Teaching Statement

My goal in teaching is to help students become critical, independent thinkers in statistics and data science, equipped with solid theoretical foundations, computational skills, and an awareness of the opportunities and risks of modern AI tools. I want students to see statistics not as a list of formulas but as a coherent way of reasoning under uncertainty—connected to real scientific, engineering, and policy problems. My teaching and mentoring experience spans Ph.D.-level courses in data science computing, asymptotic statistics, and natural language processing, as well as volunteer teaching in under-resourced high schools in China. At Tsinghua, I am prepared and enthusiastic to teach both core undergraduate courses (probability, mathematical statistics, regression, statistical learning) and advanced graduate courses in high-dimensional and distributed methods, causal inference, and genAI-enhanced statistical workflows.

1 Teaching Philosophy

When preparing a class, I start from three questions:

1. What is the core idea I want students to take away?
2. How does this idea connect to real-world applications and data-analytic practice?
3. What obstacles—conceptual, technical, or psychological—might students face in learning it?

These questions keep the students and their long-term learning at the center. I organize each lecture around a storyline: beginning with an intuitive question or motivating example, moving to formal definitions and theorems, and ending with a concrete data or algorithmic example that students can manipulate themselves. This narrative approach helps students understand why a concept matters before they invest effort in learning how it works technically.

A central principle in my teaching is to connect theory and practice tightly. For example, when introducing maximum likelihood and M-estimation, I not only present asymptotic results but also show how these estimators arise from optimization problems, how they behave under misspecification, and how they perform on real or simulated data. Students see theoretical results such as consistency and asymptotic normality come to life in simulation studies they implement themselves. This bridges abstract probability with the computational workflow they will use in research and industry.

2 Teaching Methods and Classroom Practices

I strive to design courses that are rigorous, interactive, and inclusive.

- Layered explanations. I typically introduce a concept informally, then formalize it, and finally return to an example. For instance, when teaching central limit theorems or high-dimensional asymptotics, I start with simulation-based visualizations, then move to assumptions and proofs, and then discuss failure modes in heavy-tailed or dependent settings.

- Active learning. To avoid long stretches of passive listening, I incorporate short in-class activities: quick derivations, think–pair–share questions on model assumptions, or small group discussions about why a particular estimator might fail in a given scenario. These provide real-time feedback on students' understanding and create space for quieter students to participate.

- Theory + computation in assignments. Homework sets mix proof-based questions (e.g., verifying regularity conditions, deriving influence functions, analyzing convergence) with coding tasks (implementing distributed gradient methods, bootstrap procedures, or mediation estimators in R or Python). Students learn that careful reasoning and careful implementation are both essential.

- Open-ended projects. In more advanced courses, I assign projects where students choose a dataset—such as a health, environmental, or social dataset—and must: formulate a question, justify their modeling choices, assess assumptions, and communicate their findings clearly. This prepares them for real research and interdisciplinary collaboration.

Creating a supportive environment is equally important. I work to normalize questions and confusion: I explicitly tell students that struggling with a concept is expected in a rigorous statistics course, and I share strategies for breaking large problems into manageable steps. I emphasize constructive feedback in office hours and grading, aiming to show students how they can improve and not only where they fell short.

3 Teaching in the Era of genAI

Large language models and other genAI tools are already affecting how students learn programming, statistics, and data science. I believe our responsibility as educators is to teach students to use these tools critically and responsibly, not to ignore them.

In my courses, I plan to:

- Make limitations visible. I will demonstrate cases where genAI suggestions for code or analysis are subtly incorrect or statistically invalid—e.g., mis-specified models, misuse of p-values, or unjustified causal claims—and ask students to diagnose the problems.

- Clarify allowed uses. I will design assignments where limited use of genAI (for brainstorming or minor code assistance) is permitted, but where core reasoning, derivations, and interpretation must be the student's own work.

- Integrate genAI into evaluation. In advanced classes, I would consider assignments that explicitly ask students to critique an AI-generated analysis using the tools from the course (diagnostics, model checks, robustness analysis), turning genAI from a shortcut into a learning object.

This approach helps students see genAI as a tool to be guided by statistical principles, not as an authority.

4 Teaching and Mentoring Experience

My formal roles include serving as a teaching assistant for:

- Data Science Computing / Computer Skills for Data Science (Ph.D. level). I led lab sessions on R and Python programming, simulation studies, and reproducible workflows. I designed exercises where students implemented methods related to my research, such as robust regression or simple distributed algorithms, and then compared them empirically.

- Asymptotic Statistics (Ph.D. level). I ran discussion sections covering convergence concepts, likelihood theory, M-estimation, and bootstrap methods. I prepared solution sketches that highlighted the logical structure of proofs and used office hours to connect measure-theoretic probability with practical inference tasks students encountered in other courses and research.

- Natural Language Processing (graduate level). I helped students understand probabilistic modeling, regularization, and evaluation metrics, and I highlighted how statistical thinking underlies modern deep learning and language models.

Beyond formal courses, I have mentored junior students and collaborators on research projects that led to publications in journals such as JCGS and Statistica Sinica. In supervising these projects, I structure the work into stages: understanding the literature and problem formulation; designing simulations; developing theoretical results; and writing and revising manuscripts. I have found that regular, structured meetings and written milestones (short memos or draft sections) are very effective in helping students move from "following instructions" toward owning a research direction.

Earlier, as an undergraduate, I participated in volunteer teaching programs at high schools in Chongqing and Fujian. There, with students facing large resource and preparation gaps, I saw how small actions—reviewing background material, explicitly acknowledging different starting points, providing concrete learning plans—could change students' attitudes toward mathematics and science. This experience continues to inform how I support students from diverse backgrounds in university classrooms.

5 Teaching Interests at Tsinghua

At Tsinghua's Department of Statistics and Data Science, I am eager to teach both core courses and specialized electives that align with the department's strengths at the intersection of statistics, optimization, and AI.

At the undergraduate level, I am prepared to teach:

- Probability and Mathematical Statistics
- Regression Analysis and Generalized Linear Models
- Introduction to Statistical Learning / Data Science
- Statistical Computing (R / Python, reproducible workflows)

At the graduate level, I look forward to teaching and developing courses such as:

- High-Dimensional and Robust Statistics
- Distributed and Decentralized Statistical Learning
- Causal Inference and Mediation Analysis
- Quantile and Composite-Quantile Methods
- GenAI-Enhanced Statistical Inference and Modern Statistical Computing

These courses would support students in statistics, data science, and related disciplines (e.g., computer science, engineering, public health), and could be adapted for both theoretical and applied audiences.

6 Mentoring and Advising Approach

As an advisor, I aim to meet students where they are, while steadily raising expectations and fostering independence. Different students need different kinds of interaction: some benefit from frequent, hands-on guidance, while others thrive with more autonomy. My approach is to:

- Have regular one-on-one meetings to discuss progress, questions, and long-term goals.
- Encourage students to formulate and refine their own research questions, and to present ideas in short written notes or informal talks—skills essential for successful researchers.
- Provide clear, detailed feedback on drafts and proposals, helping students see how to sharpen arguments, clarify assumptions, and position their contributions.
- Support students in presenting at seminars and conferences and in applying for fellowships or awards.

In all these activities, I see teaching and mentoring as deeply connected to my research: together, they help students acquire the tools and confidence to engage with—and ultimately push forward—the evolving landscape of statistics, data science, and AI.