
# Stats News Winter 23-24

## Department of Statistics

HARVARD
Faculty of Arts and Sciences

# Table of Contents

# Letter from the Chair

**Dr. Samuel Kou,**

**Chair of the Department**

Greetings Statistics Community,

I would like to share with you some of the major themes and events from the past year (spring 2023 until now), providing you with a brief snapshot of our department. The events, interviews, and stories in this newsletter celebrate our community of scholars and demonstrate the impact of mentor-mentee relationships. From former students traveling to attend Professor Emeritus Herman Chernoff's 100th birthday celebration to a concentrator sharing valuable career advice from Professor Neil Shephard, these stories are a testament to the bonds between our faculty and students.

This makes it fitting that this newsletter is dedicated to the memory of Professor Emeritus Carl Neracher Morris, who sadly passed away in spring 2023. In the Faculty of Arts and Sciences Memorial Minute (https://news.harvard.edu/gazette/story/2023/10/carl-neracher-morris-84/), Carl was remembered for his seminal contributions to exponential family, empirical Bayes methods, hierarchical models, sports statistics, and applications in healthcare policy. Carl was also remembered for his generosity and dedication towards his students, advisees, and junior faculty. Carl's legacy continues both through the students and colleagues that he worked with and through a new generation of scholars, such as Virginia Ma (AM '23), who describes in this newsletter her use of empirical Bayes.

Throughout the interviews in this newsletter with Professor Xihong Lin, Professor Emeritus David Harrington, and with our four student award winners from 2023, we see a bridge spanning from one generation of statisticians to the next. Alums Virginia Ma (AM '23), Jason Zhou (AB/AM '23), Rachel Li (AB '23), and PhD student Yicong Jiang tell us stories about how their teachers and advisors challenged and inspired them.

I would also like to share with you some important updates on students, faculty, and initiatives in the department. This year, we welcomed new faculty Lecturer James Xenakis, Assistant Professor Mark Sellke, and Affiliate Faculty Anna Seigal (Assistant Professor in Applied Mathematics at SEAS), who are playing a crucial role in supporting our courses, students, and contributing to the intellectual life and community in the department. The

concentrator, joint concentrator, and secondary field programs in statistics continue to thrive under the direction of Co-Directors of Undergraduate Studies, Professor of the Practice Joe Blitzstein and Senior Lecturer Kelly McConville. In spring 2023, the programs reported 246 undergraduate students and now are up to 330 students!

For the PhD program, Joint Director of Graduate Studies Professor Susan Murphy led our admissions efforts (Professor Jun Liu co-directs the PhD program and oversees student progress) resulting in seven new PhD students matriculating this fall: Nic Fishman, Anvit Garg, Yuanchuan Guo, Benedikt Koch, Somak Laha, Sarah McDonald, and Wenqi Shi. In the past year, we celebrated six graduating PhD students, including Jonathan Che, Dieyi Chen, Cory McCartan, Feicheng Wang, Ran (Christy) Huo, and Lu Zhang. You can read more about both our G1 students and our recent graduates in this newsletter.

Our master's program, under the direction of Senior Lecturer Mark Glickman, has grown to have thirty new AM students this year. This program allows both graduate students in other programs and strong undergraduate concentrators (in statistics and other areas) to complete a Master of Arts degree in Statistics by taking additional courses.

This year there are both more students concentrating and pursuing a master's in statistics and more students from other undergraduate and graduate programs seeking to develop a statistical skillset. In fact, in academic year 2022-2023, course enrollments reached over 3,000! To meet student demand and interest, our faculty have developed and taught the following exciting, new courses in the past year:

- Stat 108 Introduction to Statistical Computing with R (by Kelly McConville)
- Stat 143 Sports Analytics (by Mark Glickman)
- Stat 160, Introduction to Survey Sampling and Estimation (by Kelly McConville)
- Stat 175 Statistics and Data Science of Networks (by Morgane Austern)
- STAT 217 Topics in High-Dimensional Statistics: Methods from Statistical Physics (by Subhabrata Sen)
- STAT 291 Random High-Dimensional Optimization: Algorithms and Barriers (by Mark Sellke)

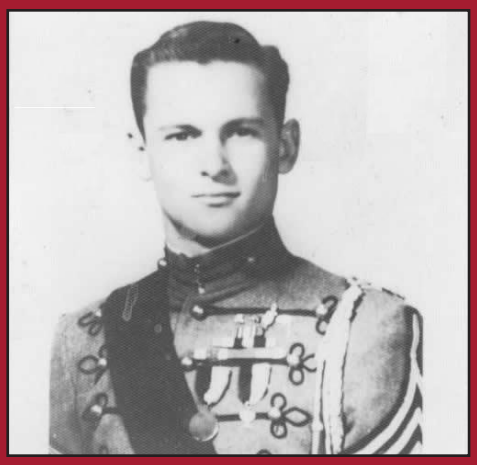
While I am proud of this rising interest in statistics and our programs, I also acknowledge that there is more work to do towards building a diverse, inclusive community. As part of this initiative, we started the Equity, Diversity, Inclusion, and Belonging Committee (EDIB), which contains staff, postdoc, faculty, and student representatives and is chaired by Dr. McConville. The EDIB Committee has made progress by creating a resource webpage, soliciting feedback from the Harvard statistics community, planning more events, and conducting community outreach. A few highlights include Dr. McConville and a committee planning Data Adventure Day, an all-day statistics event on campus for ~100 high school students, and Assistant Professor Morgane Austern and the Women in Statistics Group hosting the Healthcare Career Panel.

As Chair of the Department, it's my pleasure to share with you in this newsletter the talented work that my colleagues, students, postdocs, and staff are performing– whether in their research, teaching, or departmental service. I hope you enjoy reading about their accomplishments as much as I have.

*Samuel Kou*, **Professor**

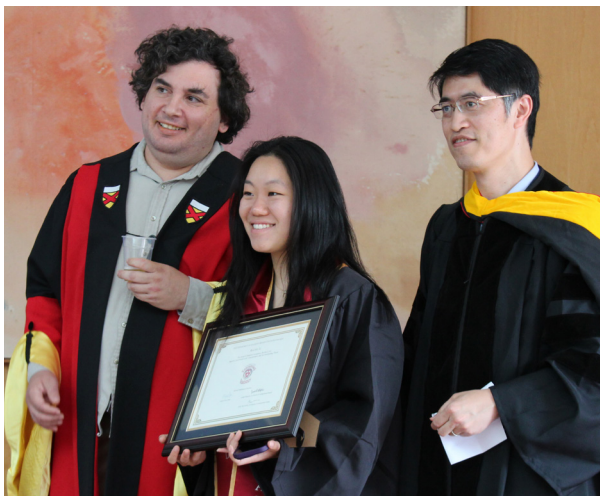
Chair of the Department

**This newsletter is dedicated in memory of Faculty Emeritus Carl Morris, a brilliant statistician who is deeply missed by his family, friends, colleagues, and mentees.**



# Profile of Rachel Li

## *Department of Statistics Senior Concentrator Prize Winner*



This year, concentrator alum Rachel Li received the 2023 Department of Statistics Senior Concentrator Prize for her superb coursework in the concentration and for an outstanding thesis. Originally named the Department of Statistics Undergraduate Prize, the award was founded in 2020 and is given annually to the graduating senior concentrator who has the best overall performance and has contributed significantly to the department. In an interview (edited and excerpted below), Rachel shared her most influential mentors and courses as well as her most valuable experiences, including co-founding the Group for Undergraduates in Statistics at Harvard (GUSH). Congratulations on your achievements, Rachel!

**When did you first become interested in statistics?**

**Li:** My high school did not offer AP Statistics, so I didn't know that much about statistics before I came to college. But my older brother Vincent, who also went to Harvard, recommended that I take Stat 110 [Probability] during my freshman fall. Following his advice, I took Stat 110 and then Stat 111 [Introduction to Statistical Inference], which taught me how to think about probability, randomness, and data within an analytical framework. Prof. Joe Blitzstein's approach to teaching – his way of demonstrating the elegant simplicity of statistical ideas – made me further enamored with the subject. The concentration also appealed to me because it's practical; data needs to be analyzed in every field, e.g., in biostats or finance, which makes statistics an extremely useful toolkit.

**What were a few highlights as a concentrator in the program? Are there certain mentors, courses or parts of the program that really shaped your experience and interest in the field?**

**Li:** The courses that had the biggest impact on me were Stat 110 and Stat 111 because they were at the beginning of my college career and introduced me to great friends and lasting mentors. There's nothing that makes you bond more with fellow students than going to three office hours in a row each week! On Thursday nights, my friend Ginnie (2023 Statistics AM graduate Virginia Ma) and I would attend Asher Spector and Kim Nguyen's office hours and would think, "This is fun! We should try to TF for this class one day." Asher and Kim's great teaching inspired me to become a TF in these courses.

Also, there's just so much to say about the welcoming and steady mentorship that Profs. Neil Shephard and Joe Blitzstein have shared with me throughout my time in college! During my senior fall, I remember a particularly

pivotal conversation I had with Neil when I was ruminating over whether or not to pursue graduate school. When I told him about my worry that I wasn't "smart enough" for a career in research or academia, Neil opened up to me about his own self-doubts that he also had from time to time. He gave me the advice to treat research like play and try to find joy and excitement in the continual process of learning.

Also, Joe, who is warm, funny, and goes above and beyond for his students, has been one of the most instrumental professors for me at Harvard. He has taught me so much, not just in terms of probabilistic thinking, but also about what it looks like to teach in a way that inspires students and makes difficult concepts approachable for everyone. As both a student and a member of the teaching staff in his courses, I have witnessed Joe's love for teaching and his dedication to his students. He truly sets a high standard of teaching that I hope to strive towards someday.

In addition to Joe and Neil, the department as a whole offers a friendly, welcoming atmosphere. I felt a sense of community in the department throughout my time at Harvard, which made my experience as a concentrator enjoyable and fulfilling. I also recognize that my experience may not be universal and that achieving inclusion and diversity are big challenges both in the field of statistics and in the department. However, I'm hopeful that we can continue to make moves to build an inclusive, diverse department. In a spontaneous office hours conversation with Joe in our freshman year, Ginnie and I were inspired to create GUSH (Group for Undergraduates in Statistics at Harvard). Our goal with this new organization was to provide resources and create a space on campus for all statistics enthusiasts to connect and share their common interests. Community-oriented events like the socials GUSH has held, as well as the department coffee hours, will hopefully help to welcome a more diverse group of students into the growing statistics community at Harvard.

**How did you select your thesis project? What questions were you asking?**

**Li:** During my junior spring, I took CS 238 [Optimized Democracy] with Professor Ariel Procaccia in the Computer Science Department. I really enjoyed the course and his teaching, so I decided to pursue a research project with Ariel as my advisor. My thesis project, which I worked on with CS graduate student Daniel Halpern, is titled "Strategy-Proofness of Voting Rules Under Voter Beliefs." Our research focuses on how to design election mechanisms that incentivize voters to report their preferences truthfully and prevent vote manipulation. Most elections are not "strategy-proof" because people often hide their true preferences in an attempt to game the system. For example, in ranked choice voting, if I have two favorite candidates, I may rank my top choice first but then rank my second choice last to boost my top choice's chances of winning. A crucial part of calculating the degree to which an election system, such as plurality [each voter votes for one candidate and the candidate with the most votes wins] or ranked choice [each voter ranks multiple candidates], is strategy-proof depends upon the probabilistic distribution of voters' beliefs. In our work, we show that when a voter believes that other participants are likely

to agree with her top choice candidate, the plurality system is more mathematically robust to protect against vote manipulation than ranked choice voting. Working on this research problem was satisfying because the theory – and crazy math behind it – helps us to understand real-world voting scenarios.

**What are a few things that you will miss the most about your Harvard experience?**

**Li:** What stands out the most to me about my experience at Harvard (and makes me a little sad!) is how college captures a specific time and place in my life that is fleeting. Part of what makes the experience on campus so special is the richness and closeness of your social interactions; you can simply walk down to the dining hall and run into ten people you know, catch up with them, and ask them about how their day has been. You might spend a couple hours "working" on a problem set with your friends, but also probably get distracted and start discussing random hypotheticals in a conversation that will stray late into the night. Harvard, in particular, is also unparalleled in its ability to foster learning and attract intelligent, passionate students. As someone who asks a lot of questions and loves learning, I have felt blessed to be in this unique environment with access to so many experts, resources, and excellent teachers.

**What are you looking forward to? What are your short-term goals?**

**Li:** I'm looking forward to starting adult life in New York City in the fall, including beginning a job as a fundamental analyst at Citadel Securities. After a lot of reflection in the past year about my goals, I've settled on something that I am certain about: I am curious; I enjoy learning and thinking about interesting ideas; and I want to continue doing this for the rest of my life. Having had a really great experience with my thesis research, I am definitely considering exploring graduate school in the future.

# Profile of Virginia Ma

## *Concurrent Masters Prize Winner*

In May 2023, Virginia (Ginnie) Linqian Ma won the inaugural Department of Statistics Concurrent Masters Prize (along with Jason Zhou '23 AB/AM). This prize is awarded annually to the graduating master's (AM) student who has the best overall performance in coursework, has demonstrated achievements in Statistics outside of coursework, and has contributed significantly to the department. In the following interview (excerpted and edited) from the summer (pardon the publishing delay!), Ginnie speaks with us about her journey as a master's student through the Stats Department. Read more to learn about Ginnie's favorite courses, thesis project, and inspiration for starting GUSH and the Florence Nightingale Day (a day of statistics outreach for middle and high school students).

**When did you first become interested in statistics?**

**Ma:** In high school, I was very lucky that we had the opportunity to take a Stats class – not your typical AP stats class – with a professor from a local university whose daughter went to our school. The course incorporated calculus and was similar in flavor to Statistics 110 [Probability] and 111 [Introduction to Statistical Inference]. I knew I already enjoyed taking math classes, but statistics interested me because you could use it to make sense of real-world phenomena. The second experience that I had in high school (in my junior year) was a summer research project working with gene expression data from pancreatic cancer patients. Working on the project made me realize that I was mostly interested in what was going on under the hood; I wanted to know why these methods worked and the principles behind them.

From these past experiences, I knew I valued a more theoretical perspective and was interested in the interplay between my pure math and statistics courses in college. As a math concentrator, I learned how to think critically; writing a proof is different from writing anything else. After taking a lot of pure math classes, I could synthesize the proof techniques and apply them to my work in statistics, particularly in proof-based stats classes. As I progressed in math and statistics, I became more interested in pursuing 200-level stats courses and the master's program in statistics. We're very lucky that the department gives us this unique opportunity to receive a concurrent master's degree.

**What were a few highlights as a master's student in the program? Are there certain mentors, courses or parts of the program that really shaped your experience and interest in the field?**

**Ma:** This is probably a frequent answer, but I think the best way to introduce someone to Stats Department courses is by taking the Stat 110 and Stat 111. The year I took Stat 111, it was co-taught by Prof. Joe Blitzstein and Prof. Neil Shephard, which produced an amazing classroom dynamic. I took several classes with Joe, and his dedication to making sure that all students enjoy and engage with the material was essential for making the Stats Department a welcoming place for me – a place where I could see myself tackling difficult problems. Also, Joe introduced me to the beauty and elegance of statistics in Stat 210 [Probability I]. For example, the course taught us the technique of using representations to solve for a distribution function. Typically, to relate the distributions of two random variables, you need to use a tedious amount of algebra or calculus. Using a representation simplifies the solution and makes it more concise (instead of writing out pages of work!), which was incredibly eye opening and satisfying for me.

Another class that I really enjoyed was Stat 171 [Introduction to Stochastic Processes] with Professor Subhabrata Sen. Stochastic processes are probabilistic models for random quantities that typically evolve over time or space. The course was a great example of a class that balances theory and coding; when learning about stochastic processes, such as Markov Chain Monte Carlo, we needed to use both theoretical and computational skills to solve the problems.

**How did you select your thesis project? What questions were you asking?**

**Ma:** I selected Prof. Lucas Janson for my math thesis advisor because I was interested in research in statistical inference (I also enjoyed his Stat 211 course on this topic) and because I had heard he was a great advisor from other students. I reached out to him to discuss a potential research project; it seemed like a good fit, so I went for it! Throughout the process, Lucas was an excellent mentor because he struck a nice balance between giving me independence to explore on my own and providing helpful guidance to make sure that I didn't stray too far off track.

My math thesis was titled "Uncertainty Quantification with Empirical Bayes." Uncertainty quantification is useful for capturing the degree of uncertainty associated with models, estimates, and other quantities of interest. Let's say that you want to estimate the average height of a population. Because you can't perform a census for every single person, you take a sample of people's heights and average them. However, your sample might be quite small or not very representative of the general population, so it can be beneficial to provide a range instead of a point estimate (e.g., estimating the average height is between 5'2" and 5'4" instead of estimating 5'3"). High dimensional data presents the challenge of identifying which covariates are relevant to the outcome variable. After identifying relevant covariates, you can perform inference such as hypothesis testing or uncertainty quantification. One appealing approach toeing the line between the two primary frameworks of statistical inference—frequentist and Bayesian—is empirical Bayes, which shares properties of

Bayesian methods, such as allowing for the use of data multiple times in constructing a model and performing analysis, but has weaker prior assumptions. My thesis examined whether empirical Bayes Methods could help with the accuracy of uncertainty quantification in these high dimensional settings. The goal was to use mathematical and statistical tools to analytically evaluate these methods and then see if they produced better interval estimates.

**What are a few things that you will miss the most about your Harvard experience?**

**Ma:** The Stats Department was a special community for me in many ways. One aspect that I appreciated was that the PhD students were incredibly friendly and kind, which was something I wasn't expecting because they are older students with more established paths in statistics. Once I started taking higher level classes, the PhD students were amazing classmates, too, and I learned a lot from them.

Another part of the Harvard community that I miss is GUSH [Group for Undergraduates in Statistics at Harvard], which was a big part of my college experience. As you know, Prof. Blitzstein helped Rachel [2023 stats concentrator alum Rachel Li] and me to get GUSH off the ground. The first official event in May 2020 was an alumni panel on Zoom, which had a great mix of speakers and strong attendance. This was an exciting event to launch the group because we saw all these students come together after being sent home at the beginning of the pandemic. GUSH was a way to stay in touch with people and remain integrated with the community. A second favorite GUSH event would be the women in statistics panels that we've hosted for two years. Held in conjunction with Women's Week (hosted by the Harvard Women's Center), these career panels present a great opportunity for students to hear from panelists at different stages in their career about varied career paths and challenges faced by women statisticians.

Another event that I was proud of working on was the Florence Nightingale Day in the Stats Department in October 2022. [Introduced by the American Statistical Association and the Caucus for Women in Statistics (CWS) in 2018, FND is part of an international effort to celebrate women in statistics, biostatistics, and data science and to cultivate student engagement in these fields. In fall 2023, we hosted a similar event for high school students from Boston and Cambridge public schools called Data Adventure Day]. I had participated in a Florence Nightingale Day [FND] when I was in high school and was interested in initiating one at Harvard, but it wasn't until I reached out to Prof. Kelly McConville that the initiative launched. Kelly was very invested in educational initiatives, particularly those focused on equity. When I emailed her about FND, she sent me a Google Doc full of ideas that same weekend! With her help and the other committee members, we were able to bring about 50 middle and high school students to campus for a full day of statistics activities.

**What are you looking forward to? What are your short-term goals?**

This summer, I'm looking forward to relaxing and recharging.One of the activities

that I picked up last summer at Harvard was running along the Charles River. While I ran track in high school and always thought of myself as a "short distance" person, I'm working on building up my stamina and running longer distances, which I hope to do at home this summer and at school in the fall. In addition, I look forward to enjoying the outdoors more by going hiking with my family and gardening. Lastly, I'm going to try to read more this summer. Joe gave me a book that details advice from statisticians about their tips for graduate school and beyond, so I'm planning on reading the book before I move across the country to start my PhD program in Statistics at Stanford. Actually, the first time that I ever traveled to California was during the visit weekend for incoming students; I'm excited about moving and getting started!

# Profile of Jason Zhou

## *Concurrent Masters Prize Winner*

Statistics concentrator and master's alum Jason Zhou was awarded the inaugural 2023 Department of Statistics Concurrent Masters Prize (along with alum Virginia Ma). To celebrate his award and learn about his thesis research and program highlights, we conducted the following interview (excerpted and edited). Congratulations, Jason!

**When did you first become interested in statistics?**

**Zhou:** I first thought I would major in economics, but when I took Stat 104 [Introduction to Quantitative Methods for Economics], I was intrigued by the emphasis the course placed on drawing nuanced quantitative conclusions from data and using these conclusions as the basis for making other qualitative arguments. Throughout the course structure, I felt encouraged to deepen my questions and examine my process for making quantitative conclusions. In general, statistics appealed to me as a concentration because I was learning quantitative tools and methods that could be applied to real-world problems.

**What were a few highlights as a concentrator/ master's student in the program? Are there certain mentors, courses, or parts of the program that really shaped your experience and interest in the field?**

**Zhou:** A clear highlight for both my undergraduate and master's programs has been how supportive the professors are in the department. The faculty want students to go to their office hours and talk about how statistical thinking can be applied to real-life problems.

In particular, the master's program was a rewarding experience because its courses included topics that broadened my view of statistics. Seeing the passion that professors have for their teaching motivated me to learn more about these topics. For example, Stat 210 [Probability I] was fun because it proposed a new way of thinking about distributions and introduced the idea of using representation. At the same time, the course rigorized statistical concepts; for example, we used mathematical axioms and tools to show how the concept of probability could be expanded. Stat 211 [Statistical Inference I] was also a formative part of my master's experience. Prof. Lucas Janson did an excellent job calibrating the level of difficulty of questions on homework and exams so that they were just a little beyond our reach and would lead to further discussion and insights in class.

**How did you select your thesis project? What questions were you asking?**

**Zhou:** Before I even started thinking about research, I kept hearing people at Harvard mention the name "Xiao-Li," which piqued my interest in him and his research. When I read a 2018 paper of Prof. Meng's on data quality, I was impressed by the idea of creating a mathematical formulation to determine what data quality is for mean estimation. In data science, there is a lot of talk about the "importance of data quality," but Xiao-Li's paper explained in detail what data quality means and how it impacts research results. After reading Xiao-Li's work, I was interested in extending this concept of data quality to other estimators, which led me to my thesis topic: "Data Quality Always Matters: An Analysis of Data Quality for Finite Population Z-Estimators" [Jason's thesis received the Hoopes Prize and was supervised by Professor Xiao-Li Meng]. My thesis project focused on defining data quality for a broad class of nonlinear estimators. One of the main goals was to show that you still need sufficient data quality for an estimator to converge; in other words, data quantity won't compensate for having bad quality data. Hopefully, this paper will encourage more researchers to think carefully about the data quantity versus quality trade-off. If I had more time to work on this project, I would explore the topic of methodological inefficiency more and would incorporate more concrete examples and use cases.

**What are a few things that you will miss the most about your Harvard experience?**

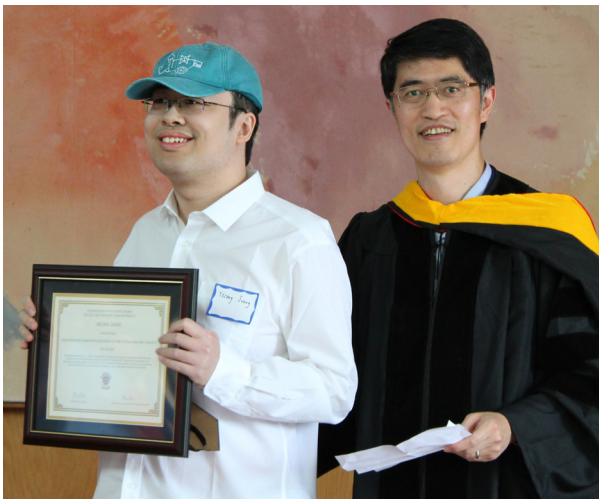
**Zhou:** I'll miss two things the most about Harvard: the commitment to learning for learning's sake and, related to this, the people. Harvard offers the kind of environment in which you can follow your passions and intellectual curiosity, which is incredibly gratifying. Students and faculty at Harvard care deeply about diverse interests, which has encouraged me to learn new skills from others; I even learned ballroom dancing in my senior year (the foxtrot was my favorite)! In college, there's just such a low activation energy to quickly get involved in fun and rewarding endeavors.

**What are you looking forward to? What are your short-term goals?**

**Zhou:** I'll be living in Chicago for the first time, so I'm excited to explore the culture and try out some deep-dish pizza. I am moving there to work at a Quant Firm called IMC – I'm looking forward to getting started!

# Profile of Yicong Jiang

## *PhD Dempster Prize Winner*



In the May 2023 Commencement Celebration, PhD student Yicong Jiang was awarded the 2022-2023 Dempster Prize for his paper co-authored with Professor Tracy Ke, "Semi-Supervised Community Detection via Structural Similarity Metrics." The Dempster Prize is named in honor of Emeritus Professor Arthur P. Dempster and is given annually to a graduate student in recognition of the student's research. In the following excerpted and edited interview, Yicong shares with us how the Stats Department helped him adjust to living in a new country, discover his passion for research, and hone his ability to collaborate with other researchers. Congrats to Yicong!

**When did you first become interested in statistics?**

**Jiang:** When I was in elementary school, my parents sent me to an additional school to learn higher level math, which is a common practice among parents in China to ensure that their kids get into a good high school. Students who take part in competitions, e.g. math competitions, and receive awards, are often assigned to a better, more competitive high school than the normal random assignment to a school. After high school, I went to Peking University, where I was a math major, which included both pure math and statistics. On the one hand, I really loved the intrinsic relationship between numbers in pure math. However, during my sophomore year, I realized that I was also interested in statistics because it reveals the relationship between data and the latent structure of data. Statistics helps us to estimate certain parameters or to discover structures like representations. I continued to be intrigued by statistics when I arrived at Harvard for my PhD because, for the first time, I could dive into research to gain novel insights into the relationship between data.

**What have been some highlights of the PhD program? Are there certain mentors, courses, or parts of the program that really shaped your experience and interest in the field?**

**Jiang:** As I mentioned, after undergrad, I came to the US for the first time to pursue my PhD at Harvard; I was excited to study at one of the best universities for statistics. While I was a little nervous at the beginning of my program, I soon realized that the people here – my advisors, professors, and cohort – were very friendly and helpful. Because the people in the department helped me to learn information about the PhD program, Harvard, and the US in general, I adapted to my new environment in about 1-2 months, and now, I've been here for three years! From the beginning of my PhD experience, Stat 210 [Probability I] stands out as a course. As you may know, Professor Blitzstein is a superstar on YouTube! In his 210 class, he did an excellent job

of teaching abstract, difficult concepts by using simple examples and stories. In doing so, Prof. Joe brought abstract concepts of probability close to us – we could almost reach out and touch them.

Another influential course for me was Stat 303 [The Art and Practice of Teaching Statistics], which I took with Professor Pragya Sur. By teaching us research communication and presentation skills, Prof. Pragya's course helped my cohort build a collaborative research community. Before taking the course, we mostly worked independently on projects without seeking much feedback, but through the course, we learned how to effectively discuss and exchange research ideas with each other. Also, the course helped me to learn professional communication skills in English, including how to highlight core ideas in my presentations instead of simply showing slides with many details and formulas.

**How did you select your research project? What questions were you asking?**

**Jiang:** After I read Professor Tracy Ke's paper on community detection in an unsupervised setting (meaning that you don't have information about the data), I was interested in her methods for clustering different people into groups and wanted to explore this concept with her. The paper that I wrote with Prof. Tracy, "Semi-supervised Community Detection via Structural Similarity Metrics" (semi-supervised means that you have a limited amount of information about the data), uses statistical methods to draw conclusions about communities within a network, such as a social media network. For example, let's say that you are interested in clustering people into political groups, Republicans and Democrats, but you only have limited information about the politicians that they follow on Twitter [now X]. If "Alice" and "Bob" follow most of the same politicians, we might expect that they are part of the same political affiliation. But how do we measure or score the closeness of their relationship to Democratic or Republican parties? Using structural similarity metrics, we can calculate the degree to which Alice and Bob exhibit the same following behavior and are likely to be a part of the same political community. These methods for community detection can be applied in other ways, such as making recommendations of artists to people based on an artist they follow or making recommendations of people to connect with based on their shared passions.

**What are you looking forward to? What are your short-term goals?**

**Jiang:** Prof. Tracy and I want to continue to work on semi-supervised community detection to develop methods for addressing the scenario in which people don't just fit one category or another. Within our current framework, we only make the judgment that people are pure Democrats or pure Republicans, for example, but sometimes people might be neutral to varying degrees (e.g., they might identify as 30% Republican leaning and 70% Democratic leaning). Because there are many possible percentage combinations where people can be considered neutral, the problem is quite difficult.

Outside of research, I am looking forward to a road trip to Yellowstone Park that my friends invited me to join (we are still in the planning phase). Our plan is to drive there; I know this sounds a little crazy, but my friends want to see the sights along the way from Boston to Yellowstone. I'm excited to go hiking because it was an activity that I enjoyed doing with my father when I was younger, but I'm also a little nervous because I haven't had time to go hiking that much since high school. As long as I have sturdy shoes, I should be fine!

# Professor Herman Chernoff

"It's fair to say that a person can only celebrate a 100th birthday once in a century," pronounced Professor of the Practice Joe Blitzstein of the Harvard Department of Statistics. Leaning towards an audience of 100 gathered at Harvard's Student Organization Center at Hilles, he added, "Today, we celebrate with our beloved colleague Herman Chernoff his birthday and 70 plus years of contributions to the field of statistics." With these words, Professor Blitzstein launched the May 5, 2023 Centennial Celebration of Emeritus Faculty Herman Chernoff and his research and teaching legacy. The day featured a full program, including a virtual tribute video from former colleagues and students, an interview with Professor Xiao-Li Meng, and research presentations from former colleagues Professors Joseph (Jay) Kadane, Tian Zheng, and Joseph Gastwirth. For making this day happen, we would like to thank the Chernoff family: his daughters, Miriam and Ellen Chernoff, and Miriam's husband, Stan Morse. We would also like to express our deepest condolences to

## *Celebrating Prof. Chernoff's 100th Birthday*

Professor Chernoff and his family for the recent loss of Judith Chernoff, his wife of over 75 years.

In acknowledgement of Professor Chernoff's birthday, we would like to provide an overview of his career highlights as well as the emerging themes from the symposium. The event celebrated Chernoff as a statistician, educator, and mentor, who has cultivated lasting relationships and possesses a great sense of humor.

**Career Highlights**

Chernoff started his career when he completed a BS in mathematics and a minor in physics in 1943 from City College in New York. For a year and a half, he worked as a physicist with the US Navy by building and fixing electronics. According to John Bather in his article "A conversation with Herman Chernoff in Statistical Science," Herman's use of statistical ideas in the Navy convinced him to return to school to first pursue a master's and then a PhD in applied math at Brown University, where he was supervised by his dissertation advisor Professor Abraham Wald. Subsequently, Chernoff held faculty positions at the University of Illinois (1949-1952), Stanford University (1952-1974), Massachusetts Institute of Technology (MIT) (1974-1985), and Harvard University (1985-1997); he continues to be Professor Emeritus at Harvard and MIT.

Chernoff's contributions to the field include work on large sample theory, experimental design, sequential analysis, methods of

presenting statistical data in visual form, and statistical decision making. In addition to his research, Professor Chernoff is known for his enthusiasm for mentoring a generation of outstanding statisticians. His interest in providing professional opportunities for young researchers led him to suggest the annual New England Statistics Symposium (NESS), established in 1987. In honor of Chernoff's contribution to NESS and to the profession in general, the New England Statistical Society established The Chernoff Excellence in Statistics Award in 2019 (the most recent awardee was Harvard statistics PhD alumna Professor Nan Laird). In recognition of his work, Chernoff has received honors from the National Academy of Sciences and the American Academy of Arts and Sciences, to name a few, and has been selected as a Fellow of The American Statistical Association and the Institute of Mathematical Statistics.

**An Expansive, Curious Mind**

In an intimate conversation with Professor Xiao-Li Meng, Chernoff revealed to the audience his reasons for switching from mathematics and physics to statistics. Chernoff described a pivotal moment in his graduate student career when he read a paper on generalizing the testing of hypotheses and estimation by the scholar Abraham Wald in a reading course: "Wald's paper struck me because, while it didn't involve intensive mathematics, it confronted the fundamental idea that the test of a hypothesis or an estimation of a parameter leads to a conclusion and that conclusion should have an economic, real-world consequence." Turning from Meng to the audience, Chernoff concluded, "and that's what converted me to being a statistician!" Chernoff added that he remembered reading papers by the statisticians Neyman and Pearson that exposed him to the idea that when evaluating a hypothesis, it was important to consider alternatives to that hypothesis. These experiences highlighted what drew Chernoff to statistical thinking: his interest in connecting theory to applications and solving and quantifying problems of uncertainty. Although he was trained as a mathematician and physicist, Chernoff relished the opportunity to tackle a new field, statistics.

In his career as a statistician, Chernoff embraced working in both applied and theoretical areas, a rare feat today because of how specialized statistics has become. From the start of the centennial celebration, Blitzstein emphasized Chernoff's multi-faceted work in statistics, pointing out his applied and theoretical, Bayesian and frequentist, parametric and nonparametric approaches. The three research talks given by Chernoff's colleagues also showcased the span of his research interests and influence. While the first talk by Professor Emeritus Jay Kadane of Carnegie Mellon University focused on the use probability theory to analyze handwriting in a court case, the second talk by Professor Tian Zheng of Columbia University discussed methods for detecting influential variables in high dimensional data, specifically genetic data. Additionally, Professor Joseph Gastwirth of George Washington University spoke about his collaboration with Chernoff on the use of L-statistics in measuring economic inequality. Representing such a wide range of interest and expertise, these talks were a reminder of

Chernoff's curiosity and expansive thinking.

In his interview with Professor Meng, Chernoff further explained why the manifold applications of statistics, such as in genetics, clinical trials, or economics, have motivated his career. Chernoff reflected on his research legacy, "People regard me as a theoretical statistician, but I've decided in recent years that I'm really an applied statistician. My theoretical insights have relied upon my work in thinking about applied problems." A salient example of Chernoff taking inspiration from applied problems was when he created Faces. Designed by Chernoff while he was working at Stanford, Faces was a data visualization tool that he developed to help researchers analyze multivariate data (by presenting data as faces). While Chernoff was still interested in developing statistical theory, he was also invested in working on more applied problems like how to effectively display multidimensional data.

The centennial celebration, from the interview with Professor Meng to the research talks to Professor Blitzstein's anecdotes, illustrated Chernoff's life-long love of learning in statistics and other fields. While his curiosity drove him to study statistics for the first time as a graduate student, it also motivated him to seek tips on the programming language R. Describing Chernoff's budding interest in R, Blitzstein said, "I was pretty impressed that he was still coding; usually brilliant mathematicians get their PhD students to do all the coding, but Herman wanted to test out his methods by carrying out the simulations himself." Blitzstein added, "One day, Herman came into my office and asked for a book on C because 'R was too slow for him,' and I was even more impressed with that!" It's clear that Chernoff has maintained a zest for learning new things and taking on new challenges.

**A Generous Educator and Mentor**

Having taught statistics in the department for many years, Chernoff still encouraged and mentored undergraduates once he became an emeritus faculty. Blitzstein reminisced about when their offices faced each other, and he would routinely return to his office and find Chernoff ensconced in a conversation with a student. "My students from Stat 110 would receive homework help from this friendly man in his 80s without even realizing that he had pioneered some of the methods they were using 60 years before," marveled Blitzstein. This story reveals how Chernoff maintained a life-long interest in nurturing students' enthusiasm for and competence in statistics. In addition, Chernoff demonstrated his interest in supporting the intellectual growth of younger students by publishing books such as *Elementary Decision Theory* in 1959 and *Algebra I for Students Comfortable with Arithmetic* in 2001. Holding his own autographed version of Elementary Decision Theory, Blitzstein commended the book for being accessible to high school students but also for addressing some of the most pressing questions in statistics.

While the centennial celebration highlighted the impact of Professor Chernoff's work, some of the most memorable moments during the conference were when colleagues and former students shared their stories about his tips and

coaching on their research and careers. For example, when Professor Gastwirth was a junior colleague at Stanford, he was thrilled to be asked for the first time to be a reviewer of an article and was willing to offer a tight turnaround with his comments. Sharing the following advice from Chernoff that he's adhered to ever since, Gastwirth said, "When I told Herman about finishing the review within a week, he said, 'No, no—you have to think about the worst possible thing that can happen to delay you (e.g., you get sick) and then you double that amount of time!'"

Harvard Statistics Professors Jun Liu and Xiao-Li Meng both shared anecdotes about Chernoff's mentorship. Shortly after Liu arrived as a junior faculty at Harvard, he was drafting his first paper in which he was a first author and Chernoff offered to read his paper. Liu recalled, "I felt a little intimidated by such a renowned statistician offering to read my paper, but Herman read it and provided such useful, line by line comments that it was ultimately accepted into the Annals of Statistics." During his interview with Chernoff, Meng highlighted an episode that occurred earlier in his career as a PhD student in the department. In preparation for his qualifying exam, Meng had printed a copy of his paper and commenced to read every page to his committee. Chernoff paused the perusal with a question that Meng promptly glossed over. Chernoff interjected, "Xiao-Li, you are not answering my question." From this moment, Meng learned that it was not sufficient to pursue important research questions; he must also learn how to communicate effectively to succeed as a researcher.

## Lasting Relationships with Colleagues, Family, and Friends

The warm outpouring of thanks and congratulations to Professor Chernoff is a testament to the close-knit relationships that he has maintained. During a tribute video and in person toasts, friends; family; and former colleagues, students, and postdocs acknowledged Chernoff's impact on them. Attendees reminisced about grad student lunches at a pizza parlor and cozy gatherings at the Chernoff home, and they shared some of Chernoff's favorite interests and things: politics and travel; dogs and Swiss orange chip ice-cream. One attendee, the son of a former high school classmate, marveled at the journey that Chernoff had taken as a child during the Great Depression from a family of modest means to becoming an influential statistician. Sharing their regret that she could not participate in the symposium, many attendees paid tribute to Judith Chernoff by reflecting on her warmth and humor and lauding the couple's marriage of over 75 years.

## A Sense of Humor

Throughout the centennial celebration, there were also many moments that brought Professor Chernoff's sense of humor into the foreground. A primary example was when he narrated a story about his analysis of the Massachusetts lottery. When Chernoff arrived in Boston to work at MIT, the lottery had only been in place for about 500 days. After a statistician had shown that the lottery was likely to have had some repetitions but didn't, it dawned on Chernoff that the

lottery was probably fixed. Chernoff elaborated, "When I realized that the lottery was fixed, I thought that I should announce it, but then maybe the gangsters would not like that and would assassinate me. On the other hand, when I told the area chair at MIT, Harvey Greenspan, he suggested that if it was fixed, I should take advantage of it!" Instead of "taking advantage of it," Chernoff started to analyze the lottery and generated a paper that provided advice to people on how to select winning combinations.

Chernoff's words are emblematic of his approach to life and his relationships with others. While he cares deeply about the theory and applications of statistics and has committed his career to pursuing meaningful questions, Chernoff also appreciates doses of levity and humor and relished engaging with Meng and others at the symposium. The celebration was truly a once in a century opportunity to recognize Chernoff's contributions as a researcher, educator, and mentor. We look forward to celebrating with Professor Chernoff again at his 105th birthday!

**References:**

Bather, John. *A conversation with Herman Chernoff*. Statistical Science 11(4): 335–350.

**Thank you to Professor (and PhD alum) Sally Thurston and PhD student Matthew Mallory for providing the photos!**



# Professor David Harrington

## *Marvin Zelen Leadership Award*

During a lunch at the University of Maryland with the biostatistician Marvin Zelen, David ("Dave") Harrington discussed his thesis and future goals with Zelen, who quipped, "If you don't come to Buffalo to work in my group, you won't amount to anything!" After spending 6 years on the faculty of the University of Virginia, Harrington began his career as a biostatistician by joining Zelen's group in 1984, which had moved to the Dana-Farber Cancer Institute (DFCI).

When receiving the Marvin Zelen Leadership Award in Statistical Science this spring, Harrington, an Emeritus Professor of Statistics at FAS, the Department of Biostatistics at the Harvard T.H. Chan School of Public Health (HSPH) and the DFCI, fondly recounted this first memory of Professor Zelen and added, "Marvin always provided helpful, but not prescriptive advice." An inspirational leader in the field of biostatistics, Zelen chaired HSPH's Department of Biostatistics from 1981-1990 and the DFCI Department of Data Science (formerly the Department of Biostatistics and Computational Biology) from 1977 – 1998. Founded by friends, family, and colleagues in Zelen's honor (he died in 2014), the award "recognizes an individual in government, industry, or academia, who by virtue of his/her outstanding leadership, has greatly impacted the theory and practice of statistical science" (source: Department of Biostatistics website). Harrington received this award because of his contributions to nurturing a strong statistics community through his research, teaching, and administrative roles. Previous awardees from the Department of Statistics include founding Chair of the Department Frederick Mosteller (1997), Professor Xiao-Li Meng (2014), PhD alumna Professor Nan Laird (2015), and Professor Xihong Lin (2022).

In honor of this award, we would like to share with you some of the highlights from Prof. Harrington's career, his Zelen Award talk on "The Special Relationship between Survival Analysis and Cancer Research – Successes and Persistent Problems," and from his one-on-one conversation with us.

**Research Career Highlights**

Over a span of 45 years, Harrington has been a leader in developing and applying statistical methods to medical research (source: Department of Biostatistics website). Since 1984, Harrington has been a Harvard faculty member; from 1998-2009, he chaired the Department of Biostatistics and Computational Biology (now the Department of Data Science) at the Dana-Farber Cancer Institute, and from 2012-2014, he chaired the Department of Statistics at FAS. Harrington also helped found the Dana Farber/Harvard Cancer Center's (HCC) Biostatistics research program and led the Biostatistics Core,

a resource that supports biomedical research by implementing best practices for efficient data management and analysis. He is a fellow of the American Statistical Association, the Institute of Mathematical Statistics, and an elected fellow of the International Statistical Institute.

In his research, Harrington has focused on developing statistical methods for designing and analyzing results in clinical trials for treating lymphoma, leukemia, colorectal, and lung cancer. His expertise and leadership in this area led him to be the principal investigator for the NCI (National Cancer Institute) sponsored Statistical Coordinating Center for the Cancer Care Outcomes Research and Surveillance (CanCORS) Consortium from 2001 to 2014. The CanCORS grant provided for one Statistical Coordinating Center and six Primary Data Collection and Research Sites to collect and then analyze data about the medical care and treatment outcomes of 10,000 lung and colorectal patients (source: National Cancer Institute, Division of Cancer Control & Population Sciences website).

**Personal Reflections on Research Career**

In a recent conversation with us, Harrington reflected on some of the people and experiences that have influenced his career. As a math major, Harrington had always gravitated towards quantitative work, but it was his time in the US Coast Guard that convinced him to pursue graduate school. Towards the end of his three-year term, Harrington worked with a group tasked with finding the optimal placement of rescue planes and ships on the East Coast. Probability theory was an important part of the solution. He realized that research in statistics would allow him to work on interesting and meaningful problems.

From graduate school onwards, Harrington was increasingly interested in addressing statistical problems within the biomedical field. As a PhD student at the University of Maryland, he forged a life-long connection with fellow student Thomas Fleming, currently a Professor of Biostatistics and Statistics at the University of Washington School of Public Health. Remaining in close touch with Fleming after graduate school, Harrington was intrigued by Fleming's work at Mayo Clinic on analysis of clinical trials for cancer treatment. Through this partnership, Harrington was introduced to the field of survival analysis, which he considered to be an "exciting mix of mathematical theory and compelling applications." Because of Harrington's interest in statistics in cancer research, he jumped at the opportunity to join DFCI, the premier institution for cancer research, as a faculty member in 1984, where he was exposed to the cancer research and mentorship of senior biostatistics and clinical colleagues.

**Teaching Career Highlights**

While Harrington's research focuses on cancer and public health, he also enjoys working with the next generation of statistical scientists. He redesigned and modernized Stat 102 Introduction to Statistics for Life Sciences, a course targeted towards undergraduates in the life sciences. A skilled teacher, Harrington was awarded the Nichols Award for Teaching Excellence (from HSPH), the Levenson Memorial

Teaching Prize (Harvard College), the Hoopes Prize for Senior Thesis Mentoring, and the Herman Callaert Leadership Award (Hasselt University), among others. Channeling his experience as an educator, Harrington wrote the seminal book "Counting Processes and Survival Analysis" with Tom Fleming as well as the book "Introduction to Statistics for the Life and Biomedical Sciences" with Julie Vu.

Harrington's former students testify to his positive influence on the community. At Harvard, he mentored over 14 Biostatistics PhD students and many concentrators in his role as Co-Director of Undergraduate Studies in the Department of Statistics from 2008-2015, a period of exciting growth from under ten to almost 200 concentrators. Statistics concentrator Michelle Zemplenyi (AB 2013, PhD in Biostatistics 2020) described a conversation with Harrington several years after she graduated: "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." Through such thoughtful mentorship of students, Harrington encapsulates the essence of the Zelen award.

**Personal Reflections on Teaching Career**

When asked about the most rewarding moments of his career, Harrington immediately volunteered "teaching." For Harrington, the reward of teaching does not materialize from a specific moment; rather, it comes from the whole process of working with students and developing personal connections with them – whether discussing a thesis with a graduate student or talking with an undergraduate student going through a difficult time. While no longer teaching courses, Harrington interacts with students at the John D. O'Bryant High School in Roxbury by tutoring AP statistics. He was introduced to this opportunity by Pie R Squared, a non-profit organization focused on building students' confidence and problem-solving skills through a community-oriented tutoring experience. Reflecting on what he's learned from students over the years, Harrington said, "My students have always been smarter, more creative than I am, so I've learned to listen to them. They will share great ideas with you if you give them a chance."

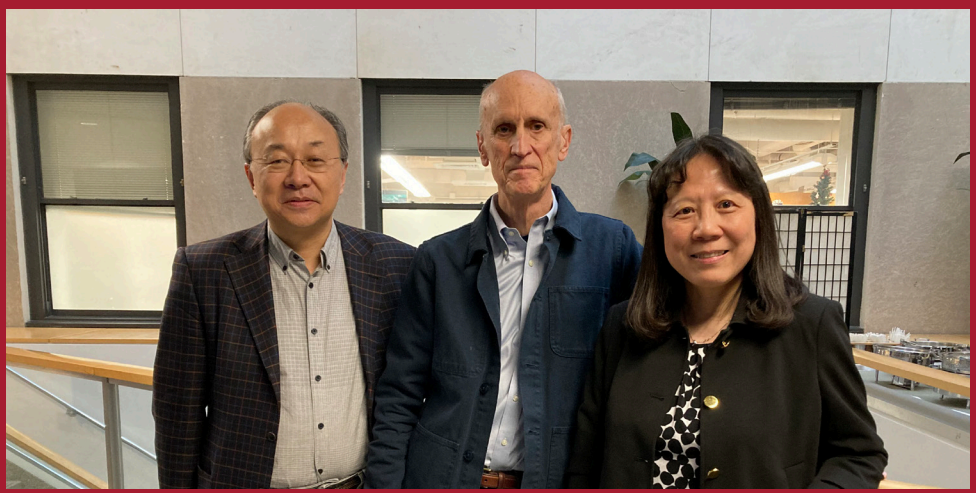
**Zelen Talk Highlights**

In her introduction to Harrington's talk, his colleague Professor Xihong Lin described him as a "visionary, thoughtful person, and strong communicator." Lin highlighted his community contributions through his support of students, postdocs, and junior faculty and through his efforts to increase the diversity of editors at the New England Journal of Medicine when he was the statistical editor from 2010-2021.

The lecture, "The Special Relationship between Survival Analysis and Cancer Research – Successes and Persistent Problems," presented the fruitful collaboration between statisticians and cancer research scientists, which led to an

increase in the cure rate in pediatric acute lymphoblastic leukemia from 10% to 90% between 1967 and 2010. Harrington traced the advances in the treatment of pediatric leukemia from the initial work of Sidney Farber to the current, biologically driven studies conducted by national and international consortia. Each stage of development has involved close collaboration between statisticians and clinical scientists, which led to many advances in survival analysis in statistics as well as cancer research. The Kaplan-Meier estimator, the log rank and Gehan-Wilcoxon test statistics, and the Cox proportional hazards model were all methods proposed and used during this period. Remarkably, the increase in cure rate was the result of a series of randomized trials in the sequencing and dosing of existing therapies – no new effective drugs for leukemia were discovered during the 43-year period.

At the close of the talk, Harrington celebrated with colleagues and former students in a reception and dinner hosted by the Biostatistics Department at HSPH. Surveying the colleagues and former students around the room, we could clearly see that Harrington's talk and his personal and professional connections to them animated their conversations. While Harrington has been recognized for his achievements in developing and analyzing statistics in the biomedical field, the Zelen Award highlights his lasting impact on the scientific community and the next generation of scientists.



# GUSH

## *Women in Statistics & Data Science Panel*

In a cozy alcove at Henrietta's Table in Harvard Square, a group of Harvard undergraduates and graduate students chatted about their classes, hobbies, and career aspirations. They were dining with speakers from a Women in Statistics and Data Science Panel hosted by the Group for Undergraduates in Statistics at Harvard (GUSH) and the Harvard College Women's Center (HCWC) as part of Women's Week 2023 at Harvard and Women's History Month. The panel brought together a wide range of speakers: Amanda Cox, former Head of Special Data Projects at USAFacts and Editor at the New York Times; Kelly McConville, Senior Lecturer in Statistics at Harvard University; Sowmya R. Rao, Senior Research and Data Analysis Scientist in the Department of Global Health at Boston University and a consultant for multiple organizations; and Kim Nguyen, Quantitative Investment Analyst at Alphataraxia Management (and a Master of Statistics 2021 alum!). We decided to revisit this pivotal event for and produced by Harvard students, Alice Wu, Ginnie Ma, Joyce Park, Michelle Qin, Sarah Cao, and Srihari Ganesh, to celebrate their achievements and share their personal career insights.

**Panelists Share Their "Statistics Journeys"**

Co-moderators Alice Wu and Joyce Park of GUSH launched the conversation by asking panelists to take the audience on their personal journeys through the field of statistics. One of the themes that emerged was that the panelists often had winding instead of direct career paths. When Nguyen started out in college, she wasn't sure exactly where she could envision herself working and in what role. As a result, she experimented as an undergraduate, trying out software engineering internships at an early stage start-up and a quant trader internship at a medium sized trading firm. Nguyen learned that she was interested in exploring other types of roles and companies, which led her to apply for a Quantitative Investment Analyst position at Alphataraxia Management, a small energy-focused quantitative investment management group with a start-up environment. An interview process that incorporated fun, challenging problem-solving, and a collaborative vibe was enough to convince Nguyen that this was the next step for her career. As Nguyen's experiences illustrate, internships and the interview process can help students decide both what they are and are not seeking in a job, thus reshaping students' plans for their future jobs.

In Cox's and Rao's career paths, they returned to school to pursue graduate programs in statistics after working for the federal reserve (Cox) and for a banking software development company (Rao). Reflecting on her decision to return to school (first a Master's in Statistics from SUNY, Buffalo, and then a PhD in Biostatistics

from Boston University) and leave behind her computer programming role, Rao said, "I pursued computer programming because my father always said that the future was in computer science, but what he didn't realize was that the future was in computer science and statistics. I also discovered that my interest in statistics lay outside of banking and finance in biostatistics because I wanted to make a difference in global health." While Cox initially considered pursuing a PhD program, she finished her Master's in Statistics at University of Washington to then follow her passion for data visualization as a graphics editor for the New York Times. The career trajectories of Cox and the other panelists demonstrate how they often don't follow a linear path; these panelists periodically reassessed their professional goals and happiness at work to inform their career decisions.

**Day in the Life of Our Panelists**

The panelists provided the audience with a window into their day-to-day lives. While they all work on tough statistics and data-science related problems that require frequent communication with colleagues, students, clients, and other stakeholders, there were also substantial variations in their schedules and how they spent their time. For example, McConville structures her mostly in person work during the semester around her meetings, teaching, advising, department administration (particularly as the Co-Director of Undergraduate Studies), and research. On the other hand, Rao, as a consultant for UNICEF, Massachusetts General Hospital, and other organizations, frequently creates her own schedule and often works remotely. As our students start navigating different jobs, it's helpful for them to consider questions, such as, what balance of work, independent vs. collaborative, administrative vs. core content area, remote vs. in-person, and scheduled vs. flexible, is most suited to me and my work style?

**Advice on Seeking Mentorship**

The panelists shared how mentorship can arise from different types of relationships: student to student, faculty to student, colleague to colleague, senior employee to junior employee, etc. "When seeking mentorship," said Cox, "Pay attention to whom you're drawn to and why, whether because of their personal qualities, professional skills and interests, or both." Cox described a mentor at the Federal Reserve who made charts in his free time, inspiring her own interest in visual representations of data. Nguyen's moment of stats inspiration arrived when she was a junior in high school. She complained to a friend in his senior year about her terrible sleep patterns, and in response, the friend coded a graph that demonstrated that less sleep correlated with the winter season. Nguyen was transfixed; this is what she wanted to do, so she asked her friend to teach her. In her freshman year of college, Nguyen took Statistics 110 (Introduction to Probability) with Professor of the Practice Joe Blitzstein. While Nguyen felt that she struggled in the course, she took advantage of Joe's office hours and advising and persevered, eventually returning to the course as a teaching fellow. Reflecting on these different mentoring experiences, Nguyen

explained, "It's important to have both mentors that push me (e.g. professors) and peer mentors whom I just feel comfortable with (and can ask all of my Gen Z questions!)."

On the topic of mentorship, Rao and McConville emphasized the importance of building and maintaining a professional network. These professional networks are not only important when searching for jobs but are also important for supporting a sense of community and intellectual growth. McConville spoke about when she met Blitzstein at an event at one of her former institutions, Swarthmore College: "Joe and I chatted about his visit as well as about his kids and Winnie the Pooh! We stayed in each other's professional network, but I had no idea that later I would become a Senior Lecturer at Harvard and would be a Co-Director of Undergraduate Studies with him." Similarly, Rao has worked on maintaining her professional relationships and described how she frequently consults her former PhD advisor and postdoc advisor on certain problems and for letters of reference. To continue to have a strong network, Rao recommended that students reach out at least once a year to professional contacts to exchange updates on their lives and careers.

**Advice for Young Data Scientists and Statisticians**

At the end of the panel, audience members asked how to prepare for the next steps in their careers and decide whether to pursue graduate school. Related to this question, McConville quipped, "When the refrain 'let epsilon be greater than zero' still rolled off my tongue, I knew grad school was in the cards!" McConville elaborated on her decision process, "Some students reach the end of their senior year and they feel burnt out or need a change of setting, but other students still enjoy the rhythm and unconventional schedule of school and are eager to continue. There isn't a right or wrong since students can also have very satisfying careers without obtaining a master's or PhD." To help determine whether grad school is the right next step, our undergraduates can take several different actions, including speaking with our master's and PhD students and faculty advisors and pursuing a thesis and other research-related projects. Cox and Nguyen both advised that students take their time making this decision because grad school, particularly a PhD program, requires substantial commitment and drive. While some students find it easier and more natural to transition directly from undergrad to grad school, other students may work outside of academia before deciding that graduate school is the right option for them.

The panelists acknowledged the importance of developing technical skills, but they also emphasized cultivating certain social skills, which can sometimes be overlooked when preparing for careers in data science and statistics. Rao and McConville revealed how communication skills have mattered to them in their careers. Practicing and refining her communication skills has helped McConville to code, collaborate, teach, and present research in an effective way for different audiences, thus contributing to her career advancement in academia. During her work as a consultant for organizations like Massachusetts General

Hospital, the VA Medical Center, UNICEF, and the World Health Organization, Rao has exercised her skill at collaborating with and communicating statistical methods and findings to a broad stakeholder audience.

Panelists also commented on the skill of self-reflection to determine strengths and what you value most in your job. For example, Rao's self-reflection early in her career led to the realization that while she enjoyed working with and analyzing data, she sought a larger purpose: "It was important to me to think beyond the data about the larger impact of my work – I wanted my work to focus on helping the health of the community rather than on simply publishing papers."

In Nguyen's case, she discovered that some of the qualities that she prized most in a job were an intellectually stimulating environment with co-workers who similarly valued learning new skills and concepts. Nguyen also shared how reflecting on the skills she had acquired as a teaching fellow helped her to develop confidence in herself as a statistician and to counter imposter syndrome, a feeling of self-doubt and lack of belonging. Through their discussion of self-reflection and communication skills, the panelists highlighted the value of nurturing less technically oriented skills to their future careers.

By the end of the event, it was clear that GUSH's Women in Statistics and Data Science Panel was a success in showcasing the varied careers and perspectives of the speakers. Despite their varied career pathways and work in different industries, the panelists' stories often reflected common themes, including the non-linear nature of career paths, the importance of developing many types of mentor-mentee relationships, and the importance of honing both technical and social skills. Surveying the room at Henrietta's Table, we could already see students soaking in the experience, applying these lessons, and building their own networks.

# Professor Xihong Lin

## *Reflects on NAS Election & Career Highlights*

On a spring afternoon in the Program in Quantitative Genomics Seminar at the T.H. Chan School of Public Health, Professor Xihong Lin felt a buzz in her pocket. Silencing the pesky phone, she was surprised to see a text saying that she was now a member of the National Academy of Sciences. The reality of this only set in when the official announcement was released at 3:00 pm and phone calls and emails started rolling in – work would have to be on pause for the day.

Professor Lin was elected as one of 120 new members of the National Academy of Sciences this year in recognition of her continuing achievements in original research focusing on developing and applying scalable statistical and machine learning methods for big genetics and health data. Her research helps identify potential drug targets and precision intervention and treatment strategies. In a conversation with Dr. Lin, she shared her reaction and reflected on her sources of inspiration, highlights from her career and work, and advice for young researchers seeking to make an impact. (Note: This interview has been revised and condensed).

**What has this honor from the National Academy of Sciences meant to you?**

**Lin:** Receiving a congratulatory text about being elected to the NAS at a seminar was a genuinely special surprise! I am deeply honored and humbled to receive this prestigious recognition.

I would like to thank the Academy and its members for this incredible honor. My heartfelt gratitude also goes out to the many collaborators, students, and postdocs whose contributions over the years have been instrumental. Their talent and creativity have been a constant source of inspiration to me. This achievement reflects what we have accomplished collectively, rather than only my individual efforts.

**What is most rewarding about your work? Which role models inspired you to become a statistician?**

**Lin:** The most fulfilling aspect of my work is the collaboration with many colleagues, students, and postdocs over years to contribute to scientific research by aiming to benefit society, improve lives, and pave the way for a better tomorrow.

When I was young, I was inspired by my grandparents, who were faculty members in preventative medicine in the US. They dedicated their careers to studying and developing a vaccine for schistosomiasis, a parasitic disease that affected many areas of the world, especially before 1960. During graduate school, my dissertation advisor, Professor Norman Breslow,

the former Chair of the Biostatistics Department at the University of Washington [Professor Breslow passed away in 2015], had a substantial impact on my development as a biostatistician. I learned a lot from Norm, not only through his remarkable statistical sense, high scientific standards, and commitment to solving pressing health problems using statistics, but also through his dedication to building a strong scientific community. In the 1970s, he was the leading statistician of the National Wilms Tumor Study, which helped improve the two-year survival rate of pediatric kidney cancer from 80% to 100%. Inspired by Norm, I find fulfillment in statistical methodological research driven by real-world problems and its applications in health science.

**What have you learned from the challenges that you faced in your career?**

**Lin:** I would say embracing risks, even if it means experiencing a touch of embarrassment, and stepping beyond one's comfort zone can pave the way for significant professional growth and new opportunities. When I joined the faculty at Harvard in 2005, I made the decision to transition into a new field where I had little prior expertise: the realm of statistical genetics and genomics. I started collaborating with Dr. David Christiani, Professor of Environmental Genetics, in the Department of Environmental Health. When I first attended his lab meetings, everything was jargon to me! To become more familiar with the field and its terms, I audited a first-year, graduate molecular biology course (I was the only full professor in the course!) during my sabbatical in 2008. While I felt like I was starting from square one and was embarrassed to ask basic questions, these efforts ultimately played a crucial role in repositioning me and opening a new door of opportunities. I began working on developing methods for analyzing rare variants in candidate gene sequencing studies well before there was sufficient whole genome sequencing data ['a process to determine all of approximately 3 billion nucleotides of an individual's DNA sequence', according to NIH's National Cancer Institute]. However, in time our intuition that the area of whole genome sequencing studies would grow proved to be correct; our methods are now being widely used by the research community to analyze large amounts of whole genome sequencing data and identify genetic variants that may cause diseases. Taking a risk to embark on a journey in a field in its infancy can yield significant long-term rewards. Having such courage is not an easy decision.

**Describe a project that has been important to you and your career.**

**Lin:** During the spring of 2020, I basically put aside all the other ongoing work to focus on how to help the world respond to COVID-19. I co-led two papers in JAMA and Nature with a former postdoc of mine, Chaolong Wang, and his colleagues at Huazhong University of Science and Technology in Wuhan on analyzing the Wuhan COVID-19 epidemic data. I had not worked on epidemic modeling before, and it was a steep learning curve. However, when a pressing global public health crisis emerged, impacting countless lives, we rose to the occasion, learning and growing together. Chaolong and I shared with Nature the story behind this

collaboration in the article "Learning from the Wuhan COVID-19 Data on the Fly." The article describes how we analyzed on-the-ground data about the spread of the virus in the early phase of the pandemic to provide recommendations for public health intervention strategies, such as mask wearing, quarantine, and isolation, before the vaccine was available. Our paper garnered global attention and helped many countries develop their public health responses. I was interviewed by many news outlets, invited to be on the Massachusetts COVID-19 Task Force, and testified in front of the UK Parliament's science and technology committee. My goal in undertaking this public outreach was to swiftly disseminate the most up-to-date knowledge to the public and policymakers across cities, states, and countries worldwide to help control outbreaks and save lives.

To support communicating about COVID-19 with the public, a student and postdoc in my lab took the lead in launching a real-time dashboard for calculating the Rt values [the rate at which one person with COVID-19 spreads the virus to other people] for different countries, states, and cities. The Rt value needed to be less than one to meet the criteria for having the spread of the virus under control in a region. Two additional postdocs in my lab took the lead on developing transmission dynamic models for analyzing more complex US COVID-19 data, and the work was published in a JASA discussion paper.

In the initial stages of the pandemic, I was deeply moved by the willingness of so many individuals to step up and lend a hand without seeking individual credit. This demonstrated an incredible sense of community, rather than competition, and was truly heartwarming to witness amidst such a challenging time.

**Describe a project that you are currently working on and would like to highlight.**

**Lin:** I am working on a project related to federated and distributed learning. Individual-level data in many healthcare systems cannot be shared due to concerns about patient privacy and federal regulations. The goal of federated learning is to develop machine learning methods using a common analysis protocol in individual healthcare systems, and then efficiently combine the site-specific results through a very small number of communications. We would like the results obtained from federated learning (with no sharing of individual-level data across different healthcare systems) to be similar to the results obtained using pooled data, while still ensuring efficient and scalable computation. The field of privacy-preserved data science is rapidly evolving and presents a lot of opportunity and promise.

**What advice would you give to young statisticians who are hoping to one day have a similar type of impact in their careers?**

**Lin:** I would tell younger researchers to identify important and promising areas at an early point in your career and to build a reputation by pinpointing a research niche with the potential for making a significant impact. It is often easier to build a reputation by getting involved in a new, understudied (but significant) area from the outset. It is beneficial to maintain an open mind,

develop independence, follow your curiosity, and demonstrate a readiness to learn on the go and from setbacks. Also, having a clear sense of a purpose and a focus will provide a compass to navigate negative feedback and distractions. Keep believing in yourself and stay positive and forward-looking! I found that having this kind of perspective was helpful both when I started working on statistical genetics and on epidemic modeling for the first time.

**Stats:** Whether taking a first-year grad course in molecular biology as a full professor or modeling infectious diseases for the first time, Professor Lin has demonstrated a zest for jumping into the unknown. By choosing to take these risks and pursue fruitful collaborations, Prof. Lin has had a profound impact on research related to statistical science and public health. A clear example of her impact was when she worked with collaborators to model the spread of COVID-19 and suggest prevention strategies. At the beginning of our conversation, Prof. Lin was not sure of the specific reason for why she was elected to the Academy, but we think that her work speaks for itself.

# Welcome Prof. Mark Sellke

## *Interview on his research, middle school math competitions & More!*

After completing his PhD at Stanford University in 2022 and a postdoc position at the Institute of Advanced Study in Princeton and at Amazon, Dr. Mark Sellke joined the Department of Statistics in July 2023 as an Assistant Professor. To introduce Prof. Sellke to the department, we conducted the following interview (revised and excerpted) in which Mark shares about his research interests, middle school math competitions, and what he's looking forward to the most this year. Welcome Prof. Sellke!

**Please introduce yourself (e.g. where you are from, your current work, a hobby that's important to you, and a fun fact about you).**

I grew up in the Midwest, attended MIT as an undergraduate, and completed my PhD at Stanford University. One of my main research interests is spin glasses, which is a high dimensional [high dimensional data occurs when the number of features in a data set is higher than the number of observations] probability topic that is also closely related to questions in statistical inference. The area of spin glasses has its roots in statistical physics. The initial motivation was to study and develop models for certain magnetic materials with unusual properties in 3-dimensional space. A fruitful avenue for making progress has been to treat the objects as high dimensional (as opposed to 3-dimensional), which makes them relevant to modern questions in machine learning and statistics. In my research, I'm also interested in other types of machine learning theory, including multi-armed bandits.

For hobbies, I enjoy racket sports. I used to play tennis a lot, and I recently started playing squash. Another favorite hobby of mine is chess, which I hear is popular around the department! A fun fact about me is that I was once featured on the social media page of a local food truck in Cambridge. They took a photo of me when I was waiting in their line on a below-freezing winter day, wearing just a t-shirt, shorts, and flip-flops (I wasn't very sensitive to cold when I was younger).

**Describe some experiences (or mentors) that sparked your interest in statistics.**

I've enjoyed probability since middle school because the problems are a lot of fun and there are many interesting paradoxes, such as Simpson's paradox and the Monty Hall problem. In these problems, you distill a complicated situation into a single number to gain some understanding about it. While in middle school, I also had a lot of fun participating in math competitions, starting with MATHCOUNTS, which was the primary math competition for middle school kids in the US.

One formative experience after college was when I spent a summer at Microsoft Research just before starting my PhD. During this time, I explored several interesting problems in statistics and machine learning, including problems on high dimensional optimization and real-time decision making. While my research has evolved since this experience, I still enjoy thinking about some of these themes in my research.

**Describe your current research and/ or teaching interests. You can discuss a particular project, if you'd like.**

To simply describe my work, I am trying to understand high dimensional optimization from the average case perspective rather than the worst-case perspective. Traditionally in optimization theory, you are trying to optimize a high dimensional function so that the algorithm works in all possible cases. From this perspective, our algorithm might have to bend over backwards to do well for an unrealistic worst-case scenario, which we can avoid by studying the average case. I am in the process of developing a general theory that identifies the best stable algorithms (meaning that if you change the input, the output of the algorithm doesn't fluctuate very much) for a wide class of such optimization problems. This spring I'm looking forward to teaching a class with a broad focus on average-case optimization. I plan to survey interesting qualitative behaviors that can appear in these types of problems.

**What made you decide to join Harvard Stats? What are you looking forward to the most for this year?**

I am excited to interact and collaborate with the brilliant students and colleagues in the department and campus-wide. Everyone has been very friendly and welcoming!

# Welcome Dr. James Xenakis

## *Interview on his research, reading American Fiction & More!*

The Department of Statistics would like to welcome Dr. James Xenakis, Lecturer in Statistics, to our teaching faculty! Dr. Xenakis joined the department in August 2023 after completing a postdoctoral fellowship at the University of North Carolina's Gillings School of Global Public Health, where he had previously completed a PhD in Biostatistics. In the edited and excerpted interview below, Dr. Xenakis spoke with us about his statistics projects in a mouse lab, reasons for joining the department, and favorite hobby (reading American fiction!).

**Please introduce yourself (e.g. where you are from, your current work, a hobby that's important to you, and a fun fact about you).**

**Xenakis:** I grew up not very far from Harvard in Concord, Massachusetts and completed my undergraduate degree at that other school in Cambridge (MIT). After college, I worked for about six years at a small pharmaceutical consulting firm, where I specialized in cost-effectiveness modeling of drugs and medical devices.

I've spent the last decade or so at the University of North Carolina (UNC) at Chapel Hill. After completing my PhD in Biostatistics there, I took a postdoc position in the Genetics department, where I was the sole statistician in a mouse lab. The lab specializes in breeding inbred strains of mice to help elucidate the genetic underpinnings of diseases. One of my main projects over the last couple years was using mouse resources we bred to study arsenic-associated type-2 diabetes. Hundreds of millions of people are exposed to unsafe levels of arsenic in their drinking water and, as a result, may also be at elevated risk of developing type-2 diabetes.

My main hobby that's stuck with me as I've gotten older is reading; no matter how busy or tired I am, I always do a little bit of reading every night. I've amassed a large collection of books, and figuring out where to put them all was a particular challenge in moving from my house in North Carolina to an apartment in Boston. My favorite genre is American fiction, and there's one book that I read almost every year: A *Confederacy of Dunces* by John Kennedy Toole. This book has an interesting history, too; it didn't get published until about twenty years after the author's death, and only due to the persistence of his mother.

A fun fact about me is that I'm a new uncle. My niece was born in April and lives in Massachusetts, which is an extra benefit because now I can look forward to watching her grow up in person rather than on Zoom (even in the past several months, she has changed so much)!

**Describe some experiences (or mentors) that sparked your interest in statistics.**

**Xenakis:** Statistics appeals to me because I see it as the glue that holds a research project together. For example, in my lab, when a biologist had a research question, the statistician would relate that question to parameters in a statistical model. Then, the technicians on the project would collect the data, and statisticians would become involved again to properly clean, merge, and analyze the data. Since statistics brings together all aspects of the project, a statistician is a natural point of contact for multiple team members with different areas of expertise. The opportunity for collaboration between statisticians and domain scientists always makes the work quite interesting and rewarding.

I had two mentors at my first job out of college at a pharmaceutical consulting firm who sparked my interest in the field and influenced my decision to pursue stats as a career. At the firm, I did cost-effectiveness modeling to support regulatory decisions. I built interfaces that allowed clients to run simulations that were essentially extrapolations and sensitivity analyses (e.g., looking at different model time horizons or populations) based on risk equations developed from their clinical trials. When the lead statistician at the company saw that I was drawn to the equations behind the models, he took me under his wing. He was very talented, trilingual, and probably one of the most prolific readers I've ever met. He was a positive role model in that he was incredibly good at statistics but was also an interesting, polymathic type of person.

Another important mentor was my boss and the founder of the company. While my boss was a medical doctor by training, he was also passionate about and quite good at statistics, programming, and economic modeling. To a fault, he would get deeply involved in projects, even at the level of coding (which he probably didn't have the time for, but he enjoyed it and was skilled at it). From his example, I learned that one of the most important things that statisticians do is to bring statistical thinking, even basic statistical thinking, to a field. This perspective aligns with the ethos of the department here at Harvard.

**Describe your current research and/or teaching interests. You can discuss a particular project, if you'd like.**

**Xenakis:** As I mentioned earlier, I was the sole statistician in a mouse lab that over the past 20 years has bred approximately 65 new inbred mouse strains. The main highlight of my time in the lab was the collaborative nature of the work; they taught me more science and I taught them some statistics (I hope). My main project in this lab was part of the Superfund Research Program at UNC, which is funded by the NIH to find scientific solutions to protect people's health, the environment, and communities. I recently helped the lab use these engineered mouse strains to study the genetics of arsenic associated type-2 diabetes.

Working in this lab also sparked an interest in data management, which will be particularly relevant to my teaching this year. I quickly learned that the data generated in a mouse

lab are particularly messy, partly because they come from living animals. On top of this, many scientists and technicians use Excel as if it were simply an electronic notebook, instead of thinking about best practices for recording experimental data. My experiences in this lab really helped me understand the challenges scientists face in managing data, and the importance of bringing statistical expertise to a project. It might seem like a humble aspect of our jobs, but you can't analyze data that isn't organized and cleaned! I'm interested in addressing the topic of data management in my courses because, while statisticians end up devoting a lot of time to this, it's also a topic that is often overlooked in our coursework.

**What made you decide to join Harvard Stats? What are you looking forward to the most for this year?**

**Xenakis:** Hands down, it was the people I met here in the spring that made me decide to join the department. Every time I have an important decision to make, I try to approach it both from an intellectual and an emotional perspective. Often, I find myself having to weigh the two sides against each other. Obviously, Harvard is the "right" choice - this was a unique opportunity to work at such a prestigious institution. But on top of that, being in the department just felt right. Having spent time in many different departments, this one felt uncommonly warm and inviting, with kind and engaging faculty, students, and staff. Fortunately, my logical and emotional sides agreed that joining the department was the best decision to make! In particular, I'm looking forward to learning from the incredibly talented and more experienced faculty this year.

# Welcoming G1 PhD Students

## Nic Fishman

**Previous Institution:**
Oxford University (master's)

**Research Interests:**
(non-)convex optimization, machine learning for causal inference, geometric deep learning for physics and biology

**Hobbies:**
running, table top games of various sorts, and reading

## Anvit Garg

**Previous Institution:**
Indian Statistical Institute (master's)

**Research Interests:**
statistical ML, high dimensional statistics

**Hobbies:**
cooking, ping-pong, hiking, programming, chess, and video games

## Yuanchuan Guo

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

**Research Interests:**
statistical machine learning, high dimensions, causal inference, and computational statistics

**Hobbies:**
swimming, hiking, climbing, and listening to music

## Benedikt Koch

**Previous Institution:**
University of Oxford (master's)

**Research Interests:**
causal inference, statistical machine learning, high-dimensional statistics

**Hobbies:**
ski tours, running, classical dance, philosophy, and history

## Somak Laha

**Previous Institution:**
Indian Statistical Institute, Kolkata (master's)

**Research Interests:**
probability theory, stochastic processes, non-parametric methods, machine learning techniques, neural networks

**Hobbies:**
cooking and video games

## Sarah McDonald

**Previous Institution:**
Stanford Graduate School of Business (Predoctoral Research Fellow)

**Research Interests:**
causal inference, experimental design, reinforcement learning, statistical methods for the social and health sciences

**Hobbies**: going to concerts, reading, playing piano, hiking, thrifting

## Wenqi Shi

**Previous Institution:**
Tsinghua University (undergrad)

**Research Interests:**
causality, Bayesian statistics, statistical machine learning

**Hobbies:**
dancing (dancesport and modern dance)

# Awards, Appointments & Honors

## NEW FACULTY

**Mark Sellke, Assistant Professor:** Dr. Sellke completed his PhD in the Stanford Math Department and was a postdoc at the Institute of Advanced Studies at Princeton. His research interests include spin glasses, machine learning theory, and high-dimensional statistics.

**Anna Seigal, Affiliate Faculty, Assistant Professor of Applied Mathematics at SEAS:** Dr. Seigal was 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 began her Assistant Professor position in Applied Math at Harvard SEAS. Her research interests include algebraic statistics, tensors and multilinear algebra, and applied algebraic geometry.

**James Xenakis, Lecturer:** Dr. Xenakis completed a postdoctoral fellowship at the University of North Carolina's Gillings School of Global Public Health, where he had previously completed a PhD in Biostatistics. His research interests are in statistics pedagogy, statistical genetics, and compositional data analysis.

## NEW POSTDOCTORAL FELLOWS (starting summer & fall 2023)

**Daiqi Gao** has joined Dr. Susan Murphy's research group. Dr. Gao's research interes are in statistical reinforcement learning, machine learning, and personalized medicine.

**Asim Gazi** has joined Dr. Susan Murphy's research group. Dr. Gazi's research interests are in digital health, just-in-time adaptive interventions, sensor informatics, psychophysiology, and multitask learning.

**Melody Huang** has joined Dr. Kosuke Imai's research group. Dr. Huang's research interests are in developing robust statistical methods to credibly estimate causal effects under real-world complications.

**Yichao Li** has joined Dr. Jun S. Liu's research group. Dr. Li's research interests are in Bayesian analysis, computational statistics, especially SMC, and network analysis.

**Shahriar Talebi** works with Dr. Lucas Janson's research group. His research interests are in learning in control and games, applied differential geometry, and reinforcement learning.

**Jiajun Tang** has joined Dr. Tracy Ke's research group. Dr. Tang's research interests are in social network data, nonparametric statistics, and statistics theory.

**Ziping Xu** has joined Susan Murphy's research group. Dr. Xu's research interests are in machine learning theory, especially on reinforcement learning (RL), transfer learning, and multitask learning.

## FACULTY AWARDS & APPOINTMENTS

**Morgane Austern** was invited to be a panelist at the National Academies of Science, Medicine and Engineering on the Mathematical Foundations of AI. Dr. Austern was also selected by the National Academy of Sciences to be a 2022 Kavli Fellow. Kavli Fellows represent "the brightest young scientists from industry, academia, and government" and are invited to a symposia.

**Joe Blitzstein** was elected to a divisional seat for Science and Engineering on the FAS Faculty Council for a term of three years, starting in the fall of 2023.

**Herman Chernoff**, Professor Emeritus, celebrated with the department his 100th birthday and was honored for his teaching and research legacy of over 70 years.

**David Harrington,** Professor Emeritus, received the 2023 Marvin Zelen Leadership Award in Statistical Science for his contributions to nurturing a strong statistics community through his research, teaching, and administrative roles.

**Lucas Janson** was promoted to associate professor in spring 2023 and received the 2023 Bernoulli Society New Researcher Award. According to the Bernoulli Society's website, the award "recognizes the work of outstanding young researchers in the fields of Mathematical Statistics and Probability."

**Tracy Ke** was promoted to associate professor in spring 2023 and was selected to be a 2023 Alfred P. Sloan Research Fellow in Mathematics for the next two years. The fellowship recognizes early-career scholars whose "achievements and potential place them among the next generation of scientific leaders in the U.S. and Canada."

**Xihong Lin** was elected as a 2023 member of the National Academy of Sciences in recognition of her continuing achievements in original research and received a grant renewal of the National Cancer Institute's Outstanding Investigator Award.

**Kelly McConville** was the Co-Program Chair of Cal Poly-San Luis Obispo for the United States Conference on Teaching Statistics (USCOTS) in May/June 2023, with 321 educators in attendance. In addition, Dr. McConville received the Alpha-Iota Prize for Excellence in Teaching from the undergraduate members of Phi Beta Kappa at Harvard.

**Xiao-Li Meng**, founding Editor-in-Chief for Harvard Data Science Review (HDSR), was reappointed for another 5 years (2023-2028). Launched on July 2, 2019, HDSR is a diamond open access platform that features "Everything Data Science and Data Science for Everyone." HDSR's achievements include receiving the Best New Journal in Science, Technology, and Medicine Award by the Association of American Publishers in 2021.

**Susan Murphy** was selected by the ASA Boston Chapter as the recipient of the 2024 Mosteller Statistician of the Year Award. In addition, Dr. Murphy delivered one of the six invited talks at

NeurIPS 2023 in New Orleans, a conference with approximately 16,000 attendees. Prof. Murphy's talk, titled "Online Reinforcement Learning in Digital Health Interventions," focused on solutions to some of the challenges in developing online RL algorithms for use in digital health interventions targeting patients struggling with health problems such as substance misuse, hypertension, and bone marrow transplantation.

**Subhabrata Sen** received the National Science Foundation's (NSF) Faculty Early Career Development Program (CAREER) Award, one of the NSF's most prestigious awards, for his project on "Statistical Inference in High Dimensions using Variational Approximations."

**Pragya Sur** is the President of the New Researchers Group (NRG) at the Institute of Mathematical Statistics (IMS) and was chosen to be an International Strategy Forum (ISF) 2023 Asia Fellow by Schmidt Futures, a philanthropic initiative founded by Eric and Wendy Schmidt.

## CONCENTRATOR AWARDS

**Dinan Hamdi Elsyad** received the best video presentation at the 2023 Electronic Undergraduate Statistics Research Conference. Elsyad received the award for her video titled, "Small Area Estimation," which describes her summer research project to evaluate current estimators being used to estimate tree biomass across the US and proposes a new Zero Inflation Estimator.

**Rachel Li** was awarded the 2023 Department of Statistics Senior Concentrator Prize, which is given annually to the graduating senior concentrator who has the best overall performance (as indicated by coursework results and thesis) and who has contributed significantly to the department.

### Hoopes Prizes:

The following students received the award for their outstanding scholarly work and theses:

- **Jason Zhou** ('23 AB/AM) for his project entitled, "Data Quality Always Matters: An Analysis of Data Quality for Finite Population Z-Estimator," supervised and nominated by Professor Xiao-Li Meng.
- **James Celi Kitch** ('23 AB) for his project entitled, "Taming the Beast of Phylogenetic Trees: Learning Evolutionary Patterns of Mutation with Volume-Regularized NMF," supervised and nominated by Professor Giovanni Parmigiani and Professor Shamil R. Sunyaev.

### Sophia Freund Prize:

The award is given annually to the highest ranking undergraduate as determined at the final degree meeting of the Faculty. The following 2023 recipients were statistics concentrators:

- **Kavya Mehul Shah ('23 AB)**
- **Jason Zhou**

### Marshall Scholarship:

There were 40 recipients from U.S. colleges to pursue graduate studies in the United Kingdom for 2 years. The following recipient was a statistics concentrator:

- **kavya Mehul Shah**

## MASTERS STUDENT AWARDS

The Department of Statistics Concurrent Masters Prize will be awarded annually to the graduating student having completed the Concurrent Masters program in Statistics who has the best overall performance (as indicated by coursework results), who has demonstrated achievements in Statistics outside of coursework, and who has contributed significantly to the department. The following students received the inaugural 2023 award:

- **Virginia (Ginnie) Linqian Ma** ('23 AB in Mathematics and AM in Statistics)
- **Jason Zhou**

## PHD STUDENT AWARDS

**Yicong Jiang** received the 2022-2023 Dempster Award for his paper co-authored with Professor Tracy Ke, "Semi-Supervised Community Detection via Structural Similarity Metrics." The Dempster Prize is named in honor of Emeritus Professor Arthur P. Dempster.

**Tianle Liu** won a 2023 New England Statistics Symposium (NESS) Student Research Award on his paper co-authored with Prof. Morgane Austern, "Wasserstein p Bounds in the Central Limit Theorem under Weak Dependence." His paper “Towards Understanding the Dynamics of Gaussian-Stein Variational Gradient Descent” with Prof. Natesh Pillai was accepted as a poster for NeurIPS (Conference on Neural Information Processing Systems) 2023.

**Haoyu Ye** received a 2023 NESS Student Poster Award for her poster "Poisson Limit Theorems for Structured Random Objects," supervised by Prof. Morgane Austern.

**Yi Zhang** won the student paper competition sponsored by the Social Statistics Section (SSS), Government Statistics Section (GSS), and Survey Research Methods Section (SRMS) of the American Statistical Association. She presented her winning paper, "Safe Policy Learning under Regression Discontinuity Designs," at the Joint Statistical Meeting in August 2023 in Toronto ON, Canada.

## ALUMNI AWARDS

**Nan Laird**, PhD alumna (’75), was awarded the 2023 Chernoff Excellence in Statistics Award by NESS. Laird, Harvey V. Fineberg Professor of Public Health, Emerita, in Harvard T.H. Chan School of Public Health’s Department of Biostatistics.

# Celebrating PhD Graduates

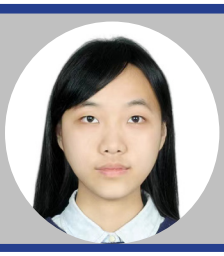
## Jonathan Che

**Dissertation Title:**
"Game, Site, Match: Topics in Causal Inference and Sports Statistics"

**Dissertation Committee:**
Luke Miratrix, Mark Glickman, José Zubizarreta

**What's Next:**
I'll be starting a position with Exponent in the new year.

**Favorite Stats Memory:**
meeting everyone as a G1, holiday social hours, graduating!



## Dieyi Chen

**Dissertation Title:**
"Clustering of Single-Cell and Text Data"

**Dissertation Committee:**
Tracy Ke, Xihong Lin, Subhabrata Sen

**Favorite Stats Memory:**
department retreat, holiday party, women in stats event!

## Cory McCartan

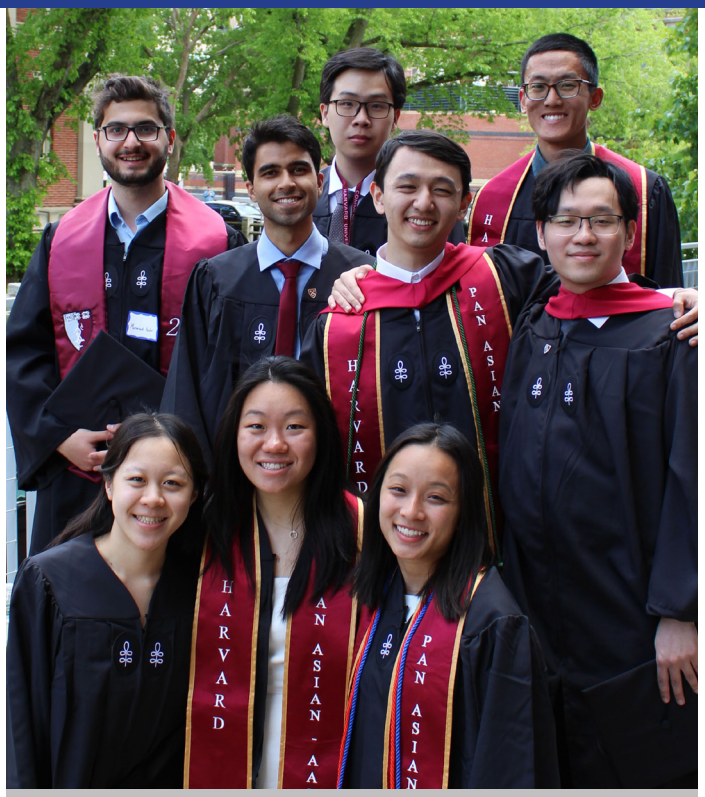
**Dissertation Title:**
"Computational and Bayesian Methods for Geographic Data in the Social Sciences"

**Dissertation Committee:**
Kosuke Imai, Xiao-Li Meng, Gary King

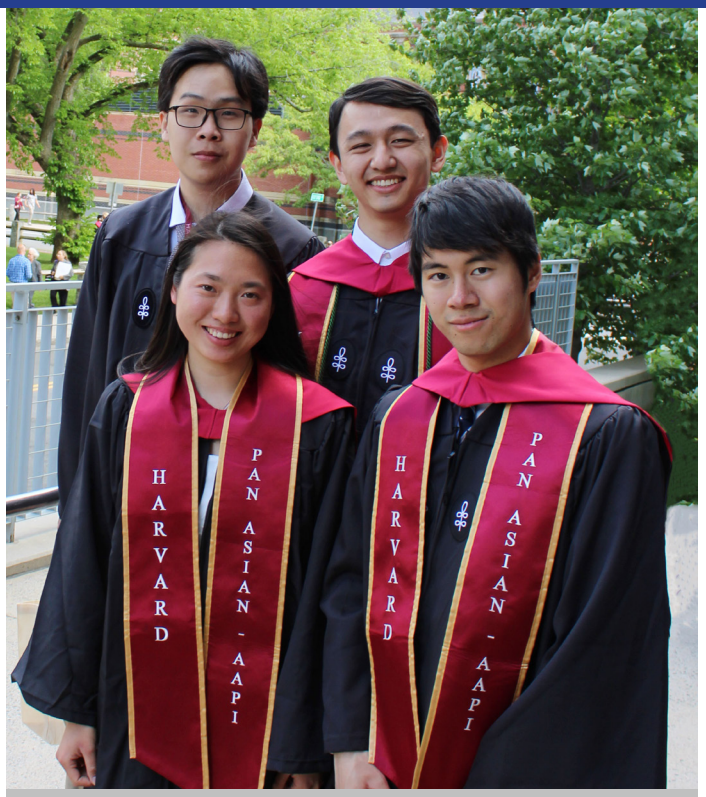
**What's Next:**
postdoc at NYU, then assistant prof. at Penn State

**Favorite Stats Memory:**
retreat days: spending time with students & trying to absorb as much as I can when faculty condense their research into 5 min.

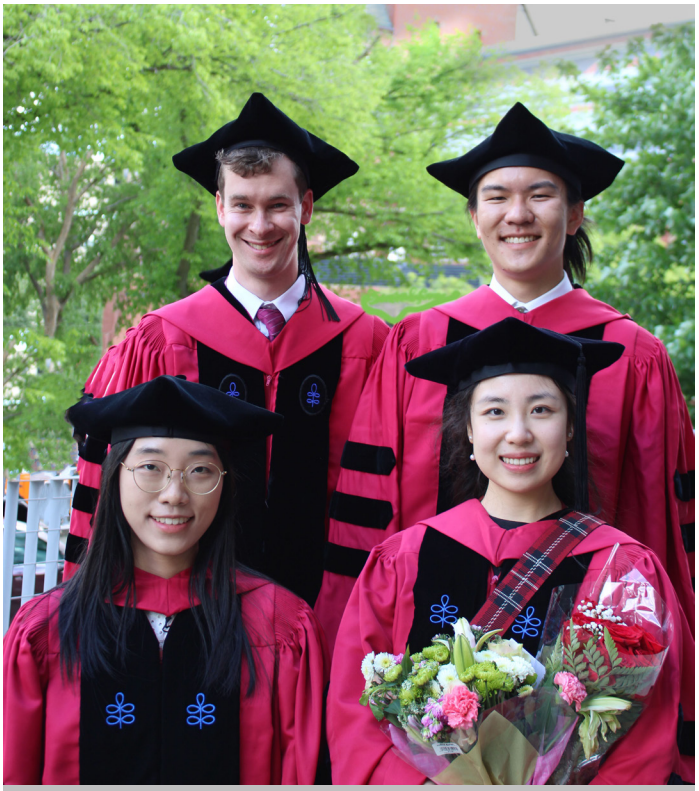
# Commencement Celebration - May 25, 2023




Bachelor of Arts Graduates




Master of Arts Graduates



PhD Graduates

PhD Graduates with Faculty

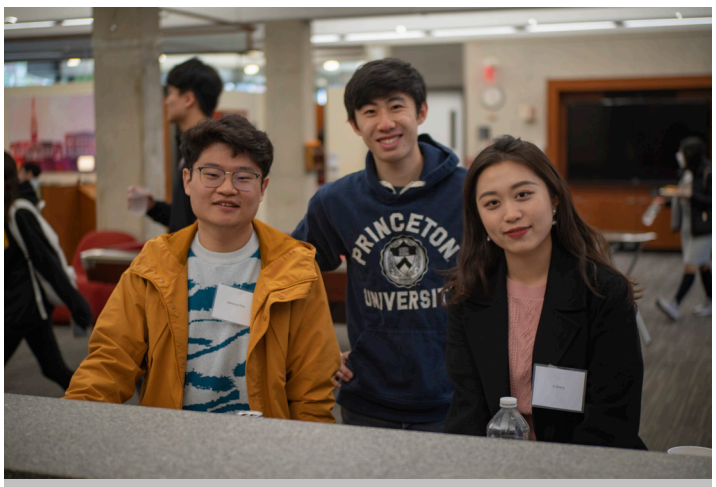
**Thank you to Lorna Blocksma for the Commencement photos!**


2023

# Community Events


GUSH & Grad Council Social

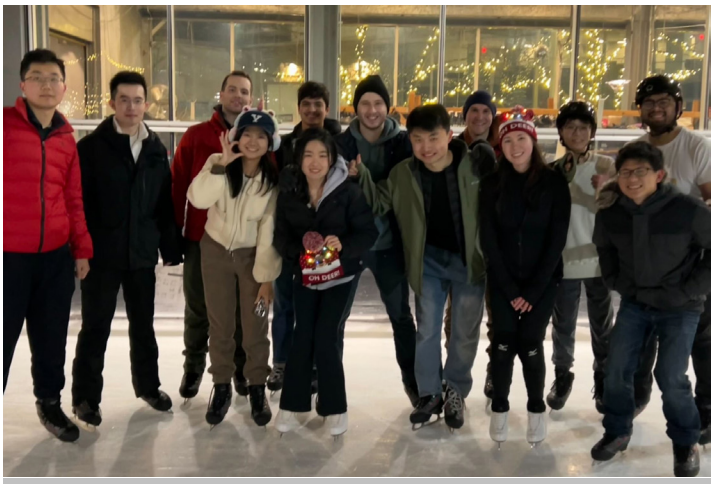
End of Semester ggplot Party




Herman Chernoff Celebration

Data Adventure Day

Holiday Party



Grad Council Ice Skating Trip

# 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:**