STAT 7301 — Advanced Statistical Theory

Syllabus · Autumn 2026

- Autumn 2026, 3 credit hours
- MWF 9:10–10:05 in Cockins Hall 228

Course overview

Instructor

Vincent Q. Vu, Ph.D.

- Ph.D., Statistics, University of California, Berkeley
- M.A., Statistics, University of California, Berkeley
- B.A., Statistics, University of California, Berkeley
- Associate Professor, Department of Statistics, The Ohio State University

Contacting me

- Email: vqv@stat.osu.edu
- Office: Cockins Hall 428B

When emailing, please include "STAT 7301" in the subject line so I can find it.

Office hours:

- Wednesdays, 3:00 – 4:00 PM (September 2 – November 18)
- Mondays, 3:00 – 4:00 PM (November 23 – December 7)

See the class calendar on CarmenCanvas for the up to date schedule.

Office hours will be in person, Cockins Hall 428B. They are drop-in and you do not need a reason to come. If the posted time does not work for you, email me and we will find another one.

About this course

Course description

Fundamental concepts from mathematical statistics, derivation and classification of estimators, large sample asymptotic analysis, and non-asymptotic analysis of high-dimensional estimation.

STAT 7301 is a course on the fundamentals of statistical theory, intended for second-year Ph.D. students in Statistics. It develops the theory of point estimation in three movements: the finite-sample, decision-theoretic core (models, sufficiency, exponential families, information); the classical large-sample asymptotic theory of M- and Z-estimators and the MLE; and the non-asymptotic, high-dimensional theory built on concentration inequalities, where the dimension of the parameter is allowed to grow with the sample size.

The lectures and the notes distributed with them are the canonical source for the course. The textbooks listed below are supplemental reading, and the notation, nomenclature, and depth of treatment in lecture will sometimes differ from them. I will try to flag those differences; you are ultimately responsible for noticing them yourself.

What I assume you already have

Mathematical analysis and probability theory are the working tools of statistical theory. **You are expected to be able to read and write rigorous mathematical proofs.** If you are unsure whether your background is sufficient, talk to me in the first week rather than the fifth.

Attendance and participation

I expect you to attend every class meeting and to participate — ask questions, answer them, object when something does not follow. Much of what this course teaches is what a complete mathematical argument sounds like, and that is transmitted in the room. I may occasionally take attendance. If you must miss a class, let me know in advance when you can, and get the notes from a classmate.

Course expected learning outcomes

By the end of this course, students should successfully be able to:

1. Understand important and fundamental concepts of mathematical statistics.
2. Formulate new and classify existing estimators in a variety of statistical problems.
3. Analyze theoretical properties of estimators in large sample settings using asymptotics.
4. Analyze theoretical properties of estimators in high-dimensional settings using concentration inequalities.
5. Understand the assumptions that underly basic tools of theoretical analysis, and how those assumptions affect the use and results of these tools.

Course materials and technologies
Required
Lecture notes written by the instructor for the exclusive use of students enrolled in this course are distributed through CarmenCanvas and are the canonical source for the material.

Recommended
- [**Keener**] Keener, R. W. (2010). *Theoretical Statistics: Topics for a Core Course.* Springer. link.springer.com/book/10.1007/978-0-387-93839-4
- [**Vershynin**] Vershynin, R. (2018). *High-Dimensional Probability: An Introduction with Applications in Data Science.* Cambridge University Press. The author's final pre-publication version is freely available at www.math.uci.edu/~rvershyn/papers/HDP-book/HDP-book.html.
- [**Wainwright**] Wainwright, M. J. (2019). *High-Dimensional Statistics: A Non-Asymptotic Viewpoint.* Cambridge University Press. Free full text via OSU Libraries.

Course technology
- **CarmenCanvas** (carmen.osu.edu) carries the schedule, assignments, notes, and announcements. Check your notification preferences so that you receive announcements.
- **Gradescope**, accessed through CarmenCanvas, is where you submit homework and where graded work is returned.
- A way to produce legible mathematics on paper. Handwriting is entirely acceptable. If you prefer to typeset, LaTeX (tug.org/texlive) or Typst (typst.app) will both do, and learning one now is a good investment for your dissertation. Overleaf (overleaf.com) is a convenient online LaTeX editor.

Grading and assessment

How your grade is calculated

Assignment category	Weight
Homework *(8 assignments, equally weighted)*	20%
Midterm exam *(in class, Friday November 13)*	40%
Final exam *(Monday December 14, 10:00-11:45 AM)*	40%
Total	**100%**

STAT 7301 is a core course for the PhD qualifying exam, and the assessment is shaped to match: most of your grade comes from two closed-book exams, written under the same conditions the qualifier is.

Tell me about any scheduling conflict with a scheduled exam **at least three weeks in advance**.

Homework

There are **eight graded homework assignments**, roughly one every other week — due on Fridays until Thanksgiving and on Wednesdays after. (Homework 0, in the first week, is a warm-up: work it, but do not hand it in.) **Assignments are posted on CarmenCanvas**, where each one appears when it is assigned — they are not handed out in class. Each is due at **9:10 AM (the start of class)** on its due date; the dates are in the schedule below and in CarmenCanvas. **Submit one PDF to Gradescope through CarmenCanvas**, tagging the pages of each problem when you upload. Handwritten work is fine — export a PDF from your tablet or scan paper with the Gradescope mobile app, not a phone photo. **I do not accept homework by email.**

Proof-writing is the skill this course exists to build, and it is built by writing proofs — not by reading them. Write your solutions in complete sentences. An answer that is correct but unreadable is not finished.

Collaboration. You may work with other students on homework problems, and I encourage it. Each of you must write up your own solutions in your own words, from your own understanding. If you collaborate, **name your collaborators on your submission.**

Old solutions. I reuse problems from earlier offerings of this and similar courses. **Do not look at or copy solutions from previous offerings**, from solution manuals, or from the internet. A copied proof fails the standard in the AI policy below for the same reason a generated one does: you cannot defend what you did not build.

Late assignments

In general, **late homework will not be accepted**. In the case of an emergency (medical, family, etc.), please let me know as soon as possible so that I can work with you to give a reasonable extension.

Exams

There are **two closed-book exams**: a midterm exam, and a final exam. All exams will take place in the meeting classroom. The dates are:

- **Midterm exam — Friday, November 13**, in class.
- **Final exam — Monday, December 14, 10:00–11:45 AM**.

What each exam covers is announced in class and on CarmenCanvas well in advance of the sitting. I would rather tell you where an exam actually lands, a week or two out, than commit in August to a boundary the semester may not respect.

The midterm is based largely on the material of the homework assignments preceding it. If you can do the homework without looking anything up, you are ready.

No make-up exams are given. This is why conflicts must reach me three weeks ahead, while there is still time to arrange an alternative sitting.

The final-exam slot is fixed by the Registrar from our meeting pattern — a Columbus MWF 9:10 class sits at Monday, December 14, 10:00–11:45 AM — and it is not mine to move. **Do not book travel out of Columbus before the afternoon of December 14.**

Grading scale

A	A-	B+	B	B-	C+	C	C-	D+	D	E
93+	90	87	83	80	77	73	70	67	60	<60

Each number is the minimum percentage earning that grade.

Academic integrity and the use of AI

Generative AI and large language models

You may use an LLM while working on homework, and you do not need permission. Use one to review background you are missing, to hunt for a counterexample, to check an algebraic step, or to ask whether a proof you have written is correct. What you may not do is submit reasoning you cannot reconstruct and defend on your own.

The standard is concrete. For every proof you hand in, you must be able to reproduce its argument at the board, without notes and without a machine, and say why each hypothesis is needed. I will occasionally ask you to do exactly that. If you cannot, the work is not yours, whatever produced it.

Disclosure. At the end of each assignment, add a sentence or two saying where you used an LLM and what for — *“asked a chatbot for a counterexample in 4(b); it gave a wrong one; found my own.”* This is not a confession and it does not cost you points. An assignment with no disclosure is a claim that you used none. Undisclosed use is a separate problem from misuse, and a worse one.

Why this rule and not a ban. The mathematics in this course is exactly the kind of writing current models are most fluent at and least reliable in. A chatbot will produce a proof of a false statement in the correct register, with the right vocabulary and a confident tone, and the error will be a single unjustified step in the middle. Learning to *find that step* is not a distraction from learning mathematical statistics — it is the same skill as reading a paper, refereeing, and checking your own work, and it is the skill this course is actually for. A ban would be unenforceable and would teach you nothing about it.

Exams are different. All exams are closed book: no notes, no devices, no tools of any kind.

The policy above appears verbatim in the header of **every homework assignment** — it is written once and included in both places, so the two can never drift apart. Assignments from Homework 1 onward also carry a recurring problem that asks you to have an LLM critique one of your own proofs and then to rule on the critique — you are graded on the ruling, not on whether the critique was right.

If you are unsure whether a particular use is within the rule, ask me. That question has never gotten anyone in trouble. Work that fails the standard above is academic misconduct and is referred to the Committee on Academic Misconduct.

Academic integrity in this course

The work you submit must be your own. In this course specifically that means:

- **Homework** may be discussed with classmates; the write-up must be yours, and collaborators must be named.
- **Solutions from previous offerings**, from solution manuals, and from the internet are off limits while you are working on an assignment.
- **All exams** are individual and closed book. No collaboration, no notes, no devices, no AI tools, no exceptions.
- **Disclosure** of collaboration and of LLM use is required on every submission, and disclosing costs you nothing.
- **The test in every case** is the one in the AI policy above: you must be able to reproduce and defend, unaided, any derivation/argument with your name on it.

I am required to report suspected academic misconduct to the Committee on Academic Misconduct, and I do.

Course schedule

Semester at a glance

The semester in outline. Assignment due dates and the meeting-by-meeting topics are in the detailed schedule below and in CarmenCanvas.

Week	Unit
1-2	Foundations
2-4	Exponential families
4-5	Risk and information
5-7	Methods of estimation
8-11	Asymptotic theory
12	Concentration inequalities
12	Midterm exam
13	Concentration inequalities
13-15	Sparse mean estimation
15	Sparse linear models (lasso)
16	Principal component analysis
17	Final exam

This schedule represents the plan as of the first day of the semester and is subject to revision. You are expected to attend class and to check CarmenCanvas regularly for updates.

Detailed schedule

Assignment due dates are also in CarmenCanvas, which is authoritative for them — including any individually adjusted deadline.

Week	Date	Class	Topics and assignments
1	8/26	1	Course overview; statistical models; estimation and inference
	8/28	2	Decision theory
2	8/31	3	Sufficiency and minimal sufficiency
	9/2	4	Exponential families I: definitions and examples
			HW 0 (warm-up) due
	9/4	5	Exponential families II: minimal representations
3	9/7		**No class** (Labor Day)

Week	Date	Class	Topics and assignments
	9/9	6	Exponential families III: regularity and overcompleteness (multinomial)
	9/11	7	Exponential families IV: moments and mean parameters
			HW 1 due
4	9/14	8	Exponential families V: worked examples and catch-up
	9/16	9	Convex loss functions; Jensen's inequality
	9/18	10	Rao–Blackwell theorem; randomization and sufficiency
5	9/21	11	Fisher information
	9/23	12	Information inequality; the Cramér–Rao lower bound
	9/25	13	Methods of estimation: plug-in, method of moments; MLE as a derivation principle
			HW 2 due
6	9/28	14	MLE under differentiability (score / likelihood equations) and under convexity
	9/30	15	MLE in exponential families: existence, uniqueness; MLE = MOM
	10/2	16	Generalized linear models
			HW 3 due
7	10/5	17	M- and Z-estimation
	10/7	18	Catch-up and flex
	10/9	19	Catch-up and flex
8	10/12	20	Asymptotics: meaning and uses; convergence concepts; WLLN, CLT
	10/14	21	Consistency of the MLE in exponential families
	10/15		**No class** (Autumn break)
	10/16		**No class** (Autumn break)
9	10/19	22	Stochastic O/o calculus; Cramér–Wold, the multivariate CLT, Slutsky
	10/21	23	Delta method: Taylor expansion, theorem, proof, examples
	10/23	24	Delta method examples; variance stabilization; catch-up
			HW 4 due
10	10/26	25	Consistency of M-estimators
	10/28	26	ULLN for random functions on C(K)

Week	Date	Class	Topics and assignments
	10/30	27	Consistency of the MLE under correct specification and under misspecification
11	11/2	28	Asymptotic normality of Z-estimators
	11/4	29	Asymptotic normality and efficiency of the MLE
	11/6	30	One-step and Newton estimators
			HW 5 due
12	11/9	31	Concentration I: sub-Gaussian distributions and norms
	11/11		**No class** (Veterans Day)
	11/13	32	**Midterm exam**
13	11/16	33	Concentration II: Hoeffding and Chernoff bounds
	11/18	34	Concentration III: MGF method; maxima of sub-Gaussians; union bound
	11/20	35	Sparse mean estimation I: thresholding; risk via Hoeffding and the union bound
			HW 6 due
14	11/23	36	Sparse mean estimation II: risk over $\ell 0$ balls; comparison with the oracle
	11/25		**No class** (Thanksgiving break)
	11/26		**No class** (Thanksgiving break)
	11/27		**No class** (Thanksgiving break)
15	11/30	37	Catch-up and flex
	12/2	38	Lasso I: the high-dimensional linear model; $\ell 1$ relaxation; subgradients and KKT; the basic inequality and the slow rate
			HW 7 due
	12/4	39	Lasso II: the master inequality; the cone condition; the fast rate under a restricted eigenvalue condition
16	12/7	40	PCA I: matrix concentration — sub-exponential and Bernstein bounds, ε-nets, the operator norm of the sample covariance
	12/9	41	PCA II: the master inequality reused; the curvature lemma; Davis–Kahan; wrap-up
			HW 8 due
17	12/14		**Final exam** 10:00 AM

University policies

This course is governed by the university-wide academic policies described in the syllabus policies and statements maintained by Ohio State, including academic misconduct, accommodations for disability and illness, religious accommodations, mental health support, and harassment, discrimination and sexual misconduct. Those statements are current on that page; this syllabus does not restate them.

Where this syllabus states a course-specific policy — LLM use, late work, collaboration, exams — that policy applies in addition.