Statistical Inference

GR5204

Johannes Wiesel

Department of Statistics, Columbia University

General

- **Lectures:** Tuesdays & Thursdays, 2:40pm-3:55pm, 702 Hamilton Hall, in person

- **Instructors:** Johannes Wiesel, `johannes.wiesel@columbia.edu`

- **TA:** Wribhu Banik, `wb2387@columbia.edu` _(12-1)_

- **Instructor office hours:** Thursday, 1-2pm, Office 1016, School of Social Work _(before 1)_

- **TA office hours & location:** Friday 12-1pm, TBD

(after 2)

Outline

1 What is this course about?

2 Organization

Outline

Applications of statistics

- Medicine:
  - Tests: Covid, pregnancy, HIV, etc.
  - Effectiveness of drugs, vaccines, treatments, etc.
  - Risks of: behaviours, drugs, treatments, etc.
- Meteorology:
  - Weather forecast: understanding and predicting local weather
  - Weather risk: predicting floods, hurricanes, tornadoes
  - Climate: understanding and predicting global climate phenomena
- Polling:
  - Election polls
  - Official statistics (e.g. census)
  - Course evaluations
- Finance and Insurance: ← *me!*
  - prices of stocks, options, etc.
  - Insurance premium
- Data science:
  - Autonomous driving
  - Voice and pattern recognition
  - Personalized ads and recommendations
  - Networks (social, financial, neural,...)

The role of Statistics

Statistics is the discipline that studies the collection, organization, analysis, interpretation, and presentation of data.

The role of Statistics

Statistics is the discipline that studies the collection, organization, analysis, interpretation, and presentation of data.

- In **theoretical statistics**, one aims to derive unified concepts and frameworks for analyzing and interpreting data, often using mathematical tools.

The role of Statistics

Statistics is the discipline that studies the collection, organization, analysis, interpretation, and presentation of data.
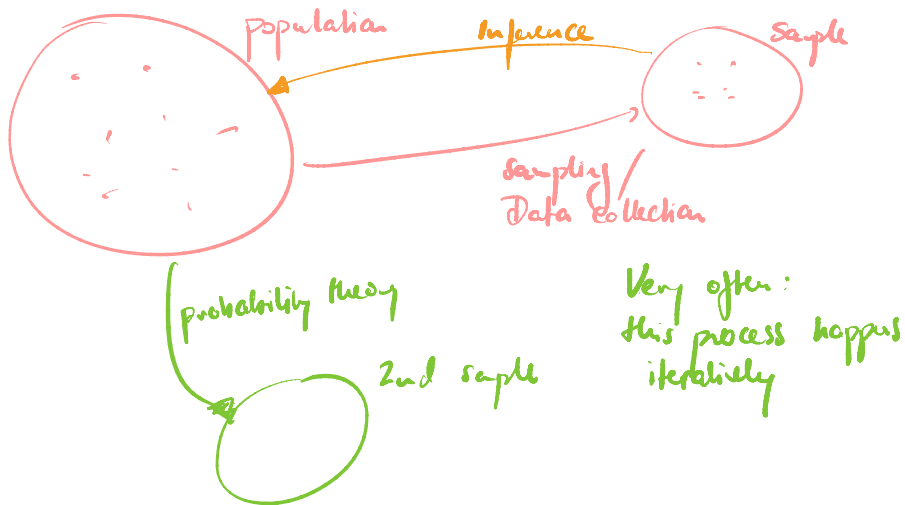
- In **theoretical statistics**, one aims to derive unified concepts and frameworks for analyzing and interpreting data, often using mathematical tools.
- In **applied statistics**, the methods studied in theoretical statistics are applied to specific problems in the sciences, technology, or engineering.

Branches of Statistics
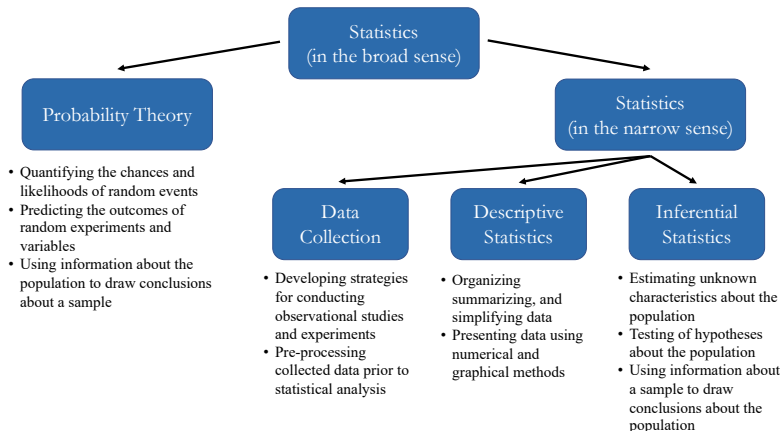
The dogma of statistics

Probability vs Statistics

Probability:

- Given a probability model P for the population, find the probability $P(A)$ of observing a particular event A when one or several subjects are picked/sampled from the population.
- Drawing conclusions about subjects/individuals from knowledge about the population.

Probability vs Statistics

Probability:

- Given a probability model P for the population, find the probability $P(A)$ of observing a particular event A when one or several subjects are picked/sampled from the population.
- Drawing conclusions about subjects/individuals from knowledge about the population.

Statistics:

- Given the behavior of several subjects from a population, find information about the (unknown) probability model P of the overall population.
- Drawing conclusions about the population from knowledge of certain subjects/individuals.

Probability vs Statistics

Example

Let $X_1, ..., X_{10}$ denote the results of flipping a coin ten times, with

$$X_i = \begin{cases} 0 & \text{if heads} \\ 1 & \text{if tails} \end{cases} \quad , \quad i = 1, ..., 10.$$

A plausible model is $X_i \overset{iid}{\sim} \text{Bernoulli}(\theta)$. We record the outcome

$$\boldsymbol{X} = (X_1, \ldots, X_{10}) = (0, 0, 0, 1, 0, 1, 1, 1, 1, 1).$$

Probability vs Statistics

Questions for a Probabilist:

- What is the probability of the outcome $\boldsymbol{X}$ as a function of θ?
- What is the probability of observing 5 heads in a row?
- If we keep tossing the coin, what can we say about the observations in the long run?

Probability vs Statistics

> **Questions for a Statistician:**
>
> - What is the true value of θ? Is the coin fair ($\theta = \frac{1}{2}$)?
> - How reliable is the estimate that we provide based on the observations?
> - Is there a "best" solution to the above problems?
> - How sensitive are our answers to departures from the assumption $X_i \overset{iid}{\sim} \text{Bernoulli}(\theta)$?
> - How do our "answers" behave as the number of tosses increases?
> - How many tosses do we need to get "accurate answers"?
> - Does our model $X_i \overset{iid}{\sim} \text{Bernoulli}(\theta)$ agree with the data?

Outline

Materials and Prerequisites

- Materials:
  - Lecture notes
  - *Probability and Statistics*, Morris H. DeGroot and Mark J. Schervish, 4th edition   *Ch. 7 - 12*

- Acknowledgment: notes adapted from Prof. Sen's and Prof. Vatter's courses

Materials and Prerequisites

- Materials:
  - Lecture notes
  - *Probability and Statistics*, Morris H. DeGroot and Mark J. Schervish, 4th edition

- Acknowledgment: notes adapted from Prof. Sen's and Prof. Vatter's courses

- Prerequisites:
  - **Calculus (single & multi-variable)**: Differentiation, integration, infinite sums, Taylor expansions, limits
  - **Linear algebra:** Vectors, matrices, eigenvectors, quadratic forms
  - **Probability theory:** STAT GR5203 or equivalent (i.e., content of chapters 1 to 6 of the textbook)

- To enjoy the course, you should have a solid interest in the **theory** and **proofs** behind the topics discussed in class.

Grading

- 20% HWs
- 40% Midterm
- 40% Final

Grading

- 20% HWs

- 40% Midterm

- 40% Final

- Remarks:
  - Grades are based on academic performance only.

  - **Disability-related accommodations:**
    health.columbia.edu/disability-services

  - **Academic integrity:**
    college.columbia.edu/academics/academicintegrity

Midterm & Final exam

- Midterm: **March 09, 2:40pm-3:55pm, during class**.
- Final exam: **May 11, 1:10–4:00pm**.

Midterm & Final exam

- Midterm: **March 09, 2:40pm-3:55pm, during class**.

- Final exam: **May 11, 1:10–4:00pm**.

- **If you have a conflict with any of the exams (for example, due to a religious holiday), please contact the instructor immediately.**

Midterm & Final exam

- Midterm: **March 09, 2:40pm-3:55pm, during class**.

- Final exam: **May 11, 1:10–4:00pm**.

- **If you have a conflict with any of the exams (for example, due to a religious holiday), please contact the instructor immediately.**

- Exam problems will be similar to those given in the problem sets and worked out in the lectures.

- You can bring 1 double-sided standard letter size sheet and 1 non-graphing calculator to the exams.

Homework

- There will be a HW problem set due roughly every 2 weeks.

- Homework assignments will be posted on Thursdays and are due on Friday, 11:59pm, in the following week. Solutions to the assignments will be posted on CourseWorks.

- Homework assignments have to be uploaded to CourseWorks as a **single pdf file**. Phone apps such as Adobe Scan work well to produce pdf scans.

Handwritten!

Homework

- There will be a HW problem set due roughly every 2 weeks.

- Homework assignments will be posted on Thursdays and are due on Friday, 11:59pm, in the following week. Solutions to the assignments will be posted on CourseWorks.

- Homework assignments have to be uploaded to CourseWorks as a **single pdf file**. Phone apps such as Adobe Scan work well to produce pdf scans.

- Late homework will not be accepted.

- However, your lowest homework score will be dropped before calculating your total homework score for the final grade. It is expected that this will take care of any homework you may fail to hand in on time because of the various difficulties you may face during the current pandemic (e.g., internet problems, traveling, mild illness, ...).

- Further accommodations can only be provided because of serious illness or another emergency of similar gravity; a note from a doctor or from a Dean will be required.

Homework

- There will be a HW problem set due roughly every 2 weeks.

- Homework assignments will be posted on Thursdays and are due on Friday, 11:59pm, in the following week. Solutions to the assignments will be posted on CourseWorks.

- Homework assignments have to be uploaded to CourseWorks as a **single pdf file**. Phone apps such as Adobe Scan work well to produce pdf scans.

- Late homework will not be accepted.

- However, your lowest homework score will be dropped before calculating your total homework score for the final grade. It is expected that this will take care of any homework you may fail to hand in on time because of the various difficulties you may face during the current pandemic (e.g., internet problems, traveling, mild illness, ...).

- Further accommodations can only be provided because of serious illness or another emergency of similar gravity; a note from a doctor or from a Dean will be required.

- Please make sure that you actually submit the solutions you want to submit (and not a corrupt file, an empty file, another assignment, ...). Submissions can be updated on CourseWorks until the due date/time.

Preparation of solutions (for HW & exams)

- Unless otherwise stated, you always have to justify your solutions.
- We may deduct points even if your final result is correct if you do not provide sufficient justification and/or intermediate steps.
- If you have to compute a quantity, do not simply state the final result. Instead, write down the formula you use, plug in the numbers, and then state the final result.

Participation

- Class participation and discussion are highly encouraged but not mandatory.

Participation

- Class participation and discussion are highly encouraged but not mandatory.

- It is in your responsibility to keep up with the course in a timely fashion.

Participation

- Class participation and discussion are highly encouraged but not mandatory.

- It is in your responsibility to keep up with the course in a timely fashion.

- In particular, if I make announcements during class, I expect that everybody s aware of them.

- Both the midterm and the final exam are **in person**: you must be able to take them at one of the dates/times specified above.

CourseWorks

Class announcements will be sent from and important material is available on CourseWorks:

- You are expected to check the CourseWorks course page regularly.
- A copy of the most recently updated **syllabus** will be on CourseWorks under "Files."
- Lecture notes, assignments and solutions will be posted on CourseWorks under "Files."
- Occasionally, there will be other course related handouts posted on CourseWorks under "Files."
- You are responsible for making sure that CourseWorks announcements are going to an e-mail you check at least once daily.

Campuswire

Campuswire is a Q&A platform designed to get you great answers from classmates and instructors fast:

https://campuswire.com/p/GB4B84815
and use the code: 7076

- Ask questions!
    - ▶ The best way to get answers is to ask questions!
    - ▶ Ask questions on Campuswire instead of e-mailing your teaching staff so everyone can benefit from the response (and so you can get answers from classmates who are up as late as you are).

Campuswire

Campuswire is a Q&A platform designed to get you great answers from classmates and instructors fast:

https://campuswire.com/p/GB4B84815
and use the code: 7076

- Ask questions!
  - The best way to get answers is to ask questions!
  - Ask questions on Campuswire instead of e-mailing your teaching staff so everyone can benefit from the response (and so you can get answers from classmates who are up as late as you are).
- Edit questions and answers wiki-style.
  - Think of Campuswire as a Q&A wiki for your class.
  - By selecting a category when posting a question, you can specifically ask questions about, for example, "Homework 1", "Midterm", or "General."

Campuswire

Campuswire is a Q&A platform designed to get you great answers from classmates and instructors fast:

https://campuswire.com/p/GB4B84815
and use the code: 7076

- Ask questions!
  - ▶ The best way to get answers is to ask questions!
  - ▶ Ask questions on Campuswire instead of e-mailing your teaching staff so everyone can benefit from the response (and so you can get answers from classmates who are up as late as you are).
- Edit questions and answers wiki-style.
  - ▶ Think of Campuswire as a Q&A wiki for your class.
  - ▶ By selecting a category when posting a question, you can specifically ask questions about, for example, "Homework 1", "Midterm", or "General."
- Add a follow-up to comment on or ask further questions.
  - ▶ To comment on or ask further questions about a post, start a follow-up discussion.
  - ▶ Mark it resolved when the issue has been addressed, and add any relevant information back into the Q&A above.