}{p} \\ &= \frac{3}{0.2} \\ &= 15 \end{aligned}$$
   $$\begin{aligned} \sigma_x^2 &= \frac{r(1-p)}{p^2} \\ &= \frac{3(0.8)}{0.2^2} \\ &= 60 \end{aligned}$$

8.6 Multinomial Distribution

Recall: A Bernoulli trial is a process that results in one of two possible outcomes. A multinomial trial is a generalization of the Bernoulli trial and is a process that can result in any of k outcomes, where $k \geq 2$.

Definition 32. Assume that n independent multinomial trials are conducted, each with the same k possible outcomes and with the same probability $p_1, \ldots, p_k$. Number the outcomes $1, 2, \ldots, k$. For each outcome, let X_i denote the number of trials that result in that outcome. Then $X_1, \ldots, X_k$ are discrete random variables.

The collection $X_1, \ldots, X_k$ is said to have the **multinomial** distribution with parameters $n, p_1, \ldots, p_k$. We write $X_1, \ldots, X_k \sim MN(n, p_1, \ldots, p_k)$.

The probability mass function of $X_1, \ldots, X_k$ is

$$p(x_1, \ldots, x_k) = P(X_1 = x_1, \ldots, X_k = x_k) = \begin{cases} \frac{n!}{x_1! x_2! \ldots x_k!} p_1^{x_1} p_2^{x_2} \ldots p_k^{x_k} & x_i = 0, 1, 2, \ldots, n \text{ and } \sum x_i = n \\ 0 & \text{otherwise} \end{cases}$$

Furthermore, if $X_1, \ldots, X_k \sim MN(n, p_1, \ldots, p_k)$, then for each i:

$$X_i \sim Bin(n, p_i)$$

Note 10. The first part of the equation is "the number of ways of dividing a group of n objects into groups of $x_1, \ldots, x_k$ objects (with $x_1 + \cdots + x_k = n$)" This gives us the number of outcomes with the desired quality.

The second part of the equation gives us probability for a particular outcome.

Example 50. Alkaptonuria is a genetic disease that results in the lack of an enzyme necessary to break down homogentistic acid. Some people are carriers of alkaptonuria, which means that they do not have the disease themselves, but they can potentially transmit it to their offspring. According to the laws of genetic inheritance, an offspring both of whose parents are carriers of alkaptonuria has a probability of 0.25 of being unaffected, 0.5 of being a carrier, and 0.25 of having the disease.

1. In a sample of 10 offspring of carriers of alkaptonuria, what is the probability that 3 are unaffected, 5 are carriers, and 2 have the disease?

   Sample size of $n = 10$. Lets define our variables, let X_1, X_2, X_3 denote the numbers among the 10 offspring who are unaffected, carriers, and diseased respectively.

   So $P_1 = 0.25$ is the probability offspring is unaffected, $P_2 = 0.5$ is probability offspring is carrier, and $P_3 = 0.25$ is the probability offspring is diseased.

   We are looking for:

$$X_1, X_2, X_3 \sim MN(10, 0.25, 0.5, 0.25)$$

$$\begin{aligned} P(X_1 = 3, X_2 = 5, X_3 = 2) &= \frac{10!}{3!5!2!} \cdot (0.25)^3 (0.5)^5 (0.25)^2 \\ &\approx 0.07690 \end{aligned}$$

2. Find the probability that exactly 4 of 10 offspring are unaffected.

   So our sample size is $n = 10$. We have a success probability for unaffected of $p = 0.25$. Let $Y =$ number of offspring unaffected among 10.

$$Y \sim Bin(10, 0.25)$$

$$\begin{aligned} P(Y = 4) &= \binom{10}{4}(0.25)^4(0.75)^6 \\ &\approx 0.146 \end{aligned}$$

8.7 Poisson Distribution

A Poisson experiment has the following properties:

1. The probability that a single event occurs in a given interval (of time, length, volume, etc.) is the same for all intervals.
2. The number of events that occur in any interval is independent of the number that occur in any other interval.

These properties are often referred to as a "Poisson process" and can be difficult to verify.

Definition 33. The **Poisson random variable** X is a count of the number of times the specific event occurs during a given interval.

The Poisson distribution is completely determined by the mean, denoted by the Greek latter lambda, λ. (Since the Poisson distribution is often used to count rare events, the mean number of events per interval is usually small.) If X has a Poisson distribution with parameter λ, we write $X \sim Poisson(\lambda)$.

A Poisson distribution can be thought of as an approximation to the binomial distribution with n is large and p is small. Specifically, if n is very large and p is very small, and we let $\lambda = np$, is can be shown by advanced methods that for all x,

$$\frac{n!}{x!(n-x)!}p^x(1-p)^{n-x} \approx e^{-\lambda}\frac{\lambda^x}{x!}$$

The approximation above leads to the Poisson probability mass function given below.

If $X \sim Poisson(\lambda)$, then

- X is a discrete random variable whose possible values are the non-negative integers.
- The parameter λ is a positive constant.
- The probability mass function of X is

$$p(x) = P(X = x) = \begin{cases} e^{-\lambda}\frac{\lambda^x}{x!} & \text{if } x \text{ is a non-negative integer} \\ 0 & \text{otherwise} \end{cases}$$

- The Poisson probability mass function is very close to the binomial probability mass function when n is large, p is small, and $\lambda = np$.
- The mean and variance of X are given by

$$\mu_x = \lambda \qquad \sigma_x^2 = \lambda$$

Example 51. Skydiving is the ultimate thrill for some, and there were over 3.1 million jumps in 2012. Despite an improved safety record, there are approximately 2 skydiving fatalities each month. Suppose this is the mean number of fatalities per month and a random month is selected.

1. Find the probability that exactly three fatalities will occur.

   Let X = number of skydiving fatalities that occur in a month. We know that the mean number of fatalities per month is 2, so $\lambda = 2$. So,

$$X \sim Poisson(2)$$

   So,

$$\begin{aligned} P(X = 3) &= e^{-2} \cdot \frac{2^3}{3!} \\ &\approx 0.1804 \end{aligned}$$

2. Find the probability that at least five fatalities occur.

$$
\begin{aligned}
P(X \geq 5) &= 1 - P(X < 5) \\
&= 1 - P(X \leq 4) \\
&\approx 1 - .9473 \\
&\approx 0.0527
\end{aligned}
$$

3. Find the mean and variance of the number of skydiving fatalities per month.

$$
\begin{aligned}
\mu_x &= \lambda = 2 \\
\sigma_x^2 &= \lambda = 2
\end{aligned}
$$

Example 52. Particles are suspended in a liquid medium at a concentration of 6 particles per mL. A large volume of suspension is thoroughly agitated, and then 3 mL are withdrawn. What is the probability that exactly 15 particles are withdrawn?

Let X = number of particles withdrawn in 3 mL. Mean number of particles in 3 mL: $6 \cdot 3 = 18$ mL. So $\lambda = 18$.

$$X \sim Poisson(18)$$

$$
\begin{aligned}
P(X = 15) &= e^{-18}\frac{18^{15}}{15!} \\
&\approx .07857
\end{aligned}
$$

Example 53. Assume that the number of hits on a certain website during a fixed time interval follows a Poisson distribution. Assume that the mean rate of hits is 5 per minute.

1. Find the probability that there will be exactly 17 hits in the next three minutes.

   Let X = number of hits in 3 minutes. $\lambda_X = 5 \cdot 3 = 15$, 15 hits per minute on average in 3 minutes.

   $$X \sim Poisson(15)$$

   $$
   \begin{aligned}
   P(X = 17) &= e^{-15}\frac{15^{17}}{17!} \\
   &\approx 0.08473
   \end{aligned}
   $$

2. Let X be the number of hits in t minutes. Find the probability mass function of X in terms of t.

   $\lambda_X = 5t$.

   $$X \sim Poisson(5t)$$

   $$P(X = x) = e^{-5t}\frac{5t^x}{x!}$$

where $x = 0, 1, 2, \cdots$

8.7.1 Using the Poisson Distribution to estimate rate and uncertainty in the estimated rate

Let λ denote the mean number of events that occur in one unit of time or space. Let X enote the number of events that are observed to occur in t units of time or space. Then if $X \sim Poisson(\lambda t)$, λ is estimated with $\hat{\lambda} = \frac{X}{t}$. In general:

If $X \sim Poisson(\lambda t)$, we estimate the rate λ with $\hat{\lambda} = \frac{X}{t}$.

- $\hat{\lambda}$ is unbiased, (i.e. $\sigma_{\hat{\lambda}} = \lambda$).
- The uncertainty in $\hat{\lambda}$ is $\sigma_{\hat{\lambda}} = \sqrt{\frac{\hat{\lambda}}{t}}$

In practice, we substitute $\hat{\lambda}$ for λ in the equation for $\sigma_{\hat{\lambda}}$ above, since λ is unknown.

Example 54. A 5 mL sample of a suspension is withdrawn, and 47 particles are counted. Estimate the mean number of particles per mL, and find the uncertainty in the estimate.

So our estimated average number of particles per mL is $\lambda \approx \hat{\lambda} = \frac{47}{5} = 9.4$. The uncertainty in our estimate is

$$\begin{aligned} \sigma_{\hat{\lambda}} &= \sqrt{\frac{\lambda}{t}} \\ &\approx \sqrt{\frac{\hat{\lambda}}{t}} \\ &\approx \sqrt{\frac{9.4}{5}} \\ &\approx 1.371 \end{aligned}$$

Example 55. A certain mass of a radioactive substance emits alpha particles at a mean rate of λ particles per second. A physicist counts 1594 emissions in 100 seconds.

1. Estimate λ, and find the uncertainty in the estimate.

   The unit rate $\lambda \approx \hat{\lambda} = \frac{1594}{100} = 15.94$ emissions per second. The uncertainty in our estimate is

$$\begin{aligned} \sigma_{\hat{\lambda}} &= \sqrt{\frac{\lambda}{100}} \\ &\approx \sqrt{\frac{\hat{\lambda}}{100}} \\ &\approx \sqrt{\frac{15.94}{100}} \\ &\approx 0.399 \end{aligned}$$

2. For how many seconds should emissions be counted to reduce the uncertainty to 0.3 emissions per second?

We want to find time t, so that our uncertainty $\sigma_{\hat{\lambda}} = 0.3$.

$$\begin{aligned} 0.3 &= \sqrt{\frac{\lambda}{t}} \\ &\approx \sqrt{\frac{\hat{\lambda}}{t}} \\ &= \sqrt{\frac{15.94}{t}} \\ t &\approx 177.11 \text{ seconds} \end{aligned}$$

8.8 Exponential Distribution

The **exponential distribution** is a continuous distribution that is sometimes used to model the time that elapses before an event occurs. Such a time is often called a **waiting time**. For example, the exponential distribution is often used to model the time to failure, or lifetime, of an electronic part. The probability density function of the exponential edistribution involves a parameter, which is a positive constant λ whose value determines the density function's location and shape.

Definition 34. The **probability density function** of the exponential distribution with paramter $\lambda > 0$ is

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & x > 0 \\ 0 & x \leq 0 \end{cases}$$

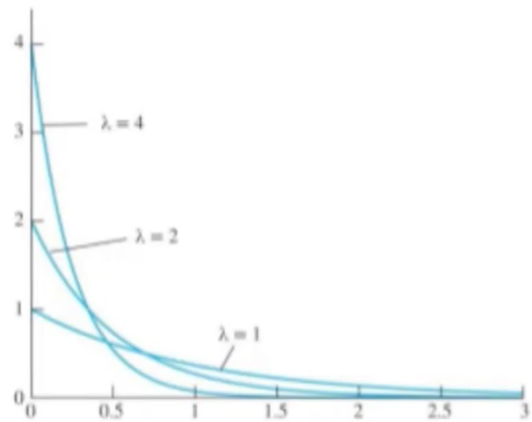
If X is a random variable whose distribution is the exponential distribution with paramter λ, we write $X \sim Exp(\lambda)$.

The mean and variance of an exponential distribution is

$$\mu_X = \frac{1}{\lambda} \qquad \sigma_X^2 = \frac{1}{\lambda^2}$$

The mean and variance can be calculated by using integration by parts.

Plots of the exponential probability density function for various values of λ:



The cumulative distribution function is defined as

$$F(x) = P(X \leq x) = \begin{cases} 1 - e^{-\lambda x} & x > 0 \\ 0 & x \leq 0 \end{cases}$$

Note 11. The pdf is easily integrated to obain the above cdf:

$$\begin{aligned} F(x) = P(X \leq x) &= \int_0^x \lambda e^{-\lambda t} \, \mathrm{d}t \\ &= -e^{-\lambda t} \mid_0^x \\ &= 1 - e^{-\lambda x} \end{aligned}$$

Example 56. Some pharmacists recommend Zicam to relieve many symptoms due to the common cold. After the prescribed dose is taken, suppose the length of time (in hours) until the symptoms return is a random variable X, that has a exponential distribution with paramter $\lambda = 0.1$.

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

   So we can see $X \sim Exp(0.1)$. So,

$$\begin{aligned} \mu_X &= \frac{1}{\lambda} \\ &= \frac{1}{0.1} \\ &= 10 \text{ hours} \end{aligned}$$

$$\begin{aligned} \sigma_X^2 &= \frac{1}{\lambda^2} \\ &= \frac{1}{0.1^2} \\ &= 100 \end{aligned}$$

So the standard deviation is

$$\begin{aligned} \sigma_X &= \sqrt{100} \\ &= 10 \end{aligned}$$

2. What is the probability that the length of time until symptoms return is less than the mean?

   We are looking for $P(X < 10)$. Since X is a continuous variable we can use the cdf.

$$\begin{aligned} P(X < 10) &= P(X \leq 10) \\ &= 1 - e^{-0.1(10)} \\ &\approx 1 - 0.3679 \\ &= 0.6321 \end{aligned}$$

3. What is the probability that the length of time until symptoms return is at least 12 hours?

   We are looking for $P(X \geq 12)$. We can use the complementation rule to calculate it.

$$\begin{aligned} P(X \geq 12) &= 1 - P(X < 12) \\ &= 1 - P(X \leq 12) \\ &= 1 - \left(1 - e^{-0.1(12)}\right) \\ &= e^{-0.1(12)} \\ &\approx 0.3012 \end{aligned}$$

4. What is the probability that the length of time until symptoms return is between 8 and 16 hours?

   We are looking for $P(8 \leq X \leq 16)$.

$$\begin{aligned} P(8 \leq X \leq 16) &= P(X \leq 16) - P(X \leq 8) \\ &= \left(1 - e^{-0.1(16)}\right) - \left(1 - e^{-0.1(8)}\right) \\ &= e^{-0.1(8)} - e^{-0.1(16)} \\ &\approx 0.2474 \end{aligned}$$

5. What is the value of the median length of time until symptoms return?

   Recall that the median is the 50th percentile. It is the number that separates the bottom 50% from the top 50%. So we are looking for t such that the $P(X \leq t) = 0.5$.

$$
\begin{aligned}
1 - e^{-0.1t} &= 0.5 \\
e^{-0.1t} &= 0.5 \\
t &= \frac{\ln(0.5)}{-0.1} \\
&\approx 6.9 \text{ hours}
\end{aligned}
$$

8.8.1 The Exponential Distribution and the Poisson Process

The exponential distribution can be used as a model for waiting times whenever the events follow a Poisson process. Recall that events follow a Poisson process with rate parameter λ when the number of events in disjoint intervals are independent, and the number of X events that occur on any time interval of length t has a Poisson distribution with mean λt, for example, $X \sim Poisson(\lambda t)$. The connection between the exponential distribution and the Poisson process are as follows:

If any events follow a Poisson process with rate paramter λ, and if T represents the waiting time from any starting point until the next event, then $T \sim Exp(\lambda)$.

Definition 35. The exponential distribution also has a property known as the memorless property:

If $T \sim Exp(\lambda)$, and t and s are positive numbers, then

$$P(T > t + s \mid T > s) = P(T > t)$$

The probability that we must wait an additional t units, given that we have already waited s units, is the same as the probability that we must wait t units from the start.

Example 57. The probability that a 4 year old circuit lasts 3 additional years is exactly the same as the probability that a new circuit will last 3 years.

Example 58. A radioactive mass emits particles according to a Poisson process at a mean rate of 15 particles per minute (60 seconds). At some point, a clock is started.

1. What is the probability that more than 5 seconds will elapse before the next emission?

   Let T = the time in seconds that elapses before the next particle is emitted. The mean rate of emissions is $\lambda = \frac{15}{60} = 0.25$. So, $T \sim Exp(0.25)$.

$$
\begin{aligned}
P(T > 5) &= 1 - P(T \leq 5) \\
&= 1 - \left(1 - e^{-0.25(5)}\right) \\
&\approx 0.2865
\end{aligned}
$$

2. What is the mean waiting time (in seconds) until the next particle is emitted?

   The mean waiting time in seconds is

$$\begin{aligned}\mu_T &= \frac{1}{\lambda} \\ &= \frac{1}{0.25} \\ &= 4 \text{ seconds}\end{aligned}$$

Example 59. The lifetime of a an integrated circuit has an exponential distribution with mean 2 years.

1. Find the probability that the circuit lasts longer than three years.

   Let $T =$ lifetime of a circuit. The mean of T is $\mu_T = 2$. We can find λ by solving

$$\begin{aligned}\mu_T &= 2 \\ \frac{1}{\lambda} &= 2 \\ \lambda &= 0.5\end{aligned}$$

   So we have $T \sim Exp(0.5)$.

$$\begin{aligned}P(T > 3) &= e^{-0.5(3)} \\ &\approx 0.223\end{aligned}$$

2. Assume the circuit is now four years old and is still functioning. Find the probability that it functions for more than three additional years.

   We are looking for $P(T > 7 \mid T > 4)$.

$$\begin{aligned}P(T > 7 \mid T > 4) &= \frac{P(T > 7)}{P(T > 4)} \\ &= \frac{e^{-0.5(7)}}{e^{-0.5(4)}} \\ &\approx 0.223\end{aligned}$$

Example 60. The number of hits on a website follows a Poisson process with a rate of 3 per minute.

1. What is the probability that more than a minute goes by without a hit?

   So let $T =$ waiting time in minutes until the next hit. Our rate parameter is $\lambda = 3$. So,

$$T \sim Exp(3)$$

$$
\begin{aligned}
P(T > 1) &= 1 - P(T \leq 1) \\
&= 1 - [1 - e^{-3 \cdot 1}] \\
&\approx 0.0498
\end{aligned}
$$

2. If 2 minutes have gone by without a hit, what is the probability that a hit will occur in the next minute?

   The waiting time is at most 3 minutes given at least 2 minutes have gone by. We can use the memoryless property to find this.

$$
\begin{aligned}
P(T \leq 3 \mid T > 2) &= 1 - P(T > 3 \mid T > 2) \\
&= 1 - P(T > 1) \\
&= 1 - 0.0498 \\
&= 0.9502
\end{aligned}
$$

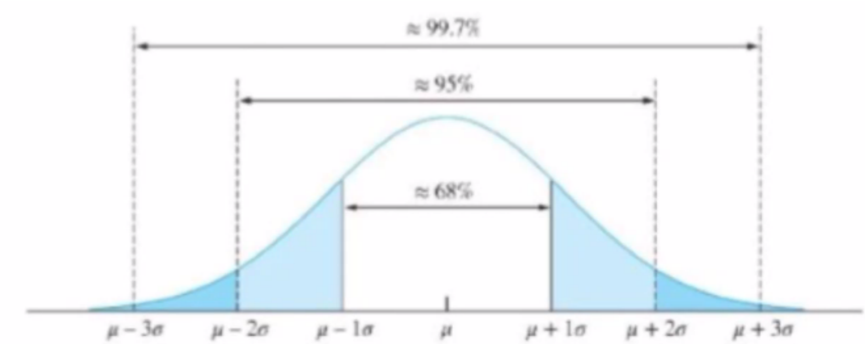
8.9 Normal Distribution

Definition 36. The **normal distribution** (also called the Gaussian distribution) is the most commonly used distribution in statistics. It provides a good model for many, although not all, continuous populations.

The normal distribution is a continuous distribution. The mean of a normal random variable may have any value, and the variance may have any positive value. The probability density function of a normal random variable with mean mu and variance σ^2 is given by

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/(2\sigma^2)}$$

The normal distribution density function is sometimes called the normal curve. The normal curve is a bell-shaped curve that is symmetric about and centered at the mean μ and its apread depends on the standard deviation σ, the larger the standard deviation, the flatter and more spread out is the distribution. The inflection points of a normal curve occur at $\mu - \sigma$ and $\mu + \sigma$.

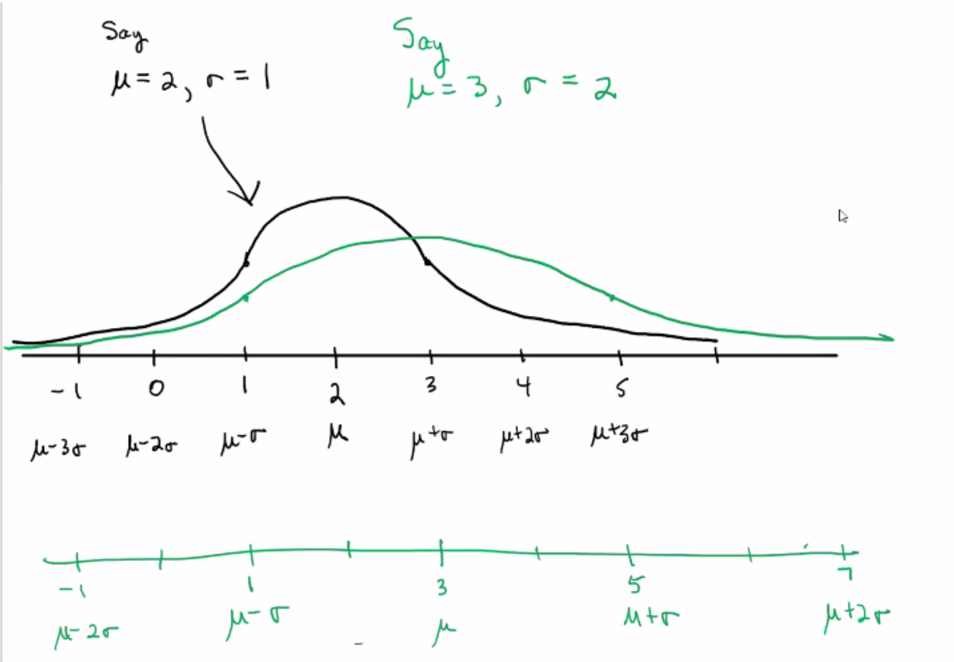


It is also the case that for any normal population

- About 68% of the population is in the interval $\mu \pm \sigma$
- About 95% of the population is in the interval $\mu \pm 2\sigma$

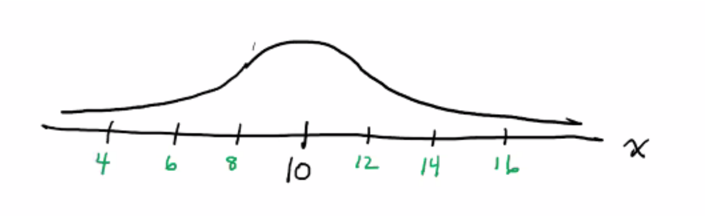
- About 99.7% of the population is in the interval $\mu \pm 3\mu$

There are infinitely many normal curves. The total area under a Normal curve must be 1.



The proportion of a normal population that list within a given interval is equal to the area under the normal probability density curve above that interval. Areas under a normal curve must be approximated numerically. We will use our calculators to approximate it. We will actually never use the normal pdf function on our calculators.

Example 61. Suppose X is normally distributed with a $\mu = 10$ and $\sigma^2 = 4$. Then, $X \sim N(10, 4)$. So, remember the standard deviation is $\sigma = \sqrt{4} = 2$.



1. Find the probability $P(X > 12.5)$.

$$\begin{aligned} P(X > 12.5) &= \text{normcdf}(12.5, \infty, 10, 2) \\ &\approx 0.1056 \end{aligned}$$

2. Find the probability $P(9 \leq X \leq 10)$.

$$\begin{aligned} P(9 \leq X \leq 10) &= \text{normcdf}(9, 10, 10, 2) \\ &\approx 0.1915 \end{aligned}$$

We can find the z-score for $x = 9$ and $x = 10$.

$$\begin{aligned} x &= 9 \\ z &= \frac{9 - 10}{2} \\ &= -0.5 \end{aligned}$$

This means that $x = 9$ is half a standard deviation below the mean.

$$\begin{aligned} x &= 10 \\ z &= \frac{10 - 10}{2} \\ &= 0 \end{aligned}$$

3. Find the value of c such that the probability $P \leq c = 0.75$. (Find the 75th percentile).

   We can use the inverse normal function to find the upper bound c. The inverse normal function gives you the value c with the area p to its left under the normal curve.

$$\begin{aligned} c = 0.75 &= \text{invNorm}(0.75, 10, 2) \\ &\approx 11.349 \end{aligned}$$

   We can also do this by using the z-score. Recall:

$$\begin{aligned} z &= \frac{x - \mu}{\sigma} \\ x &= \mu + z\sigma \end{aligned}$$

   So we can find the 75th percentile for the z-values:

$$\begin{aligned} z_{75} &= \text{invNorm}(0.75, 0, 1) \\ &\approx 0.67 \end{aligned}$$

   So for our distribution in terms of x, we can see that the 75th percentile is 0.67 standard deviations above the mean.

$$\begin{aligned} x_{75} &= 10 + 0.67(2) \\ &= 11.34 \end{aligned}$$

8.9.1 Standard units and the Standard Normal Distribution

The proportion of a normal population that is within a given number of standard deviation of the mean is the same for any normal population. For this reason, when dealing with normal populations we often convert from units in which the population items were originally measured to **standard units**. Standard units tell us how many standard deviations an observation is from the population mean.

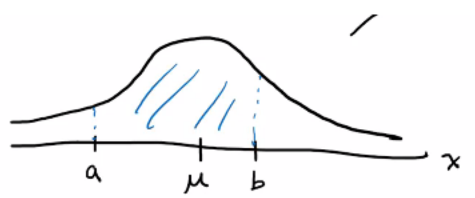
If x is an item sampled from a normal population with mean μ and variance σ^2, the standard unit is equivalent of x is the number z, where

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

The number z is sometimes called the z-score of x. The z-score is an item sampled from a normal population with mean 0 and standard deviation 1. This normal population is called the **standard normal population**.

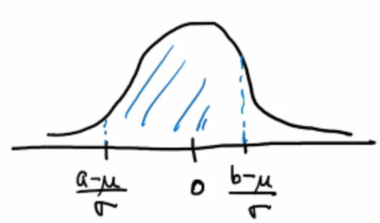
The standard normal distribution has a mean of $\mu = 0$ and a standard deviation of $\sigma = 1$. We write this as $Z \sim N(0, 1^2)$.

Example 62. Lets say we have $N(\mu, \sigma^2)$.



We can apply the z variable and come up with this curve. Where

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



These 2 curves will have the same area in the interval.

$$P(a < X < b) = P\left(\frac{a - \mu}{\sigma} < Z < \frac{b - \mu}{\sigma}\right)$$

Example 63. Lifetimes of batteries in a certain application are normally distributed with mean 50 hours and standard deviation of 5 hours.

1. Find the probability that a randomly chosen batter lasts between 42 and 52 hours.

   Let X = lifetime of a randomly chosen battery. Then $X \sim N(50, 5^2)$.

   So we are looking for

$$\begin{aligned} P(42 < X < 52) &= \text{normalcdf}(42, 52, 50, 5) \\ &\approx 0.6006 \end{aligned}$$

2. Find the 40th percentile of battery lifetimes.

$$\begin{aligned} X_{40} &= \text{invNorm}(0.4, 50, 5) \\ &\approx 48.73 \end{aligned}$$

Example 64. A process manufactures ball bearings whose diameters are normally distributed with mean 2.505 cm and a standard deviation of 0.008 cm. Specifications call for the diameter to be in the interval 2.5 ± 0.01 cm.

1. What proportion of the ball bearings will meet the specifications?

   Let X = the diameter of a randomly chosen ball bearing. So, $X \sim N(2.505, 0.008^2)$.

   The ball bearing diameters must be in the range from 2.49 to 2.51.

$$\begin{aligned} P(2.49 < X < 2.51) &= \text{normalcdf}(2.49, 2.51, 2.505, 0.008) \\ &\approx 0.7036 \end{aligned}$$

2. The process can be recalibrated so that the mean will equal 2.5 cm, the center of the specification interval. The standard deviation of a process remains 0.008 cm. What proportion of the diameter will meet specifications?

   We are looking for $P(2.49 < X < 2.51)$ but our new mean is $\mu_X = 2.5$ cm.

$$\begin{aligned} P(2.49 < X < 2.51) &= \text{normalcdf}(2.49, 2.51, 2.5, 0.008) \\ &\approx 0.7887 \end{aligned}$$

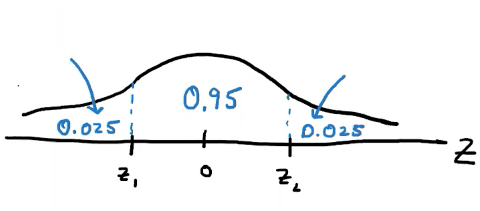
3. Assume that the process has been recalibrated so that the mean diameter is now 2.5 cm. To what value must the standard deviation be lowered so that 95% of the diameters will meet the specifications?

   Now we assume $X \sim N(2.5, 0.008^2)$. We are looking for σ so that $P(2.49 < X < 2.51) = 0.95$. We have to use the standardized version of x:

$$Z = \frac{x - \mu}{\sigma} = \frac{x - 2.5}{0.008}$$

Here we are looking for σ.

$$Z = \frac{x - 2.5}{\sigma}$$



$$\begin{aligned} Z_1 &= \text{invNorm}(0.025, 0, 1) \\ &\approx -1.96 \\ Z_2 &= \text{invNorm}(0.975, 0, 1) \\ &\approx 1.96 \end{aligned}$$

So we can apply it to the Z equation above using $x = 2.51$ and $Z = 1.96$:

$$\begin{aligned} 1.96 &= \frac{2.51 - 2.5}{\sigma} \\ \sigma &\approx 0.005 \end{aligned}$$

8.9.2 The Distribution of the Sample Mean and the Central Limit Theorem

The case of a normal population distribution:

Let $X_1, \ldots, X_n$ be a simple random sample from a *normal* distribution with mean μ and variance σ^2.

Let $\bar{X} = \frac{X_1 + \cdots + X_n}{n}$ be the sample mean.

Let $S_n = X_1 + \cdots + X_n$ be the sum of the sample observations.

Then, for any n:

$$\bar{X} \text{ is normally distributed with mean } \mu \text{ and standard deviation } \frac{\sigma}{\sqrt{n}}$$
$$S_n \text{ is normally distributed with mean } n\mu \text{ and standard deviation } \sqrt{n}\sigma$$

That is,

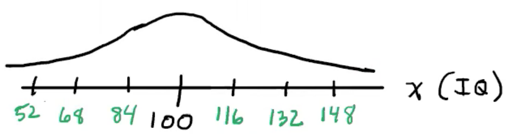
$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right) \qquad\qquad S_n \sim N(n\mu, n\sigma^2)$$

Example 65. Intelligence quotients (IQs) measured on the Standford Revision of the Binet-Simon Intelligence Scale are normally distributed with a mean of 100 and a standard deviation of 16.

1. Determine the probability that a randomly selected person has an IQ between 90 and 110.

   Let X = the IQ of a person. The mean is $\mu = 100$ and the standard devitation $\sigma = 16$. So, $X \sim N(100, 16^2)$.

$$\begin{aligned} P(90 < X < 110) &= \text{normalcdf}(90, 110, 100, 16) \\ &\approx 0.4680 \end{aligned}$$



2. Determine the probability that the mean IQ of 4 randomly selected people will be between 90 and 110.

   Let $X_1, \ldots, X_4$ be the IQx of the 4 people in the sample. $\bar{X} = \frac{X_1 + \cdots + X_4}{4}$ is the mean IQ of the sample of 4 people. So we are looking $P(90 < \bar{X} < 110)$. So since $X \sim N(100, 16^2)$, we know that $\bar{X} \sim N(100, \frac{16^2}{4})$. That is, $\bar{X} \sim N(100, 64)$.

$$\begin{aligned} P(90 < \bar{X} < 110) &= \text{normalcdf}(90, 110, 100, 8) \\ &\approx 0.7887 \end{aligned}$$

   Standardized version of $\bar{X}$:

$$\begin{aligned} Z &= \frac{\bar{X} - \mu_{\bar{X}}}{\sigma_{\bar{X}}} \\ &= \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \\ &= \frac{\bar{X} - 100}{\frac{16}{\sqrt{4}}} \\ &= \frac{\bar{X} - 100}{8} \end{aligned}$$

   So when $\bar{X} = 90$, $Z = -1.25$. When $\bar{X} = 110$, $Z = 1.25$.

   So if we were doing this problem in standard units it would look like:

$$\begin{aligned} P(90 < \bar{X} < 110) &= P(-1.25 < Z < 1.25) \\ &= \text{normalcdf}(-1.25, 1.25, 0, 1) \\ &\approx 0.7887 \end{aligned}$$

8.9.3 The Central Limit Theorem

Definition 37. Let $X_1, \ldots, X_n$ be a simple random sample from a population with mean μ and variance σ^2.

Let $\bar{X} = \frac{X_1+\cdots+X_n}{n}$ be a sample mean.

Let $S_n = X_1 + \cdots + X_n$ be the sum of the sample observations.

Then if n is sufficiently large,

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right) \text{ approximately}$$

and

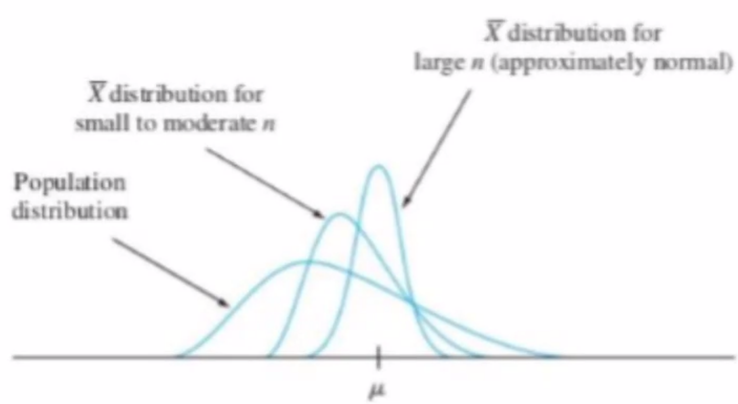
$$S_n \sim N(n\mu, n\sigma^2) \text{ approximately}$$

The larger the value of n, the better the approximation.

Note 12. If the sample is drawn from a nearly symmetric distribution, the normal approximation can be good even for a fairly small value of n. However, if the population is heavily skewed, a fairly large n may be necessary. For most populations, if the sample size n is greater than 30, the Central Limit Theorem approximation is good.

The Central Limit Theorem tells us that if we draw a large enough sample from a population, then the distribution of the sample mean is at least approximately normal, no matter what population the sample was drawn from.

Having the normal population is not required.



Example 66. At a large university, the mean age of the students is 22.3 years, and the standard deviation is 4 years. A random sample of 64 students is drawn.

1. What is the probability that the average age of these students is greater than 23 years?

   Our population is all of the students. Our population variable X = the age of a student. We know the mean is $\mu_X = 22.3$ years. Our standard deviation is $\sigma_X = 4$ years. However, we do not know how this population is distributed. Our sample size is $n = 64$ students. We can let $X_1, \ldots, X_{64}$ represent the ages of the 64 students in the sample. Then we can define an average age (a sample mean age) as $\bar{X} = X_1 + \cdots + X_{64}$. That is, $\bar{X}$ = the mean age of a sample of 64 students. By the central limit theorem, since $n = 64$ and $64 > 30$, we can say that $\bar{X} \sim N\left(22.3, \frac{4^2}{64}\right)$, at least approximately.

$$\begin{aligned} P(\bar{X} > 23) &= \text{normalcdf}(23, \infty, 22.3, \sqrt{0.25}) \\ &\approx 0.0808 \end{aligned}$$

2. How many students must be sampled so that the probability is 0.05 that the sample mean age exceeds 23?

   So we are looking for the sample size n so that $P(\bar{X} > 23) = 0.05$. We can use the standardized values to find this (the z-score).

$$\begin{aligned} Z &= \text{invNorm}(0.95, 0, 1) \\ &\approx 1.645 \end{aligned}$$

   So using $Z = 1.645$, $\bar{X} = 23$, $\mu = 22.3$ and $\sigma = 4$

$$\begin{aligned} 1.645 &= \frac{23 - 22.3}{\frac{4}{\sqrt{n}}} \\ \sqrt{n} &= \frac{1.645(4)}{0.7} \\ n &= 88.36 \end{aligned}$$

   So to guarantee we have a 0.05 probability of having sample mean age we need to sample 89 students.

Example 67. The manufacturer of a certain part requires two different machine operations. The time on machine 1 has a mean of 0.4 hours and a standard deviation of 0.1 hours. The time on machine 2 has a mean of 0.45 hours and a standard deviation of 0.15 hours. The times needed on the machines are independent. Suppose that 65 parts are manufactured.

1. What is the distribution of the total time on machine 1?

   Let X = time on machine 1, so the mean is $\mu_X = 0.4$ and standard deviation $\sigma_X = 0.1$.

   Let $X_1, \ldots, X_{65}$ be the times of the 65 parts on machine 1. We can call the sum of the times, the total time on machine 1, $S_X = X_1 + \cdots + X_{65}$. Remember, our mean $\mu = 65(0.4)$ and our standard deviation is $\sigma = 65(0.1)$.

   Remember

$$\mu_{S_x} = \mu_{x_1 + \cdots + x_{65}} = \mu_{x_1} + \cdots + \mu_{x_{65}}$$

   and since there is 65 means, that it why it equals 65(0.4).

   By the central limit theorem, since our sample size $n = 65 > 30$, we can say that $S_X \sim N(26, 0.65)$ approximately.

2. What is the probability that the total time on machine 1 is less than 27 hours?

$$\begin{aligned} P(S_x < 27) &= \text{normalcdf}(-\infty, 27, 26, \sqrt{0.65}) \\ &\approx 0.8926 \end{aligned}$$

3. What is the distribution of the total time on machine 2?

   Let Y = time on machine 2, so the mean is $\mu_Y = 0.45$ and the standard deviation $\sigma_Y = 0.15$.

   Let $Y_1, \ldots, Y_{65}$ be the times of the 65 parts on machine 2. We can call the sum of the times, the total time on machien 2, $S_Y = Y_1 + \cdots + Y_{65}$.

   So by the central limit theorem, the distribution of the time on machine 2 is $S_Y \sim N(65(0.45), 65(0.15)^2)$ approximately. This can be cleaned up as $S_Y \sim N(29.25, 1.4625)$.

4. What is the probability that the total time, used by both machines together, is between 50 and 55 hours?

   Let T = total time used by both machines. So

$$T = \underbrace{S_X}_{\text{time on machine 1}} + \overbrace{S_Y}^{\text{time on machine 2}}$$

   The mean of the total time is

$$\begin{aligned} \mu_T &= \mu_{S_X} + \mu_{S_Y} \\ &= 26 + 29.25 \\ &= 55.25 \text{ hours} \end{aligned}$$

   The variance of the total time is

$$\begin{aligned} \sigma_T^2 &= \sigma_{S_X}^2 + \sigma_{S_Y}^2 \\ &= 0.65 + 1.4625 \\ &= 2.1125 \end{aligned}$$

   So we can say $T \sim N(55.25, 2.1125)$ approximately. So we can find the probability that the time is between 50 and 55 hours

$$\begin{aligned} P(50 < T < 55) &= \text{normalcdf}(50, 55, 55.25, \sqrt{2.1125}) \\ &\approx 0.4316 \end{aligned}$$

5. Find the probability that the total time on machine 2 is greater than the total time on machine 1.

   Recall

$$S_X \sim N(26, 0.65) \qquad S_Y \sim N(29.25, 1.4625)$$

We are looking for $P(S_Y > S_X)$. Let $D = S_Y - S_X$

$$\begin{aligned} P(S_Y > S_X) &= P(S_Y - S_X > 0) \\ &= P(D > 0) \end{aligned}$$

We can find the difference in means

$$\begin{aligned} \mu_D &= \mu_{S_Y} - \mu_{S_X} \\ &= 29.25 - 26 \\ &= 3.25 \end{aligned}$$

and the different in standard deviations

$$\begin{aligned} \sigma_D^2 &= \sigma_{S_Y}^2 - \sigma_{S_X}^2 \\ &= 2.1125 \end{aligned}$$

So we can say that the different has a distribution of $D \sim N(3.25, 2.1125)$ approximately. So,

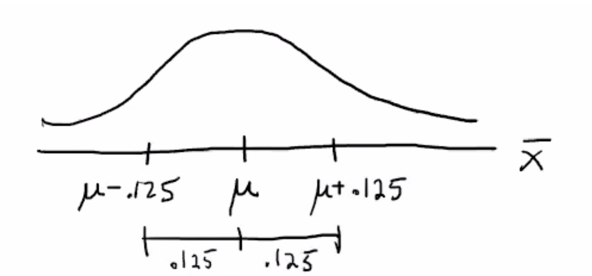
$$\begin{aligned} P(D > 0) &= \text{normalcdf}(0, \infty, 3.25, \sqrt{2.1125}) \\ &\approx 0.9873 \end{aligned}$$

Example 68. Birth weights of male babies have a standard deviation of 1.33 lb. Determine the percentage of all samples of 400 male babies that have mean birth weights within 0.125 lb (2 oz.) of the population mean birth weight of all male babies.

Here the population consists of all male babies. Let X = birth weight of a male babie in pounds. The population mean μ_X is unknown, however the standard deviation is $\sigma_X = 1.33$ lbs. We have a sample size of $n = 400$.

Let $X_1, \ldots, X_{400}$ be the birth weights of each of 400 male babies. The mean birth weight of the sample is $\bar{X} = \frac{X_1 + \cdots + X_{400}}{400}$. By the central limit theorem, since the sample size is large enough, we can describe the distribution of all possible sample means is at least approximately normal with mean $\mu_{\bar{X}} = \mu$ and standard deviation $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} = \frac{1.33}{\sqrt{400}} = 0.0665$.

That is, $\bar{X} \sim N(\mu, 0.0665^2)$.



So we are looking for the area under the curve between $\mu \pm 0.125$. We can find this by using standardized units (converting to z-score).

$$
\begin{aligned}
Z &= \frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \\
&= \frac{\bar{X}-\mu}{\frac{1.33}{\sqrt{400}}} \\
&= \frac{\bar{X}-\mu}{0.0665}
\end{aligned}
$$

So we can plug this in

$$
\begin{aligned}
P(\mu-0.125<\bar{X}<\mu+0.125) &= P\left(\frac{(\mu-0.125)-\mu}{0.0665}<Z<\frac{(\mu+0.125)-\mu}{0.0665}\right) \\
&= P\left(\frac{-0.125}{0.0665}<Z<\frac{0.125}{0.0665}\right) \\
&= P(-1.88<Z<1.88) \\
&= \text{normalcdf}(-1.88, 1.88, 0, 1) \\
&\approx 0.9399
\end{aligned}
$$

So 93.99% of all samples of all 400 male babies are within 0.125 lbs of the sample mean. In terms of sample error: there is a 93.99% chance that the sampling error made in estimating the mean birth weight of all male babies by that of a sample of 400 male babies will be at most 0.125 lbs.

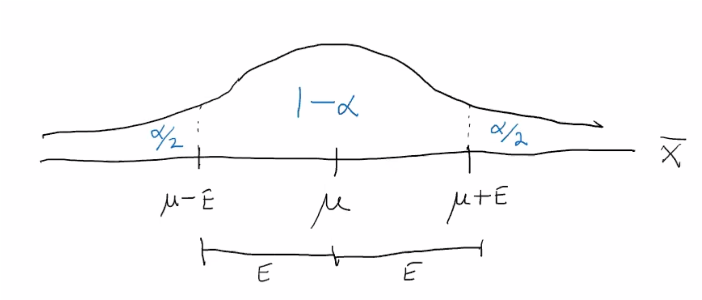
9 Confidence Intervals

Some useful definitions:

- A **parameter** is a descriptive measure for a population. For example, a population mean μ and a standard deviation σ are parameters.
- A **statistic** is a descriptive measure for a sample from a population. For example, a sample mean $\bar{X}$ and a sample standard deviation s are statistics.
- A **sampling error** is the error resulting from using a sample to estimate a population characteristic. For example, (informally) when a particular sample mean $\bar{X}$ is used to estimate a population μ, there is a certain amount of error E involved. Usually shown as

$$\mu = \bar{X} \pm E$$
$$|\bar{X} - \mu| = E$$

- A **point estimate** of a parameter is the value of a statistic used to estimate the value of the parameter. For example, a particular value of a sample mean $\bar{X}$ is often used as a "point estimate" of a mean μ.
- A **confidence interval** estimate for a paramter (like μ) provides a range of numbers along with a percentage confidence that the parameter lies in that range. Suppose we have a distribution of $\bar{X}$ values of size n, and suppose it is normally distributed



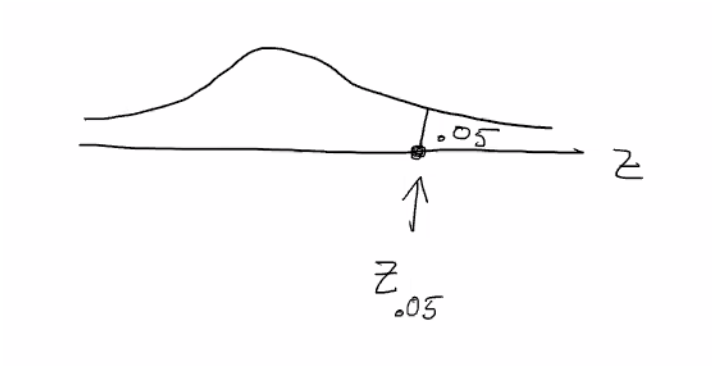
where $\frac{\alpha}{2} + \frac{\alpha}{2} = \alpha$. So $100(1-\alpha)\%$ of all sample means of size n have sample means $\bar{X}$ in the interval $(\mu - E, \mu + E)$. Equivalently, for samples of size n, $100(1-\alpha)\%$ of all intervals of the form $(\bar{X} - E, \bar{X} + E)$ will cover (or contain) μ.

Aside 3. Z_α notation (α is a proportion)

Z_α is the z-score with area α to its right under the standard normal curve.

Example 69. Find $Z_{0.05}$.

This is the z-score with the area of 0.05 area to its right under the standard normal curve.



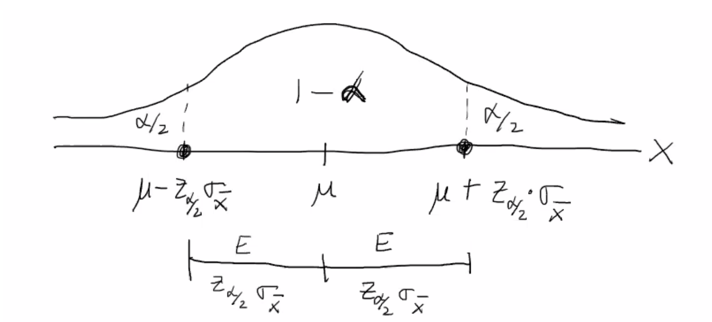
We could either calculate the area to the left of the z-score, 0.95

$$\text{invNorm}(0.95, 0, 1) \approx 1.645$$

We could also use symmetry and look for $-Z_{0.05}$

$$\begin{aligned} -Z_{0.05} &= \text{invNorm}(0.05, 0, 1) \\ &\approx -1.645 \\ &= 1.645 \end{aligned}$$

So $\mu \pm E$ can be written as $\mu \pm Z_{\frac{\alpha}{2}} \cdot \sigma_{\bar{X}}$.



Here E is called the "margin of error" for the confidence interval estimate of μ.

$$\begin{aligned} E &= Z_{\frac{\alpha}{2}} \sigma_{\bar{X}} \\ &= Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \end{aligned}$$

Definition 38. Let $X_1, \ldots, X_n$ be a random sample from a *normal* population with mean μ and a standard deviation σ so that $\bar{X}$ has a normal distribution. A $100(1-\alpha)\%$ **confidence interval** for the mean μ of a **normal** population when the value of σ is known is given by

$$\bar{X} \pm Z_{\frac{\alpha}{2}} \sigma_{\bar{X}} \qquad \text{where } \sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$$

For example, a level $100(1-\alpha)\%$ confidence interval for μ is given by

$$(\bar{X} - Z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}, \bar{X} + Z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}})$$

Definition 39. Large-Sample Confidence Intervals for a Population Mean

Let $X_1, \ldots, X_n$ be a *large* $(n > 30)$ random sample from a population with mean μ and a standard deviation σ, so that $\bar{X}$ is approximately normal. Then a level $100(1-\alpha)\%$ confidence interval for μ is given by

$$\bar{X} \pm Z_{\frac{\alpha}{2}} \sigma_{\bar{X}} \qquad \text{where } \sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$$

When the value is σ is unknown ans the sample size is large enough, σ can be replaced with the sample standard deviation s.

Definition 40. One Sided Confidence Intervals (Confidence Bounds)

Let $X_1, \ldots, X_n$ be a *large* $(n > 30)$ random sample from a population with mean μ and a standard deviation σ, so that $\bar{X}$ is approximately normal. Then a level $100(1-\alpha)\%$ lower confidence bound for μ is given by

$$\bar{X} - Z_{\alpha} \sigma_{\bar{X}}$$

and level $100(1-\alpha)\%$ upper confidence bound for μ is given by

$$\bar{X} + Z_{\alpha} \sigma_{\bar{X}}$$

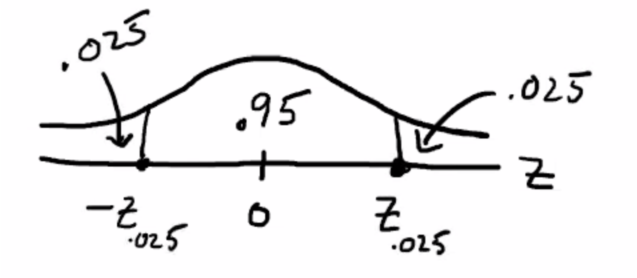
where $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$. When the value of σ is unknown and the sample size is large enough, σ can be replaced with the sample standard deviation s.

Example 70. In a sample of 50 microdrills drilling a low-carbon alloy steel, the average lifetime (expressed as the number of holes drilled before failure) was 12.68 with a standard deviation of 6.83.

1. Find a 95% confidence interval for the mean lifetime of microdrills under these conditions.

   The population here is all microdrills, with a sample of $n = 50$ microdrills. So we have a SRS of

$X_1, \ldots, X_{50}$ of lifetimes of the 50 microdrills sample. Our sample mean is $\bar{X} = 12.68$. Our sample standard deviation is $s = 6.83$. We are looking to solve the confidence interval $\bar{X} \pm Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}$.



$$\begin{aligned} -Z_{0.025} &= \text{invNorm}(0.025, 0, 1) \\ &\approx -1.96 \\ Z_{0.025} &\approx 1.96 \end{aligned}$$

So we can solve

$$\begin{aligned} \bar{X} - Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} &= 12.68 - 1.96 \left(\frac{\sigma}{\sqrt{50}} \right) \\ &= 12.68 - 1.96 \left(\frac{6.83}{\sqrt{50}} \right) \\ &\approx 10.79 \end{aligned}$$

The right endpoint

$$\begin{aligned} \bar{X} + Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} &= 12.68 + 1.96 \left(\frac{6.83}{\sqrt{50}} \right) \\ &\approx 14.57 \end{aligned}$$

So our 95% confidence interval would be $(10.79, 14.57)$. "Using samples of size 50, we are 95% confident that the mean lifetime of all microdrills is somewhere between 10.79 and 14.57 holes."

2. Find an 80% confidence interval for the mean lifetime of microdrills under these conditions.

   Using our same values, $n = 50$, $\bar{X} = 12.68$, $s = 6.83$, but we are looking for the confidence level of 0.8.

   Our z-score will be

$$\begin{aligned} -Z_{0.1} &= \text{invNorm}(0.1, 0, 1) \\ &\approx -1.28 \\ Z_{0.1} &= 1.28 \end{aligned}$$

$$\begin{aligned}\bar{X} - Z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} &\approx 12.68 - (1.28)\left(\frac{6.83}{\sqrt{50}}\right) \\ &\approx 11.44 \\ \bar{X} + Z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} &\approx 12.68 + (1.28)\left(\frac{6.83}{\sqrt{50}}\right) \\ &\approx 13.92\end{aligned}$$

So our confidence interval for 80% confidence is $(11.44, 13.92)$.

"The wider the confidence interval, the more confident we are that the mean exists in the interval." The margin of error $E = Z_{\frac{\alpha}{2}}\sigma_{\bar{X}} = Z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}$

For the bigger confidence interval

$$\begin{aligned}E &= 1.96\left(\frac{6.83}{\sqrt{50}}\right) \\ &\approx 1.893\end{aligned}$$

For the smaller confidence interval

$$\begin{aligned}E &= 1.28\left(\frac{6.83}{\sqrt{50}}\right) \\ &\approx 1.236\end{aligned}$$

Remember that the sample mean $\bar{X}$ is the midpoint of each confidence interval.

3. How many microdrills must be sampled to obtain a 95% confidence interval that specifies the mean to with ± 0.5?

   Recall, sample mean $\bar{X} = 12.68$, sample standard deviation $s = 6.83$.

   We are looking for n so that $E = 0.5$. Our z-score is $Z = 1.96$. Since we don't know σ we can use the sample standard deviation s and solve for n.

$$\begin{aligned}E &= Z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}} \\ 0.5 &= 1.96\left(\frac{6.83}{\sqrt{n}}\right) \\ n &\approx 716.8\end{aligned}$$

   We will round up to the nearest integer so our sample size $n = 717$ to have a 95% confidence interval.

4. Based on microdrill lifetime data presented above, an engineer reported a confidence interval of (11.09, 14.27) but neglected to specify the level. What is the level of this confidence interval?

   Recall, $n = 50$, $\bar{X} = 12.68$, $s = 6.83$.

$$\bar{X} \pm Z_{\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} = 12.68 \pm Z_{\frac{\alpha}{2}} \left(\frac{6.83}{\sqrt{50}}\right)$$

$$\begin{aligned} 11.09 &= 12.68 - Z_{\frac{\alpha}{2}} \left(\frac{6.83}{\sqrt{50}}\right) \\ 14.27 &= 12.68 + Z_{\frac{\alpha}{2}} \left(\frac{6.83}{\sqrt{50}}\right) \end{aligned}$$

Lets chose the second one and solve for $Z_{\frac{\alpha}{2}}$.

$$\begin{aligned} 14.27 &= 12.68 + Z_{\frac{\alpha}{2}} \left(\frac{6.83}{\sqrt{50}}\right) \\ Z_{\frac{\alpha}{2}} &= \frac{14.27 - 12.68}{\frac{6.83}{\sqrt{50}}} \\ &\approx 1.646 \end{aligned}$$

We can use a normal cdf to find our confidence interval

$$\text{normalcdf}(-1.646, 1.646, 0, 1) \approx 0.9002$$

5. Find both 95% lower confidence bound and a 99% upper confidence bound for the mean lifetime of the microdrills.

   Recall, $n = 50$, $\bar{X} = 12.68$, $s = 6.83$. For the 95% lower we can solve with a z-score for 95% using $Z = \text{invNorm}(0.95, 0, 1) \approx 1.645$.

$$\begin{aligned} \bar{X} - Z_{\alpha} \frac{\sigma}{\sqrt{n}} &\approx \bar{X} - Z_{\alpha} \frac{s}{\sqrt{n}} \\ &= 12.68 - Z_{\alpha} \frac{6.83}{\sqrt{50}} \\ &= 12.68 - 1.645 \left(\frac{6.83}{\sqrt{50}}\right) \\ &\approx 11.09 \end{aligned}$$

   Since this is a lower bound, the interval would be $(11.09, \infty)$.

   For the 99% upper bound we can solve with a z-score of 99% using $Z = \text{invNorm}(0.99, 0, 1) \approx 2.326$.

$$\begin{aligned}\bar{X} + Z_\alpha \frac{\sigma}{\sqrt{n}} &\approx \bar{X} + Z_\alpha \frac{s}{\sqrt{n}} \\ &= 12.68 + Z_\alpha \frac{6.83}{\sqrt{50}} \\ &= 12.68 + 2.326 \left(\frac{6.83}{\sqrt{50}}\right) \\ &\approx 14.93\end{aligned}$$

Since this ia a upper bound, the interval for 99% confidence would be $(-\infty, 14.93)$.

Confidence, accuracy (precision) and sample size:

- For a fixed sample size, decreasing the confidence level decreases the margin of error, and, hence, improves the accuracy (precision) of a confidence interval estimate.
- For a fixed confidence level, increasing the sample size decreases the margin of error, and, hence, improves the accuracy (precision) of a confidence interval estimate.

9.1 Confidence intervals for population proportions

Recall that if $X \sim Bin(n,p)$, then the estimate for p is $\hat{p} = \frac{X}{n}$. When the sample size is large enough ($n > 30$) it follows that from the Central Limit Theorem that $\hat{p}$ is at least approximately normally distributed. That is $\hat{p} \sim N\left(p, \frac{p(1-p)}{n}\right)$ approximately, where the mean of $\hat{p}$ is p, and the uncertainty (standard deviation) of $\hat{p}$ is $\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$.

A traditional $100(1-\alpha)\%$ confidence interval for $\hat{p}$ would be in the form

$$\hat{p} \pm Z_{\frac{\alpha}{2}} \sigma_{\hat{p}}$$

Definition 41. Let X be the number of successes in n independent Bernoulli trials with a success probability p, so that $X \sim Bin(n,p)$.

Define $\tilde{n} = n + 4$ and $\tilde{p} = \frac{X+2}{\tilde{n}}$.

Then, a level $100(1-\alpha)\%$ **confidence interval for** p is given by

$$\tilde{p} \pm Z_{\alpha/2} \sqrt{\frac{\tilde{p}(1-\tilde{p})}{\tilde{n}}}$$

Definition 42. For one sided intervals:

Let X be the number of successes in n independent Bernoulli trials with a success probability p, so that $X \sim Bin(n, p)$.

Define $\tilde{n} = n + 4$ and $\tilde{p} = \frac{X+2}{\tilde{n}}$. Then a level $100(1-\alpha)\%$ lower bound for p is given by

$$\tilde{p} - Z_{\alpha/2}\sqrt{\frac{\tilde{p}(1-\tilde{p})}{\tilde{n}}}$$

and a $100(1-\alpha)\%$ upper bound for p is given by

$$\tilde{p} + Z_{\alpha/2}\sqrt{\frac{\tilde{p}(1-\tilde{p})}{\tilde{n}}}$$

Note 13. If the lower limit is less than 0, replace it with 0. If the upper limit is greater than 1, replace it with 1.

Definition 43. The traditional method for computing Confidence Intervals for a proportion (widely used but not recommended).

Let $\hat{p}$ be the proportion of successes in a *large* number n of independent Bernoulli trials with success probability p. Then the traditional level $100(1-\alpha)\%$ confidence interval for p is given by

$$\hat{p} \pm Z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

The method cannot be used unless the sample contains at least 10 successes and 10 failures. (Both $n\hat{p}$ and $n\hat{p}(1-\hat{p})$ should be greater than 10.)

Example 71. An anterior cruciate ligament (ACL) runs diagonally through the middle of the knee. A research article reported results for 85 young athletes who suffered ACL injuries. Of the 85 injuries, 51 were to the left knee and 34 were to the right knee.

1. Find a 90% confidence interval for the proportion of ACL injuries that are to the left knee.

   We know $n = 85$ injuries (our number of trials). Our number of successes is (an injury to the left knee), $X = 51$. So we can find $\tilde{n}$ and $\tilde{p}$ by

$$\begin{aligned}\tilde{n} &= 85 + 4\\ &= 89\\ \tilde{p} &= \frac{51+2}{89}\\ &\approx 0.5955\end{aligned}$$

   Next we need to find $Z_{\alpha/2}$

$$
\begin{aligned}
1-\alpha &= 0.9 \\
\alpha &= 0.1 \\
\frac{\alpha}{2} &= 0.05 \\
\text{invNorm}(0.95, 0, 1) &\approx 1.645
\end{aligned}
$$

So for our confidence interval

$$
\begin{aligned}
\tilde{p} - Z_{\alpha/2}\sqrt{\frac{\tilde{p}(1-\tilde{p})}{\tilde{n}}} &= 0.5955 - 1.645\sqrt{\frac{0.5955(0.4045)}{89}} \\
&\approx 0.5099 \\
\tilde{p} + Z_{\alpha/2}\sqrt{\frac{\tilde{p}(1-\tilde{p})}{\tilde{n}}} &= 0.5955 + 1.645\sqrt{\frac{0.5955(0.4045)}{89}} \\
&\approx 0.6811
\end{aligned}
$$

So the 90% confidence interval is from $(0.5099, 0.6811)$. “We can be 90% confidence that the true proportion p of ACL injuries to the left knee for all athletes that suffered ACL injuries is somewhere between 0.5099 and 0.6811.”

2. What sample size is needed to obtain a 95% confidence interval with width ± 0.08?

   Our margin of error is $E = Z_{\alpha/2}\sqrt{\frac{\tilde{p}(1-\tilde{p})}{\tilde{n}}}$. Our z-score is $Z = 1.96$. We are looking for n

$$
\begin{aligned}
0.08 &= 1.96\sqrt{\frac{0.5955(0.4045)}{n+4}} \\
n &\approx 140.6
\end{aligned}
$$

   So we round up to the next whole number and our sample size is $n = 141$.

3. How large a sample is needed to guarantee that thw width of the 95% confidence interval will be no greater than ± 0.08 if no preliminary sample has been taken? Note: Here, we can use the fact that the quantity $\tilde{p}(1-\tilde{p})$, which determines the width of the confidence interval, is maximized for $\tilde{p} = 0.5$. Since the width is greatest when $\tilde{p}(1-\tilde{p})$ is greatest, we can compute a conservative sample size estimate by substituting $\tilde{p} = 0.5$ and proceeding as in part 2 above.

   Aside 4.

$$
\begin{aligned}
\sigma_{\tilde{p}} &= \sqrt{\frac{\tilde{p}(1-\tilde{p})}{\tilde{n}}} \\
f(x) &= x - x^2
\end{aligned}
$$

   This parabola has a max of 0.5, which is the idea behind why $\tilde{p} = 0.5$.

   We can use $Z = 1.96$, and $\tilde{p} = 0.5$, and setup the problem like this to solve for n

$$0.08 = 1.96\sqrt{\frac{0.5(1-0.5)}{n+4}}$$
$$n = 146.0625$$

So our $n = 147$.

9.2 t-distributions

Definition 44. If X is a normally distributed variable with mean μ and standard deviation σ, then, for sample size of n, the variable $\bar{X}$ is also normally distributed with mean μ and standard deviation $\frac{\sigma}{\sqrt{n}}$. Equivalently, the **standardized version** of $\bar{X}$

$$Z = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

has the standard normal distribution.

On the other hand, when the population standard deviation σ is unknown, the best we can do is estimate it by a sample standard deviation s. When σ is replaced with s, the resulting variable

$$t = \frac{\bar{X} - \mu}{\frac{s}{\sqrt{n}}}$$

is called the **studentized version** of $\bar{X}$.

The Student's t-distribution has $n-1$ degrees of freedom ($df = n - 1$).

A variable with t-distribution has an associated curve, called the t-curve. Although there is a different t-curve for each number of degrees of freedom, all t-curves are similar and resemble the standard normal curve. A t-curve with ν degrees of freedom is sometimes denoted by t_ν.

Its worth noting that

- The total area under a t-curve equals 1.
- A t-curve extends indefinitely in both directions, approaching, but never touching, the horizontal axis as it does so.
- A t-curve is symmetric about 0.
- As the numbers of degrees of freedom becomes larger, t-curves increasingly look like the standard normal curve.

t-curves have more spread than the standard normal curve. This property follows from that fact that, for a t-curve with ν degrees of freedom, where $\nu > 2$, the standard deviation is $\sqrt{\frac{\nu}{\nu-1}}$. This quantity always exceeds 1, which is the standard deviation of the standard normal curve.

Percentages (and probabilities) for a variable having a t-distribution equal areas under the variables associated t-curve.

Our goal will be to find the confidence intervals and for this we will need to be able to find t-values. The symbol t_α denotes the t-value having area α to its right under a t-curve. For this, you will need to use the "invT" function on your calculator (or a t-table). A t-value having area α to its right under a t-curve with $n-1$ degrees of freedom is sometimes denoted $t_{n-1,\alpha}$.

Note 14. When n is large, the distribution of $\frac{\bar{X}-\mu}{\frac{s}{\sqrt{n}}}$ is very close to normal, so the normal curve can be used instead of the Student's t. (We did this in the first Confidence Inteval section.)

Example 72. For a t-curve with 13 degrees of freedom, determine $t_{0.05}$.

We want to find $t_{13,0.05}$. We can use our calculator for this problem:

$$\begin{aligned} t_{n-1,\alpha} &= \text{invT}(1-\alpha, n-1) \\ &= \text{invT}(1-.05, 13) \\ &\approx 1.771 \end{aligned}$$

The α is a right tail area, but our calculators view it as a left tail area, hence the $1-\alpha$. Optionally you could also use symmetry.

Definition 45. Confidence intervals using the Student's t distribution.

Let $X_1, \ldots, X_n$ be a *small* random sample from a *normal* population with mean μ. Then a level $100(1-\alpha)\%$ confidence interval for μ is given by

$$\bar{X} \pm t_{n-1,\alpha/2} \frac{s}{\sqrt{n}}$$

Note 15. If σ is known, we use Z instead of t.

Definition 46. Sample mean variance and standard deviation.

Let $X_1, \ldots, X_n$ be a sample. The **sample mean** is

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

The **sample variance** is the quantity

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$$

An equivalent formula, which can be easier to compute, is

$$s^2 = \frac{1}{n-1} \left(\sum_{i=1}^{n} X_i^2 - n\bar{X}^2 \right)$$

The **sample standard deviation** is the quantity

$$s = \sqrt{\frac{1}{n-1} \sum_{i=1}^{2} (X_i - \bar{X})^2}$$

An equivalent formula, which can be easier to compute, is

$$s = \sqrt{\frac{1}{n-1} \left(\sum_{i=1}^{n} X_i^2 - n\bar{X}^2 \right)}$$

The sample standard deviation is the square root of the sample variance.

Example 73. A metallurgist is studying a new welding process. He manufactures five welded joins and measures the yield strength of each. The five values (in ksi) are 56.3, 65.4, 58.7, 70.1, and 63.9. Assume that these values are a random sample from an approximately normal distribution. Find a 95% confidence interval for the mean strength of weld made by this process.

We have a sample size of $n = 5$. Our X represents the yield strength. The first thing we can do is find the sum of the X and the sum of the X^2.

$$\begin{aligned}\sum X &= 56.3 + 65.4 + 58.7 + 70.1 + 63.9\\ &= 314.4\\ \sum X^2 &= 56.3^2 + 65.4^2 + 58.7^2 + 70.1^2 + 63.9^2\\ &= 19889.6\end{aligned}$$

Then we can plug those into the formula

$$\begin{aligned}\bar{X} &= \frac{\sum X}{n}\\ &= \frac{314.4}{5}\\ &\approx 62.88\end{aligned}$$

$$\begin{aligned}s^2 &= \frac{\sum X^2 - \frac{(\sum X)^2}{n}}{n-1}\\ &= \frac{19889.76 - \frac{314.4^2}{5}}{4}\\ &\approx 30.072\\ s &= \sqrt{30.072}\\ &\approx 5.484\end{aligned}$$

Now that we have $\bar{X}$ and s we just need to find t to find our confidence interval Our degrees of freedom $df = 5 - 1 = 4$. Confidence level is $1 - \alpha = .95$, so $\frac{\alpha}{2} = \frac{0.05}{2} = 0.025$.

$$\begin{aligned}t_{4,0.025} &= \text{invT}(0.975, 4)\\ &\approx 2.776\end{aligned}$$

So now to find the left and right endpoint of our interval

$$\bar{X} \pm t_{n-1,\alpha/2}\frac{s}{\sqrt{n}} = 62.88 \pm 2.776\left(\frac{5.484}{\sqrt{5}}\right)$$

So our interval is $(56.07, 69.688)$.

10 Hypothesis Testing

A **hypothesis** is a statement that something is true. In statistics, a hypothesis is a declaration, or a claim, in the form of mathematical statement, about the value of a specific population parameter (or about the values of several population characteristics).

A hypothesis test typically involves two hypotheses:

- A **null hypothesis** is denoted by H_0, is the hypothesis to be tested.
- The **Alternative hypothesis**, denoted by H_1, is the hypothesis to be considered as an alternative to the null hypothesis. It is the assertion that is contradictory to H_0.

A **test of hypothesis** is a method for using sample data to decide whether the null hypothesis should be rejected. The null hypothesis will be rejected in favor of the alternative hypothesis only if sample evidence suggests that H_0 is false. If the sample does not strongly contradict H_0, we will continue to believe in that plausibility of the null hypothesis. The two possible conclusions from a hypothesis-testing analysis are then we **reject** H_0, or **fail** H_0.

> **Note 16.** If we fail to reject H_0, this just means that H_0 is "plausible". In fact, H_0 may or may not actually be true.

In performing a hypothesis test, we essentially put the null hypothesis on trial. We begin by assuming that H_0 is true, just as we begin a trial by assuming a defendant to be innocent. The random sample provides the evidence. The hypothesis test involves measuring the strength of 0 and 1, called a P-value. The P-value measures the plausibility of H_0. The smaller the P-value, the stronger the evidence is against H_0. If the P-value is sufficiently small, we may be willing to abandon our assumption that H_0 is true and believe that H_1 instead. This is referred to as **rejecting** the null hypothesis.

In general, a hypothesis test is done by first computing the distribution $\bar{X}$ under the assumption that H_0 is true. This distribution is called the **null distribution** of $\bar{X}$.

Then the P-value is computed using an observed value of $\bar{X}$. The P-value is the probability, under the assumption that H_0 is true, of observing a value of $\bar{X}$ whose disagreement with H_0 is as least as great as that of the observed value of $\bar{X}$ from the random sample.

Steps in performing a hypothesis test:

1. Define H_0 and H_1.
2. Assume H_0 to be true.
3. Compute a **test statistic**. A test statistic is a statistic that is used to assess the strength of the evidence against H_0.
4. Compute the P-value of the test statistic. The P-value is the probability, assuming H_0 to be true, that the test statistic would have a value whose disagreement with H_0 is as great or greater than that actually observed. The P-value is also called the **observed significance level**.
5. State a conclusion about the strength of evidence against H_0.

Definition 47. Procedure used to perform a large-sample hypothesis test for a population mean.

Let $X_1, \ldots, X_n$ be a *large* ($n > 30$) sample from a population with mean μ and standard deviation σ.

To test a null hypothesis of the form

$$H_0 : \mu \leq \mu_0, \qquad \text{or} \qquad H_0 : \mu \geq \mu_0, \qquad \text{or} \qquad H_0 : \mu = \mu_0$$

- Compute the Z-score

$$Z = \frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$$

  If σ is unknown it may be approximated with s.

- Compute the P-value. The P-value is an area under the normal curve, which depends on the alternative hypothesis as follows:

$H_1 : \mu > \mu_0$	Area to the right of Z
$H_1 : \mu < \mu_0$	Area to the left of Z
$H_1 : \mu \neq \mu_0$	Sum of the areas in the tails cut off by Z and $-Z$

Example 74. A certain type of automobile engine emits a mean of 100 mg of oxides of nitrogen (NOx) per second at 100 horsepower. A modification to the engine design has been proposed that may reduce NOx emissions. A new design will be put into production if it can be demonstrated that its mean emission rate is less than 100 mg/s. A sample of 50 modified engines are built and tested. The sample mean NOx emissions is 92 mg/s, and the sample standard deviation is 21 mg/s.

Note that the population in this case consists of the emission rates from the engines that would be built if this modified design is put into production. The manufacturers are concerned that the modified engines might not reduce emissions at all, that is, that the population mean might be 100 or more. They want to know whether this concern is justified. The question, therefore, is this: Is it plausible that this sample, with its mean of 92, could have come from a population whose mean is 100 or more?

For the sample of 92, there are two possible interpretations of this observations:

1. The population mean is actually greater than or equal to 100, that the sample mean is lower than this only because of random variation from the population mean. Thus the emissions will not go down if this new design is put into production, and the sample is misleading.
2. The population mean is actually less than 100, and the sample mean reflects this fact. Thus the sample represents a real different, that can be expected if the new design is put into production.

How might a hypothesis test be performed for the above sample?

1. First we define H_0 and H_1.

$$H_0 : \mu \geq 100 \qquad\qquad H_1 : \mu < 100$$

   The manufacturers are hoping to demonstrate H_1.

2. Next we assume that H_0 is true. Take as the assumed value of μ the value closest to the alternate hypothesis H_1. Use $\mu = 100$, (this is the μ_0 in the procedure defined above, $\mu_0 = 100$). Note our sample size is $n = 50$, we have $\bar{X} = 92$, and we have $s = 21$. $\bar{X}$ = the mean NO_x emission for a sample of size 50 modified engines. By the Central Limit Theorem, (since $n > 30$), $\bar{X}$ is approximately normally distributed with mean $\mu_{\bar{X}} = \mu = 100$ and $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \approx \frac{21}{\sqrt{50}} \approx 2.97$.

$$\bar{X} \sim N(100, 2.97^2)$$

   That is, the "null distribution" is $N(100, 2.97^2)$.

3. Compute the test statistic. We can find our Z-score.

$$\begin{aligned} Z &= \frac{\bar{X} - 100}{2.97} \\ &= \frac{92 - 100}{2.97} \\ &\approx -2.69 \end{aligned}$$

   Our test stat is $Z = -2.69$. (This corresponds to an $\bar{X}$ value of 92).

4. Compute the P-value. The P-value is the probability that $\bar{X}$ takes on a value as extreme as (or more extreme than) the observed value of 92. We can find this P-value using either the Z-score or the $\bar{X}$ value.

$$\begin{aligned} P\text{-value} &= P(\bar{X} \leq 92) \\ &= P(Z \leq -2.69) \\ &= \text{normalcdf}(-\infty, -2.69, 0, 1) \\ &\approx 0.0036 \end{aligned}$$

   Using the $\bar{X}$ value we get ≈ 0.0035. Most of the time we can do these directly (without converting to Z-score). So, the probability that $\bar{X}$ takes on a value as extreme as (or more extreme than) the observed value of 92, is 0.0036.

5. In conclusion, since the probability is so low, we should reject our null hypothesis H_0. As a rule of thumb, we reject H_0 if the P-value is ≤ 0.05. Therefore, we have pretty strong evidence that the mean NOx emission rate of the modified engines will be less than 100 mg/s.

Example 75. In a certain experiment, 45 steel balls lubricated with purified paraffin were subjected to a 40 kg load at 600 rpm for 60 minutes. The average wear, measured by the reduction in diameter, was 673.2 μm and the standard deviation is 14.9 μm. Assume that the specification for a lubricate is that the mean wear be less than 675 μm. Find the P-value for testing $H_0 : \mu \leq 675$ versus $H_1 : \mu < 675$.

We know $n = 45$ steel balls, and $\bar{X}$ = average wear for a sample of 45 steel balls. We have a particular value of $\bar{X} = 673.2$, and a sample standard deviation of $s = 14.9$. We can find our mean and standard deviation for the null distribution:

$$\mu_{\bar{X}} = \mu = 675$$
$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} \approx \frac{14.9}{\sqrt{45}} \approx 2.22$$

So our null distribution is $N(675, 2.22^2)$. We can find our Z-score:

$$\begin{aligned} Z &= \frac{673.2 - 675}{2.22} \\ &\approx -0.81 \end{aligned}$$

So our P-value is the area in the left tail.

$$\begin{aligned} P(\bar{X} \leq 673.2) &= \text{normalcdf}(-\infty, 673.2, 675, 2.22) \\ &\approx 0.2087 \end{aligned}$$

We can also find it using the Z-score:

$$\begin{aligned} P(Z \leq -0.81) &= \text{normalcdf}(-\infty, -0.81, 0, 1) \\ &\approx 0.2090 \end{aligned}$$

Since our P-value is over 0.05, the evidence is not strong enough to reject H_0. That is, we do not reject H_0. This means that H_1 is "plausible". The average wear may, or may not, be greater than or equal to 675 μm. (Note, this also implies that the alternative hypothesis is plausible too. The average wear may or may not be less than 675 μm.)

Example 76. A scale is to be calibrated by weighing a 1000 g test weight 60 times. The 60 scale readings have mean 1000.6 g and a standard deviation of 2 g. Find P-value for testing $H_0 : \mu = 1000$ versus $H_1 : \mu \neq 1000$.

Note: $\mu \neq 1000$ means that it is either $\mu < 1000$ or $\mu > 1000$, and the scale is out of calibration.

We know our sample size is $n = 60$, and $\bar{X}$ = average weight of 60 scale readings. Our particular value of $\bar{X} = 1000.6$ and a sample standard deviation of $s = 2$. We can find our mean and standard deviation of $\bar{X}$:

$$\mu_{\bar{X}} = \mu = 1000$$
$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}} \approx \frac{2}{\sqrt{60}} \approx 0.258$$

So our null distribution is $N(1000, 0.258^2)$.

We can find our Z-score

$$\begin{aligned} Z &= \frac{1000.6 - 1000}{0.258} \\ &\approx 2.33 \end{aligned}$$

We are looking at a two tail test, the P-value will be a sum of the area in both tails (0.6 away from the center of 1000). Using symmetry we can see that the left hand Z-score will be -2.33.

Using $\bar{X}$ values:

$$\begin{aligned} P(\bar{X} \leq 999.4 \text{ OR } \bar{X} \geq 1000.6) &= P(\bar{X} \leq 999.4) + P(\bar{X} \geq 1000.6) \\ &= 2 \cdot P(\bar{X} \geq 1000.6) \\ &= 2 \cdot \text{normalcdf}(1000.6, \infty, 1000, 0.258) \\ &\approx 0.0200 \end{aligned}$$

We can multiply one of the tails by 2 because of symmetry. Lets use the Z-score to see if its much different

$$\begin{aligned} P(Z \leq -2.33 \text{ OR } Z \geq 2.33) &= P(Z \leq -2.33) + P(Z \geq 2.33) \\ &= 2 \cdot P(Z \geq 2.33) \\ &= 2 \cdot \text{normalcdf}(2.33, \infty, 0, 1) \\ &\approx 0.0198 \end{aligned}$$

This is a very small probability ($P \leq 0.05$), so we would reject H_0. So, we consider H_0 to be false, and we conclude that H_1 is true. There is strong evidence that $\mu \neq 1000$, and the scale is out of calibration.

Example 77. To understand how to calculate right tail tests: use $n = 60$, $\bar{X} = 1000.6$, $s = 2$, and $\mu_0 = 1000$, find the P-value for testing $H_0 : \mu \leq 1000$ and $H_1 : \mu > 1000$.

The null distribution is $N(1000, 0.258^2)$. The P-value is the area in the right tail. The Z-score is 2.33.

$$\begin{aligned} P(\bar{X} \geq 1000.6) &= P(Z \geq 2.33) \\ &= \text{normalcdf}(2.33, \infty, 0, 1) \\ &\approx 0.0099 \end{aligned}$$

So rounding to two decimal places $P = 0.01$.

10.1 Conclusions for Hypothesis Tests

Definition 48. Drawing conclusions from the results of hypothesis tests.

The only two conclusions that can be reached in a hypothesis test are that H_0 is false or that H_0 is plausible.

- The smaller the P-value, the more certain we can be that H_0 is false.
- The larger the P-value, the more plausible that H_0 becomes, but we can never be certain that H_0 is true.
- A rule of thumb suggests to reject H_0 whenever $P \leq 0.05$. While this rule is convenient, it has no scientific basis.

If we reject H_0 : we conclude that H_0 is false, and thus, conclude that H_1 is true.

If we do not reject H_0 : we conclude that H_0 is plausible (it may or may not be true.) This implies that H_1 is plausible as well.

Definition 49. Statistical significance.

Let α be any value between 0 and 1. Then, if $P \leq \alpha$,

- The result of the test is said to be statistically significant at the $100\alpha\%$ level.
- The null hypothesis is rejected at the $100\alpha\%$ level.
- When reporting the result of a hypothesis test, report the P-value, rather than just comparing it to 5% or 1%.

Note 17. Statistical significance is not the same as practical significance, that is, it is possible for a statistically significant result to not have any scientific or practical importance.

Example 78. A hypothesis test is performed of the null hypothesis $H_0 : \mu = 0$. The P-value turns out to be 0.03.

1. Is the result statistically significant at the 10% level?

   This means that $\alpha = 0.1$.

   The 5% level?

   This means that $\alpha = 0.05$.

   The 1% level?

   This means that $\alpha = 0.01$.

   Because our P-value is $P = 0.03$, this means that the value is statistically significant at the 10% and the 5% level, but not at the 1% level.

   $P \leq 0.1$ and $P \leq 0.05$, but $P > 0.01$.

2. Is the null hypothesis rejected at the 10% level? The 5% level? The 1% level?

   With a $P = 0.03$, our null hypothesis H_0 is rejected at the 10% level and the 5% level, but not at the 1% level.

Choosing H_0 to answer the right question:

When performing a hypothesis test, it is important to choose H_0 and H_1 appropriately so that the result of the test can be useful in forming a conclusion.

Example 79. Specifications for a water pipe call for a mean breaking strength μ of more than 2000 lb per linear foot. Engineers will perform a hypothesis test to decide whether or not to use a certain kine of pipe. They will select a random sample of 1 ft sections of pipe, measure their breaking strengths, and perform a hypothesis test. The pipe will not be used unless the engineers can conclude that $\mu > 2000$.

1. Assume they test $H_0 : \mu \leq 2000$ versus $H_1 : \mu > 2000$. Will the engineers decide to use the pipe if H_0 is rejected? What if H_0 is not rejected?

   If H_0 is rejected, the engineers will conclude that $\mu > 2000$, and they will use the pipe. If H_0 is not rejected, it is plausible that $\mu > 2000$, but it is also plausible that $\mu \leq 2000$, so the engineers don't have enough evidence, and will not use the pipe.

2. Assume the engineers test $H_0 : \mu \geq 2000$ versus $H_1 : \mu < 2000$. Will the engineers decide to use the pipe if H_0 is rejected? What if H_0 is not rejected?

   If H_0 is rejected, they would conclude that $\mu < 2000$, and therefore would not use the pipe. If H_0 is not rejected, then H_0 is plausible, but H_1 is also plausible, and so the engineers don't have enough evidence, and will not use the pipe.

So we would use the first test, because the second is not a useful test.

Relationship between hypothesis tests and confidence intervals:

In a hypothesis test for a population mean μ, we specify a particular value of μ (the null hypothesis) and determine whether that value is plausible. In contrast, a confidence interval for a population mean μ can be thought of as the collection of all values for μ that meet a certain criterion of probability, specified by the confidence level $100(1 - \alpha\%$.

The values contained within a two-sided level $100(1 - \alpha)\%$ confidence interval for a population mean μ are precisely those values for which the P-value of a two-tailed test will be greater than α. (Similarly, a one-sided level $100(1-\alpha)\%$ confidence interval consists of all the values for which the P-value in a one-tailed test would be greater than α.)

Example 80. In a sample of 50 microdrills drilling a low-carbon alloy steel, the average lifetime (expressed as the number of holes drilled before failure) was 12.68 with a standard deviation of 6.83. Setting α to 0.05, the 95% confidence interval for the population mean life μ is (10.79, 14.57), found by computing $12.68 \pm 1.96 \left(\frac{6.83}{\sqrt{50}}\right)$.

1. Suppose we tested $H_0 : \mu = 10.79$ versus $H_1 : \mu \neq 10.79$. What is the P-value?

   With $\bar{X} = 12.68$, $n = 50$, and a null distribution of $N(10.79, \frac{6.83^2}{50})$. We know $\sigma_{\bar{X}} \approx \frac{6.83}{\sqrt{50}} \approx 0.9659$. So we can find the Z-score:

$$\begin{aligned} Z &= \frac{12.68 - 10.79}{0.9659} \\ &\approx 1.96 \end{aligned}$$

   So to find our P-value:

$$\begin{aligned} P &= 2 \cdot P(\bar{X} \geq 12.68) \\ &= 2 \cdot P(Z \geq 1.96) \\ &= 2 \cdot \text{normalcdf}(1.96, \infty, 0, 1) \\ &\approx 0.0500 \end{aligned}$$

2. Suppose we tested $H_0 : \mu = 14.57$ and $H_1 : \mu \neq 14.57$. What is the P-value?

   Our null distribution is $N(14.57, 0.9659^2)$.

$$\begin{aligned} Z &= \frac{12.68 - 14.57}{0.9659} \\ &\approx -1.96 \end{aligned}$$

   So to find our P-value:

$$\begin{aligned} P &= 2 \cdot P(\bar{X} \leq 12.68) \\ &= 2 \cdot P(Z \leq -1.96) \\ &\approx 0.0500 \end{aligned}$$

   Both 1 and 2 show that there is a significance level of 5% for hypothesis tests.

3. Suppose we tested $H_0 : \mu = 11.56$ versus $H_1 : \mu \neq 11.56$. What is the P-value?

   Our null distribution is $N(11.56, 0.9659^2)$.

   Our Z-score is

$$\begin{aligned} Z &= \frac{12.68 - 11.56}{0.9659} \\ &\approx 1.16 \end{aligned}$$

   Our P-value is

$$\begin{aligned} P &= 2 \cdot P(Z \geq 1.16) \\ &\approx 0.2460 \end{aligned}$$

4. Suppose we tested $H_0 : \mu = 14.92$ versus $H_1 : \mu \neq 14.92$. What is our P-value?

   Our null distribution is $N(14.92, 0.9659^2)$.

$$\begin{aligned} Z &= \frac{12.68 - 14.92}{0.9659} \\ &\approx -2.32 \end{aligned}$$

   So our P-value is

$$\begin{aligned} P &= 2 \cdot P(Z \leq -2.32) \\ &\approx 0.0203 \end{aligned}$$

If we choose any value μ_0 in the interval (10.79, 14.56) and test $H_0 : \mu = \mu_0$ versus $H_1 : \mu \neq \mu_0$, the P-value will be greater than 0.05. On the other hand, if we choose $\mu_0 < 10.79$ or $\mu_0 > 14.57$, the P-value will be less than 0.05.

The 95% confidence interval consists of precisely those values of μ whose P-values are greater than 0.05 in a hypothesis test. In this sense, the confidence interval contains all the values that are plausible for the population mean μ.

10.2 Small sample tests for a population mean

Definition 50. Let $X_1, \dots, X_n$ be a sample from a *normal* population with mean μ and standard deviation σ, where σ is unknown. To test the null hypothesis in the form $H_0 : \mu \leq \mu_0$, $H_0 : \mu \geq \mu_0$, or $H_0 : \mu = \mu_0$

- Compute the test statistic

$$t = \frac{\bar{X} - \mu_0}{\frac{s}{\sqrt{n}}}$$

- Compute the P-value. The P-value is an area under the Student's t curve with $n-1$ degrees of freedom, which depends on the alternate hypothesis as follows

$H_1 : \mu > \mu_0$	Area to the right of t
$H_1 : \mu < \mu_0$	Area to the left of t
$H_1 : \mu \neq \mu_0$	Sum of the ares in the tails cut off by t and $-t$

- If σ is known, the test statistic $Z = \frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}}$, and a Z test should be performed.

Example 81. An article describes a method of measuring precipitation by combining satellite measures with measurements made on the ground. Errors were recorded monthly for 16 months, and averaged 0.393 mm with a standard deviation of 0.368 mm. Can we conclude that the mean error is less than 0.6 mm? (Assume error measurements are approximately normally distributed.)

Let μ represent the mean error. The null and alternative hypotheses are

$$H_0 : \mu \geq 0.6$$
$$H_1 : \mu < 0.6$$

Our $n = 16$, $\bar{X} = 0.393$, and $s = 0.368$. Since our sample size $n = 16$ is less than 30, we must use the Student's t-distribution. So we can find t:

$$\begin{aligned} t &= \frac{\bar{X} - 0.6}{\frac{s}{\sqrt{n}}} \\ &= \frac{0.393 - 0.6}{\frac{0.368}{\sqrt{16}}} \\ &\approx -2.25 \end{aligned}$$

The null distribution here is the Student's t with 15 degrees of freedom ($df = 15$). The P-value is the probability that we find t as as extreme (or more extreme) than that observed. We are looking for the area in the left tail.

$$\begin{aligned} P &= P(t \leq -2.25) \\ &= \text{t cdf}(-\infty, -2.25, 15) \\ &\approx .01994 \end{aligned}$$

Since our P-value is less than 0.05, we can reject H_0, therefore H_1 is true, and the mean $\mu < 0.6$ mm. (Note also that this result is statistically significant at the 2% level too, since $P < 0.02$.)

Note 18. Normal probability plots are often used to help decide if a sample might be from a normal (or approximately normal) population. This is usually a graph of normal scores over values in a sample, and if the plots are in a relatively linear fashion then the population is approximately normal.

Example 82. Space collars for a transmission countershaft have a thickness specification of 38.98 to 39.02 mm. The process that manufactures the collars is supposed to be calibrated so that the mean thickness is 39.00 mm, which is the center of the specification window. A sample of six collars is drawn and measured for thickness. The six thicknesses are: 39.030, 38.997, 39.012, 39.008, 39.019, and 39.002.

Assume the population of thicknesses of collars is approximately normal. Can we conclude that the process needs recalibration?

Lets start by finding our hypotheses:

$$H_0 : \mu = 39.00 \qquad\qquad H_1 : \mu \neq 39.00$$

$$\begin{aligned}
\sum X &= 234.068 \\
\sum X^2 &= 9131.305482 \\
\bar{X} &= \frac{234.068}{6} \\
&\approx 39.01133 \\
s &= \sqrt{\frac{\sum X^2 - \frac{(\sum X)^2}{n}}{n-1}} \\
&\approx 0.011928
\end{aligned}$$

Our sample size is $n = 6$ collars, $\bar{X} = 39.01133$, $s = 0.011928$ We can find t

$$\begin{aligned}
t &= \frac{\bar{X} - \mu}{\frac{s}{\sqrt{n}}} \\
&\approx 2.327
\end{aligned}$$

Our null distribution is a Student's t distribution with $df = 5$.

We are looking for our P-value for the two tails now using our t-value.

$$\begin{aligned}
P &= 2 \cdot P(T \leq -2.327) \\
&= 2 \cdot \text{tcdf}(-\infty, -2.327, 5) \\
&\approx 0.0675
\end{aligned}$$

Our P-value is greater than 0.05, so we can't reject H_0. The evidence isn't strong enough to conclude that the process is out of calibration. It is plausible that it is in calibration.

10.3 Fixed level testing

Definition 51 (Fixed level test)**.** When a decision is going to be made on the basis of a hypothesis test, there is no choice but to pick a cutoff point for the P-value. When this is done, the test is referred to as a **fixed level test**.

To conduct a fixed level test:

- We first choose a number α, where $0 < \alpha < 1$. This is called the significance level.
- Compute the P-value in the usual way.
- If $P \leq \alpha$, reject H_0. If $P > \alpha$, do not reject H_0.

Definition 52 (Critical points and rejection regions)**.** In a fixed level test, a **critical point** is a value of the test statistic that produces a P-value exactly equal to α, where α is the significance level of the test. If the test stat is on one side of the critical point, the P-value will be less than α and H_0 will be rejected. If the test stat is on the other side of the critical point, the P-value will be greater than α and H_0 will not be rejected.

The region on the side of the critical point that leads to rejection of H_0 is the **rejection region**. Note, the critical point itself is also in the rejection region.

Example 83. A new concrete mix is being evaluated. The plan is to sample 100 concrete blocks made with the new mix, compute the sample mean $\bar{X}$, and then test.

$$H_0 : \mu \leq 1350 \qquad\qquad H_1 : \mu > 1350$$

The units are in MPa. It is assumed from previous tests of this sort that the population standard deviation σ will be close to 70 MPa. Find the critical point and the rejection region if the test will be conducted at a significant level of 5%.

We have a sample size of $n = 100$ concrete blocks. We can find the standard deviation of $\bar{X}$:

$$\begin{aligned} \sigma_{\bar{X}} &= \frac{\sigma}{\sqrt{n}} \\ &= \frac{70}{\sqrt{100}} \\ &= 7 \end{aligned}$$

So our null distribution is $N(1350, 7^2)$. We have $\alpha = 0.05$.

We can find the critical point by first finding a Z-score:

$$\begin{aligned} Z &= \text{invNorm}(0.95, 0, 1) \\ &\approx 1.645 \end{aligned}$$

We can find the critical point by adding the corresponding number of standard deviations to our sample mean.

$$\begin{aligned} \bar{X} &= 1350 + 1.645(7) \\ &= 1361.5 \end{aligned}$$

Our rejection region is all of the values beyond the critical point. The rejection region consists of all values of $\bar{X}$ that are greater than or equal to 1361.5. That is, we reject H_0 if we get $\bar{X} \geq 1361.5$.

Notice that the rejection region for this one-tailed test consists of the upper 5% of the null distribution.

Also worth trying to get the value directly using $\bar{X}$ values:

$$\text{invNorm}(0.95, 1350, 7) \approx 1361.5$$

Example 84. In a hypothesis test to determine whether or not a scale is in calibration, the null hypothesis is $H_0 : \mu = 1000$. The null distribution of $\bar{X}$ is $N(1000, 0.26^2)$. Find the rejection region if the test will be conducted at a significance level of 5%.

We know that $H_1 : \mu \neq 1000$. We know via our significance level that our $\alpha = 0.05$. Since this is a two-tailed test, it means that both tails will have an area of 0.05, $\frac{\alpha}{x} = 0.025$. So we can find the Z value for those areas:

$$\text{invNorm}(0.025, 0, 1) = -1.96$$

So our $\bar{X}$ value for that Z is:

$$\begin{aligned}\bar{X} &= 1000 \pm 1.96(0.26) \\ &\approx 999.5, 1000.5\end{aligned}$$

So our area of rejection is $\bar{X} \leq 999.5$ or $\bar{X} \geq 1000.5$. H_0 will be rejected if $\bar{X}$ is in the upper or lower 2.5% of the null distribution.

Note that we can also do this directly with $\bar{X}$ values:

$$\text{invNorm}(0.025, 1000, 0.26) \approx 999.5$$

10.4 Type I and Type II errors

There are two types of incorrect decisions that can happen when performing a hypothesis test, called **Type I error** and **Type II error**.

Definition 53 (Type I error)**.** Rejecting the null hypothesis H_0 when it is in fact true.

Definition 54 (Type II error)**.** Failing to reject the null hypothesis H_0 when it is in fact false.

These can be visualized in table form:

	H_0 is True	H_0 is False
Do not reject H_0	Correct	Incorrect (II)
Reject H_0	Incorrect (I)	Correct

Example 85. A snack-food company produces 454 g bags of pretzels. Although the actual net weights deviate slightly from 454 g and vary from one bag to another, the company insists that the mean net weight of the bags be 454 g. A hypothesis test is performed:

$$H_0 : \mu = 454 \qquad H_1 : \mu \neq 454$$

H_0 is “the packaging machine is working correctly”. H_1 is “the packaging machine is not working correctly.”

Explain what each of the following would be:

1. Type I error: we conclude that the machine isn’t working, when in fact it is working properly.
2. Type II Error: we fail to conclude that the packaging machine is not working properly when in fact it is not working properly.
3. Correct decision:
   (a) A true null hypothesis is not rejected.
   (b) A false null hypothesis is rejected.

   A correct decision occurs if either “we fail to conclude the packaging machine is not working properly (we conclude it is plausible it is working properly) when it is in fact working properly.” Or, “we conclude that the packaging machine is not working properly when it is in fact not working properly.”
4. Now suppose the results of the hypothesis test lead to rejection of the null hypothesis $\mu = 454$, so that the conclusion is $\mu \neq 454$. Is this a correct conclusion or is there an error?
   (a) If the mean net weight μ is in fact 454g: we rejected H_0 when in fact it is true, Type I error.
   (b) If the mean net weight μ is in fact not 454g: we rejected H_0 when in fact it is false, correct decision.

Definition 55 (Probability of a Type I error)**.** If α is the significance level that has been chosen for the hypothesis test, then the probability of a Type I error is never greater than α. When μ is on the boundary of H_0, for example ($\mu = \mu_0$), then the probability of a Type I error is equal to α.

When testing

$$H_0 : \mu \leq \mu_0 \qquad H_1 : \mu > \mu_0$$

μ_0 is on the boundary of H_0, closest to the alternative hypothesis.

$$P(\text{Type I error}) \leq \alpha$$

10.5 Power

Definition 56. The **power** of a hypothesis test is the probability of not making a Type II error. That is, it is the probability of rejecting the null hypothesis H_0 when it is false.

Let $\beta = P(\text{Type II error})$. Then, power $= 1 - \beta$.

The power measures the ability of a hypothesis test to detect a false null hypothesis.

- When power is close to 0, the test is not very good at detecting a false null hypothesis.

- When power is close to 1, the test is very good at detecting a false null hypothesis.

This will be a conditional probability, where

$$P(\text{reject } H_0 \mid H_0 \text{ is false})$$

To compute the power:

1. Compute the rejection region for the test.
2. Compute the probability that the test stat falls in the rejection region if the alternate hypothesis is true.

   For example, let μ_1 be a particular value of the mean belonging to the alternative hypothesis. Then, the power $= P(\text{rejecting } H_0 \mid \mu = \mu_1)$.

   If our tests are

   $$H_0 : \mu \leq \mu_0 \qquad\qquad H_1 : \mu > \mu_0$$

   When we perform our tests for a significance level of α, we can find the critical point that contains the area of α in the tail. If the mean lands in that area, then we reject H_0. If we pick an alternative mean of μ_1 that is greater than μ_0, we have an alternative distribution, where $\mu_{\bar{X}} = \mu_1$, and $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$. The area under this alternative curve to the right of the critical point found earlier is the power.

Example 86. Assume that a new chemical process has been developed that may increase a yield over that of the current process. The current process is known to have a mean yield of 80, and a standard deviation of 5, where the units are a percentage of a theoretical maximum. If the mean yield of the new process is shown to be greater than 80, the new process will be put into production. Let μ denote the mean yield of the new process. It is proposed to run the new process 50 times and then to test the hypothesis

$$H_0 : \mu \leq 80 \qquad\qquad H_1 : \mu > 80$$

at a significance level of 5%. If H_0 is rejected, it will be concluded that $\mu > 80$, and the new process will be put into production.

1. Let us assume that if the new process had a mean yield of 81, then it would be a substantial benefit to put this process into production. If it is in fact the case that $\mu = 81$, what is the power of the test, that is, the probability that H_0 will be rejected. (Assume the standard deviation for the new process is $\sigma = 5$ and that $\sigma_{\bar{X}} = \frac{5}{\sqrt{50}} = 0.707$.)

   We can find power $= P(\text{reject } H_0 \mid \mu = 81)$. We know our null distribution is $N(80, 0.707^2)$. Our right tail area is 0.05 because that is our significance level. Our critical point is the $\bar{X}$ value that splits the area. We can find that critical point:

   $$\begin{aligned}\text{critical point} &= \text{invNorm}(0.95, 80, 0.707)\\ &\approx 81.16\end{aligned}$$

So we can find our alternative distribution, which assumes that our mean $\mu = 81$. Our rejection region is all values where $\bar{X} \geq 81.16$. Our alternative distribution is $N(81, 0.707^2)$.

We can find our power

$$\begin{aligned}\text{Power} &= P(\bar{X} \geq 81.16 \mid \mu = 81) \\ &= \text{normalcdf}(81.16, \infty, 81, 0.707) \\ &\approx 0.4105\end{aligned}$$

"There is a 41.05% chance of rejecting H_0 if $\mu = 81$." This is considered a relatively low power. So, the test may not be all that good at detecting a false null hypothesis. This means that if the mean yield of the new process is actually equal to 81, there is only a 41.05% chance that the proposed experiment will detect the improvement over the old process, and allow the new process to be put into production.

2. Find the power of the 5% level test of $H_0 : \mu \leq 80$ versus $H_1 : \mu > 80$ for the mean yield of the new process under the alternative $\mu = 82$, assuming $n = 50$ and $\sigma = 5$.

   Our null distribution is $N(80, 0.707^2)$. Our alternative distribution is $N(82, 0.707^2)$. We can use the same critical point as before, $\bar{X} = 81.16$. Our rejection region is $\bar{X} \geq 81.16$. We can find the area under the alternative curve, that is in the rejection region:

$$\begin{aligned}\text{Power} &= P(\bar{X} \geq 81.16 \mid \mu = 82) \\ &= \text{normalcdf}(81.16, \infty, 82, 0.707) \\ &\approx 0.8826\end{aligned}$$

   "There is a 88.26% chance of rejecting H_0 if $\mu = 82$."

   Lets do this same problem but with Z-scores:

$$\begin{aligned}Z &= \text{invNorm}(0.95, 0, 1) \\ &\approx 1.645\end{aligned}$$

   So to find our critical point:

$$\begin{aligned}\bar{X} &= 80 + 1.645(0.707) \\ &\approx 81.16\end{aligned}$$

   So this $\bar{X}$ value corresponds to the Z-score of

$$\begin{aligned}Z &= \frac{81.16 - 82}{0.707} \\ &\approx -1.19\end{aligned}$$

So we can find the power, the area under the alternative distribution that lies in the rejection region:

$$\begin{aligned}\text{Power} &= P(\bar{X} \geq 81.16 \mid \mu = 81)\\ &= P(Z \geq -1.19)\\ &= \text{normalcdf}(-1.19, \infty, 0, 1)\\ &\approx 0.8830\end{aligned}$$

Remember, the answers will differ slightly depending on if you use Z-scores or if you calculate directly with $\bar{X}$.

3. In testing the hypothesis $H_0 : \mu \leq 80$ versus $H_1 : \mu > 80$ regarding the mean yield of the new process, how many times must the new process be run so that a test conducted at a significance level of 5% will have power 0.90 against the alternative $\mu = 81$, if it is assumed that $\sigma = 5$?

   We are looking for n so that $P(\text{Reject } H_0 \mid \mu_1 = 81) = 0.90$. We have a null distribution with $\mu_{\bar{X}} = 80, \sigma_{\bar{X}} = \frac{5}{\sqrt{n}}$. We have an alternative distribution with $\mu_{\bar{X}} = 81, \sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$.

   The area to the right of the critical point is the rejection region. We can't yet find the critical point because we don't have a σ value for the alternative distribution. We know that the area under the curve in the rejection region on the alternative distribution is going to be the power, $= 0.9$.

   We can find the critical point using the null distribution's $\bar{X}$ value:

$$\begin{aligned}\text{Critical point}_0 &= \mu + Z\sigma_{\bar{X}}\\ &= 80 + Z\left(\frac{5}{\sqrt{n}}\right)\end{aligned}$$

   We can find the Z-score for the null distribution, called Z_0 :

$$\begin{aligned}Z_0 &= \text{invNorm}(0.95, 0, 1)\\ &\approx 1.645\end{aligned}$$

   So our critical point is 1.645 standard deviations above the null distributions mean.

$$\text{Critical point}_1 = 80 + 1.645\left(\frac{5}{\sqrt{n}}\right)$$

   We can also find the Z-score for the alternative distribution, that we can call Z_1:

$$\begin{aligned}Z_1 &= \text{invNorm}(0.10, 0, 1)\\ &\approx -1.28\end{aligned}$$

So the critical point n the alternative distribution is

$$\text{Critical point} = 81 - 1.28\left(\frac{5}{\sqrt{n}}\right)$$

Since we have 2 expressions that have the variable n, we can now solve for n :

$$\begin{aligned}
\text{Critical point}_0 &= \text{Critical point}_1 \\
80 + 1.645\left(\frac{5}{\sqrt{n}}\right) &= 81 - 1.28\left(\frac{5}{\sqrt{n}}\right) \\
2.925\left(\frac{5}{\sqrt{n}}\right) &= 1 \\
\sqrt{n} &= 2.925(5) \\
n &= 213.89
\end{aligned}$$

We always round sample size up, so $n = 214$. "The new process must be run 214 times to get a 90% power."

4. Finding the power for a left-tailed and two-tailed tests will be similar to finding the power of a right-tailed test (like we just did), but the picture will be different.

   Test $H_0 : \mu = 80$ versus $H_1 : \mu \neq 80$. $n = 50, \sigma = 5, \sigma_{\bar{X}} = \frac{5}{\sqrt{50}} \approx 0.707$. Use a significance level of $\alpha = 0.05$.

   Our null distribution has $\mu_{\bar{X}} = 80, \sigma_{\bar{X}} = 0.707$. Our alternative distribution has $\mu_{\bar{X}} = 82, \sigma_{\bar{X}} = 0.707$.

   Since this is a two-tailed test, our tails have the area of $\frac{\alpha}{2} = 0.025$. Our rejection region is the area of those tails. We can find the left critical point with:

$$\text{invNorm}(0.025, 80, 0.707) = 78.61$$

   And find our right critical point with symmetry, $= 81.39$. The power is the area under the alternative distribution curve in the rejection region, which is in in **both** tails!

$$\begin{aligned}
\text{Power} &= P(\bar{X} \leq 78.61 \text{ OR } \bar{X} \geq 81.39 \mid \mu = 82) \\
&= \text{normalcdf}(-\infty, 78.16, 82, 0.707) + \text{normalcdf}(81.38, \infty, 82, 0.707) \\
&\approx 0.8059
\end{aligned}$$

Aside 5.

$$H_0 : \mu \leq 80 \qquad\qquad H_1 : \mu > 80$$

μ_1	Power
80.5	0.1753
81.0	0.4105
81.5	0.6847
82.0	0.8826
82.5	0.9710
83.0	0.9954
83.5	0.9995
84.0	0.99997

When graphed, these values are approaching 1.

Definition 57 (Relationship between Type I and Type II errors)**.** Let $\alpha = P(\text{Type I error})$ and $\beta = P(\text{Type II error})$. The area in the null hypothesis curve that is **not** in the rejection region is β. For a fixed sample size, the smaller we specify the significance level α, the larger will be the probability β of not rejecting a false null hypothesis.

$$P(\text{Type II error}) = P(\text{Not rejecting } H_0 \mid H_0 \text{ is false})$$

Definition 58 (Relationship between the Power of a Test and Sample Size)**.** For a fixed significance level α, the power of a hypothesis increases as the sample size n increases. The area in the alternative curve that is in the rejection region is the power, so since our standard deviation is $\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$ (often called standard error). $\sigma_{\bar{X}}$ gets smaller as n increases.

11 Jointly Distributed Random Variables

Definition 59. When two or more variables are associated with each item in a population, the random variables are said to be **jointly distributed**.

If the random variables are discrete, they are said to be **jointly discrete**. If the random variables are continuous, they are said to be **jointly continuous**.

We will be focusing on jointly discrete random variables and their probability distributions.

Example 87. Rectangular plastic covers for a compact disc (CD) tray have specifications regarding length and width. Measurements are rounded to the nearest millimeter. Let X represent the measured length and let Y represent the measured width. The possible values of X are 129, 130, and 131. The possible values of Y are 15 and 16. Both X and Y are discrete, so X and Y are **jointly discrete**. There are six possible values for the ordered pair (X, Y) :

x	y	$P(X = x)$ and $P(Y = y)$
129	15	0.12
129	16	0.08
130	15	0.42
130	16	0.28
131	15	0.06
131	16	0.04

The **joint probability mass function** is the function $p(x, y) = P(X = x \text{ and } Y = y)$. For example, the probability of

$$\begin{aligned} p(129, 15) &= P(X = 129 \text{ and } Y = 15) \\ &= 0.12 \end{aligned}$$

1. Find the probability that a CD cover has the length of 129mm.

$$\begin{aligned} P(X = 129) &= P(X = 129 \text{ and } Y = 15) + P(X = 129 \text{ and } Y = 16) \\ &= 0.12 + 0.08 \\ &= 0.2 \end{aligned}$$

2. Find the probability that a CD cover has a width of 16.

$$\begin{aligned} P(Y = 16) &= 0.08 + 0.28 + 0.04 \\ &= 0.4 \end{aligned}$$

We can organize the information from the previous table into a 2-way table called a **contingency table**. This table shows the joint probability distribution. X is the column and Y is the rows.

	15	16	$P_X(x) = P(X = x)$
129	0.12	0.08	0.2
130	0.42	0.28	0.7
131	0.06	0.04	0.1
$P_Y(y) = P(Y = y)$	0.6	0.4	1

The sum of the rows is actually the probability of that specific X value happening, for example $P(X = 129) = 0.2$. The sum of the column is the probability of that specific Y value happening, for example $P(Y = 15) = 0.6$. Since the probability mass functions for X and Y appear in the margins of our table, they are often referred to as the "marginal probability mass functions".

Note:

$$\sum_x P_X(x) = 1 \qquad\qquad \sum_y P_Y(y) = 1$$

So our probability mass function for X is

x	$p_X(x) = P(X = x)$
129	0.2
130	0.7
131	0.1
	1

The mean and variance of the random variable X is

$$\begin{aligned}
\mu_X &= \sum_x x p_X(x) \\
&= 129(0.2) + 130(0.7) + 131(0.1) \\
&= 129.9 \\
\sigma_X^2 &= \sum_x (x - \mu)^2 p_X(x) \\
&= \sum_x x^2 p_X(x) - \mu_X^2 \\
&= \left(129^2(0.2) + 130^2(0.7) + 131^2(0.1)\right) - 129.9^2 \\
&\approx 0.29
\end{aligned}$$

Our probability mass function for Y is

y	$p_Y(y) = P(Y = y)$
15	0.6
16	0.4
	1

$$\begin{aligned}
\mu_Y &= 15(0.6) + 16(0.4) \\
&= 15.4 \\
\sigma_Y^2 &\approx 0.24
\end{aligned}$$

Definition 60. If X and Y are jointly discrete random variables:

- The joint probability mass function of X and Y is the function:

$$p(x, y) = P(X = x \text{ and } Y = y)$$

- The marginal probability mass functions of X and of Y can be obtained from the joint probability mass function as follows:

$$p_X(x) = P(X = x) = \sum_y p(x, y) \qquad\qquad p_Y(y) = P(Y = y) = \sum_x p(x, y)$$

  where the sums are taken over all possible values of Y and of X, respectively.

- The joint probability mass function has the property

$$\sum_x \sum_y p(x, y) = 1$$

  where the sum is taken over all possible values of X and Y.

Note 19. The idea of a joint probability mass function extends easily to more than two variables. In general, we have the following definition:

If the random variables $X_1, \ldots, X_n$ are jointly discrete, the joint probability mass function is

$$p(x_1, \ldots, x_n) = P(X_1 = x_1, \ldots, X_n = x_n)$$

Definition 61 (Mean of a Random Variable)**.** Let X be a random variable, and let $h(X)$ be a function of X.

- If X is discrete with probability mass function $p(x)$, the mean of $h(X)$ is given by

$$\mu_{h(X)} = \sum_x h(x)p(x)$$

This idea can also be used with joint random variables.

If X and Y are jointly distributed random variables, and $h(X, Y)$ is a function of X and Y, then if X and Y are jointly discrete with joint probability mass function $p(x, y)$,

$$\mu_{h(X,Y)} = \sum_x \sum_y h(x, y)p(x, y)$$

So if $h(X) = (X - \mu_X)^2$, then the expected value is

$$\begin{aligned} E\left[(X-\mu_X)^2\right] &= \mu_{(X-\mu_X)^2} \\ &= \sum_x (x-\mu_X)^2 p(x) \\ &= \sigma_X^2 \end{aligned}$$

Recall that this is the "defining formula."

In general, if X and Y are jointly discrete with a probability mass function $p(x, y)$:

$$\mu_{h(X,Y)} = \sum_x \sum_y h(x,y)p(x,y)$$

Example 88. 1. X and Y are discrete random variables with the following joint distribution (X is column, Y is row.)

	-1	1	3
1	0.2	0.15	0.05
4	0.3	0.05	0.25

(a) Construct the marginal probability mass function for X and Y

	-1	1	3	$P_X(x)$
1	0.2	0.15	0.05	0.4
4	0.3	0.05	0.25	0.6
$P_Y(y)$	0.5	0.2	0.3	1

(b) Find the mean values for X and Y.

$$\begin{aligned} \mu_X &= \sum_x x P_X(x) \\ &= 1(0.4) + 4(0.6) \\ &= 2.8 \\ \mu_Y &= \sum_y y P_Y(y) \\ &= -1(0.5) + 1(0.2) + 3(0.3) \\ &= 0.6 \end{aligned}$$

(c) Find the variances for X and Y.

Aside 6.

$$\begin{aligned}
E(X^2) &= 1^2(0.4) + 4^2(0.6) \\
&= 10 \\
E(Y^2) &= (-1)^2(0.5) + 1^2(0.2) + 3^2(0.3) \\
&= 3.4
\end{aligned}$$

$$\begin{aligned}
\sigma_X^2 &= E(X^2) - (\mu_X)^2 \\
&= 10 - 2.8^2 \\
&= 2.16 \\
\sigma_Y^2 &= E(Y^2) - (\mu_Y)^2 \\
&= 3.4 - (0.6)^2 \\
&= 3.04
\end{aligned}$$

(d) Compute $Cov(X,Y)$. (Definition below)

We know that

$$Cov(X,Y) = \mu_{XY} - \mu_X \mu_Y$$

So lets first start by computing the μ_{XY}, this is simply multiplying the X value times the Y value, and multiply it by its weight (use table above):

$$\begin{aligned}
\mu_{XY} &= \sum_x \sum_y xy \cdot p(x,y) \\
&= (1)(-1)(0.2) + (1)(1)(0.15) + (1)(3)(0.05) \\
&+ (4)(-1)(0.3) + (4)(1)(0.05) + (4)(3)(0.25) \\
&= 2.1
\end{aligned}$$

So now we can calculate the covariance (defined in next section):

$$\begin{aligned}
Cov(X,Y) &= 2.1 - (2.8)(0.6) \\
&= 0.42
\end{aligned}$$

(e) Find $\rho_{X,Y}$. (Correlation defined in next section).

We need to calculate the covariance divided by the product of the standard deviations:

$$\begin{aligned}\rho_{X,Y} &= \frac{Cov(X,Y)}{\sigma_X \sigma_Y} \\ &= \frac{0.42}{\sqrt{2.16}\sqrt{3.04}} \\ &\approx 0.1639\end{aligned}$$

The correlation is close to 0, therefore there is at most a weak linear relationship between X and Y.

(f) Are X and Y independent? Explain.

We have to check all values and each product to see if they are equal to the marginals (definition in following subsections). We are checking $P(X = x) \cdot P(Y = y) = p(x, y)$ for all x and y.

$$\begin{aligned}P(X = 1) \cdot P(Y = -1) &= (0.4)(0.5) \\ p(1, -1) &= 0.2\end{aligned}$$

Notice that the joint probability for $(1, -1) = 0.2$, so that checks out.

$$\begin{aligned}P(X = 4) \cdot P(Y = -1) &= (0.6)(0.5) \\ p(4, -1) &= 0.3\end{aligned}$$

This continues for all values of x and y.

$$\begin{aligned}P(X = 1) \cdot P(Y = 1) &= (0.4)(0.2) \\ p(1, 1) &\neq 0.15\end{aligned}$$

Since $p(1, 1) = 0.15 \neq 0.08$, X and Y are **not** independent. The joint probability mass function is **not** equal to the product of the marginals.

(g) Construct the conditional distribution of X given that $Y = 1$.

We can use the values from our table above to create a new table:

x	$P_{X\mid Y}(X \mid 1)$
1	0.75
4	0.25
	1

Shown below:

$$\begin{aligned}
P_{X|Y}(1 \mid 1) &= \frac{p(1,1)}{p_Y(1)} \\
&= \frac{0.15}{0.2} \\
&\approx 0.75 \\
P_{X|Y}(4 \mid 1) &= \frac{p(4,1)}{p_Y(1)} \\
&= \frac{0.05}{0.2} \\
&\approx 0.25
\end{aligned}$$

(h) Conditional probability of Y given $X = 4$?

y	$P_{Y\|X}(y \mid 4)$
-1	0.5
1	0.083
3	0.417

The work to get the above values are below:

$$\begin{aligned}
P_{Y|X}(-1 \mid 4) &= \frac{p(4,-1)}{p_X(4)} \\
&= \frac{0.3}{0.6} \\
&\approx 0.5 \\
P_{Y|X}(1 \mid 4) &= \frac{p(4,1)}{0.6} \\
&= \frac{0.05}{0.6} \\
&\approx 0.083 \\
P_{Y|X}(3 \mid 4) &= \frac{p(4,3)}{p_X(4)} \\
&= \frac{0.25}{0.6} \\
&\approx 0.417
\end{aligned}$$

(i) Use your distribution in the last part to find $E(X \mid Y = 1)$.

$$\begin{aligned}
E(X \mid Y = 1) &= \sum_x x \cdot P_{X|Y}(x \mid 1) \\
&= 1(0.75) + 4(0.25) \\
&= 1.75
\end{aligned}$$

11.1 Covariance

Definition 62 (Covariance)**.** Let X and Y be random variables with mean μ_X and μ_Y. The covariance of X and Y is

$$\begin{aligned} Cov(X,Y) &= \mu_{(X-\mu_X)(Y-\mu_Y)} \\ &= E\left((X-\mu_X)(Y-\mu_Y)\right) \end{aligned}$$

An alternate (computing) formula is

$$\begin{aligned} Cov(X,Y) &= \mu_{XY} - \mu_X\mu_Y \\ &= E(XY) - \mu_X\mu_Y \end{aligned}$$

Note 20.

$$Cov(X,X) = \sigma_X^2$$

In general, the population covariance is a measure of the joint variability between two random variables, a measure of the total variation of two random variables from their expected values. In particular, the covariance is a measure of a certain type of relationship called a *linear relationship.* A positive covariance indicates that the two variables tend to move in the same direction; for example as one variable increases the other variable tends to increase as well. A negative covariance indicates that the two variables tend to move in inverse directions; for example, as one variable increases, the other variable tends to decrease. The strength of the linear relationship is also measured but is not easy to interpret since the covariance is not a unitless measure.

To see how the covariance works:

The covariance is the mean of the product of deviations $(X-\mu_X)(Y-\mu_Y)$. If a Cartesian coordinate system is constructed with the origin at (μ_X,μ_Y), this product will be positive in the first and third quadrants, and negative in the second and fourth quadrants.

It follows that if $Cov(X,Y)$ is strongly positive, then values of (X,Y) in the first and third quadrants will be observed much more often than values in the second and fourth quadrants. In a random sample of points, therefore, larger values of X would tend to be paired with larger values of Y, while smaller values of X would tend to be paired with smaller values of Y. (This shows an increase from left to right in the scatter plot.)

Similarly, if $Cov(X,Y)$ is strongly negative, the points in a random sample would be more likely to lie in the second and fourth quadrants, so larger values of X would tend to be paired with smaller values of Y. (This shows a decrease from left to right in the scatter plot.)

Finally, if $Cov(X,Y)$ is near 0, there would be little tendency for larger values of X to be paired with either larger *or* smaller values of Y. (This shows a random scatter of points in all quadrants in the scatter plot.)

Mini-proof for mean of random variable computing formula:

$$
\begin{aligned}
E[(X-\mu)^2] &= E(X^2) - (\mu_X)^2 \\
\sigma_X^2 &= E[(X-\mu)^2] \\
&= E(X^2 - 2\mu X + \mu^2) \\
&= E(X^2) - 2\mu E(X) + \mu^2 \\
&= E(X^2) - 2\mu \cdot \mu + \mu^2 \\
&= E(X^2) - \mu^2 \\
&= \mu_{X^2} - \mu_X^2
\end{aligned}
$$

Mini-proof for covariance computing formula:

$$
\begin{aligned}
Cov(X,Y) &= E\left((X-\mu_X)(Y-\mu_Y)\right) \\
&= E(XY - \mu_Y X - \mu_X Y + \mu_X\mu_Y) \\
&= E(XY) - \mu_Y E(X) - \mu_X E(Y) + \mu_X\mu_Y \\
&= E(XY) - \mu_Y\mu_X - \mu_X\mu_Y + \mu_X\mu_Y \\
&= E(XY) - 2\mu_X\mu_Y + \mu_X\mu_Y \\
&= E(XY) - \mu_X\mu_Y \\
&= \mu_{XY} - \mu_X\mu_Y
\end{aligned}
$$

11.2 Correlation

The population correlation is a unitless measure of the strength of a linear relationship between two variables.

Definition 63 (Correlation)**.** Let X and Y be jointly distributed random variables with standard deviation σ_X and σ_Y. The correlation between X and Y is denoted $\rho_{X,Y}$ and is given by

$$\rho_{X,Y} = \frac{Cov(X,Y)}{\sigma_X\sigma_Y}$$

For any two random variables X and Y :

$$-1 \le \rho_{X,Y} \le 1$$

The closer the correlation value is to either -1 or 1, the stronger the linear relationship is between the variables. For example, a correlation value close to -1 indicates a strong negative linear relationship between X and Y, meaning there is a strong tendency for Y to decrease linearly as X increases. A correlation value close to 1 indicates a strong positive linear relationship, meaning there is a strong tendency for Y to increase linearly as X increases. A correlation value near 0 indicates at most a weak linear relationship between X and Y.

11.3 Conditional distribution

Suppose X and Y are jointly discrete. Let x be any value for which $P(X = x) > 0$. Then the conditional probability that $Y = y$ is $P(Y = y \mid X = x)$. If we let $p(x, y)$ denote the joint probability mass function of X and Y, and let $p_X(x)$ denote the marginal probability mass function of X. Then the conditional probability is

$$\begin{aligned} P(Y = y \mid X = x) &= \frac{P(X = x \text{ and } Y = y)}{P(X = x)} \\ &= \frac{p(x, y)}{p_X(x)} \end{aligned}$$

The conditional probability mass function of Y given $X = x$ is the conditional probability $P(Y = y \mid X = x)$, considered as a function of y and x.

Definition 64. Let X and Y be jointly discrete random variables with joint probability mass function $p(x, y)$. Let $p_X(x)$ denote the marginal probability mass function of X and let x be any number for which $p_X(x) > 0$.

The **conditional probability mass function** of Y given $X = x$ is

$$p_{Y|X}(y \mid x) = \frac{p(x, y)}{p_X(x)}$$

Note, that for any particular values of x and y, the value of $p_{Y|X}(y \mid X)$ is just the conditional probability $P(Y = y \mid X = x)$.

Definition 65 (Conditional expectation)**.** Recall that "expectation" is another term for mean. For jointly discrete variables, a **conditional expectation** is an expectation, or mean, calculated using a conditional probability mass function.

The conditional expectation of Y given $X = x$ is denoted $E(Y \mid X = x)$ or $\mu_{Y|X=x}$.

11.4 Independent random variables

Recall

Definition 66. Two random variables X and Y are independent, provided that if X and Y are jointly discrete, the joint probability mass function is equal to the product of the marginals:

$$p(x, y) = p_X(x) p_Y(y)$$

Note 21. In general, random variables $X_1, \ldots, X_n$ are independent, provided that if $X_1, \ldots, X_n$ are jointly discrete, the joint probability mass function is equal to the product of the marginals:

$$p(x_1, \ldots, x_n) = p_{X_1}(x_1) \cdots p_{X_n}(x_n)$$

If X and Y are independent, then if X and Y are jointly discrete, and x is a value for which $p_X(x) > 0$, then

$$p_{Y|X}(y \mid x) = p_Y(y)$$

If the jointly discrete variables X and Y are independent random variables, then the conditional distribution of Y given X is the same as the marginal distribution of Y.

11.5 Relationships between Covariance, Correlation, and Independence

Relationships

- If $Cov(X,Y) = \rho_{X,Y} = 0$, then X and Y are uncorrelated.

  When both the covariance and correlation are equal to 0, there is no linear relationship between X and Y. In this case we say that X and Y are **uncorrelated**. Note that if $Cov(X,Y) = 0$, then it is always the case that $\rho_{X,Y} = 0$, and vice versa.

- If X and Y are independent, then X and Y are uncorrelated.

  If X and Y are independent random variables, then X and Y are always uncorrelated, since there is no relationship, linear or otherwise, between them.

  In particular, if X and Y are independent, it can be shown that $\mu_{XY} = \mu_X \mu_Y$, from which is follows that the $Cov(X,Y) = \rho_{X,Y} = 0$.

- It is mathematically possible for X and Y to be uncorrelated without being independent. This rarely occurs in practice.

  It is mathematically possible to construct random variables that are uncorrelated but not independent. This phenomenon is rarely seen in practice, however. So, having $Cov(X,Y) = \rho_{X,Y} = 0$ does not necessarily mean that X and Y are independent.

When X and Y are independent:

$$\sigma^2_{aX+bY} = a^2\sigma^2_X + b^2\sigma^2_Y$$

Without the assumption that X and Y are independent, we have the following:

$$\sigma^2_{aX+bY} = a^2\sigma^2_X + b^2\sigma^2_Y + 2abCov(X,Y)$$

Here is a mini-proof:

$$\begin{aligned}
\sigma^2_{aX+bY} &= E[(aX+bY)^2] - \mu^2_{aX+bY} \\
&= E[a^2X^2 + 2abXY + b^2Y^2] - (a\mu_X + b\mu_Y)^2 \\
&= a^2E(X^2) + 2abE(XY) + b^2E(Y^2) - (a^2\mu^2_X + 2ab\mu_X\mu_Y + b^2\mu^2_Y) \\
&= a^2[E(X^2) - \mu^2_X] + b^2[E(Y^2) - \mu^2_Y] + 2ab[E(XY) - \mu_X\mu_Y] \\
&= a^2\sigma^2_X + b^2\sigma^2_Y + 2ab \cdot Cov(X,Y)
\end{aligned}$$