Week 2: Random Variables

Brandon Stewart[1]

Princeton

September 17/19, 2018

[1] These slides are heavily influenced by Adam Glynn, Justin Grimmer and Jens Hainmueller. Many illustrations by Shay O'Brien.

Where We've Been and Where We're Going...

- Last Week

Where We've Been and Where We're Going...

- Last Week
  - welcome and outline of course
  - described uncertain outcomes with probability.

Where We've Been and Where We're Going...

- Last Week
  - welcome and outline of course
  - described uncertain outcomes with probability.
- This Week

Where We've Been and Where We're Going...

- Last Week
  - welcome and outline of course
  - described uncertain outcomes with probability.
- This Week
  - Monday:
    - summarize one random variable using expectation and variance

Where We've Been and Where We're Going...

- Last Week
  - welcome and outline of course
  - described uncertain outcomes with probability.
- This Week
  - Monday:
    - summarize one random variable using expectation and variance
    - show how to condition on a variable

Where We've Been and Where We're Going...

- Last Week
  - welcome and outline of course
  - described uncertain outcomes with probability.
- This Week
  - Monday:
    - summarize one random variable using expectation and variance
    - show how to condition on a variable
  - Wednesday:
    - properties of joint distributions

Where We've Been and Where We're Going...

- Last Week
  - welcome and outline of course
  - described uncertain outcomes with probability.
- This Week
  - Monday:
    - summarize one random variable using expectation and variance
    - show how to condition on a variable
  - Wednesday:
    - properties of joint distributions
    - conditional expectations

Where We've Been and Where We're Going...

- Last Week
  - welcome and outline of course
  - described uncertain outcomes with probability.
- This Week
  - Monday:
    - summarize one random variable using expectation and variance
    - show how to condition on a variable
  - Wednesday:
    - properties of joint distributions
    - conditional expectations
    - covariance, correlation, independence

Where We've Been and Where We're Going...

- Last Week
  - welcome and outline of course
  - described uncertain outcomes with probability.
- This Week
  - Monday:
    - summarize one random variable using expectation and variance
    - show how to condition on a variable
  - Wednesday:
    - properties of joint distributions
    - conditional expectations
    - covariance, correlation, independence
- Next Week

Where We've Been and Where We're Going...

- Last Week
  - welcome and outline of course
  - described uncertain outcomes with probability.
- This Week
  - Monday:
    - summarize one random variable using expectation and variance
    - show how to condition on a variable
  - Wednesday:
    - properties of joint distributions
    - conditional expectations
    - covariance, correlation, independence
- Next Week
  - estimating these features from data

Where We've Been and Where We're Going...

- Last Week
  - welcome and outline of course
  - described uncertain outcomes with probability.
- This Week
  - Monday:
    - summarize one random variable using expectation and variance
    - show how to condition on a variable
  - Wednesday:
    - properties of joint distributions
    - conditional expectations
    - covariance, correlation, independence
- Next Week
  - estimating these features from data
  - estimating uncertainty

Where We've Been and Where We're Going...

- Last Week
  - welcome and outline of course
  - described uncertain outcomes with probability.
- This Week
  - Monday:
    - summarize one random variable using expectation and variance
    - show how to condition on a variable
  - Wednesday:
    - properties of joint distributions
    - conditional expectations
    - covariance, correlation, independence
- Next Week
  - estimating these features from data
  - estimating uncertainty
- Long Run

Where We've Been and Where We're Going...

- Last Week
  - welcome and outline of course
  - described uncertain outcomes with probability.
- This Week
  - Monday:
    - summarize one random variable using expectation and variance
    - show how to condition on a variable
  - Wednesday:
    - properties of joint distributions
    - conditional expectations
    - covariance, correlation, independence
- Next Week
  - estimating these features from data
  - estimating uncertainty
- Long Run
  - probability $\rightarrow$ inference $\rightarrow$ regression $\rightarrow$ causal inference

Where We've Been and Where We're Going...

- Last Week
  - welcome and outline of course
  - described uncertain outcomes with probability.
- This Week
  - Monday:
    - summarize one random variable using expectation and variance
    - show how to condition on a variable
  - Wednesday:
    - properties of joint distributions
    - conditional expectations
    - covariance, correlation, independence
- Next Week
  - estimating these features from data
  - estimating uncertainty
- Long Run
  - probability $\rightarrow$ inference $\rightarrow$ regression $\rightarrow$ causal inference

Questions?

Admin

Admin

- Notation guide

Admin

- Notation guide
- Using the slides (links, what's contained in a single deck etc.)

Admin

- Notation guide
- Using the slides (links, what's contained in a single deck etc.)
- Any logistical hiccups?

1. Random Variables and Distributions
   - What is a Random Variable?
   - Discrete Distributions
   - Continuous Distributions
2. Characteristics of Distributions
   - Central Tendency
   - Measures of Dispersion
3. Conditional Distributions
4. Fun with Averages
5. Fun with Sensitive Questions
6. Appendix: Why the Mean?
7. Joint Distributions
   - Discrete Random Variable
   - Continuous Random Variable
8. Conditional Expectation
9. Properties
   - Independence
   - Covariance and Correlation
   - Conditional Independence
10. Famous Distributions
11. Fun With Spam

1. Random Variables and Distributions
   - What is a Random Variable?
   - Discrete Distributions
   - Continuous Distributions
2. Characteristics of Distributions
   - Central Tendency
   - Measures of Dispersion
3. Conditional Distributions
4. Fun with Averages
5. Fun with Sensitive Questions
6. Appendix: Why the Mean?
7. Joint Distributions
   - Discrete Random Variable
   - Continuous Random Variable
8. Conditional Expectation
9. Properties
   - Independence
   - Covariance and Correlation
   - Conditional Independence
10. Famous Distributions
11. Fun With Spam

Example: Ballot Order

Example: Ballot Order

Evidence suggests that candidates gain a small advantage from ballot order.

Example: Ballot Order

Evidence suggests that candidates gain a small advantage from ballot order.

As a response, in 2008 New Hampshire chose a letter from the alphabet and then listed the candidates in alphabetical order starting with that letter.

Example: Ballot Order

Evidence suggests that candidates gain a small advantage from ballot order.

As a response, in 2008 New Hampshire chose a letter from the alphabet and then listed the candidates in alphabetical order starting with that letter.

We can use probability to assess the "fairness" of this process.

Example: Ballot Order

Evidence suggests that candidates gain a small advantage from ballot order.

As a response, in 2008 New Hampshire chose a letter from the alphabet and then listed the candidates in alphabetical order starting with that letter.

We can use probability to assess the "fairness" of this process.

We will do this by introducing a random variable X to be Barack Obama's position on the 2008 New Hampshire primary ballot.

Example: Assessing Racial Prejudice

Example: Assessing Racial Prejudice

- We often want to ask sensitive questions which a survey respondent is unlikely to honestly answer

Example: Assessing Racial Prejudice

- We often want to ask sensitive questions which a survey respondent is unlikely to honestly answer
- A list experiment asks respondents how many items on a list they agree with

Example: Assessing Racial Prejudice

- We often want to ask sensitive questions which a survey respondent is unlikely to honestly answer
- A list experiment asks respondents how many items on a list they agree with
  - for example, what proportion of people would be upset by a black family moving in next door to them (Kuklinski et al 1997).

Example: Assessing Racial Prejudice

- We often want to ask sensitive questions which a survey respondent is unlikely to honestly answer
- A list experiment asks respondents how many items on a list they agree with
  - for example, what proportion of people would be upset by a black family moving in next door to them (Kuklinski et al 1997).
  - randomly split survey into two halves

Example: Assessing Racial Prejudice

- We often want to ask sensitive questions which a survey respondent is unlikely to honestly answer
- A list experiment asks respondents how many items on a list they agree with
  - for example, what proportion of people would be upset by a black family moving in next door to them (Kuklinski et al 1997).
  - randomly split survey into two halves
  - first half ask how many of the following items upset you:

Example: Assessing Racial Prejudice

- We often want to ask sensitive questions which a survey respondent is unlikely to honestly answer
- A list experiment asks respondents how many items on a list they agree with
  - for example, what proportion of people would be upset by a black family moving in next door to them (Kuklinski et al 1997).
  - randomly split survey into two halves
  - first half ask how many of the following items upset you:
    1. the federal government increasing the tax on gasoline

Example: Assessing Racial Prejudice

- We often want to ask sensitive questions which a survey respondent is unlikely to honestly answer
- A list experiment asks respondents how many items on a list they agree with
  - for example, what proportion of people would be upset by a black family moving in next door to them (Kuklinski et al 1997).
  - randomly split survey into two halves
  - first half ask how many of the following items upset you:
    1. the federal government increasing the tax on gasoline
    2. professional athletes getting million-dollar salaries

Example: Assessing Racial Prejudice

- We often want to ask sensitive questions which a survey respondent is unlikely to honestly answer
- A list experiment asks respondents how many items on a list they agree with
  - for example, what proportion of people would be upset by a black family moving in next door to them (Kuklinski et al 1997).
  - randomly split survey into two halves
  - first half ask how many of the following items upset you:
    1. the federal government increasing the tax on gasoline
    2. professional athletes getting million-dollar salaries
    3. large corporations polluting the environment.

Example: Assessing Racial Prejudice

- We often want to ask sensitive questions which a survey respondent is unlikely to honestly answer
- A list experiment asks respondents how many items on a list they agree with
  - for example, what proportion of people would be upset by a black family moving in next door to them (Kuklinski et al 1997).
  - randomly split survey into two halves
  - first half ask how many of the following items upset you:
    1. the federal government increasing the tax on gasoline
    2. professional athletes getting million-dollar salaries
    3. large corporations polluting the environment.
  - second half, add a fourth item

Example: Assessing Racial Prejudice

- We often want to ask sensitive questions which a survey respondent is unlikely to honestly answer
- A list experiment asks respondents how many items on a list they agree with
  - for example, what proportion of people would be upset by a black family moving in next door to them (Kuklinski et al 1997).
  - randomly split survey into two halves
  - first half ask how many of the following items upset you:
    1. the federal government increasing the tax on gasoline
    2. professional athletes getting million-dollar salaries
    3. large corporations polluting the environment.
  - second half, add a fourth item
    4. a black family moving in next door

Example: Assessing Racial Prejudice

- We often want to ask sensitive questions which a survey respondent is unlikely to honestly answer
- A list experiment asks respondents how many items on a list they agree with
  - for example, what proportion of people would be upset by a black family moving in next door to them (Kuklinski et al 1997).
  - randomly split survey into two halves
  - first half ask how many of the following items upset you:
    1. the federal government increasing the tax on gasoline
    2. professional athletes getting million-dollar salaries
    3. large corporations polluting the environment.
  - second half, add a fourth item
    4. a black family moving in next door
  - use the answers to infer the proportion upset by the fourth item.

Example: Assessing Racial Prejudice

- We often want to ask sensitive questions which a survey respondent is unlikely to honestly answer
- A list experiment asks respondents how many items on a list they agree with
  - for example, what proportion of people would be upset by a black family moving in next door to them (Kuklinski et al 1997).
  - randomly split survey into two halves
  - first half ask how many of the following items upset you:
    1. the federal government increasing the tax on gasoline
    2. professional athletes getting million-dollar salaries
    3. large corporations polluting the environment.
  - second half, add a fourth item
    4. a black family moving in next door
  - use the answers to infer the proportion upset by the fourth item.
- To do this we need to understand random variables

What is a Random Variable?

Intuition: functions that map outcomes to numbers.

What is a Random Variable?

Intuition: functions that map outcomes to numbers.

Formal: X is a function that maps the sample space to the real numbers.

What is a Random Variable?

Intuition: functions that map outcomes to numbers.

Formal: X is a function that maps the sample space to the real numbers.

Imagine an experiment of two coin flips

$$\Omega = \big\{\{heads, heads\}, \{heads, tails\}, \{tails, heads\}, \{tails, tails\}\big\}$$

What is a Random Variable?

Intuition: functions that map outcomes to numbers.

Formal: X is a function that maps the sample space to the real numbers.

Imagine an experiment of two coin flips

$$\Omega = \big\{\{heads, heads\}, \{heads, tails\}, \{tails, heads\}, \{tails, tails\}\big\}$$

we could define a random variable $X(\omega)$ to be the function that returns the number of heads for each element of Ω.

What is a Random Variable?

Intuition: functions that map outcomes to numbers.

Formal: X is a function that maps the sample space to the real numbers.

Imagine an experiment of two coin flips

$$\Omega = \big\{\{heads, heads\}, \{heads, tails\}, \{tails, heads\}, \{tails, tails\}\big\}$$

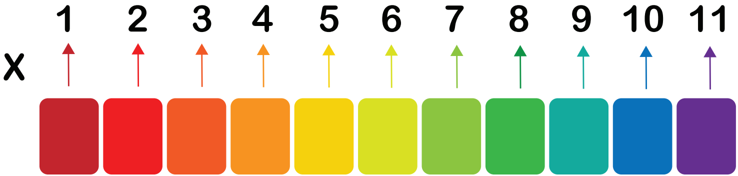
we could define a random variable $X(\omega)$ to be the function that returns the number of heads for each element of Ω.

- $X(\{heads, heads\}) = 2$
- $X(\{heads, tails\}) = 1$
- $X(\{tails, heads\}) = 1$
- $X(\{tails, tails\}) = 0$

A Visual Example

A Visual Example

A Visual Example



A Brief Note on Notation

A Brief Note on Notation

- We almost always use capital roman letters for the “name” of the random variable such as X

A Brief Note on Notation

- We almost always use capital roman letters for the "name" of the random variable such as X
- We refer to a particular value with a lower case letter x

A Brief Note on Notation

- We almost always use capital roman letters for the “name” of the random variable such as X
- We refer to a particular value with a lower case letter x
- So we might write $P(X = x)$ to be the probability that the number of heads we observe is equal to x.

A Brief Note on Notation

- We almost always use capital roman letters for the "name" of the random variable such as X
- We refer to a particular value with a lower case letter x
- So we might write $P(X = x)$ to be the probability that the number of heads we observe is equal to x.
- For more complicated random variables we often write out values as follows

$$X = \begin{cases} 1 \text{ if heads} \\ 0 \text{ if tails} \end{cases}$$

A Brief Note on Notation

- We almost always use capital roman letters for the "name" of the random variable such as X
- We refer to a particular value with a lower case letter x
- So we might write $P(X = x)$ to be the probability that the number of heads we observe is equal to x.
- For more complicated random variables we often write out values as follows

$$X = \begin{cases} 1 \text{ if heads} \\ 0 \text{ if tails} \end{cases}$$

- Sometimes the sample space is already numeric so its more obvious (e.g. how long until the train arrives)

Quick FAQ

Quick FAQ

- Why have random variables at all?

Quick FAQ

- Why have random variables at all?
  it makes the math easier, even across very different sample spaces.

Quick FAQ

- Why have random variables at all?
  it makes the math easier, even across very different sample spaces.
- Why are they random variables?

Quick FAQ

- Why have random variables at all?
  it makes the math easier, even across very different sample spaces.
- Why are they random variables?
  realizations of a stochastic process (i.e. randomness in the outcome, not the mapping)

Quick FAQ

- Why have random variables at all?
  it makes the math easier, even across very different sample spaces.
- Why are they <u>random</u> variables?
  realizations of a stochastic process (i.e. randomness in the outcome, not the mapping)
- Is it really easier this way? It seems hard.

Quick FAQ

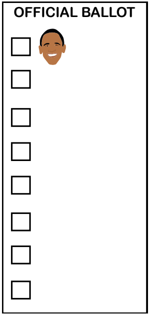
- Why have random variables at all?
  it makes the math easier, even across very different sample spaces.
- Why are they random variables?
  realizations of a stochastic process (i.e. randomness in the outcome, not the mapping)
- Is it really easier this way? It seems hard.
  yep. seriously. let's do an example!

NH Ballot Order Example

Candidates:

- Joe Biden
- Hillary Clinton
- Chris Dodd
- John Edwards
- Mike Gravel
- Dennis Kucinich
- Barack Obama
- Bill Richardson

$$X = \begin{cases} 1 \\ \\ \\ \end{cases}$$



A,B,C,D,E,F,G,H,I,J,K,L,M,N,O,P,Q,R,S,T,U,V,W,X,Y,Z

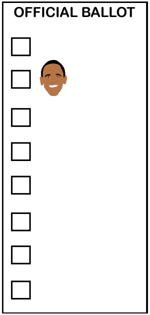
X is a random variable indicating Obama's position on the ballot. Highlighted letters are those leading to a given ballot position. Highlighted individual is first.

NH Ballot Order Example

Candidates:

- Joe Biden
- Hillary Clinton
- Chris Dodd
- John Edwards
- Mike Gravel
- Dennis Kucinich
- Barack Obama
- Bill Richardson

$$X = \begin{cases} 1 \\ 2 \end{cases}$$



A,B,C,D,E,F,G,H,I,J,K,L,M,N,O,P,Q,R,S,T,U,V,W,X,Y,Z

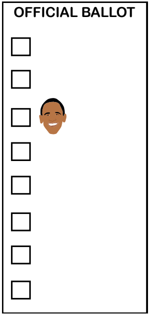
X is a random variable indicating Obama's position on the ballot. Highlighted letters are those leading to a given ballot position. Highlighted individual is first.

NH Ballot Order Example

Candidates:

- Joe Biden
- Hillary Clinton
- Chris Dodd
- John Edwards
- Mike Gravel
- Dennis Kucinich
- Barack Obama
- Bill Richardson

$$X = \begin{cases} 1 \\ 2 \\ 3 \end{cases}$$



A,B,C,D,E,F,G,H,I,J,K,L,M,N,O,P,Q,R,S,T,U,V,W,X,Y,Z

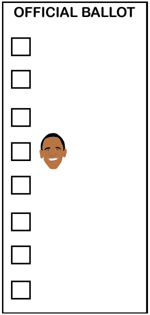
X is a random variable indicating Obama's position on the ballot. Highlighted letters are those leading to a given ballot position. Highlighted individual is first.

NH Ballot Order Example

Candidates:

- Joe Biden
- Hillary Clinton
- Chris Dodd
- John Edwards
- Mike Gravel
- Dennis Kucinich
- Barack Obama
- Bill Richardson

$$X = \begin{cases} 1 \\ 2 \\ 3 \\ 4 \\ \\ \\ \\ \end{cases}$$



A,B,C,D,E,F,G,H,I,J,K,L,M,N,O,P,Q,R,S,T,U,V,W,X,Y,Z

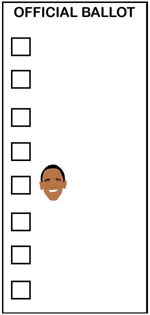
X is a random variable indicating Obama's position on the ballot. Highlighted letters are those leading to a given ballot position. Highlighted individual is first.

NH Ballot Order Example

Candidates:

- Joe Biden
- Hillary Clinton
- Chris Dodd
- John Edwards
- Mike Gravel
- Dennis Kucinich
- Barack Obama
- Bill Richardson

$$X = \begin{cases} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{cases}$$



A,B,C,D,E,F,G,H,I,J,K,L,M,N,O,P,Q,R,S,T,U,V,W,X,Y,Z

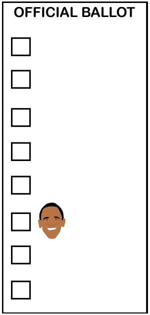
X is a random variable indicating Obama's position on the ballot. Highlighted letters are those leading to a given ballot position. Highlighted individual is first.

NH Ballot Order Example

Candidates:

- Joe Biden
- Hillary Clinton
- Chris Dodd
- John Edwards
- Mike Gravel
- Dennis Kucinich
- Barack Obama
- Bill Richardson

$$X = \begin{cases} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \end{cases}$$



A,B,C,D,E,F,G,H,I,J,K,L,M,N,O,P,Q,R,S,T,U,V,W,X,Y,Z

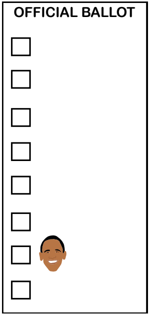
X is a random variable indicating Obama's position on the ballot. Highlighted letters are those leading to a given ballot position. Highlighted individual is first.

NH Ballot Order Example

Candidates:

- Joe Biden
- Hillary Clinton
- Chris Dodd
- John Edwards
- Mike Gravel
- Dennis Kucinich
- Barack Obama
- Bill Richardson

$$X = \begin{cases} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \\ \end{cases}$$



A,B,C,D,E,F,G,H,I,J,K,L,M,N,O,P,Q,R,S,T,U,V,W,X,Y,Z

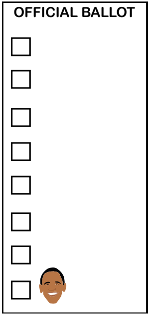
X is a random variable indicating Obama's position on the ballot. Highlighted letters are those leading to a given ballot position. Highlighted individual is first.

NH Ballot Order Example

Candidates:

- Joe Biden
- Hillary Clinton
- Chris Dodd
- John Edwards
- Mike Gravel
- Dennis Kucinich
- Barack Obama
- Bill Richardson

$$X = \begin{cases} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \\ 8 \end{cases}$$



A,B,C,D,E,F,G,H,I,J,K,L,M,N,O,P,Q,R,S,T,U,V,W,X,Y,Z

X is a random variable indicating Obama's position on the ballot. Highlighted letters are those leading to a given ballot position. Highlighted individual is first.

Discrete Distributions

Discrete Distributions

- For discrete distributions, the random variable X takes on a finite, or a countably infinite number of values.

Discrete Distributions

- For discrete distributions, the random variable X takes on a finite, or a countably infinite number of values.
- A common shorthand is to think of discrete RVs taking on distinct values.

Discrete Distributions

- For discrete distributions, the random variable X takes on a finite, or a countably infinite number of values.
- A common shorthand is to think of discrete RVs taking on distinct values.
- A probability mass function (pmf) and a cumulative distribution function (cdf) are two common ways to define the probability distribution for a discrete RV.

Discrete Distributions

- For discrete distributions, the random variable X takes on a finite, or a countably infinite number of values.
- A common shorthand is to think of discrete RVs taking on distinct values.
- A probability mass function (pmf) and a cumulative distribution function (cdf) are two common ways to define the probability distribution for a discrete RV.
- Probability mass functions provide a compact way to represent information about how likely various outcomes are.

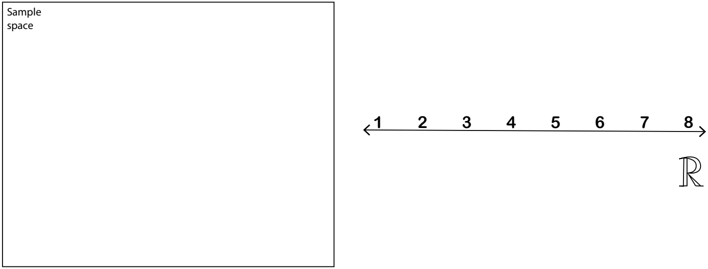
Where do Distributions Come From?

Where do Distributions Come From?

The probabilities associated with each realization of the r.v. come from the underlying experiment and sample space.

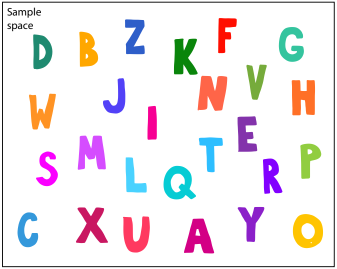
Where do Distributions Come From?

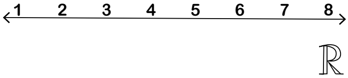
The probabilities associated with each realization of the r.v. come from the underlying experiment and sample space.



Where do Distributions Come From?

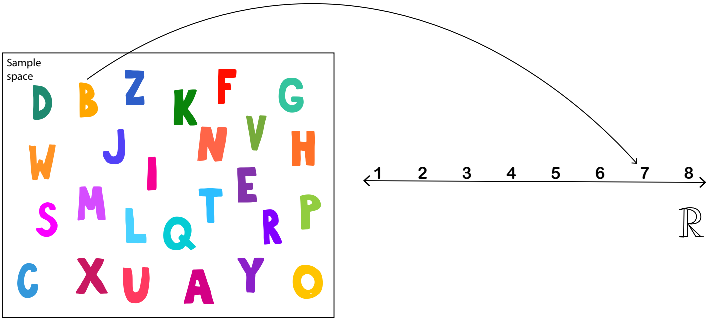
The probabilities associated with each realization of the r.v. come from the underlying experiment and sample space.





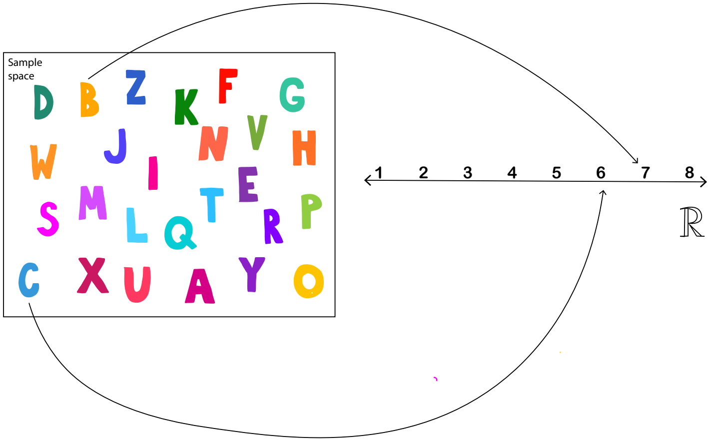
Where do Distributions Come From?

The probabilities associated with each realization of the r.v. come from the underlying experiment and sample space.



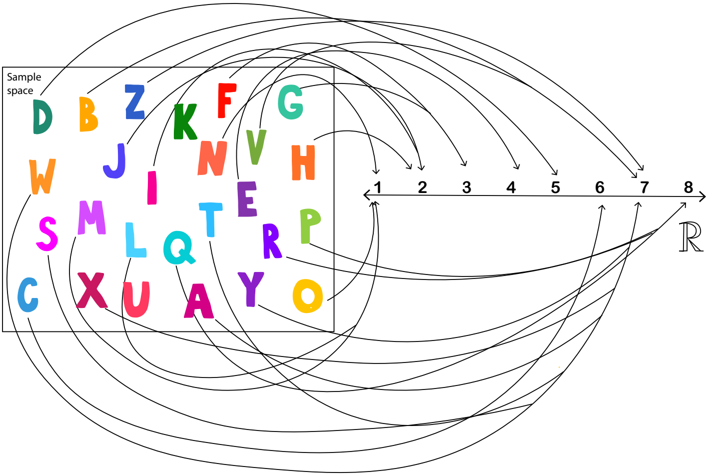
Where do Distributions Come From?

The probabilities associated with each realization of the r.v. come from the underlying experiment and sample space.



Where do Distributions Come From?

The probabilities associated with each realization of the r.v. come from the underlying experiment and sample space.

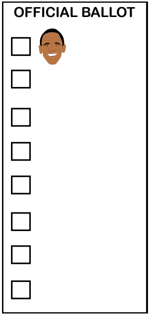


Example: New Hampshire

Candidates:

- Joe Biden
- Hillary Clinton
- Chris Dodd
- John Edwards
- Mike Gravel
- Dennis Kucinich
- Barack Obama
- Bill Richardson

$$f(x) = \begin{cases} 4/26 & x = 1 \\ & \end{cases}$$



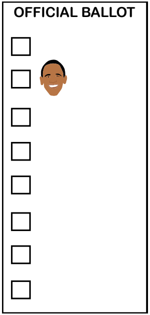
A,B,C,D,E,F,G,H,I,J,K,L,M,N,O,P,Q,R,S,T,U,V,W,X,Y,Z

Example: New Hampshire

Candidates:

- Joe Biden
- Hillary Clinton
- Chris Dodd
- John Edwards
- Mike Gravel
- Dennis Kucinich
- Barack Obama
- Bill Richardson

$$f(x) = \begin{cases} 4/26 & x = 1 \\ 4/26 & x = 2 \end{cases}$$



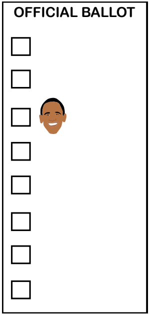
A,B,C,D,E,F,G,H,I,J,K,L,M,N,O,P,Q,R,S,T,U,V,W,X,Y,Z

Example: New Hampshire

Candidates:

- Joe Biden
- Hillary Clinton
- Chris Dodd
- John Edwards
- Mike Gravel
- Dennis Kucinich
- Barack Obama
- Bill Richardson

$$f(x) = \begin{cases} 4/26 & x = 1 \\ 4/26 & x = 2 \\ 2/26 & x = 3 \end{cases}$$



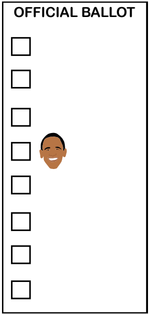
A,B,C,D,E,F,G,H,I,J,K,L,M,N,O,P,Q,R,S,T,U,V,W,X,Y,Z

Example: New Hampshire

Candidates:

- Joe Biden
- Hillary Clinton
- Chris Dodd
- John Edwards
- Mike Gravel
- Dennis Kucinich
- Barack Obama
- Bill Richardson

$$f(x) = \begin{cases} 4/26 & x = 1 \\ 4/26 & x = 2 \\ 2/26 & x = 3 \\ 1/26 & x = 4 \end{cases}$$



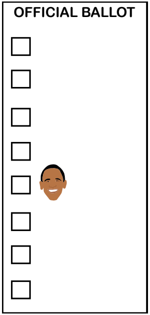
A,B,C,D,E,F,G,H,I,J,K,L,M,N,O,P,Q,R,S,T,U,V,W,X,Y,Z

Example: New Hampshire

Candidates:

- Joe Biden
- Hillary Clinton
- Chris Dodd
- John Edwards
- Mike Gravel
- Dennis Kucinich
- Barack Obama
- Bill Richardson

$$f(x) = \begin{cases} 4/26 & x = 1 \\ 4/26 & x = 2 \\ 2/26 & x = 3 \\ 1/26 & x = 4 \\ 1/26 & x = 5 \end{cases}$$



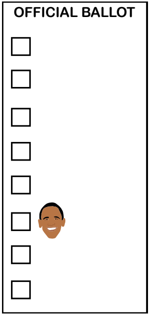
A,B,C,D,E,F,G,H,I,J,K,L,M,N,O,P,Q,R,S,T,U,V,W,X,Y,Z

Example: New Hampshire

Candidates:

- Joe Biden
- Hillary Clinton
- Chris Dodd
- John Edwards
- Mike Gravel
- Dennis Kucinich
- Barack Obama
- Bill Richardson

$$f(x) = \begin{cases} 4/26 & x = 1 \\ 4/26 & x = 2 \\ 2/26 & x = 3 \\ 1/26 & x = 4 \\ 1/26 & x = 5 \\ 1/26 & x = 6 \end{cases}$$



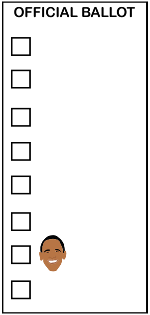
A,B,C,D,E,F,G,H,I,J,K,L,M,N,O,P,Q,R,S,T,U,V,W,X,Y,Z

Example: New Hampshire

Candidates:

- Joe Biden
- Hillary Clinton
- Chris Dodd
- John Edwards
- Mike Gravel
- Dennis Kucinich
- Barack Obama
- Bill Richardson

$$f(x) = \begin{cases} 4/26 & x = 1 \\ 4/26 & x = 2 \\ 2/26 & x = 3 \\ 1/26 & x = 4 \\ 1/26 & x = 5 \\ 1/26 & x = 6 \\ 10/26 & x = 7 \end{cases}$$



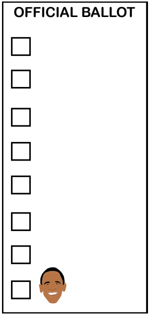
A,B,C,D,E,F,G,H,I,J,K,L,M,N,O,P,Q,R,S,T,U,V,W,X,Y,Z

Example: New Hampshire

Candidates:

- Joe Biden
- Hillary Clinton
- Chris Dodd
- John Edwards
- Mike Gravel
- Dennis Kucinich
- Barack Obama
- Bill Richardson

$$f(x) = \begin{cases} 4/26 & x = 1 \\ 4/26 & x = 2 \\ 2/26 & x = 3 \\ 1/26 & x = 4 \\ 1/26 & x = 5 \\ 1/26 & x = 6 \\ 10/26 & x = 7 \\ 3/26 & x = 8 \end{cases}$$



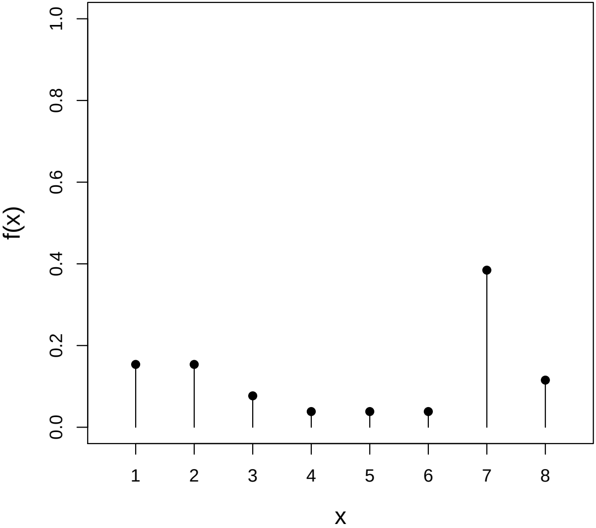
A,B,C,D,E,F,G,H,I,J,K,L,M,N,O,P,Q,R,S,T,U,V,W,X,Y,Z

Discrete Probability Mass Functions

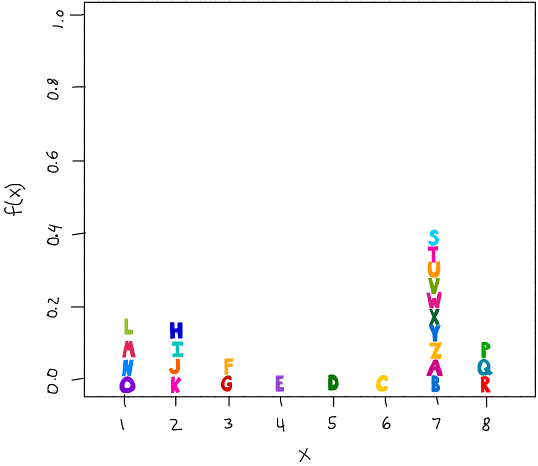
Discrete Probability Mass Functions

A probability mass function $f(x)$ of a random variable X is a non-negative function that gives the probability that $X = x$ and $\sum_x f(x) = 1$.

NH Obama Ballot Position PMF Plot



NH Obama Ballot Position PMF Plot

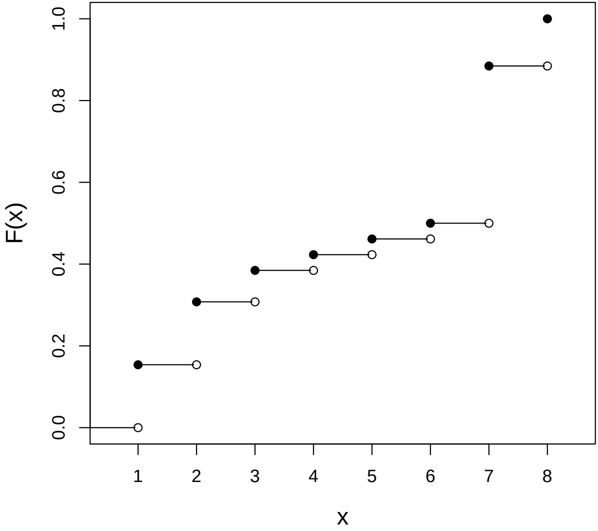


Discrete Cumulative Distribution Function

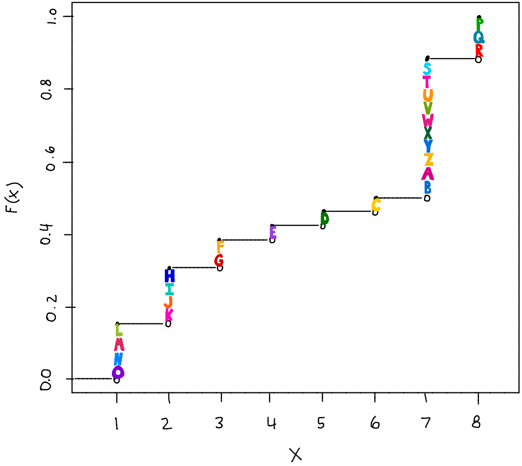
Discrete Cumulative Distribution Function

A cumulative distribution function $F(x)$ of a random variable X is a non-decreasing function that gives the probability that $X \leq x$.

NH Obama Ballot Position CDF Plot



NH Obama Ballot Position CDF Plot



Some Important Discrete Distributions

Some Important Discrete Distributions

- Let X be a binary variable with $P(X = 1) = p$ and, thus, $P(X = 0) = 1 - p$, where $p \in [0, 1]$. Then we say that X follows a Bernoulli distribution with the following pmf:

Some Important Discrete Distributions

- Let X be a binary variable with $P(X = 1) = p$ and, thus, $P(X = 0) = 1 - p$, where $p \in [0, 1]$. Then we say that X follows a Bernoulli distribution with the following pmf:

$$f_X(x) = p^x(1-p)^{1-x} \qquad \text{for } x \in \{0, 1\}.$$

Some Important Discrete Distributions

- Let X be a binary variable with $P(X = 1) = p$ and, thus, $P(X = 0) = 1 - p$, where $p \in [0, 1]$. Then we say that X follows a Bernoulli distribution with the following pmf:

$$f_X(x) = p^x(1-p)^{1-x} \qquad \text{for } x \in \{0, 1\}.$$

- Probably the most famous distribution for a discrete r.v. is the discrete uniform distribution that puts equal probability on each value that X can take:

Some Important Discrete Distributions

- Let X be a binary variable with $P(X = 1) = p$ and, thus, $P(X = 0) = 1 - p$, where $p \in [0, 1]$. Then we say that X follows a Bernoulli distribution with the following pmf:

$$f_X(x) = p^x(1-p)^{1-x} \qquad \text{for } x \in \{0, 1\}.$$

- Probably the most famous distribution for a discrete r.v. is the discrete uniform distribution that puts equal probability on each value that X can take:

$$f_X(x) = \begin{cases} 1/k & \text{for } x = 1, \ldots, k \\ 0 & \text{otherwise} \end{cases}$$

Some Important Discrete Distributions

- Let X be a binary variable with $P(X = 1) = p$ and, thus, $P(X = 0) = 1 - p$, where $p \in [0, 1]$. Then we say that X follows a Bernoulli distribution with the following pmf:

$$f_X(x) = p^x(1-p)^{1-x} \qquad \text{for } x \in \{0, 1\}.$$

- Probably the most famous distribution for a discrete r.v. is the discrete uniform distribution that puts equal probability on each value that X can take:

$$f_X(x) = \begin{cases} 1/k & \text{for } x = 1, \ldots, k \\ 0 & \text{otherwise} \end{cases}$$

- We can summarize these distributions with one number (e.g. the probability of variables being 1)

Empirical Distributions

Empirical Distributions

An empirical mass function $\widehat{f}(x)$ of a variable X is a non-negative function that gives the frequency of the value x from data on X.

Empirical Distributions

An empirical mass function $\widehat{f}(x)$ of a variable X is a non-negative function that gives the frequency of the value x from data on X.

An empirical cumulative distribution function $\widehat{F}(x)$ of a variable X is a non-decreasing function that gives the frequency of values of X less than x.

Example: Assessing Racial Prejudice

- We often want to ask sensitive questions which a survey respondent is unlikely to honestly answer
- A list experiment asks respondents how many items on a list they agree with
  - for example, what proportion of people would be upset by a black family moving in next door to them (Kuklinski et al 1997).
  - randomly split survey into two halves
  - first half ask how many of the following items upset you:
    1. the federal government increasing the tax on gasoline
    2. professional athletes getting million-dollar salaries
    3. large corporations polluting the environment.
  - second half, add a fourth item
    4. a black family moving in next door
  - use the answers to infer the proportion upset by the fourth item.
- To do this we need to understand random variables

Racial Prejudice Example (Kuklinski et al, 1997)

Racial Prejudice Example (Kuklinski et al, 1997)

$X = \#$ of angering items on the baseline list for Southerners:

x	0	1	2	3
$f(x)$	?	?	?	?
$\widehat{f}(x)$	0.02	0.27	0.43	0.28
$\widehat{F}(x)$	0.02	0.29	0.72	1.00

Racial Prejudice Example (Kuklinski et al, 1997)

$X = \#$ of angering items on the baseline list for Southerners:

x	0	1	2	3
$f(x)$	?	?	?	?
$\widehat{f}(x)$	0.02	0.27	0.43	0.28
$\widehat{F}(x)$	0.02	0.29	0.72	1.00

$Y = \#$ of angering items on the treatment list for Southerners:

y	0	1	2	3	4
$f(y)$	?	?	?	?	?
$\widehat{f}(y)$	0.02	0.20	0.40	0.28	0.10
$\widehat{F}(y)$	0.02	0.22	0.62	0.90	1.00

Continuous Distributions

Continuous Distributions

- Continuous random variables take on an uncountably infinite number of values.

Continuous Distributions

- Continuous random variables take on an uncountably infinite number of values.
- This is often a useful approximation when a variable takes on many values.

Continuous Distributions

- Continuous random variables take on an uncountably infinite number of values.
- This is often a useful approximation when a variable takes on many values.
- A probability density function (pdf) and a cumulative distribution function (cdf) are two common ways to define the distribution for a continuous RV.

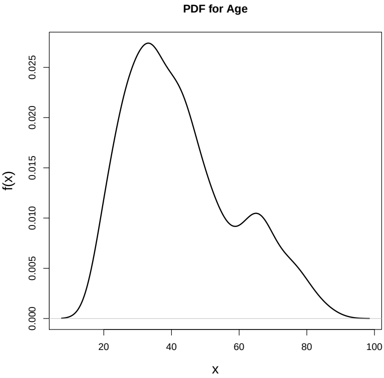
Example: Age in the Racial Prejudice Example

Example: Age in the Racial Prejudice Example

Let X be the age of a randomly selected individual from the Kuklinski et al. (1997) data set.

Example: Age in the Racial Prejudice Example

Let X be the age of a randomly selected individual from the Kuklinski et al. (1997) data set.
The probability distribution for this variable is well approximated by a probability density function.

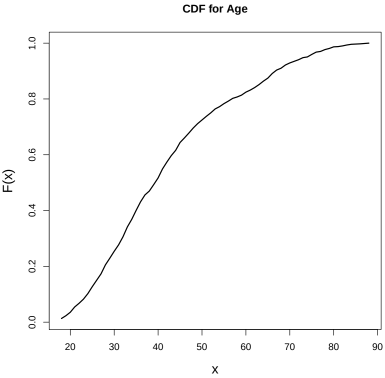


Continuous Cumulative Distribution Functions

Continuous Cumulative Distribution Functions

A cumulative distribution function $F(x)$ of a random variable X is a non-decreasing function that gives the probability that $X \leq x$. For a continuous RV, the cdf is continuous.

$$F(x) = \int_{-\infty}^{x} f(z)dz$$

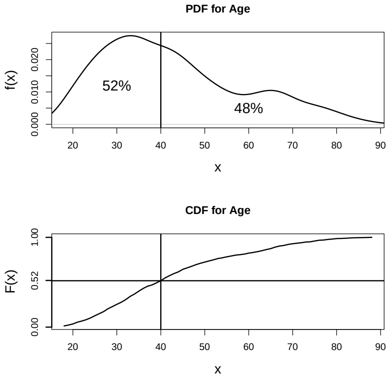


From PDFs to CDFs

From PDFs to CDFs

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(z)dz$$

$$.52 = P(X \leq 40) = \int_{-\infty}^{40} f(z)dz$$

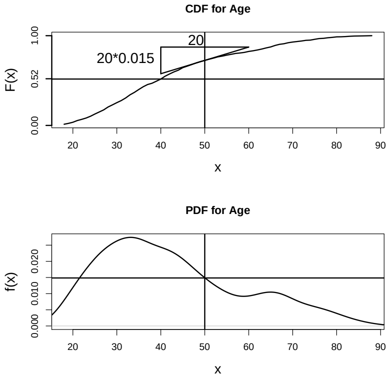


From CDFs to PDFs

From CDFs to PDFs

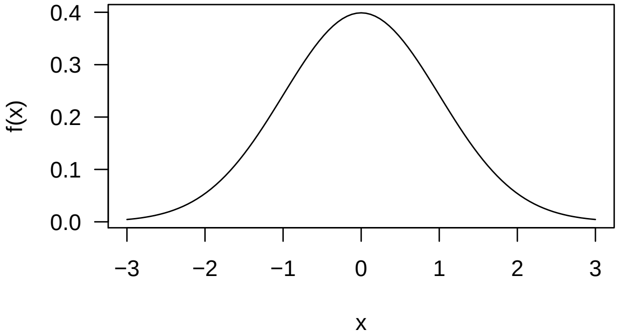
$$f(x) = \frac{dF(x)}{dx}$$

$$.015 = \frac{dF(50)}{dx}$$



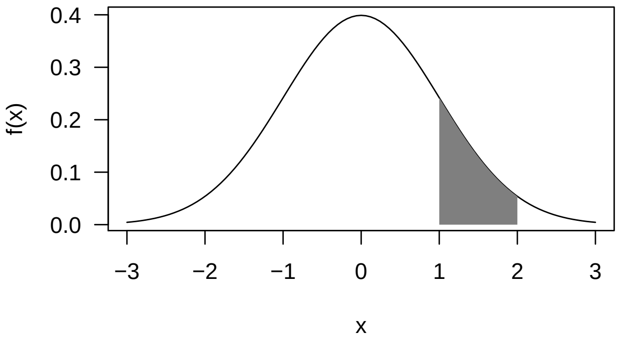
Subtleties of Continuous Densities

Remember- the height of the curve is not the probability of x occurring.



Subtleties of Continuous Densities

Remember- the height of the curve is not the probability of x occurring. To get the probability that X will fall in some region, you need the area under the curve.



1. Random Variables and Distributions
   - What is a Random Variable?
   - Discrete Distributions
   - Continuous Distributions
2. Characteristics of Distributions
   - Central Tendency
   - Measures of Dispersion
3. Conditional Distributions
4. Fun with Averages
5. Fun with Sensitive Questions
6. Appendix: Why the Mean?
7. Joint Distributions
   - Discrete Random Variable
   - Continuous Random Variable
8. Conditional Expectation
9. Properties
   - Independence
   - Covariance and Correlation
   - Conditional Independence
10. Famous Distributions
11. Fun With Spam

1 Random Variables and Distributions
- What is a Random Variable?
- Discrete Distributions
- Continuous Distributions

2 Characteristics of Distributions
- Central Tendency
- Measures of Dispersion

3 Conditional Distributions

4 Fun with Averages

5 Fun with Sensitive Questions

6 Appendix: Why the Mean?

7 Joint Distributions
- Discrete Random Variable
- Continuous Random Variable

8 Conditional Expectation

9 Properties
- Independence
- Covariance and Correlation
- Conditional Independence

10 Famous Distributions

11 Fun With Spam

Expectation

Expectation

The expected value of a random variable X is denoted by $E[X]$ and is a measure of **central tendency** of X. Roughly speaking, an expected value is like a weighted average of all of the values weighted by probability of occurrence.

The expected value of a discrete random variable X is defined as

$$E[X] = \sum_{\text{all } x} x \cdot f_X(x).$$

Expectation

The expected value of a random variable X is denoted by $E[X]$ and is a measure of **central tendency** of X. Roughly speaking, an expected value is like a weighted average of all of the values weighted by probability of occurrence.

The expected value of a discrete random variable X is defined as

$$E[X] = \sum_{\text{all } x} x \cdot f_X(x).$$

The expected value of a continuous random variable X is defined as

$$E[X] = \int_{-\infty}^{\infty} x \cdot f_X(x)dx.$$

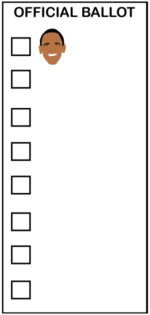
What did we expect for Obama's NH position?

Candidates:

- Joe Biden
- Hillary Clinton
- Chris Dodd
- John Edwards
- Mike Gravel
- Dennis Kucinich
- Barack Obama
- Bill Richardson

$4/26 \quad \times 1$

$+$



A,B,C,D,E,F,G,H,I,J,K,L,M,N,O,P,Q,R,S,T,U,V,W,X,Y,Z

What did we expect for Obama's NH position?

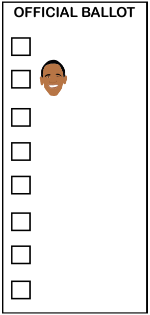
Candidates:

- Joe Biden
- Hillary Clinton
- Chris Dodd
- John Edwards
- Mike Gravel
- Dennis Kucinich
- Barack Obama
- Bill Richardson

$4/26 \quad \times 1$

$4/26 \quad \times 2$

$+$ ____________



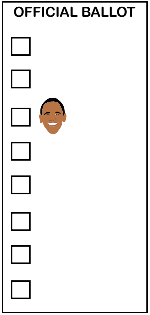
A,B,C,D,E,F,G,H,I,J,K,L,M,N,O,P,Q,R,S,T,U,V,W,X,Y,Z

What did we expect for Obama's NH position?

Candidates:

- Joe Biden
- Hillary Clinton
- Chris Dodd
- John Edwards
- Mike Gravel
- Dennis Kucinich
- Barack Obama
- Bill Richardson

$$
\begin{array}{ll}
4/26 & \times 1 \\
4/26 & \times 2 \\
2/26 & \times 3 \\
\\
+ & \\
\hline
\end{array}
$$



A,B,C,D,E,F,G,H,I,J,K,L,M,N,O,P,Q,R,S,T,U,V,W,X,Y,Z

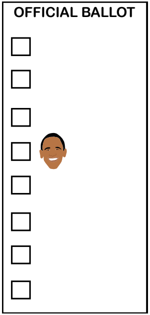
What did we expect for Obama's NH position?

Candidates:

- Joe Biden
- Hillary Clinton
- Chris Dodd
- John Edwards
- Mike Gravel
- Dennis Kucinich
- Barack Obama
- Bill Richardson

$4/26 \quad \times 1$
$4/26 \quad \times 2$
$2/26 \quad \times 3$
$1/26 \quad \times 4$

$+$
___



A,B,C,D,E,F,G,H,I,J,K,L,M,N,O,P,Q,R,S,T,U,V,W,X,Y,Z

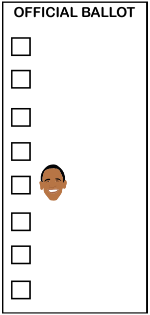
What did we expect for Obama's NH position?

Candidates:

- Joe Biden
- Hillary Clinton
- Chris Dodd
- John Edwards
- Mike Gravel
- Dennis Kucinich
- Barack Obama
- Bill Richardson

$4/26 \quad \times 1$
$4/26 \quad \times 2$
$2/26 \quad \times 3$
$1/26 \quad \times 4$
$1/26 \quad \times 5$

$+$ ______________



A,B,C,D,E,F,G,H,I,J,K,L,M,N,O,P,Q,R,S,T,U,V,W,X,Y,Z

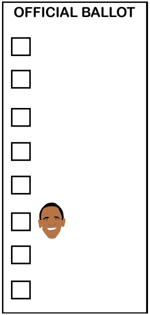
What did we expect for Obama's NH position?

Candidates:

- Joe Biden
- Hillary Clinton
- Chris Dodd
- John Edwards
- Mike Gravel
- Dennis Kucinich
- Barack Obama
- Bill Richardson

$4/26 \quad \times 1$
$4/26 \quad \times 2$
$2/26 \quad \times 3$
$1/26 \quad \times 4$
$1/26 \quad \times 5$
$1/26 \quad \times 6$

$+$ ____________



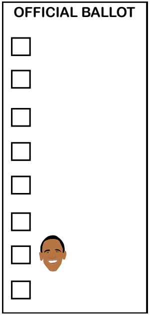
A,B,C,D,E,F,G,H,I,J,K,L,M,N,O,P,Q,R,S,T,U,V,W,X,Y,Z

What did we expect for Obama's NH position?

Candidates:

- Joe Biden
- Hillary Clinton
- Chris Dodd
- John Edwards
- Mike Gravel
- Dennis Kucinich
- Barack Obama
- Bill Richardson

$$\begin{array}{rr} 4/26 & \times 1 \\ 4/26 & \times 2 \\ 2/26 & \times 3 \\ 1/26 & \times 4 \\ 1/26 & \times 5 \\ 1/26 & \times 6 \\ 10/26 & \times 7 \\ + & \\ \hline \end{array}$$



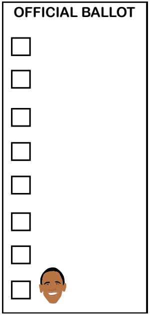
A,B,C,D,E,F,G,H,I,J,K,L,M,N,O,P,Q,R,S,T,U,V,W,X,Y,Z

What did we expect for Obama's NH position?

Candidates:

- Joe Biden
- Hillary Clinton
- Chris Dodd
- John Edwards
- Mike Gravel
- Dennis Kucinich
- Barack Obama
- Bill Richardson

$$
\begin{array}{rrr}
 & 4/26 & \times 1 \\
 & 4/26 & \times 2 \\
 & 2/26 & \times 3 \\
 & 1/26 & \times 4 \\
 & 1/26 & \times 5 \\
 & 1/26 & \times 6 \\
 & 10/26 & \times 7 \\
+ & 3/26 & \times 8 \\
\hline
\end{array}
$$



A,B,C,D,E,F,G,H,I,J,K,L,M,N,O,P,Q,R,S,T,U,V,W,X,Y,Z

What did we expect for Obama's NH position?

Candidates:

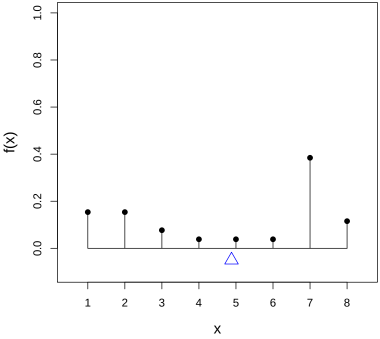
- Joe Biden
- Hillary Clinton
- Chris Dodd
- John Edwards
- Mike Gravel
- Dennis Kucinich
- Barack Obama
- Bill Richardson

$$
\begin{array}{rrr}
 & 4/26 & \times 1 \\
 & 4/26 & \times 2 \\
 & 2/26 & \times 3 \\
 & 1/26 & \times 4 \\
 & 1/26 & \times 5 \\
 & 1/26 & \times 6 \\
 & 10/26 & \times 7 \\
+ & 3/26 & \times 8 \\
\hline
 & & 4.88
\end{array}
$$

A,B,C,D,E,F,G,H,I,J,K,L,M,N,O,P,Q,R,S,T,U,V,W,X,Y,Z

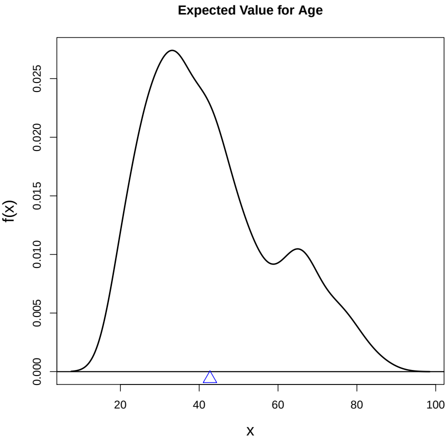
Interpreting Discrete Expected Value

The expected value for a discrete random variable is the balance point of the mass function.



Interpreting Continuous Expected Value

The expected value for a continuous random variable is the balance point of the density function.



Why the Expected Value (Balance Point)?

Why the Expected Value (Balance Point)?

- It is the probabilistic equivalent of the sample average (mean).

Why the Expected Value (Balance Point)?

- It is the probabilistic equivalent of the sample average (mean).
- It is a reasonable measure for the "center" of the data.

Why the Expected Value (Balance Point)?

- It is the probabilistic equivalent of the sample average (mean).
- It is a reasonable measure for the "center" of the data.
- We have some intuition about balance points.

Why the Expected Value (Balance Point)?

- It is the probabilistic equivalent of the sample average (mean).
- It is a reasonable measure for the "center" of the data.
- We have some intuition about balance points.
- It has some useful and convenient properties.

Appendix

Population Mean as an Expected Value

Population Mean as an Expected Value

Let $x_1, \ldots, x_N$ be our population. Then the population mean is the following

Population Mean as an Expected Value

Let $x_1, \ldots, x_N$ be our population. Then the population mean is the following

$$\bar{x} = \frac{1}{N} \sum_{i=1}^{N} x_i$$

Population Mean as an Expected Value

Let $x_1, \ldots, x_N$ be our population. Then the population mean is the following

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i$$

This can be re-written in the following form:

$$\bar{x} = \sum_{i=1}^{N}\left\{\frac{1}{N}x_i\right\}$$

Population Mean as an Expected Value

Let $x_1, \ldots, x_N$ be our population. Then the population mean is the following

$$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i$$

This can be re-written in the following form:

$$\bar{x} = \sum_{i=1}^{N}\left\{\frac{1}{N}x_i\right\}$$

Note how this resembles the definition of discrete expected value. If all values distinct (i.e. $x_i \neq x_j$ for all $i \neq j$).

Population Mean as an Expected Value

Let $x_1, \ldots, x_N$ be our population. Then the population mean is the following

$$\bar{x} = \frac{1}{N} \sum_{i=1}^{N} x_i$$

This can be re-written in the following form:

$$\bar{x} = \sum_{i=1}^{N} \left\{ \frac{1}{N} x_i \right\}$$

Note how this resembles the definition of discrete expected value. If all values distinct (i.e. $x_i \neq x_j$ for all $i \neq j$).

$$\bar{x} = \sum_{\text{all } x_i} x_i f(x_i), \text{ where } f(x_i) = \frac{1}{N}$$

Property 1 of Expected Value: Additivity

Property 1 of Expected Value: Additivity

Expectations of sums are sums of expectations.

Property 1 of Expected Value: Additivity

Expectations of sums are sums of expectations.

Suppose we have k random variables $X_1, \ldots, X_k$. If $E[X_i]$ exists for all $i = 1, \ldots, k$, then

Property 1 of Expected Value: Additivity

Expectations of sums are sums of expectations.

Suppose we have k random variables $X_1, \ldots, X_k$. If $E[X_i]$ exists for all $i = 1, \ldots, k$, then

$$E\left[\sum_{i=1}^{k} X_i\right] = E[X_1] + \cdots + E[X_k]$$

Property 1 of Expected Value: Additivity

Expectations of sums are sums of expectations.

Suppose we have k random variables $X_1, \ldots, X_k$. If $E[X_i]$ exists for all $i = 1, \ldots, k$, then

$$E\left[\sum_{i=1}^{k} X_i\right] = E[X_1] + \cdots + E[X_k]$$

Property 2 of Expected Value: Homogeneity

Property 2 of Expected Value: Homogeneity

- The expected value of a constant is the constant.

Property 2 of Expected Value: Homogeneity

- The expected value of a constant is the constant.
- The expectation of a constant times a RV is the constant times the expectation of the RV.

Property 2 of Expected Value: Homogeneity

- The expected value of a constant is the constant.
- The expectation of a constant times a RV is the constant times the expectation of the RV.

Property 2 of Expected Value: Homogeneity

- The expected value of a constant is the constant.
- The expectation of a constant times a RV is the constant times the expectation of the RV.

Suppose a and b are constants and X is a random variable. Then

Property 2 of Expected Value: Homogeneity

- The expected value of a constant is the constant.
- The expectation of a constant times a RV is the constant times the expectation of the RV.

Suppose a and b are constants and X is a random variable. Then

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

Property 2 of Expected Value: Homogeneity

- The expected value of a constant is the constant.
- The expectation of a constant times a RV is the constant times the expectation of the RV.

Suppose a and b are constants and X is a random variable. Then

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

Together properties 1 and 2 are linearity (and this is sometimes presented as Linearity of Expectations).

Property 3 of Expected Value: LOTUS

Property 3 of Expected Value: LOTUS

Law of the Unconscious Statistician: If $g(X)$ is a function of a discrete random variable, then

$$E[g(X)] = \sum_x g(x) f_X(x),$$

Property 3 of Expected Value: LOTUS

Law of the Unconscious Statistician: If $g(X)$ is a function of a discrete random variable, then

$$E[g(X)] = \sum_x g(x) f_X(x),$$

essentially the expected value of the transformation of the random variable is just the weighted average of the transformed outcomes.

Property 3 of Expected Value: LOTUS

Law of the Unconscious Statistician: If $g(X)$ is a function of a discrete random variable, then

$$E[g(X)] = \sum_x g(x) f_X(x),$$

essentially the expected value of the transformation of the random variable is just the weighted average of the transformed outcomes.

We will come back to this later. But it means that we can can calculate the expected value of $g(X)$ without explicitly knowing the distribution of $g(X)$!

Summary of Expected Value Properties

Summary of Expected Value Properties

The three properties:

1) Additivity: expectation of sums are sums of expectations

$$E[X + Y] = E[X] + E[Y]$$

Summary of Expected Value Properties

The three properties:

1) Additivity: expectation of sums are sums of expectations

$$E[X + Y] = E[X] + E[Y]$$

2) Homogeneity: expected value of a constant is the constant

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

Summary of Expected Value Properties

The three properties:

1) Additivity: expectation of sums are sums of expectations

$$E[X + Y] = E[X] + E[Y]$$

2) Homogeneity: expected value of a constant is the constant

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

3) LOTUS: Law of the Unconscious Statistician

$$E[g(X)] = \sum_x g(x) f_X(x)$$

Summary of Expected Value Properties

The three properties:

1) Additivity: expectation of sums are sums of expectations

$$E[X + Y] = E[X] + E[Y]$$

2) Homogeneity: expected value of a constant is the constant

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

3) LOTUS: Law of the Unconscious Statistician

$$E[g(X)] = \sum_x g(x) f_X(x)$$

However,

- $E[g(X)] \neq g(E[X])$ unless $g(\cdot)$ is a linear function

Summary of Expected Value Properties

The three properties:

1) Additivity: expectation of sums are sums of expectations

$$E[X + Y] = E[X] + E[Y]$$

2) Homogeneity: expected value of a constant is the constant

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

3) LOTUS: Law of the Unconscious Statistician

$$E[g(X)] = \sum_x g(x)f_X(x)$$

However,

- $E[g(X)] \neq g(E[X])$ unless $g(\cdot)$ is a linear function
- $E[XY] \neq E[X]E[Y]$ unless X and Y are independent

Racial Prejudice Example

$X = \#$ of angering items on the baseline list for Southerners:

x	0	1	2	3	Sum
$\widehat{f}(x)$	0.02	0.27	0.43	0.28	1.00
$x \cdot \widehat{f}(x)$	0.00	0.27	0.86	0.84	1.97

$Y = \#$ of angering items on the treatment list for Southerners:

y	0	1	2	3	4	Sum
$\widehat{f}(y)$	0.03	0.20	0.40	0.28	0.10	1.00
$y \cdot \widehat{f}(y)$	0.00	0.20	0.80	0.84	0.40	2.24

Identifying the Percent Angry

Assume that $Y = X + A$, where for a randomly sampled respondent,

- $Y =$ the number of total angering items
- $X =$ the number of angering items on baseline list
- $A = 1$ if angered by a black family moving in next door
- $A = 0$ if not angered by a black family moving in next door.

Identifying the Percent Angry

Assume that $Y = X + A$, where for a randomly sampled respondent,

- $Y =$ the number of total angering items
- $X =$ the number of angering items on baseline list
- $A = 1$ if angered by a black family moving in next door
- $A = 0$ if not angered by a black family moving in next door.

Exercises for Later:

- Then we know that $E[Y] - E[X] = E[A]$, but can you prove it?

Identifying the Percent Angry

Assume that $Y = X + A$, where for a randomly sampled respondent,

- $Y =$ the number of total angering items
- $X =$ the number of angering items on baseline list
- $A = 1$ if angered by a black family moving in next door
- $A = 0$ if not angered by a black family moving in next door.

Exercises for Later:

- Then we know that $E[Y] - E[X] = E[A]$, but can you prove it?
- Noting that A is a Bernoulli RV, how can we interpret $E[A]$?

Identifying the Percent Angry

Assume that $Y = X + A$, where for a randomly sampled respondent,

- $Y =$ the number of total angering items
- $X =$ the number of angering items on baseline list
- $A = 1$ if angered by a black family moving in next door
- $A = 0$ if not angered by a black family moving in next door.

Exercises for Later:

- Then we know that $E[Y] - E[X] = E[A]$, but can you prove it?
- Noting that A is a Bernoulli RV, how can we interpret $E[A]$?
- What properties and assumptions were necessary?

Variance

The expected value of a function $g()$ of the random variable X, written $g(X)$, is denoted by $E[g(X)]$ and is a measure of central tendency of $g(X)$.

Variance

The expected value of a function $g()$ of the random variable X, written $g(X)$, is denoted by $E[g(X)]$ and is a measure of central tendency of $g(X)$.

The variance is a special case of this, and the variance of a random variable X (a measure of its dispersion) is given by

$$V[X] = E[(X - E[X])^2]$$

It is the expectation of the squared distances from the mean.

For a discrete random variable X

$$V[X] = \sum_{\text{all } x} (x - E[X])^2 f_X(x)$$

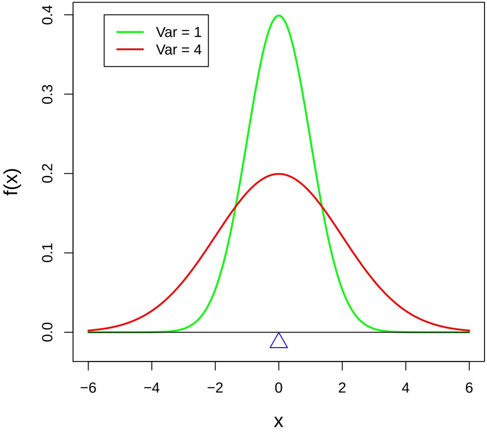
For a discrete random variable X

$$V[X] = \sum_{\text{all } x} (x - E[X])^2 f_X(x)$$

For a continuous random variable X

$$V[X] = \int_{-\infty}^{\infty} (x - E[X])^2 f_X(x) dx$$

Variance Measures the Spread of a Distribution



Why the Variance?

Why the Variance?

- It is a reasonable measure for the “spread” of a distribution.

Why the Variance?

- It is a reasonable measure for the "spread" of a distribution.
- The Normal distribution (bell shaped with thin tails) is completely determined by its expected value (location) and variance (spread).

Why the Variance?

- It is a reasonable measure for the “spread” of a distribution.
- The Normal distribution (bell shaped with thin tails) is completely determined by its expected value (location) and variance (spread).
- The square root of the variance is the standard deviation.

Why the Variance?

- It is a reasonable measure for the “spread” of a distribution.
- The Normal distribution (bell shaped with thin tails) is completely determined by its expected value (location) and variance (spread).
- The square root of the variance is the standard deviation.
- The variance and standard deviation have some useful properties.

Property 1 of Variance: Behavior with Constants

Suppose a and b are constants and X is a random variable. Then

Property 1 of Variance: Behavior with Constants

Suppose a and b are constants and X is a random variable. Then

- The variance of a constant is zero.

Property 1 of Variance: Behavior with Constants

Suppose a and b are constants and X is a random variable. Then

- The variance of a constant is zero.
- The variance of a constant times a RV is the constant squared times the variance of the RV.

Property 1 of Variance: Behavior with Constants

Suppose a and b are constants and X is a random variable. Then

- The variance of a constant is zero.
- The variance of a constant times a RV is the constant squared times the variance of the RV.

$$V[b] = 0$$
$$V[aX] = a^2 V[X]$$
$$V[aX + b] = a^2 V[X] + 0$$

Property 2 of Variance: Additivity for Independent Random Variables

Property 2 of Variance: Additivity for Independent Random Variables

Variances of sums of independent RVs are sums of variances.

Property 2 of Variance: Additivity for Independent Random Variables

Variances of sums of independent RVs are sums of variances.

Suppose we have k independent random variables $X_1, \ldots, X_k$. If $V[X_i]$ exists for all $i = 1, \ldots, k$, then

$$V\left[\sum_{i=1}^{k} X_i\right] = V[X_1] + \cdots + V[X_k]$$

NB: Technically independence is sufficient but not necessary.

What was the variance of Obama's NH position?

Candidates:

- Joe Biden
- Hillary Clinton
- Chris Dodd
- John Edwards
- Mike Gravel
- Dennis Kucinich
- Barack Obama
- Bill Richardson

$$4/26 \quad \times (1 - 4.88)^2$$

$$+$$

A,B,C,D,E,F,G,H,I,J,K,L,M,N,O,P,Q,R,S,T,U,V,W,X,Y,Z

What was the variance of Obama's NH position?

Candidates:

- Joe Biden
- Hillary Clinton
- Chris Dodd
- John Edwards
- Mike Gravel
- Dennis Kucinich
- Barack Obama
- Bill Richardson

$$4/26 \quad \times (1 - 4.88)^2$$
$$4/26 \quad \times (2 - 4.88)^2$$

$+$

A,B,C,D,E,F,G,H,I,J,K,L,M,N,O,P,Q,R,S,T,U,V,W,X,Y,Z

What was the variance of Obama's NH position?

Candidates:

- Joe Biden
- Hillary Clinton
- Chris Dodd
- John Edwards
- Mike Gravel
- Dennis Kucinich
- Barack Obama
- Bill Richardson

$$
\begin{array}{ll}
4/26 & \times (1 - 4.88)^2 \\
4/26 & \times (2 - 4.88)^2 \\
2/26 & \times (3 - 4.88)^2 \\
 & \\
+ & \\
\hline
\end{array}
$$

A,B,C,D,E,F,G,H,I,J,K,L,M,N,O,P,Q,R,S,T,U,V,W,X,Y,Z

What was the variance of Obama's NH position?

Candidates:

- Joe Biden
- Hillary Clinton
- Chris Dodd
- John Edwards
- Mike Gravel
- Dennis Kucinich
- Barack Obama
- Bill Richardson

$$
\begin{aligned}
4/26 &\quad \times (1-4.88)^2 \\
4/26 &\quad \times (2-4.88)^2 \\
2/26 &\quad \times (3-4.88)^2 \\
1/26 &\quad \times (4-4.88)^2 \\
+ &
\end{aligned}
$$

A,B,C,D,E,F,G,H,I,J,K,L,M,N,O,P,Q,R,S,T,U,V,W,X,Y,Z

What was the variance of Obama's NH position?

Candidates:

- Joe Biden
- Hillary Clinton
- Chris Dodd
- John Edwards
- Mike Gravel
- Dennis Kucinich
- Barack Obama
- Bill Richardson

$$\begin{array}{ll} 4/26 & \times (1-4.88)^2 \\ 4/26 & \times (2-4.88)^2 \\ 2/26 & \times (3-4.88)^2 \\ 1/26 & \times (4-4.88)^2 \\ 1/26 & \times (5-4.88)^2 \\ & \\ + & \\ \hline \end{array}$$

A,B,C,D,E,F,G,H,I,J,K,L,M,N,O,P,Q,R,S,T,U,V,W,X,Y,Z

What was the variance of Obama's NH position?

Candidates:

- Joe Biden
- Hillary Clinton
- Chris Dodd
- John Edwards
- Mike Gravel
- Dennis Kucinich
- Barack Obama
- Bill Richardson

$$
\begin{array}{ll}
4/26 & \times (1 - 4.88)^2 \\
4/26 & \times (2 - 4.88)^2 \\
2/26 & \times (3 - 4.88)^2 \\
1/26 & \times (4 - 4.88)^2 \\
1/26 & \times (5 - 4.88)^2 \\
1/26 & \times (6 - 4.88)^2 \\
+ & \\
\hline
\end{array}
$$

A,B,C,D,E,F,G,H,I,J,K,L,M,N,O,P,Q,R,S,T,U,V,W,X,Y,Z

What was the variance of Obama's NH position?

Candidates:

- Joe Biden
- Hillary Clinton
- Chris Dodd
- John Edwards
- Mike Gravel
- Dennis Kucinich
- Barack Obama
- Bill Richardson

$$
\begin{array}{rl}
4/26 & \times (1 - 4.88)^2 \\
4/26 & \times (2 - 4.88)^2 \\
2/26 & \times (3 - 4.88)^2 \\
1/26 & \times (4 - 4.88)^2 \\
1/26 & \times (5 - 4.88)^2 \\
1/26 & \times (6 - 4.88)^2 \\
10/26 & \times (7 - 4.88)^2 \\
+ & \\
\hline
\end{array}
$$

A,B,C,D,E,F,G,H,I,J,K,L,M,N,O,P,Q,R,S,T,U,V,W,X,Y,Z

What was the variance of Obama's NH position?

Candidates:

- Joe Biden
- Hillary Clinton
- Chris Dodd
- John Edwards
- Mike Gravel
- Dennis Kucinich
- Barack Obama
- Bill Richardson

$$\begin{array}{rrl} & 4/26 & \times (1-4.88)^2 \\ & 4/26 & \times (2-4.88)^2 \\ & 2/26 & \times (3-4.88)^2 \\ & 1/26 & \times (4-4.88)^2 \\ & 1/26 & \times (5-4.88)^2 \\ & 1/26 & \times (6-4.88)^2 \\ & 10/26 & \times (7-4.88)^2 \\ + & 3/26 & \times (8-4.88)^2 \\ \hline \end{array}$$

A,B,C,D,E,F,G,H,I,J,K,L,M,N,O,P,Q,R,S,T,U,V,W,X,Y,Z

What was the variance of Obama's NH position?

Candidates:

- Joe Biden
- Hillary Clinton
- Chris Dodd
- John Edwards
- Mike Gravel
- Dennis Kucinich
- Barack Obama
- Bill Richardson

$$
\begin{array}{rrl}
 & 4/26 & \times (1 - 4.88)^2 \\
 & 4/26 & \times (2 - 4.88)^2 \\
 & 2/26 & \times (3 - 4.88)^2 \\
 & 1/26 & \times (4 - 4.88)^2 \\
 & 1/26 & \times (5 - 4.88)^2 \\
 & 1/26 & \times (6 - 4.88)^2 \\
 & 10/26 & \times (7 - 4.88)^2 \\
+ & 3/26 & \times (8 - 4.88)^2 \\
\hline
 & & 2.93
\end{array}
$$

A,B,C,D,E,F,G,H,I,J,K,L,M,N,O,P,Q,R,S,T,U,V,W,X,Y,Z

What was the variance of Obama's NH position?

Candidates:

- Joe Biden
- Hillary Clinton
- Chris Dodd
- John Edwards
- Mike Gravel
- Dennis Kucinich
- Barack Obama
- Bill Richardson

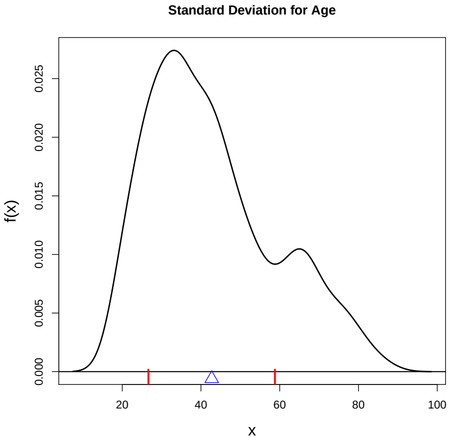
$$
\begin{array}{rr}
 & 4/26 \times (1-4.88)^2 \\
 & 4/26 \times (2-4.88)^2 \\
 & 2/26 \times (3-4.88)^2 \\
 & 1/26 \times (4-4.88)^2 \\
 & 1/26 \times (5-4.88)^2 \\
 & 1/26 \times (6-4.88)^2 \\
 & 10/26 \times (7-4.88)^2 \\
+ & 3/26 \times (8-4.88)^2 \\
\hline
 & 2.93
\end{array}
$$

A,B,C,D,E,F,G,H,I,J,K,L,M,N,O,P,Q,R,S,T,U,V,W,X,Y,Z

Does variance matter for fairness?

Interpreting Continuous Standard Deviation

The standard deviation for a continuous random variable is a measure of the spread of the pdf.



Do we lose anything when we use the list experiment?

$Y = \#$ of angering items on the treatment list for Southerners:

y	0	1	2	3	4	Sum
$\widehat{f}(y)$	0.03	0.20	0.40	0.28	0.10	1.00
$(y - 2.24)^2 \cdot \widehat{f}(y)$	0.15	0.31	0.02	0.16	0.31	0.95

More on this next week when we talk about estimator properties!

1. Random Variables and Distributions
   - What is a Random Variable?
   - Discrete Distributions
   - Continuous Distributions
2. Characteristics of Distributions
   - Central Tendency
   - Measures of Dispersion
3. Conditional Distributions
4. Fun with Averages
5. Fun with Sensitive Questions
6. Appendix: Why the Mean?
7. Joint Distributions
   - Discrete Random Variable
   - Continuous Random Variable
8. Conditional Expectation
9. Properties
   - Independence
   - Covariance and Correlation
   - Conditional Independence
10. Famous Distributions
11. Fun With Spam

1 Random Variables and Distributions
- What is a Random Variable?
- Discrete Distributions
- Continuous Distributions

2 Characteristics of Distributions
- Central Tendency
- Measures of Dispersion

3 Conditional Distributions

4 Fun with Averages

5 Fun with Sensitive Questions

6 Appendix: Why the Mean?

7 Joint Distributions
- Discrete Random Variable
- Continuous Random Variable

8 Conditional Expectation

9 Properties
- Independence
- Covariance and Correlation
- Conditional Independence

10 Famous Distributions

11 Fun With Spam

Joint and Conditional Distributions

- We can describe more than one random variable with joint and conditional distributions.

Joint and Conditional Distributions

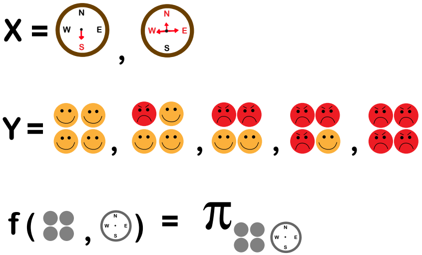
- We can describe more than one random variable with joint and conditional distributions.
- For example, suppose we define $X = 0$ (Non-southern), 1 (Southern) and $Y =$ "number of angering items" for a randomly selected respondent receiving the treatment list.

Joint and Conditional Distributions

- We can describe more than one random variable with joint and conditional distributions.
- For example, suppose we define $X = 0$ (Non-southern), 1 (Southern) and $Y =$ "number of angering items" for a randomly selected respondent receiving the treatment list.
- Furthermore, we define the probability that this respondent will have the values $X = x$ and $Y = y$ to be $f(y, x) = \pi_{yx}$

Joint and Conditional Distributions

- We can describe more than one random variable with joint and conditional distributions.
- For example, suppose we define $X = 0$ (Non-southern), 1 (Southern) and $Y =$ "number of angering items" for a randomly selected respondent receiving the treatment list.
- Furthermore, we define the probability that this respondent will have the values $X = x$ and $Y = y$ to be $f(y, x) = \pi_{yx}$



Example Conditional Distribution: Binary X, Discrete Y

Example Conditional Distribution: Binary X, Discrete Y

Although we cannot observe the responses for the entire population, we can imagine what they might look like as a joint distribution.

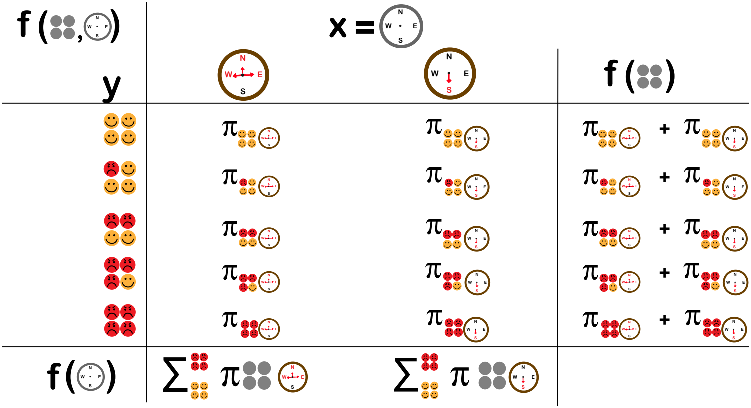
Example Conditional Distribution: Binary X, Discrete Y

Although we cannot observe the responses for the entire population, we can imagine what they might look like as a joint distribution.

$f(y,x)$	x		
y	0	1	$f(y)$
0	π_{00}	π_{01}	$\pi_{00} + \pi_{01}$
1	π_{10}	π_{11}	$\pi_{00} + \pi_{01}$
2	π_{20}	π_{21}	$\pi_{00} + \pi_{01}$
3	π_{30}	π_{31}	$\pi_{00} + \pi_{01}$
4	π_{40}	π_{41}	$\pi_{00} + \pi_{01}$
$f(x)$	$\sum_{y=0}^{4} \pi_{y0}$	$\sum_{y=0}^{4} \pi_{y1}$	

Example Conditional Distribution: Binary X, Discrete Y

Although we cannot observe the responses for the entire population, we can imagine what they might look like as a joint distribution.

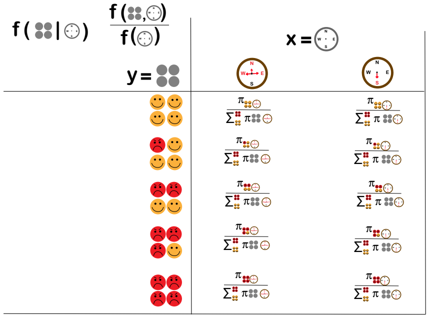


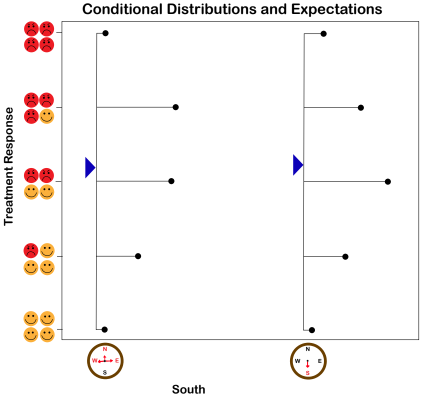
Discrete Conditional Distribution

Given the joint distribution, we can imagine what the conditional distribution and the conditional expectations would look like.

Discrete Conditional Distribution

Given the joint distribution, we can imagine what the conditional distribution and the conditional expectations would look like.





(More on conditional expectations on Wednesday)

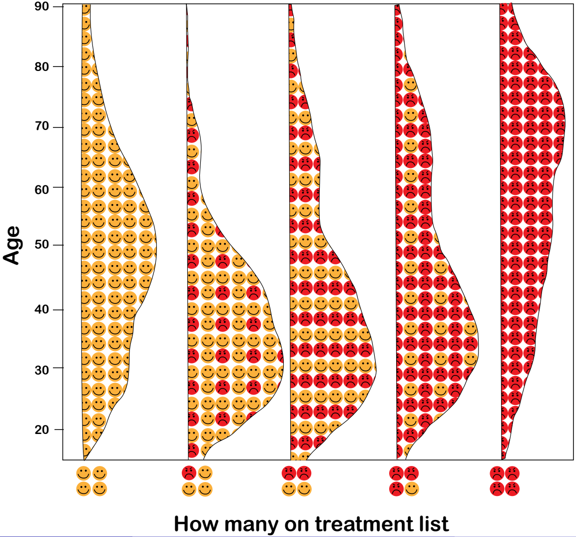
Example: Conditional Distribution with "Continuous" Y

Example: Conditional Distribution with "Continuous" Y

Suppose we define $X =$ "number of angering items" and $Y =$ "age" for a randomly selected respondent receiving the treatment list.

Example: Conditional Distribution with "Continuous" Y

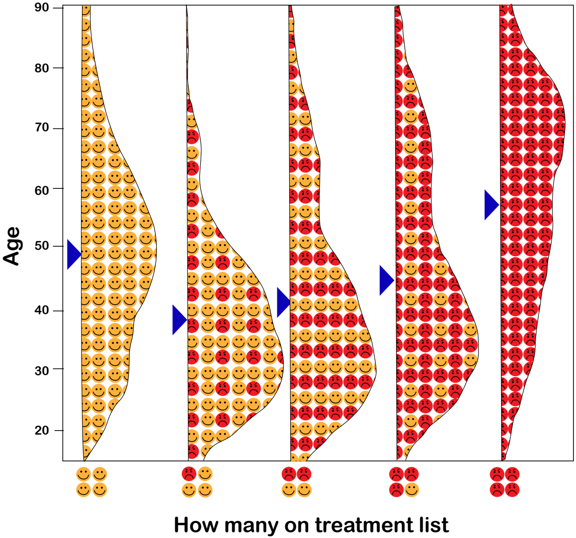
Suppose we define $X =$ "number of angering items" and $Y =$ "age" for a randomly selected respondent receiving the treatment list.



Conditional Expectation Function (CEF)

Conditional Expectation Function (CEF)

The conditional expectations form a CEF: $E[Y|X = x] = h(x)$



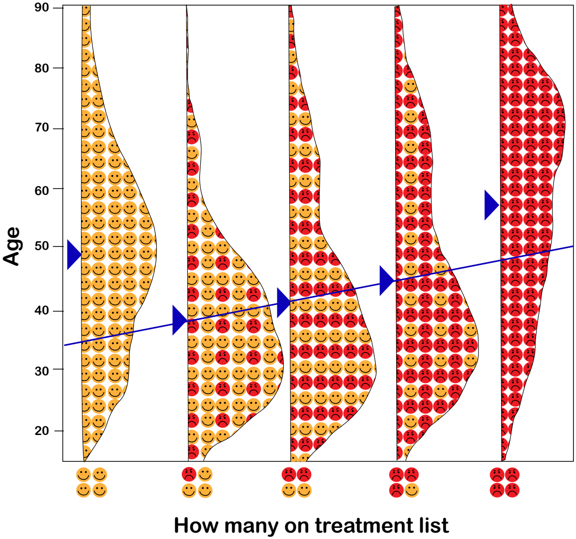
Linear CEF Assumption

Linear CEF Assumption

Often we will assume that the CEF is linear: $E[Y|X = x] = \beta_0 + \beta_1 x$

Linear CEF Assumption

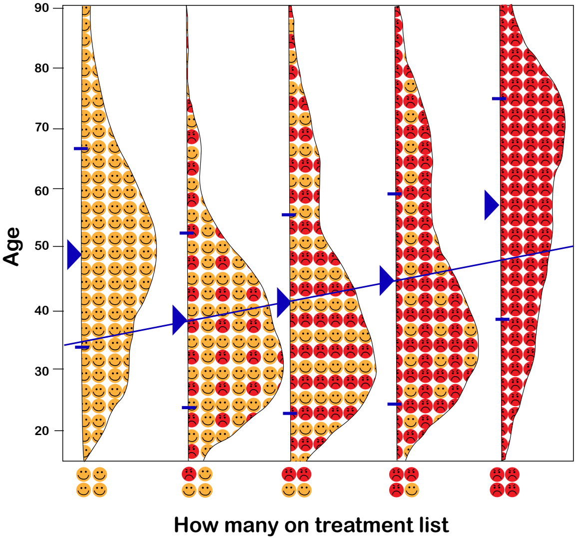
Often we will assume that the CEF is linear: $E[Y|X = x] = \beta_0 + \beta_1 x$



Conditional Variance and Standard Deviation

Conditional Variance and Standard Deviation

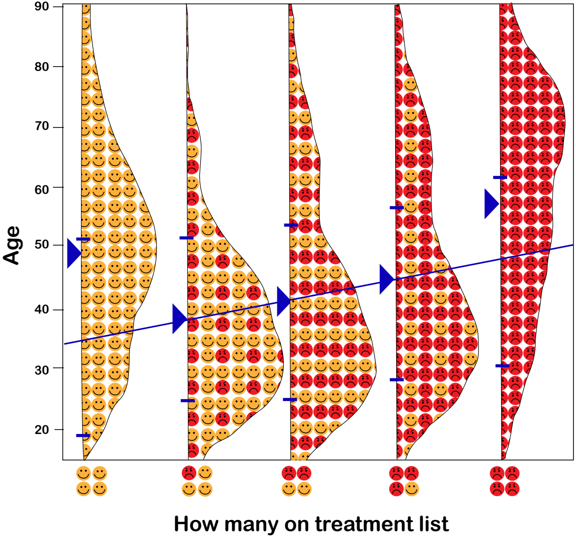
Similarly, we can assess the conditional standard deviation



Linear CEF and Constant Variance Assumptions

Linear CEF and Constant Variance Assumptions

Often, we assume that variance is the same for all values of x.



Preview: Interpreting the CEF

Preview: Interpreting the CEF

Because the CEF is defined merely in terms of the larger population and not in terms of a causal effect (e.g., the causal effect of "number of angering items" on Age), we will utilize a descriptive interpretation of β_0 and β_1.

Preview: Interpreting the CEF

Because the CEF is defined merely in terms of the larger population and not in terms of a causal effect (e.g., the causal effect of "number of angering items" on Age), we will utilize a descriptive interpretation of β_0 and β_1.

- For this example, β_0 is the expected age for an individual that is angered by zero items

Preview: Interpreting the CEF

Because the CEF is defined merely in terms of the larger population and not in terms of a causal effect (e.g., the causal effect of "number of angering items" on Age), we will utilize a descriptive interpretation of β_0 and β_1.

- For this example, β_0 is the expected age for an individual that is angered by zero items
- β_1 is the expected difference in age between two individuals that have a one unit difference in the number of angering items.

Summary

Summary

- Random variables and probability distributions provide useful infrastructure for everything we will do this year.

Summary

- Random variables and probability distributions provide useful infrastructure for everything we will do this year.
- Expected value and variance are two useful characteristics of the probability distributions associated with random variables.

Summary

- Random variables and probability distributions provide useful infrastructure for everything we will do this year.
- Expected value and variance are two useful characteristics of the probability distributions associated with random variables.
- These concepts can be extended by conditioning on other variables.

Summary

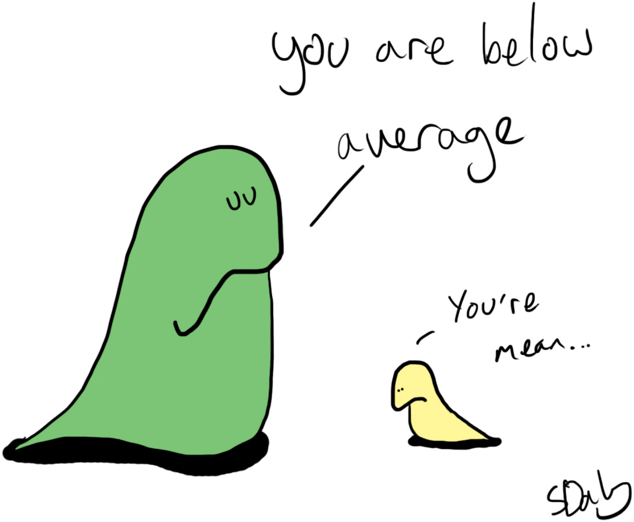
- Random variables and probability distributions provide useful infrastructure for everything we will do this year.
- Expected value and variance are two useful characteristics of the probability distributions associated with random variables.
- These concepts can be extended by conditioning on other variables.
- Next class we will cover joint distributions and conditional expectations in more depth.

Fun with

Fun with Averages

Fun with Averages

Central Tendency



The Story of Averages

Measurements

MESURES de la POITRINE.	NOMBRE d'hommes.	NOMBRE PROPORTIONNEL.	PROBABILITÉ d'après L'OBSERVATION.	RANG dans LA TABLE.	RANG d'après le CALCUL.	PROBABILITÉ d'après LA TABLE.	NOMBRE D'OBSERVATIONS calculé.
Pouces.							
33	3	5	0,5000			0,5000	7
34	18	31	0,4995	52	50	0,4993	29
35	81	141	0,4964	42,5	42,5	0,4964	110
36	185	322	0,4823	33,5	34,5	0,4854	323
37	420	732	0,4501	26,0	26,5	0,4531	732
38	749	1305	0,3769	18,0	18,5	0,3799	1333
39	1073	1867	0,2464	10,5	10,5	0,2466	1838
			0,0597	2,5	2,5	0,0628	
40	1079	1882	0,1285	5,5	5,5	0,1359	1987
41	934	1628	0,2913	13	13,5	0,3034	1675
42	658	1148	0,4061	21	21,5	0,4130	1096
43	370	645	0,4706	30	29,5	0,4690	560
44	92	160	0,4866	35	37,5	0,4911	221
45	50	87	0,4953	41	45,5	0,4980	69
46	21	38	0,4991	49,5	53,5	0,4996	16
47	4	7	0,4998	56	61,8	0,4999	3
48	1	2	0,5000			0,5000	1
	5738	1,0000					1,0000

Social Physics

Social Physics

The determination of the average man is not merely a matter of speculative curiosity; it may be of the most important service to the science of man and the social system. It ought necessarily to precede every other inquiry into social physics, since it is, as it were, the basis. The average man, indeed, is in a nation what the centre of gravity is in a body; it is by having that central point in view that we arrive at the apprehension of all the phenomena of equilibrium and motion
- Quetelet

The Military Takes to the Idea

The Problem with Averages

The Average Man

THE EDGE (HEALTH)

Stacking Up

Ever wonder how you measure up against the average American Joe? As a weight-training and fitness enthusiast, some of these stats may be a bit surprising...and pathetic.

69% Percentage of men who consider themselves "physically fit"

13% Percentage who actually are fit

$874 Amount spent on vitamins and supplements annually

45% Percentage of men who use their fitness equipment as a place to hang their clothes

52 Beats per minute of a fit man's resting heart rate

2.87 Hours per week of exercise

1.1 lbs. Amount of fat gained each year

1 Average number of pull-ups

77 yrs. Average life expectancy

23 Age when the average guy is in the best shape of his life

12:17 Average time to run a mile

1 Pounds of muscle the average sedentary guy loses each year

Size Yourself Up

Age (yrs.)	20s	30s	40s	50s	60s
Weight (lbs.)	168	179	182	185	184
Sit-ups (1 min.)	40	36	31	26	20
Push-ups (1 min.)	33	27	21	15	15
Bench Press (lbs.) (one time)	180	158	143	128	116

5'9.2" The average American male height

40" Size of the average guy's chest (inhaled) This is also the bodypart men spend the most time trying to develop

17.6% Average bodyfat

13" Size of the average guy's biceps (flexed)

34" Average waist size

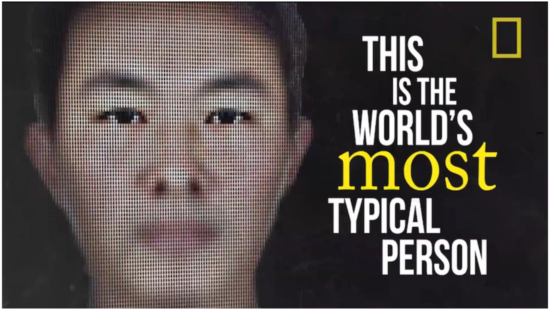
5.08 The average penis length in inches, according to *The Journal of Urology*

The *International Journal of Impotence Research* finds it slightly longer at 5.35 inches.

Sex Stats

- The average **sex session** lasts 3-10 minutes.
- The average **hotel porn-viewer** watches 12 minutes.
- The average **male orgasm** lasts six seconds.
- Number of men who would **rather work out than have sex**: 1 in 7

The Face of the Average Man



Fun with Sensitive Questions

Fun with Sensitive Questions

Graeme Blair
(slides that follow from Graeme)

Fun with Sensitive Questions

Cannot ask direct questions when there are **incentives to conceal sensitive responses**

Fun with Sensitive Questions

Cannot ask direct questions when there are **incentives to conceal sensitive responses**

1. Social pressure

Fun with Sensitive Questions

Cannot ask direct questions when there are **incentives to conceal sensitive responses**

1. Social pressure
2. Physical retaliation

Fun with Sensitive Questions

Cannot ask direct questions when there are **incentives to conceal sensitive responses**

1. Social pressure
2. Physical retaliation
3. Legal jeopardy

How to Address Incentives to Conceal

How to Address Incentives to Conceal

Develop trust with respondents, ask directly

How to Address Incentives to Conceal

Develop trust with respondents, ask directly

Survey experimental methods

1. **Endorsement experiment** Evaluation bias
2. **List experiment** Aggregation

How to Address Incentives to Conceal

Develop trust with respondents, ask directly

Survey experimental methods

1. **Endorsement experiment** Evaluation bias
2. **List experiment** Aggregation
3. **Randomized response** Random noise

Bias in Direct Questions on Vote Buying

Estimated rate of vote buying from direct survey item

2.4%

Gonzalez-Ocantos et al. 2011, *AJPS*

Question text: "they gave you a gift or did you a favor"

Bias in Direct Questions on Vote Buying

Estimated rate of vote buying from direct survey item

2.4%

Estimate using list experiment

24.3%

Gonzalez-Ocantos et al. 2011, *AJPS*

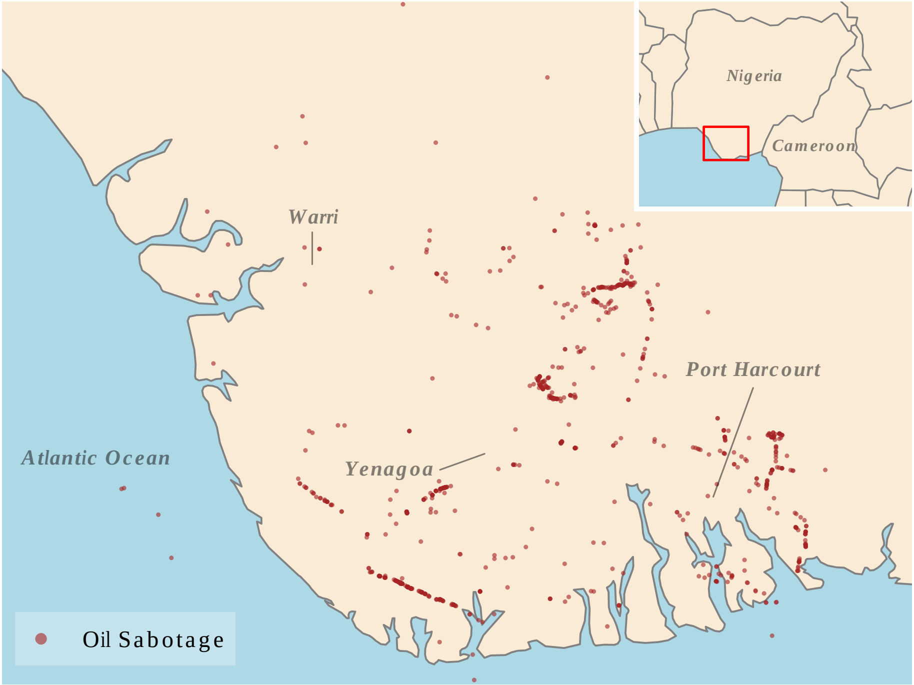
Question text: "they gave you a gift or did you a favor"

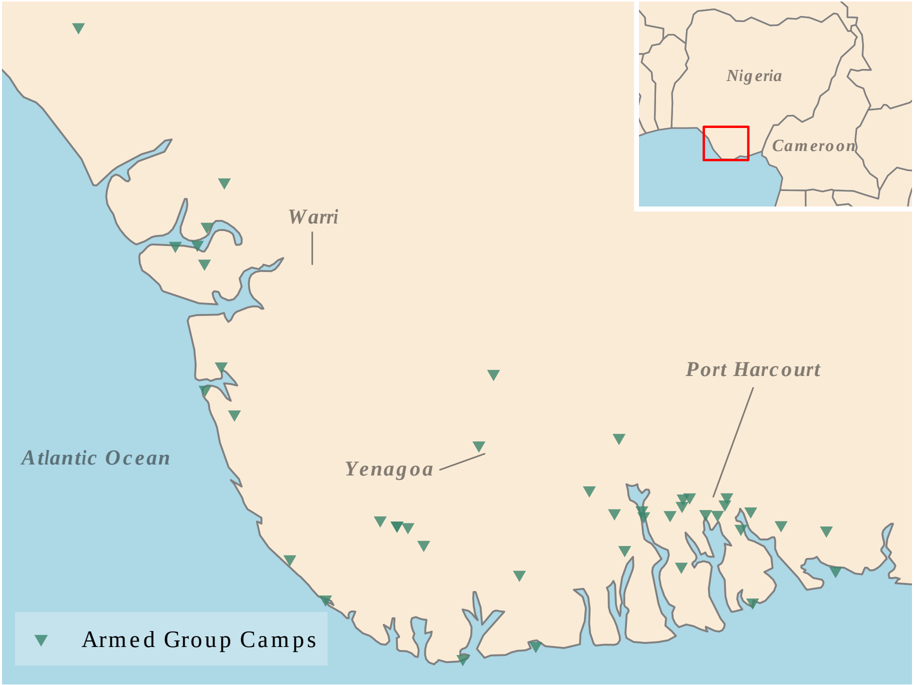
Survey

- Survey of 2,448 civilians in the Niger Delta

Survey

- Survey of 2,448 civilians in the Niger Delta
- Randomly sampled 204 communities near oil interruption sites and camps of armed groups


Nigeria
Cameroon
Warri
Port Harcourt
Atlantic Ocean
Yenagoa
Oil Sabotage


Nigeria
Cameroon
Warri
Atlantic Ocean
Yenagoa
Port Harcourt
Armed Group Camps

Survey

- Survey of 2,448 civilians in the Niger Delta
- Random sample of 204 communities near and far from oil interruption sites and armed group camps

Survey

- Survey of 2,448 civilians in the Niger Delta
- Random sample of 204 communities near and far from oil interruption sites and armed group camps
- Interviewed 12 people per community
  Random walk pattern to select households; Kish grid within household

Funded by the International Growth Centre

Outcome

"Did you share information with **militants** about their enemies in the community, state counterinsurgency forces, or oil facility activities?"

Problems with using list or endorsement experiments

Too sensitive for list experiment

Often difficult to define "control" condition in endorsement experiment for behaviors

Problems with using list or endorsement experiments

Too sensitive for list experiment

Often difficult to define "control" condition in endorsement experiment for behaviors

Alternative: **Randomized response technique**

Randomized response technique

How? Introducing random noise

Randomized response technique

How? Introducing random noise

- Roll the dice in private

Randomized response technique

How? Introducing random noise

- Roll the dice in private
- If you roll a 1, tell me "no"

Randomized response technique

How? Introducing random noise

- Roll the dice in private
- If you roll a 1, tell me "no"
- If you roll a 6, tell me "yes"

Randomized response technique

How? Introducing random noise

- Roll the dice in private
- If you roll a 1, tell me "no"
- If you roll a 6, tell me "yes"
- Otherwise, answer: "Did you share information with armed groups"

Analysis of the randomized response technique

1. Used fair dice, and actually rolled it.

Analysis of the randomized response technique

1. Used fair dice, and actually rolled it.
2. **Compliance.** Complied with "forced" response.

Analysis of the randomized response technique

1. Used fair dice, and actually rolled it.
2. **Compliance.** Complied with "forced" response.
3. **No Liars.** When not forced, answered truthfully.

Analysis of the randomized response technique

1. Used fair dice, and actually rolled it.
2. **Compliance.** Complied with "forced" response.
3. **No Liars.** When not forced, answered truthfully.

Proportion answered yes

$= 2/3 \cdot \text{Proportion yes to sensitive item} + 1/6$

Analysis of the randomized response technique

1. Used fair dice, and actually rolled it.
2. **Compliance.** Complied with "forced" response.
3. **No Liars.** When not forced, answered truthfully.

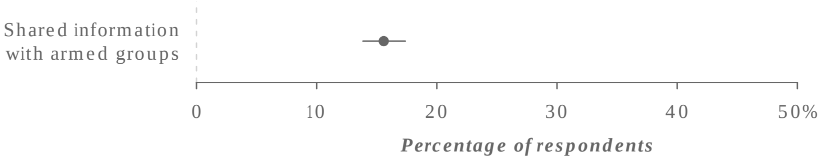
Proportion answered yes

$= 2/3 \cdot$ Proportion yes to sensitive item $+ 1/6$

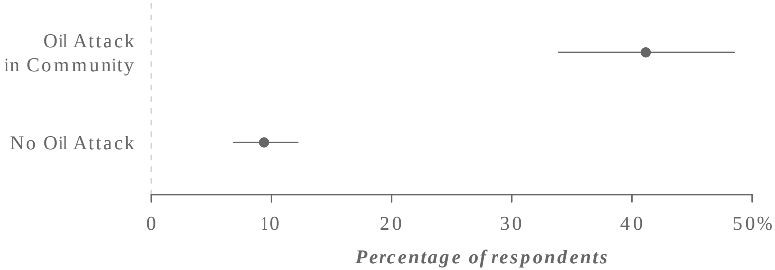
Proportion yes to sensitive item

$= 3/2 \cdot ($ Proportion answered yes $- 1/6)$

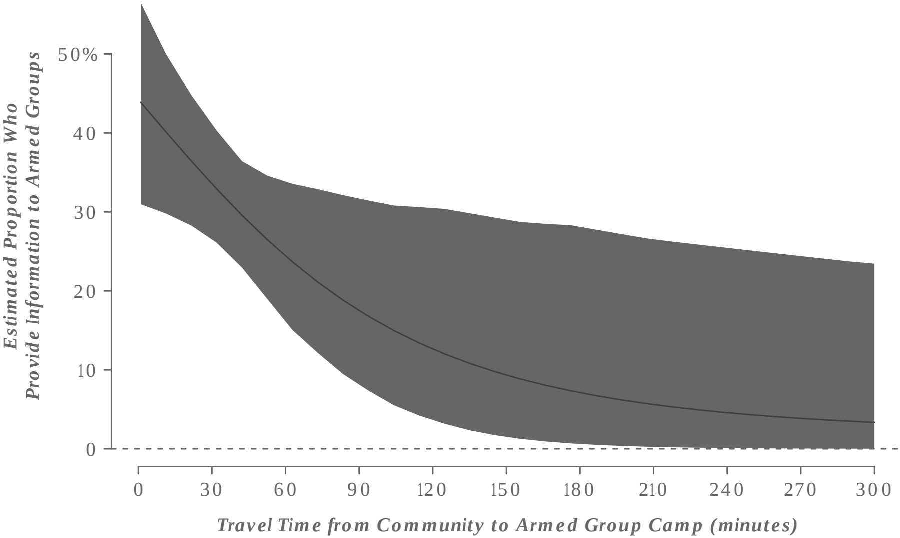
1. Civilians share information regularly with armed groups



2. Civilians near oil interruptions dominate collaboration



3. Civilians near armed group camps dominate collaboration



Software

- `rr` package in R for randomized response

  Blair with Yang-Yang Zhou and Kosuke Imai

- `list` package in R for list experiments

  Blair with Kosuke Imai

- `endorse` package in R for endorsement experiments

  Yuki Shiraito and Kosuke Imai

1. Random Variables and Distributions
   - What is a Random Variable?
   - Discrete Distributions
   - Continuous Distributions
2. Characteristics of Distributions
   - Central Tendency
   - Measures of Dispersion
3. Conditional Distributions
4. Fun with Averages
5. Fun with Sensitive Questions
6. Appendix: Why the Mean?
7. Joint Distributions
   - Discrete Random Variable
   - Continuous Random Variable
8. Conditional Expectation
9. Properties
   - Independence
   - Covariance and Correlation
   - Conditional Independence
10. Famous Distributions
11. Fun With Spam

1 Random Variables and Distributions
- What is a Random Variable?
- Discrete Distributions
- Continuous Distributions

2 Characteristics of Distributions
- Central Tendency
- Measures of Dispersion

3 Conditional Distributions

4 Fun with Averages

5 Fun with Sensitive Questions

6 Appendix: Why the Mean?

7 Joint Distributions
- Discrete Random Variable
- Continuous Random Variable

8 Conditional Expectation

9 Properties
- Independence
- Covariance and Correlation
- Conditional Independence

10 Famous Distributions

11 Fun With Spam

Expected Value as Mean Square Error Minimizer

Back

Suppose we want to pick a single number (c) that summarizes a random variable X. What we mean by summarizes determines the best choice of c.

Generally speaking we want a summary that is in the "center" of the data, i.e. that is as close as possible to all possible datapoints. Again though, the choice turns on what we mean by close.
Let's say we want to minimize:

- Mean Squared Error: $E(X-c)^2$
  This leads to choosing the mean of X: μ
- Mean Absolute Error: $E[|X-c|]$
  This leads to choosing the median of X: m

Let's prove the first result (see Blitzstein and Hwang 2014 Theorem 6.1.4 on pg 245 for this proof and the proof on mean absolute error).

Proof of Mean as Mean Square Error Minimizer

Let X be a random variable with mean μ. We want to show that the value of c that minimizes the mean squared error $E(X-c)^2$ is the mean, μ (Blitzstein and Hwang Theorem 6.1.4).

We will prove the following identity below:

$$E(X-c)^2 = \text{Var}(X) + (\mu - c)^2 \tag{1}$$

We are trying to choose c to minimize this term. The choice cannot affect $\text{Var}(X)$. Setting $c = \mu$ sets $(\mu - c)^2 = 0$ and any other choice makes $(\mu - c)^2 > 0$. Therefore (assuming the identity holds), $c = \mu$ minimizes Eq 1.

Now to prove the identity:

$$\begin{aligned}
\text{Var}(X) &= \text{Var}(X - c) && \text{(Prop 1 of Variance)}\\
&= E(X-c)^2 - (E[X-c])^2 && \text{(Defn of Variance)}\\
&= E(X-c)^2 - (\mu - c)^2 && \text{(Linearity of Exp)}\\
\text{Var}(X) + (\mu - c)^2 &= E(X-c)^2
\end{aligned}$$

▸▸ Back

References

- Kuklinski et al. 1997 "Racial prejudice and attitudes toward affirmative action" *American Journal of Political Science*
- Glynn 2013 "What can we learn with statistical truth serum? Design and analysis of the list experiment"
- All the Blair papers above.

Where We've Been and Where We're Going...

- Last Week

Where We've Been and Where We're Going...

- Last Week
  - welcome and outline of course
  - described uncertain outcomes with probability.

Where We've Been and Where We're Going...

- Last Week
  - welcome and outline of course
  - described uncertain outcomes with probability.
- This Week

Where We've Been and Where We're Going...

- Last Week
  - welcome and outline of course
  - described uncertain outcomes with probability.
- This Week
  - Monday:
    - summarize one random variable using expectation and variance

Where We've Been and Where We're Going...

- Last Week
  - welcome and outline of course
  - described uncertain outcomes with probability.
- This Week
  - Monday:
    - summarize one random variable using expectation and variance
    - show how to condition on a variable

Where We've Been and Where We're Going...

- Last Week
  - welcome and outline of course
  - described uncertain outcomes with probability.
- This Week
  - Monday:
    - summarize one random variable using expectation and variance
    - show how to condition on a variable
  - Wednesday:
    - properties of joint distributions

Where We've Been and Where We're Going...

- Last Week
  - welcome and outline of course
  - described uncertain outcomes with probability.
- This Week
  - Monday:
    - summarize one random variable using expectation and variance
    - show how to condition on a variable
  - Wednesday:
    - properties of joint distributions
    - conditional expectations

Where We've Been and Where We're Going...

- Last Week
  - welcome and outline of course
  - described uncertain outcomes with probability.
- This Week
  - Monday:
    - summarize one random variable using expectation and variance
    - show how to condition on a variable
  - Wednesday:
    - properties of joint distributions
    - conditional expectations
    - covariance, correlation, independence

Where We've Been and Where We're Going...

- Last Week
  - welcome and outline of course
  - described uncertain outcomes with probability.
- This Week
  - Monday:
    - summarize one random variable using expectation and variance
    - show how to condition on a variable
  - Wednesday:
    - properties of joint distributions
    - conditional expectations
    - covariance, correlation, independence
- Next Week

Where We've Been and Where We're Going...

- Last Week
  - welcome and outline of course
  - described uncertain outcomes with probability.
- This Week
  - Monday:
    - summarize one random variable using expectation and variance
    - show how to condition on a variable
  - Wednesday:
    - properties of joint distributions
    - conditional expectations
    - covariance, correlation, independence
- Next Week
  - estimating these features from data

Where We've Been and Where We're Going...

- Last Week
  - welcome and outline of course
  - described uncertain outcomes with probability.
- This Week
  - Monday:
    - summarize one random variable using expectation and variance
    - show how to condition on a variable
  - Wednesday:
    - properties of joint distributions
    - conditional expectations
    - covariance, correlation, independence
- Next Week
  - estimating these features from data
  - estimating uncertainty

Where We've Been and Where We're Going...

- Last Week
  - welcome and outline of course
  - described uncertain outcomes with probability.
- This Week
  - Monday:
    - summarize one random variable using expectation and variance
    - show how to condition on a variable
  - Wednesday:
    - properties of joint distributions
    - conditional expectations
    - covariance, correlation, independence
- Next Week
  - estimating these features from data
  - estimating uncertainty
- Long Run

Where We've Been and Where We're Going...

- Last Week
  - welcome and outline of course
  - described uncertain outcomes with probability.
- This Week
  - Monday:
    - summarize one random variable using expectation and variance
    - show how to condition on a variable
  - Wednesday:
    - properties of joint distributions
    - conditional expectations
    - covariance, correlation, independence
- Next Week
  - estimating these features from data
  - estimating uncertainty
- Long Run
  - probability $\rightarrow$ inference $\rightarrow$ regression $\rightarrow$ causal inference

Where We've Been and Where We're Going...

- Last Week
  - welcome and outline of course
  - described uncertain outcomes with probability.
- This Week
  - Monday:
    - summarize one random variable using expectation and variance
    - show how to condition on a variable
  - Wednesday:
    - properties of joint distributions
    - conditional expectations
    - covariance, correlation, independence
- Next Week
  - estimating these features from data
  - estimating uncertainty
- Long Run
  - probability $\rightarrow$ inference $\rightarrow$ regression $\rightarrow$ causal inference

Questions?

1. Random Variables and Distributions
   - What is a Random Variable?
   - Discrete Distributions
   - Continuous Distributions
2. Characteristics of Distributions
   - Central Tendency
   - Measures of Dispersion
3. Conditional Distributions
4. Fun with Averages
5. Fun with Sensitive Questions
6. Appendix: Why the Mean?
7. Joint Distributions
   - Discrete Random Variable
   - Continuous Random Variable
8. Conditional Expectation
9. Properties
   - Independence
   - Covariance and Correlation
   - Conditional Independence
10. Famous Distributions
11. Fun With Spam

1 Random Variables and Distributions
- What is a Random Variable?
- Discrete Distributions
- Continuous Distributions

2 Characteristics of Distributions
- Central Tendency
- Measures of Dispersion

3 Conditional Distributions

4 Fun with Averages

5 Fun with Sensitive Questions

6 Appendix: Why the Mean?

7 Joint Distributions
- Discrete Random Variable
- Continuous Random Variable

8 Conditional Expectation

9 Properties
- Independence
- Covariance and Correlation
- Conditional Independence

10 Famous Distributions

11 Fun With Spam

Joint Distributions

Joint Distributions

- We've talked about joint probabilities of events—what was the probability of A and B occurring: $P(A \cap B)$

Joint Distributions

- We've talked about joint probabilities of events—what was the probability of A and B occurring: $P(A \cap B)$
- We also talked about the conditional probability of A given that B occurred.

Joint Distributions

- We've talked about joint probabilities of events—what was the probability of A and B occurring: $P(A \cap B)$
- We also talked about the conditional probability of A given that B occurred.
- We also need to think about more than one r.v. at the same time.

Joint Distributions

- We've talked about joint probabilities of events—what was the probability of A and B occurring: $P(A \cap B)$
- We also talked about the conditional probability of A given that B occurred.
- We also need to think about more than one r.v. at the same time.
  - in regression we think about how the distribution of one variable changes under different values of another variable

Joint Distributions

- We've talked about joint probabilities of events—what was the probability of A and B occurring: $P(A \cap B)$
- We also talked about the conditional probability of A given that B occurred.
- We also need to think about more than one r.v. at the same time.
  - in regression we think about how the distribution of one variable changes under different values of another variable
  - e.g. does running more negative ads decrease election turnout?

Joint Distributions

- We've talked about joint probabilities of events—what was the probability of A and B occurring: $P(A \cap B)$
- We also talked about the conditional probability of A given that B occurred.
- We also need to think about more than one r.v. at the same time.
  - in regression we think about how the distribution of one variable changes under different values of another variable
  - e.g. does running more negative ads decrease election turnout?
- The joint distribution of two (or more) variables describes the pairs of observations that we are more or less likely to see.

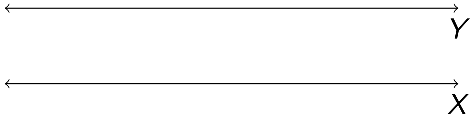
Understanding Joint Distributions

Understanding Joint Distributions

- Consider two r.v.s now, X and Y, each on the real line, $\mathbb{R}$.

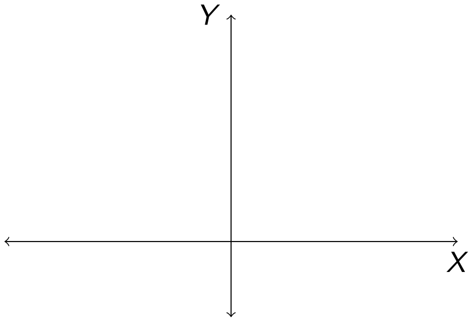
Understanding Joint Distributions

- Consider two r.v.s now, X and Y, each on the real line, $\mathbb{R}$.



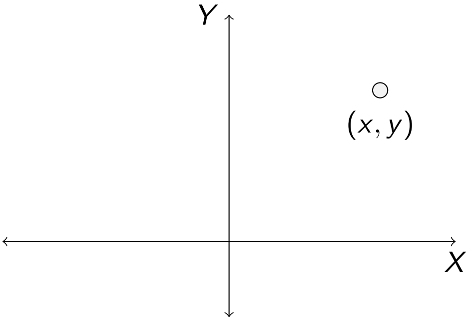
Understanding Joint Distributions

- Consider two r.v.s now, X and Y, each on the real line, $\mathbb{R}$.
- The pair form a two-dimensional space, or $\mathbb{R} \times \mathbb{R}$



Understanding Joint Distributions

- Consider two r.v.s now, X and Y, each on the real line, $\mathbb{R}$.
- The pair form a two-dimensional space, or $\mathbb{R} \times \mathbb{R}$
- One realization of the r.v. is a point in that space



Understanding Joint Distributions

Understanding Joint Distributions

- Imagine we are throwing darts on a two-dimensional board: the joint distribution tells us where the darts are more likely to land.

Understanding Joint Distributions

- Imagine we are throwing darts on a two-dimensional board: the joint distribution tells us where the darts are more likely to land.
- Distributions can be limited to a subset of the real line

Understanding Joint Distributions

- Imagine we are throwing darts on a two-dimensional board: the joint distribution tells us where the darts are more likely to land.
- Distributions can be limited to a subset of the real line
  - e.g. two uniform random variables might be between 0 and 1

Understanding Joint Distributions

- Imagine we are throwing darts on a two-dimensional board: the joint distribution tells us where the darts are more likely to land.
- Distributions can be limited to a subset of the real line
  - e.g. two uniform random variables might be between 0 and 1
  - e.g. discrete random variables typically only include integers

Understanding Joint Distributions

- Imagine we are throwing darts on a two-dimensional board: the joint distribution tells us where the darts are more likely to land.
- Distributions can be limited to a subset of the real line
  - e.g. two uniform random variables might be between 0 and 1
  - e.g. discrete random variables typically only include integers
- With two r.vs. there are now two dimensions to deal with.

Understanding Joint Distributions

- Imagine we are throwing darts on a two-dimensional board: the joint distribution tells us where the darts are more likely to land.
- Distributions can be limited to a subset of the real line
  - e.g. two uniform random variables might be between 0 and 1
  - e.g. discrete random variables typically only include integers
- With two r.vs. there are now two dimensions to deal with.
- Often, we are interested in two random variables that are qualitatively different:

Understanding Joint Distributions

- Imagine we are throwing darts on a two-dimensional board: the joint distribution tells us where the darts are more likely to land.
- Distributions can be limited to a subset of the real line
  - e.g. two uniform random variables might be between 0 and 1
  - e.g. discrete random variables typically only include integers
- With two r.vs. there are now two dimensions to deal with.
- Often, we are interested in two random variables that are qualitatively different:
  - Y (response, outcome, dependent variable, etc.)
    $=$ the random variable we want to explain, or predict.

Understanding Joint Distributions

- Imagine we are throwing darts on a two-dimensional board: the joint distribution tells us where the darts are more likely to land.
- Distributions can be limited to a subset of the real line
  - e.g. two uniform random variables might be between 0 and 1
  - e.g. discrete random variables typically only include integers
- With two r.vs. there are now two dimensions to deal with.
- Often, we are interested in two random variables that are qualitatively different:
  - Y (response, outcome, dependent variable, etc.)
    $=$ the random variable we want to explain, or predict.
  - X (predictor, explanatory/independent variable, covariate, etc.)
    $=$ the random variable with which we want to explain Y.

1. Random Variables and Distributions
   - What is a Random Variable?
   - Discrete Distributions
   - Continuous Distributions
2. Characteristics of Distributions
   - Central Tendency
   - Measures of Dispersion
3. Conditional Distributions
4. Fun with Averages
5. Fun with Sensitive Questions
6. Appendix: Why the Mean?
7. Joint Distributions
   - Discrete Random Variable
   - Continuous Random Variable
8. Conditional Expectation
9. Properties
   - Independence
   - Covariance and Correlation
   - Conditional Independence
10. Famous Distributions
11. Fun With Spam

1. Random Variables and Distributions
   - What is a Random Variable?
   - Discrete Distributions
   - Continuous Distributions
2. Characteristics of Distributions
   - Central Tendency
   - Measures of Dispersion
3. Conditional Distributions
4. Fun with Averages
5. Fun with Sensitive Questions
6. Appendix: Why the Mean?
7. Joint Distributions
   - Discrete Random Variable
   - Continuous Random Variable
8. Conditional Expectation
9. Properties
   - Independence
   - Covariance and Correlation
   - Conditional Independence
10. Famous Distributions
11. Fun With Spam

Joint Probability Mass Function

Definition

For two discrete random variables X and Y the joint PMF $P_{X,Y}(x, y)$ gives the probability that $X = x$ and $Y = y$ for all x and y:

$$P_{X,Y}(x, y) \; = \; \Pr(X = x \text{ and } Y = y)$$

Restrictions:

Joint Probability Mass Function

Definition

For two discrete random variables X and Y the joint PMF $P_{X,Y}(x,y)$ gives the probability that $X = x$ and $Y = y$ for all x and y:

$$P_{X,Y}(x,y) \;=\; \Pr(X = x \text{ and } Y = y)$$

Restrictions:

- $P_{X,Y}(x,y) \geq 0$ and $\sum_x \sum_y P_{X,Y}(x,y) = 1$.

Joint Probability Mass Function

Definition

For two discrete random variables X and Y the joint PMF $P_{X,Y}(x,y)$ gives the probability that $X = x$ and $Y = y$ for all x and y:

$$P_{X,Y}(x,y) \;=\; \Pr(X = x \text{ and } Y = y)$$

Should the U.S. allow more immigrants to come and live here?

		X: Education			
		less HS	HS	College	BA
Y: Support	oppose	0.07	0.22	0.18	0.15
	neutral	0.02	0.06	0.05	0.05
	favor	0.01	0.03	0.04	0.11

Joint Probability Mass Function

Definition

For two discrete random variables X and Y the joint PMF $P_{X,Y}(x, y)$ gives the probability that $X = x$ and $Y = y$ for all x and y:

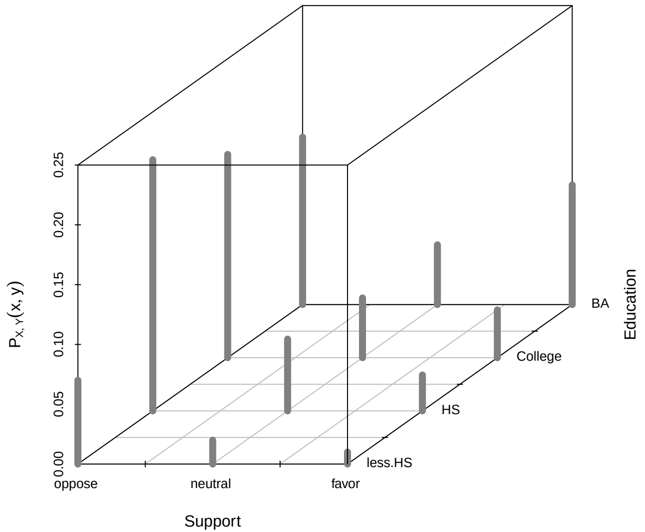
$$P_{X,Y}(x, y) \ = \ \Pr(X = x \text{ and } Y = y)$$

Should the U.S. allow more immigrants to come and live here?

		X: Education			
		less HS	HS	College	BA
Y: Support	oppose	0.07	0.22	0.18	0.15
	neutral	0.02	0.06	0.05	0.05
	favor	0.01	0.03	0.04	0.11

With discrete r.v.s this is very similar to thinking about a cross-tab, with frequencies/ probabilities in the cells instead of raw numbers.

Joint Probability Mass Function



From Joint to Marginal PMF

Given the joint PMF $P_{X,Y}(x,y)$ can we recover the marginal PMF $P_Y(y)$ (distribution over a single variable)?

		X: Education				
		less HS	HS	College	BA	
Y: Support	oppose	0.07	0.21	0.17	0.14	
	neutral	0.02	0.06	0.05	0.05	
	favor	0.01	0.03	0.04	0.10	

From Joint to Marginal PMF

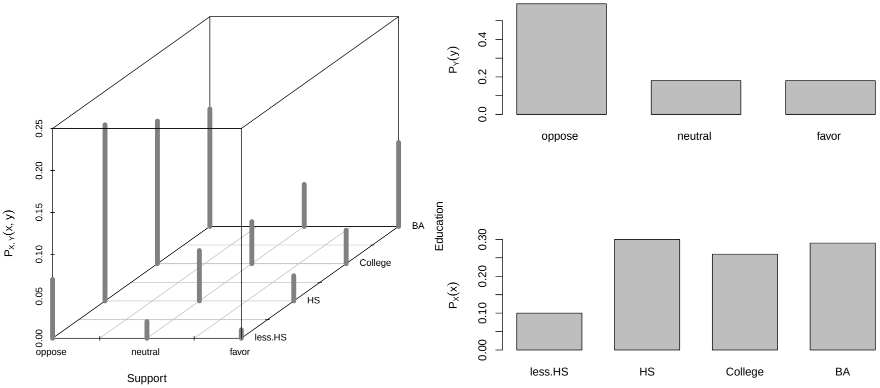
Given the joint PMF $P_{X,Y}(x,y)$ can we recover the marginal PMF $P_Y(y)$ (distribution over a single variable)?

		X: Education				
		less HS	HS	College	BA	$P_Y(y)$
Y: Support	oppose	0.07	0.21	0.17	0.14	0.62
	neutral	0.02	0.06	0.05	0.05	0.19
	favor	0.01	0.03	0.04	0.10	0.19

To obtain $P_Y(y)$ we marginalize the joint probability function $P_{X,Y}(x,y)$ over X:

$$P_Y(y) = \sum_x P_{X,Y}(x,y) = \sum_x \Pr(X = x, Y = y)$$

Joint and Marginal Probability Mass Functions



Why Does Marginalization Work?

Why Does Marginalization Work?

Begin with discrete case.

Why Does Marginalization Work?

Begin with discrete case. Consider jointly distributed discrete random variables, X and Y. We'll suppose they have joint pmf,

$$P(X = x, Y = y) = p(x, y)$$

Why Does Marginalization Work?

Begin with discrete case. Consider jointly distributed discrete random variables, X and Y. We'll suppose they have joint pmf,

$$P(X = x, Y = y) = p(x, y)$$

Suppose that the distribution allocates its mass at $x_1, x_2, \ldots, x_M$ and $y_1, y_2, \ldots, y_N$.

Why Does Marginalization Work?

Begin with discrete case. Consider jointly distributed discrete random variables, X and Y. We'll suppose they have joint pmf,

$$P(X = x, Y = y) = p(x, y)$$

Suppose that the distribution allocates its mass at $x_1, x_2, \ldots, x_M$ and $y_1, y_2, \ldots, y_N$.
Define the conditional mass function $P(X = x|Y = y)$ as,

$$\begin{aligned} P(X = x|Y = y) \equiv &= p(x|y) \\ &= p(x, y)/p(y) \end{aligned}$$

Why Does Marginalization Work?

Begin with discrete case. Consider jointly distributed discrete random variables, X and Y. We'll suppose they have joint pmf,

$$P(X = x, Y = y) = p(x, y)$$

Suppose that the distribution allocates its mass at $x_1, x_2, \ldots, x_M$ and $y_1, y_2, \ldots, y_N$.

Define the conditional mass function $P(X = x|Y = y)$ as,

$$\begin{aligned} P(X = x|Y = y) \equiv \quad &= \quad p(x|y) \\ &= \quad p(x, y)/p(y) \end{aligned}$$

Then it follows that:

$$p(x, y) \quad = \quad p(x|y)p(y)$$

Why Does Marginalization Work?

Begin with discrete case. Consider jointly distributed discrete random variables, X and Y. We'll suppose they have joint pmf,

$$P(X = x, Y = y) = p(x, y)$$

Suppose that the distribution allocates its mass at $x_1, x_2, \ldots, x_M$ and $y_1, y_2, \ldots, y_N$.
Define the conditional mass function $P(X = x|Y = y)$ as,

$$\begin{aligned} P(X = x|Y = y) \equiv & = p(x|y) \\ & = p(x, y)/p(y) \end{aligned}$$

Then it follows that:

$$p(x, y) = p(x|y)p(y)$$

Marginalizing over y to get $p(x)$ is then,

$$p(x_j) = \sum_{i=1}^{N} p(x_j|y_i)p(y_i)$$

A Table

	$Y = 0$	$Y = 1$	
$X = 0$	$p(0,0)$	$p(0, 1)$	$p_X(0)$
$X = 1$	$p(1,0)$	$p(1,1)$	$p_X(1)$
	$p_Y\ (0)$	$p_Y\ (1)$	

A Table

	Y = 0	Y = 1	
X = 0	0.01	0.05	?
X = 1	0.25	0.69	?
	0.26	0.74	

$$
\begin{aligned}
p_X(0) &= p(0|y=0)p(y=0) + p(0|y=1)p(y=1) \\
&= \frac{0.01}{0.26} \times 0.26 + \frac{0.05}{0.74} \times 0.74 \\
&= 0.06
\end{aligned}
$$

A Table

	Y = 0	Y = 1	
X = 0	0.01	0.05	?
X = 1	0.25	0.69	?
	0.26	0.74	

$$
\begin{aligned}
p_X(0) &= p(0|y=0)p(y=0) + p(0|y=1)p(y=1) \\
&= \frac{0.01}{0.26} \times 0.26 + \frac{0.05}{0.74} \times 0.74 \\
&= 0.06
\end{aligned}
$$

$$
\begin{aligned}
p_X(1) &= p(1|y=0)p(y=0) + p(1|y=1)p(y=1) \\
&= \frac{0.25}{0.26} \times 0.26 + \frac{0.69}{0.74} \times 0.74 \\
&= 0.94
\end{aligned}
$$

Conditional PMF

Conditional PMF

Definition

The conditional PMF of Y given X, $P_{Y|X}(y|x)$, is the PMF of Y when X is known to be at a particular value $X = x$:

$$P_{Y|X}(y|x) = \frac{\Pr(X = x \text{ and } Y = y)}{\Pr(X = x)} = \frac{P_{X,Y}(x, y)}{P_X(x)}$$

Conditional PMF

Definition

The conditional PMF of Y given X, $P_{Y|X}(y|x)$, is the PMF of Y when X is known to be at a particular value $X = x$:

$$P_{Y|X}(y|x) = \frac{\Pr(X = x \text{ and } Y = y)}{\Pr(X = x)} = \frac{P_{X,Y}(x,y)}{P_X(x)}$$

Key relationships:

- $P_{X,Y}(x,y) = P_{Y|X}(y|x)P_X(x)$ (multiplicative rule)
- $P_{Y|X}(y|x) = P_{X|Y}(x|y)P_Y(y)/P_X(x)$ (Bayes' rule)

Conditional PMF

Definition

The conditional PMF of Y given X, $P_{Y|X}(y|x)$, is the PMF of Y when X is known to be at a particular value $X = x$:

$$P_{Y|X}(y|x) = \frac{\Pr(X = x \text{ and } Y = y)}{\Pr(X = x)} = \frac{P_{X,Y}(x,y)}{P_X(x)}$$

Key relationships:

- $P_{X,Y}(x,y) = P_{Y|X}(y|x)P_X(x)$ (multiplicative rule)
- $P_{Y|X}(y|x) = P_{X|Y}(x|y)P_Y(y)/P_X(x)$ (Bayes' rule)

Conditional PMFs are just like ordinary PMFs, but refer to a universe where the "conditioning event" $(X = x)$ is known to have occurred.

Conditional PMF

Definition

The conditional PMF of Y given X, $P_{Y|X}(y|x)$, is the PMF of Y when X is known to be at a particular value $X = x$:

$$P_{Y|X}(y|x) = \frac{\Pr(X = x \text{ and } Y = y)}{\Pr(X = x)} = \frac{P_{X,Y}(x,y)}{P_X(x)}$$

Key relationships:

- $P_{X,Y}(x,y) = P_{Y|X}(y|x)P_X(x)$ (multiplicative rule)
- $P_{Y|X}(y|x) = P_{X|Y}(x|y)P_Y(y)/P_X(x)$ (Bayes' rule)

Conditional PMFs are just like ordinary PMFs, but refer to a universe where the "conditioning event" $(X = x)$ is known to have occurred.

Conditional distributions are key in statistical modeling because they inform us how the distribution of Y varies across different levels of X.

From Joint to Conditional: $P_{Y|X}(y|x) = \frac{P_{X,Y}(x,y)}{P_X(x)}$

Table: Joint PMF $P_{X,Y}(x,y)$ and Marginal PMFs $P_X(x), P_Y(y)$

		Education				
	$P_{X,Y}(x,y)$	less HS	HS	College	BA	$P_Y(y)$
Support	oppose	0.07	0.22	0.18	0.15	0.62
	neutral	0.02	0.06	0.05	0.05	0.19
	favor	0.01	0.03	0.04	0.11	0.19
	$P_X(x)$	0.11	0.32	0.27	0.31	1.00

From Joint to Conditional: $P_{Y|X}(y|x) = \frac{P_{X,Y}(x,y)}{P_X(x)}$

Table: Joint PMF $P_{X,Y}(x,y)$ and Marginal PMFs $P_X(x), P_Y(y)$

		Education				
	$P_{X,Y}(x,y)$	less HS	HS	College	BA	$P_Y(y)$
Support	oppose	0.07	0.22	0.18	0.15	0.62
	neutral	0.02	0.06	0.05	0.05	0.19
	favor	0.01	0.03	0.04	0.11	0.19
	$P_X(x)$	0.11	0.32	0.27	0.31	1.00

Table: Conditional PMF $P_{Y|X}(y|x)$

		Education				
	$P_{Y|X}(y|x)$	less HS	HS	College	BA	
Support	oppose	0.70	0.70	0.65	0.48	0.62
	neutral	0.20	0.20	0.19	0.17	0.19
	favor	0.10	0.10	0.15	0.34	0.19
		1.00	1.00	1.00	1.00	1.00

Joint and Conditional Probability Mass Functions

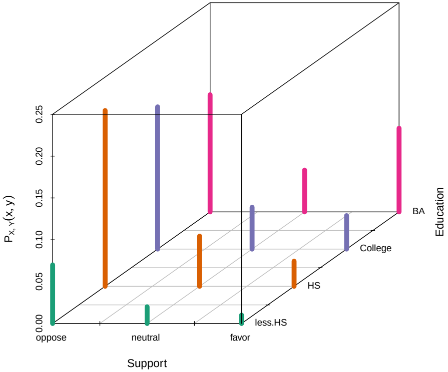


Figure: Joint

Joint and Conditional Probability Mass Functions

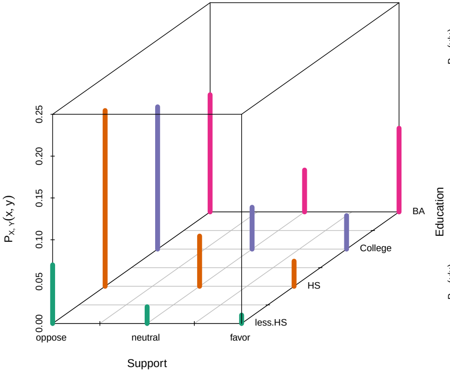


Figure: Joint

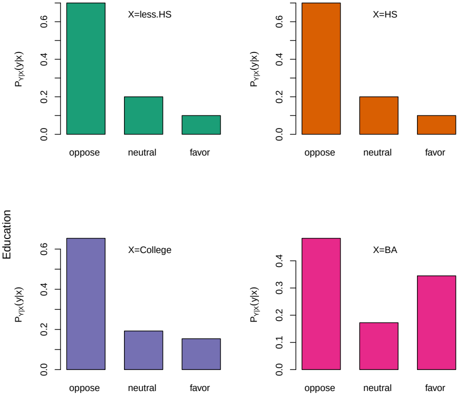


Figure: Conditional

1. Random Variables and Distributions
    - What is a Random Variable?
    - Discrete Distributions
    - Continuous Distributions
2. Characteristics of Distributions
    - Central Tendency
    - Measures of Dispersion
3. Conditional Distributions
4. Fun with Averages
5. Fun with Sensitive Questions
6. Appendix: Why the Mean?
7. Joint Distributions
    - Discrete Random Variable
    - Continuous Random Variable
8. Conditional Expectation
9. Properties
    - Independence
    - Covariance and Correlation
    - Conditional Independence
10. Famous Distributions
11. Fun With Spam

1 Random Variables and Distributions
- What is a Random Variable?
- Discrete Distributions
- Continuous Distributions

2 Characteristics of Distributions
- Central Tendency
- Measures of Dispersion

3 Conditional Distributions

4 Fun with Averages

5 Fun with Sensitive Questions

6 Appendix: Why the Mean?

7 Joint Distributions
- Discrete Random Variable
- Continuous Random Variable

8 Conditional Expectation

9 Properties
- Independence
- Covariance and Correlation
- Conditional Independence

10 Famous Distributions

11 Fun With Spam

Joint Probability Density Function

Definition

For two continuous random variables X and Y the joint PDF $f_{X,Y}(x, y)$ gives the density height where $X = x$ and $Y = y$ for all x and y.

Joint Probability Density Function

Definition

For two continuous random variables X and Y the joint PDF $f_{X,Y}(x, y)$ gives the density height where $X = x$ and $Y = y$ for all x and y.

The multiplicative rule:

$$f_{X,Y}(x, y) = f_{Y|X}(y|x) f_X(x)$$

where

- $f_{Y|X}(y|x)$: Conditional PDF of Y given $X = x$
- $f_X(x)$: Marginal PDF of X

Restrictions:

Joint Probability Density Function

Definition

For two continuous random variables X and Y the joint PDF $f_{X,Y}(x,y)$ gives the density height where $X = x$ and $Y = y$ for all x and y.

The multiplicative rule:

$$f_{X,Y}(x,y) = f_{Y|X}(y|x) f_X(x)$$

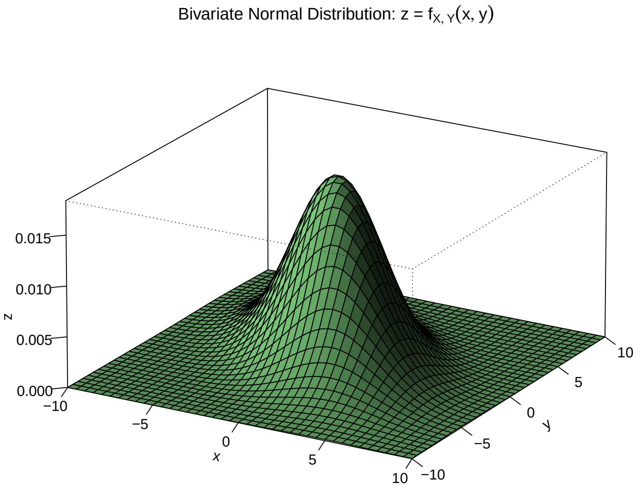
where

- $f_{Y|X}(y|x)$: Conditional PDF of Y given $X = x$
- $f_X(x)$: Marginal PDF of X

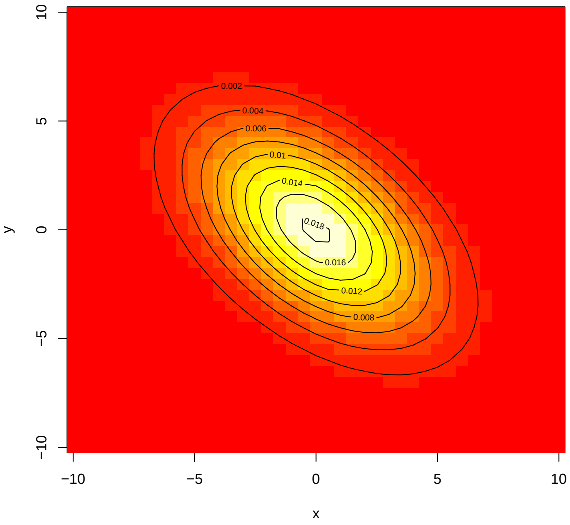
Restrictions:

- $\int_x \int_y f_{X,Y}(x,y)\, dy\, dx = 1$

3D Plot of a Joint Probability Density Function



Contour Plot of a Joint Probability Density Function



From Joint to Marginal PDF

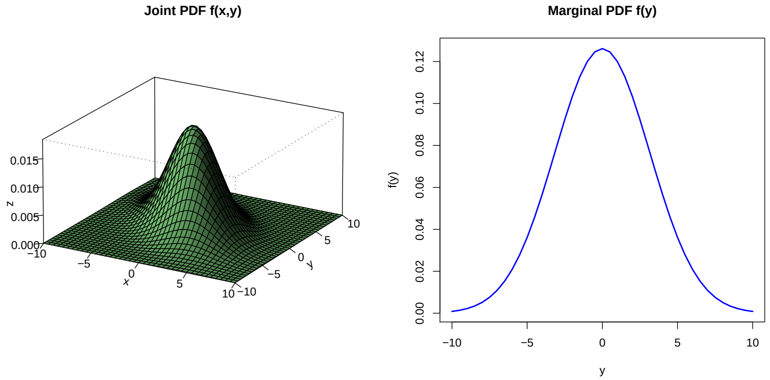
How can we obtain $f_Y(y)$ from $f_{X,Y}(x,y)$?

From Joint to Marginal PDF

How can we obtain $f_Y(y)$ from $f_{X,Y}(x,y)$?

We marginalize the joint probability function $f_{X,Y}(x,y)$ over X:

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)dx$$



1. Random Variables and Distributions
    - What is a Random Variable?
    - Discrete Distributions
    - Continuous Distributions
2. Characteristics of Distributions
    - Central Tendency
    - Measures of Dispersion
3. Conditional Distributions
4. Fun with Averages
5. Fun with Sensitive Questions
6. Appendix: Why the Mean?
7. Joint Distributions
    - Discrete Random Variable
    - Continuous Random Variable
8. Conditional Expectation
9. Properties
    - Independence
    - Covariance and Correlation
    - Conditional Independence
10. Famous Distributions
11. Fun With Spam

1. Random Variables and Distributions
   - What is a Random Variable?
   - Discrete Distributions
   - Continuous Distributions
2. Characteristics of Distributions
   - Central Tendency
   - Measures of Dispersion
3. Conditional Distributions
4. Fun with Averages
5. Fun with Sensitive Questions
6. Appendix: Why the Mean?
7. Joint Distributions
   - Discrete Random Variable
   - Continuous Random Variable
8. **Conditional Expectation**
9. Properties
   - Independence
   - Covariance and Correlation
   - Conditional Independence
10. Famous Distributions
11. Fun With Spam

Conditioning on X

Conditioning on X

- Goal in statistical modeling is often to characterize the conditional distribution of the outcome variable $f_{Y|X}(y|x)$ across different levels of $X = x$.

Conditioning on X

- Goal in statistical modeling is often to characterize the conditional distribution of the outcome variable $f_{Y|X}(y|x)$ across different levels of $X = x$.

- Typically, we summarize the conditional distributions with a few parameters such as the conditional mean of $E[Y|X = x]$ and the conditional variance $V[Y|X = x]$

Conditioning on X

- Goal in statistical modeling is often to characterize the conditional distribution of the outcome variable $f_{Y|X}(y|x)$ across different levels of $X = x$.
- Typically, we summarize the conditional distributions with a few parameters such as the conditional mean of $E[Y|X = x]$ and the conditional variance $V[Y|X = x]$
- Moreover, we are often interested in estimating $E[Y|X]$, i.e. the conditional expectation function that describes how the conditional mean of Y varies across all possible values of X (we sometimes call this the population regression function)

Conditional Expectation

Definition (Conditional Expectation (Discrete))

Let Y and X be discrete random variables. The conditional expectation of Y given $X = x$ is defined as:

$$E[Y|X=x] = \sum_y y \Pr(Y=y|X=x) = \sum_y y\, P_{Y|X}(y|x)$$

Conditional Expectation

Definition (Conditional Expectation (Discrete))

Let Y and X be discrete random variables. The conditional expectation of Y given $X = x$ is defined as:

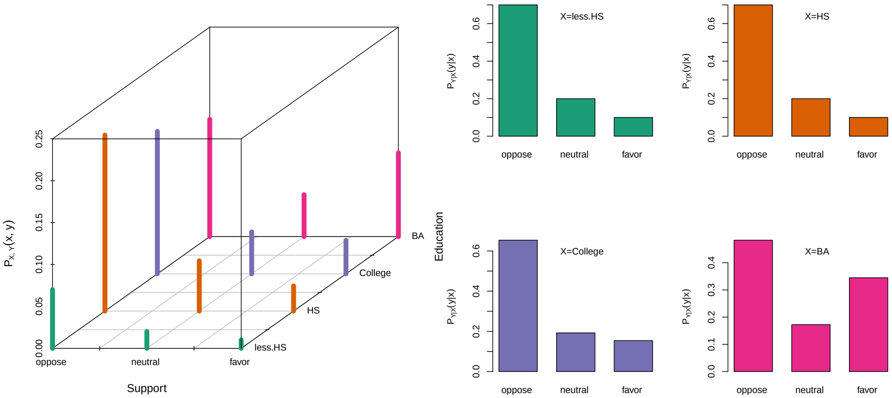
$$E[Y|X = x] = \sum_y y \Pr(Y = y|X = x) = \sum_y y \, P_{Y|X}(y|x)$$

Definition (Conditional Expectation (Continuous))

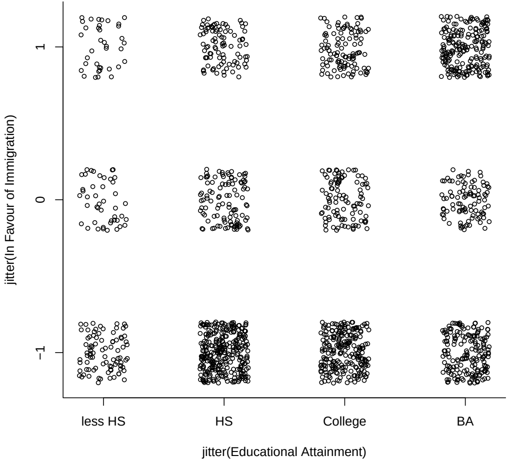
Let Y and X be continuous random variables. The conditional expectation of Y given $X = x$ is given by:

$$E[Y|X = x] = \int_{-\infty}^{\infty} y \, f_{Y|X}(y|x) dy$$

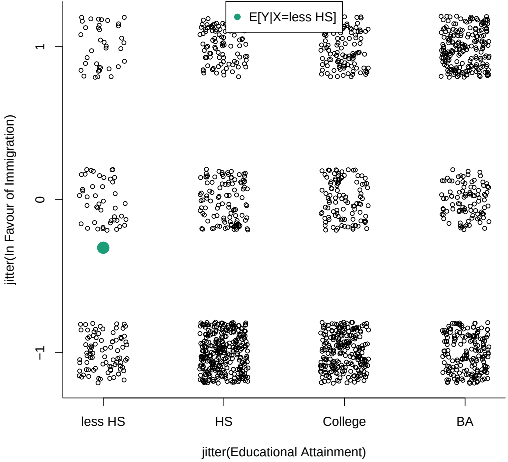
Joint and Conditional Probability Mass Functions



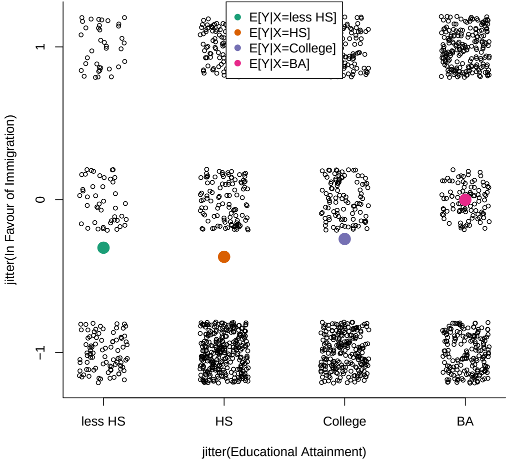
Conditional PMF $P_{Y|X}(y|x)$



Conditional Expectation $E[Y|X = 1]$



Conditional Expectation Function $E[Y|X]$



Law of Iterated Expectations

Theorem (Law of Iterated Expectations)

For two random variables X and Y,

$$E[Y] = E[E[Y|X]] = \begin{cases} \displaystyle\sum_{\text{all } x} E[Y|X=x] \cdot P_X(x) & (\text{discrete } X) \\ \displaystyle\int_{-\infty}^{\infty} E[Y|X=x] \cdot f_X(x)dx & (\text{continuous } X) \end{cases}$$

Note that the outer expectation is taken with respect to the distribution of X.

Law of Iterated Expectations

Theorem (Law of Iterated Expectations)

For two random variables X and Y,

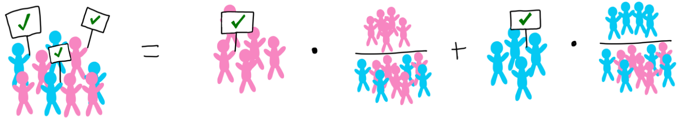
$$E[Y] = E[E[Y|X]] = \begin{cases} \sum_{\text{all } x} E[Y|X=x] \cdot P_X(x) & (\textit{discrete } X) \\ \int_{-\infty}^{\infty} E[Y|X=x] \cdot f_X(x)dx & (\textit{continuous } X) \end{cases}$$

Note that the outer expectation is taken with respect to the distribution of X.

Example: Y (support) and $X \in \{1, 0\}$ (gender). Then, the LIE tells us:

$$\underbrace{E[Y]}_{\text{Average Support}} = E[E[Y|X]]$$

$$= \underbrace{E[Y|X=1]}_{\text{Average Support}|\text{Woman}} \cdot \underbrace{P_X(1)}_{\text{Pr(Woman)}} + \underbrace{E[Y|X=0]}_{\text{Average Support}|\text{Man}} \cdot \underbrace{P_X(0)}_{\text{Pr(Man)}}$$



Properties of Conditional Expectation

Properties of Conditional Expectation

Conditional expectations have some convenient properties

Properties of Conditional Expectation

Conditional expectations have some convenient properties

1. $E[c(X)|X] = c(X)$ for any function $c(X)$.

Properties of Conditional Expectation

Conditional expectations have some convenient properties

1. $E[c(X)|X] = c(X)$ for any function $c(X)$.
   - Basically, any function of X is a constant with regard to the conditional expectation. If we know X, then we also know X^2, for instance.

Properties of Conditional Expectation

Conditional expectations have some convenient properties

1. $E[c(X)|X] = c(X)$ for any function $c(X)$.
   - Basically, any function of X is a constant with regard to the conditional expectation. If we know X, then we also know X^2, for instance.
2. If $E[Y^2] < \infty$ and $E[g(X)^2] < \infty$ for some function g, then $E[(Y - E[Y|X])^2|X] \leq E[(Y - g(X))^2|X]$ and $E[(Y - E[Y|X])^2] \leq E[(Y - g(X))^2]$

Properties of Conditional Expectation

Conditional expectations have some convenient properties

1. $E[c(X)|X] = c(X)$ for any function $c(X)$.
   - Basically, any function of X is a constant with regard to the conditional expectation. If we know X, then we also know X^2, for instance.
2. If $E[Y^2] < \infty$ and $E[g(X)^2] < \infty$ for some function g, then $E[(Y - E[Y|X])^2|X] \leq E[(Y - g(X))^2|X]$ and $E[(Y - E[Y|X])^2] \leq E[(Y - g(X))^2]$

The second property is quite important. It says that the conditional expectation is the function of X that minimizes the squared prediction error for Y across any possible function of X.

Conditional Variance

Conditional Variance

Conditional expectation gives us information about the central tendency of a random variable given another random variable.

Conditional Variance

Conditional expectation gives us information about the central tendency of a random variable given another random variable.

We also want to know the conditional variance to understand our uncertainty about the conditional distribution.

Conditional Variance

Conditional expectation gives us information about the central tendency of a random variable given another random variable.

We also want to know the conditional variance to understand our uncertainty about the conditional distribution.

Remember, the conditional distribution of $Y|X$ is basically like any other probability distribution, so we are going to want to summarize the center *and* spread.

Conditional Variance

Definition

The conditional variance of Y given $X = x$ is defined as:

$$V[Y|X=x] = \begin{cases} \sum_{\text{all } y} (y - E[Y|X=x])^2 P_{Y|X}(y|x) & \text{(discrete } Y\text{)} \\ \int_{-\infty}^{\infty} (y - E[Y|X=x])^2 f_{Y|X}(y|x) dy & \text{(continuous } Y\text{)} \end{cases}$$

Conditional Variance

Definition

The conditional variance of Y given $X = x$ is defined as:

$$V[Y|X=x] = \begin{cases} \sum_{\text{all } y} (y - E[Y|X=x])^2 P_{Y|X}(y|x) & \text{(discrete } Y\text{)} \\ \int_{-\infty}^{\infty} (y - E[Y|X=x])^2 f_{Y|X}(y|x) dy & \text{(continuous } Y\text{)} \end{cases}$$

A useful rule related to conditional variance is the law of total variance:

$$\underbrace{V[Y]}_{\text{Total variance}} = \underbrace{E[V[Y|X]]}_{\text{Average of Group Variances}} + \underbrace{V[E[Y|X]]}_{\text{Variance in Group Averages}}$$

Conditional Variance

Definition

The conditional variance of Y given $X = x$ is defined as:

$$V[Y|X=x] = \begin{cases} \sum_{\text{all } y} (y - E[Y|X=x])^2 P_{Y|X}(y|x) & \text{(discrete } Y\text{)} \\ \int_{-\infty}^{\infty} (y - E[Y|X=x])^2 f_{Y|X}(y|x) dy & \text{(continuous } Y\text{)} \end{cases}$$

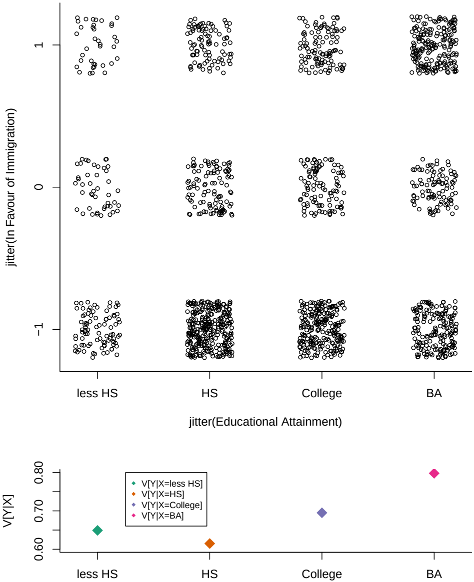
A useful rule related to conditional variance is the law of total variance:

$$\underbrace{V[Y]}_{\text{Total variance}} = \underbrace{E[V[Y|X]]}_{\text{Average of Group Variances}} + \underbrace{V[E[Y|X]]}_{\text{Variance in Group Averages}}$$

Example: Y (support) and $X \in \{1, 0\}$ (gender). The LTV says that the total variance in support can be decomposed into two parts:

1. On average, how much support varies within gender groups (within variance)
2. How much average support varies between gender groups (between variance)

Conditional Variance Function $V[Y|X]$



Important Subtleties

Important Subtleties

- It is important to distinguish between what is random/stochastic and what is constant. However, this can be tricky at first.

Important Subtleties

- It is important to distinguish between what is random/stochastic and what is constant. However, this can be tricky at first.
- If X is a random variable, generally a function of X ($g(X)$) is also a random variable.

Important Subtleties

- It is important to distinguish between what is random/stochastic and what is constant. However, this can be tricky at first.
- If X is a random variable, generally a function of X ($g(X)$) is also a random variable.
- $E[X]$ is a constant though (we sometimes refer to $E[\cdot]$ as an operator to make clear it doesn't behave the same as $g(\cdot)$).
- $E[X|Y]$ is random though.

Important Subtleties

- It is important to distinguish between what is random/stochastic and what is constant. However, this can be tricky at first.
- If X is a random variable, generally a function of X ($g(X)$) is also a random variable.
- $E[X]$ is a constant though (we sometimes refer to $E[\cdot]$ as an operator to make clear it doesn't behave the same as $g(\cdot)$).
- $E[X|Y]$ is random though.
- Why?

Important Subtleties

- It is important to distinguish between what is random/stochastic and what is constant. However, this can be tricky at first.
- If X is a random variable, generally a function of X ($g(X)$) is also a random variable.
- $E[X]$ is a constant though (we sometimes refer to $E[\cdot]$ as an operator to make clear it doesn't behave the same as $g(\cdot)$).
- $E[X|Y]$ is random though.
- Why? There is no longer anything stochastic in $E[X]$. Take the discrete case: $E[X] = \sum_x xp(X = x)$. Note that this is entirely in terms of realized values.

Important Subtleties

- It is important to distinguish between what is random/stochastic and what is constant. However, this can be tricky at first.
- If X is a random variable, generally a function of X ($g(X)$) is also a random variable.
- $E[X]$ is a constant though (we sometimes refer to $E[\cdot]$ as an operator to make clear it doesn't behave the same as $g(\cdot)$).
- $E[X|Y]$ is random though.
- Why? There is no longer anything stochastic in $E[X]$. Take the discrete case: $E[X] = \sum_x xp(X = x)$. Note that this is entirely in terms of realized values.
- By contrast $E[X|Y]$ is a function into which one can plug a value of $Y = y$ and get the expectation of X conditional on that value. Thus the randomness 'comes from' Y.

Important Subtleties

- It is important to distinguish between what is random/stochastic and what is constant. However, this can be tricky at first.
- If X is a random variable, generally a function of X ($g(X)$) is also a random variable.
- $E[X]$ is a constant though (we sometimes refer to $E[\cdot]$ as an operator to make clear it doesn't behave the same as $g(\cdot)$).
- $E[X|Y]$ is random though.
- Why? There is no longer anything stochastic in $E[X]$. Take the discrete case: $E[X] = \sum_x xp(X = x)$. Note that this is entirely in terms of realized values.
- By contrast $E[X|Y]$ is a function into which one can plug a value of $Y = y$ and get the expectation of X conditional on that value. Thus the randomness 'comes from' Y.

Let's look at this in pictures.
(If you want to know more: Blitzstein and Hwang pg 392-393 is great.)

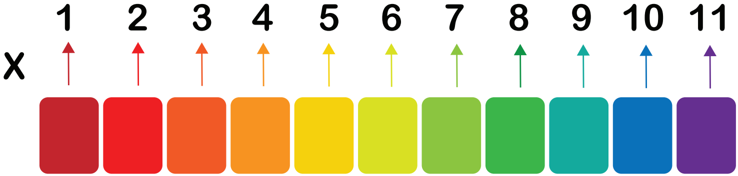
Important Subtleties in Pictures

Sample space

Important Subtleties in Pictures

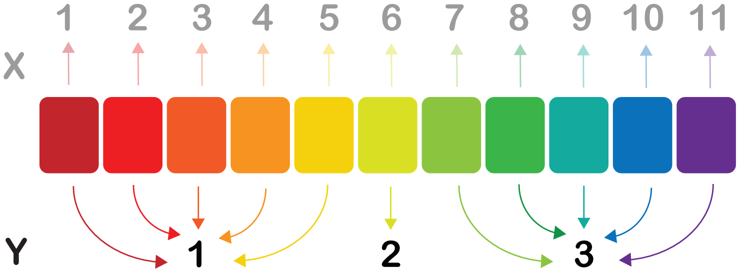
Sample space

Important Subtleties in Pictures



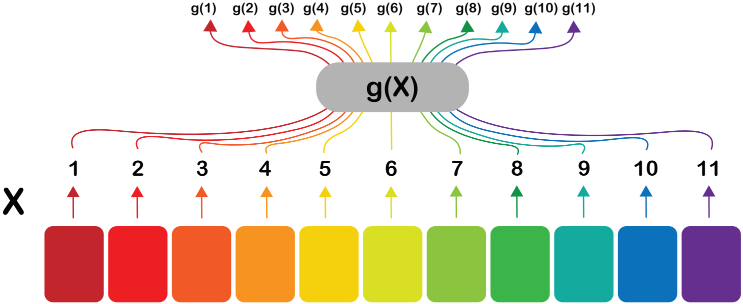
Random variable

Important Subtleties in Pictures



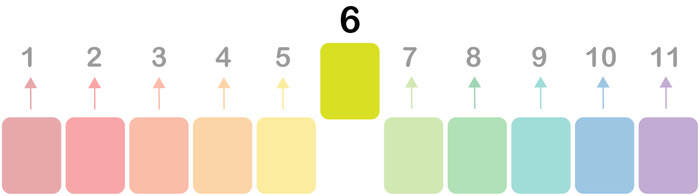
Random variable

Important Subtleties in Pictures



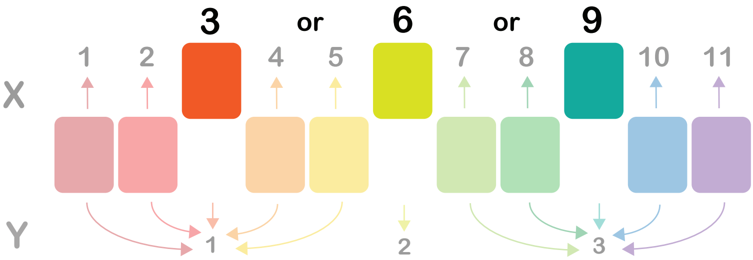
Function of a random variable is a random variable

Important Subtleties in Pictures



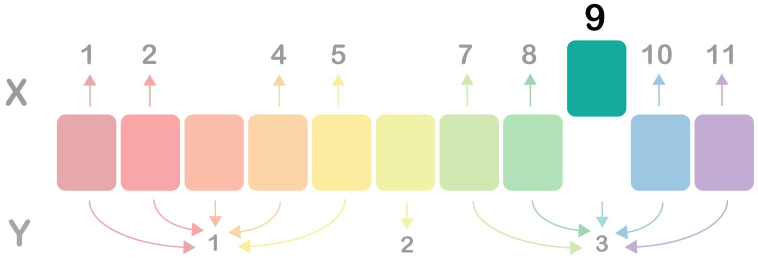
$E[X]$

Important Subtleties in Pictures



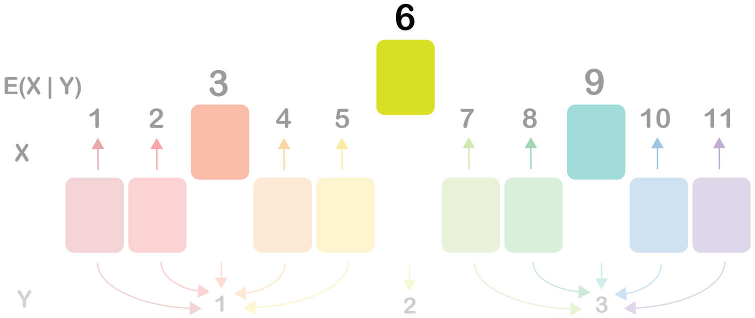
$E[X|Y]$

Important Subtleties in Pictures



$$E[X|Y=3]$$

Important Subtleties in Pictures



$$E[E[X|Y]] = E[X]$$

1. Random Variables and Distributions
   - What is a Random Variable?
   - Discrete Distributions
   - Continuous Distributions
2. Characteristics of Distributions
   - Central Tendency
   - Measures of Dispersion
3. Conditional Distributions
4. Fun with Averages
5. Fun with Sensitive Questions
6. Appendix: Why the Mean?
7. Joint Distributions
   - Discrete Random Variable
   - Continuous Random Variable
8. Conditional Expectation
9. Properties
   - Independence
   - Covariance and Correlation
   - Conditional Independence
10. Famous Distributions
11. Fun With Spam

1 Random Variables and Distributions
- What is a Random Variable?
- Discrete Distributions
- Continuous Distributions

2 Characteristics of Distributions
- Central Tendency
- Measures of Dispersion

3 Conditional Distributions

4 Fun with Averages

5 Fun with Sensitive Questions

6 Appendix: Why the Mean?

7 Joint Distributions
- Discrete Random Variable
- Continuous Random Variable

8 Conditional Expectation

9 Properties
- Independence
- Covariance and Correlation
- Conditional Independence

10 Famous Distributions

11 Fun With Spam

Independence

Definition (Independence of Random Variables)

Two random variables Y and X are independent if

$$f_{X,Y}(x,y) = f_X(x)f_Y(y)$$

for all x and y. We write this as $Y \perp\!\!\!\perp X$.

Independence

Definition (Independence of Random Variables)

Two random variables Y and X are independent if

$$f_{X,Y}(x,y) = f_X(x)f_Y(y)$$

for all x and y. We write this as $Y \perp\!\!\!\perp X$.

Independence implies

$$f_{Y|X}(y|x) = f_Y(y)$$

and thus

$$E[Y|X=x] =$$

Independence

Definition (Independence of Random Variables)

Two random variables Y and X are independent if

$$f_{X,Y}(x,y) = f_X(x) f_Y(y)$$

for all x and y. We write this as $Y \perp\!\!\!\perp X$.

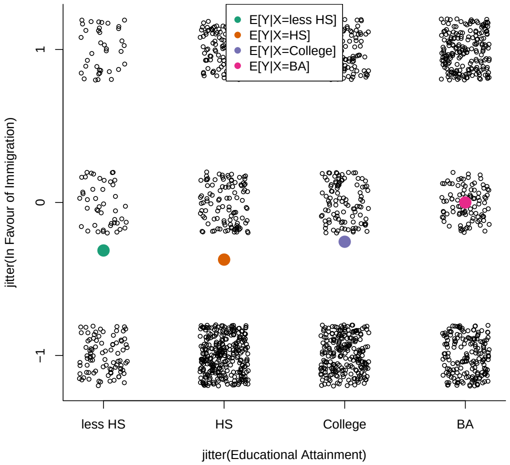
Independence implies

$$f_{Y|X}(y|x) = f_Y(y)$$

and thus

$$E[Y|X = x] = E[Y]$$

Is $Y \perp\!\!\!\perp X$?



Expected Values with Independent Random Variables

If random variables X and Y are independent, then

$$E[XY] = E[X]E[Y]$$

Expected Values with Independent Random Variables

If random variables X and Y are independent, then

$$E[XY] = E[X]E[Y]$$

Proof: For discrete X and Y,

$$E[XY] =$$

Expected Values with Independent Random Variables

If random variables X and Y are independent, then

$$E[XY] = E[X]E[Y]$$

Proof: For discrete X and Y,

$$E[XY] = \sum_{\text{all } x} \sum_{\text{all } y} x\, y\, P_{X,Y}(x, y)$$

Expected Values with Independent Random Variables

If random variables X and Y are independent, then

$$E[XY] = E[X]E[Y]$$

Proof: For discrete X and Y,

$$\begin{aligned} E[XY] &= \sum_{\text{all } x} \sum_{\text{all } y} x\, y\, P_{X,Y}(x, y) \\ &= \sum_{\text{all } x} \sum_{\text{all } y} x\, y\, P_X(x) P_Y(y) \end{aligned}$$

Expected Values with Independent Random Variables

If random variables X and Y are independent, then

$$E[XY] = E[X]E[Y]$$

Proof: For discrete X and Y,

$$\begin{aligned} E[XY] &= \sum_{\text{all } x} \sum_{\text{all } y} x\, y\, P_{X,Y}(x, y) \\ &= \sum_{\text{all } x} \sum_{\text{all } y} x\, y\, P_X(x) P_Y(y) \\ &= \sum_{\text{all } x} x\, P_X(x) \sum_{\text{all } y} y\, P_Y(y) \end{aligned}$$

Expected Values with Independent Random Variables

If random variables X and Y are independent, then

$$E[XY] = E[X]E[Y]$$

Proof: For discrete X and Y,

$$\begin{aligned}
E[XY] &= \sum_{\text{all } x} \sum_{\text{all } y} x\, y\, P_{X,Y}(x, y) \\
&= \sum_{\text{all } x} \sum_{\text{all } y} x\, y\, P_X(x) P_Y(y) \\
&= \sum_{\text{all } x} x\, P_X(x) \sum_{\text{all } y} y\, P_Y(y) \\
&= E[X]E[Y]
\end{aligned}$$

Expected Values with Independent Random Variables

If random variables X and Y are independent, then

$$E[XY] = E[X]E[Y]$$

Proof: For discrete X and Y,

$$\begin{aligned} E[XY] &= \sum_{\text{all } x}\sum_{\text{all } y} x\, y\, P_{X,Y}(x,y) \\ &= \sum_{\text{all } x}\sum_{\text{all } y} x\, y\, P_X(x) P_Y(y) \\ &= \sum_{\text{all } x} x\, P_X(x) \sum_{\text{all } y} y\, P_Y(y) \\ &= E[X]E[Y] \end{aligned}$$

We can prove the continuous case by following the same steps, with $\sum$ replaced by $\int$.

Covariance

Covariance

Definition

The covariance of X and Y is defined as:

$$\begin{aligned} Cov[X, Y] &= E[(X - E[X])(Y - E[Y])] \\ &= E[XY] - E[X]E[Y] \end{aligned}$$

Covariance

Definition

The covariance of X and Y is defined as:

$$\begin{aligned} Cov[X, Y] &= E[(X - E[X])(Y - E[Y])] \\ &= E[XY] - E[X]E[Y] \end{aligned}$$

- Covariance measures the linear association between two random variables

Covariance

Definition

The covariance of X and Y is defined as:

$$
\begin{aligned}
Cov[X, Y] &= E[(X - E[X])(Y - E[Y])] \\
&= E[XY] - E[X]E[Y]
\end{aligned}
$$

- Covariance measures the linear association between two random variables .
- If $Cov[X, Y] > 0$, observing an X value greater than $E[X]$ makes it more likely to also observe a Y value greater than $E[Y]$, and vice versa.

Covariance

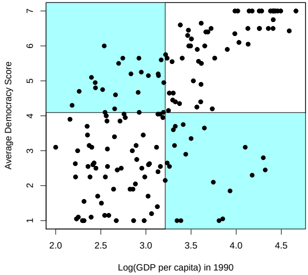
Definition

The covariance of X and Y is defined as:

$$Cov[X,Y] = E[(X - E[X])(Y - E[Y])] \\ = E[XY] - E[X]E[Y]$$

- Covariance measures the linear association between two random variables .
- If $Cov[X,Y] > 0$, observing an X value greater than $E[X]$ makes it more likely to also observe a Y value greater than $E[Y]$, and vice versa.

  - Points in upper right and lower left quadrants (relative to the means) add to the covariance.
  - Points in the upper left and lower right quadrants subtract from the covariance.



Covariance and Independence

Does $X \perp\!\!\!\perp Y$ imply $Cov[X, Y] = 0$?

Covariance and Independence

Does $X \perp\!\!\!\perp Y$ imply $Cov[X, Y] = 0$? Yes!

Covariance and Independence

Does $X \perp\!\!\!\perp Y$ imply $Cov[X, Y] = 0$? Yes!

Proof:

$$\begin{aligned} Cov[X,Y] &= E[XY] - E[X]E[Y] \\ &= \end{aligned}$$

Covariance and Independence

Does $X \perp\!\!\!\perp Y$ imply $Cov[X, Y] = 0$? Yes!

Proof:

$$
\begin{aligned}
Cov[X, Y] &= E[XY] - E[X]E[Y] \\
&= E[X]E[Y] - E[X]E[Y] \quad \text{(independence)} \\
&= 0.
\end{aligned}
$$

Covariance and Independence

Does $X \perp\!\!\!\perp Y$ imply $Cov[X, Y] = 0$? Yes!

Proof:

$$\begin{aligned} Cov[X, Y] &= E[XY] - E[X]E[Y] \\ &= E[X]E[Y] - E[X]E[Y] \quad \text{(independence)} \\ &= 0. \end{aligned}$$

Does $Cov[X, Y] = 0$ imply $X \perp\!\!\!\perp Y$?

Covariance and Independence

Does $X \perp\!\!\!\perp Y$ imply $Cov[X, Y] = 0$? Yes!

Proof:

$$\begin{aligned} Cov[X, Y] &= E[XY] - E[X]E[Y] \\ &= E[X]E[Y] - E[X]E[Y] \quad \text{(independence)} \\ &= 0. \end{aligned}$$

Does $Cov[X, Y] = 0$ imply $X \perp\!\!\!\perp Y$? No!

Counterexample: Suppose $X \in \{-1, 0, 1\}$ with $P_X(x) = 1/3$ and $Y = X^2$.

Is $X \perp\!\!\!\perp Y$?

Covariance and Independence

Does $X \perp\!\!\!\perp Y$ imply $Cov[X, Y] = 0$? Yes!

Proof:

$$\begin{aligned} Cov[X, Y] &= E[XY] - E[X]E[Y] \\ &= E[X]E[Y] - E[X]E[Y] \quad \text{(independence)} \\ &= 0. \end{aligned}$$

Does $Cov[X, Y] = 0$ imply $X \perp\!\!\!\perp Y$? No!

Counterexample: Suppose $X \in \{-1, 0, 1\}$ with $P_X(x) = 1/3$ and $Y = X^2$.

Is $X \perp\!\!\!\perp Y$? No, because $P_{Y|X}(y \mid x) \neq P_Y(y)$
(Learning about X gives meaningful information about Y.)

Covariance and Independence

Does $X \perp\!\!\!\perp Y$ imply $Cov[X, Y] = 0$? Yes!

Proof:

$$\begin{aligned} Cov[X, Y] &= E[XY] - E[X]E[Y] \\ &= E[X]E[Y] - E[X]E[Y] \quad \text{(independence)} \\ &= 0. \end{aligned}$$

Does $Cov[X, Y] = 0$ imply $X \perp\!\!\!\perp Y$? No!

Counterexample: Suppose $X \in \{-1, 0, 1\}$ with $P_X(x) = 1/3$ and $Y = X^2$.

Is $X \perp\!\!\!\perp Y$? No, because $P_{Y|X}(y \mid x) \neq P_Y(y)$
(Learning about X gives meaningful information about Y.)

What is $Cov[X, Y]$?

Covariance and Independence

Does $X \perp\!\!\!\perp Y$ imply $Cov[X, Y] = 0$? Yes!

Proof:

$$\begin{aligned} Cov[X, Y] &= E[XY] - E[X]E[Y] \\ &= E[X]E[Y] - E[X]E[Y] \quad \text{(independence)} \\ &= 0. \end{aligned}$$

Does $Cov[X, Y] = 0$ imply $X \perp\!\!\!\perp Y$? No!

Counterexample: Suppose $X \in \{-1, 0, 1\}$ with $P_X(x) = 1/3$ and $Y = X^2$.

Is $X \perp\!\!\!\perp Y$? No, because $P_{Y|X}(y \mid x) \neq P_Y(y)$
(Learning about X gives meaningful information about Y.)

What is $Cov[X, Y]$?

$$\begin{aligned} Cov[X, Y] &= E[XX^2] - E[X]E[X^2] = E[X^3] - E[X]E[X^2] \\ &= \end{aligned}$$

Covariance and Independence

Does $X \perp\!\!\!\perp Y$ imply $Cov[X,Y] = 0$? Yes!

Proof:

$$\begin{aligned} Cov[X,Y] &= E[XY] - E[X]E[Y] \\ &= E[X]E[Y] - E[X]E[Y] \quad \text{(independence)} \\ &= 0. \end{aligned}$$

Does $Cov[X,Y] = 0$ imply $X \perp\!\!\!\perp Y$? No!

Counterexample: Suppose $X \in \{-1, 0, 1\}$ with $P_X(x) = 1/3$ and $Y = X^2$.

Is $X \perp\!\!\!\perp Y$? No, because $P_{Y|X}(y \mid x) \neq P_Y(y)$
(Learning about X gives meaningful information about Y.)

What is $Cov[X,Y]$?

$$\begin{aligned} Cov[X,Y] &= E[XX^2] - E[X]E[X^2] = E[X^3] - E[X]E[X^2] \\ &= E[X] - E[X]E[X^2] = \end{aligned}$$

Covariance and Independence

Does $X \perp\!\!\!\perp Y$ imply $Cov[X, Y] = 0$? Yes!

Proof:

$$\begin{aligned} Cov[X, Y] &= E[XY] - E[X]E[Y] \\ &= E[X]E[Y] - E[X]E[Y] \quad \text{(independence)} \\ &= 0. \end{aligned}$$

Does $Cov[X, Y] = 0$ imply $X \perp\!\!\!\perp Y$? No!

Counterexample: Suppose $X \in \{-1, 0, 1\}$ with $P_X(x) = 1/3$ and $Y = X^2$.

Is $X \perp\!\!\!\perp Y$? No, because $P_{Y|X}(y \mid x) \neq P_Y(y)$
(Learning about X gives meaningful information about Y.)

What is $Cov[X, Y]$?

$$\begin{aligned} Cov[X, Y] &= E[XX^2] - E[X]E[X^2] = E[X^3] - E[X]E[X^2] \\ &= E[X] - E[X]E[X^2] = 0 - 0 \cdot E[X^2] = 0. \end{aligned}$$

Covariance and Independence

Does $X \perp\!\!\!\perp Y$ imply $Cov[X, Y] = 0$? Yes!

Proof:

$$\begin{aligned} Cov[X, Y] &= E[XY] - E[X]E[Y] \\ &= E[X]E[Y] - E[X]E[Y] \quad \text{(independence)} \\ &= 0. \end{aligned}$$

Does $Cov[X, Y] = 0$ imply $X \perp\!\!\!\perp Y$? No!

Counterexample: Suppose $X \in \{-1, 0, 1\}$ with $P_X(x) = 1/3$ and $Y = X^2$.

Is $X \perp\!\!\!\perp Y$? No, because $P_{Y|X}(y \mid x) \neq P_Y(y)$
(Learning about X gives meaningful information about Y.)

What is $Cov[X, Y]$?

$$\begin{aligned} Cov[X, Y] &= E[XX^2] - E[X]E[X^2] = E[X^3] - E[X]E[X^2] \\ &= E[X] - E[X]E[X^2] = 0 - 0 \cdot E[X^2] = 0. \end{aligned}$$

Therefore, $X \perp\!\!\!\perp Y \implies Cov[X, Y] = 0$, but not vice versa.

Important Identities for Variances and Covariances

Important Identities for Variances and Covariances

1. For random variables X and Y and constants a, b and c,

$$V[aX + bY + c] = a^2 V[X] + b^2 V[Y] + 2ab\, Cov[X, Y]$$

Important Identities for Variances and Covariances

1. For random variables X and Y and constants a, b and c,

$$V[aX + bY + c] = a^2 V[X] + b^2 V[Y] + 2ab\, Cov[X, Y]$$

2. Important special cases:

$$V[X + Y] = V[X] + V[Y] + 2Cov[X, Y]$$
$$V[X - Y] = V[X] + V[Y] - 2Cov[X, Y]$$

Important Identities for Variances and Covariances

1. For random variables X and Y and constants a, b and c,

$$V[aX + bY + c] = a^2 V[X] + b^2 V[Y] + 2ab\, Cov[X, Y]$$

2. Important special cases:

$$V[X + Y] = V[X] + V[Y] + 2Cov[X, Y]$$
$$V[X - Y] = V[X] + V[Y] - 2Cov[X, Y]$$

3. Furthermore, if X and Y are independent,

$$V[X \pm Y] = V[X] + V[Y]$$

Important Identities for Variances and Covariances

1. For random variables X and Y and constants a, b and c,

$$V[aX + bY + c] = a^2 V[X] + b^2 V[Y] + 2ab\, Cov[X, Y]$$

2. Important special cases:

$$V[X + Y] = V[X] + V[Y] + 2Cov[X, Y]$$
$$V[X - Y] = V[X] + V[Y] - 2Cov[X, Y]$$

3. Furthermore, if X and Y are independent,

$$V[X \pm Y] = V[X] + V[Y]$$

Proof: Plug in to the definition of variance and expand (try it yourself!)

Correlation

Correlation

- $Cov[X, Y]$ depends not only on the strength of (linear) association between X and Y, but also the scale of X and Y.

Correlation

- $Cov[X, Y]$ depends not only on the strength of (linear) association between X and Y, but also the scale of X and Y.
- Can we have a pure measure of association that is scale-independent?

Correlation

- $Cov[X, Y]$ depends not only on the strength of (linear) association between X and Y, but also the scale of X and Y.
- Can we have a pure measure of association that is scale-independent?

Definition (Correlation)

The correlation between two random variables X and Y is defined as

$$Cor[X, Y] = \frac{Cov[X, Y]}{\sqrt{V[X]V[Y]}} = \frac{Cov[X, Y]}{SD[X]SD[Y]}.$$

Correlation

- $Cov[X, Y]$ depends not only on the strength of (linear) association between X and Y, but also the scale of X and Y.
- Can we have a pure measure of association that is scale-independent?

Definition (Correlation)

The correlation between two random variables X and Y is defined as

$$Cor[X, Y] = \frac{Cov[X, Y]}{\sqrt{V[X]V[Y]}} = \frac{Cov[X, Y]}{SD[X]SD[Y]}.$$

- $Cor[X, Y]$ is a standardized measure of linear association between X and Y.

Correlation

- $Cov[X, Y]$ depends not only on the strength of (linear) association between X and Y, but also the scale of X and Y.
- Can we have a pure measure of association that is scale-independent?

Definition (Correlation)

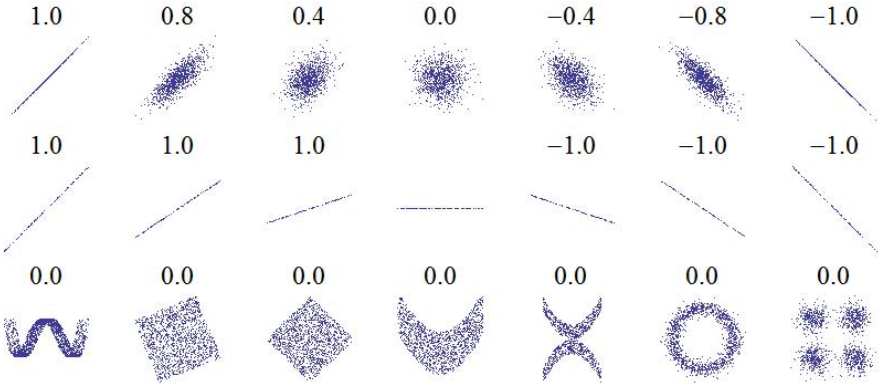
The correlation between two random variables X and Y is defined as

$$Cor[X, Y] = \frac{Cov[X, Y]}{\sqrt{V[X]V[Y]}} = \frac{Cov[X, Y]}{SD[X]SD[Y]}.$$

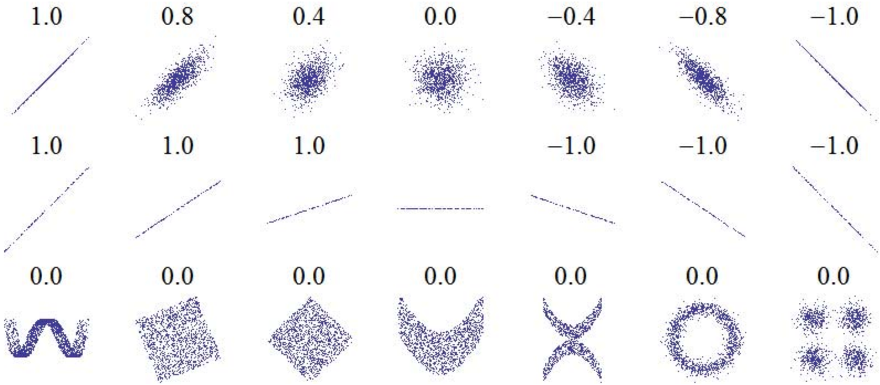
- $Cor[X, Y]$ is a standardized measure of linear association between X and Y.
- Always satisfies: $-1 \leq Cor[X, Y] \leq 1$.

Correlation is *Linear*

Correlation is *Linear*

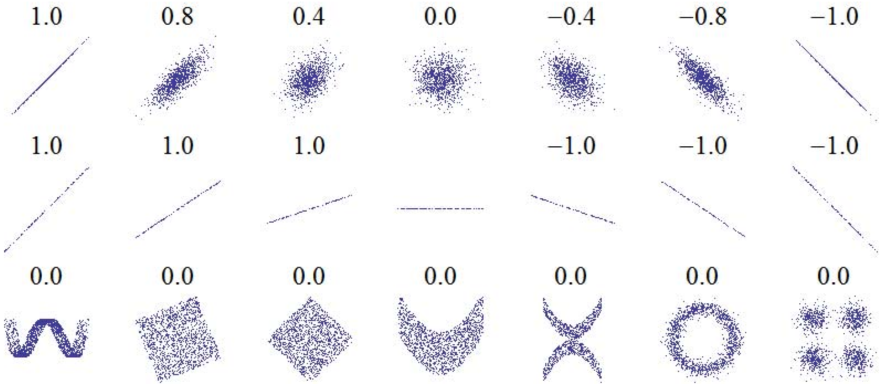


Correlation is *Linear*



- $Cor[X, Y] = \pm 1$ iff $Y = aX + b$ where $a \neq 0$.

Correlation is *Linear*



- $Cor[X, Y] = \pm 1$ iff $Y = aX + b$ where $a \neq 0$.
- Like covariance, correlation measures the linear association between X and Y.

Conditional Independence

Conditional Independence

Definition (Conditional Independence of Random Variables)

Random variables Y and X are conditionally independent given Z iff

$$f_{X,Y|Z}(x,y|z) \;=\; f_{Y|Z}(y|z) \cdot f_{X|Z}(x|z)$$

for all x, y, and z. This is often written as $Y \perp\!\!\!\perp X \mid Z$.

Conditional Independence

Definition (Conditional Independence of Random Variables)

Random variables Y and X are conditionally independent given Z iff

$$f_{X,Y|Z}(x,y|z) \;=\; f_{Y|Z}(y|z) \cdot f_{X|Z}(x|z)$$

for all x, y, and z. This is often written as $Y \perp\!\!\!\perp X \mid Z$.

- Can also be written as

$$f_{Y|X,Z}(y \mid x,z) \;=\; f_{Y|Z}(y \mid z)$$

Conditional Independence

Definition (Conditional Independence of Random Variables)

Random variables Y and X are conditionally independent given Z iff

$$f_{X,Y|Z}(x, y|z) \;=\; f_{Y|Z}(y|z) \cdot f_{X|Z}(x|z)$$

for all x, y, and z. This is often written as $Y \perp\!\!\!\perp X \mid Z$.

- Can also be written as

$$f_{Y|X,Z}(y \mid x, z) \;=\; f_{Y|Z}(y \mid z)$$

- Interpretation: Once we know Z, X contains no meaningful information about likely values of Y.
(Z has all the information about Y contained in X, if any.)

Conditional Independence

Definition (Conditional Independence of Random Variables)

Random variables Y and X are conditionally independent given Z iff

$$f_{X,Y|Z}(x,y|z) \; = \; f_{Y|Z}(y|z) \cdot f_{X|Z}(x|z)$$

for all x, y, and z. This is often written as $Y \perp\!\!\!\perp X \mid Z$.

- Can also be written as

$$f_{Y|X,Z}(y \mid x,z) \; = \; f_{Y|Z}(y \mid z)$$

- Interpretation: Once we know Z, X contains no meaningful information about likely values of Y.
  (Z has all the information about Y contained in X, if any.)
- $Y \perp\!\!\!\perp X \mid Z$ implies

$$E[Y|X=x, Z=z] =$$

Conditional Independence

Definition (Conditional Independence of Random Variables)

Random variables Y and X are conditionally independent given Z iff

$$f_{X,Y|Z}(x, y|z) \; = \; f_{Y|Z}(y|z) \cdot f_{X|Z}(x|z)$$

for all x, y, and z. This is often written as $Y \perp\!\!\!\perp X \mid Z$.

- Can also be written as

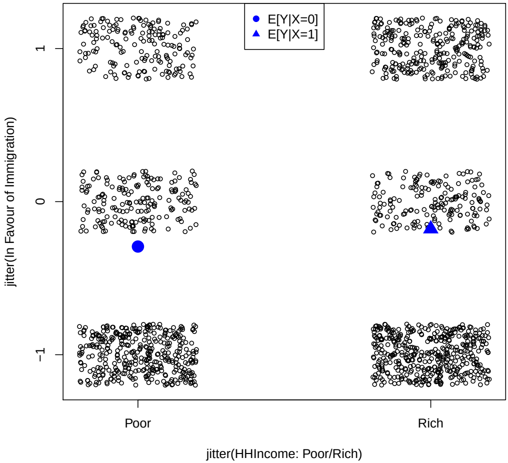
$$f_{Y|X,Z}(y \mid x, z) \; = \; f_{Y|Z}(y \mid z)$$

- Interpretation: Once we know Z, X contains no meaningful information about likely values of Y.
  (Z has all the information about Y contained in X, if any.)
- $Y \perp\!\!\!\perp X \mid Z$ implies

$$E[Y|X = x, Z = z] = E[Y|Z = z].$$

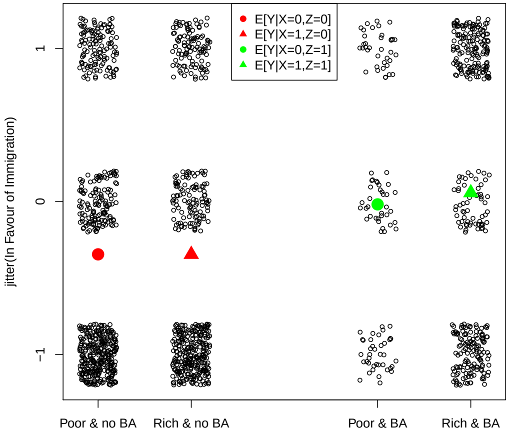
Is $Y \perp\!\!\!\perp X$?

Example: X = wealth, Y = support for immigration, Z = education.



Is $Y \perp\!\!\!\perp X | Z$?

Example: X = wealth, Y = support for immigration, Z = education.



1. Random Variables and Distributions
    - What is a Random Variable?
    - Discrete Distributions
    - Continuous Distributions
2. Characteristics of Distributions
    - Central Tendency
    - Measures of Dispersion
3. Conditional Distributions
4. Fun with Averages
5. Fun with Sensitive Questions
6. Appendix: Why the Mean?
7. Joint Distributions
    - Discrete Random Variable
    - Continuous Random Variable
8. Conditional Expectation
9. Properties
    - Independence
    - Covariance and Correlation
    - Conditional Independence
10. Famous Distributions
11. Fun With Spam

1. Random Variables and Distributions
   - What is a Random Variable?
   - Discrete Distributions
   - Continuous Distributions
2. Characteristics of Distributions
   - Central Tendency
   - Measures of Dispersion
3. Conditional Distributions
4. Fun with Averages
5. Fun with Sensitive Questions
6. Appendix: Why the Mean?
7. Joint Distributions
   - Discrete Random Variable
   - Continuous Random Variable
8. Conditional Expectation
9. Properties
   - Independence
   - Covariance and Correlation
   - Conditional Independence
10. **Famous Distributions**
11. Fun With Spam

Distributions

Distributions

- We like random variables because they take complex real world phenomena and represent them with a common mathematical infrastructure

Distributions

- We like random variables because they take complex real world phenomena and represent them with a common mathematical infrastructure
- We can work with arbitrary pmf/pdfs but we will often work with particular families of distributions

Distributions

- We like random variables because they take complex real world phenomena and represent them with a common mathematical infrastructure
- We can work with arbitrary pmf/pdfs but we will often work with particular families of distributions
  - members of the same family have similar forms determined by parameters

Distributions

- We like random variables because they take complex real world phenomena and represent them with a common mathematical infrastructure
- We can work with arbitrary pmf/pdfs but we will often work with particular families of distributions
    - members of the same family have similar forms determined by parameters
    - the parameters determine the shape of the distribution

Distributions

- We like random variables because they take complex real world phenomena and represent them with a common mathematical infrastructure
- We can work with arbitrary pmf/pdfs but we will often work with particular families of distributions
  - members of the same family have similar forms determined by parameters
  - the parameters determine the shape of the distribution
- When we can work with an existing set of distributions, it makes calculations simpler

Distributions

- We like random variables because they take complex real world phenomena and represent them with a common mathematical infrastructure
- We can work with arbitrary pmf/pdfs but we will often work with particular families of distributions
  - members of the same family have similar forms determined by parameters
  - the parameters determine the shape of the distribution
- When we can work with an existing set of distributions, it makes calculations simpler
- Examples: Bernoulli, Binomial, Gamma, Normal, Poisson, t-distribution

Bernoulli Random Variable

Definition

Suppose X is a random variable, with $X \in \{0, 1\}$ and $P(X = 1) = \pi$.
Then we will say that X is Bernoulli random variable,

$$p(X = x) \quad = \quad \pi^x (1 - \pi)^{1-x}$$

for $x \in \{0, 1\}$ and $p(X = x) = 0$ otherwise.
We will (equivalently) say that

$$X \quad \sim \quad \text{Bernoulli}(\pi)$$

$\sim$ means equality in distribution (not values!). Often $X \sim$ Bernoulli(π) would be read 'X is distributed Bernoulli with parameter π'

Bernoulli Random Variable Mean and Variance

Suppose $X \sim \text{Bernoulli}(\pi)$

Bernoulli Random Variable Mean and Variance

Suppose $X \sim \text{Bernoulli}(\pi)$

$$\begin{aligned} E[X] &= 1 \times P(X = 1) + 0 \times P(X = 0) \\ &= \pi + 0(1 - \pi) = \pi \end{aligned}$$

$E[X] = \pi$

Bernoulli Random Variable Mean and Variance

Suppose $X \sim \text{Bernoulli}(\pi)$

$$\begin{aligned} E[X] &= 1 \times P(X=1) + 0 \times P(X=0) \\ &= \pi + 0(1-\pi) = \pi \\ \text{var}(X) &= E[X^2] - E[X]^2 \end{aligned}$$

$E[X] = \pi$

Bernoulli Random Variable Mean and Variance

Suppose $X \sim \text{Bernoulli}(\pi)$

$$\begin{aligned} E[X] &= 1 \times P(X=1) + 0 \times P(X=0) \\ &= \pi + 0(1-\pi) = \pi \\ \text{var}(X) &= E[X^2] - E[X]^2 \\ E[X^2] &= 1^2 P(X=1) + 0^2 P(X=0) \end{aligned}$$

$E[X] = \pi$

Bernoulli Random Variable Mean and Variance

Suppose $X \sim \text{Bernoulli}(\pi)$

$$\begin{aligned}
E[X] &= 1 \times P(X = 1) + 0 \times P(X = 0) \\
&= \pi + 0(1 - \pi) = \pi \\
\text{var}(X) &= E[X^2] - E[X]^2 \\
E[X^2] &= 1^2 P(X = 1) + 0^2 P(X = 0) \\
&= \pi
\end{aligned}$$

$E[X] = \pi$

Bernoulli Random Variable Mean and Variance

Suppose $X \sim \text{Bernoulli}(\pi)$

$$\begin{aligned} E[X] &= 1 \times P(X = 1) + 0 \times P(X = 0) \\ &= \pi + 0(1 - \pi) = \pi \\ \text{var}(X) &= E[X^2] - E[X]^2 \\ E[X^2] &= 1^2 P(X = 1) + 0^2 P(X = 0) \\ &= \pi \\ \text{var}(X) &= \pi - \pi^2 \end{aligned}$$

$E[X] = \pi$

Bernoulli Random Variable Mean and Variance

Suppose $X \sim \text{Bernoulli}(\pi)$

$$\begin{aligned} E[X] &= 1 \times P(X=1) + 0 \times P(X=0) \\ &= \pi + 0(1-\pi) = \pi \\ \text{var}(X) &= E[X^2] - E[X]^2 \\ E[X^2] &= 1^2 P(X=1) + 0^2 P(X=0) \\ &= \pi \\ \text{var}(X) &= \pi - \pi^2 \\ &= \pi(1-\pi) \end{aligned}$$

$E[X] = \pi$
$\text{var}(X) = \pi(1-\pi)$

Bernoulli Random Variable Mean and Variance

Suppose $X \sim \text{Bernoulli}(\pi)$

$$\begin{aligned}
E[X] &= 1 \times P(X = 1) + 0 \times P(X = 0) \\
&= \pi + 0(1 - \pi) = \pi \\
\text{var}(X) &= E[X^2] - E[X]^2 \\
E[X^2] &= 1^2 P(X = 1) + 0^2 P(X = 0) \\
&= \pi \\
\text{var}(X) &= \pi - \pi^2 \\
&= \pi(1 - \pi)
\end{aligned}$$

$E[X] = \pi$
$\text{var}(X) = \pi(1 - \pi)$
Importantly, we can also just look this up!

Normal/Gaussian Random Variables

Definition

Suppose X is a random variable with $X \in \mathbb{R}$ and density

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

Then X is a normally distributed random variable with parameters μ and σ^2.
Equivalently, we'll write

$$X \sim \text{Normal}(\mu, \sigma^2)$$

Expected Value/Variance of Normal Distribution

Z is a standard normal distribution if

Expected Value/Variance of Normal Distribution

Z is a standard normal distribution if

$$Z \sim \text{Normal}(0, 1)$$

Expected Value/Variance of Normal Distribution

Z is a standard normal distribution if

$$Z \sim \text{Normal}(0, 1)$$

We'll call the cumulative distribution function of Z,

Expected Value/Variance of Normal Distribution

Z is a standard normal distribution if

$$Z \sim \text{Normal}(0, 1)$$

We'll call the cumulative distribution function of Z,

$$F_Z(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp(-z^2/2) dz$$

Expected Value/Variance of Normal Distribution

Z is a standard normal distribution if

$$Z \sim \text{Normal}(0, 1)$$

We'll call the cumulative distribution function of Z,

$$F_Z(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp(-z^2/2) dz$$

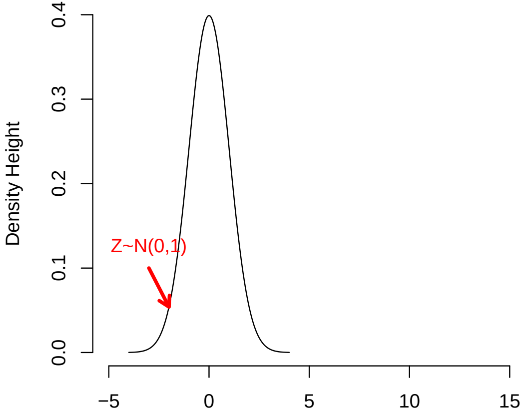
Proposition

Scale/Location. If $Z \sim N(0, 1)$, then $X = aZ + b$ is,

$$X \sim Normal(b, a^2)$$

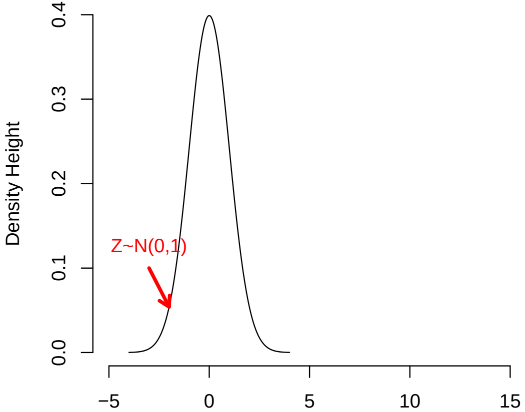
Intuition

Suppose $Z \sim \text{Normal}(0, 1)$.



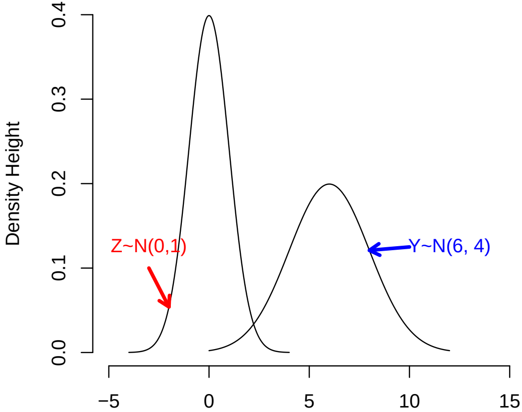
Intuition

Suppose $Z \sim \text{Normal}(0, 1)$.
$Y = 2Z + 6$



Intuition

Suppose $Z \sim \text{Normal}(0,1)$.
$Y = 2Z + 6$
$Y \sim \text{Normal}(6,4)$



Proof: $Z \sim N(0, 1)$ and $Y = aZ + b$, then $Y \sim N(b, a^2)$

To prove

Proof: $Z \sim N(0, 1)$ and $Y = aZ + b$, then $Y \sim N(b, a^2)$

To prove we need to show that density for Y is a normal distribution.

Proof: $Z \sim N(0,1)$ and $Y = aZ + b$, then $Y \sim N(b, a^2)$

To prove we need to show that density for Y is a normal distribution.
That is, we'll show $F_Y(x)$ is Normal cdf.

Proof: $Z \sim N(0, 1)$ and $Y = aZ + b$, then $Y \sim N(b, a^2)$

To prove we need to show that density for Y is a normal distribution.
That is, we'll show $F_Y(x)$ is Normal cdf.
Call $F_Z(x)$ cdf for standardized normal.

Proof: $Z \sim N(0, 1)$ and $Y = aZ + b$, then $Y \sim N(b, a^2)$

To prove we need to show that density for Y is a normal distribution.
That is, we'll show $F_Y(x)$ is Normal cdf.
Call $F_Z(x)$ cdf for standardized normal.

$$F_Y(x) = P(Y \leq x)$$

Proof: $Z \sim N(0,1)$ and $Y = aZ + b$, then $Y \sim N(b, a^2)$

To prove we need to show that density for Y is a normal distribution.
That is, we'll show $F_Y(x)$ is Normal cdf.
Call $F_Z(x)$ cdf for standardized normal.

$$\begin{aligned} F_Y(x) &= P(Y \leq x) \\ &= P(aZ + b \leq x) \end{aligned}$$

Proof: $Z \sim N(0,1)$ and $Y = aZ + b$, then $Y \sim N(b, a^2)$

To prove we need to show that density for Y is a normal distribution.
That is, we'll show $F_Y(x)$ is Normal cdf.
Call $F_Z(x)$ cdf for standardized normal.

$$\begin{aligned} F_Y(x) &= P(Y \leq x) \\ &= P(aZ + b \leq x) \\ &= P(Z \leq \left[\frac{x-b}{a}\right]) \end{aligned}$$

Proof: $Z \sim N(0,1)$ and $Y = aZ + b$, then $Y \sim N(b, a^2)$

To prove we need to show that density for Y is a normal distribution.
That is, we'll show $F_Y(x)$ is Normal cdf.
Call $F_Z(x)$ cdf for standardized normal.

$$
\begin{aligned}
F_Y(x) &= P(Y \leq x) \\
&= P(aZ + b \leq x) \\
&= P(Z \leq \left[\frac{x-b}{a}\right]) \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{x-b}{a}} \exp(-\frac{z^2}{2}) dz
\end{aligned}
$$

Proof: $Z \sim N(0,1)$ and $Y = aZ + b$, then $Y \sim N(b, a^2)$

To prove we need to show that density for Y is a normal distribution.
That is, we'll show $F_Y(x)$ is Normal cdf.
Call $F_Z(x)$ cdf for standardized normal.

$$
\begin{aligned}
F_Y(x) &= P(Y \leq x) \\
&= P(aZ + b \leq x) \\
&= P(Z \leq \left[\frac{x-b}{a}\right]) \\
&= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{x-b}{a}} \exp(-\frac{z^2}{2}) dz \\
&= F_Z(\frac{x-b}{a})
\end{aligned}
$$

Proof: $Z \sim N(0,1)$ and $Y = aZ + b$, then $Y \sim N(b, a^2)$

So, we can work with $F_Z(\frac{x-b}{a})$.

Proof: $Z \sim N(0,1)$ and $Y = aZ + b$, then $Y \sim N(b, a^2)$

So, we can work with $F_Z(\frac{x-b}{a})$.

$$\frac{\partial F_Y(x)}{\partial x} = \frac{\partial F_Z(\frac{x-b}{a})}{\partial x}$$

Proof: $Z \sim N(0,1)$ and $Y = aZ + b$, then $Y \sim N(b, a^2)$

So, we can work with $F_Z(\frac{x-b}{a})$.

$$\begin{aligned} \frac{\partial F_Y(x)}{\partial x} &= \frac{\partial F_Z(\frac{x-b}{a})}{\partial x} \\ &= f_Z(\frac{x-b}{a})\frac{1}{a} \text{ By the chain rule} \end{aligned}$$

Proof: $Z \sim N(0,1)$ and $Y = aZ + b$, then $Y \sim N(b, a^2)$

So, we can work with $F_Z(\frac{x-b}{a})$.

$$\begin{aligned}
\frac{\partial F_Y(x)}{\partial x} &= \frac{\partial F_Z(\frac{x-b}{a})}{\partial x} \\
&= f_Z(\frac{x-b}{a})\frac{1}{a} \text{ By the chain rule} \\
&= \frac{1}{\sqrt{2\pi}a}\exp\left[-\frac{\left(\frac{x-b}{a}\right)^2}{2}\right] \text{ By definition of } f_Z(x) \text{ or FTC}
\end{aligned}$$

Proof: $Z \sim N(0,1)$ and $Y = aZ + b$, then $Y \sim N(b, a^2)$

So, we can work with $F_Z(\frac{x-b}{a})$.

$$\begin{aligned}
\frac{\partial F_Y(x)}{\partial x} &= \frac{\partial F_Z(\frac{x-b}{a})}{\partial x} \\
&= f_Z(\frac{x-b}{a})\frac{1}{a} \text{ By the chain rule} \\
&= \frac{1}{\sqrt{2\pi}a}\exp\left[-\frac{\left(\frac{x-b}{a}\right)^2}{2}\right] \text{ By definition of } f_Z(x) \text{ or FTC} \\
&= \frac{1}{\sqrt{2\pi}a}\exp\left[-\frac{(x-b)^2}{2a^2}\right]
\end{aligned}$$

Proof: $Z \sim N(0,1)$ and $Y = aZ + b$, then $Y \sim N(b, a^2)$

So, we can work with $F_Z(\frac{x-b}{a})$.

$$\begin{aligned}
\frac{\partial F_Y(x)}{\partial x} &= \frac{\partial F_Z(\frac{x-b}{a})}{\partial x} \\
&= f_Z(\frac{x-b}{a})\frac{1}{a} \text{ By the chain rule} \\
&= \frac{1}{\sqrt{2\pi}a}\exp\left[-\frac{\left(\frac{x-b}{a}\right)^2}{2}\right] \text{ By definition of } f_Z(x) \text{ or FTC} \\
&= \frac{1}{\sqrt{2\pi}a}\exp\left[-\frac{(x-b)^2}{2a^2}\right] \\
&= \text{Normal}(b, a^2)
\end{aligned}$$

Expectation and Variance

Assume we know:

$$
\begin{aligned}
E[Z] &= 0 \\
Var(Z) &= 1
\end{aligned}
$$

Expectation and Variance

Assume we know:

$$
\begin{aligned}
E[Z] &= 0 \\
Var(Z) &= 1
\end{aligned}
$$

This implies that, for $Y \sim \text{Normal}(\mu, \sigma^2)$

Expectation and Variance

Assume we know:

$$
\begin{aligned}
E[Z] &= 0 \\
Var(Z) &= 1
\end{aligned}
$$

This implies that, for $Y \sim \text{Normal}(\mu, \sigma^2)$

$$
E[Y] = E[\sigma Z + \mu]
$$

Expectation and Variance

Assume we know:

$$\begin{aligned} E[Z] &= 0 \\ Var(Z) &= 1 \end{aligned}$$

This implies that, for $Y \sim \text{Normal}(\mu, \sigma^2)$

$$\begin{aligned} E[Y] &= E[\sigma Z + \mu] \\ &= \sigma E[Z] + \mu \end{aligned}$$

Expectation and Variance

Assume we know:

$$\begin{aligned} E[Z] &= 0 \\ Var(Z) &= 1 \end{aligned}$$

This implies that, for $Y \sim \text{Normal}(\mu, \sigma^2)$

$$\begin{aligned} E[Y] &= E[\sigma Z + \mu] \\ &= \sigma E[Z] + \mu \\ &= \mu \end{aligned}$$

Expectation and Variance

Assume we know:

$$\begin{aligned} E[Z] &= 0 \\ Var(Z) &= 1 \end{aligned}$$

This implies that, for $Y \sim \text{Normal}(\mu, \sigma^2)$

$$\begin{aligned} E[Y] &= E[\sigma Z + \mu] \\ &= \sigma E[Z] + \mu \\ &= \mu \\ Var(Y) &= Var(\sigma Z + \mu) \end{aligned}$$

Expectation and Variance

Assume we know:

$$\begin{aligned} E[Z] &= 0 \\ Var(Z) &= 1 \end{aligned}$$

This implies that, for $Y \sim \text{Normal}(\mu, \sigma^2)$

$$\begin{aligned} E[Y] &= E[\sigma Z + \mu] \\ &= \sigma E[Z] + \mu \\ &= \mu \\ Var(Y) &= Var(\sigma Z + \mu) \\ &= \sigma^2 Var(Z) + Var(\mu) \end{aligned}$$

Expectation and Variance

Assume we know:

$$\begin{aligned} E[Z] &= 0 \\ Var(Z) &= 1 \end{aligned}$$

This implies that, for $Y \sim \text{Normal}(\mu, \sigma^2)$

$$\begin{aligned} E[Y] &= E[\sigma Z + \mu] \\ &= \sigma E[Z] + \mu \\ &= \mu \\ Var(Y) &= Var(\sigma Z + \mu) \\ &= \sigma^2 Var(Z) + Var(\mu) \\ &= \sigma^2 + 0 \end{aligned}$$

Expectation and Variance

Assume we know:

$$\begin{aligned} E[Z] &= 0 \\ Var(Z) &= 1 \end{aligned}$$

This implies that, for $Y \sim \text{Normal}(\mu, \sigma^2)$

$$\begin{aligned} E[Y] &= E[\sigma Z + \mu] \\ &= \sigma E[Z] + \mu \\ &= \mu \\ Var(Y) &= Var(\sigma Z + \mu) \\ &= \sigma^2 Var(Z) + Var(\mu) \\ &= \sigma^2 + 0 \\ &= \sigma^2 \end{aligned}$$

Multivariate Normal

Definition

Suppose $\boldsymbol{X} = (X_1, X_2, \ldots, X_N)$ is a vector of random variables. If $\boldsymbol{X}$ has pdf

$$f(\boldsymbol{x}) \quad = \quad (2\pi)^{-N/2} \det(\boldsymbol{\Sigma})^{-1/2} \exp\left(-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu})' \boldsymbol{\Sigma}(\boldsymbol{x} - \boldsymbol{\mu})\right)$$

Then we will say $\boldsymbol{X}$ has a Multivariate Normal Distribution,

$$\boldsymbol{X} \quad \sim \quad \text{Multivariate Normal}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$$

Multivariate Normal Distribution

Consider the (bivariate) special case where $\boldsymbol{\mu} = (0, 0)$ and

$$\boldsymbol{\Sigma} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

Multivariate Normal Distribution

Consider the (bivariate) special case where $\boldsymbol{\mu} = (0, 0)$ and

$$\boldsymbol{\Sigma} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

Then

$$f(x_1, x_2) = (2\pi)^{-2/2} 1^{-1/2} \exp\left(-\frac{1}{2}\left((\boldsymbol{x} - \mathbf{0})' \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} (\boldsymbol{x} - \mathbf{0})\right)\right)$$

Multivariate Normal Distribution

Consider the (bivariate) special case where $\boldsymbol{\mu} = (0, 0)$ and

$$\boldsymbol{\Sigma} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

Then

$$\begin{aligned} f(x_1, x_2) &= (2\pi)^{-2/2} 1^{-1/2} \exp\left(-\frac{1}{2}\left((\boldsymbol{x} - \mathbf{0})' \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} (\boldsymbol{x} - \mathbf{0})\right)\right) \\ &= \frac{1}{2\pi} \exp\left(-\frac{1}{2}(x_1^2 + x_2^2)\right) \end{aligned}$$

Multivariate Normal Distribution

Consider the (bivariate) special case where $\boldsymbol{\mu} = (0, 0)$ and

$$\boldsymbol{\Sigma} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

Then

$$\begin{aligned} f(x_1, x_2) &= (2\pi)^{-2/2} 1^{-1/2} \exp\left(-\frac{1}{2}\left((\boldsymbol{x} - \mathbf{0})' \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} (\boldsymbol{x} - \mathbf{0})\right)\right) \\ &= \frac{1}{2\pi} \exp\left(-\frac{1}{2}(x_1^2 + x_2^2)\right) \\ &= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x_1^2}{2}\right) \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x_2^2}{2}\right) \end{aligned}$$

Multivariate Normal Distribution

Consider the (bivariate) special case where $\boldsymbol{\mu} = (0, 0)$ and

$$\boldsymbol{\Sigma} = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

Then

$$\begin{aligned} f(x_1, x_2) &= (2\pi)^{-2/2} 1^{-1/2} \exp\left(-\frac{1}{2}\left((\boldsymbol{x} - \boldsymbol{0})' \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} (\boldsymbol{x} - \boldsymbol{0})\right)\right) \\ &= \frac{1}{2\pi} \exp\left(-\frac{1}{2}(x_1^2 + x_2^2)\right) \\ &= \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x_1^2}{2}\right) \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x_2^2}{2}\right) \end{aligned}$$

$\rightsquigarrow$ product of univariate standard normally distributed random variables

Properties of the Multivariate Normal Distribution

Suppose $\boldsymbol{X} = (X_1, X_2, \ldots, X_N)$

$$
\begin{aligned}
E[\boldsymbol{X}] &= \boldsymbol{\mu} \\
\text{cov}(\boldsymbol{X}) &= \boldsymbol{\Sigma}
\end{aligned}
$$

So that,

$$
\boldsymbol{\Sigma} = \begin{pmatrix}
\text{var}(X_1) & \text{cov}(X_1, X_2) & \ldots & \text{cov}(X_1, X_N) \\
\text{cov}(X_2, X_1) & \text{var}(X_2) & \ldots & \text{cov}(X_2, X_N) \\
\vdots & \vdots & \ddots & \vdots \\
\text{cov}(X_N, X_1) & \text{cov}(X_N, X_2) & \ldots & \text{var}(X_N)
\end{pmatrix}
$$

One Step Deeper: Exponential Family

Nearly every distribution we will discuss is in the exponential family. An exponential family distribution has the density of the following form:

$$f_Y(y;\theta,\phi) = \exp\left\{\frac{y\theta - b(\theta)}{a(\phi)} + c(y,\phi)\right\}$$

One Step Deeper: Exponential Family

Nearly every distribution we will discuss is in the exponential family. An exponential family distribution has the density of the following form:

$$f_Y(y;\theta,\phi) = \exp\left\{\frac{y\theta - b(\theta)}{a(\phi)} + c(y,\phi)\right\}$$

Example: Poisson(μ):

$$\Pr(Y_i = y \mid \mu) \ = \ \exp\{y \log \mu - \exp(\log \mu) - \log y!\}$$

One Step Deeper: Exponential Family

Nearly every distribution we will discuss is in the exponential family. An exponential family distribution has the density of the following form:

$$f_Y(y;\theta,\phi) = \exp\left\{\frac{y\theta - b(\theta)}{a(\phi)} + c(y,\phi)\right\}$$

Example: Poisson(μ):

$$\Pr(Y_i = y \mid \mu) \;=\; \exp\{y \log\mu - \exp(\log\mu) - \log y!\}$$

One Step Deeper: Exponential Family

Nearly every distribution we will discuss is in the exponential family. An exponential family distribution has the density of the following form:

$$f_Y(y;\theta,\phi) = \exp\left\{\frac{y\theta - b(\theta)}{a(\phi)} + c(y,\phi)\right\}$$

Example: Poisson(μ):

$$\Pr(Y_i = y \mid \mu) \;=\; \exp\left\{y \log\mu - \exp(\log\mu) - \log y!\right\}$$

One Step Deeper: Exponential Family

Nearly every distribution we will discuss is in the exponential family. An exponential family distribution has the density of the following form:

$$f_Y(y;\theta,\phi) = \exp\left\{\frac{y\theta - b(\theta)}{a(\phi)} + c(y,\phi)\right\}$$

Example: Poisson(μ):

$$\Pr(Y_i = y \mid \mu) \; = \; \exp\{y\log\mu - \exp(\log\mu) - \log y!\}$$

$\Longrightarrow \theta = \log\mu$, $\phi = 1$, $a(\phi) = \phi, b(\theta) = \exp(\theta)$, and $c = -\log y!$

One Step Deeper: Exponential Family

Nearly every distribution we will discuss is in the exponential family. An exponential family distribution has the density of the following form:

$$f_Y(y;\theta,\phi) = \exp\left\{\frac{y\theta - b(\theta)}{a(\phi)} + c(y,\phi)\right\}$$

Example: Poisson(μ):

$$\Pr(Y_i = y \mid \mu) \;=\; \exp\left\{y \log \mu - \exp(\log \mu) - \log y!\right\}$$

$\Longrightarrow \theta = \log \mu$, $\phi = 1$, $a(\phi) = \phi, b(\theta) = \exp(\theta)$, and $c = -\log y!$

Many other examples, including: Normal, Bernoulli/binomial, Gamma, multinomial, exponential, negative binomial, beta, uniform, chi-squared, etc.

This slide and the following based on material from Teppei Yamamoto

One Step Deeper: Properties of the Exponential Family

- Mean is a function of θ and given by

$$\mathbb{E}(Y) \equiv \mu = b'(\theta)$$

One Step Deeper: Properties of the Exponential Family

- Mean is a function of θ and given by

$$\mathbb{E}(Y) \equiv \mu = b'(\theta)$$

- Variance is a function of θ and ϕ and given by

$$\mathbb{V}(Y) \equiv V = b''(\theta)a(\phi)$$

One Step Deeper: Properties of the Exponential Family

- Mean is a function of θ and given by

$$\mathbb{E}(Y) \equiv \mu = b'(\theta)$$

- Variance is a function of θ and ϕ and given by

$$\mathbb{V}(Y) \equiv V = b''(\theta)a(\phi)$$

- Common forms of $a(\phi)$: 1 (Poisson, Bernoulli), ϕ (normal, Gamma), and ϕ/ω_i (binomial)

One Step Deeper: Properties of the Exponential Family

- Mean is a function of θ and given by

$$\mathbb{E}(Y) \equiv \mu = b'(\theta)$$

- Variance is a function of θ and ϕ and given by

$$\mathbb{V}(Y) \equiv V = b''(\theta)a(\phi)$$

- Common forms of $a(\phi)$: 1 (Poisson, Bernoulli), ϕ (normal, Gamma), and ϕ/ω_i (binomial)
- $b''(\theta)$ is called the variance function

One Step Deeper: Properties of the Exponential Family

- Mean is a function of θ and given by

$$\mathbb{E}(Y) \equiv \mu = b'(\theta)$$

- Variance is a function of θ and ϕ and given by

$$\mathbb{V}(Y) \equiv V = b''(\theta)a(\phi)$$

- Common forms of $a(\phi)$: 1 (Poisson, Bernoulli), ϕ (normal, Gamma), and ϕ/ω_i (binomial)
- $b''(\theta)$ is called the variance function
- In the Poisson model, $\theta_i = \log \mu_i$, $a(\phi) = 1$ and $b(\theta_i) = \exp(\theta_i)$

One Step Deeper: Properties of the Exponential Family

- Mean is a function of θ and given by

$$\mathbb{E}(Y) \equiv \mu = b'(\theta)$$

- Variance is a function of θ and ϕ and given by

$$\mathbb{V}(Y) \equiv V = b''(\theta)a(\phi)$$

- Common forms of $a(\phi)$: 1 (Poisson, Bernoulli), ϕ (normal, Gamma), and ϕ/ω_i (binomial)
- $b''(\theta)$ is called the variance function
- In the Poisson model, $\theta_i = \log \mu_i$, $a(\phi) = 1$ and $b(\theta_i) = \exp(\theta_i)$
  $\Rightarrow \mathbb{E}(Y_i) = \frac{db(\theta_i)}{d\theta_i} = \exp(\theta_i) = \mu_i$ and $\mathbb{V}(Y_i) = \frac{d^2 b(\theta_i)}{d\theta_i^2} = \exp(\theta_i) = \mu_i$

Summary

Summary

- Random variables and probability distributions provide useful models of the world

Summary

- Random variables and probability distributions provide useful models of the world
- We can characterize distributions in terms of their expectation (location) and variance (spread).

Summary

- Random variables and probability distributions provide useful models of the world
- We can characterize distributions in terms of their expectation (location) and variance (spread).
- Joint and conditional distributions capture the relationship between random variables.

Summary

- Random variables and probability distributions provide useful models of the world
- We can characterize distributions in terms of their expectation (location) and variance (spread).
- Joint and conditional distributions capture the relationship between random variables.
- There is a common set of famous distributions such as the Normal distribution.

Where We've Been and Where We're Going . . .

This week:

- Monday:
  - summarize one random variable using expectation and variance
  - show how to condition on a variable
- Wednesday:
  - properties of joint distributions
  - conditional expectations
  - covariance, correlation, independence

Next week:

Where We've Been and Where We're Going . . .

This week:

- Monday:
  - summarize one random variable using expectation and variance
  - show how to condition on a variable
- Wednesday:
  - properties of joint distributions
  - conditional expectations
  - covariance, correlation, independence

Next week:

- estimating these features from data

Where We've Been and Where We're Going . . .

This week:

- Monday:
  - summarize one random variable using expectation and variance
  - show how to condition on a variable
- Wednesday:
  - properties of joint distributions
  - conditional expectations
  - covariance, correlation, independence

Next week:

- estimating these features from data
- estimating uncertainty

Where We've Been and Where We're Going . . .

This week:

- Monday:
  - summarize one random variable using expectation and variance
  - show how to condition on a variable
- Wednesday:
  - properties of joint distributions
  - conditional expectations
  - covariance, correlation, independence

Next week:

- estimating these features from data
- estimating uncertainty

New reading:

Where We've Been and Where We're Going . . .

This week:

- Monday:
  - summarize one random variable using expectation and variance
  - show how to condition on a variable
- Wednesday:
  - properties of joint distributions
  - conditional expectations
  - covariance, correlation, independence

Next week:

- estimating these features from data
- estimating uncertainty

New reading:

- Aronow and Miller Chapter 3.1-3.2.6, 3.4.1
- Optional: Fox Chapter 3: Examining Data

1. Random Variables and Distributions
   - What is a Random Variable?
   - Discrete Distributions
   - Continuous Distributions
2. Characteristics of Distributions
   - Central Tendency
   - Measures of Dispersion
3. Conditional Distributions
4. Fun with Averages
5. Fun with Sensitive Questions
6. Appendix: Why the Mean?
7. Joint Distributions
   - Discrete Random Variable
   - Continuous Random Variable
8. Conditional Expectation
9. Properties
   - Independence
   - Covariance and Correlation
   - Conditional Independence
10. Famous Distributions
11. Fun With Spam

1 Random Variables and Distributions
- What is a Random Variable?
- Discrete Distributions
- Continuous Distributions

2 Characteristics of Distributions
- Central Tendency
- Measures of Dispersion

3 Conditional Distributions

4 Fun with Averages

5 Fun with Sensitive Questions

6 Appendix: Why the Mean?

7 Joint Distributions
- Discrete Random Variable
- Continuous Random Variable

8 Conditional Expectation

9 Properties
- Independence
- Covariance and Correlation
- Conditional Independence

10 Famous Distributions

11 Fun With Spam

Fun With Spam

Fun With: Building a Spam Filter

Suppose we have an email i, $(i = 1, \ldots, N)$ which we represent as a count of J words

Fun With: Building a Spam Filter

Suppose we have an email i, $(i = 1, \ldots, N)$ which we represent as a count of J words

$$\boldsymbol{x}_i = (x_{1i}, x_{2i}, \ldots, x_{Ji})$$

Fun With: Building a Spam Filter

Suppose we have an email i, $(i = 1, \ldots, N)$ which we represent as a count of J words

$$\boldsymbol{x}_i = (x_{1i}, x_{2i}, \ldots, x_{Ji})$$

Set of K categories. Category k $(k = 1, \ldots, K)$

$$\{C_1, C_2, \ldots, C_K\}$$

Fun With: Building a Spam Filter

Suppose we have an email i, $(i = 1, \ldots, N)$ which we represent as a count of J words

$$\boldsymbol{x}_i = (x_{1i}, x_{2i}, \ldots, x_{Ji})$$

Set of K categories. Category k $(k = 1, \ldots, K)$

$$\{C_1, C_2, \ldots, C_K\}$$

Subset of labeled documents $\boldsymbol{Y} = (Y_1, Y_2, \ldots, Y_N)$ where $Y_i \in \{C_1, C_2, \ldots, C_K\}$.

Fun With: Building a Spam Filter

Suppose we have an email i, $(i = 1, \ldots, N)$ which we represent as a count of J words

$$\boldsymbol{x}_i = (x_{1i}, x_{2i}, \ldots, x_{Ji})$$

Set of K categories. Category k $(k = 1, \ldots, K)$

$$\{C_1, C_2, \ldots, C_K\}$$

Subset of labeled documents $\boldsymbol{Y} = (Y_1, Y_2, \ldots, Y_N)$ where $Y_i \in \{C_1, C_2, \ldots, C_K\}$.

Goal: classify every document into one category.

Fun With: Building a Spam Filter

Suppose we have an email i, ($i = 1, \ldots, N$) which we represent as a count of J words

$$\boldsymbol{x}_i = (x_{1i}, x_{2i}, \ldots, x_{Ji})$$

Set of K categories. Category k ($k = 1, \ldots, K$)

$$\{C_1, C_2, \ldots, C_K\}$$

Subset of labeled documents $\boldsymbol{Y} = (Y_1, Y_2, \ldots, Y_N)$ where $Y_i \in \{C_1, C_2, \ldots, C_K\}$.

Goal: classify every document into one category.

- Learn a function that maps from space of (possible) documents to categories

Fun With: Building a Spam Filter

Suppose we have an email i, ($i = 1, \ldots, N$) which we represent as a count of J words

$$\boldsymbol{x}_i = (x_{1i}, x_{2i}, \ldots, x_{Ji})$$

Set of K categories. Category k ($k = 1, \ldots, K$)

$$\{C_1, C_2, \ldots, C_K\}$$

Subset of labeled documents $\boldsymbol{Y} = (Y_1, Y_2, \ldots, Y_N)$ where $Y_i \in \{C_1, C_2, \ldots, C_K\}$.

Goal: classify every document into one category.

- Learn a function that maps from space of (possible) documents to categories
- Use documents with known categories to estimate function

Fun With: Building a Spam Filter

Suppose we have an email i, $(i = 1, \ldots, N)$ which we represent as a count of J words

$$\boldsymbol{x}_i = (x_{1i}, x_{2i}, \ldots, x_{Ji})$$

Set of K categories. Category k $(k = 1, \ldots, K)$

$$\{C_1, C_2, \ldots, C_K\}$$

Subset of labeled documents $\boldsymbol{Y} = (Y_1, Y_2, \ldots, Y_N)$ where $Y_i \in \{C_1, C_2, \ldots, C_K\}$.

Goal: classify every document into one category.

- Learn a function that maps from space of (possible) documents to categories
- Use documents with known categories to estimate function
- Then apply model to new data, classify those observations

Example: Building a Spam Filter

Goal: For each document $\boldsymbol{x}_i$, we want to infer most likely category

Example: Building a Spam Filter

Goal: For each document $\boldsymbol{x}_i$, we want to infer most likely category

$$C_{\text{Max}} = \arg\max_k p(C_k|\boldsymbol{x}_i)$$

Example: Building a Spam Filter

Goal: For each document $\boldsymbol{x}_i$, we want to infer most likely category

$$C_{\text{Max}} = \arg\max_k p(C_k|\boldsymbol{x}_i)$$

We're going to use Bayes' rule to estimate $p(C_k|\boldsymbol{x}_i)$.

Example: Building a Spam Filter

Goal: For each document $\boldsymbol{x}_i$, we want to infer most likely category

$$C_{\text{Max}} = \arg\max_k p(C_k|\boldsymbol{x}_i)$$

We're going to use Bayes' rule to estimate $p(C_k|\boldsymbol{x}_i)$.

$$p(C_k|\boldsymbol{x}_i) = \frac{p(C_k, \boldsymbol{x}_i)}{p(\boldsymbol{x}_i)}$$

Example: Building a Spam Filter

Goal: For each document $\boldsymbol{x}_i$, we want to infer most likely category

$$C_{\text{Max}} = \arg\max_k p(C_k|\boldsymbol{x}_i)$$

We're going to use Bayes' rule to estimate $p(C_k|\boldsymbol{x}_i)$.

$$\begin{aligned} p(C_k|\boldsymbol{x}_i) &= \frac{p(C_k, \boldsymbol{x}_i)}{p(\boldsymbol{x}_i)} \\ &= \frac{p(C_k)p(\boldsymbol{x}_i|C_k)}{p(\boldsymbol{x}_i)} \end{aligned}$$

Example: Building a Spam Filter

Goal: For each document $\boldsymbol{x}_i$, we want to infer most likely category

$$C_{\text{Max}} = \arg\max_k p(C_k|\boldsymbol{x}_i)$$

We're going to use Bayes' rule to estimate $p(C_k|\boldsymbol{x}_i)$.

$$\begin{aligned} p(C_k|\boldsymbol{x}_i) &= \frac{p(C_k, \boldsymbol{x}_i)}{p(\boldsymbol{x}_i)} \\ &= \frac{\overbrace{p(C_k)}^{\text{Baseline Proportion}} \underbrace{p(\boldsymbol{x}_i|C_k)}_{\text{Words Given Category}}}{p(\boldsymbol{x}_i)} \end{aligned}$$

Example: Building a Spam Filter

Example: Building a Spam Filter

$$C_{\text{Max}} = \arg\max_k \frac{p(C_k)p(\boldsymbol{x}_i|C_k)}{p(\boldsymbol{x}_i)}$$

Example: Building a Spam Filter

$$
\begin{aligned}
C_{\text{Max}} &= \arg\max_k \frac{p(C_k)p(\boldsymbol{x}_i|C_k)}{p(\boldsymbol{x}_i)} \\
C_{\text{Max}} &= \arg\max_k \, p(C_k)p(\boldsymbol{x}_i|C_k)
\end{aligned}
$$

Example: Building a Spam Filter

$$\begin{aligned} C_{\text{Max}} &= \arg\max_k \frac{p(C_k)p(\boldsymbol{x}_i|C_k)}{p(\boldsymbol{x}_i)} \\ C_{\text{Max}} &= \arg\max_k p(C_k)p(\boldsymbol{x}_i|C_k) \end{aligned}$$

Two probabilities to estimate:

Example: Building a Spam Filter

$$C_{\text{Max}} = \arg\max_k \frac{p(C_k)p(\boldsymbol{x}_i|C_k)}{p(\boldsymbol{x}_i)}$$

$$C_{\text{Max}} = \arg\max_k p(C_k)p(\boldsymbol{x}_i|C_k)$$

Two probabilities to estimate:

$p(C_k) = \frac{\text{No. Documents in } k}{\text{No. Documents}}$ (from our labeled set)

Example: Building a Spam Filter

$$C_{\text{Max}} = \arg\max_k \frac{p(C_k)p(\boldsymbol{x}_i|C_k)}{p(\boldsymbol{x}_i)}$$

$$C_{\text{Max}} = \arg\max_k p(C_k)p(\boldsymbol{x}_i|C_k)$$

Two probabilities to estimate:

$p(C_k) = \frac{\text{No. Documents in } k}{\text{No. Documents}}$ (from our labeled set)

$p(\boldsymbol{x}_i|C_k)$ complicated without heroic assumptions

Example: Building a Spam Filter

$$C_{\text{Max}} = \arg\max_k \frac{p(C_k)p(\boldsymbol{x}_i|C_k)}{p(\boldsymbol{x}_i)}$$

$$C_{\text{Max}} = \arg\max_k p(C_k)p(\boldsymbol{x}_i|C_k)$$

Two probabilities to estimate:

$p(C_k) = \frac{\text{No. Documents in } k}{\text{No. Documents}}$ (from our labeled set)

$p(\boldsymbol{x}_i|C_k)$ complicated without heroic assumptions

- Even if x_{ij} is binary . Then 2^J possible $\boldsymbol{x}_i$ documents

Example: Building a Spam Filter

$$C_{\text{Max}} = \arg\max_k \frac{p(C_k)p(\boldsymbol{x}_i|C_k)}{p(\boldsymbol{x}_i)}$$
$$C_{\text{Max}} = \arg\max_k p(C_k)p(\boldsymbol{x}_i|C_k)$$

Two probabilities to estimate:

$p(C_k) = \frac{\text{No. Documents in } k}{\text{No. Documents}}$ (from our labeled set)

$p(\boldsymbol{x}_i|C_k)$ complicated without heroic assumptions

- Even if x_{ij} is binary . Then 2^J possible $\boldsymbol{x}_i$ documents
- Simplify: assume each word is independent given class

Example: Building a Spam Filter

$$C_{\text{Max}} = \arg\max_k \frac{p(C_k)p(\boldsymbol{x}_i|C_k)}{p(\boldsymbol{x}_i)}$$
$$C_{\text{Max}} = \arg\max_k p(C_k)p(\boldsymbol{x}_i|C_k)$$

Two probabilities to estimate:

$p(C_k) = \frac{\text{No. Documents in } k}{\text{No. Documents}}$ (from our labeled set)

$p(\boldsymbol{x}_i|C_k)$ complicated without heroic assumptions

- Even if x_{ij} is binary . Then 2^J possible $\boldsymbol{x}_i$ documents
- Simplify: assume each word is independent given class

$$p(\boldsymbol{x}_i|C_k) = \prod_{j=1}^{J} p(x_{ij}|C_k)$$

Example: Building a Spam Filter

$$C_{\text{Max}} = \arg\max_k \frac{p(C_k)p(\boldsymbol{x}_i|C_k)}{p(\boldsymbol{x}_i)}$$
$$C_{\text{Max}} = \arg\max_k p(C_k)p(\boldsymbol{x}_i|C_k)$$

Two probabilities to estimate:

$p(C_k) = \frac{\text{No. Documents in } k}{\text{No. Documents}}$ (from our labeled set)

$p(\boldsymbol{x}_i|C_k)$ complicated without heroic assumptions

- Even if x_{ij} is binary . Then 2^J possible $\boldsymbol{x}_i$ documents
- Simplify: assume each word is independent given class

$$p(\boldsymbol{x}_i|C_k) = \prod_{j=1}^{J} p(x_{ij}|C_k)$$

This is called a Naïve Bayes classifier.

Estimating the Naïve Bayes Classifier

Two components to estimate:

Estimating the Naïve Bayes Classifier

Two components to estimate:

- $p(C_k) = \frac{\text{No. Documents in } k}{\text{No. Documents}}$

Estimating the Naïve Bayes Classifier

Two components to estimate:

- $p(C_k) = \frac{\text{No. Documents in } k}{\text{No. Documents}}$
- $p(\boldsymbol{x}_i|C_k) = \prod_{j=1}^{J} p(x_{ij}|C_k)$

Estimating the Naïve Bayes Classifier

Two components to estimate:

- $p(C_k) = \frac{\text{No. Documents in } k}{\text{No. Documents}}$
- $p(\boldsymbol{x}_i | C_k) = \prod_{j=1}^{J} p(x_{ij}|C_k)$

$$p(x_{im} = z | C_k) = \frac{\text{No(Docs}_{ij} = \text{z and C} = \text{C}_k\text{)}}{\text{No(C= C}_k\text{)}}$$

Estimating the Naïve Bayes Classifier

Two components to estimate:

- $p(C_k) = \frac{\text{No. Documents in } k}{\text{No. Documents}}$
- $p(\boldsymbol{x}_i|C_k) = \prod_{j=1}^{J} p(x_{ij}|C_k)$

$$p(x_{im} = z|C_k) = \frac{\text{No(Docs}_{ij} = \text{z and C} = \text{C}_k \text{)}}{\text{No(C= C}_k\text{)}}$$

Algorithm steps:

Estimating the Naïve Bayes Classifier

Two components to estimate:

- $p(C_k) = \frac{\text{No. Documents in } k}{\text{No. Documents}}$
- $p(\boldsymbol{x}_i|C_k) = \prod_{j=1}^{J} p(x_{ij}|C_k)$

$$p(x_{im} = z|C_k) = \frac{\text{No(Docs}_{ij} = \text{z and C} = \text{C}_k\text{)}}{\text{No(C= C}_k\text{)}}$$

Algorithm steps:

1) Learn $\hat{p}(C)$ and $\hat{p}(\boldsymbol{x}_i|C_k)$ on labeled data

Estimating the Naïve Bayes Classifier

Two components to estimate:

- $p(C_k) = \frac{\text{No. Documents in } k}{\text{No. Documents}}$
- $p(\boldsymbol{x}_i | C_k) = \prod_{j=1}^{J} p(x_{ij} | C_k)$

$$p(x_{im} = z | C_k) = \frac{\text{No(Docs}_{ij} = \text{z and C} = \text{C}_k \text{)}}{\text{No(C= C}_k\text{)}}$$

Algorithm steps:

1) Learn $\hat{p}(C)$ and $\hat{p}(\boldsymbol{x}_i | C_k)$ on labeled data
2) Use this to identify most likely C_k for each document i in unlabeled data

Estimating the Naïve Bayes Classifier

Two components to estimate:

- $p(C_k) = \frac{\text{No. Documents in } k}{\text{No. Documents}}$
- $p(\boldsymbol{x}_i | C_k) = \prod_{j=1}^{J} p(x_{ij} | C_k)$

$$p(x_{im} = z | C_k) = \frac{\text{No(Docs}_{ij} = \text{z and C} = \text{C}_k\text{)}}{\text{No(C= C}_k\text{)}}$$

Algorithm steps:

1) Learn $\hat{p}(C)$ and $\hat{p}(\boldsymbol{x}_i | C_k)$ on labeled data
2) Use this to identify most likely C_k for each document i in unlabeled data

Simple intuition about Naïve Bayes:

Estimating the Naïve Bayes Classifier

Two components to estimate:

- $p(C_k) = \frac{\text{No. Documents in } k}{\text{No. Documents}}$
- $p(\boldsymbol{x}_i | C_k) = \prod_{j=1}^{J} p(x_{ij} | C_k)$

$$p(x_{im} = z | C_k) = \frac{\text{No(Docs}_{ij} = \text{z and C} = \text{C}_k\text{)}}{\text{No(C= C}_k\text{)}}$$

Algorithm steps:

1) Learn $\hat{p}(C)$ and $\hat{p}(\boldsymbol{x}_i | C_k)$ on labeled data
2) Use this to identify most likely C_k for each document i in unlabeled data

Simple intuition about Naïve Bayes:

- Learn what documents in class j look like

Estimating the Naïve Bayes Classifier

Two components to estimate:

- $p(C_k) = \frac{\text{No. Documents in } k}{\text{No. Documents}}$
- $p(\boldsymbol{x}_i|C_k) = \prod_{j=1}^{J} p(x_{ij}|C_k)$

$$p(x_{im} = z|C_k) = \frac{\text{No(Docs}_{ij} = \text{z and C} = \text{C}_k\text{)}}{\text{No(C= C}_k\text{)}}$$

Algorithm steps:

1) Learn $\hat{p}(C)$ and $\hat{p}(\boldsymbol{x}_i|C_k)$ on labeled data
2) Use this to identify most likely C_k for each document i in unlabeled data

Simple intuition about Naïve Bayes:

- Learn what documents in class j look like
- Find class k that document i is most similar to

Example: Building a Spam Filter

Example: Building a Spam Filter

Scoring the algorithm is easy.

$$p(C_k|\boldsymbol{x}_i) \propto p(C_k) \prod_{j=1}^{J} p(x_{i,j}|C_k)^{x_{ij}}$$

which is simply the probability of the class multiplied by the product of the probabilities for the words that are observed in the test document.

Example: Building a Spam Filter

Example: Building a Spam Filter

- Learn the most probable class using Bayes Rule and a powerful but "naïve" independence assumption

Example: Building a Spam Filter

- Learn the most probable class using Bayes Rule and a powerful but "naïve" independence assumption
- Despite that the model is "wrong" it classifies spam quite well

Example: Building a Spam Filter

- Learn the most probable class using Bayes Rule and a powerful but "naïve" independence assumption
- Despite that the model is "wrong" it classifies spam quite well
- Shares the basic structure of many models, is a building block for more complex models

Example: Building a Spam Filter

- Learn the most probable class using Bayes Rule and a powerful but "naïve" independence assumption
- Despite that the model is "wrong" it classifies spam quite well
- Shares the basic structure of many models, is a building block for more complex models
- This was a complicated example, it is okay if you didn't follow all of it.

Example: Building a Spam Filter

- Learn the most probable class using Bayes Rule and a powerful but "naïve" independence assumption
- Despite that the model is "wrong" it classifies spam quite well
- Shares the basic structure of many models, is a building block for more complex models
- This was a complicated example, it is okay if you didn't follow all of it.
- More on estimators next week!