

Math 4261: Mathematical Statistics I

Course introduction and survey sampling
(Rice 7.1-7.2)

Fall 2026

Outline

Course introduction

Logistics

Policies

What is this course about?

Simple random sample

Understanding the bias

Understanding the variability

Understanding the sampling distribution

Encouragement

Outline

Course introduction

Logistics

Policies

What is this course about?

Simple random sample

Understanding the bias

Understanding the variability

Understanding the sampling distribution

Encouragement

Course introduction

This course introduces **mathematical statistics**.

Intended audience

- ▶ Students studying statistics, data science, machine learning, and AI who want to understand the foundations of statistical inference.
- ▶ Students and researchers in areas where statistics is applied who want the first course in a two semester sequence on mathematical statistics.

Prerequisites

Math 3215 or Math 3235 or Math 3670. Some multivariate calculus, linear algebra, and maybe some simple programming. Willingness to learn!

Logistics

Lecture

► Mon/Wed 2:00–3:15pm, Skiles 270

Instructor: Timothy Wee

Email: timothy.wee@gatech.edu

Office hours: Clough 248, Wednesdays 4pm-6pm

Logistics

Course homepage

- ▶ Canvas: notes, homework, announcements, and everything else will be posted here
- ▶ Please read the `Syllabus.pdf`

Grading

- ▶ Homework 35%
- ▶ Two midterms: the higher score contributes 25%, the lower contributes 15%
- ▶ Final exam 25%

Logistics

Homework

- ▶ Approximately weekly; at most 10 total assignments
- ▶ Submitted and graded on Gradescope
- ▶ Due dates will be posted on Canvas and Gradescope
- ▶ Each assignment has 20 points: 10 for completeness and 10 graded
- ▶ Your lowest homework score will be dropped

Logistics

Midterms and Final

- ▶ Held in person; see Canvas for the schedule
- ▶ One crib sheet for Midterm 1, two for Midterm 2, and three for the Final
- ▶ Each crib sheet is two-sided
- ▶ No electronic devices except calculators

Policies

Homework policy

- ▶ You have 8 late days total.
- ▶ A late day is counted as 24 hours after the homework due date.
- ▶ Should you exhaust all late days, every assignment thereafter will accrue a 20% penalty for that assignment per late day.
- ▶ Any particular assignment later than 4 days will not be accepted (**even if you have spare late days**).
- ▶ Please indicate on the first page of any late submission the total number of late days you have accumulated up to and including that assignment.

Resources

Ed discussion forum

- ▶ Link provided on Canvas
- ▶ Strongly encouraged to post questions and answers here!

References

- ▶ Rice, *Mathematical Statistics and Data Analysis*, 3rd edition
- ▶ DeGroot and Schervish *Probability and Statistics*, 4th edition
- ▶ Hogg, McKean, and Craig, *Introduction to Mathematical Statistics* 8th edition

Outline

Course introduction

Logistics

Policies

What is this course about?

Simple random sample

Understanding the bias

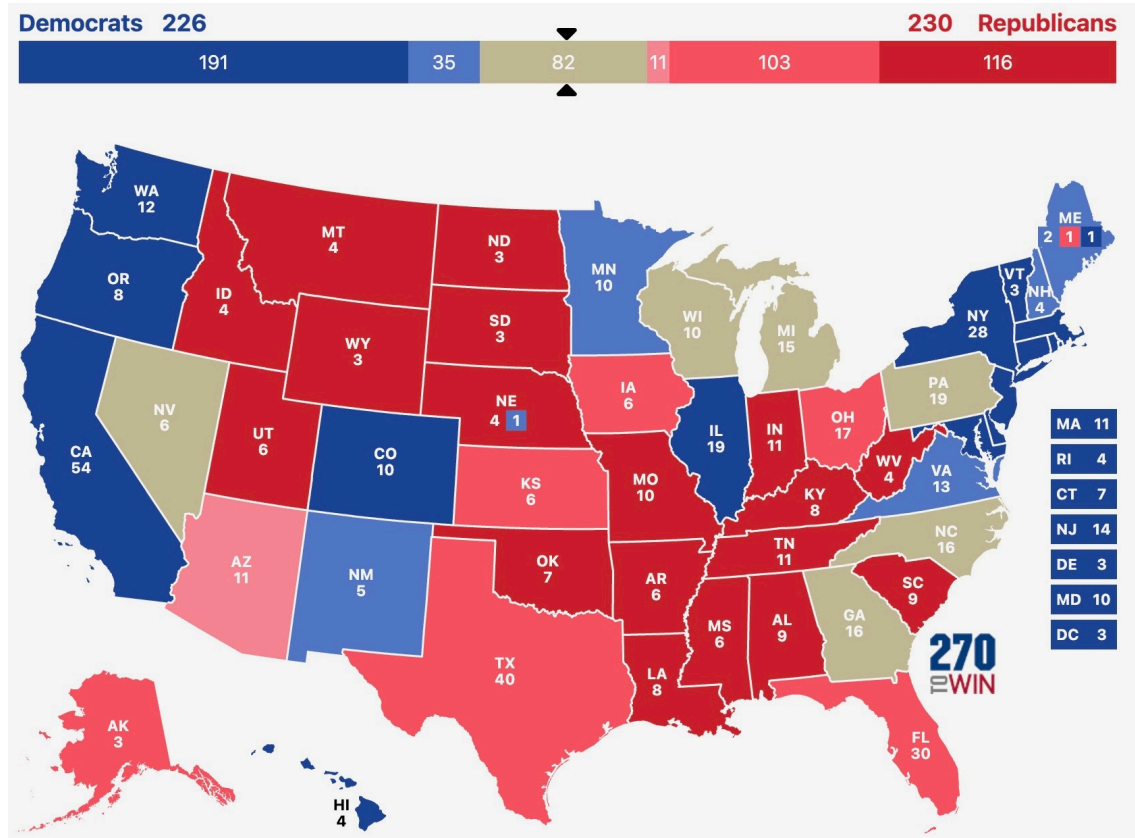
Understanding the variability

Understanding the sampling distribution

Encouragement

Example: polling

2024 Presidential election consensus polling on Nov 4:



270towin.com/maps/consensus-2024-presidential-election-forecast

Example: polling

Estimating the support for Kamala Harris:

$N = 7,007,154$ active voters in Georgia

θ = fraction of active voters who support Harris

$1 - \theta$ = fraction of active voters who support another candidate

$\theta \in [0, 1]$ is an **unknown parameter** we want to understand.

Example: polling

Collecting data:

We could:

1. Survey every active voter (Infeasible, expensive)
2. Survey “representative” voters, determined and selected by an investigator (Needs to be free from bias).
3. Survey a **random** subset of active voters in Georgia.

Advantage of **random** sample

Enables mathematical quantification of the uncertainty and magnitudes of error in inferences drawn from the data.

Simple random sample

Suppose we poll a **simple random sample** of $n = 1000$ people from Georgia. This means:

- ▶ Person 1 is chosen randomly (equally likely) from all N registered voters in Georgia. Then person 2 is chosen at random from the remaining $N - 1$ people. Then person 3 chosen at random from remaining $N - 2$ people, etc.
- ▶ Equivalently, all $\binom{N}{n} = \frac{N!}{n!(N-n)!}$ possible sets of n people are equally likely to be chosen as our sample.

Simple random sample

Suppose 470 out of a random sample of 1000 active voters say they support Harris. Possible estimate:

$$\hat{\theta} = \frac{\# \text{ in sample that support Harris}}{1000} = 0.47.$$

We want to use $\hat{\theta}$ as a proxy for the unknown

$$\theta = \frac{\# \text{ of active voters in GA that support Harris}}{N}.$$

Example: polling

Questions of interest:

- ▶ Is $\theta > 0.5$? (Hypothesis testing)
- ▶ How close is $\hat{\theta}$ to θ ? How much uncertainty is in the estimate? (Parameter estimation, confidence intervals)
- ▶ Other ways to estimate θ ? (Maximum likelihood, method of moments, Bayesian approach)
- ▶ Given a new voter with certain characteristics, can we predict if they support Harris? (Regression)

Simple random sample

Let's call our **data** $X_1, \dots, X_n$:

$$X_i = \begin{cases} 1 & \text{if the } i\text{-th sampled person supports Harris} \\ 0 & \text{if the } i\text{-th sampled person does not support Harris} \end{cases}$$

$$\text{Then } \hat{\theta} = \frac{X_1 + X_2 + \dots + X_n}{n}.$$

The data $X_1, \dots, X_n$ are random, because we took a random sample. Therefore $\hat{\theta}$ is a **random variable**.

Example: marginals in sampling

Setting: the population contains N people, and a proportion θ of them support Harris. Let

$$X_i = \begin{cases} 1 & \text{if the } i\text{-th sampled person supports Harris} \\ 0 & \text{if the } i\text{-th sampled person does not support Harris} \end{cases}$$

Compare the following two sampling schemes:

1. Sample n people **with replacement**
2. Sample n people **without replacement**.

For each scheme, find the **distribution** of X_i .

Example: Marginals in sampling

$$X_i = \mathbf{1}\{\text{the } i\text{-th sampled person supports Harris}\}$$

Sampling with replacement

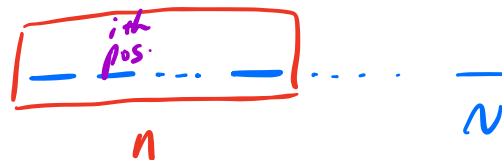
$$X_i \sim \text{Ber}(\theta)$$

Example: Marginals in sampling

$$X_i = \mathbf{1}\{\text{the } i\text{-th sampled person supports Harris}\}$$

Sampling without replacement

$$X_i \sim \text{Ber}(\theta) \quad \text{too!}$$



Any of the N people is equally likely to be the i -th sampled person. This probability is $\frac{1}{N}$.

There are $N\theta$ Harris supporters.

$$P[X_i = 1] = N\theta \cdot \frac{1}{N} = \theta.$$

Joint distribution in sampling

By definition

$$\sum_{i=1}^n X_i = \text{number of sampled people who support Harris.}$$

With replacement, the X_i 's are **independent**. Without replacement, they are **dependent**. This leads to different joint distributions.

$$\sum_{i=1}^n X_i \sim \begin{cases} \text{Binomial}(n, \theta), & \text{with replacement,} \\ \text{Hypergeometric}(N, N\theta, n), & \text{without replacement}^1. \end{cases}$$

¹It is not necessary to memorize every named distribution. Before each test you will be informed which ones are important. 

The sampling distribution of $\hat{\theta}$

$\hat{\theta}$ is a value computed from our observed data. It is a **statistic**. The probability distribution of $\hat{\theta}$ is called its **sampling distribution**.

We may be interested in the following properties:

- ▶ What is the mean of $\hat{\theta}$? If the mean is equal to θ , the $\hat{\theta}$ is an **unbiased** estimator for θ . Otherwise $\hat{\theta}$ is biased.
- ▶ What is the standard deviation of $\hat{\theta}$? This is called the **standard error** and it quantifies the variability of our estimate $\hat{\theta}$.
- ▶ What are the quantiles of $\hat{\theta}$? This information allows us to create **confidence intervals** for θ .

Understanding the bias

For each $i = 1, \dots, n$, by symmetry, the i^{th} person sampled is equally likely to be each of the N individuals.

Probability that person i supports Harris is $\frac{N\theta}{N} = \theta$. So

$$X_i \sim \text{Bernoulli}(\theta), \quad \mathbb{E} X_i = \theta \quad \text{for all } i = 1, \dots, n$$

Recall that expectation is **linear**:

$$\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y], \quad \mathbb{E}[cX] = c\mathbb{E}[X] \quad \text{for constant } c.$$

Thus

$$\mathbb{E}[\hat{\theta}] = \mathbb{E}\left[\frac{X_1 + \dots + X_n}{n}\right] = \frac{1}{n} \cdot n \cdot \mathbb{E}[X_i] = \theta$$

The mean of $\hat{\theta}$ is θ , so $\hat{\theta}$ is **unbiased**. (If we were to repeat the survey many times, on average the value of $\hat{\theta}$ would be θ).

Understanding the variability

Unbiasedness doesn't tell us how far off $\hat{\theta}$ is from θ , for a single survey. For this we should look at the **standard deviation**

$$\sqrt{\text{Var}[\hat{\theta}]}$$

(also called std. err)

Consider first a simpler setting, where we sample $n = 1000$ individuals **with replacement** from a population of size N .

In this case the $X_1, \dots, X_n$ would be **independent**. The event that the i^{th} sampled person supports Harris is independent of all other samples.

Understanding the variability *(sampling with replacement)*

Recall that

$$\text{Var}[cX] = c^2 \text{Var}[X] \quad \text{for constant } c.$$

If X and Y are **independent** then also

$$\text{Var}[X + Y] = \text{Var}[X] + \text{Var}[Y]$$

$$\text{So } \text{Var}[\hat{\theta}] = \text{Var}\left[\frac{X_1 + \dots + X_n}{n}\right]$$

$$= \frac{1}{n^2} (\text{Var}[X_1] + \dots + \text{Var}[X_n])$$

$$= \frac{1}{n^2} n \cdot \text{Var}[X_1]$$

$$= \frac{\theta(1-\theta)}{n}$$

$$X_1 \sim \text{Bernoulli}(\theta)$$

$$\text{i.e. } \begin{aligned} P[X_1 = 1] &= \theta \\ P[X_1 = 0] &= 1 - \theta \end{aligned}$$

$$\begin{aligned} \text{Var } X_1 &= E[X_1^2] - (E[X_1])^2 \\ &= \theta - \theta^2 = \theta(1-\theta) \end{aligned}$$

Understanding the variability

In the original setting, the $n = 1000$ individuals are sampled **without replacement**. This introduces dependence between the X_i 's:

Suppose $X_1 = 1$, i.e. person 1 supports Harris. Conditional on $X_1 = 1$, the probability that $X_2 = 1$ is now $\frac{N\theta-1}{N-1}$ instead of θ . So X_1 and X_2 are dependent.

Another way to compute $\text{Var}[\hat{\theta}]$ is

$$\text{Var}[\hat{\theta}] = \mathbb{E}[\hat{\theta}^2] - (\mathbb{E}\hat{\theta})^2.$$

The second term is just $(\mathbb{E}\hat{\theta})^2 = \theta^2$.

Understanding the variability

For the first term,

$$\mathbb{E}[\hat{\theta}^2] = \mathbb{E}\left[\left(\frac{X_1 + \dots + X_n}{n}\right)^2\right]$$

$$= \frac{1}{n^2} \mathbb{E}\left[\sum_{i=1}^n X_i^2 + \sum_{i \neq j} X_i X_j\right]$$

$$= \frac{1}{n^2} \cdot n \mathbb{E}[X_1^2] + \frac{1}{n^2} \sum_{i \neq j} \mathbb{E}[X_i X_j]$$

$$= \frac{1}{n} \mathbb{E}[X_1^2] + \frac{1}{n^2} n(n-1) \mathbb{E}[X_1 X_2]$$

$\underbrace{\quad}_{= X_1}$

$$= \begin{cases} 1 & \text{if both } X_1, X_2 = 1 \\ 0 & \text{otherwise} \end{cases}$$

$$\left(\sum_i a_i\right)^2 = \sum_i a_i^2 + \sum_{i \neq j} a_i a_j$$

Understanding the variability

$$\begin{aligned}\mathbb{E}(X_1 X_2) &= P[X_1 = 1, X_2 = 1] \\ &= P[X_1 = 1] P[X_2 = 1 \mid X_1 = 1] \\ &= \theta \cdot \frac{N\theta - 1}{N-1}\end{aligned}$$

Hence

$$\begin{aligned}\text{Var } \hat{\theta} &= \mathbb{E}[\hat{\theta}^2] - (\mathbb{E} \hat{\theta})^2 \\ &= \frac{1}{n} \mathbb{E}[X_1^2] + \frac{n-1}{n} \mathbb{E}[X_1 X_2] - (\mathbb{E} \hat{\theta})^2 \\ &= \frac{1}{n} \theta + \frac{n-1}{n} \theta \cdot \frac{N\theta - 1}{N-1} - \theta^2 \\ &= \frac{\theta(N-\theta)}{n} \left(1 - \frac{n-1}{N-1}\right)\end{aligned}$$

Understanding the variability

$$\text{with replacement: } \text{Var}[\hat{\theta}] = \frac{\theta(1-\theta)}{n}$$

Combining the above, for sampling without replacement

$$\text{Var}[\hat{\theta}] = \frac{\theta(1-\theta)}{n} \left(1 - \frac{n-1}{N-1} \right).$$

If $N \gg n$, variances coincide.

For $N = 7,007,154$, $n = 1000$, and $\hat{\theta} = 0.47$, the standard error is approximately

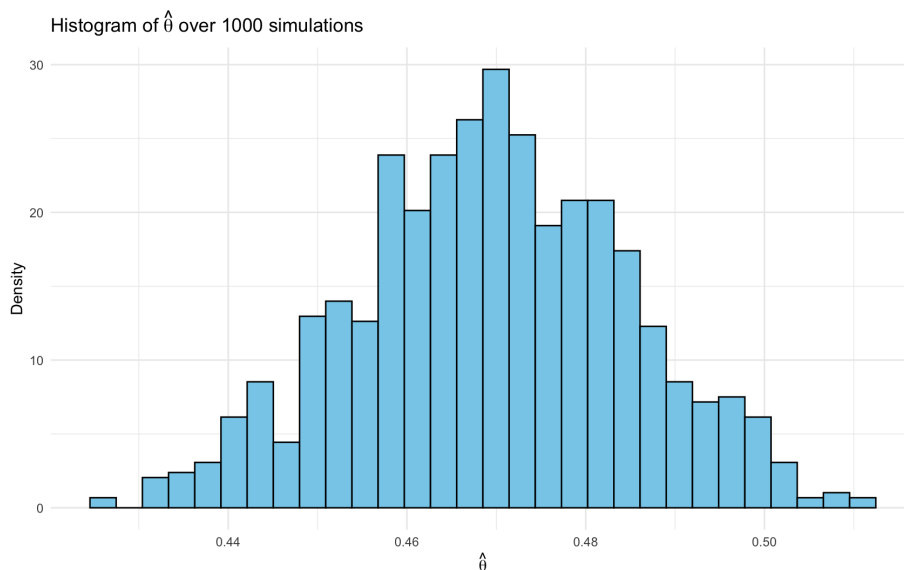
$$\sqrt{\text{Var}[\hat{\theta}]} \approx 0.016$$

It is ‘not unlikely’ for $\theta \approx 0.45$ and ‘highly unlikely’ that $\theta \approx 0.41$.

Understanding the sampling distribution

To make ‘not/highly unlikely’ precise, we consider the distribution of $\hat{\theta}$. Let’s **simulate** $X_1, \dots, X_n$ from a population of N people, of which a fraction θ support Harris (supposing $\theta = 0.47$). Then compute $\hat{\theta}$.

Histogram of values of $\hat{\theta}$:



Understanding the sampling distribution

The histogram of $\hat{\theta}$ looks like a **normal bell curve**, with mean 0.47 and standard deviation 0.016. Why?

Again, suppose we had sampled **with replacement**, Then $X_1, \dots, X_n$ are independent. By the **Central Limit Theorem**, if n is large, then

$$\sqrt{n}(\hat{\theta} - \theta) = \sqrt{n}\left(\frac{X_1 + \dots + X_n}{n} - \theta\right) \sim \mathcal{N}(0, \theta(1 - \theta))$$

approximately. Some rearranging shows that the distribution of $\hat{\theta}$ is approximately $\mathcal{N}\left(\theta, \frac{\theta(1-\theta)}{n}\right)$.

A confidence statement

For sampling **without replacement**, the same normal approximation holds, provided the sample size n is large but much smaller than population size N .

Roughly 95% of the probability density of a normal distribution is within two standard deviations of the mean. Thus, approximately,

$$(0.47 - 2(0.016), 0.47 + 2(0.016)) = (0.438, 0.502)$$

is a 95% confidence interval for θ .

Statistical paradigm

In statistical analyses, we often assume that

Data is a realization of a random process

Reasonable in many circumstances, for instance

- ▶ We introduced randomness by design. E.g. sampling, randomized clinical trials, A/B testing.
- ▶ We are studying actual random phenomena. E.g. coin flips, Brownian motion.
- ▶ We want to model phenomena we cannot explain. E.g. measurement error, noise in signal transmission.

Statistics and probability

Probability provides the language and tools to perform statistical analysis.

One viewpoint is that

$$\boxed{\text{Statistics} = \text{Probability}^{-1}}$$

- ▶ **Probability:** given a random model, what can we say about the data generated from it?
- ▶ **Statistics:** Given observed data, what random model does it come from? Can we learn it?

Topics

- ▶ Probability tools: joint distributions, moment generating functions, limit theorems, and the delta method
- ▶ Sampling distributions: normal, χ^2 , t , F , and order statistics
- ▶ Parameter estimation: method of moments, maximum likelihood, Bayesian methods, consistency, and asymptotic normality
- ▶ Confidence intervals and the basics of hypothesis testing

Outline

Course introduction

Logistics

Policies

What is this course about?

Simple random sample

Understanding the bias

Understanding the variability

Understanding the sampling distribution

Encouragement

Encouragement

‘Young man, in mathematics you don’t understand things. You just get used to them.’

John von Neumann to Felix Smith

‘Mathematics, you see, is not a spectator sport.’

Georgia Pólya

Encouragement

‘Using fancy tools like neural nets, boosting, and support vector machines without understanding basic statistics is like doing brain surgery before knowing how to use a band-aid.’

Larry Wasserman, preface to *All of Statistics*

