Stats News 2022

Harvard University
Department of Statistics

HARVARD
Faculty of Arts and Sciences
DEPARTMENT OF STATISTICS

Table of Contents

Letter from the Chair

Dr. Samuel Kou,

Chair of the Department

Greetings Statistics Community,

As I embark on this journey as the new chair, I want to first acknowledge the contributions of former Chair Neil Shephard to the department for the past seven years. Neil did a remarkable job of recruiting and mentoring a cohort of outstanding faculty members (including five new women faculty members) and deftly steered us through pandemic times. Thanks to Neil (and Xiao-Li Meng's encouragement!), we are launching our department newsletter to help alums and current students, postdocs, and faculty stay connected at Harvard and beyond.

This year Neil also oversaw planning for the Visiting Committee, a group of faculty and distinguished professionals from different institutions who met with our chair and statistics community in March 2022 to assess departmental needs and make recommendations. Remarking that "the Department is on the cusp of becoming a real powerhouse," the committee's report highlighted our award-winning faculty (for both research and teaching), career support for junior faculty, and a growing demand for our undergraduate courses. The committee's report also commented on areas of growth, such as increasing full-time faculty positions, graduate student career and funding resources, machine learning course options, and representation of women and underrepresented minorities at all levels.

I see immense opportunities to make progress in these areas:

- With our next round of faculty searches, we will strategize to draw the most talented, diverse candidates. Leveraging the interdisciplinary nature of our field, we will also seek out candidates for joint appointments with other departments.

- For our graduate students' success, we have hired Senior Research and Financial Administrator Laverne Martinez to assist with grant applications. We are planning more professional development opportunities to connect with alums,

including a new LinkedIn group.

- With the growing importance of machine learning, the department has built a new Statistical Machine Learning track for undergraduates and will prioritize developing future course options for graduate students.

- We will collaborate with the Dean of FAS to continue to follow best practices to recruit a diverse student body and faculty and to foster an inclusive community in the department.

As chair, I am focused on supporting a modern statistics department that creates and applies groundbreaking statistical methods, bridging theory and practice. But the true success of our department derives from the people in it. Last year we had three excellent faculty join us: Morgane Austern, an Assistant Professor, Sham Kakade, the Gordon McKay Professor of Computer Science and Statistics, and Kelly McConville, Senior Lecturer and Co-Director of Undergraduate Studies. We also have a new cohort of postdocs and PhD students (whom you can read about in this newsletter) and new administrative staff. In summer 2023, we look forward to having Mark Sellke join us as an Assistant Professor and Anna Seigal as an Affiliate Faculty.

This newsletter highlights some key moments from the year, including our graduation events to celebrate with our graduating concentrators, master's, and PhD students from 2020-2022 in person (finally!). In several interviews with alumni, students, and faculty, you can discover the exciting projects that they are working on related to research, teaching, or a career in industry. In a section on "Awards and Appointments," we highlight some of the accomplishments of our faculty, postdocs, and students. To keep you informed about our concentrators and PhD students, we give you an update from the directors of undergraduate and graduate study.

My hope is that the stories in this newsletter will inspire you to be an active participant in our thriving, intellectually stimulating community this year. We encourage you to attend our events and seminars and to join our online community on Facebook and Twitter. In addition, alumni and students can extend their professional network online by joining our LinkedIn group. If you want to reach out to us or share a story for our next newsletter, please email us (statistics@fas.harvard.edu) – we'd love to hear from you!

I am ready for this journey, ready to see the novel, exciting directions you will take in your research, courses, and careers this year.

Samuel Kou, **Professor**

Chair of the Department

Undergraduate Program Highlights

Dr. Kelly McConville & Dr. Joe Blitzstein,

Co-Directors of Undergraduate Studies

Dear Statistics Community,

As the Co-Directors of Undergraduate Study, we would like to share with you some highlights about our undergraduate program this year, including the accomplishments of our graduates and the innovations of our faculty in curriculum, program design, and advising. With approximately 170 concentrators in any given year, Statistics is currently the 2nd largest concentration in the Science Division. Although we are impressed by these numbers, we are not surprised because of the growing importance of statistics in all fields.

Despite all the challenges during the pandemic, including attending many classes on Zoom, our students have consistently impressed us with their perseverance and intellectual curiosity. Setting a record, we had 40 graduating concentrators who wrote theses in spring 2022. It is likely that the thesis option has become more attractive after faculty (and undergraduate advisor) Alex Young transformed the tutorial STAT 99r into a course that nurtures a community of thesis writers who can brainstorm (and commiserate) together, present, and get feedback throughout the process.

Keeping pace with the latest developments in the field, we are thrilled to report that the faculty are launching this year a 5th concentrator track in machine learning and artificial intelligence to add to our general and other specialized (data science, quantitative finance, and bioinformatics and computational biology) tracks. This track will allow concentrators to couple their statistical analysis skills with machine learning skills, preparing them for data science careers in different industries. To support this track, faculty will teach courses such as STAT 185 Unsupervised Learning, STAT 195 Supervised Learning, and STAT 184 Reinforcement Learning, a new course designed by faculty Lucas Janson and Sham Kakade.

We've also revamped our STAT 100 and STAT 111 courses in the past few years by incorporating aspects of data science.

STAT 100, now known as Introduction to Statistics and Data Science, was redesigned by Kelly McConville to have a data science "flavor;" the course develops students' data acumen and emphasizes topics such as data ethics and reproducible workflows. Simultaneously, the course covers classical topics for an introductory statistics course, such as statistical inference and linear models. STAT 111 Introduction to Statistical Inference was completely reimagined over the last few years by faculty Neil Shephard, Susan Murphy, and Joe Blitzstein to emphasize the combined approach of statistical theory, simulation, and data. Rather than just calculating math, the course shows the power of simulation and the applicability of the theoretical concepts to case studies with real data.

Moving forward, the faculty will continue to develop a curriculum that is responsive to the latest advancements in statistics and related fields in order to prepare students for their future careers. This summer Lucas Janson and Kelly McConville supported a cohort of undergraduate researchers; Kelly's researchers worked on projects for the U.S. Forest Service's Forest Inventory and Analysis Program. We are in the process of finding strategies to connect our concentrators with more research opportunities both on and off campus in the future.

When we returned to in person events and classes in 2021-2022, we felt an irrepressible energy from students, faculty, and advisors. We enjoyed seeing students at the game night that faculty Alex Young and Kevin Rader planned, at our daily coffee hours in the spring, and at our commencement celebrations for both 2022 and 2020/2021 grads. We also appreciated the community building and academic mentoring provided by the student-run Group for Undergraduates in Statistics at Harvard (GUSH), including running panel discussions, socials, and a peer mentoring program. Building a successful undergraduate program is truly a joint effort, so thank you students, faculty, and alums!

During this academic year, we are excited to launch an advisee lunch program in which all concentrators have the opportunity to go out to lunch or coffee with their advisor and other advisees. Our hope is that this initiative will help students to strengthen ties to their advisor and other concentrators, building their professional network.

We look forward to working with all of you this year and to staying connected with our alums.

Joe Blitzstein, **Professor of Practice**

Kelly McConville, **Senior Lecturer**

Co-Directors of Undergraduate Studies

PhD Program Highlights

Dr. Susan Murphy & Dr. Jun Liu,

Joint Directors of Graduate Studies

Dear Statistics Community,

As the Joint Directors of Graduate Studies, we would like to share with you some of the highlights from students' research and teaching and from our curriculum design. From the STAT 300 Research in Statistics seminar to the graduation celebrations to the department's coffee hours, the faculty are thrilled to be engaging with our students in person again. It's these spontaneous meetings – walking down the stairwell, grabbing lunch after STAT 300, or writing out an algorithm on the walls of 316 – that allow ideas to germinate and our community to thrive. The past few years during the pandemic have shown that our grad students are up to any challenge, and this fall, our 37 PhD students are charging full steam ahead, publishing papers, winning awards, and eagerly contributing to the intellectual life of the department.

We would like to celebrate the accomplishments of our PhD students and to welcome the incoming PhD students. This year, we have eight PhD graduates: Niloy Biswas, Ambarish Chattopadhyay, Kristen Hunter, Phyllis Ju, Han Yan, Azeem Zaman, Sanqian Zhang, and Lu Zhang. Having joined our community this August, our 11 G1 scholars represent a diverse cohort from around the country and world with academic interests that range from statistical physics to differential privacy to Bayesian modeling. They are also excellent cooks, pianists, and one is even a stellar cocktail maker!

When reflecting on the past year within the graduate program, we are most proud of the excellent research and teaching conducted by our students. Examples of influential research projects that PhD students have recently worked on include developing statistical methods to forecast infectious diseases from online search data, delineating causality from association, building a software package for evaluating redistricting plans in 50 states, leveraging machine learning algorithms for model free inference, and employing a randomization-

based inference framework to study peer effects.

Our PhD students excel not only at conducting research but also at communicating about their research. A pillar of the program, the STAT 300 course is instrumental for building PhD students' communication skills for their future success as researchers or educators in academia or in careers in industry. In this weekly seminar, students present their research to their peers and faculty and receive feedback. In addition, our STAT 303 The Art of Teaching Statistics course, which is co-taught by a senior faculty member and a new tenure-track faculty member, helps our students to acquire the necessary teaching and communication skills to instruct introductory students in complex statistical ideas. Our students routinely win teaching awards; this course deserves major credit for helping our PhD students feel supported and well-prepared to become good teaching fellows.

To further our PhD students' professional growth and prepare them for potential careers in industry, we offer an exciting, relatively new course STAT 305R Statistical Consulting, designed and taught by Lucas Janson. Each semester 5-10 Statistics PhD students enroll in this course in which they provide free (both in terms of money and co-authorship) statistical advice to approximately 50 researchers from all over campus and meet weekly to discuss consulting strategies.

We look forward to discovering what our PhD students will accomplish this year and seeing our graduating students launch the next stage in their careers. As we look to spring 2023, we are excited by the prospect of building a new class for fall 2023 (we can't believe that we are already thinking about 2023!).

Jun Liu, **Professor**

Susan Murphy, **Professor**

Joint Directors of Graduate Studies

Masters Program Highlights

Dr. Mark Glickman

Director of Masters Studies

Dear Statistics Community,

As the Director of Masters Studies, I would like to share some of the exciting developments of our program this year. This is a unique program that allows both graduate students in other programs and strong undergraduate concentrators (both in statistics and other areas) to complete a Master of Arts degree in Statistics by taking additional courses. Undergraduates must complete the program during their four years of college. Over time the program has gained popularity; the program admitted 16 students into our AB/AM cohort by spring 2022, compared to an average of 9 students enrolling per year from 2017 to 2021.

Undergraduate students who apply for the program are talented students who can articulate their program goals and set a course plan. About half of the students are concentrators in Statistics, and the other half come from related disciplines, often computer science, math, and applied math; the range of disciplines reflects the interdisciplinary nature of statistics. Our graduating students from this year went on to positions such as data scientists and quantitative researchers in industry. Other students from our program typically go on to PhD programs or pursue jobs in the financial sector, software development, or government agencies.

This fall, I look forward to welcoming the new batch of Masters students to the program and helping them to achieve their academic and career goals.

Mark Glickman, **Senior Lecturer**

Director of Masters Studies

Profile of Yash Nair

Undergraduate Thesis Prize Winner

"Every weekend I go down to the river and try to learn the first trick that everyone does. When you see it, it doesn't look impressive but it's very scary. I need all the knee pads and elbow pads before I am comfortable doing it."

Stats concentrator alum Yash Nair is not afraid to tackle new challenges, whether it's landing that first skateboarding trick or officially becoming a stats concentrator only this year (it helped that he took many relevant classes along the way!). Yash capped off his undergraduate experience by receiving the Undergraduate Thesis Prize at our Statistics Commencement Celebration on May 26th for his thesis on "Randomization Tests for Adaptively Collected Data," supervised by Professor Lucas Janson. In a recent Zoom conversation with Yash, we talked about his award-winning thesis, program highlights, and tips for incoming concentrators.

Tell us about your thesis project: How did you select your project and what questions were you asking?

Nair: Sure, I can give you some background on my project and the motivation behind it. In clinical trials, to figure out if a treatment is effective, the gold standard is the randomized control trial, but it's not always ethical to assign a treatment based on essentially flipping a coin. In these trials, you often learn new information over time, like patient A has certain characteristics that make them respond better to a treatment than patient B. Instead of assigning the treatment randomly (to either patient A or B), you would assign the treatment adaptively to patient A. But, when you assign treatments adaptively, it's not clear whether the positive result is because of the treatment itself or because it's a treatment based on previous data. When decisions are built on previous data, there are a lot of complicated dependencies that make it harder to know what influences the outcome of the patient. My project with Lucas was to figure out in a regime where you are assigning treatments adaptively, how can you still infer which treatments are working or if their impact is changing over time?

My thesis was more focused on methodology than any particular application but finding the answer to this question might have some useful applications. For example, a mobile health application that learns that a patient's response to a treatment is changing over

time could provide the patient with a useful intervention.

What were some of the highlights as a concentrator in the program? Are there certain people, courses or parts of the program that really shaped your experience?

Nair: My turning point was meeting Lucas; he's the main reason why I've decided to continue studying statistics in grad school [Yash started a PhD program in Statistics at Stanford this fall]. Originally, I had thought of myself as more of a CS person, but after meeting Lucas, I was open to taking more stats courses, and I ultimately became a stats concentrator. He was incredibly supportive as an advisor and was very patient with the mistakes that I inevitably made during the project."

Also, the quality of teaching in the Stats Department is unique; every time I take a stats course, I feel lucky because the professors and lecturers are absolutely incredible. Some courses that had the biggest impact on me and my ability to do research were the intro grad stats series: 210 [Probability I], 211 [Statistical Inference I], 212 [Probability II], and 213 [Statistical Inference II], taught by Professors Joe Blitzstein, Lucas Janson, Morgane Austern, and Pragya Sur. The faculty were very supportive - willing to meet with students and help them outside of class - and were just generally very excited about research. In some of these classes, we would talk about how these techniques can be used in disciplines ranging from genomics to clinical trials to the sciences at large, showing how the methods we learn are used to address real-world problems."

Another highlight was how collaborative the classes were. During "pencil problems," the prof. would stop the class and allow students to talk through a problem together as a group. This would A) help students with problem-solving, and B) help students to get to know each other. When I TF'd in 110, I found that students were very collaborative and were interested in content, not obsessed with grades. Stats seems to draw both undergrads and grads from many different departments, which also contributes to the community and brings in other perspectives.

What advice would you give to future concentrators?

Nair: The biggest thing for me was to be open to trying and learning new things (I only became a stats concentrator this year). My path to stats was a long and winding one. I would say, don't be afraid to try new things; don't be afraid to change paths, especially if you're not sure what you're really interested in.

Stats: We think that Yash has certainly followed his own advice! The Statistics Department wishes him well on his adventures (both research and skateboarding-related) at Stanford.

Profile of Ambarish Chattopadhyay

PhD Student Dempster Prize Winner

"I'm from a place called, it's a very fancy name, Diamond Harbour, in India. There are different theories about the origin of the name. I don't know which one is correct (maybe all of them are wrong!), but one story is that, as the European traders approached the harbor, they saw the shore glittered like diamonds in the rays of the blazing sun and named the place Diamond Harbour."

Spring 2022 PhD alum Ambarish Chattopadhyay offers an inspirational image of his hometown. When he arrived at the Harvard Statistics Department in 2017, he found new inspiration through his learning about research, teaching, and collaboration. Having recently received the Dempster Prize (named in honor of Emeritus Professor Arthur P. Dempster) at the May 26th Commencement Celebration for his paper "On the implied weights of linear regression for causal inference" with advisor José Zubizarreta, Ambarish spoke with us about what motivated his dissertation work and about other "gems" of his experience in the Statistics Department.

Tell us about your dissertation project: How did you select your topic and what questions were you asking?

Chattopadhyay: For my PhD, I wanted to work on something that bridged the gap between theory and practice. Very broadly, I was interested in working on causal inference because of how fundamental it is (and relevant to practical applications). People make causal statements all the time: for example, my headache or my blood pressure was reduced because I took this medicine. In the department, there has been significant work in the field of causal inference, which is one of the reasons why I wanted to join the program. I also like how causal inference is very interdisciplinary; many disciplines (e.g. in the medical field and social sciences) are exploring causal relationships. From a statistical point of view, there were many interesting connections to statistical topics that I was familiar with, like sample surveys and missing data.

Broadly speaking, in my dissertation, I explored how to efficiently learn about causal effects. There are two main ways to study causal effects: by conducting an experiment or by observing groups. For example, if you wanted to learn about the effect of medicine on reducing blood pressure, you could conduct an experiment with control and treatment groups, or you could observe groups of people who did or didn't take the

medicine. In both types of studies, I wanted to know: how do I infer about the causal effects? A fundamental concept that drove the work was the idea of balance – the idea of creating an apples-to-apples comparison. For example, if I wanted to learn the effect of medicine on Person A's blood pressure, ideally, I would have a clone of person A so that one person could be in the treatment group and one person in the control group. If the responses of Person A and the clone to the medicine are different, then there is a causal effect. Another key idea was that of representativeness – the groups should not only be similar to each other, but also similar to the population that you want to learn about. Governed by these principles of balance and representativeness, I wanted to develop and analyze methodologies to set up similar groups relative to the population of interest and then compare the results.

What were some of the highlights as a PhD student in the program? Are there certain people, courses or parts of the program that really shaped your experience?

Chattopadhyay: The best part was meeting with so many brilliant people working in several different fields in the department and learning the importance of collaboration was eye-opening. I learned so much from conversations with my committee members, Carl Morris, Kosuke Imai, and of course, my advisor José Zubizarreta; they were instrumental in shaping my research experience. When I joined here, I had a very vague idea about causal inference, but José introduced me to the formal research, helping me to understand what is of essence in a research problem. Most importantly, José taught me about how to communicate through writing, redirecting me from writing papers that were similar to lecture notes.

An important aspect of my experience in the department was a retreat where faculty give lightning research talks about what they are interested in and what they are doing. From watching José's presentation, I knew I was interested in working with him on causal inference. It's the same experience with many other students; the retreat, along with other departmental seminars (such as STAT 300), is an opportunity to learn about research in the department, and it is an environment that encourages you to speak to faculty members. This is a place where everyone – students and faculty – come together.

I must also mention the importance of teaching during my time at Harvard and for my future career. I was highly impressed with how the Harvard Statistics Department emphasizes teaching and, more generally, emphasizes communicating statistics. Teaching is one of the best ways to communicate ideas. Explaining concepts to a class also helps me to gain understanding; whenever I teach, I learn something new. In my first year, I took the required course STAT 303 The Art and Practice of Teaching Statistics. For the course, we had to record several teaching presentations of

different durations (5 minutes, 15 minutes, an hour) in front of faculty members and students, which we then reviewed and discussed with the teaching fellows to determine what went right or wrong in the class. Hearing objective opinions about my teaching was very helpful.

What are your goals moving forward?

Chattopadhyay: My goal with taking a Postdoc position at Stanford in data science is to continue to work on causal inference problems in a collaborative way. Causal inference is not just the property of statisticians; other fields like computer science and economics have made significant contributions in answering causal questions. Because data science is interdisciplinary, this Postdoc position will be a great way to collaborate on developing and applying methodologies to different problems across disciplines.

What advice would you give to future or 1st year PhD students?

Chattopadhyay: Honestly, I don't think I'm in a position to give advice! Even though it's cliché, it's true that you should do what you love and love what you do – having fun (both inside and outside of research) is important! Having a few close friends (you don't need 20 or 50!) to talk to will also make a big difference in your experience.

I've learned that communicating well about your research is just as important as the research itself. While you are here over the next 5 or 6 years, I would recommend taking opportunities, like presenting in the STAT 300 Seminar series, to improve your communication. As I mentioned before, when I write things down, it requires a lot of thinking and makes me understand my work better.

Lastly, grad school is hard; sometimes progress can be stalled for days or weeks, but don't get discouraged, and seek help when you need it. In the past I've been a lone wolf, but whenever I did ask for help, it worked. There are also a lot of resources in the department and at the university, so if you are experiencing a problem, like a mental health issue, please reach out for help.

Stats: Despite his protest, we think that these are some sage words of advice from Ambarish and have served him well during his time at Harvard. We wish Ambarish well during his next stage in his career at Stanford – please stay in touch, Ambarish!

Dr. Kelly McConville

Reimagines the STAT 100 Course

"The exciting challenge of revamping STAT 100, which is our most general audience intro stats class, was an opportunity that I couldn't say no to," reflects Dr. Kelly McConville, a Senior Lecturer in the Harvard Statistics Department, on her decision to join the department. A survey statistician with interests in machine learning and statistics and data science education, McConville arrived in January 2022 from Reed College, where she was an Associate Professor of Statistics. We reached out to Kelly to learn more about her reimagined course and her new experiences at Harvard to share with the community. During a lively conversation, Dr. McConville highlighted her passion for course design, teaching, and mentoring undergraduate researchers.

What brought you to the Harvard Statistics Department and what are your favorite parts about the department so far?

McConville: This is a fun question to think about! For the past 10 years, I taught in math departments at small liberal arts colleges that would have somewhere between one and four statisticians. The idea of joining a statistics department, specifically, where I would be surrounded by statisticians and could attend many weekly seminars, was enticing, especially because Harvard has one of the best statistics departments in the country.

An aspect that I value – it's hard to put it into words – is this aura within the department. If you want to do something, whether it's teaching a new class, hosting an event, forging a new collaboration, or starting a book club, the department will do whatever they can to support you and make it happen – I love that! Different perspectives are valued and encouraged here as part of the department's efforts to continue learning and improving; as a new faculty member, I really appreciate that kind of support.

Additionally, I value that the department authentically cares about both research and teaching, especially as someone who is on the teaching end of the spectrum.

Stats: One of the main ways that McConville advanced her pedagogical ideas this spring was by redesigning STAT 100 Introduction to Statistics and Data Science (formerly known as "Introduction to Quantitative Methods for Social Sciences and Humanities"). In her conversation with us, she highlighted how the course is designed for all students to learn how to effectively analyze data,

whether they think they are a "math/stats person" or not.

Why is STAT 100 an important course for students to take?

McConville: I bet that all students will end up using data in their jobs or lives to make decisions, but because data is often messy and incomplete, it's tough to know how to use it to make decisions. STAT 100 is about helping students to learn how to extract knowledge from messy data; along the way, they acquire good data habits, and they develop their statistical thinking skills.

Because a lot of the students come in with mathematical anxiety, one of my less advertised goals with the course is to dispel the notion that this class isn't for everyone. I believe in the growth mindset and want students to finish the class thinking that they're a statistical rockstar (or at least to see how statistics is worthwhile in their daily lives!).

What changes did you make to the STAT 100 curriculum?

McConville: The original version of the course had a title that identified the target population of students, but we changed this because I wanted this class to be for any student, regardless of their concentration. As you would guess from the title, we've also added more data science topics into the course like "data wrangling."

Another important change includes taking a more computational route (as opposed to a more mathematical approach) to understand foundational statistical ideas. A benefit of using a computational approach is that the barrier to entry is a little bit lower; I assume that students have had zero programming experience before the class. From week one, students use code (I teach Tidyverse Packages in R) to help them understand a new statistical concept. With these packages, there's not as steep a learning curve, which helps students to develop their computational skills faster. We generate, graph, and visualize distributions and look at what happens when we tweak different pieces of the code. We use computation to simultaneously learn the concepts and how to do statistical analysis.

Was there a specific part of the STAT 100 course experience that stood out?

McConville: For the first time, I was teaching a class of over 130, so I was worried that I'd struggle to connect with my students and to build a supportive learning community. However, my STAT 100 TFs/CAs and students did an amazing job of helping me build a fun, engaging environment and made me feel so welcome at Harvard. My favorite moments were all the little conversations that I had with students about the books they were reading, their career aspirations, the ways they were using stats, or even about a lackluster menu in the dining hall. At the end of my last lecture, this was also the first time that I had a line of students

who wanted to take their picture with me – that was pretty fun!

What kinds of exciting projects were you working on during the summer?

McConville: To help me with forestry data science projects, I employed a group of undergraduate researchers. These projects were directly motivated by researchers at the U.S. Forest Service Forest Inventory and Analysis Program, the U.S.'s survey agency for our nation's trees. This awesome group of five students worked on different problems in pairs; they ran simulation studies to compare estimators, explored ways to measure and report uncertainty, and created interactive dashboards for disseminating the work. One of our projects centered on producing estimates and uncertainty estimates of above ground carbon, which is important data to know when considering climate change. Through these projects, I emphasized how to share estimates in a way that is digestible to a non-statistician (to the people who own or manage forests). We tried to think through questions like, "how do you give people an intuitive understanding of uncertainty?" The projects focused on statistical communication for a broad audience.

What is something new that you would like to achieve in the next year?

McConville: I'm excited about designing a new statistical computing course for the spring. This course, STAT 108: Introduction to Statistical Computing with R, will have no coding prerequisites. Students will progress from learning how to use someone else's code to imagining and creating their own code, all with an eye towards best practices for statistical analysis.

The other project that I'm looking forward to is related to undergraduate research. As you know, I'm a big fan of undergraduate research for many reasons: students learn statistics by doing statistics, they develop communication skills, they strengthen their connection to the field, and they benefit from greater career clarity. I feel energized by a proposal (from former Chair Neil Shephard) to serve as the department's undergraduate research catalyst to help students learn about statistics related undergraduate research experiences on and off campus.

Stats: As we move further into the school year, we anticipate sharing with the community more about Dr. McConville's new computing course and her initiative to connect more undergraduates with research opportunities.

Profs. Lucas Janson & Sham Kakade

Create a New Reinforcement Learning Course

Prof. Lucas Janson

Prof. Sham Kakade

It's the start of the fall semester and time to map out a course schedule, but what will work best for you? You love math, so maybe you should stock up on quantitative courses; on the other hand, it might be a good idea to get some course requirements out of the way. You're a night owl, but maybe you should take those morning classes so that you can play badminton later in the day. While the new course CS/STAT 184 "Introduction to Reinforcement Learning" may not solve these scheduling dilemmas, the course will teach undergrads how to think through these types of problems. Through the study of reinforcement learning, a sub-field of machine learning, students will learn how to state a problem with different actions and rewards, implement algorithms to solve these problems, and identify why the algorithms work.

To get a preview of CS/STAT 184 Reinforcement Learning, we spoke with Dr. Lucas Janson, Assistant Professor of Statistics and Affiliate in Computer Science, and Dr. Sham Kakade, Professor of Computer Science and Statistics, about their inspiration for the course and their goals for students. Janson arrived at Harvard Statistics five years ago after his PhD and has focused mainly on research in high dimensional inference, uncertainty quantification, and reinforcement learning. Joining Harvard in fall 2021, Kakade is the Co-Director of the Kempner Institute for the Study of Artificial Intelligence and is broadly interested in the areas of machine learning and AI. More specifically, Kakade is curious about how to improve methods "to push the frontier for computers to complete the kinds of tasks that humans excel at." Kakade's and Janson's responses below are edited excerpts from a conversation with them on these topics.

What led you to pair up to design a new course on RL for this year?

Janson: To put this course in context, I have been trying to incorporate more machine learning courses for the past several years

into the undergrad curriculum in the Statistics Department. Machine learning is often categorized into three sub-categories: unsupervised learning [uses unlabeled data], supervised learning [uses labeled data], and reinforcement learning (RL). In both unsupervised and supervised learning, models are used to provide you with information, but you must act based on the information. In contrast, in RL, the model incorporates acting into the feedback loop, so it's the only branch of machine learning that affects the real world directly. A few years ago, I created STAT 195 Statistical Machine Learning [now Introduction to Supervised Learning], and my colleague Dr. Alex Young simultaneously started STAT 185 Introduction to Unsupervised Learning, so it seemed to be a natural progression to introduce a reinforcement learning course.

The idea for this joint venture came about during one of Sham's visits to Harvard (before he joined the Department). We were walking around Harvard Square and realized that we both intended to independently teach a course like this one.

Kakade: After chatting with Lucas, I thought we should clearly do this together both because the collaboration would help us to reduce the prep work and to source the best material – and it's simply more fun! My own path to this course started after I taught a graduate class on the theoretical foundations of RL at Cornell Tech in New York, and I thought it would be a good idea to bring an undergraduate version of the course to Harvard [Professor Susan Murphy teaches a graduate RL course at Harvard].

Why do you think this is an important course for students to take?

Kakade: In this course, we want students to understand how algorithms can solve problems both by studying the underlying mathematical formulas as well as by implementing the algorithms and analyzing the results. Students will see how the algorithms work in both abstract and hands on ways, which is a pretty good model of learning, not just for this class, but also for more advanced machine learning courses.

As Lucas mentioned previously, there are a couple of different sub-areas of machine learning – unsupervised, supervised, and RL. As we move into RL and more realistic applications, the deployment becomes interactive in nature; our computers don't just sit there, but instead, they interact with the world to learn. For any effective deployment, there is some aspect of interaction, and this course provides students with the foundation to think carefully about this process at multiple levels. In reinforcement learning, often there is not a teacher in the system telling you what to do, which to a large extent reflects the dilemmas we face in the real-world when we start deploying technological and scientific systems. Because of the increasing number of systems deployed that utilize these ideas and impact our world, students recognize the value of this style of thinking and want to learn more about RL.

Provide a brief overview of the course's content, structure, and goals.

Kakade: Broadly speaking, we're approaching this course topic and its structure by focusing on the different levels of complexity in how a system might interact with the world and the algorithms used to solve these problems. To begin with, we will explore simple problems with short-term decision making, then move to long-term decision making, and, finally, advance to long-term decision making in complex environments. There's a natural progression of problem complexity and solution complexity that will be adjusted throughout the course. By the end of the course, students will be able to implement algorithms and to analyze why they provide a correct solution.

Janson: I would just add that one advantage of starting with the simpler problems is that we can usually draw more conclusions about them. As problems advance in complexity, you may be able to determine the algorithms that solve them, but it becomes harder to draw rigorous conclusions about why these algorithms work. While the cutting edge of the field lies in trying to better comprehend the complex algorithms that solve problems in complex environments, students can start demonstrating interesting results in these simpler problems. This lays the groundwork for thinking about more challenging problems. Another goal of ours with this course was to make it accessible to more students by requiring minimal prerequisites. Exposure to probability, programming, and linear algebra is about all you need for a course focused on teaching the first principles of reinforcement learning.

Which part of the course do you think students will look forward to the most?

Janson: In terms of the big picture, I think this course will continue to generate interest with students because there is a lot of pioneering work with reinforcement learning in the real world now. Because RL problems act in the real world, even their simple versions can look intimidating to solve. However, even before approaching the math portion of RL, you can teach students to state the problem very simply and intuitively. In real life, we encounter scenarios all the time in which we can take different actions with different rewards and try to choose the option with the maximum reward. For instance, when deciding on the fastest route to work, you might select a route that has been fast in your experience, but maybe there's a better route that you haven't tried out. When determining the different rewards, you need to strike a balance between selecting an option that you already know versus a potentially better option that you need to spend time learning about. After stating the problem, then students can delve into the math and implementation of the algorithms. Students will be challenged by this course but will also quickly understand how to approach RL problems and how to determine why certain algorithms work, which will be satisfying and exciting for them.

Stats: Based on our conversation with Prof. Janson and Prof. Kakade, it sounds like students will have an enriching experience this fall in CS/STAT 184. Thank you to Prof. Janson and Prof. Kakade for sharing your insights!

Profile of Alumna Michele Zemplenyi

Fellow at the U.S. Department of Energy

After graduating from Harvard with an A.B. in Statistics in 2013, alumna Michele Zemplenyi had a plethora of career options to consider. Dr. Zemplenyi was not only a statistics concentrator but also completed premed requirements and a thesis in conjunction with the Stem Cell and Regenerative Biology Department. Choice is great, but it can also be intimidating (as our senior concentrators probably know!). Fortunately, Michele shared with us how she carved out a career path, starting with statistics in undergrad to a PhD in biostats to a current Fellow position in the Federal Energy Management Program at the U.S. Department of Energy. Her journey has been sustained by a love of math and the biological sciences, and a desire to help solve climate issues. The responses below are revised excerpts from our interview with Dr. Zemplenyi.

Please introduce your current work.

Zemplenyi: For a year now, I've been working as an AAAS [the American Association for the Advancement of Science] Science and Technology Policy Fellow within the U.S. Department of Energy. The goal of our office is to help the whole federal government reach energy- and water-related efficiency goals. To give some context, the U.S. government is the largest energy consumer in the U.S., with a footprint of approximately 600,000 vehicles and 300,000 buildings. Under this administration, our office's main objective is to decarbonize the federal government's assets.

I'm currently working on a project for the White House Council on Environmental Quality analyzing which electric utility companies (out of the hundreds that serve federal facilities) carry the greatest loads. This will help us craft a government-wide strategy to procure more carbon pollution-free electricity by first engaging with the utility companies where we can make the biggest impact on the government's carbon footprint.

Another priority in my work is to develop tools that help the government plan for and install electric vehicle charging infrastructure. The optimal charging infrastructure for a fleet of vehicles depends on many factors, including the size of the vehicles, daily mileage traveled, parking lot configuration, and the capacity of existing electric service panels and transformers. One of the teams at the National Renewable Energy Laboratory that I work with is developing a software tool to help facilities weigh the tradeoffs between different charging infrastructure designs

and estimate the associated costs. My role is to think carefully about the data required for these cost estimates, look for new datasets, and work with federal partners to apply our team's tools during site assessments. So far, I've been involved in assessments at several National Park Service sites, including Grand Teton National Park and Yellowstone National Park.

What aspects of your time in the Harvard Statistics Department influenced you the most?

Zemplenyi: I benefitted from some very invested professors who were not only passionate about teaching but also wanted to ensure our future success in life. Professors Joe Blitzstein, Dave Harrington, Kevin Rader, and my undergraduate thesis advisor Tirthankar Dasgupta helped shape my career path. While I was working at my first job after undergrad at a software consulting firm, I was on my way to the American Statistical Association's Joint Statistical Meeting in Boston when I ran into Dave Harrington on the silver line. It was very fortuitous; over coffee, we discussed what I liked/didn't like about my job and what I missed about school. Dave said, "Have you considered biostatistics?" I really hadn't, but a few weeks after that conversation, I started applying to PhD programs in biostatistics. Dave helped me realize that there are ways to incorporate my interests in math, biology, and science through a biostatistics degree. I'm really grateful to Dave and my professors for continuing our relationship after graduation to help guide me on my career path.

How has your career path evolved and been informed by your degree programs?

Zemplenyi: My positive experience in the Statistics Department in undergrad influenced my desire to continue on that path with a graduate degree in biostats. My PhD program, in turn, led to my interest in researching ways to reduce air pollution, which is what drew me towards working on climate and energy issues.

One of the great things about going into statistics has been that it is a very applied and flexible degree, so you can choose how to use it. While I don't have specific training in energy-related issues, I can still apply my math, programming, and critical thinking skills gained from my statistics degrees in my current role at the Department of Energy. Even though I worked on a relatively narrow topic in my PhD program, focusing on the effects of air pollution on human health, the challenge of developing new statistical methods and learning new programming languages in graduate school gave me the confidence that I could teach myself the skills necessary to work on problems in other fields.

As far as my decision to pursue a fellowship in the federal government, well, I've long been inspired by past accomplishments of the government, say during the Manhattan

Project and space race, and today, transitioning to a low-carbon economy is the next big challenge for the government to take on. It's such exciting work to be a part of and people with a statistics skillset are valuable to the cause.

What advice would you give to current undergrads in statistics who are undecided about their careers?

Zemplenyi: 10 years out from college, I'm still not "there" yet! I'm still trying to define my career goals and the skills that I want to develop. For a lot of undergrads, the first job that you have out of college is most likely not going to be the one that you stick with for your career, so don't stress about getting it perfect right away. There's time for experimentation and iteration. Before the summer of senior year, consider internships that you really enjoyed (or didn't enjoy) and issues that are interesting to you – these could be topics you find yourself reading about in your spare time. To learn about possible jobs in those fields, try to talk to people at organizations that seem interesting – you don't have to wait for well-crafted experiences like the big recruitment career fairs. Instead, you can try reaching out to someone on LinkedIn, Twitter, or a Slack community to learn about their path. If you're interested in a lot of different areas, like I am, it's helpful to start cutting off the pathways simply by experiencing it all (both what you like and don't like) and talking to other people about their careers. In time, you'll get closer and closer to finding the right path for yourself.

Looking back on your Harvard experience, what do you miss the most?

Zemplenyi: I miss the fun of living in the houses with my closest friends and the vibrancy of campus, including all the concerts, dance performances, and incredible artists and speakers who would visit. In the average week, I could hear Yo-Yo Ma perform or listen to Andre Agassi discuss his favorite moments on the tennis court! When will I get opportunities like that again?

Stats: We appreciate Dr. Zemplenyi taking the time to speak with us about her career path and offer insights into the options available to our students. Through Michele's work on the government's plans to decarbonize in the coming decades, we see how our alums can leverage statistical thinking to have a meaningful impact on our world (and have fun in the process!).

Cory McCartan & Professor Imai

Create an Algorithm for Evaluating Redistricting Plans

Cory McCartan

Prof. Kosuke Imai

With the mid-term elections coming up this fall, debate has been brewing in states over drawing new congressional district maps, which occurs every 10 years after the completion of the census (also known as "redistricting"). In multiple court cases, voters have challenged the new maps with claims of gerrymandering (intentionally favoring one political party over another). In Alabama, voters filed a lawsuit contending that the new plan failed to provide enough majority-minority districts, and thus, hurt black voters. Voters in New York filed a lawsuit claiming Democratic partisan gerrymandering while voters in Ohio filed a lawsuit claiming Republican partisan gerrymandering. With the stakes so high, where could the plaintiffs, defendants, and courts turn to for supporting evidence of bias (or a lack thereof) in a new redistricting plan?

50-State Redistricting Simulation Project

Enter the ALARM (Algorithm-Assisted Redistricting Methodology) Project's R software package "redist" used to simulate thousands of sample districting plans for all 50 states as part of the 50-State Redistricting Simulation Project. The ALARM Project is a research group, including graduate and undergraduate students, and one high school student, led by Professor of Statistics and Government Kosuke Imai. The research group focuses on "redistricting sampling algorithms, best practices and workflows for redistricting analysis, and tools to visualize, explore, and understand redistricting plans," and receives computational support from the Harvard Data Science Initiative and from Microsoft (ALARM Project website). To further understand the impact of the ALARM group's 50-State Project, the Statistics Department spoke with Professor Kosuke Imai and 4th year Statistics PhD student Cory McCartan, Kosuke's advisee, who helped develop the simulation algorithms, design the R software package, and introduce the idea of expanding the project to all 50 states and more. The contents of this article are based on conversations with Kosuke Imai and Cory McCartan.

Software Addresses Redistricting Problem

According to Kosuke, historically, researchers would consult older districting

plans or plans from other states to measure the fairness of the proposed or enacted plan. However, these methods were not very useful for rating a plan both because demographics and political views often change over time and because states can vary substantially in terms of their demographics, politics, and redistricting rules. Instead, researchers needed a method to obtain a random sample of redistricting plans that followed state rules and could be used as a neutral baseline to measure the fairness of an enacted or proposed redistricting plan.

"To address this problem," explains Kosuke, "Cory and I came up with the new simulation algorithm, called the Sequential Monte Carlo (SMC) algorithm, and implemented a software package to make these cutting-edge algorithms accessible to a broad audience - social scientists, policy makers, journalists, etc." Expanding on Cory's role in the project, Kosuke says, "Cory is a unique scholar because he brings together both the mathematical and statistical expertise and a passion for politics and the social sciences. This kind of applied work requires an immense amount of attention to detail while innovating new methods that are actually able to solve real-world problems. For example, it was important for the algorithm to be able to incorporate criteria such as not splitting county boundaries or not having two incumbents face off. These details have a significant impact on redistricting and have real-world political implications, even if they might not be math problems that can be easily solved. Cory understood this and cared about getting this right."

In recent years, researchers have developed other algorithmic tools to draw proposed districts, but these tools didn't have a way of measuring bias in the plans. With the SMC algorithm in the software package, "now, you have a mathematical algorithm, something you can write out on a whiteboard and talk about and study, which helps us to be more confident that the method is fair and doesn't have bias built in," explains Cory. The mathematical algorithm that Cory refers to generates a random sample of plans that incorporate common state redistricting rules, such as that the plans have equal population across districts, follow existing county boundaries, be contiguous, and be geographically compact.

To help us visualize the algorithm, Cory describes how the formula starts with a blank map, draws one district that it keeps and then, out of the remaining area in the state, continues to draw a 2nd district, 3rd district, and so on. This process is completed many times over in parallel, each time starting over with a blank map – "similar to how an etch a sketch works," explains Cory. The method that the SMC algorithm follows is an improvement over some other algorithms because it results in a high level of efficiency. Cory elaborates, "When we talk about the efficiency of the algorithm, it's about how can I get the most information and explore more possibilities within a given amount of time. While it's faster to just change a single border than it is to draw a whole plan from scratch, when we start from scratch each time, we get a whole new map that has a lot more information."

Communicating the Data

While the SMC algorithm excels in efficiently producing a high-quality sample

of redistricting plans, this would not be sufficient to make a difference in court cases, in policy, in the real-world in general. Cory highlights the importance of providing tools that clearly communicate the data: "If you simply make 10,000 plans, that will just show you every single precinct in the state and which district it's assigned to - for a human that's completely useless. We need to think about how to pull information out and visualize the data in one or two helpful charts. There's a technical stats side to this project but also a communication side." Specifically, the software package helps users communicate data by including graph plotting functions in the code, including graphs that show the geographic compactness and Republican/Democratic vote share by district.

Kosuke shares his thoughts about Cory's contributions to the project in the area of communication, "Cory is an incredible mathematician and statistician; while he can understand the math and write out the algorithm, he can also make it usable and understandable to others." The software package is readily accessible to undergraduate students (and even high school students!), who have used the algorithm for ALARM's 50 states project to simulate sample plans for all 50 states. Because the software is open source, it is available to a wide range of researchers, policy makers, journalists, and anyone else interested in the redistricting process.

Real-World Impact of 50-State Project

Now that the R software package is accessible to a wide range of stakeholders, it's easier to hold states accountable for drawing fair, unpartisan voting districts. The package has been used by many researchers and journalists. It has also played a significant role in several recent court cases, including Alabama, New York, and Ohio. Recently, United States Supreme Court justices discussed the simulation work presented in Kosuke's expert witness report during the oral argument about the important racial gerrymandering case in Alabama. Cory elaborates on the significance of the project to him: "It's been great to see how the simulation evidence, either from our team or other teams, has been critical in these court cases to prove that districts are fair or not fair. This has a big impact on who gets elected and whether voters have a fair chance of electing their preferred candidate. I've been in grad school for a little more than 3 years, and it's gone from a very small project to a large team where we get to work on these important cases – it's been a very cool project to be a part of!"

Regarding the next steps for the project, Cory expresses his enthusiasm for continuing to learn more: "As the dust settles now, it's going to be helpful to do a post-mortem to see what things worked (or didn't). We are currently working on an analytical framework for looking at the impact of different types of gerrymandering beyond partisan gerrymandering – like racial gerrymandering and gerrymandering at the local level. We now have a whole other cycle on our hands and that's going to be very exciting." We look forward to seeing what Cory, Kosuke and the rest of the ALARM Project do next and how their work will continue to impact the redistricting process.

Welcoming G1 PhD Students

Kyla Chasalow

Previous Institution:
Cornell University (undergrad)

Research Interests:
statistical methods for the social sciences, causality, privacy, algorithmic fairness, history of statistics, history of science

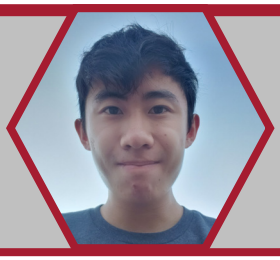
Hobbies:
painting, swing dance, everything equestrian

Nathan Cheng

Previous Institution:
University of California, Berkeley (undergrad)

Research Interests:
causal inference, high-dimensional statistics

Hobbies:
Doing crosswords, baking



Alan Chung

Previous Institution: Princeton University (undergrad)

Research Interests: probability theory, theoretical ML. I'm looking forward to learning about random matrix theory, mathematical statistics, statistical physics.

Hobbies: tennis, rock climbing, jazz piano, cooking, poker, video games (e.g. Age of Empires II and Valorant). If anyone else plays these games, please let me know :D.

Nowell Closser

Previous Institution/Past Work:
Rensselaer Polytechnic Institute (undergrad). Worked as data engineer and at a variety of tech startups.

Research Interests: MCMC, causal inference, statistical machine learning

Hobbies: write and play music for the piano, outdoor activities (ski, bike, climb, etc), and make the occasional high-quality cocktail

Johann Gaebler

Previous Institution: Stanford University (PhD program in Computational and Mathematical Engineering)

Research Interests: the intersection of statistics, CS, and policy, including causal inference, computational social science, Bayesian statistics, ML, algorithmic fairness, and statistical computing

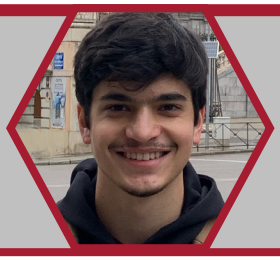
Hobbies: cooking, learning new languages, playing the guitar (not very well), running

Yuzhou Lin

Previous Institution:
University of Pennsylvania (undergrad)

Research Interests:
causal inference, Bayesian modeling, design of experiments and observational studies, applications of statistics in public health and public policy

Hobbies:
clarinet, go, poker, ping-pong, badminton, hiking



Matthew Esmaili Mallory

Previous Institution: UCLA (undergrad)

Research Interests: broadly interested in the intersection of mathematics and statistics, such as probability theory, algebraic, and asymptotic statistics

Hobbies: cooking (I say that I will become a chef if statistics doesn't work out!), piano and guitar, listening to music (The Beatles & Pink Floyd), film, photography, videography, chess, and many more :^)

Philip O'Sullivan

Previous Institution:
University of Chicago (undergrad)

Research Interests:
applications of statistics in law, causal inference, text processing

Hobbies:
cooking, video games, programming, a small but growing interest in gardening

Benjamin Schiffer

Previous Institution:
Princeton University (undergrad)

Research Interests:
probability theory, information theory, theoretical machine learning

Hobbies:
tennis, ping-pong, reading

Souhardya Sengupta

Previous Institution:
Indian Statistical Institute, Kolkata (undergrad)

Research Interests:
methodological and theoretical problems in statistical machine learning, high-dimensional inference and non-parametric statistics

Xiaodong Yang

Previous Institution:
University of Science and Technology of China (undergrad)

Research Interests:
MCMC, stochastic optimization, and differential privacy. I'm looking forward to exploring more during graduate studies!

Awards, Appointments & Honors

NEW FACULTY

Morgane Austern, Assistant Professor: Dr. Austern graduated with a PhD in statistics from Columbia University and was a Postdoctoral Researcher at Microsoft Research New England. Her research interests include applied probability, learning on structured data, machine learning theory, high dimensional statistics, and information and ergodic theory.

Sham Kakade, Gordon McKay Professor of Computer Science and Statistics: Dr. Kakade completed his PhD in computational neuroscience at the Gatsby Unit at University College London. Most recently, he was Professor of Computer Science and Statistics at the University of Washington and a Senior Principal Researcher at Microsoft Research. His research interests include machine learning and AI theory, reinforcement learning, deep learning, natural language processing, and robotics.

Kelly McConville, Senior Lecturer and Co-Director of Undergraduate Studies: Dr. McConville completed her PhD in statistics at Colorado State University, Fort Collins, and was an Associate Professor of Statistics at Reed College. Her research interests include statistics and data science education, survey statistics, statistical learning, and official statistics.

Anna Seigal, Affiliate Faculty (starting July 2023): Dr. Seigal is currently a Junior Fellow at the Society of Fellows at Harvard and received her PhD in mathematics from the University of California. In July 2023, she will begin her Assistant Professor position in Applied Math at Harvard SEAS. Her research interests include tensors and multilinear algebra, applied algebraic geometry, and algebraic statistics.

Mark Sellke, Assistant Professor (starting July 2023): Dr. Sellke completed his PhD in the Stanford Math Department and is currently a Postdoctoral member at the Institute of Advanced Studies at Princeton. His research interests include probability, high-dimensional geometry, statistics, optimization, and machine learning.

NEW POSTDOCTORAL FELLOWS (starting summer & fall 2022)

Julien Chhor has joined Dr. Subhabrata Sen's and Dr. Rajarshi Mukherjee's (Department of Biostatistics) research group. Dr. Chhor's research interests are in high-dimensional and non-parametric statistics, robust statistics, local differential privacy, and benign overfitting.

Kyra Gan has joined Dr. Susan Murphy's research group. Dr. Gan's research interests are in causal inference, optimization, and machine learning.

Yongyi Guo has joined Dr. Susan Murphy's research group. Dr. Guo's research interests are in reinforcement learning, data-driven decision making, and statistical machine learning.

Samriddha Lahiry has joined Dr. Pragya Sur's research group. Dr. Lahiry's research interests

are in high-dimensional statistics, exact asymptotics, nonparametric estimation, and quantum statistical inference.

Shuangning Li has joined Dr. Susan Murphy's research group. Dr. Li's research interests are in causal inference, multiple testing, network interference, and statistical reinforcement learning.

Ivana Malenica has joined Dr. Susan Murphy's research group. Dr. Malenica's research interests are in causal inference, machine learning, non/semiparametric inference, and reinforcement learning.

FACULTY AWARDS & APPOINTMENTS

Iavor Bojinov received the Richard Hodgson Fellowship from 2021-2025. This award is given to one tenure track faculty member per year at the Harvard Business School. According to HBS, "The Richard Hodgson Endowment Fund for Faculty Development was established in 1997 by Richard Hodgson, MBA 1939, in support of excellence in management education."

Joe Blitzstein set a record for his celebrated Stat 110 course, Introduction to Probability, which enrolled 815 students – the highest ever for a statistics course at Harvard!

Stephen Blyth was elected Principal of Lady Margaret Hall, University of Oxford. He takes up his post in October 2022. Professor Blyth will remain an Associate of the Statistics Department.

Mark Glickman was appointed Chair of the American Statistical Association's Committee on Data Science and Artificial Intelligence. In addition, Dr. Glickman was reappointed this summer as Chair of the US Chess ratings committee. He has served as a member continuously since 1985 and was previously Chair from 1992-2019.

Kosuke Imai: Clarivate Analytics awarded him as a Highly Cited Researcher (cross-field category) for ``production of multiple highly cited papers that rank in the top 1% by citations for field and year in Web of Science." In 2021, he received an Excellence in Mentoring Award from the Society for Political Methodology. Professor Imai's forthcoming paper "Experimental Evaluation of Algorithm-Assisted Human Decision-Making: Application to Pretrial Public Safety Assessment" (with Z. Jiang, D.J. Greiner, R. Halen, and S. Shin) was selected as a discussion paper by the Research Section of the Royal Statistical Society.

Sham Kakade is co-leading the "Kempner Institute for the Study of Natural and Artificial Intelligence."

Samuel Kou was appointed the new Department of Statistics Chair in July 2022.

Xihong Lin received the 2022 Marvin Zelen Leadership Award in Statistical Science. According to an article in the Gazette, the "award recognized her advances in statistical methods, particularly as applied to genetics and genomics, as well as her role as a leader and mentor in the field." At the Joint Statistical Meeting this August in D.C., Professor Xihong Lin was presented with the National Institute of Statistical Sciences' (NISS) Jerome Sacks Award for Outstanding Cross-Disciplinary Research for 2022.

Jun Liu was selected to be one of the 11 International Society for Computational Biology (ISCB) Fellows for 2022. ISCB's newsletter recognizes Jun, "as an influential leader in Bayesian statistical modeling and computation with seminal contributions in the theory and method of Markov Chain Monte Carlo and sequential Monte Carlo, which have been created and applied to scientific problems such as motif discovery and cancer genomics, and for his development of widely used tools in bioinformatics and data science."

Xiao-Li Meng resumed writing articles as part of his "XL-Files" series for the Institute of Mathematical Statistics this summer. In his latest article, "I am Not Yo-Yo Ma," he warns against jumping to conclusions, whether it's regarding people's identities or the data you're analyzing. Dr. Meng culls these lessons from being mistaken as Yo-Yo Ma while at the induction ceremony of the American Academy of Arts and Sciences this year (Dr. Meng was elected in 2020 to the AAAS, but the ceremony was postponed until this year due to COVID).

Kelly McConville was selected as a Fellow of the American Statistical Association "for excellence in the dissemination of statistical knowledge, notable contributions to the modernization of survey methods in official statistics, outstanding mentorship of undergraduates and junior researchers, and service to the profession." She was also elected to be the 2023 chair-elect of the ASA Section on Statistics and Data Science Education.

Susan Murphy was selected to give the S. Wilks Memorial Lecture for the Department of Operations Research & Financial Engineering at Princeton University in 2022. In addition, Dr. Murphy gave the Opening Plenary (virtual) at the Society for Ambulatory Assessment Conference and the Charles L. Odoroff Memorial Lecture at the Department of Biostatistics and Computational Biology, University of Rochester Medical Center.

Giovanni Parmigiani was awarded the W.J. Youden Award in Interlaboratory Testing by the American Statistical Association for his paper on "Multi-study factor analysis." This award was created in 1985 to recognize researchers who make significant contributions to the design and analysis of interlaboratory tests or to the evaluation of data from such tests.

Natesh Pillai was elected to be a Fellow of the Institute of Mathematical Statistics in spring 2021. Dr. Pillai received the award for significant contributions to Markov Chain Monte Carlo Methods, mixing times of Markov chains, and bridging the gap between applied probability, statistics, and computation. In addition, he gave the keynote lecture for the "Frontier Probability Days 2021" regional workshop in Las Vegas.

Pragya Sur was invited to speak by the U.S. National Academies' Board on Mathematical Sciences and Analytics on her works in high-dimensional statistics and statistical machine learning. This fall, Dr. Sur was selected to participate in the 2023 International Strategy Forum (ISF) by Schmidt Futures, which is an 11-month, non-residential fellowship program for rising leaders from ages 25-35 from around the world. In addition, Dr. Sur was invited to preside over the Institute of Mathematical Statistics New Researchers Group—her term started in August 2022.

José Zubizarreta was named a Fellow of the American Statistical Association (ASA) for 2022. Specifically, Dr. Zubizarreta was recognized "for outstanding contributions to the design and analysis of observational studies, particularly on matching and weighting methodologies; for excellence in teaching of causal inference; and for dedicated mentoring of young scholars" (Harvard Medical School News, May 2022).

CONCENTRATOR AWARDS

Yash Nair, supervised by Professor Lucas Jansen, received the 2022 Undergraduate Department of Statistics Prize for his superb coursework in the concentration and for an outstanding thesis.

Hoopes Prizes:

The following student winners were awarded for demonstrating excellence in their research projects.

- **Jay Chandra** was awarded for his project entitled "Classification of Continuous Natural Human Behavior Using Intracranial Field Potentials"—supervised and nominated by Professor Gabriel Kreiman.
- **Rachel Guo** was awarded for her project entitled "On the Effect of Ranger Patrols on Deterring Poaching: A Bayesian Approach for Causal Inference Using Field Tests as an Instrument"—supervised and nominated by Professor Milind Tambe.
- **Sreekar Mantena** was awarded for his project entitled "Model-Based Exploration Algorithms for Diagnostic Probe Design"—supervised and nominated by Professor Pardis Sabeti.
- **Yash Nair** was awarded for his project entitled "Randomization Tests for Adaptively Collected Data"—supervised and nominated by Professor Lucas Janson.
- **James Roney** was awarded for his project entitled "Evidence for and Applications of Physics-Based Reasoning in AlphaFold"—supervised and nominated by Dr. Sergey Ovchinnikov.

PHD STUDENT AWARDS

Niloy Biswas, a 2022 graduate, was given the 2021 Laplace Award by the American Statistical Association for his paper titled "Coupled Markov Chain Monte Carlo for High-Dimensional Regression with Half-t Priors" authored with Dr. Anirban Bhattacharya, Dr. Pierre E. Jacob, and Dr. James E. Johndrow.

Ambarish Chattopadhyay, a 2022 graduate, received the Dempster Prize (named in honor of Emeritus Professor Arthur P. Dempster) for his paper "On the Implied Weights of Linear Regression for Causal Inference" with advisor Dr. José Zubizarreta.

Kuanhao Jiang, a 3rd year student, received the 35th New England Statistics Symposium Student Research Award. Kuanhao received this honor for his paper on "A New Central Limit Theorem for the Augmented IPW Estimator: Variance Inflation, Cross-Fit Covariance and Beyond" with Statistics Professors Pragya Sur and Subhabrata Sen and Biostatistics Professor Rajarshi Mukherjee.

Biyonka Liang, a 3rd year student, received an honorable mention for the 2022 Gertrude M. Cox Scholarship! The award was created in 1989 by the ASA's Committee on Women in Statistics and Caucus for Women in Statistics to encourage more women to pursue statistics-related careers.

Xiang Meng, a 3rd year PhD student, was selected to be a Bok Pedagogy Fellow for the department for 2022-2023. The Derek Bok Center for Teaching and Learning is a resource dedicated to supporting innovative teaching pedagogy in the Faculty of Arts and Sciences. As a Bok Pedagogy Fellow, Meng will serve as a liaison between the Department of Statistics and the Bok Center, assist with the teaching of STAT 303 The Art and Practice of Teaching Statistics, and provide support for our Teaching Fellows in the department.

Kelly Zhang, a 5th year SEAS Computer Science PhD student who works with both Profs. Susan Murphy and Lucas Janson, is a 2023 Siebel Scholars award winner. According to the Siebel Scholars website, "Each year, more than 90 exceptional students are selected as Siebel Scholars based on academic excellence and leadership potential."

Lu Zhang, a fall 2022 graduate, won the 2022 Statistical Learning and Data Science (SLDS) Student Paper Award for her paper entitled "Floodgate: Inference for Model-Free Variable Importance" and presented it at the SLDS section in JSM 2022. This paper is co-authored with her PhD advisor, Dr. Lucas Janson, and proposes a new inferential approach called floodgate.

POSTDOCTORAL FELLOW AWARDS

Raaz Dwivedi received the Best Student Paper Award (January 2022) for joint work on "Distribution Compression in Near-Linear Time" with Dr. Abhishek Shetty and Dr. Lester Mackey. Dr. Dwivedi received the Best Presentation Award (January 2022) for the talk on "Near-optimal Compression in Near-linear Time" at the 27th Annual Machine Learning and Statistics Session at MIT (2022). He also received a Certificate of Distinction and Excellence in Teaching (July 2022) by Harvard University.

Jingming Wang won the Best Poster Award for the 2022 Workshop on Statistical Network Analysis and Beyond. Dr. Wang's poster presentation was based on her paper "Optimal Network Membership Estimation Under Severe Degree Heterogeneity."

Celebrating PhD Graduates

Niloy Biswas

Dissertation Title:
"Contributions to Scalable Bayesian Computation"

Dissertation Committee:
Xiao-Li Meng, Lester Mackey, Neil Shephard

Ambarish Chattopadhyay

Dissertation Title:
"New Methods for Causal Inference in Randomized Experiments and Observational Studies"

Dissertation Committee:
José Zubizarreta, Carl Morris, Kosuke Imai

What's Next:
Starting in September 2022, I am a Postdoctoral Fellow in Data Science at Stanford.

Kristen Hunter

Dissertation Title:
"Topics in Randomized Experiments: Design, Modeling, and Power"

Dissertation Committee:
Mark Glickman, Luke Miratrix, Johann Gagnon-Bartsch

What's Next:
I am a faculty member at the University of New South Wales.

Phyllis Ju

Dissertation Title:
"Statistical Methods in Infectious Disease Modeling: Agent-based Models, Computations, and Selection Biases"

Dissertation Committee:
Pierre Jacob, Xiao-Li Meng, Neil Shephard

Celebrating PhD Graduates

Han Yan

Dissertation Title:
"Bayesian Models to Identify Hidden Patterns with Applications in Biology"

Dissertation Committee:
Jun Liu, Samuel Kou, Scott Edwards

Azeem Zaman

Dissertation Title:
“Statistical Advances in Bayesian Logistic Regression, Sports Forecasting, and the Theory of Distributed Estimation"

Dissertation Committee:
Natesh Pillai, Mark Glickman, Botond Szabó

Sanqian Zhang

Dissertation Title:
“Building upon Bradley-Terry and Plackett-Luce: Some Methods for Modeling Paired Comparison and Rank Order Data”

Dissertation Committee:
Mark Glickman, Jun Liu, Joe Blitzstein

What’s Next:
I am a Data Scientist at Robinhood.

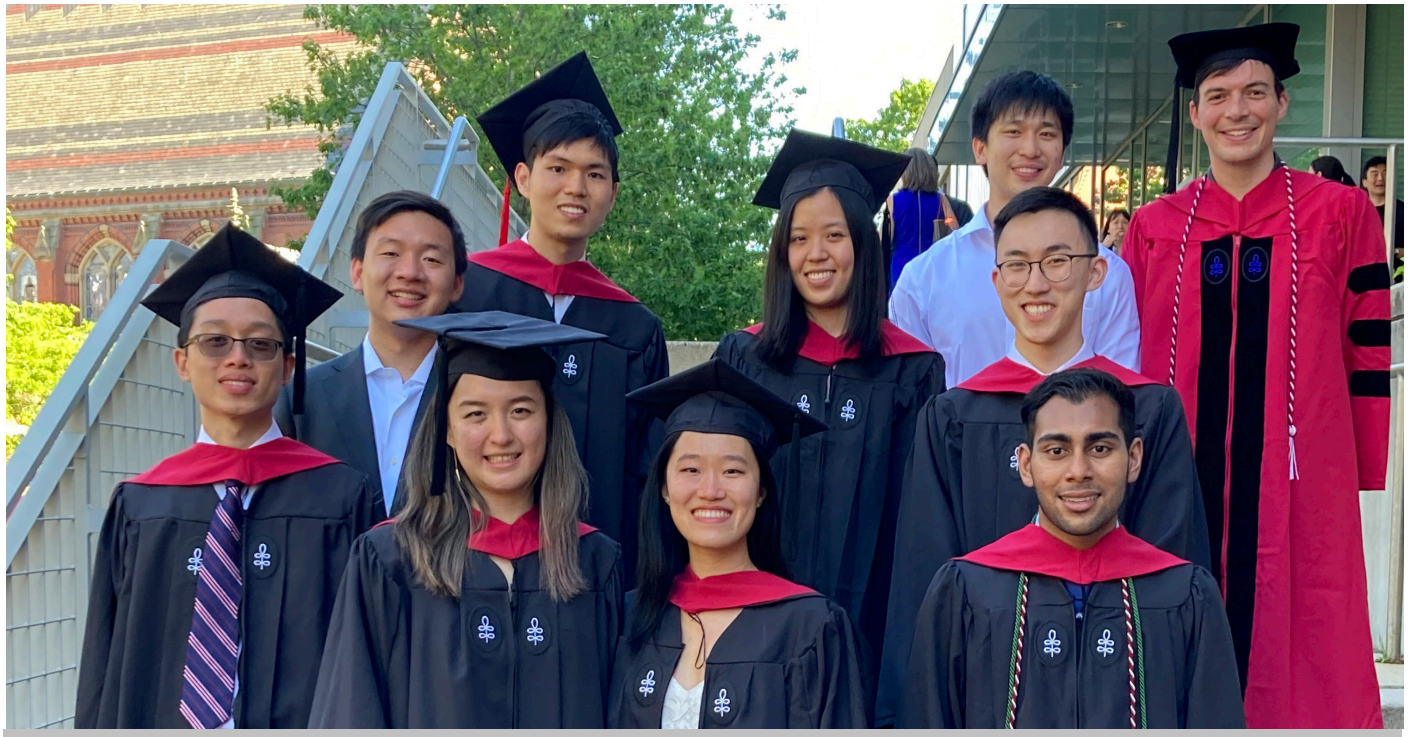
Celebrating AB, AM & PhD Graduates

Bachelor of Arts Graduates (2022), May 26th Commencement Celebration

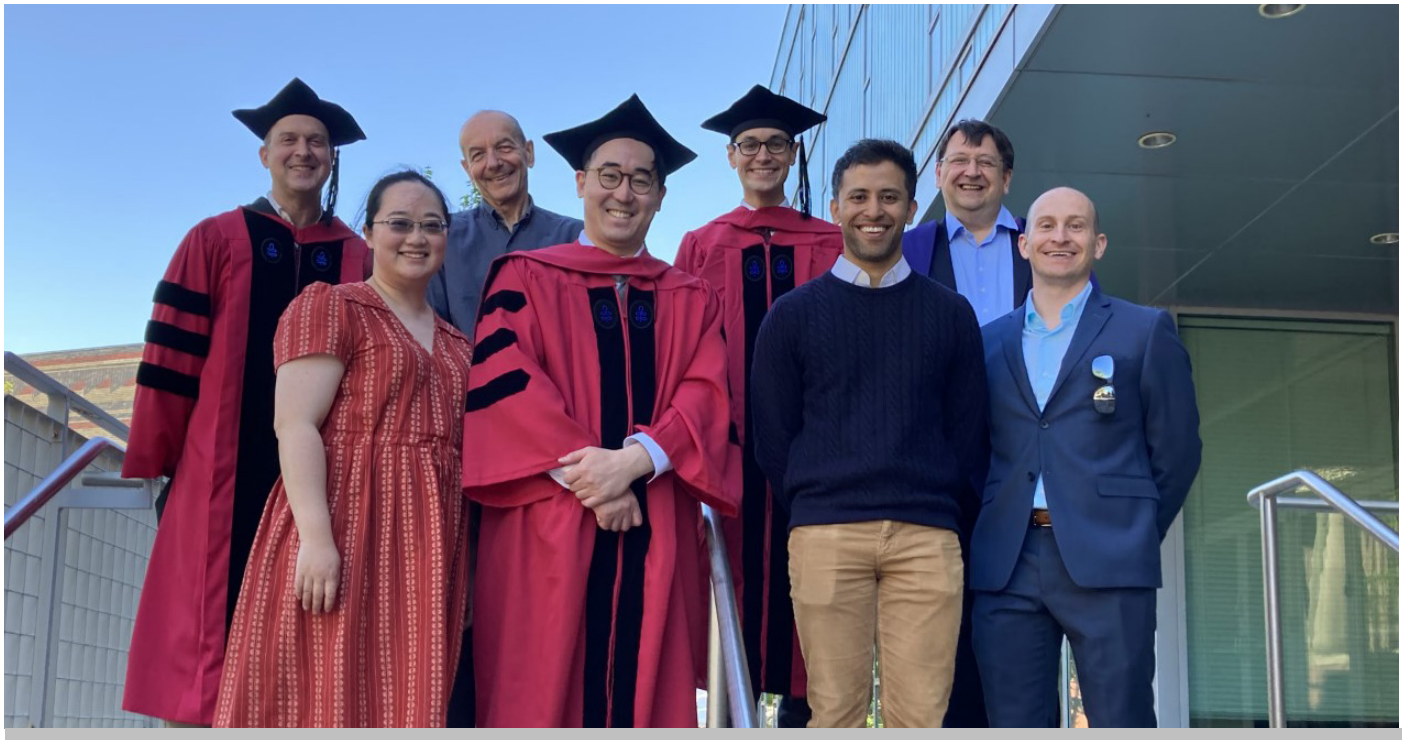
Master of Arts Graduates (2022), May 26th Commencement Celebration

PhD Graduates (2022) with Faculty, May 26th Commencement Celebration

Bachelor of Arts Graduates (2020/2021), May 29th Commencement Celebration



Master of Arts Graduates (2020/2021), May 29th Commencement Celebration



PhD Graduates (2020/2021) with Faculty, May 29th Commencement Celebration

HARVARD
Faculty of Arts and Sciences
DEPARTMENT OF STATISTICS

ADDRESS: 1 OXFORD STREET, SUITE 400
CAMBRIDGE, MA 02138

PHONE: 617-495-5496

EMAIL: statistics@fas.harvard.edu

WEBSITE: https://statistics.fas.harvard.edu

SOCIAL MEDIA: