

Advanced Quantitative Research Methodology, Lecture Notes: Introduction

Gary King¹
Institute for Quantitative Social Science
Harvard University

January 9, 2018

¹GaryKing.org

What is the Field of Statistical Analysis?

What is the Field of Statistical Analysis?

- **An Old Field:** originates in the study of politics and government: “*state-istics*” (circa 1662)

What is the Field of Statistical Analysis?

- **An Old Field:** originates in the study of politics and government: “*state*-istics” (circa 1662)
- **Recent Progress:** increasing exponentially

What is the Field of Statistical Analysis?

- **An Old Field:** originates in the study of politics and government: “*state*-istics” (circa 1662)
- **Recent Progress:** increasing exponentially
 - **Mid-1930s:** Experiments and random assignment

What is the Field of Statistical Analysis?

- **An Old Field:** originates in the study of politics and government: “*state*-istics” (circa 1662)
- **Recent Progress:** increasing exponentially
 - **Mid-1930s:** Experiments and random assignment
 - **1950s:** The modern theory of inference

What is the Field of Statistical Analysis?

- **An Old Field:** originates in the study of politics and government: “*state*-istics” (circa 1662)
- **Recent Progress:** increasing exponentially
 - **Mid-1930s:** Experiments and random assignment
 - **1950s:** The modern theory of inference
 - **In your lifetime:** Modern causal inference

What is the Field of Statistical Analysis?

- **An Old Field:** originates in the study of politics and government: “*state*-istics” (circa 1662)
- **Recent Progress:** increasing exponentially
 - **Mid-1930s:** Experiments and random assignment
 - **1950s:** The modern theory of inference
 - **In your lifetime:** Modern causal inference
 - **Since you started school:** Part of a monumental societal change, “big data”, and the march of quantification through academic, professional, commercial, government and policy fields.

What is the Field of Statistical Analysis?

- **An Old Field:** originates in the study of politics and government: “*state*-istics” (circa 1662)
- **Recent Progress:** increasing exponentially
 - **Mid-1930s:** Experiments and random assignment
 - **1950s:** The modern theory of inference
 - **In your lifetime:** Modern causal inference
 - **Since you started school:** Part of a monumental societal change, “big data”, and the march of quantification through academic, professional, commercial, government and policy fields.
- **The number of new methods:** increasing fast

What is the Field of Statistical Analysis?

- **An Old Field:** originates in the study of politics and government: “*state-istics*” (circa 1662)
- **Recent Progress:** increasing exponentially
 - **Mid-1930s:** Experiments and random assignment
 - **1950s:** The modern theory of inference
 - **In your lifetime:** Modern causal inference
 - **Since you started school:** Part of a monumental societal change, “big data”, and the march of quantification through academic, professional, commercial, government and policy fields.
- **The number of new methods:** increasing fast
- **Most important methods:** originate *outside* the discipline of statistics (random assignment, experimental design, survey research, machine learning, MCMC methods, ...).

What is the Field of Statistical Analysis?

- **An Old Field:** originates in the study of politics and government: “*state-istics*” (circa 1662)
- **Recent Progress:** increasing exponentially
 - **Mid-1930s:** Experiments and random assignment
 - **1950s:** The modern theory of inference
 - **In your lifetime:** Modern causal inference
 - **Since you started school:** Part of a monumental societal change, “big data”, and the march of quantification through academic, professional, commercial, government and policy fields.
- **The number of new methods:** increasing fast
- **Most important methods:** originate *outside* the discipline of statistics (random assignment, experimental design, survey research, machine learning, MCMC methods, ...).
- **Statistics:** abstracts, proves formal properties, generalizes, and distributes results back out.

A 2nd Course in *Political Methodology*

A 2nd Course in *Political Methodology*

- Political science's methodological subfield

A 2nd Course in *Political Methodology*

- Political science's methodological subfield
- Second largest APSA section, essentially no unemployment

A 2nd Course in *Political Methodology*

- Political science's methodological subfield
- Second largest APSA section, essentially no unemployment
- Akin to:

A 2nd Course in *Political Methodology*

- Political science's methodological subfield
- Second largest APSA section, essentially no unemployment
- Akin to: econometrics within economics,

A 2nd Course in *Political Methodology*

- Political science's methodological subfield
- Second largest APSA section, essentially no unemployment
- Akin to: econometrics within economics, psychometrics within psychology,

A 2nd Course in *Political Methodology*

- Political science's methodological subfield
- Second largest APSA section, essentially no unemployment
- Akin to: econometrics within economics, psychometrics within psychology, sociological methodology within sociology,

A 2nd Course in *Political Methodology*

- Political science's methodological subfield
- Second largest APSA section, essentially no unemployment
- Akin to: econometrics within economics, psychometrics within psychology, sociological methodology within sociology, biostatistics within public health and medicine,

A 2nd Course in *Political Methodology*

- Political science's methodological subfield
- Second largest APSA section, essentially no unemployment
- Akin to: econometrics within economics, psychometrics within psychology, sociological methodology within sociology, biostatistics within public health and medicine, many others

A 2nd Course in *Political Methodology*

- Political science's methodological subfield
- Second largest APSA section, essentially no unemployment
- Akin to: econometrics within economics, psychometrics within psychology, sociological methodology within sociology, biostatistics within public health and medicine, many others
- Union of Methods Subfields:

A 2nd Course in *Political Methodology*

- Political science's methodological subfield
- Second largest APSA section, essentially no unemployment
- Akin to: econometrics within economics, psychometrics within psychology, sociological methodology within sociology, biostatistics within public health and medicine, many others
- Union of Methods Subfields: known as data science,

A 2nd Course in *Political Methodology*

- Political science's methodological subfield
- Second largest APSA section, essentially no unemployment
- Akin to: econometrics within economics, psychometrics within psychology, sociological methodology within sociology, biostatistics within public health and medicine, many others
- Union of Methods Subfields: known as data science, applied statistics,

A 2nd Course in *Political Methodology*

- Political science's methodological subfield
- Second largest APSA section, essentially no unemployment
- Akin to: econometrics within economics, psychometrics within psychology, sociological methodology within sociology, biostatistics within public health and medicine, many others
- Union of Methods Subfields: known as data science, applied statistics, computational social science

A 2nd Course in *Political Methodology*

- Political science's methodological subfield
- Second largest APSA section, essentially no unemployment
- Akin to: econometrics within economics, psychometrics within psychology, sociological methodology within sociology, biostatistics within public health and medicine, many others
- Union of Methods Subfields: known as data science, applied statistics, computational social science
- Part of a massive change in the evidence base of the social sciences, from: (a) surveys, (b) end of period government stats, and (c) one-off studies of people, places, or events ~> numerous new types and huge quantities of (big) data

A 2nd Course in *Political Methodology*

- Political science's methodological subfield
- Second largest APSA section, essentially no unemployment
- Akin to: econometrics within economics, psychometrics within psychology, sociological methodology within sociology, biostatistics within public health and medicine, many others
- Union of Methods Subfields: known as data science, applied statistics, computational social science
- Part of a massive change in the evidence base of the social sciences, from: (a) surveys, (b) end of period government stats, and (c) one-off studies of people, places, or events ~> numerous new types and huge quantities of (big) data
- Political science: an unusually diverse discipline, cross-roads of other fields, great place for a broad perspective on methods

What's *This* Course in Political Methodology About?

What's *This* Course in Political Methodology About?

1. Understanding statistical inference

What's *This* Course in Political Methodology About?

1. Understanding statistical inference
2. Making novel contributions to a scholarly literature.

What's *This* Course in Political Methodology About?

1. Understanding statistical inference
2. Making novel contributions to a scholarly literature.
3. Learning how to choose a topic for research, solve hard problems by changing the question, and identify a big idea.

What's *This Course in Political Methodology* About?

1. Understanding statistical inference

- “Using facts you know to learn about facts you don’t know”

2. Making novel contributions to a scholarly literature.

3. Learning how to choose a topic for research, solve hard problems by changing the question, and identify a big idea.

What's *This Course in Political Methodology* About?

1. Understanding statistical inference

- “Using facts you know to learn about facts you don’t know”
- Understand statistical theory underlying specific methods

2. Making novel contributions to a scholarly literature.

3. Learning how to choose a topic for research, solve hard problems by changing the question, and identify a big idea.

What's *This Course in Political Methodology* About?

1. Understanding statistical inference

- “Using facts you know to learn about facts you don’t know”
- Understand statistical theory underlying specific methods
- Feel comfortable using these methods in your own work

2. Making novel contributions to a scholarly literature.

3. Learning how to choose a topic for research, solve hard problems by changing the question, and identify a big idea.

What's *This Course in Political Methodology* About?

1. Understanding statistical inference

- “Using facts you know to learn about facts you don’t know”
- Understand statistical theory underlying specific methods
- Feel comfortable using these methods in your own work
- Learn, implement, apply, interpret, and explain methods invented after class ends

2. Making novel contributions to a scholarly literature.

3. Learning how to choose a topic for research, solve hard problems by changing the question, and identify a big idea.

What's *This Course in Political Methodology* About?

1. Understanding statistical inference

- “Using facts you know to learn about facts you don’t know”
- Understand statistical theory underlying specific methods
- Feel comfortable using these methods in your own work
- Learn, implement, apply, interpret, and explain methods invented after class ends
- **The practical tools of research**: theory, applications, simulation, programming — whatever is useful

2. Making novel contributions to a scholarly literature.

3. Learning how to choose a topic for research, solve hard problems by changing the question, and identify a big idea.

What's *This* Course in Political Methodology About?

1. Understanding statistical inference

- “Using facts you know to learn about facts you don’t know”
- Understand statistical theory underlying specific methods
- Feel comfortable using these methods in your own work
- Learn, implement, apply, interpret, and explain methods invented after class ends
- **The practical tools of research**: theory, applications, simulation, programming — whatever is useful

2. Making novel contributions to a scholarly literature.

- Sounds hard, but almost everyone gets there

3. Learning how to choose a topic for research, solve hard problems by changing the question, and identify a big idea.

What's *This* Course in Political Methodology About?

1. Understanding statistical inference

- “Using facts you know to learn about facts you don’t know”
- Understand statistical theory underlying specific methods
- Feel comfortable using these methods in your own work
- Learn, implement, apply, interpret, and explain methods invented after class ends
- **The practical tools of research**: theory, applications, simulation, programming — whatever is useful

2. Making novel contributions to a scholarly literature.

- Sounds hard, but almost everyone gets there
- Revised versions of class papers (by grads and undergrads) $\rightsquigarrow$ articles, books, senior theses, dissertations, blog posts, conference presentations, many awards

3. Learning how to choose a topic for research, solve hard problems by changing the question, and identify a big idea.

What's *This Course in Political Methodology* About?

1. Understanding statistical inference

- “Using facts you know to learn about facts you don’t know”
- Understand statistical theory underlying specific methods
- Feel comfortable using these methods in your own work
- Learn, implement, apply, interpret, and explain methods invented after class ends
- **The practical tools of research**: theory, applications, simulation, programming — whatever is useful

2. Making novel contributions to a scholarly literature.

- Sounds hard, but almost everyone gets there
- Revised versions of class papers (by grads and undergrads) ~> articles, books, senior theses, dissertations, blog posts, conference presentations, many awards

3. Learning how to choose a topic for research, solve hard problems by changing the question, and identify a big idea.

- Biggest factor impacting success: the question

What's *This Course in Political Methodology* About?

1. Understanding statistical inference

- “Using facts you know to learn about facts you don’t know”
- Understand statistical theory underlying specific methods
- Feel comfortable using these methods in your own work
- Learn, implement, apply, interpret, and explain methods invented after class ends
- **The practical tools of research**: theory, applications, simulation, programming — whatever is useful

2. Making novel contributions to a scholarly literature.

- Sounds hard, but almost everyone gets there
- Revised versions of class papers (by grads and undergrads) ~> articles, books, senior theses, dissertations, blog posts, conference presentations, many awards

3. Learning how to choose a topic for research, solve hard problems by changing the question, and identify a big idea.

- Biggest factor impacting success: the question
- Nursery school until today: questions are fixed, immutable

What's *This Course in Political Methodology* About?

1. Understanding statistical inference

- “Using facts you know to learn about facts you don’t know”
- Understand statistical theory underlying specific methods
- Feel comfortable using these methods in your own work
- Learn, implement, apply, interpret, and explain methods invented after class ends
- **The practical tools of research**: theory, applications, simulation, programming — whatever is useful

2. Making novel contributions to a scholarly literature.

- Sounds hard, but almost everyone gets there
- Revised versions of class papers (by grads and undergrads) ~> articles, books, senior theses, dissertations, blog posts, conference presentations, many awards

3. Learning how to choose a topic for research, solve hard problems by changing the question, and identify a big idea.

- Biggest factor impacting success: the question
- Nursery school until today: questions are fixed, immutable
- Today: you can change or choose the question!

Who takes this course?

Who takes this course?

- Gov2001

Who takes this course?

- Gov2001
 - Most Gov Dept grad students doing empirical work

Who takes this course?

- Gov2001
 - Most Gov Dept grad students doing empirical work
 - Grad students from other Harvard departments & schools

Who takes this course?

- Gov2001
 - Most Gov Dept grad students doing empirical work
 - Grad students from other Harvard departments & schools
- Gov1002 Undergrads

Who takes this course?

- Gov2001
 - Most Gov Dept grad students doing empirical work
 - Grad students from other Harvard departments & schools
- Gov1002 Undergrads
- StatE-200 Non-Harvard students, visitors, faculty, & others (online through the Harvard Extension school)

Who takes this course?

- Gov2001
 - Most Gov Dept grad students doing empirical work
 - Grad students from other Harvard departments & schools
- Gov1002 Undergrads
- StatE-200 Non-Harvard students, visitors, faculty, & others (online through the Harvard Extension school)

Some of the best experiences here: getting to know people in other fields

Requirements

Requirements

1. Engage in the collective experience:

Requirements

1. Engage in the collective experience:
2. Weekly assignments

Requirements

1. Engage in the collective experience:
2. Weekly assignments
3. One “publishable” coauthored paper. (Easier than you think!)

Requirements

1. Engage in the collective experience:
 - Many assignments: carefully choreographed dances
2. Weekly assignments
3. One “publishable” coauthored paper. (Easier than you think!)

Requirements

1. Engage in the collective experience:

- Many assignments: carefully choreographed dances
- Other students depend on you, and you on them

2. Weekly assignments

3. One “publishable” coauthored paper. (Easier than you think!)

Requirements

1. Engage in the collective experience:

- Many assignments: carefully choreographed dances
- Other students depend on you, and you on them
- Come to class, meet all deadlines, support your classmates

2. Weekly assignments

3. One “publishable” coauthored paper. (Easier than you think!)

Requirements

1. Engage in the collective experience:

- Many assignments: carefully choreographed dances
- Other students depend on you, and you on them
- Come to class, meet all deadlines, support your classmates
- We learn better collectively (did MOOCs replace Harvard?)

2. Weekly assignments

3. One “publishable” coauthored paper. (Easier than you think!)

Requirements

1. Engage in the collective experience:

- Many assignments: carefully choreographed dances
- Other students depend on you, and you on them
- Come to class, meet all deadlines, support your classmates
- We learn better collectively (did MOOCs replace Harvard?)

2. Weekly assignments

- Readings via Perusall

3. One “publishable” coauthored paper. (Easier than you think!)

Requirements

1. Engage in the collective experience:

- Many assignments: carefully choreographed dances
- Other students depend on you, and you on them
- Come to class, meet all deadlines, support your classmates
- We learn better collectively (did MOOCs replace Harvard?)

2. Weekly assignments

- Readings via Perusall
- Annotate, help others, don't skip equations

3. One “publishable” coauthored paper. (Easier than you think!)

Requirements

1. Engage in the collective experience:

- Many assignments: carefully choreographed dances
- Other students depend on you, and you on them
- Come to class, meet all deadlines, support your classmates
- We learn better collectively (did MOOCs replace Harvard?)

2. Weekly assignments

- Readings via Perusall
- Annotate, help others, don't skip equations
- Problem sets & assessment questions: try it yourself, work in groups of 3-4, write it up yourself

3. One “publishable” coauthored paper. (Easier than you think!)

Requirements

1. Engage in the collective experience:

- Many assignments: carefully choreographed dances
- Other students depend on you, and you on them
- Come to class, meet all deadlines, support your classmates
- We learn better collectively (did MOOCs replace Harvard?)

2. Weekly assignments

- Readings via Perusall
- Annotate, help others, don't skip equations
- Problem sets & assessment questions: try it yourself, work in groups of 3-4, write it up yourself
- All questions sent via Perusall

3. One “publishable” coauthored paper. (Easier than you think!)

Requirements

1. Engage in the collective experience:

- Many assignments: carefully choreographed dances
- Other students depend on you, and you on them
- Come to class, meet all deadlines, support your classmates
- We learn better collectively (did MOOCs replace Harvard?)

2. Weekly assignments

- Readings via Perusall
- Annotate, help others, don't skip equations
- Problem sets & assessment questions: try it yourself, work in groups of 3-4, write it up yourself
- All questions sent via Perusall
- Peer grading and peer instruction (“teaching teaches the teacher”)

3. One “publishable” coauthored paper. (Easier than you think!)

Requirements

1. Engage in the collective experience:

- Many assignments: carefully choreographed dances
- Other students depend on you, and you on them
- Come to class, meet all deadlines, support your classmates
- We learn better collectively (did MOOCs replace Harvard?)

2. Weekly assignments

- Readings via Perusall
- Annotate, help others, don't skip equations
- Problem sets & assessment questions: try it yourself, work in groups of 3-4, write it up yourself
- All questions sent via Perusall
- Peer grading and peer instruction (“teaching teaches the teacher”)

3. One “publishable” coauthored paper. (Easier than you think!)

- Draft submission and replication exercise helps a lot.

Requirements

1. Engage in the collective experience:

- Many assignments: carefully choreographed dances
- Other students depend on you, and you on them
- Come to class, meet all deadlines, support your classmates
- We learn better collectively (did MOOCs replace Harvard?)

2. Weekly assignments

- Readings via Perusall
- Annotate, help others, don't skip equations
- Problem sets & assessment questions: try it yourself, work in groups of 3-4, write it up yourself
- All questions sent via Perusall
- Peer grading and peer instruction (“teaching teaches the teacher”)

3. One “publishable” coauthored paper. (Easier than you think!)

- Draft submission and replication exercise helps a lot.
- See “Publication, Publication”

Requirements

1. Engage in the collective experience:

- Many assignments: carefully choreographed dances
- Other students depend on you, and you on them
- Come to class, meet all deadlines, support your classmates
- We learn better collectively (did MOOCs replace Harvard?)

2. Weekly assignments

- Readings via Perusall
- Annotate, help others, don't skip equations
- Problem sets & assessment questions: try it yourself, work in groups of 3-4, write it up yourself
- All questions sent via Perusall
- Peer grading and peer instruction (“teaching teaches the teacher”)

3. One “publishable” coauthored paper. (Easier than you think!)

- Draft submission and replication exercise helps a lot.
- See “Publication, Publication”
- You won't be alone: you'll work with each other and us

Requirements

1. Engage in the collective experience:

- Many assignments: carefully choreographed dances
- Other students depend on you, and you on them
- Come to class, meet all deadlines, support your classmates
- We learn better collectively (did MOOCs replace Harvard?)

2. Weekly assignments

- Readings via Perusall
- Annotate, help others, don't skip equations
- Problem sets & assessment questions: try it yourself, work in groups of 3-4, write it up yourself
- All questions sent via Perusall
- Peer grading and peer instruction (“teaching teaches the teacher”)

3. One “publishable” coauthored paper. (Easier than you think!)

- Draft submission and replication exercise helps a lot.
- See “Publication, Publication”
- You won't be alone: you'll work with each other and us
- Peer grading and peer instruction

Course Information

Course Information

- All class materials: j.mp/G2001

Course Information

- All class materials: j.mp/G2001
- Find [Study Group](#) of 3–4, and [Coauthor Group](#) of 2–3

Course Information

- All class materials: j.mp/G2001
- Find [Study Group](#) of 3–4, and [Coauthor Group](#) of 2–3
- Browse archive of [previous years' emails](#) (Note which now-famous scholar is asking the question!)

Course Information

- All class materials: j.mp/G2001
- Find [Study Group](#) of 3–4, and [Coauthor Group](#) of 2–3
- Browse archive of [previous years' emails](#) (Note which now-famous scholar is asking the question!)

Course Information

- All class materials: j.mp/G2001
- Find [Study Group](#) of 3–4, and [Coauthor Group](#) of 2–3
- Browse archive of [previous years' emails](#) (Note which now-famous scholar is asking the question!)
- In-

Course Information

- All class materials: j.mp/G2001
- Find [Study Group](#) of 3–4, and [Coauthor Group](#) of 2–3
- Browse archive of [previous years' emails](#) (Note which now-famous scholar is asking the question!)
- In-ter-

Course Information

- All class materials: j.mp/G2001
- Find [Study Group](#) of 3–4, and [Coauthor Group](#) of 2–3
- Browse archive of [previous years' emails](#) (Note which now-famous scholar is asking the question!)
- [In-ter-rupt](#)

Course Information

- All class materials: j.mp/G2001
- Find [Study Group](#) of 3–4, and [Coauthor Group](#) of 2–3
- Browse archive of [previous years' emails](#) (Note which now-famous scholar is asking the question!)
- [In-ter-rupt](#) me as often as necessary

Course Information

- All class materials: j.mp/G2001
- Find [Study Group](#) of 3–4, and [Coauthor Group](#) of 2–3
- Browse archive of [previous years' emails](#) (Note which now-famous scholar is asking the question!)
- [In-ter-rupt](#) me as often as necessary
- (Got a dumb question? Assume you are the smartest person in class and you eventually will be!)

Course Information

- All class materials: j.mp/G2001
- Find [Study Group](#) of 3–4, and [Coauthor Group](#) of 2–3
- Browse archive of [previous years' emails](#) (Note which now-famous scholar is asking the question!)
- [In-ter-rupt](#) me as often as necessary
- (Got a dumb question? Assume you are the smartest person in class and you eventually will be!)
- [When are Gary's office hours?](#)

Course Information

- All class materials: j.mp/G2001
- Find [Study Group](#) of 3–4, and [Coauthor Group](#) of 2–3
- Browse archive of [previous years' emails](#) (Note which now-famous scholar is asking the question!)
- [In-ter-rupt](#) me as often as necessary
- (Got a dumb question? Assume you are the smartest person in class and you eventually will be!)
- [When are Gary's office hours?](#)
 - (Big secret: Office hours prevent visits at other times)

Course Information

- All class materials: j.mp/G2001
- Find [Study Group](#) of 3–4, and [Coauthor Group](#) of 2–3
- Browse archive of [previous years' emails](#) (Note which now-famous scholar is asking the question!)
- [In-ter-rupt](#) me as often as necessary
- (Got a dumb question? Assume you are the smartest person in class and you eventually will be!)
- [When are Gary's office hours?](#)
 - (Big secret: Office hours prevent visits at other times)
 - Come when you like; if you can't find me or I'm in a meeting, come back, talk to my assistant in the office next to me, or send chat via Perusall

Course Information

- All class materials: j.mp/G2001
- Find [Study Group](#) of 3–4, and [Coauthor Group](#) of 2–3
- Browse archive of [previous years' emails](#) (Note which now-famous scholar is asking the question!)
- [In-ter-rupt](#) me as often as necessary
- (Got a dumb question? Assume you are the smartest person in class and you eventually will be!)
- [When are Gary's office hours?](#)
 - (Big secret: Office hours prevent visits at other times)
 - Come when you like; if you can't find me or I'm in a meeting, come back, talk to my assistant in the office next to me, or send chat via Perusall
- [Focus, like I will, on learning \(not grades\)](#): Especially when we work on papers, I will treat you like a [colleague](#), not a [student](#)

Course Strategy

Course Strategy

- We could teach you the latest and greatest methods,

Course Strategy

- We could teach you the latest and greatest methods, but when you graduate *they will be old*

Course Strategy

- We could teach you the latest and greatest methods, but when you graduate *they will be old*
- We could teach you all the methods that might prove useful during your career,

Course Strategy

- We could teach you the latest and greatest methods, but when you graduate *they will be old*
- We could teach you all the methods that might prove useful during your career, but when you graduate *you will be old*

Course Strategy

- We could teach you the latest and greatest methods, but when you graduate *they will be old*
- We could teach you all the methods that might prove useful during your career, but when you graduate *you will be old*
- Instead, we teach you the *fundamentals*,

Course Strategy

- We could teach you the latest and greatest methods, but when you graduate *they will be old*
- We could teach you all the methods that might prove useful during your career, but when you graduate *you will be old*
- Instead, we teach you the *fundamentals*, the underlying *theory of inference*, from which statistical models are developed

Course Strategy

- We could teach you the latest and greatest methods, but when you graduate *they will be old*
- We could teach you all the methods that might prove useful during your career, but when you graduate *you will be old*
- Instead, we teach you the *fundamentals*, the underlying *theory of inference*, from which statistical models are developed
- *Balance of emphasis: proofs v. concepts & intuition*

Course Strategy

- We could teach you the latest and greatest methods, but when you graduate *they will be old*
- We could teach you all the methods that might prove useful during your career, but when you graduate *you will be old*
- Instead, we teach you the *fundamentals*, the underlying *theory of inference*, from which statistical models are developed
- *Balance of emphasis: proofs v. concepts & intuition*
 - *Baby Stats*: dumbed down proofs, vague intuition

Course Strategy

- We could teach you the latest and greatest methods, but when you graduate *they will be old*
- We could teach you all the methods that might prove useful during your career, but when you graduate *you will be old*
- Instead, we teach you the *fundamentals*, the underlying *theory of inference*, from which statistical models are developed
- *Balance of emphasis: proofs v. concepts & intuition*
 - *Baby Stats*: dumbed down proofs, vague intuition
 - *Math Stats*: rigorous mathematical proofs, no data experience

Course Strategy

- We could teach you the latest and greatest methods, but when you graduate *they will be old*
- We could teach you all the methods that might prove useful during your career, but when you graduate *you will be old*
- Instead, we teach you the **fundamentals**, the underlying **theory of inference**, from which statistical models are developed
- **Balance of emphasis: proofs v. concepts & intuition**
 - **Baby Stats**: dumbed down proofs, vague intuition
 - **Math Stats**: rigorous mathematical proofs, no data experience
 - **Practical Stats**: math when needed, deep concepts and intuition

Course Strategy

- We could teach you the latest and greatest methods, but when you graduate *they will be old*
- We could teach you all the methods that might prove useful during your career, but when you graduate *you will be old*
- Instead, we teach you the **fundamentals**, the underlying **theory of inference**, from which statistical models are developed
- **Balance of emphasis: proofs v. concepts & intuition**
 - **Baby Stats**: dumbed down proofs, vague intuition
 - **Math Stats**: rigorous mathematical proofs, no data experience
 - **Practical Stats**: math when needed, deep concepts and intuition
 - **Ultimate Goal**: how to do empirical research, in depth

Course Strategy

- We could teach you the latest and greatest methods, but when you graduate *they will be old*
- We could teach you all the methods that might prove useful during your career, but when you graduate *you will be old*
- Instead, we teach you the *fundamentals*, the underlying *theory of inference*, from which statistical models are developed
- *Balance of emphasis: proofs v. concepts & intuition*
 - *Baby Stats*: dumbed down proofs, vague intuition
 - *Math Stats*: rigorous mathematical proofs, no data experience
 - *Practical Stats*: math when needed, deep concepts and intuition
 - *Ultimate Goal*: how to do empirical research, in depth
 - *Procedure*: Traverse from theoretical foundations to practical application (including “how to” computations)

Course Strategy

- We could teach you the latest and greatest methods, but when you graduate *they will be old*
- We could teach you all the methods that might prove useful during your career, but when you graduate *you will be old*
- Instead, we teach you the **fundamentals**, the underlying **theory of inference**, from which statistical models are developed
- **Balance of emphasis: proofs v. concepts & intuition**
 - **Baby Stats**: dumbed down proofs, vague intuition
 - **Math Stats**: rigorous mathematical proofs, no data experience
 - **Practical Stats**: math when needed, deep concepts and intuition
 - **Ultimate Goal**: how to do empirical research, in depth
 - **Procedure**: Traverse from theoretical foundations to practical application (including “how to” computations)
- Do you have the background for this class?

Course Strategy

- We could teach you the latest and greatest methods, but when you graduate *they will be old*
- We could teach you all the methods that might prove useful during your career, but when you graduate *you will be old*
- Instead, we teach you the **fundamentals**, the underlying **theory of inference**, from which statistical models are developed
- **Balance of emphasis: proofs v. concepts & intuition**
 - **Baby Stats**: dumbed down proofs, vague intuition
 - **Math Stats**: rigorous mathematical proofs, no data experience
 - **Practical Stats**: math when needed, deep concepts and intuition
 - **Ultimate Goal**: how to do empirical research, in depth
 - **Procedure**: Traverse from theoretical foundations to practical application (including “how to” computations)
- Do you have the background for this class? **A Test: What's this?**

$$b = (X'X)^{-1}X'y$$

The Big Picture

The Big Picture

Just because inference is impossible, doesn't mean you shouldn't do it!

The Big Picture

Just because inference is impossible, doesn't mean you shouldn't do it!

- **Assume:** substantive question of interest

The Big Picture

Just because inference is impossible, doesn't mean you shouldn't do it!

- **Assume:** substantive question of interest
- **Formalize:** quantity of interest (QOI), **given question**

The Big Picture

Just because inference is impossible, doesn't mean you shouldn't do it!

- **Assume:** substantive question of interest
- **Formalize:** quantity of interest (QOI), given question
- **Collect:** data, given QOI, question

The Big Picture

Just because inference is impossible, doesn't mean you shouldn't do it!

- **Assume:** substantive question of interest
- **Formalize:** quantity of interest (QOI), **given question**
- **Collect:** data, **given QOI, question**
- **Assume:** class of models, **given data, QOI, question**

The Big Picture

Just because inference is impossible, doesn't mean you shouldn't do it!

- **Assume:** substantive question of interest
- **Formalize:** quantity of interest (QOI), given question
- **Collect:** data, given QOI, question
- **Assume:** class of models, given data, QOI, question
- **Estimate from data:** best model in class, given of all the above

The Big Picture

Just because inference is impossible, doesn't mean you shouldn't do it!

- **Assume:** substantive question of interest
- **Formalize:** quantity of interest (QOI), given question
- **Collect:** data, given QOI, question
- **Assume:** class of models, given data, QOI, question
- **Estimate from data:** best model in class, given of all the above
- **Present results:** given all of the above

The Big Picture

Just because inference is impossible, doesn't mean you shouldn't do it!

- **Assume:** substantive question of interest
- **Formalize:** quantity of interest (QOI), **given question**
- **Collect:** data, **given QOI, question**
- **Assume:** class of models, **given data, QOI, question**
- **Estimate from data:** best model **in class, given of all the above**
- **Present results:** **given all of the above**
 - QOI Estimates: interpretable by anyone

The Big Picture

Just because inference is impossible, doesn't mean you shouldn't do it!

- **Assume:** substantive question of interest
- **Formalize:** quantity of interest (QOI), **given question**
- **Collect:** data, **given QOI, question**
- **Assume:** class of models, **given data, QOI, question**
- **Estimate from data:** best model **in class, given of all the above**
- **Present results:** **given all of the above**
 - QOI Estimates: interpretable by anyone
 - Uncertainty estimates: CIs, SEs, etc.

The Big Picture

Just because inference is impossible, doesn't mean you shouldn't do it!

- **Assume:** substantive question of interest
- **Formalize:** quantity of interest (QOI), **given question**
- **Collect:** data, **given QOI, question**
- **Assume:** class of models, **given data, QOI, question**
- **Estimate from data:** best model **in class, given of all the above**
- **Present results:** **given all of the above**
 - QOI Estimates: interpretable by anyone
 - Uncertainty estimates: CIs, SEs, etc.
- **Question:**

The Big Picture

Just because inference is impossible, doesn't mean you shouldn't do it!

- **Assume:** substantive question of interest
- **Formalize:** quantity of interest (QOI), given question
- **Collect:** data, given QOI, question
- **Assume:** class of models, given data, QOI, question
- **Estimate from data:** best model in class, given of all the above
- **Present results:** given all of the above
 - QOI Estimates: interpretable by anyone
 - Uncertainty estimates: CIs, SEs, etc.
- **Question:**
 - Better ways of presenting results

The Big Picture

Just because inference is impossible, doesn't mean you shouldn't do it!

- **Assume:** substantive question of interest
- **Formalize:** quantity of interest (QOI), given question
- **Collect:** data, given QOI, question
- **Assume:** class of models, given data, QOI, question
- **Estimate from data:** best model in class, given of all the above
- **Present results:** given all of the above
 - QOI Estimates: interpretable by anyone
 - Uncertainty estimates: CIs, SEs, etc.
- **Question:**
 - Better ways of presenting results
 - Model dependence

The Big Picture

Just because inference is impossible, doesn't mean you shouldn't do it!

- **Assume:** substantive question of interest
- **Formalize:** quantity of interest (QOI), given question
- **Collect:** data, given QOI, question
- **Assume:** class of models, given data, QOI, question
- **Estimate from data:** best model in class, given of all the above
- **Present results:** given all of the above
 - QOI Estimates: interpretable by anyone
 - Uncertainty estimates: CIs, SEs, etc.
- **Question:**
 - Better ways of presenting results
 - Model dependence
 - Data problems

The Big Picture

Just because inference is impossible, doesn't mean you shouldn't do it!

- **Assume:** substantive question of interest
- **Formalize:** quantity of interest (QOI), given question
- **Collect:** data, given QOI, question
- **Assume:** class of models, given data, QOI, question
- **Estimate from data:** best model in class, given of all the above
- **Present results:** given all of the above
 - QOI Estimates: interpretable by anyone
 - Uncertainty estimates: CIs, SEs, etc.
- **Question:**
 - Better ways of presenting results
 - Model dependence
 - Data problems
 - More interesting QOIs

The Big Picture

Just because inference is impossible, doesn't mean you shouldn't do it!

- **Assume:** substantive question of interest
- **Formalize:** quantity of interest (QOI), **given question**
- **Collect:** data, **given QOI, question**
- **Assume:** class of models, **given data, QOI, question**
- **Estimate from data:** best model **in class, given of all the above**
- **Present results:** **given all of the above**
 - QOI Estimates: interpretable by anyone
 - Uncertainty estimates: CIs, SEs, etc.
- **Question:**
 - Better ways of presenting results
 - Model dependence
 - Data problems
 - More interesting QOIs
 - More impactful questions to avoid problems in the first place

[Change of subject:] What is this?

[Change of subject:] What is this?



[Change of subject:] What is this?



- Now you know what a **model** is:

[Change of subject:] What is this?



- Now you know what a **model** is: **It's an abstraction**

[Change of subject:] What is this?



- Now you know what a **model** is: **It's an abstraction**
- Is this model true or false?

[Change of subject:] What is this?



- Now you know what a **model** is: **It's an abstraction**
- Is this model true or false? Crazy question!

[Change of subject:] What is this?



- Now you know what a **model** is: **It's an abstraction**
- Is this model true or false? Crazy question!
- Are models realistic or not?

[Change of subject:] What is this?



- Now you know what a **model** is: **It's an abstraction**
- Is this model true or false? Crazy question!
- Are models realistic or not? OK for features we care about

[Change of subject:] What is this?



- Now you know what a **model** is: **It's an abstraction**
- Is this model true or false? Crazy question!
- Are models realistic or not? OK for features we care about
- Are models useful or not?

[Change of subject:] What is this?



- Now you know what a **model** is: **It's an abstraction**
- Is this model true or false? Crazy question!
- Are models realistic or not? OK for features we care about
- Are models useful or not? Always a good question

[Change of subject:] What is this?



- Now you know what a **model** is: **It's an abstraction**
- Is this model true or false? Crazy question!
- Are models realistic or not? OK for features we care about
- Are models useful or not? Always a good question
- E.g.: Models of dirt on airplanes v. aerodynamics

Overview and Logistics

Statistical Models

Data Generation Processes (with Simulation)

Probability as a Model of the DGP

Statistical Models: Notation

Statistical Models: Notation

- Dependent (or “outcome”) variable

Statistical Models: Notation

- Dependent (or “outcome”) variable
 - Y is $n \times 1$.

Statistical Models: Notation

- Dependent (or “outcome”) variable
 - Y is $n \times 1$.
 - y_i , a number (after we know it)

Statistical Models: Notation

- Dependent (or “outcome”) variable
 - Y is $n \times 1$.
 - y_i , a number (after we know it)
 - Y_i , a random variable (before we know it)

Statistical Models: Notation

- Dependent (or “outcome”) variable
 - Y is $n \times 1$.
 - y_i , a number (after we know it)
 - Y_i , a random variable (before we know it)
 - **Commonly misunderstood**: a “dependent variable” can be

Statistical Models: Notation

- Dependent (or “outcome”) variable
 - Y is $n \times 1$.
 - y_i , a number (after we know it)
 - Y_i , a random variable (before we know it)
 - **Commonly misunderstood**: a “dependent variable” can be
 - a column of numbers in your data set

Statistical Models: Notation

- Dependent (or “outcome”) variable
 - Y is $n \times 1$.
 - y_i , a number (after we know it)
 - Y_i , a random variable (before we know it)
 - **Commonly misunderstood**: a “dependent variable” can be
 - a column of numbers in your data set
 - the random variable *for each unit i* .

Statistical Models: Notation

- Dependent (or “outcome”) variable
 - Y is $n \times 1$.
 - y_i , a number (after we know it)
 - Y_i , a random variable (before we know it)
 - **Commonly misunderstood**: a “dependent variable” can be
 - a column of numbers in your data set
 - the random variable *for each unit i* .
- Explanatory variables

Statistical Models: Notation

- Dependent (or “outcome”) variable
 - Y is $n \times 1$.
 - y_i , a number (after we know it)
 - Y_i , a random variable (before we know it)
 - **Commonly misunderstood**: a “dependent variable” can be
 - a column of numbers in your data set
 - the random variable *for each unit i* .
- Explanatory variables
 - aka “covariates,” “independent,” or “exogenous” variables

Statistical Models: Notation

- Dependent (or “outcome”) variable
 - Y is $n \times 1$.
 - y_i , a number (after we know it)
 - Y_i , a random variable (before we know it)
 - **Commonly misunderstood**: a “dependent variable” can be
 - a column of numbers in your data set
 - the random variable *for each unit i* .
- Explanatory variables
 - aka “covariates,” “independent,” or “exogenous” variables
 - $X = \{x_{ij}\}$ is $n \times k$ (observations by variables)

Statistical Models: Notation

- Dependent (or “outcome”) variable
 - Y is $n \times 1$.
 - y_i , a number (after we know it)
 - Y_i , a random variable (before we know it)
 - **Commonly misunderstood**: a “dependent variable” can be
 - a column of numbers in your data set
 - the random variable *for each unit i* .
- Explanatory variables
 - aka “covariates,” “independent,” or “exogenous” variables
 - $X = \{x_{ij}\}$ is $n \times k$ (observations by variables)
 - A set of columns (variables): $X = \{x_1 \dots, x_k\}$

Statistical Models: Notation

- Dependent (or “outcome”) variable
 - Y is $n \times 1$.
 - y_i , a number (after we know it)
 - Y_i , a random variable (before we know it)
 - **Commonly misunderstood**: a “dependent variable” can be
 - a column of numbers in your data set
 - the random variable *for each unit i* .
- Explanatory variables
 - aka “covariates,” “independent,” or “exogenous” variables
 - $X = \{x_{ij}\}$ is $n \times k$ (observations by variables)
 - A set of columns (variables): $X = \{x_1 \dots, x_k\}$
 - Row (observation) i : $x_i = \{x_{i1}, \dots, x_{ik}\}$

Statistical Models: Notation

- Dependent (or “outcome”) variable
 - Y is $n \times 1$.
 - y_i , a number (after we know it)
 - Y_i , a random variable (before we know it)
 - **Commonly misunderstood**: a “dependent variable” can be
 - a column of numbers in your data set
 - the random variable *for each unit i* .
- Explanatory variables
 - aka “covariates,” “independent,” or “exogenous” variables
 - $X = \{x_{ij}\}$ is $n \times k$ (observations by variables)
 - A set of columns (variables): $X = \{x_1 \dots, x_k\}$
 - Row (observation) i : $x_i = \{x_{i1}, \dots, x_{ik}\}$
 - X is fixed (not random).

Equivalent Linear Regression Notation

Equivalent Linear Regression Notation

- Standard Notation

Equivalent Linear Regression Notation

- Standard Notation

$$Y_i = x_i\beta + \epsilon_i \quad = \text{systematic} + \text{stochastic}$$

Equivalent Linear Regression Notation

- Standard Notation

$$Y_i = x_i\beta + \epsilon_i \quad = \text{systematic} + \text{stochastic}$$
$$\epsilon_i \sim f_N(0, \sigma^2)$$

Equivalent Linear Regression Notation

- Standard Notation

$$Y_i = x_i\beta + \epsilon_i \quad = \text{systematic} + \text{stochastic}$$
$$\epsilon_i \sim f_N(0, \sigma^2)$$

- Alternative Notation

Equivalent Linear Regression Notation

- Standard Notation

$$Y_i = x_i\beta + \epsilon_i \quad = \text{systematic} + \text{stochastic}$$
$$\epsilon_i \sim f_N(0, \sigma^2)$$

- Alternative Notation

$$Y_i \sim f_N(\mu_i, \sigma^2) \quad \text{stochastic}$$

Equivalent Linear Regression Notation

- Standard Notation

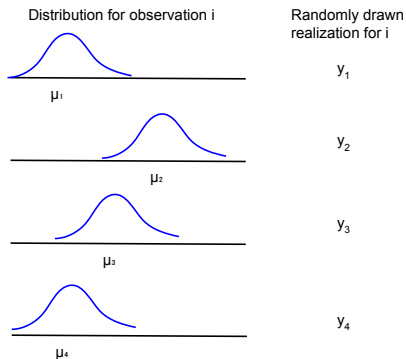
$$Y_i = x_i\beta + \epsilon_i \quad = \text{systematic} + \text{stochastic}$$
$$\epsilon_i \sim f_N(0, \sigma^2)$$

- Alternative Notation

$$Y_i \sim f_N(\mu_i, \sigma^2) \quad \text{stochastic}$$
$$\mu_i = x_i\beta \quad \text{systematic}$$

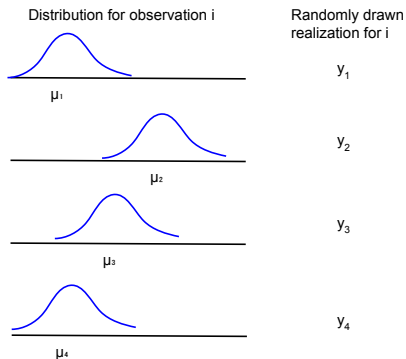
Understanding Alternative Regression Notation

Understanding Alternative Regression Notation



where $\mu_i = X_i\beta$.

Understanding Alternative Regression Notation



where $\mu_i = X_i\beta$.

Quiz: Is a histogram of y a test of normality?

Generalized Alternative Notation for Most Models

Generalized Alternative Notation for Most Models

$$Y_i \sim f(\theta_i, \alpha)$$

stochastic

Generalized Alternative Notation for Most Models

$$Y_i \sim f(\theta_i, \alpha)$$

stochastic

$$\theta_i = g(X_i, \beta)$$

systematic

Generalized Alternative Notation for Most Models

$$Y_i \sim f(\theta_i, \alpha)$$

stochastic

$$\theta_i = g(X_i, \beta)$$

systematic

where

Generalized Alternative Notation for Most Models

$$Y_i \sim f(\theta_i, \alpha) \quad \text{stochastic}$$

$$\theta_i = g(X_i, \beta) \quad \text{systematic}$$

where

Y_i random outcome variable

Generalized Alternative Notation for Most Models

$$Y_i \sim f(\theta_i, \alpha) \quad \text{stochastic}$$

$$\theta_i = g(X_i, \beta) \quad \text{systematic}$$

where

Y_i random outcome variable

$f(\cdot)$ probability density

Generalized Alternative Notation for Most Models

$$Y_i \sim f(\theta_i, \alpha) \quad \text{stochastic}$$

$$\theta_i = g(X_i, \beta) \quad \text{systematic}$$

where

Y_i random outcome variable

$f(\cdot)$ probability density

θ_i a systematic feature of the density that varies over i

Generalized Alternative Notation for Most Models

$$Y_i \sim f(\theta_i, \alpha) \quad \text{stochastic}$$

$$\theta_i = g(X_i, \beta) \quad \text{systematic}$$

where

Y_i random outcome variable

$f(\cdot)$ probability density

θ_i a systematic feature of the density that varies over i

α ancillary parameter (feature of the density constant over i)

Generalized Alternative Notation for Most Models

$$Y_i \sim f(\theta_i, \alpha) \quad \text{stochastic}$$

$$\theta_i = g(X_i, \beta) \quad \text{systematic}$$

where

Y_i random outcome variable

$f(\cdot)$ probability density

θ_i a systematic feature of the density that varies over i

α ancillary parameter (feature of the density constant over i)

$g(\cdot)$ functional form

Generalized Alternative Notation for Most Models

$$Y_i \sim f(\theta_i, \alpha) \quad \text{stochastic}$$

$$\theta_i = g(\textcolor{blue}{X}_i, \beta) \quad \text{systematic}$$

where

Y_i random outcome variable

$f(\cdot)$ probability density

θ_i a systematic feature of the density that varies over i

α ancillary parameter (feature of the density constant over i)

$g(\cdot)$ functional form

$\textcolor{blue}{X}_i$ explanatory variables

Generalized Alternative Notation for Most Models

$$Y_i \sim f(\theta_i, \alpha) \quad \text{stochastic}$$

$$\theta_i = g(X_i, \beta) \quad \text{systematic}$$

where

Y_i random outcome variable

$f(\cdot)$ probability density

θ_i a systematic feature of the density that varies over i

α ancillary parameter (feature of the density constant over i)

$g(\cdot)$ functional form

X_i explanatory variables

β effect parameters

Forms of Uncertainty

Forms of Uncertainty

$$Y_i \sim f(\theta_i, \alpha)$$

stochastic

Forms of Uncertainty

$$Y_i \sim f(\theta_i, \alpha)$$

stochastic

$$\theta_i = g(X_i, \beta)$$

systematic

Forms of Uncertainty

$$Y_i \sim f(\theta_i, \alpha)$$

stochastic

$$\theta_i = g(X_i, \beta)$$

systematic

- **Estimation uncertainty:** Lack of knowledge of β and α .
Vanishes as n gets larger.

Forms of Uncertainty

$$Y_i \sim f(\theta_i, \alpha)$$

stochastic

$$\theta_i = g(X_i, \beta)$$

systematic

- Estimation uncertainty: Lack of knowledge of β and α . Vanishes as n gets larger.
- **Fundamental uncertainty**: Represented by the stochastic component. (Exists no matter what the researcher does; no matter how large n is.)

Forms of Uncertainty

$$Y_i \sim f(\theta_i, \alpha)$$

stochastic

$$\theta_i = g(X_i, \beta)$$

systematic

- Estimation uncertainty: Lack of knowledge of β and α .
Vanishes as n gets larger.
- Fundamental uncertainty: Represented by the stochastic component. (Exists no matter what the researcher does; no matter how large n is.)
- **Model dependence**: Maybe the whole specification is wrong?

Forms of Uncertainty

$$Y_i \sim f(\theta_i, \alpha)$$

stochastic

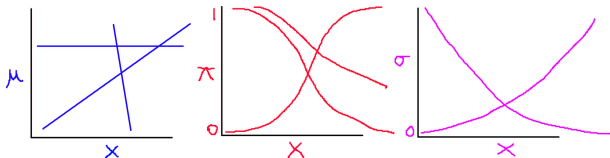
$$\theta_i = g(X_i, \beta)$$

systematic

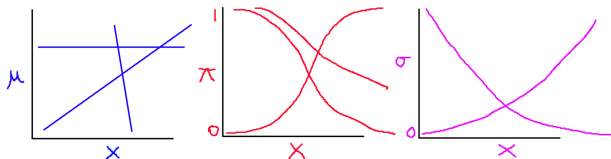
- Estimation uncertainty: Lack of knowledge of β and α .
Vanishes as n gets larger.
- Fundamental uncertainty: Represented by the stochastic component. (Exists no matter what the researcher does; no matter how large n is.)
- Model dependence: Maybe the whole specification is wrong?
- Quiz: If you know the model, including α and β , is $R^2 = 1$?

Systematic Components: Examples

Systematic Components: Examples

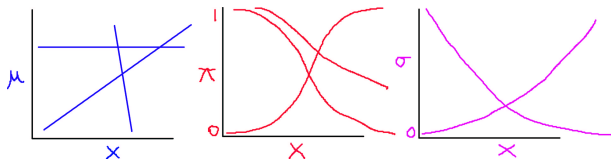


Systematic Components: Examples



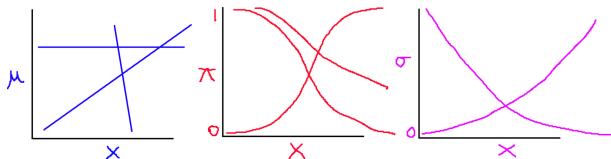
- $E(Y_i) \equiv \mu_i = X_i\beta = \beta_0 + \beta_1 X_{1i} + \dots + \beta_k X_{ki}$

Systematic Components: Examples



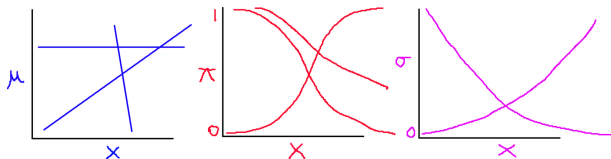
- $E(Y_i) \equiv \mu_i = X_i\beta = \beta_0 + \beta_1 X_{1i} + \dots + \beta_k X_{ki}$
- $\Pr(Y_i = 1) \equiv \pi_i = \frac{1}{1+e^{-x_i\beta}}$

Systematic Components: Examples



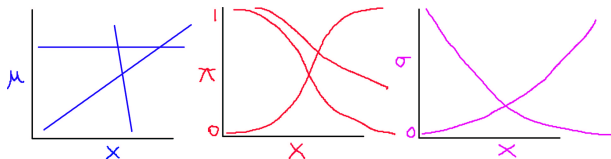
- $E(Y_i) \equiv \mu_i = X_i\beta = \beta_0 + \beta_1 X_{1i} + \dots + \beta_k X_{ki}$
- $\Pr(Y_i = 1) \equiv \pi_i = \frac{1}{1+e^{-x_i\beta}}$
- $V(Y_i) \equiv \sigma_i^2 = e^{x_i\beta}$

Systematic Components: Examples



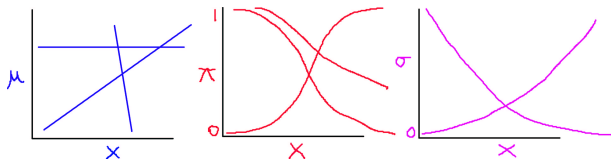
- $E(Y_i) \equiv \mu_i = X_i\beta = \beta_0 + \beta_1 X_{1i} + \dots + \beta_k X_{ki}$
- $\Pr(Y_i = 1) \equiv \pi_i = \frac{1}{1+e^{-x_i\beta}}$
- $V(Y_i) \equiv \sigma_i^2 = e^{x_i\beta}$
- Interpretation:

Systematic Components: Examples



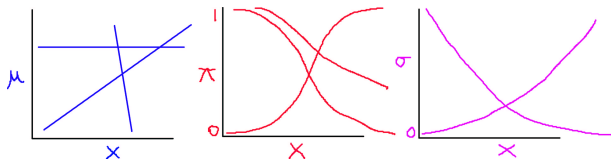
- $E(Y_i) \equiv \mu_i = X_i\beta = \beta_0 + \beta_1 X_{1i} + \dots + \beta_k X_{ki}$
- $\Pr(Y_i = 1) \equiv \pi_i = \frac{1}{1+e^{-x_i\beta}}$
- $V(Y_i) \equiv \sigma_i^2 = e^{x_i\beta}$
- Interpretation:
 - Each is a class of functional forms

Systematic Components: Examples



- $E(Y_i) \equiv \mu_i = X_i\beta = \beta_0 + \beta_1 X_{1i} + \dots + \beta_k X_{ki}$
- $\Pr(Y_i = 1) \equiv \pi_i = \frac{1}{1+e^{-x_i\beta}}$
- $V(Y_i) \equiv \sigma_i^2 = e^{x_i\beta}$
- Interpretation:
 - Each is a **class of functional forms**
 - β in each is an “effect parameter” vector, with a different meaning

Systematic Components: Examples



- $E(Y_i) \equiv \mu_i = X_i\beta = \beta_0 + \beta_1 X_{1i} + \dots + \beta_k X_{ki}$
- $\Pr(Y_i = 1) \equiv \pi_i = \frac{1}{1+e^{-x_i\beta}}$
- $V(Y_i) \equiv \sigma_i^2 = e^{x_i\beta}$
- Interpretation:
 - Each is a **class of functional forms**
 - β in each is an “effect parameter” vector, with a different meaning
 - Set β and it picks out one **member of the class**

Systematic Components: Examples

Systematic Components: Examples

- Standard procedure:

Systematic Components: Examples

- Standard procedure:
 - Use theory: Assume a class of functional forms (that delineate all *remaining* possible relationships, indexed by parameters)

Systematic Components: Examples

- Standard procedure:
 - Use theory: Assume a class of functional forms (that delineate all *remaining* possible relationships, indexed by parameters)
 - Use data: Estimate the values of the parameters (to pick out one member of the class)

Systematic Components: Examples

- **Standard procedure:**
 - **Use theory:** Assume a class of functional forms (that delineate all *remaining* possible relationships, indexed by parameters)
 - **Use data:** Estimate the values of the parameters (to pick out one member of the class)
 - **Remain uncertain:** Because of estimation uncertainty, fundamental uncertainty, model dependence

Systematic Components: Examples

- Standard procedure:
 - Use theory: Assume a class of functional forms (that delineate all *remaining* possible relationships, indexed by parameters)
 - Use data: Estimate the values of the parameters (to pick out one member of the class)
 - Remain uncertain: Because of estimation uncertainty, fundamental uncertainty, model dependence
- If we choose the family of functional forms wrong, we:

Systematic Components: Examples

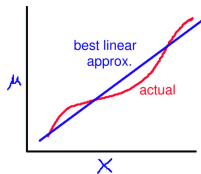
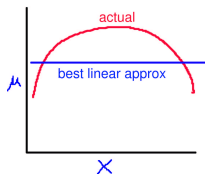
- **Standard procedure:**
 - **Use theory:** Assume a class of functional forms (that delineate all *remaining* possible relationships, indexed by parameters)
 - **Use data:** Estimate the values of the parameters (to pick out one member of the class)
 - **Remain uncertain:** Because of estimation uncertainty, fundamental uncertainty, model dependence
- **If we choose the family of functional forms wrong, we:**
 - Have **specification error**, and potentially bias

Systematic Components: Examples

- **Standard procedure:**
 - **Use theory:** Assume a class of functional forms (that delineate all *remaining* possible relationships, indexed by parameters)
 - **Use data:** Estimate the values of the parameters (to pick out one member of the class)
 - **Remain uncertain:** Because of estimation uncertainty, fundamental uncertainty, model dependence
- **If we choose the family of functional forms wrong, we:**
 - Have **specification error**, and potentially bias
 - Still get the **best [linear,logit,etc] approximation** to the correct functional form.

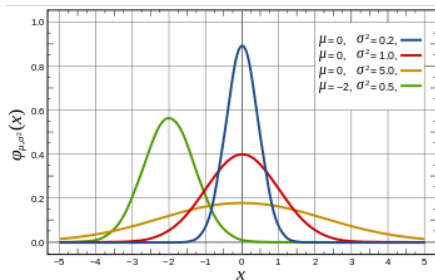
Systematic Components: Examples

- **Standard procedure:**
 - **Use theory:** Assume a class of functional forms (that delineate all *remaining* possible relationships, indexed by parameters)
 - **Use data:** Estimate the values of the parameters (to pick out one member of the class)
 - **Remain uncertain:** Because of estimation uncertainty, fundamental uncertainty, model dependence
- **If we choose the family of functional forms wrong, we:**
 - Have **specification error**, and potentially bias
 - Still get the **best [linear,logit,etc] approximation** to the correct functional form.
 - May be **close or far from the truth:**



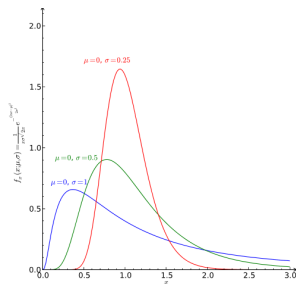
Overview of Stochastic Components

Overview of Stochastic Components



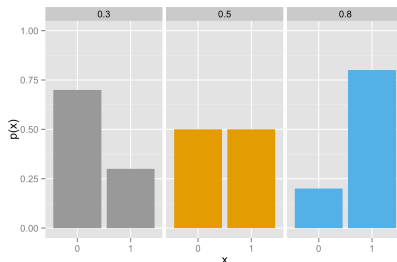
- Normal — continuous, unimodal, symmetric, unbounded

Overview of Stochastic Components



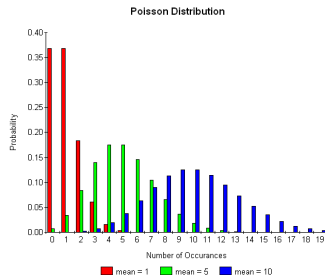
- Normal — continuous, unimodal, symmetric, unbounded
- Log-normal — continuous, unimodal, skewed, bounded from below by zero

Overview of Stochastic Components



- Normal — continuous, unimodal, symmetric, unbounded
- Log-normal — continuous, unimodal, skewed, bounded from below by zero
- Bernoulli — discrete, binary outcomes

Overview of Stochastic Components



- Normal — continuous, unimodal, symmetric, unbounded
- Log-normal — continuous, unimodal, skewed, bounded from below by zero
- Bernoulli — discrete, binary outcomes
- Poisson — discrete, countably infinite on the nonnegative integers

Choosing systematic and stochastic components

Choosing systematic and stochastic components

- If one is bounded, so is the other

Choosing systematic and stochastic components

- If one is bounded, so is the other
- If the stochastic component is bounded, the systematic component must be globally nonlinear (tho possibly locally linear)

Choosing systematic and stochastic components

- If one is bounded, so is the other
- If the stochastic component is bounded, the systematic component must be globally nonlinear (tho possibly locally linear)
- All modeling decisions are about the **data generation process** — from how the world produces data to how we observe it

Choosing systematic and stochastic components

- If one is bounded, so is the other
- If the stochastic component is bounded, the systematic component must be globally nonlinear (tho possibly locally linear)
- All modeling decisions are about the **data generation process** — from how the world produces data to how we observe it
- **What if we don't know the DGP (& we usually don't)?**

Choosing systematic and stochastic components

- If one is bounded, so is the other
- If the stochastic component is bounded, the systematic component must be globally nonlinear (tho possibly locally linear)
- All modeling decisions are about the **data generation process** — from how the world produces data to how we observe it
- **What if we don't know the DGP (& we usually don't)?**
 - **The problem:** model dependence

Choosing systematic and stochastic components

- If one is bounded, so is the other
- If the stochastic component is bounded, the systematic component must be globally nonlinear (tho possibly locally linear)
- All modeling decisions are about the **data generation process** — from how the world produces data to how we observe it
- **What if we don't know the DGP (& we usually don't)?**
 - **The problem:** model dependence
 - **Our first approach:** make “reasonable” assumptions and check fit (& other observable implications of the assumptions)

Choosing systematic and stochastic components

- If one is bounded, so is the other
- If the stochastic component is bounded, the systematic component must be globally nonlinear (tho possibly locally linear)
- All modeling decisions are about the **data generation process** — from how the world produces data to how we observe it
- **What if we don't know the DGP (& we usually don't)?**
 - **The problem:** model dependence
 - **Our first approach:** make “reasonable” assumptions and check fit (& other observable implications of the assumptions)
 - **Later:**

Choosing systematic and stochastic components

- If one is bounded, so is the other
- If the stochastic component is bounded, the systematic component must be globally nonlinear (tho possibly locally linear)
- All modeling decisions are about the **data generation process** — from how the world produces data to how we observe it
- **What if we don't know the DGP (& we usually don't)?**
 - **The problem:** model dependence
 - **Our first approach:** make “reasonable” assumptions and check fit (& other observable implications of the assumptions)
 - **Later:**
 - Avoid it: relax assumptions (functional form, distribution, etc)

Choosing systematic and stochastic components

- If one is bounded, so is the other
- If the stochastic component is bounded, the systematic component must be globally nonlinear (tho possibly locally linear)
- All modeling decisions are about the **data generation process** — from how the world produces data to how we observe it
- **What if we don't know the DGP (& we usually don't)?**
 - **The problem:** model dependence
 - **Our first approach:** make “reasonable” assumptions and check fit (& other observable implications of the assumptions)
 - **Later:**
 - Avoid it: relax assumptions (functional form, distribution, etc)
 - Detect remaining model dependence

Choosing systematic and stochastic components

- If one is bounded, so is the other
- If the stochastic component is bounded, the systematic component must be globally nonlinear (tho possibly locally linear)
- All modeling decisions are about the **data generation process** — from how the world produces data to how we observe it
- **What if we don't know the DGP (& we usually don't)?**
 - **The problem:** model dependence
 - **Our first approach:** make “reasonable” assumptions and check fit (& other observable implications of the assumptions)
 - **Later:**
 - Avoid it: relax assumptions (functional form, distribution, etc)
 - Detect remaining model dependence
 - Remove model dependence: preprocess data (via matching, etc.)

Overview and Logistics

Statistical Models

Data Generation Processes (with Simulation)

Probability as a Model of the DGP

What's the probability that these goats are real?

What's the probability that these goats are real?



What's the probability that these goats are real?



What's the probability that these goats are real?



[Hey! What are you looking at?]

What's the probability that these goats are real?



[Hey! What are you looking at?]

- How I can convince you the goats are real?

What's the probability that these goats are real?



[Hey! What are you looking at?]

- How I can convince you the goats are real?
 - I took the photos myself

What's the probability that these goats are real?



[Hey! What are you looking at?]

- How I can convince you the goats are real?
 - I took the photos myself (OK, but you have to trust me)

What's the probability that these goats are real?



[Hey! What are you looking at?]

- How I can convince you the goats are real?
 - I took the photos myself (OK, but you have to trust me)
 - On a family vacation

What's the probability that these goats are real?



[Hey! What are you looking at?]

- How I can convince you the goats are real?
 - I took the photos myself (OK, but you have to trust me)
 - On a family vacation (The story helps some)

What's the probability that these goats are real?



[Hey! What are you looking at?]

- How I can convince you the goats are real?
 - I took the photos myself (OK, but you have to trust me)
 - On a family vacation (The story helps some)
 - Here's some evidence

What's the probability that these goats are real?



[Hey! What are you looking at?]

- How I can convince you the goats are real?
 - I took the photos myself (OK, but you have to trust me)
 - On a family vacation (The story helps some)
 - Here's some evidence (How good is Gary at photoshop?)

What's the probability that these goats are real?



[Hey! What are you looking at?]

- How I can convince you the goats are real?
 - I took the photos myself (OK, but you have to trust me)
 - On a family vacation (The story helps some)
 - Here's some evidence (How good is Gary at photoshop?)
 - Would it help you if I told you why the goats are in the trees?

What's the probability that these goats are real?



[Hey! What are you looking at?]

- How I can convince you the goats are real?
 - I took the photos myself (OK, but you have to trust me)
 - On a family vacation (The story helps some)
 - Here's some **evidence** (How good is Gary at photoshop?)
 - Would it help you if I told you **why** the goats are in the trees?
- Key points:

What's the probability that these goats are real?



[Hey! What are you looking at?]

- How I can convince you the goats are real?
 - I took the photos myself (OK, but you have to trust me)
 - On a family vacation (The story helps some)
 - Here's some **evidence** (How good is Gary at photoshop?)
 - Would it help you if I told you **why** the goats are in the trees?
- Key points:
 - DGP Uncertainty $\rightsquigarrow$ Photo (Substantive) Uncertainty

What's the probability that these goats are real?



[Hey! What are you looking at?]

- How I can convince you the goats are real?
 - I took the photos myself (OK, but you have to trust me)
 - On a family vacation (The story helps some)
 - Here's some evidence (How good is Gary at photoshop?)
 - Would it help you if I told you why the goats are in the trees?
- Key points:
 - DGP Uncertainty $\rightsquigarrow$ Photo (Substantive) Uncertainty
 - For *any* data to be useful, you need to know the DGP

What's the probability that these goats are real?



[Hey! What are you looking at?]

- How I can convince you the goats are real?
 - I took the photos myself (OK, but you have to trust me)
 - On a family vacation (The story helps some)
 - Here's some **evidence** (How good is Gary at photoshop?)
 - Would it help you if I told you **why** the goats are in the trees?
- Key points:
 - DGP Uncertainty $\rightsquigarrow$ Photo (Substantive) Uncertainty
 - For *any* data to be useful, you need to know the DGP
 - Theory is helpful for data; data is helpful for theory

What's the probability that these goats are real?



[Hey! What are you looking at?]

- How I can convince you the goats are real?
 - I took the photos myself (OK, but you have to trust me)
 - On a family vacation (The story helps some)
 - Here's some evidence (How good is Gary at photoshop?)
 - Would it help you if I told you why the goats are in the trees?
- Key points:
 - DGP Uncertainty $\rightsquigarrow$ Photo (Substantive) Uncertainty
 - For *any* data to be useful, you need to know the DGP
 - Theory is helpful for data; data is helpful for theory
- The Tools of DGPs

What's the probability that these goats are real?



[Hey! What are you looking at?]

- How I can convince you the goats are real?
 - I took the photos myself (OK, but you have to trust me)
 - On a family vacation (The story helps some)
 - Here's some **evidence** (How good is Gary at photoshop?)
 - Would it help you if I told you **why** the goats are in the trees?
- Key points:
 - DGP Uncertainty $\rightsquigarrow$ Photo (Substantive) Uncertainty
 - For *any* data to be useful, you need to know the DGP
 - Theory is helpful for data; data is helpful for theory
- The Tools of DGPs
 - **Probability**: Assumes the DGP

What's the probability that these goats are real?



[Hey! What are you looking at?]

- How I can convince you the goats are real?
 - I took the photos myself (OK, but you have to trust me)
 - On a family vacation (The story helps some)
 - Here's some **evidence** (How good is Gary at photoshop?)
 - Would it help you if I told you **why** the goats are in the trees?
- Key points:
 - DGP Uncertainty $\rightsquigarrow$ Photo (Substantive) Uncertainty
 - For *any* data to be useful, you need to know the DGP
 - Theory is helpful for data; data is helpful for theory
- The Tools of DGPs
 - **Probability**: Assumes the DGP
 - **Statistical Inference**: Learns the DGP

Simulation is used to:

Simulation is used to:

1. Understand the assumed DGP

Simulation is used to:

1. Understand the assumed DGP
2. Solve probability problems

Simulation is used to:

1. Understand the assumed DGP
2. Solve probability problems
3. Evaluate estimators

Simulation is used to:

1. Understand the assumed DGP
2. Solve probability problems
3. Evaluate estimators
4. Calculate features of probability densities

Simulation is used to:

1. Understand the assumed DGP
2. Solve probability problems
3. Evaluate estimators
4. Calculate features of probability densities
5. Transform statistical results into quantities of interest

Simulation is used to:

1. Understand the assumed DGP
2. Solve probability problems
3. Evaluate estimators
4. Calculate features of probability densities
5. Transform statistical results into quantities of interest
6. Get the right answer: easier than mathematical calculations

What is simulation?

Survey Sampling

Simulation

What is simulation?

Survey Sampling

Simulation

What is simulation?

Survey Sampling

1. Learn about a population by taking a random sample from it

Simulation

1. Learn about a distribution by taking random draws from it

What is simulation?

Survey Sampling

1. Learn about a population by taking a random sample from it
2. Use the random sample to estimate a feature of the population

Simulation

1. Learn about a distribution by taking random draws from it
2. Use the random draws to approximate a feature of the distribution

What is simulation?

Survey Sampling

1. Learn about a population by taking a random sample from it
2. Use the random sample to estimate a feature of the population
3. The estimate is arbitrarily precise for large n

Simulation

1. Learn about a distribution by taking random draws from it
2. Use the random draws to approximate a feature of the distribution
3. The approximation is arbitrarily precise for large M

What is simulation?

Survey Sampling

1. Learn about a population by taking a random sample from it
2. Use the random sample to estimate a feature of the population
3. The estimate is arbitrarily precise for large n
4. Example: estimate the mean of the population

Simulation

1. Learn about a distribution by taking random draws from it
2. Use the random draws to approximate a feature of the distribution
3. The approximation is arbitrarily precise for large M
4. Example: Approximate the mean of the distribution

Monte Hall's Let's Make a Deal



Monte Hall's Let's Make a Deal



You choose among 3 doors. Behind one is a **car**; behind two are **goats**. You choose one at random. Monte peeks at the 2 remaining and opens the one (or one of the two) with the **goat** and asks if you'll switch to the door that hasn't been opened yet. **Should you switch?**

Monte Hall's Let's Make a Deal



You choose among 3 doors. Behind one is a **car**; behind two are **goats**. You choose one at random. Monte peeks at the 2 remaining and opens the one (or one of the two) with the **goat** and asks if you'll switch to the door that hasn't been opened yet. **Should you switch?**

```
sims<-1000; WinNoSwitch<-0; WinSwitch<-0; doors<-c(1,2,3)
for (i in 1:sims) {
  WinDoor <- sample(doors, 1)
  choice <- sample(doors, 1)
  if (WinDoor == choice)                                # no switch
    WinNoSwitch <- WinNoSwitch + 1
  doorsLeft <- doors[doors != choice]                   # switch
  if (any(doorLeft == WinDoor))
    WinSwitch <- WinSwitch + 1
}
cat("Prob(Car | no switch)=", WinNoSwitch/sims, "\\n")
cat("Prob(Car | switch)=", WinSwitch/sims, "\\n")
```

Monte Hall's Let's Make a Deal

$\text{Pr}(\text{car} \text{No Switch})$	$\text{Pr}(\text{car} \text{Switch})$
.324	.676
.345	.655
.320	.680
.327	.673

The Birthday Problem

The Birthday Problem

Given a room with 24 randomly selected people, what is the probability that at least two have the same birthday?

The Birthday Problem

Given a room with 24 randomly selected people, what is the probability that at least two have the same birthday?

```
sims <- 1000
people <- 24
alldays <- seq(1, 365, 1)
sameday <- 0
for (i in 1:sims) {
  room <- sample(alldays, people, replace = TRUE)
  if (length(unique(room)) < people) \# same birthday
    sameday <- sameday+1
}

cat("Probability >=2:", sameday/sims, "\\n")
```

The Birthday Problem

Given a room with 24 randomly selected people, what is the probability that at least two have the same birthday?

```
sims <- 1000
people <- 24
alldays <- seq(1, 365, 1)
sameday <- 0
for (i in 1:sims) {
  room <- sample(alldays, people, replace = TRUE)
  if (length(unique(room)) < people) \# same birthday
    sameday <- sameday+1
}

cat("Probability >=2:", sameday/sims, "\n\n")
```

Four runs: 0.538, 0.550, 0.547, 0.524

Overview and Logistics

Statistical Models

Data Generation Processes (with Simulation)

Probability as a Model of the DGP

Probability

Probability

- A function $\Pr(y|M) \equiv \Pr(\text{data}|\text{Model})$, where $M = (f, g, X, \beta, \alpha)$.

Probability

- A function $\Pr(y|M) \equiv \Pr(\text{data}|\text{Model})$, where $M = (f, g, X, \beta, \alpha)$.
- 3 axioms define the function $\Pr(\cdot|\cdot)$

Probability

- A function $\Pr(y|M) \equiv \Pr(\text{data}|\text{Model})$, where $M = (f, g, X, \beta, \alpha)$.
- 3 axioms define the function $\Pr(\cdot|\cdot)$
 1. $\Pr(z) \geq 0$ for some event z

Probability

- A function $\Pr(y|M) \equiv \Pr(\text{data}|\text{Model})$, where $M = (f, g, X, \beta, \alpha)$.
- 3 axioms define the function $\Pr(\cdot|\cdot)$
 1. $\Pr(z) \geq 0$ for some event z
 2. $\Pr(\text{sample space}) = 1$

Probability

- A function $\Pr(y|M) \equiv \Pr(\text{data}|\text{Model})$, where $M = (f, g, X, \beta, \alpha)$.
- 3 axioms define the function $\Pr(\cdot|\cdot)$
 1. $\Pr(z) \geq 0$ for some event z
 2. $\Pr(\text{sample space}) = 1$
 3. If $z_1, \dots, z_k$ are mutually exclusive events,

Probability

- A function $\Pr(y|M) \equiv \Pr(\text{data}|\text{Model})$, where $M = (f, g, X, \beta, \alpha)$.
- 3 axioms define the function $\Pr(\cdot|\cdot)$
 1. $\Pr(z) \geq 0$ for some event z
 2. $\Pr(\text{sample space}) = 1$
 3. If $z_1, \dots, z_k$ are mutually exclusive events,

$$\Pr(z_1 \cup \dots \cup z_k) = \Pr(z_1) + \dots + \Pr(z_k),$$

Probability

- A function $\Pr(y|M) \equiv \Pr(\text{data}|\text{Model})$, where $M = (f, g, X, \beta, \alpha)$.
- 3 axioms define the function $\Pr(\cdot|\cdot)$
 1. $\Pr(z) \geq 0$ for some event z
 2. $\Pr(\text{sample space}) = 1$
 3. If $z_1, \dots, z_k$ are mutually exclusive events,

$$\Pr(z_1 \cup \dots \cup z_k) = \Pr(z_1) + \dots + \Pr(z_k),$$

- 1 & 2 imply $0 \leq \Pr(z) \leq 1$

Probability

- A function $\Pr(y|M) \equiv \Pr(\text{data}|\text{Model})$, where $M = (f, g, X, \beta, \alpha)$.
- 3 axioms define the function $\Pr(\cdot|\cdot)$
 1. $\Pr(z) \geq 0$ for some event z
 2. $\Pr(\text{sample space}) = 1$
 3. If $z_1, \dots, z_k$ are mutually exclusive events,

$$\Pr(z_1 \cup \dots \cup z_k) = \Pr(z_1) + \dots + \Pr(z_k),$$

- 1& 2 imply $0 \leq \Pr(z) \leq 1$
- Axioms are not assumptions; they can't be wrong.

Probability

- A function $\Pr(y|M) \equiv \Pr(\text{data}|\text{Model})$, where $M = (f, g, X, \beta, \alpha)$.
- 3 axioms define the function $\Pr(\cdot|\cdot)$
 1. $\Pr(z) \geq 0$ for some event z
 2. $\Pr(\text{sample space}) = 1$
 3. If $z_1, \dots, z_k$ are mutually exclusive events,

$$\Pr(z_1 \cup \dots \cup z_k) = \Pr(z_1) + \dots + \Pr(z_k),$$

- 1& 2 imply $0 \leq \Pr(z) \leq 1$
- Axioms are not assumptions; they can't be wrong.
- From the axioms come **all** rules of probability theory.

Probability

- A function $\Pr(y|M) \equiv \Pr(\text{data}|\text{Model})$, where $M = (f, g, X, \beta, \alpha)$.
- 3 axioms define the function $\Pr(\cdot|\cdot)$
 1. $\Pr(z) \geq 0$ for some event z
 2. $\Pr(\text{sample space}) = 1$
 3. If $z_1, \dots, z_k$ are mutually exclusive events,

$$\Pr(z_1 \cup \dots \cup z_k) = \Pr(z_1) + \dots + \Pr(z_k),$$

- 1& 2 imply $0 \leq \Pr(z) \leq 1$
- Axioms are not assumptions; they can't be wrong.
- From the axioms come **all** rules of probability theory.
- **Quiz:** what happens if $\Pr(\text{sample space}) \neq 1$

Probability

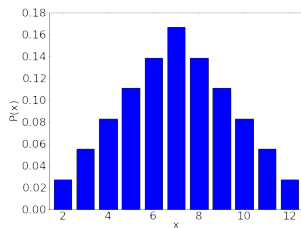
- A function $\Pr(y|M) \equiv \Pr(\text{data}|\text{Model})$, where $M = (f, g, X, \beta, \alpha)$.
- 3 axioms define the function $\Pr(\cdot|\cdot)$
 1. $\Pr(z) \geq 0$ for some event z
 2. $\Pr(\text{sample space}) = 1$
 3. If $z_1, \dots, z_k$ are mutually exclusive events,

$$\Pr(z_1 \cup \dots \cup z_k) = \Pr(z_1) + \dots + \Pr(z_k),$$

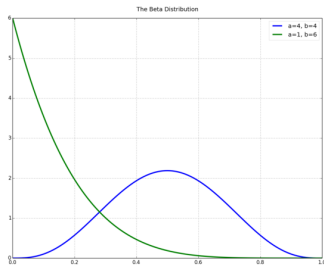
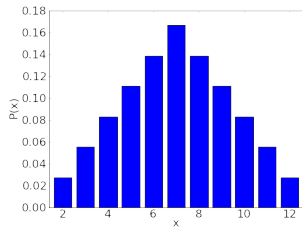
- 1& 2 imply $0 \leq \Pr(z) \leq 1$
- Axioms are not assumptions; they can't be wrong.
- From the axioms come **all** rules of probability theory.
- **Quiz:** what happens if $\Pr(\text{sample space}) \neq 1$
- Rules can be applied **analytically** or via **simulation**.

PDFs: Probability Density Functions

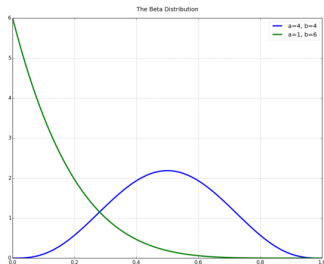
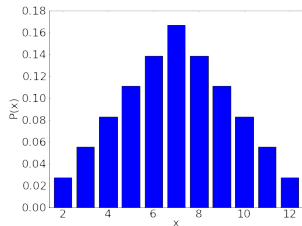
PDFs: Probability Density Functions



PDFs: Probability Density Functions

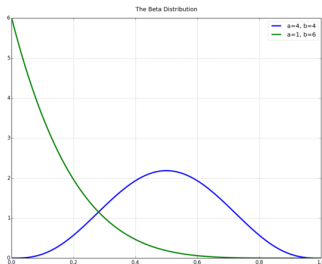
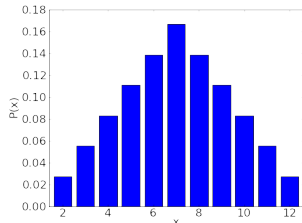


PDFs: Probability Density Functions



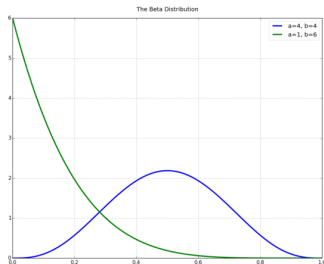
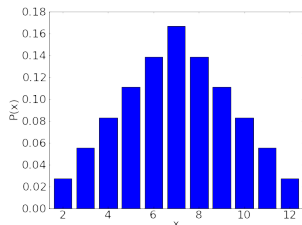
- defined for any y (outcome of the experiment)

PDFs: Probability Density Functions



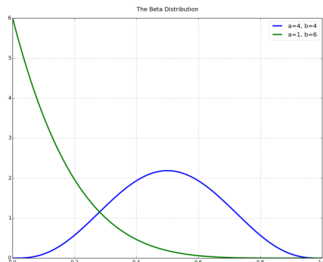
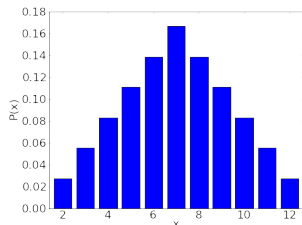
- defined for any y (outcome of the experiment)
- assigns probability to every possible y (or range of y)

PDFs: Probability Density Functions



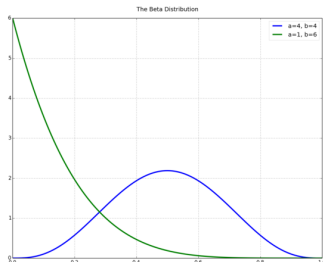
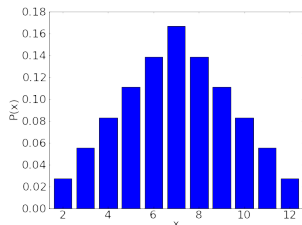
- defined for any y (outcome of the experiment)
- assigns probability to every possible y (or range of y)
- a function, $P(y)$ or $f(y)$, such that

PDFs: Probability Density Functions



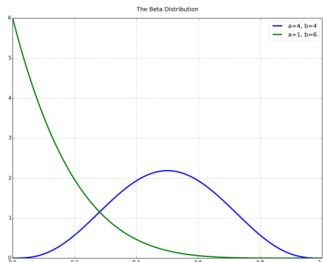
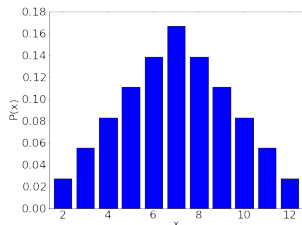
- defined for any y (outcome of the experiment)
- assigns probability to every possible y (or range of y)
- a function, $P(y)$ or $f(y)$, such that
 - $P(y) \geq 0$ for any y

PDFs: Probability Density Functions



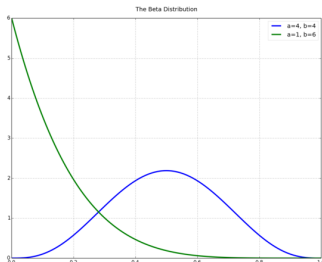
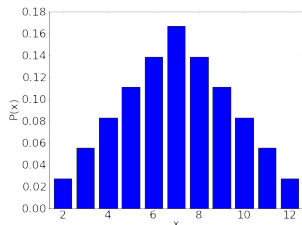
- defined for any y (outcome of the experiment)
- assigns probability to every possible y (or range of y)
- a function, $P(y)$ or $f(y)$, such that
 - $P(y) \geq 0$ for any y
 - for discrete y : $\sum_{\text{all } y} P(y) = 1$

PDFs: Probability Density Functions



- defined for any y (outcome of the experiment)
- assigns probability to every possible y (or range of y)
- a function, $P(y)$ or $f(y)$, such that
 - $P(y) \geq 0$ for any y
 - for discrete y : $\sum_{\text{all } y} P(y) = 1$
 - for continuous y : $\int_{-\infty}^{\infty} P(y) dy = 1$

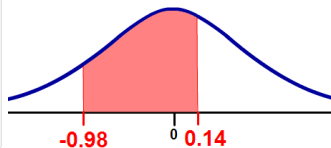
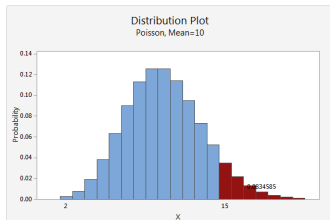
PDFs: Probability Density Functions



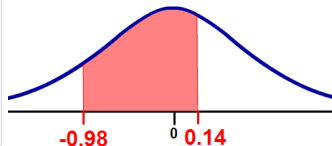
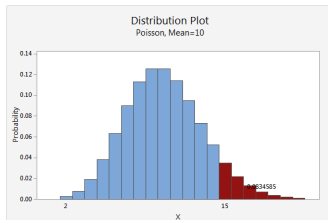
- defined for any y (outcome of the experiment)
- assigns probability to every possible y (or range of y)
- a function, $P(y)$ or $f(y)$, such that
 - $P(y) \geq 0$ for any y
 - for discrete y : $\sum_{\text{all } y} P(y) = 1$
 - for continuous y : $\int_{-\infty}^{\infty} P(y) dy = 1$
- Quiz: Are the curves above PDFs?

Computing Probabilities from PDFs

Computing Probabilities from PDFs



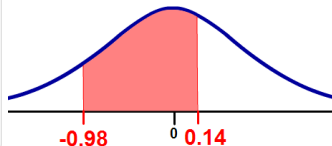
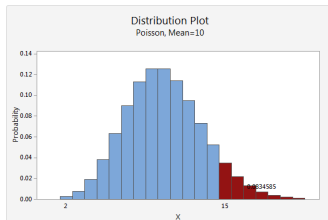
Computing Probabilities from PDFs



•

$$\Pr(a \leq Y \leq b) = \begin{cases} \sum_{a \leq y \leq b} P(y) & \text{discrete} \\ \int_a^b P(y) dy & \text{continuous} \end{cases}$$

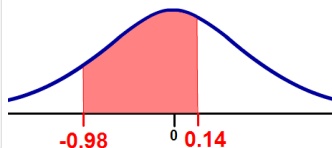
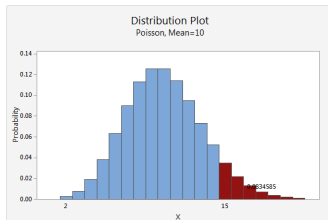
Computing Probabilities from PDFs



$$\Pr(a \leq Y \leq b) = \begin{cases} \sum_{a \leq y \leq b} P(y) & \text{discrete} \\ \int_a^b P(y) dy & \text{continuous} \end{cases}$$

$$\Pr(Y = y) = \begin{cases} P(y) & \text{discrete} \\ 0 & \text{continuous} \end{cases}$$

Computing Probabilities from PDFs



-

$$\Pr(a \leq Y \leq b) = \begin{cases} \sum_{a \leq y \leq b} P(y) & \text{discrete} \\ \int_a^b P(y) dy & \text{continuous} \end{cases}$$

-

$$\Pr(Y = y) = \begin{cases} P(y) & \text{discrete} \\ 0 & \text{continuous} \end{cases}$$

- Quiz: why?

What you should know about every pdf

What you should know about every pdf

- The **assignment** of a probability or probability density to every conceivable value of Y_i

What you should know about every pdf

- The **assignment** of a probability or probability density to every conceivable value of Y_i
- The **first principles**

What you should know about every pdf

- The **assignment** of a probability or probability density to every conceivable value of Y_i
- The **first principles**
- How to **use** the final expression (but not necessarily the full derivation)

What you should know about every pdf

- The **assignment** of a probability or probability density to every conceivable value of Y_i
- The **first principles**
- How to **use** the final expression (but not necessarily the full derivation)
- How to **simulate** from the density

What you should know about every pdf

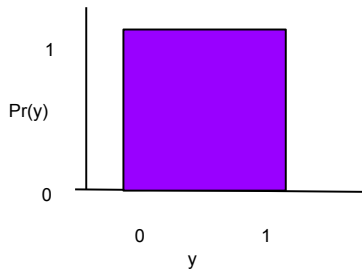
- The **assignment** of a probability or probability density to every conceivable value of Y_i
- The **first principles**
- How to **use** the final expression (but not necessarily the full derivation)
- How to **simulate** from the density
- How to **compute** features of the density such as its “moments”

What you should know about every pdf

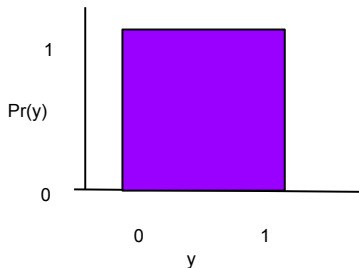
- The **assignment** of a probability or probability density to every conceivable value of Y_i
- The **first principles**
- How to **use** the final expression (but not necessarily the full derivation)
- How to **simulate** from the density
- How to **compute** features of the density such as its “moments”
- How to **verify** that the final expression is indeed a proper density

Uniform Density on the interval $[0, 1]$

Uniform Density on the interval $[0, 1]$

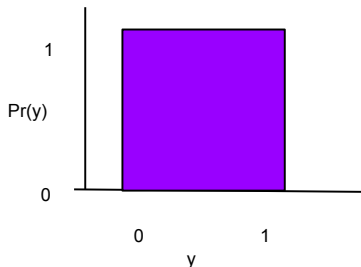


Uniform Density on the interval $[0, 1]$



First Principles about the process that generates Y_i is such that

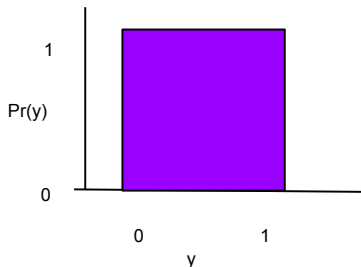
Uniform Density on the interval $[0, 1]$



First Principles about the process that generates Y_i is such that

- Y_i always falls in the “unit” interval: $\int_0^1 P(y)dy = 1$

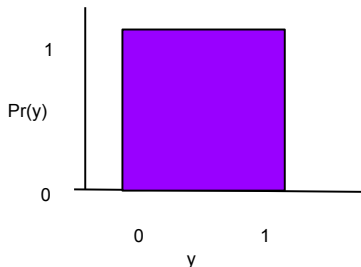
Uniform Density on the interval $[0, 1]$



First Principles about the process that generates Y_i is such that

- Y_i always falls in the “unit” interval: $\int_0^1 P(y)dy = 1$
- $\Pr(Y \in (a, b)) = \Pr(Y \in (c, d))$ if $a < b$, $c < d$, and $b - a = d - c$.

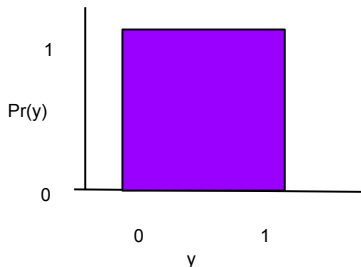
Uniform Density on the interval $[0, 1]$



First Principles about the process that generates Y_i is such that

- Y_i always falls in the “unit” interval: $\int_0^1 P(y)dy = 1$
- $\Pr(Y \in (a, b)) = \Pr(Y \in (c, d))$ if $a < b$, $c < d$, and $b - a = d - c$.
- **Quiz:** How do you know it's a pdf?

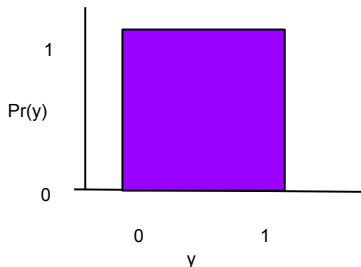
Uniform Density on the interval $[0, 1]$



First Principles about the process that generates Y_i is such that

- Y_i always falls in the “unit” interval: $\int_0^1 P(y)dy = 1$
- $\Pr(Y \in (a, b)) = \Pr(Y \in (c, d))$ if $a < b$, $c < d$, and $b - a = d - c$.
- Quiz: How do you know it's a pdf?
- Quiz 2: How to simulate?

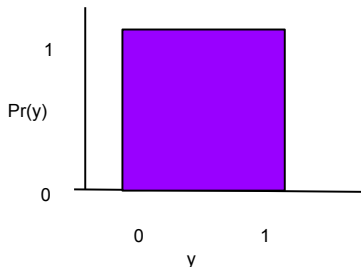
Uniform Density on the interval $[0, 1]$



First Principles about the process that generates Y_i is such that

- Y_i always falls in the “unit” interval: $\int_0^1 P(y)dy = 1$
- $\Pr(Y \in (a, b)) = \Pr(Y \in (c, d))$ if $a < b$, $c < d$, and $b - a = d - c$.
- **Quiz:** How do you know it's a pdf?
- **Quiz 2:** How to simulate? `runif(1000)`

Uniform Density on the interval $[0, 1]$

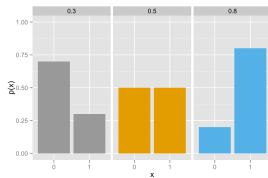


First Principles about the process that generates Y_i is such that

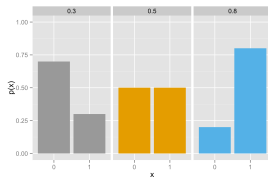
- Y_i always falls in the “unit” interval: $\int_0^1 P(y)dy = 1$
- $\Pr(Y \in (a, b)) = \Pr(Y \in (c, d))$ if $a < b$, $c < d$, and $b - a = d - c$.
- Quiz: How do you know it's a pdf?
- Quiz 2: How to simulate? `runif(1000)`
- Quiz 3: This PDF has no parameters. Could we add some?

Bernoulli pdf (or pmf)

Bernoulli pdf (or pmf)

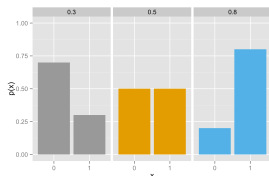


Bernoulli pdf (or pmf)



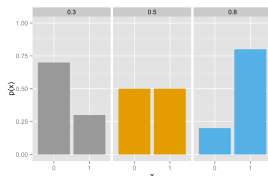
- First principles about the process that generates Y_i :

Bernoulli pdf (or pmf)



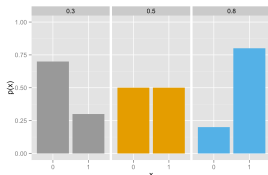
- **First principles** about the process that generates Y_i :
 - Y_i has 2 **mutually exclusive** outcomes; and

Bernoulli pdf (or pmf)



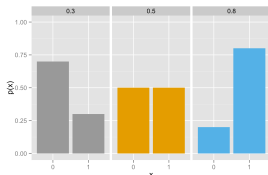
- **First principles** about the process that generates Y_i :
 - Y_i has 2 **mutually exclusive** outcomes; and
 - The 2 outcomes are **exhaustive**

Bernoulli pdf (or pmf)



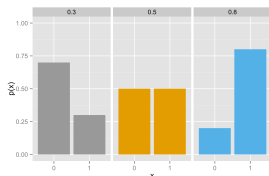
- First principles about the process that generates Y_i :
 - Y_i has 2 mutually exclusive outcomes; and
 - The 2 outcomes are exhaustive
- Quiz: What's an example that violates these rules

Bernoulli pdf (or pmf)



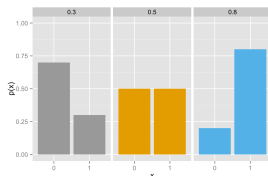
- **First principles** about the process that generates Y_i :
 - Y_i has 2 **mutually exclusive** outcomes; and
 - The 2 outcomes are **exhaustive**
- **Quiz: What's an example that violates these rules**
- In this simple case, we'll compute features analytically and by simulation.

Bernoulli pdf (or pmf)



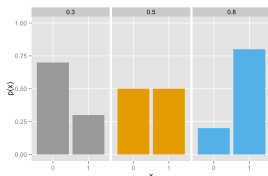
- **First principles** about the process that generates Y_i :
 - Y_i has 2 **mutually exclusive** outcomes; and
 - The 2 outcomes are **exhaustive**
- **Quiz: What's an example that violates these rules**
- In this simple case, we'll compute features analytically and by simulation.
- Mathematical expression for the pmf

Bernoulli pdf (or pmf)



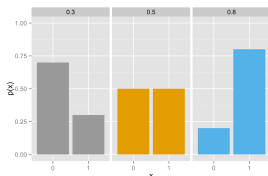
- **First principles** about the process that generates Y_i :
 - Y_i has 2 **mutually exclusive** outcomes; and
 - The 2 outcomes are **exhaustive**
- **Quiz: What's an example that violates these rules**
- In this simple case, we'll compute features analytically and by simulation.
- Mathematical expression for the pmf
 - $\Pr(Y_i = 1|\pi_i) = \pi_i$,

Bernoulli pdf (or pmf)



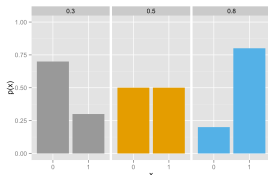
- **First principles** about the process that generates Y_i :
 - Y_i has 2 **mutually exclusive** outcomes; and
 - The 2 outcomes are **exhaustive**
- **Quiz: What's an example that violates these rules**
- In this simple case, we'll compute features analytically and by simulation.
- Mathematical expression for the pmf
 - $\Pr(Y_i = 1|\pi_i) = \pi_i$, $\Pr(Y_i = 0|\pi_i) = 1 - \pi_i$

Bernoulli pdf (or pmf)



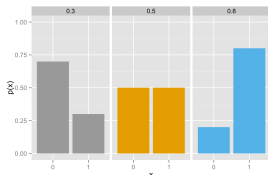
- **First principles** about the process that generates Y_i :
 - Y_i has 2 **mutually exclusive** outcomes; and
 - The 2 outcomes are **exhaustive**
- **Quiz: What's an example that violates these rules**
- In this simple case, we'll compute features analytically and by simulation.
- Mathematical expression for the pmf
 - $\Pr(Y_i = 1|\pi_i) = \pi_i$, $\Pr(Y_i = 0|\pi_i) = 1 - \pi_i$
 - The parameter π happens to be interpretable as a probability

Bernoulli pdf (or pmf)



- **First principles** about the process that generates Y_i :
 - Y_i has 2 **mutually exclusive** outcomes; and
 - The 2 outcomes are **exhaustive**
- **Quiz: What's an example that violates these rules**
- In this simple case, we'll compute features analytically and by simulation.
- Mathematical expression for the pmf
 - $\Pr(Y_i = 1|\pi_i) = \pi_i, \Pr(Y_i = 0|\pi_i) = 1 - \pi_i$
 - The parameter π happens to be interpretable as a probability
 - $\implies \Pr(Y_i = y|\pi_i) = \pi_i^y (1 - \pi_i)^{1-y}$

Bernoulli pdf (or pmf)



- **First principles** about the process that generates Y_i :
 - Y_i has 2 **mutually exclusive** outcomes; and
 - The 2 outcomes are **exhaustive**
- **Quiz: What's an example that violates these rules**
- In this simple case, we'll compute features analytically and by simulation.
- Mathematical expression for the pmf
 - $\Pr(Y_i = 1|\pi_i) = \pi_i, \Pr(Y_i = 0|\pi_i) = 1 - \pi_i$
 - The parameter π happens to be interpretable as a probability
 - $\implies \Pr(Y_i = y|\pi_i) = \pi_i^y (1 - \pi_i)^{1-y}$
 - Alternative notation: $\Pr(Y_i = y|\pi_i) = \text{Bernoulli}(y|\pi_i) = f_b(y|\pi_i)$

Features of the Bernoulli: analytically

Features of the Bernoulli: analytically

- Expected value:

Features of the Bernoulli: analytically

- Expected value:

$$E(Y) = \sum_{\text{all } y} yP(y)$$

Features of the Bernoulli: analytically

- Expected value:

$$\begin{aligned} E(Y) &= \sum_{\text{all } y} yP(y) \\ &= 0 \Pr(0) + 1 \Pr(1) \end{aligned}$$

Features of the Bernoulli: analytically

- Expected value:

$$\begin{aligned} E(Y) &= \sum_{\text{all } y} yP(y) \\ &= 0 \Pr(0) + 1 \Pr(1) \\ &= \pi \end{aligned}$$

Features of the Bernoulli: analytically

- Expected value:

$$\begin{aligned} E(Y) &= \sum_{\text{all } y} yP(y) \\ &= 0 \Pr(0) + 1 \Pr(1) \\ &= \pi \end{aligned}$$

- Variance:

Features of the Bernoulli: analytically

- Expected value:

$$\begin{aligned} E(Y) &= \sum_{\text{all } y} yP(y) \\ &= 0 \Pr(0) + 1 \Pr(1) \\ &= \pi \end{aligned}$$

- Variance:

$$V(Y) = E[(Y - E(Y))^2] \quad (\text{The definition})$$

Features of the Bernoulli: analytically

- Expected value:

$$\begin{aligned} E(Y) &= \sum_{\text{all } y} yP(y) \\ &= 0 \Pr(0) + 1 \Pr(1) \\ &= \pi \end{aligned}$$

- Variance:

$$\begin{aligned} V(Y) &= E[(Y - E(Y))^2] && \text{(The definition)} \\ &= E(Y^2) - E(Y)^2 && \text{(An easier version)} \end{aligned}$$

Features of the Bernoulli: analytically

- Expected value:

$$\begin{aligned} E(Y) &= \sum_{\text{all } y} yP(y) \\ &= 0 \Pr(0) + 1 \Pr(1) \\ &= \pi \end{aligned}$$

- Variance:

$$\begin{aligned} V(Y) &= E[(Y - E(Y))^2] && \text{(The definition)} \\ &= E(Y^2) - E(Y)^2 && \text{(An easier version)} \\ &= E(Y^2) - \pi^2 \end{aligned}$$

Features of the Bernoulli: analytically

- Expected value:

$$\begin{aligned} E(Y) &= \sum_{\text{all } y} yP(y) \\ &= 0 \Pr(0) + 1 \Pr(1) \\ &= \pi \end{aligned}$$

- Variance:

$$\begin{aligned} V(Y) &= E[(Y - E(Y))^2] && \text{(The definition)} \\ &= E(Y^2) - E(Y)^2 && \text{(An easier version)} \\ &= E(Y^2) - \pi^2 \end{aligned}$$

- How do we compute $E(Y^2)$?

Expected values of functions of random variables

Expected values of functions of random variables

$$E[g(Y)] = \sum_{\text{all } y} g(y)P(y)$$

Expected values of functions of random variables

$$E[g(Y)] = \sum_{\text{all } y} g(y)P(y)$$

or

Expected values of functions of random variables

$$E[g(Y)] = \sum_{\text{all } y} g(y)P(y)$$

or

$$E[g(Y)] = \int_{-\infty}^{\infty} g(y)P(y)$$

Expected values of functions of random variables

$$E[g(Y)] = \sum_{\text{all } y} g(y)P(y)$$

or

$$E[g(Y)] = \int_{-\infty}^{\infty} g(y)P(y)$$

For example,

Expected values of functions of random variables

$$E[g(Y)] = \sum_{\text{all } y} g(y)P(y)$$

or

$$E[g(Y)] = \int_{-\infty}^{\infty} g(y)P(y)$$

For example,

$$E(Y^2) = \sum_{\text{all } y} y^2 P(y)$$

Expected values of functions of random variables

$$E[g(Y)] = \sum_{\text{all } y} g(y)P(y)$$

or

$$E[g(Y)] = \int_{-\infty}^{\infty} g(y)P(y)$$

For example,

$$\begin{aligned} E(Y^2) &= \sum_{\text{all } y} y^2 P(y) \\ &= 0^2 \Pr(0) + 1^2 \Pr(1) \end{aligned}$$

Expected values of functions of random variables

$$E[g(Y)] = \sum_{\text{all } y} g(y)P(y)$$

or

$$E[g(Y)] = \int_{-\infty}^{\infty} g(y)P(y)$$

For example,

$$\begin{aligned} E(Y^2) &= \sum_{\text{all } y} y^2 P(y) \\ &= 0^2 \Pr(0) + 1^2 \Pr(1) \\ &= \pi \end{aligned}$$

Variance of the Bernoulli (uses above results)

Variance of the Bernoulli (uses above results)

$$V(Y) = E[(Y - E(Y))^2] \quad (\text{The definition})$$

Variance of the Bernoulli (uses above results)

$$\begin{aligned} V(Y) &= E[(Y - E(Y))^2] && \text{(The definition)} \\ &= E(Y^2) - E(Y)^2 && \text{(An easier version)} \end{aligned}$$

Variance of the Bernoulli (uses above results)

$$\begin{aligned} V(Y) &= E[(Y - E(Y))^2] && \text{(The definition)} \\ &= E(Y^2) - E(Y)^2 && \text{(An easier version)} \\ &= \pi - \pi^2 \end{aligned}$$

Variance of the Bernoulli (uses above results)

$$\begin{aligned} V(Y) &= E[(Y - E(Y))^2] && \text{(The definition)} \\ &= E(Y^2) - E(Y)^2 && \text{(An easier version)} \\ &= \pi - \pi^2 \\ &= \pi(1 - \pi) \end{aligned}$$

Variance of the Bernoulli (uses above results)

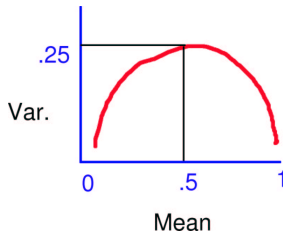
$$\begin{aligned} V(Y) &= E[(Y - E(Y))^2] && \text{(The definition)} \\ &= E(Y^2) - E(Y)^2 && \text{(An easier version)} \\ &= \pi - \pi^2 \\ &= \pi(1 - \pi) \end{aligned}$$

This makes sense:

Variance of the Bernoulli (uses above results)

$$\begin{aligned} V(Y) &= E[(Y - E(Y))^2] && \text{(The definition)} \\ &= E(Y^2) - E(Y)^2 && \text{(An easier version)} \\ &= \pi - \pi^2 \\ &= \pi(1 - \pi) \end{aligned}$$

This makes sense:



How to Simulate from the Bernoulli with parameter π

How to Simulate from the Bernoulli with parameter π

- take one draw u from a uniform density on the interval $[0,1]$

How to Simulate from the Bernoulli with parameter π

- take one draw u from a uniform density on the interval $[0,1]$
- Set π to a particular value

How to Simulate from the Bernoulli with parameter π

- take one draw u from a uniform density on the interval $[0,1]$
- Set π to a particular value
- Set $y = 1$ if $u < \pi$ and $y = 0$ otherwise

How to Simulate from the Bernoulli with parameter π

- take one draw u from a uniform density on the interval $[0,1]$
- Set π to a particular value
- Set $y = 1$ if $u < \pi$ and $y = 0$ otherwise
- In R:

```
sims <- 1000           # set parameters
bernpi <- 0.2
u <- runif(sims)        # uniform sims
y <- as.integer(u < bernpi)
y                        # print results
```

How to Simulate from the Bernoulli with parameter π

- take one draw u from a uniform density on the interval $[0,1]$
- Set π to a particular value
- Set $y = 1$ if $u < \pi$ and $y = 0$ otherwise
- In R:

```
sims <- 1000           # set parameters
bernpi <- 0.2
u <- runif(sims)        # uniform sims
y <- as.integer(u < bernpi)
y                        # print results
```

- Running the program gives:

How to Simulate from the Bernoulli with parameter π

- take one draw u from a uniform density on the interval $[0,1]$
- Set π to a particular value
- Set $y = 1$ if $u < \pi$ and $y = 0$ otherwise
- In R:

```
sims <- 1000           # set parameters
bernpi <- 0.2
u <- runif(sims)        # uniform sims
y <- as.integer(u < bernpi)
y                        # print results
```

- Running the program gives:

```
0 0 0 1 0 0 1 1 0 0 1 1 1 0 ...
```

How to Simulate from the Bernoulli with parameter π

- take one draw u from a uniform density on the interval $[0,1]$
- Set π to a particular value
- Set $y = 1$ if $u < \pi$ and $y = 0$ otherwise
- In R:

```
sims <- 1000          # set parameters
bernpi <- 0.2
u <- runif(sims)       # uniform sims
y <- as.integer(u < bernpi)
y                      # print results
```

- Running the program gives:

```
0 0 0 1 0 0 1 1 0 0 1 1 1 0 ...
```

- Quiz: What can we do with the simulations?

Binomial Distribution

Binomial Distribution

First principles:

Binomial Distribution

First principles:

- N iid Bernoulli trials, $y_1, \dots, y_N$

Binomial Distribution

First principles:

- N iid Bernoulli trials, $y_1, \dots, y_N$
- The trials are independent

Binomial Distribution

First principles:

- N iid Bernoulli trials, $y_1, \dots, y_N$
- The trials are independent
- The trials are identically distributed

Binomial Distribution

First principles:

- N iid Bernoulli trials, $y_1, \dots, y_N$
- The trials are independent
- The trials are identically distributed
- We observe $Y = \sum_{i=1}^N y_i$

Binomial Distribution

First principles:

- N iid Bernoulli trials, $y_1, \dots, y_N$
- The trials are independent
- The trials are identically distributed
- We observe $Y = \sum_{i=1}^N y_i$

Density:

Binomial Distribution

First principles:

- N iid Bernoulli trials, $y_1, \dots, y_N$
- The trials are independent
- The trials are identically distributed
- We observe $Y = \sum_{i=1}^N y_i$

Density:

$$P(Y = y|\pi) = \binom{N}{y} \pi^y (1 - \pi)^{N-y}$$

Binomial Distribution

First principles:

- N iid Bernoulli trials, $y_1, \dots, y_N$
- The trials are independent
- The trials are identically distributed
- We observe $Y = \sum_{i=1}^N y_i$

Density:

$$P(Y = y|\pi) = \binom{N}{y} \pi^y (1 - \pi)^{N-y}$$

Explanation:

Binomial Distribution

First principles:

- N iid Bernoulli trials, $y_1, \dots, y_N$
- The trials are independent
- The trials are identically distributed
- We observe $Y = \sum_{i=1}^N y_i$

Density:

$$P(Y = y|\pi) = \binom{N}{y} \pi^y (1 - \pi)^{N-y}$$

Explanation:

- $\binom{N}{y}$ because (1 0 1) and (1 1 0) are both $y = 2$.

Binomial Distribution

First principles:

- N iid Bernoulli trials, $y_1, \dots, y_N$
- The trials are independent
- The trials are identically distributed
- We observe $Y = \sum_{i=1}^N y_i$

Density:

$$P(Y = y|\pi) = \binom{N}{y} \pi^y (1 - \pi)^{N-y}$$

Explanation:

- $\binom{N}{y}$ because (1 0 1) and (1 1 0) are both $y = 2$.
- π^y because y successes with π probability each (product taken due to independence)

Binomial Distribution

First principles:

- N iid Bernoulli trials, $y_1, \dots, y_N$
- The trials are independent
- The trials are identically distributed
- We observe $Y = \sum_{i=1}^N y_i$

Density:

$$P(Y = y|\pi) = \binom{N}{y} \pi^y (1 - \pi)^{N-y}$$

Explanation:

- $\binom{N}{y}$ because (1 0 1) and (1 1 0) are both $y = 2$.
- π^y because y successes with π probability each (product taken due to independence)
- $(1 - \pi)^{N-y}$ because $N - y$ failures with $1 - \pi$ probability each

Binomial Distribution

First principles:

- N iid Bernoulli trials, $y_1, \dots, y_N$
- The trials are independent
- The trials are identically distributed
- We observe $Y = \sum_{i=1}^N y_i$

Density:

$$P(Y = y|\pi) = \binom{N}{y} \pi^y (1 - \pi)^{N-y}$$

Explanation:

- $\binom{N}{y}$ because (1 0 1) and (1 1 0) are both $y = 2$.
- π^y because y successes with π probability each (product taken due to independence)
- $(1 - \pi)^{N-y}$ because $N - y$ failures with $1 - \pi$ probability each
- Moments: Mean $E(Y) = N\pi$;

Binomial Distribution

First principles:

- N iid Bernoulli trials, $y_1, \dots, y_N$
- The trials are independent
- The trials are identically distributed
- We observe $Y = \sum_{i=1}^N y_i$

Density:

$$P(Y = y|\pi) = \binom{N}{y} \pi^y (1 - \pi)^{N-y}$$

Explanation:

- $\binom{N}{y}$ because (1 0 1) and (1 1 0) are both $y = 2$.
- π^y because y successes with π probability each (product taken due to independence)
- $(1 - \pi)^{N-y}$ because $N - y$ failures with $1 - \pi$ probability each
- Moments: Mean $E(Y) = N\pi$; Variance $V(Y) = \pi(1 - \pi)/N$

How to simulate from the Binomial distribution

How to simulate from the Binomial distribution

- To simulate from the Binomial(π ; N):

How to simulate from the Binomial distribution

- To simulate from the Binomial(π ; N):
 - Simulate N independent Bernoulli variables, $Y_1, \dots, Y_N$, each with parameter π

How to simulate from the Binomial distribution

- To simulate from the Binomial(π ; N):
 - Simulate N independent Bernoulli variables, $Y_1, \dots, Y_N$, each with parameter π
 - Add them up: $\sum_{i=1}^N Y_i$

How to simulate from the Binomial distribution

- To simulate from the Binomial(π ; N):
 - Simulate N independent Bernoulli variables, $Y_1, \dots, Y_N$, each with parameter π
 - Add them up: $\sum_{i=1}^N Y_i$
- What can you do with the simulations?

Where to get uniform random numbers

Where to get uniform random numbers

- Random is not haphazard (e.g., Benford's law)

Where to get uniform random numbers

- Random is not haphazard (e.g., Benford's law)
- Random number generators are perfectly predictable (what?)

Where to get uniform random numbers

- Random is not haphazard (e.g., Benford's law)
- Random number generators are perfectly predictable (what?)
- We use pseudo-random numbers which have (a) digits that occur with 1/10th probability, (b) no time series patterns, etc.

Where to get uniform random numbers

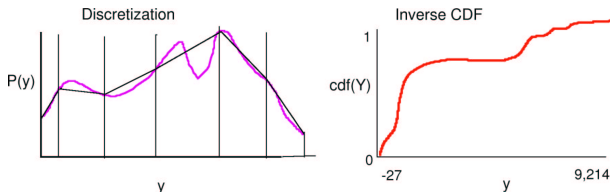
- Random is not haphazard (e.g., Benford's law)
- Random number generators are perfectly predictable (what?)
- We use pseudo-random numbers which have (a) digits that occur with 1/10th probability, (b) no time series patterns, etc.
- How to create real random numbers?

Where to get uniform random numbers

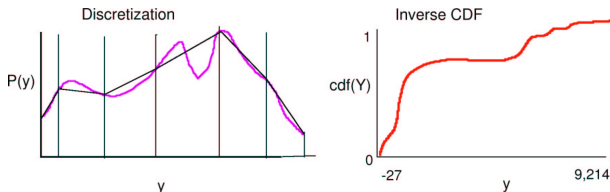
- Random is not haphazard (e.g., Benford's law)
- Random number generators are perfectly predictable (what?)
- We use pseudo-random numbers which have (a) digits that occur with 1/10th probability, (b) no time series patterns, etc.
- How to create real random numbers?
- Some chips now use quantum effects

Discretization for random draws from discrete pmfs

Discretization for random draws from discrete pmfs

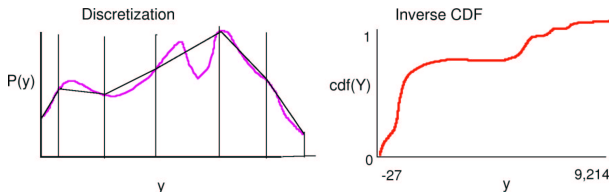


Discretization for random draws from discrete pmfs



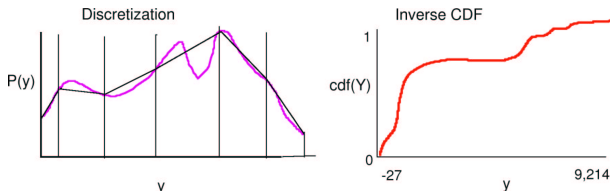
- Divide up PDF into a grid

Discretization for random draws from discrete pmfs



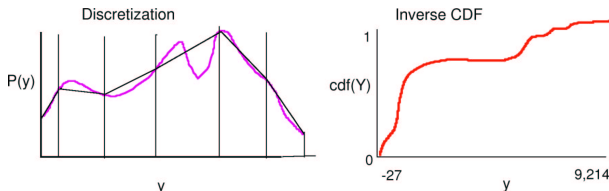
- Divide up PDF into a grid
- Approximate probabilities by trapezoids

Discretization for random draws from discrete pmfs



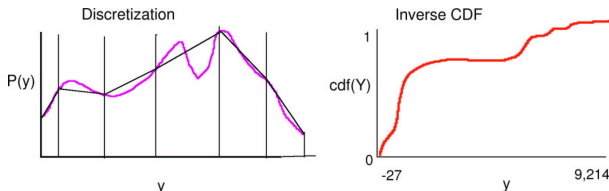
- Divide up PDF into a grid
- Approximate probabilities by trapezoids
- Map $[0,1]$ uniform draws to the proportion area in each trapezoid

Discretization for random draws from discrete pmfs



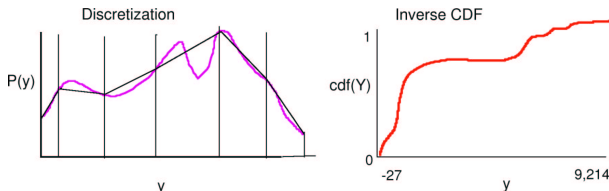
- Divide up PDF into a grid
- Approximate probabilities by trapezoids
- Map $[0,1]$ uniform draws to the proportion area in each trapezoid
- Return midpoint of each trapezoid

Discretization for random draws from discrete pmfs



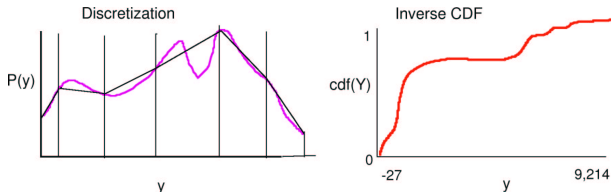
- Divide up PDF into a grid
- Approximate probabilities by trapezoids
- Map $[0,1]$ uniform draws to the proportion area in each trapezoid
- Return midpoint of each trapezoid
- More trapezoids $\rightsquigarrow$ better approximation

Discretization for random draws from discrete pmfs

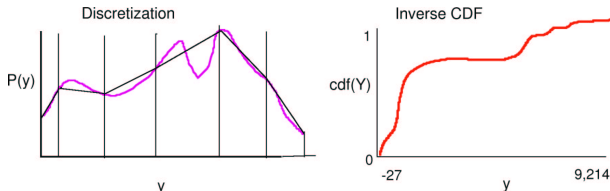


- Divide up PDF into a grid
- Approximate probabilities by trapezoids
- Map $[0,1]$ uniform draws to the proportion area in each trapezoid
- Return midpoint of each trapezoid
- More trapezoids $\rightsquigarrow$ better approximation
- (Works for a few dimensions, but infeasible for many)

Inverse CDF: drawing from arbitrary continuous pdfs



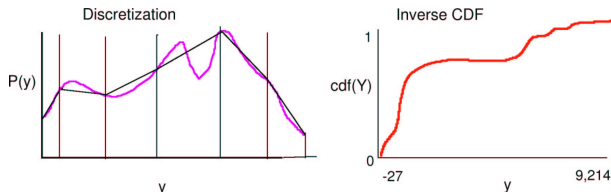
Inverse CDF: drawing from arbitrary continuous pdfs



- From the pdf $f(Y)$, compute the cdf:

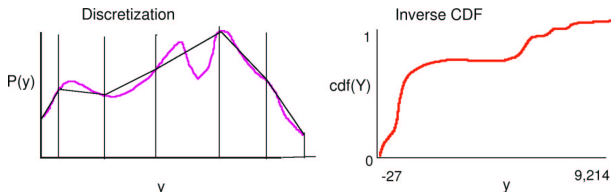
$$\Pr(Y \leq y) \equiv F(y) = \int_{-\infty}^y f(z) dz$$

Inverse CDF: drawing from arbitrary continuous pdfs



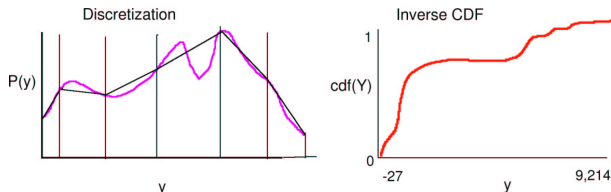
- From the pdf $f(Y)$, compute the cdf:
$$\Pr(Y \leq y) \equiv F(y) = \int_{-\infty}^y f(z) dz$$
- Define the inverse cdf $F^{-1}(y)$, such that $F^{-1}[F(y)] = y$

Inverse CDF: drawing from arbitrary continuous pdfs



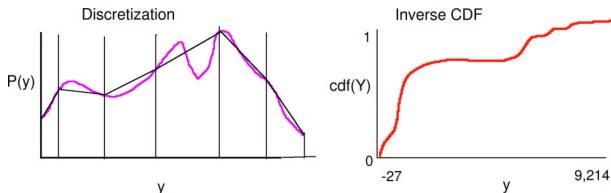
- From the pdf $f(Y)$, compute the cdf:
$$\Pr(Y \leq y) \equiv F(y) = \int_{-\infty}^y f(z) dz$$
- Define the inverse cdf $F^{-1}(y)$, such that $F^{-1}[F(y)] = y$
- Draw random uniform number, U

Inverse CDF: drawing from arbitrary continuous pdfs

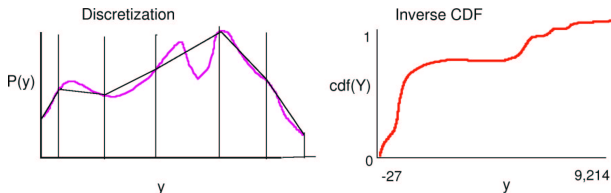


- From the pdf $f(Y)$, compute the cdf:
$$\Pr(Y \leq y) \equiv F(y) = \int_{-\infty}^y f(z) dz$$
- Define the inverse cdf $F^{-1}(y)$, such that $F^{-1}[F(y)] = y$
- Draw random uniform number, U
- Then $F^{-1}(U)$ gives a random draw from $f(Y)$.

Using Inverse CDF to Improve Discretization Method

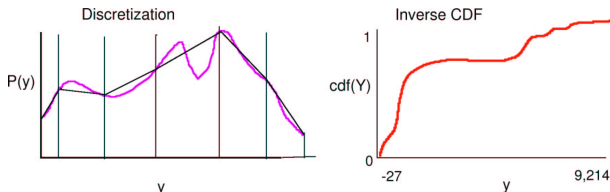


Using Inverse CDF to Improve Discretization Method



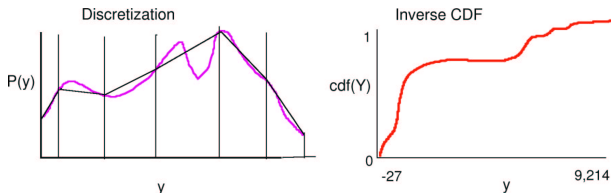
- Refined Discretization Method:

Using Inverse CDF to Improve Discretization Method



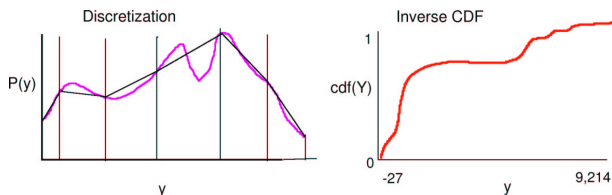
- Refined Discretization Method:
 - Choose interval randomly as above (based on area in trapezoids)

Using Inverse CDF to Improve Discretization Method



- **Refined Discretization Method:**
 - Choose interval randomly as above (based on area in trapezoids)
 - Draw a number within each trapezoid by the inverse CDF method applied to the trapezoidal approximation.

Using Inverse CDF to Improve Discretization Method



- **Refined Discretization Method:**
 - Choose interval randomly as above (based on area in trapezoids)
 - Draw a number within each trapezoid by the inverse CDF method applied to the trapezoidal approximation.
- Drawing random numbers from arbitrary multivariate densities: now an enormous literature

Normal Distribution

Normal Distribution

- Many different first principles

Normal Distribution

- Many different first principles
- A common one is the central limit theorem

Normal Distribution

- Many different first principles
- A common one is the central limit theorem
- The **univariate normal** density (with **mean** μ_i , **variance** σ^2)

Normal Distribution

- Many different first principles
- A common one is the central limit theorem
- The **univariate normal** density (with **mean** μ_i , **variance** σ^2)

$$N(y_i|\mu_i, \sigma^2) = (2\pi\sigma^2)^{-1/2} \exp\left(\frac{-(y_i - \mu_i)^2}{2\sigma^2}\right)$$

Normal Distribution

- Many different first principles
- A common one is the central limit theorem
- The **univariate normal** density (with **mean** μ_i , **variance** σ^2)

$$N(y_i|\mu_i, \sigma^2) = (2\pi\sigma^2)^{-1/2} \exp\left(\frac{-(y_i - \mu_i)^2}{2\sigma^2}\right)$$

- The **stylized normal**: $f_{stn}(y_i|\mu_i) = N(y|\mu_i, 1)$

Normal Distribution

- Many different first principles
- A common one is the central limit theorem
- The **univariate normal** density (with **mean** μ_i , **variance** σ^2)

$$N(y_i|\mu_i, \sigma^2) = (2\pi\sigma^2)^{-1/2} \exp\left(\frac{-(y_i - \mu_i)^2}{2\sigma^2}\right)$$

- The **stylized normal**: $f_{stn}(y_i|\mu_i) = N(y|\mu_i, 1)$

$$f_{stn}(y|\mu_i) = (2\pi)^{-1/2} \exp\left(\frac{-(y_i - \mu_i)^2}{2}\right)$$

Normal Distribution

- Many different first principles
- A common one is the central limit theorem
- The **univariate normal** density (with **mean** μ_i , **variance** σ^2)

$$N(y_i|\mu_i, \sigma^2) = (2\pi\sigma^2)^{-1/2} \exp\left(\frac{-(y_i - \mu_i)^2}{2\sigma^2}\right)$$

- The **stylized normal**: $f_{stn}(y_i|\mu_i) = N(y|\mu_i, 1)$

$$f_{stn}(y|\mu_i) = (2\pi)^{-1/2} \exp\left(\frac{-(y_i - \mu_i)^2}{2}\right)$$

- The **standardized normal**: $f_{sn}(y_i) = N(y_i|0, 1) = \phi(y_i)$

Normal Distribution

- Many different first principles
- A common one is the central limit theorem
- The **univariate normal** density (with **mean** μ_i , **variance** σ^2)

$$N(y_i|\mu_i, \sigma^2) = (2\pi\sigma^2)^{-1/2} \exp\left(\frac{-(y_i - \mu_i)^2}{2\sigma^2}\right)$$

- The **stylized normal**: $f_{stn}(y_i|\mu_i) = N(y|\mu_i, 1)$

$$f_{stn}(y|\mu_i) = (2\pi)^{-1/2} \exp\left(\frac{-(y_i - \mu_i)^2}{2}\right)$$

- The **standardized normal**: $f_{sn}(y_i) = N(y_i|0, 1) = \phi(y_i)$

$$f_{sn}(y_i) = (2\pi)^{-1/2} \exp\left(\frac{-y_i^2}{2}\right)$$

Reminder: Equivalent Regression Notation

- Standard version

$$Y_i = x_i\beta + \epsilon_i \quad = \text{systematic} + \text{stochastic}$$
$$\epsilon_i \sim f_N(0, \sigma^2)$$

- Alternative version

$$Y_i \sim f_N(\mu_i, \sigma^2) \quad \text{stochastic}$$
$$\mu_i = x_i\beta \quad \text{systematic}$$

- Generalized version

$$Y_i \sim f(\theta_i, \alpha) \quad \text{stochastic}$$
$$\theta_i = g(x_i, \beta) \quad \text{systematic}$$

Multivariate Normal Distribution

Multivariate Normal Distribution

- Let $Y_i \equiv \{Y_{1i}, \dots, Y_{ki}\}$ be a $k \times 1$ vector, jointly random:

Multivariate Normal Distribution

- Let $Y_i \equiv \{Y_{1i}, \dots, Y_{ki}\}$ be a $k \times 1$ vector, jointly random:

$$Y_i \sim N(y_i | \mu_i, \Sigma)$$

Multivariate Normal Distribution

- Let $Y_i \equiv \{Y_{1i}, \dots, Y_{ki}\}$ be a $k \times 1$ vector, jointly random:

$$Y_i \sim N(y_i | \mu_i, \Sigma)$$

where μ_i is $k \times 1$ and Σ is $k \times k$. For $k = 2$,

Multivariate Normal Distribution

- Let $Y_i \equiv \{Y_{1i}, \dots, Y_{ki}\}$ be a $k \times 1$ vector, jointly random:

$$Y_i \sim N(y_i | \mu_i, \Sigma)$$

where μ_i is $k \times 1$ and Σ is $k \times k$. For $k = 2$,

$$\mu_i = \begin{pmatrix} \mu_{1i} \\ \mu_{2i} \end{pmatrix} \quad \Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix}$$

Multivariate Normal Distribution

- Let $Y_i \equiv \{Y_{1i}, \dots, Y_{ki}\}$ be a $k \times 1$ vector, jointly random:

$$Y_i \sim N(y_i | \mu_i, \Sigma)$$

where μ_i is $k \times 1$ and Σ is $k \times k$. For $k = 2$,

$$\mu_i = \begin{pmatrix} \mu_{1i} \\ \mu_{2i} \end{pmatrix} \quad \Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix}$$

- Mathematical form:

Multivariate Normal Distribution

- Let $Y_i \equiv \{Y_{1i}, \dots, Y_{ki}\}$ be a $k \times 1$ vector, jointly random:

$$Y_i \sim N(y_i | \mu_i, \Sigma)$$

where μ_i is $k \times 1$ and Σ is $k \times k$. For $k = 2$,

$$\mu_i = \begin{pmatrix} \mu_{1i} \\ \mu_{2i} \end{pmatrix} \quad \Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix}$$

- Mathematical form:

$$N(y_i | \mu_i, \Sigma) = (2\pi)^{-k/2} |\Sigma|^{-1/2} \exp \left[-\frac{1}{2} (y_i - \mu_i)' \Sigma^{-1} (y_i - \mu_i) \right]$$

Multivariate Normal Distribution

- Let $Y_i \equiv \{Y_{1i}, \dots, Y_{ki}\}$ be a $k \times 1$ vector, jointly random:

$$Y_i \sim N(y_i | \mu_i, \Sigma)$$

where μ_i is $k \times 1$ and Σ is $k \times k$. For $k = 2$,

$$\mu_i = \begin{pmatrix} \mu_{1i} \\ \mu_{2i} \end{pmatrix} \quad \Sigma = \begin{pmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{pmatrix}$$

- Mathematical form:

$$N(y_i | \mu_i, \Sigma) = (2\pi)^{-k/2} |\Sigma|^{-1/2} \exp \left[-\frac{1}{2} (y_i - \mu_i)' \Sigma^{-1} (y_i - \mu_i) \right]$$

- Simulating *once* from this density produces k numbers. Special algorithms are used to generate normal random variates (in R, `mvrnorm()`, from the MASS library).

Multivariate Normal Distribution

- Moments: $E(Y) = \mu_i$, $V(Y) = \Sigma$, $\text{Cov}(Y_1, Y_2) = \sigma_{12} = \sigma_{21}$.

Multivariate Normal Distribution

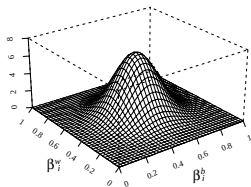
- Moments: $E(Y) = \mu_i$, $V(Y) = \Sigma$, $\text{Cov}(Y_1, Y_2) = \sigma_{12} = \sigma_{21}$.
- $\text{Corr}(Y_1, Y_2) = \frac{\sigma_{12}}{\sigma_1 \sigma_2}$

Multivariate Normal Distribution

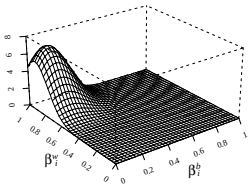
- Moments: $E(Y) = \mu_i$, $V(Y) = \Sigma$, $\text{Cov}(Y_1, Y_2) = \sigma_{12} = \sigma_{21}$.
- $\text{Corr}(Y_1, Y_2) = \frac{\sigma_{12}}{\sigma_1 \sigma_2}$
- Marginals:

$$N(Y_1 | \mu_1, \sigma_1^2) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} N(y_i | \mu_i, \Sigma) dy_2 dy_3 \cdots dy_k$$

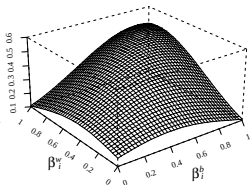
Truncated bivariate normal examples (for β^b and β^w)



(a) 0.5 0.5 0.15 0.15 0



(b) 0.1 0.9 0.15 0.15 0



(c) 0.8 0.8 0.6 0.6 0.5

Parameters are μ_1 , μ_2 , σ_1 , σ_2 , and ρ .