CSSS Math Camp Lecture 5

Introduction to Probability

Authored by: Laina Mercer, PhD

Erin Lipman and Jess Kunke

Department of Statistics

September 15, 2023

Probability

Motivating Example

A disease has a prevalence of 1% in the population. A blood test for the disease has high sensitivity (the probability of a positive test if someone is sick) and specificity (the probability of a negative test if someone is not sick).

- If someone has the disease, there is a 98% chance they will test positive.
- If someone does not have the disease, there is a 95% chance they will test negative

Suppose you test positive for the disease and you want to figure out the probability that you have the disease. That is, given someone has tested positive for the disease what is the chance that they have the disease?

Probability

Motivating Example

What information do we have?

- $P(+\text{ test}|\text{ diseased}) = 0.98$
- $P(-\text{ test}|\text{ healthy}) = 0.95$

What quantity do we want?

- $P(\text{ diseased}| + \text{ test})$

So, what is the probability of disease given a positive test?

What information do we have?

- $P(+\text{ test}|\text{ diseased}) = 0.98$
- $P(-\text{ test}|\text{ healthy}) = 0.95$

What quantity do we want?

- $P(\text{ diseased}|+\text{ test})$

So, what is the probability of disease given a positive test? 16.5%

What information do we have?

- $P(+\text{ test}|\text{ diseased}) = 0.98$
- $P(-\text{ test}|\text{ healthy}) = 0.95$

What quantity do we want?

- $P(\text{ diseased}|+\text{ test})$

So, what is the probability of disease given a positive test? 16.5%

- Assumes no additional info about the person being tested (e.g. symptoms)
- We'll build up the tools to explain this result!

Probability vs. statistics

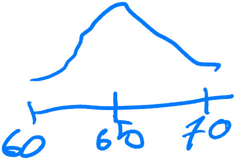
- Probability: we make a model for how uncertain/noisy/random data is generated
- Statistics: we use a probability model to **infer** how the data is generated, or to infer something about a population given only a sample

make model for how data was generated

data → model → population quantities

inference

ex heights



ex avg height in whole pop

Motivation

$\boxed{\text{population}} \xrightarrow{\text{random}} \boxed{\text{sample}}$

- Statistics enables us to draw conclusions about a **population** from a **sample**.

  For example we use the sample mean to estimate the population mean.

- Our estimator will be **random**: each time we perform the experiment, we will get a different sample and a different value.
- By understanding the probability distribution of our estimator, we can express its uncertainty (i.e. confidence intervals).

Lecture 5: Probability

Set notation and concepts

Probability and sampling

Conditional probability and independence

Random variables

Expectation, variance, and other properties

Discrete vs. continuous random variables

Lecture 5: Probability

Set notation and concepts

Probability and sampling

Conditional probability and independence

Random variables

Expectation, variance, and other properties

Discrete vs. continuous random variables

Set Notation

A **set** is a collection of elements from a population.

Examples $\nearrow$ population

- Positive Integers ≤ 5 : $A = \{1, 2, 3, 4, 5\}$
- Primary Colors: B={blue, red, yellow }
- Odd Numbers: $C = \{1, 3, 5, 7, 9...\}$

$\emptyset$

A set is an empty set if it contains no elements: written $\emptyset$ or ~~$D = \{\emptyset\}$~~. An example of an empty set would be integers that are greater than 4 and less than 1.

$\hookrightarrow$ $D = \{\,\}$ $\quad$ $D = \emptyset$

A set is called the universal set if it contains all the elements in the population: written $\boxed{\Omega}$ or ~~$E = \{\Omega\}$~~

$\hookrightarrow$ universal set $\nearrow$ population
$=$ all elements

Set Notation

Intersection, $\cap$

The **intersection** of two sets A, B is the set of all elements that are in A **AND** B. The intersection is denoted $A \cap B$.

Find the intersection:

- $A = \{1, 2, 3, 4, 5\}$, $B = \{2, 4, 6, 8, 10\}$

  $A \cap B = \{2, 4\}$

- $A = \{$ Odd numbers $\}$, $B = \{$ Even numbers $\}$ → mutually exclusive

  $A \cap B = \emptyset$ $\quad$ $A \cap B = \{\ \}$

- $A = \{$ Integers less than 5 $\}$, $B = \{$ Integers greater than 2$\}$

  $A \cap B = \{3, 4\}$

Set Notation

Intersection, $\cap$

The **intersection** of two sets A, B is the set of all elements that are in A **AND** B. The intersection is denoted $A \cap B$.

Find the intersection:

- $A = \{1, 2, 3, 4, 5\}$, $B = \{2, 4, 6, 8, 10\}$
- $A = \{$ Odd numbers $\}$, $B = \{$ Even numbers $\}$
- $A = \{$ Integers less than 5 $\}$, $B = \{$ Integers greater than 2$\}$

If the intersection of A and B is $\emptyset$ (no elements in common), A and B are called **mutually exclusive**. Were any of the pairs of sets above mutually exclusive?

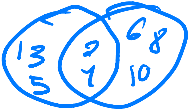
Set Notation

Union, $\cup$

Reals $\mathbb{R}$ integers $\mathbb{Z}$ natural numbers $\mathbb{N}$

The **union** of two sets A, B is the set of all elements that are in A **OR** B. The intersection is denoted $A \cup B$.

Find the union:



integers

- $A = \{1, 2, 3, 4, 5\}$, $B = \{2, 4, 6, 8, 10\}$

$A \cup B = \{1, 2, 3, 4, 5, 6, 8, 10\}$

$A \cup B = \mathbb{Z}$

- $A = \{$ Odd numbers $\}$, $B = \{$ Even numbers $\}$

$A \cup B = \{\text{all integers}\} = \{\ldots, -2, -1, 0, 1, 2, \ldots\}$

- $A = \{$ Integers less than 5 $\}$, $B = \{$ Integers greater than 2$\}$

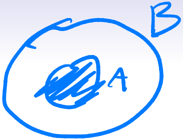
$A \cup B = \{\text{all integers}\}$

A, B is a partition of the integers

$A \cup B = \Omega$ where Ω is universal set of integers

Set Notation

Subest, $\subseteq$



$A \subseteq B$

One set can be contained inside another. If all elements of A are also in B, then A is a **subset** of B. The subset is denoted $A \subseteq B$.

Is A a subset of B?

- $A = \{1, 5\}$, $B = \{1, 2, 3, 4, 5\}$ $A \subseteq B$ (yes)
- $A = \{1, \cancel{5}\}$, $B = \{1, 4, 9, 18\}$ $A \not\subseteq B$ (No)

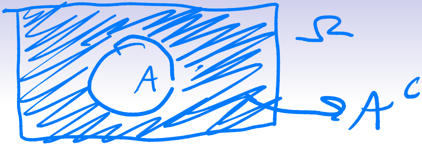
If $A \subseteq B$ AND $B \subseteq A$ then all of the elements in A are in B and all of the elements of B are in A, so $A = B$.

ex $A = \{1, 5\}$ $B = \{1, 5\}$

$A \subseteq B$
$B \subseteq A$
$A = B$

Set Notation

Complement



The **complement** of a set is the set of elements in the population that are not in A. The complement is denoted by A^c.

Example: Suppose the population is all the possible numbers you can get from rolling a fair six-sided die (where the faces are numbered 1 through 6). If $A = \{1, 3\}$, what is A^c?

$$A^c = \{2, 4, 5, 6\}$$

$$A \cup A^c = \Omega = \{1, 2, 3, 4, 5, 6\}$$

Lecture 5: Probability

Set notation and concepts

Probability and sampling

Conditional probability and independence

Random variables

Expectation, variance, and other properties

Discrete vs. continuous random variables

Experiments and sample spaces

An **experiment** is an action or process of observation. It has only one outcome but we do not know what it will be with certainty until the experiment is carried out.

Examples: rolling a die, flipping a coin, drawing a survey sample

The **sample space** is made up of all the possible outcomes of the experiment and usually denoted by S.

Experiments and sample spaces

An **experiment** is an action or process of observation. It has only one outcome but we do not know what it will be with certainty until the experiment is carried out.

Examples: rolling a die, flipping a coin, drawing a survey sample

The **sample space** is made up of all the possible outcomes of the experiment and usually denoted by S.

Example: For our experiment, we flip a coin three times.

$$S = \{HHH, TTT, HHT, THH, HTH, THT, HTT, TTH\}$$

$$2 \times 2 \times 2 = 2^3 = 8$$

An **event** is a subset of the sample space.

Ω

Examples:
$S = \{HHH, HHT, HTH, THH, TTH, THT, HTT, TTT\}$

- Getting 2 heads: $\{HHT, HTH, THH\}$
- Getting an odd number of tails:
  $\{HHT, HTH, THH, TTT\}$
- Getting more than 1 head:

Probability

The final layer is to specify the probability of each outcome. This is the probability distribution.

There are two key properties that a probability distribution (the probabilities of all the outcomes) must satisfy:

- The probability of any given outcome must be ≥ 0 and ≤ 1
- The probabilities must sum to 1

Probability

The final layer is to specify the probability of each outcome. This is the probability distribution.

There are two key properties that a probability distribution (the probabilities of all the outcomes) must satisfy:

- The probability of any given outcome must be ≥ 0 and ≤ 1
- The probabilities must sum to 1

Suppose we are rolling a six-sided die. Which of the following are valid probability distributions?

outcome	1	2	3	4	5	6
p	1/6	1/6	1/6	1/6	1/6	1/6

Probability of an event

We can find the probability of an event by adding up the probabilities of the elements in the event. Rolling a fair die:

outcome versus event

ex

outcome: TTH

event: less than 2 heads

$S = \{1, 2, 3, 4, 5, 6\}$
Each element has probability $1/6$.

Find the probability of each event:

- $A = \{\text{roll} \leq 4\} = \{1, 2, 3, 4\}$

$$P(A) = P(1) + P(2) + P(3) + P(4)$$
$$= \frac{1}{6} + \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{4}{6} = \frac{2}{3}$$

- $B = \{\text{roll odd}\} = \{1, 3, 5\}$

$$P(B) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{3}{6} = \frac{1}{2}$$

Probability
Unions

We can find the probability of the union of two events.

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

We are adding up the probability of the elements in the even A and the elements in the event B. However, in doing that we count the elements that are in both A and B ($A \cap B$) twice. So we must subtract the intersection of A and B.

$P(A \cup B) = \{1, 2, 3, 4, 5\}$
$= \frac{5}{6}$

We can find the probability of the union of two events.

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

We are adding up the probability of the elements in the even A and the elements in the event B. However, in doing that we count the elements that are in both A and B ($A \cap B$) twice. So we must subtract the intersection of A and B.

Example: let's consider again rolling a fair die, so $S = \{1, 2, 3, 4, 5, 6\}$ and each element has probability $1/6$. Let $A = \{\text{roll} \leq 4\}, B = \{1, 3, 5\}$. $P(A \cup B) = ?$

$P(A) = P(\{1, 2, 3, 4\}) = \frac{4}{6}$

$P(B) = P(\{1, 3, 5\}) = \frac{3}{6}$

$P(A \cap B) = P(\{1, 3\}) = \frac{2}{6}$

$P(A \cup B) = \frac{4}{6} + \frac{3}{6} - \frac{2}{6}$
$= \frac{5}{6}$

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

Selecting cards from a deck (52 total cards):
Each element has probability $1/52$.

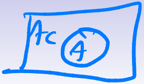
$A = \{\text{Hearts}\}$ and $B = \{\text{King}\}$. $P(A \cup B) = ?$

$P(A) = \frac{13}{52}$ $\quad P(B) = \frac{4}{52}$

$P(A \cap B) = P(\text{King of hearts}) = \frac{1}{52}$

$$P(A \cup B) = \frac{13}{52} + \frac{4}{52} - \frac{1}{52} = \frac{16}{52} = \frac{4}{13}$$

Probability

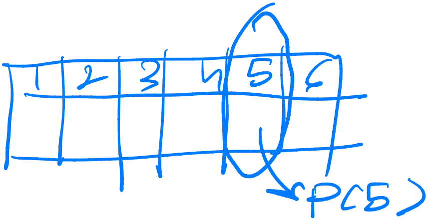


The sum of the probabilities of all elements in the sample space MUST be 1. Since the complement of a set A is everything in the population that is not in A, the probability of the complement can be found:

$$P(A^c) = 1 - P(A)$$

Example: If I roll a fair six-sided die, what is the probability I don't get a 5?

$$P(\{5\}^c) = 1 - P(\{5\}) = 1 - \frac{1}{6} = \frac{5}{6}$$



Lecture 5: Probability

- Set notation and concepts
- Probability and sampling
- Conditional probability and independence
- Random variables
- Expectation, variance, and other properties
- Discrete vs. continuous random variables

Conditional Probability

Sometimes knowing that one event has occurred changes what you know about the probability of another event. For example if the sidewalk is wet in the morning you might think it is more likely that it rained last night than if you didn't know anything about the sidewalk.

Conditional Probability

Sometimes knowing that one event has occurred changes what you know about the probability of another event. For example if the sidewalk is wet in the morning you might think it is more likely that it rained last night than if you didn't know anything about the sidewalk.

The **conditional probability** of A given B is the probability that A occurs given that B has been observed. It is denoted $P(A|B)$.

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

The denominator is the probability of B occurring. The numerator is the probability of A and B happening.

Conditional Probability

Example

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

Rolling a fair die:

$A = \{\text{odd number}\}$ and $\boxed{B = \{\text{roll} < 4\}}$ $\not\Rightarrow$ converse

What is the probability that your roll is odd if you know that it is less than 4? That is, what is $P(A|B) = P(\text{Odd} \mid < 4)$?

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{2/6}{3/6} = \frac{2}{3}$$

$$P(A) = \frac{1}{2}$$

$\hookrightarrow$ A and B are Not independent

Independence

What if knowing B does not give us any information about A? That is, if $P(A|B) = P(A)$, then we say that A and B are **independent**.

Independence

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

What if knowing B does not give us any information about A? That is, if $\boxed{P(A|B) = P(A)}$, then we say that A and B are **independent**.

Independence also means: $\cancel{P(A|B) = P(A \cap B)}$

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

$P(A) = P(A|B)$ [independence]

$P(A)P(B) = P(A|B)P(B)$

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

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

Independence

$P(B|A) = P(B)$

What if knowing B does not give us any information about A? That is, if $P(A|B) = P(A)$, then we say that A and B are **independent**.

Independence also means: $\cancel{P(A|B) = P(A \cap B)}$

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

Thus, $P(A) \cdot P(B) = P(A \cap B)$ allows us to check for independence.

Independence

Example

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

$P(B) = P(B|A)$

Rolling a fair die:
$A = \{1, 2, 3, 4\}$ and $B = \{\text{odd number}\}$ $A \cap B = \{1, 3\}$
Are A and B independent?

$P(A) = \frac{4}{6}$ $P(B) = \frac{3}{6}$ $P(A \cap B) = \frac{2}{6}$

$P(A|B) = \frac{2/6}{3/6} = \frac{2}{3} = \frac{4}{6}$

$P(A) = \frac{4}{6} = P(A|B)$ → independent

$P(A)P(B) = \left(\frac{4}{6}\right)\left(\frac{3}{6}\right) = \left(\frac{2}{3}\right)\left(\frac{1}{2}\right) = \frac{2}{6}$

$P(A \cap B) = \frac{2}{6}$ $P(A)P(B) = P(A \cap B)$ → independent

Dependent Events

Example

To check this reasoning, try this similar example yourself: Still rolling a fair die, but now $A = \{1, 2, 3, 5\}$ and $B = \{\text{odd number}\}$ Are A and B independent?

Bayes Rule

Sometimes you may know one conditional probability, but not the other. How can you use the first conditional probability to find the other one?

In general

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

ex

$$P(\text{rain last night} \mid \text{sidewalk is wet}) \neq P(\text{sidewalk wet} \mid \text{rain last night})$$

Bayes Rule

Sometimes you may know one conditional probability, but not the other. How can you use the first conditional probability to find the other one?

$\times P(B)$ $\times P(B)$

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \Rightarrow P(A|B) \cdot P(B) = \underline{P(A \cap B)}$$

$$P(B|A) = \frac{P(A \cap B)}{P(A)} \rightarrow P(B|A)P(A) = \underline{P(A \cap B)}$$

$$\rightarrow P(A|B)P(B) = P(B|A)P(A)$$

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Bayes Rule

Sometimes you may know one conditional probability, but not the other. How can you use the first conditional probability to find the other one?

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \Rightarrow P(A|B) \cdot P(B) = P(A \cap B)$$

$$P(B|A) = \frac{P(A \cap B)}{P(A)} \Rightarrow P(B|A) \cdot P(A) = P(A \cap B)$$

This implies:

Bayes Rule

Sometimes you may know one conditional probability, but not the other. How can you use the first conditional probability to find the other one?

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \Rightarrow P(A|B) \cdot P(B) = P(A \cap B)$$

$$P(B|A) = \frac{P(A \cap B)}{P(A)} \Rightarrow P(B|A) \cdot P(A) = P(A \cap B)$$

This implies:

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

Bayes Rule

Sometimes you may know one conditional probability, but not the other. How can you use the first conditional probability to find the other one?

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \Rightarrow P(A|B) \cdot P(B) = P(A \cap B)$$

$$P(B|A) = \frac{P(A \cap B)}{P(A)} \Rightarrow P(B|A) \cdot P(A) = P(A \cap B)$$

This implies:

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

$$P(A|B) = \frac{P(B|A) \cdot P(A)}{P(B)}$$

Bayes rule

Bayes Rule

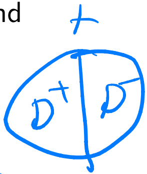
Testing Example

$A = \{+\}$ $\quad B := D^+$

$A^c = \{+\}^c = \{-\}$ $\quad B^c = D^-$

Let's return to our motivating example. What do we know? Let's use D^+ =diseased, D^- =healthy, $+$ =positive test, and $-$ =negative test.

- $P(+|D^+) = 0.98$, $P(-|D^+) = 0.02$
- $P(-|D^-) = 0.95$, $P(+|D^-) = 0.05$
- $P(D^+) = 0.01$, $P(D^-) = 0.99$



$$P(+|D^-) = 1 - P(-|D^-)$$

$$P(D^+|+) = \frac{P(+|D^+)P(D^+)}{P(+)}$$

$$P(+) = P(+ \cap D^+) + P(+ \cap D^-)$$
$$= P(+|D^+)P(D^+) + P(+|D^-)P(D^-)$$

$$P(D^+|+) = \frac{P(+|D^+)P(D^+)}{P(+|D^+)P(D^+) + P(+|D^-)P(D^-)}$$

Bayes Rule

Testing Example

Let's return to our motivating example. What do we know? Let's use D^+ =diseased, D^- =healthy, $+$ =positive test, and $-$ =negative test.

- $P(+|D^+) = 0.98,\ P(-|D^+) = 0.02$
- $P(-|D^-) = 0.95,\ P(+|D^-) = 0.05$
- $P(D^+) = 0.01,\ P(D^-) = 0.99$

$$\begin{aligned}
P(D^+|+) = \frac{P(+|D^+)P(D^+)}{P(+)} &= \frac{P(+|D^+)P(D^+)}{P(+\cap D^+) + P(+\cap D^-)} \\
&= \frac{P(+|D^+)P(D^+)}{P(+|D^+)P(D^+) + P(+|D^-)P(D^-)} \\
&= \frac{0.98 \cdot 0.01}{0.98 \cdot 0.01 + 0.05 \cdot 0.99} \\
&= 0.165
\end{aligned}$$

As the population prevalence increases the $P(D^+|+)$ (this is called the positive predictive value) increases:

Prevalence	$P(D^+\|+)$
0.01	16.5%
0.02	28.6%
0.05	50.8%
0.10	68.5%
0.25	86.7%

Lecture 5: Probability

Set notation and concepts

Probability and sampling

Conditional probability and independence

Random variables

Expectation, variance, and other properties

Discrete vs. continuous random variables

Random Variables

$\underline{X}$ (# of heads)

HHT $\longrightarrow$ 2

HTT $\longrightarrow$ 1

A **random variable** is a function which assigns a number to each element in the sample space. (Think of it as the answer to a question you are asking about each element in the space).

Variable, because the answer will be different for each element.

Random because we can't predict the answer with certainty.

Random variables are usually denoted with capital letters, such as X, Y, Z, and they are often summaries of the data.

Random variables: Examples

- Example: You flip a coin three times.
  - There are 8 possible outcomes of the experiment: HHH, HHT, etc.
  - $X =$ number of heads; there are four possible values of X (0, 1, 2, 3)

Random variables: Examples

- Example: You flip a coin three times.
  - There are 8 possible outcomes of the experiment: HHH, HHT, etc.
  - X = number of heads; there are four possible values of X (0, 1, 2, 3)
- Example: You draw a survey sample to estimate the proportion of the population that voted in the last election.
  - There are many possible samples we could draw!
  - Y = proportion that voted; Y could be one of many* numbers from 0 to 1

ex 10 people

Y can be

0, .1, .2, ..., .9, 1

Random variables: Examples

- Example: You flip a coin three times.
  - There are 8 possible outcomes of the experiment: HHH, HHT, etc.
  - X = number of heads; there are four possible values of X (0, 1, 2, 3)
- Example: You draw a survey sample to estimate the proportion of the population that voted in the last election.
  - There are many possible samples we could draw!
  - Y = proportion that voted; Y could be one of many* numbers from 0 to 1

In both experiments, many outcomes/samples might have the same X or Y value

Probability Distribution

We already introduced probability distributions for the outcomes, and now we can apply the same idea to random variables.

The **probability distribution** of a random variable is a function that assigns a probability to each possible value of X.

When you have finitely many possible values of X, the probability distribution can be written as:

ex $P(X=1)$
$P(X=2)$

X-Value	x_1	x_2	...	x_n
$P(X = x_i)$	p_1	p_2	...	p_n

where each possible value x_i for X is listed with its probability p_i. Find each p_i by summing the probabilities of the outcomes such that $X = x_i$.

Rolling a die (6 choices for 1 roll - 6^1).

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

X=the values of the roll:

X-Value	1	2	3	4	5	6
$P(X = x_i)$	$1/6$	$1/6$	$1/6$	$1/6$	$1/6$	$1/6$

Probability Distribution
Example

Flipping 3 Coins (2 choices for each coin - $2^3 = 8$).

Possible Values
$S = \{HHH, HHT, HTH, THH, TTH, THT, HTT, TTT\}$.

X=The number of Heads

X-Value	0	1	2	3
$P(X = x)$	1/8	3/8	3/8	1/8

$P(X=1) = P(\{HTT, THT, TTH\})$

Probability Distribution

Remember the two necessary conditions for probability distributions:

$$0 \leq p_i \leq 1 \text{ for all } i$$

$$\sum_{i=1}^{n} p_i = 1$$

If these conditions are not met, it is not a valid probability distribution.

Lecture 5: Probability

Set notation and concepts

Probability and sampling

Conditional probability and independence

Random variables

Expectation, variance, and other properties

Discrete vs. continuous random variables

Means & Expectations

Once we have the distribution of a random variable, we can figure out what value of X we would expect to see.

If all of the values are equally likely, we could just take the average.

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

X-Value	0	1	2	3
$P(X = x)$	1/4	1/4	1/4	1/4

$$\bar{X} = \frac{0+1+2+3}{4} = 6/4 = 1.5$$

Note: the value will often not be one of the possible x-values.

$$E(X) = \sum_{i=1}^{4} x_i p_i = 0 \cdot \frac{1}{4} + 1 \cdot \frac{1}{4} + 2 \cdot \frac{1}{4} + 3 \cdot \frac{1}{4} = \frac{0+1+2+3}{4}$$

Means & Expectations

What if all of the values were not equally likely? The average of the numbers should be closer to the values with the highest probability.

We find the **expected value** or **expectation** or **mean** of x, $E[X]$, using a weighted mean:

$$E[X] = \sum_{i=1}^{n} x_i \cdot p_i$$

X-Value	0	1	2	3
$P(X = x)$	0.3	0.45	0.2	0.05

$$E[X] = \sum_{i=1}^{4} x_i \cdot p_i = 0 \cdot 0.30 + 1 \cdot 0.45 + 2 \cdot 0.20 + 3 \cdot 0.05 = 1$$

Variance

The expected value is often the summary we will use of the center of a distribution. The **variance** is the measure we will use of the spread of the distribution.

Like the expectation, the variance can be thought of as a weighted mean.

$$Var[X] = \sum_{i=1}^{n} (x_i - E[X])^2 \cdot p_i$$

- a weighted average of the squared-difference (distance) between the x-values and the mean
- values with higher probability have more weight in the average

Variance

Example

$$Var(X) = \sum_{i=1}^{n} (x_i - E(X))^2 p_i$$

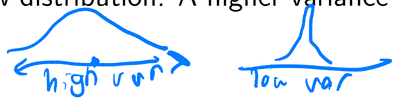
$\boxed{\text{var has units squared} \quad sd(X) = \sqrt{var(X)}}$

X-Value	0	1	2	3
$P(X = x)$	0.3	0.45	0.2	0.05

$$E[X] = \sum_{i=1}^{n} x_i \cdot p_i = 0 \cdot 0.30 + 1 \cdot 0.45 + 2 \cdot 0.20 + 3 \cdot 0.05 = 1$$

$$\begin{aligned} Var[X] &= (0-1)^2 \cdot 0.30 + (1-1)^2 \cdot 0.45 + (2-1)^2 \cdot 0.20 + (3-1)^2 \cdot 0.05 \\ &= 1^2 \cdot 0.30 + 0 \cdot 0.45 + 1^2 \cdot 0.20 + 2^2 \cdot 0.05 \\ &= 0.30 + 0.20 + 0.20 = 0.70 \end{aligned}$$

Low variance means a tight/narrow distribution. A higher variance means a wider, flatter distribution.



Linear Functions of Random Variables

Sometimes we are interested in functions of random variables.

For example, if we know the mean and variance of X, do we know anything about $Y = 2X$ or $Y = X + 4$?

Or if we have a temperature information in Fahrenheit, can we find the mean and variance in Celsius?

As it turns out, expectations (or averages) are linear. If every number undergoes the same transformation, so does the mean.

Linear Functions of Random Variables

Mean Example

If $E[X] = 1$, then the average of the random variable X is 1. Now let's look at the random variable $Y = X + 4$.

What is $E[Y]$?

$X + 4$ add the constant 4 to every value of X

X-Value	0	1	2	3
$P(X = x)$	0.3	0.45	0.2	0.05

Y-Value	4	5	6	7
$P(Y = y)$	0.3	0.45	0.2	0.05

$$E[Y] = 4 \cdot 0.30 + 5 \cdot 0.45 + 6 \cdot 0.20 + 7 \cdot 0.05 = 1.20 + 2.25 + 1.20 + 0.35 = 5$$

$$E[Y] = E[X + 4] = E[X] + 4 = 1 + 4 = 5$$

Linear Functions of Random Variables

Mean Example

If $E[X] = 1$, then the average of the random variable X is 1. Now let's look at the random variable $Y = 2X$.
What is $E[Y]$?
$2X$ multiplies every X by 2:

X-Value	0	1	2	3
$P(X = x)$	0.3	0.45	0.2	0.05

Y-Value	0	2	4	6
$P(Y = y)$	0.3	0.45	0.2	0.05

$$E[Y] = 0 \cdot 0.30 + 2 \cdot 0.45 + 4 \cdot 0.20 + 6 \cdot 0.05 = 0.90 + 0.80 + 0.30 = 2$$

$$E[Y] = E[2X] = 2E[X] = 2 \cdot 1 = 2$$

Linear Functions of Random Variables

Variance Example, Y=X+4

X-Value	0	1	2	3
$P(X = x)$	0.3	0.45	0.2	0.05

Y-Value	4	5	6	7
$P(Y = y)$	0.3	0.45	0.2	0.05

$$\begin{aligned}
Var[X] &= (0-1)^2 \cdot 0.30 + (1-1)^2 \cdot 0.45 + (2-1)^2 \cdot 0.20 + (3-1)^2 \cdot 0.05 \\
&= 1^2 \cdot 0.30 + 0 \cdot 0.45 + 1^2 \cdot 0.20 + 2^2 \cdot 0.05 \\
&= 0.30 + 0.20 + 0.20 = 0.70 \\
Var[Y] &= (4-5)^2 \cdot 0.30 + (5-5)^2 \cdot 0.45 + (6-5)^2 \cdot 0.20 + (7-5)^2 \cdot 0.05 \\
&= 1^2 \cdot 0.30 + 0 \cdot 0.45 + 1^2 \cdot 0.20 + 2^2 \cdot 0.05 \\
&= 0.30 + 0.20 + 0.20 = 0.70
\end{aligned}$$

If we add 4 to every number, the spread is completely unaffected.
$Var[Y] = Var[X+4] = Var[X]$.

Linear Functions of Random Variables

Variance Example, Y=2X

X-Value	0	1	2	3
$P(X=x)$	0.3	0.45	0.2	0.05

Y-Value	0	2	4	6
$P(Y=y)$	0.3	0.45	0.2	0.05

$$
\begin{aligned}
Var[X] &= (0-1)^2 \cdot 0.30 + (1-1)^2 \cdot 0.45 + (2-1)^2 \cdot 0.20 + (3-1)^2 \cdot 0.05 \\
&= 1^2 \cdot 0.30 + 0 \cdot 0.45 + 1^2 \cdot 0.20 + 2^2 \cdot 0.05 \\
&= 0.30 + 0.20 + 0.20 = 0.70 \\
Var[Y] &= (0-2)^2 \cdot 0.30 + (2-2)^2 \cdot 0.45 + (4-2)^2 \cdot 0.20 + (6-2)^2 \cdot 0.05 \\
&= 2^2 \cdot 0.30 + 0 \cdot 0.45 + 2^2 \cdot 0.20 + 4^2 \cdot 0.05 \\
&= 1.2 + 0.80 + 0.80 = 2.80
\end{aligned}
$$

If we multiply every value by 2 we quadruple the spread.
$Var[Y] = Var[2X] = 2^2 Var[X]$.

Summary of Expectations & Variances

data	pop
$\bar{x}$	$E(x)$
s^2	$Var(x)$

$$E[aX + b] = aE[X] + b$$

$$Var[aX + b] = a^2 Var[X]$$

Extra practice: Find the expectation and variance of Y.

- Suppose January temperatures in Seattle have $E[F] = 60$, $Var[F] = 2$ in degrees Fahrenheit. What about in Celsius, $C = 0.56F - 17.77$?

$$E(C) = E(.56F - 17.77)$$
$$= .56E(F) - 17.77$$

$$Var(C) = Var(.56F - 17.77)$$
$$= (.56)^2 Var(F)$$

Lecture 5: Probability

Set notation and concepts

Probability and sampling

Conditional probability and independence

Random variables

Expectation, variance, and other properties

Discrete vs. continuous random variables

Probability distribution functions

Random Variables can be discrete or continuous.

A **discrete random variable** only gives values that you can list or count.

- Previous examples were all discrete distributions.

A **continuous random variable** gives an infinite number of values in a range.

- We can't list all of the possible values for a continuous random variable.

Probability distribution functions

Random Variables can be discrete or continuous.

A **discrete random variable** only gives values that you can list or count.

- Previous examples were all discrete distributions.

A **continuous random variable** gives an infinite number of values in a range.

- We can't list all of the possible values for a continuous random variable.

Examples

- X-number of books in a grad student's office.
  DISCRETE $X = 0, 1, 2, 3, 4, ...$
- Y - the amount of water it takes to fill a pool.
  CONTINUOUS $Y \in (0, \infty)$
- Z - the time it takes to finish a task.
  CONTINUOUS $Z \in (0, \infty)$

Probability mass functions

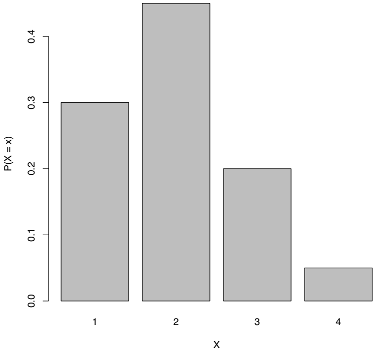
The **probability distribution function** is defined differently for discrete and continuous random variables.

Discrete Random Variables

- **probability mass function** (pmf)
- Countable number of outcomes n, can write down

$$P(X = x_i) \;\forall\, x_i, \;\; i = 1, \ldots, n$$

.

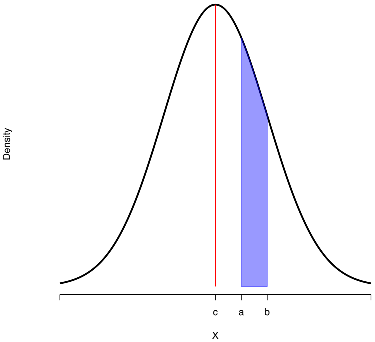


Probability density functions

The **probability distribution function** is defined differently for discrete and continuous random variables.

Continuous Random Variables

- **probability density function** (pdf)
- $P(X = c) = 0$ for any value of x.
- $P(a < X < b) > 0$, if X can take on at least some of the values in that interval.



Probability distribution examples

Binomial:

- Probability of getting k heads out of n coin tosses when the prob. of heads is p

Normal: