Yiqun T. Chen

Website: https://yiqunchen.github.io/
Google scholar: https://tinyurl.com/yiqun-scholar
Email: yiqunc@uw.edu

EDUCATION

Ph.D. Candidate in Biostatistics — GPA: 3.9/4.0
Fall 2017 – Spring 2022 (Expected) — University of Washington, Seattle
Dissertation Committee: Dr. Daniela Witten (Chair), Dr. Alex Luedtke, Dr. Amy Willis, Dr. Jamie Morgenstern.
Research Interest: Topics in selective inference: uncertainty quantification in statistical learning.
Relevant Coursework: Theoretical statistics, causal inference, deep learning, interactive learning, fairness in machine learning, computing ethics, convex optimization.

B.A. in Statistics, Computer Science, and B.S. in Chemical Biology — GPA: 3.9/4.0
Fall 2013 – Summer 2017 — University of California, Berkeley
High Distinction in General Scholarship

PUBLICATIONS AND PREPRINTS

1. Marquez C*, **Chen YT***, Atukunda M, Chamie G, Balzer LB, ..., Charlebois ED, Havlir DV, Petersen ML (2022+). The Association Between Social Network Characteristics and TB Infection Among Adults in Nine Rural Ugandan Communities. *Submitted to Journal of Infectious Diseases*; joint first authorship.
2. **Chen YT**, Smith AD, Reinecke K, To A (2022). Collecting and Reporting Race and Ethnicity Data in HCI. *To appear in CHI'22 Extended Abstracts*
3. **Chen YT**, Jewell SW, and Witten DM (2022+). More powerful selective inference for the graph fused lasso. arXiv preprint. arXiv link: `https://arxiv.org/abs/2109.10451`. [Revision submitted to *Journal of Computational and Graphical Statistics*]
4. **Chen YT**, Jewell SW, and Witten DM (2021). Quantifying uncertainty in spikes estimated from calcium imaging data. To appear in *Biostatistics*. arXiv link: `https://arxiv.org/abs/2103.07818`.
5. **Chen YT**, Brown LB, Chamie G, Kwarisiima D, Ayieko J, Kabami J, Charlebois E, Clark T, Kamya M, Havlir DV, Petersen ML, and Balzer LB (2021). Social networks and HIV care outcomes in rural Kenya and Uganda. *Epidemiology*, 32(4):551-559.
6. **Chen YT**, Gopinath R, Tadakamalla A, Ernst MD, Holmes R, Fraser G, Ammann P, Just R. Revisiting the relationship between fault detection, test adequacy criteria, and test set size. In: *2020 35th IEEE/ACM International Conference on Automated Software Engineering (ASE).* 2020:237-249.
7. Brown L, Balzer L, Kabami J, Kwarisiima D, Sang N, Ayieko J, **Chen Y**, Chamie G, Charlebois E, Camlin C, Cohen C, Bukusi E, Kamya MR, Moody J, Havlir D, Petersen M (2020). The influence of social networks on antiretroviral therapy initiation among HIV-infected antiretroviral therapy-naive youth in rural Kenya and Uganda. *J Acquir Immune Defic Syndr.* 83(1):9-15.
8. **Chen Y**, Zheng W, Brown LB, Chamie G, Kwarisiima D, Kabami J, Clark TD, Sang N, Ayieko J, Charlebois ED, Jain V, Balzer L, Kamya MR, Havlir D, Petersen M, the SEARCH Collaboration. Semi-supervised record linkage for

construction of large-scale sociocentric Networks in resource-limited settings: an application to the SEARCH study in rural Uganda and Kenya. arXiv preprint. arXiv link: `http://arxiv.org/abs/1908.09059`.

9. Jakobson C, **Chen Y**, Slininger M, Valdivia E, Kim E, Tullman-Ercek D (2016). Tuning the catalytic activity of subcellular nanoreactors. J Mol Biol. 428(15):2989-2996.

WORK IN PROGRESS

1. **Chen YT**, Witten DM (2022+). Selective inference for k-means clustering.

SELECTED TALKS

1. (August 2022; invited) "Selective inference for k-means clustering" at COMPSTAT 2022, Bologna, Italy.
2. (May 2022; contributed) "Collecting and Reporting Race and Ethnicity Data in HCI" at CHI 2022, New Orleans, LA, USA.
3. (April 2022; invited) "Selective inference for k-means clustering" at Young Data Science Researcher Seminar at ETH Zürich, virtual.
4. (April 2022; invited) "Selective inference for k-means clustering" at International Seminar on Selective Inference, virtual.
5. (December 2021; contributed) "Selective inference for k-means clustering" at UW Biostatistics student seminar, and at SLAB LAB group meeting, Seattle, WA, USA.
6. (June 2021; contributed) "Quantifying uncertainty in spikes estimated from calcium imaging data" at the 2021 WNAR Annual Meeting of International Biometric Conferences, Anchorage, AK, USA (virtual due to the COVID-19 pandemic).
7. (September 2020; contributed) "Revisiting the relationship between fault detection, test adequacy criteria, and test set size" at the 2020 35th IEEE/ACM International Conference on Automated Software Engineering (ASE), Melbourne, Australia (virtual due to the COVID-19 pandemic).
8. (May 2020; invited) "Social networks and HIV care outcomes in rural Kenya and Uganda" at the UCSF social network and HIV workshop, San Fransisco, CA, USA, (pending rescheduling due to the COVID-19 pandemic).
9. (March 2020; contributed) "HIV+ persons in rural Uganda with fewer social connections have lower HIV suppression" at the 2020 Conference on Retroviruses and Opportunistic Infections (CROI), Boston, MA, USA, (virtual due to the COVID-19 pandemic).

SELECTED SOFTWARE

1. `KmeansInference`:
   - An `R` package for testing for a difference in means between groups identified via k-means clustering, available here.
   - Tutorials available at `https://yiqunchen.github.io/KmeansInference/`.
2. `GFLassoInference`: test for equality in means between a pair of connected components estimated from the graph fused lasso.
   - An `R` package available at here.
   - Tutorials available at `https://yiqunchen.github.io/GFLassoInference/`.
3. `SpikeInference`:

- An `R` (which serves as a wrapper for `c++` code) package that quantifies uncertainty for spikes estimated from calcium imaging data, available here.
- Tutorials available at `https://yiqunchen.github.io/SpikeInference/`.

HONORS & AWARDS

- Finalist for Data Science Postdoctoral Fellowship at Stanford Spring 2022
- Biostatistics Retention and Success Scholarship Spring 2022
- Best Student Oral Presentation at WNAR 2021, for "Quantifying uncertainty in spikes estimated from calcium imaging data" Spring 2021
- Scholarship for 6th Seattle Symposium in Biostatistics Fall 2020
- New Investigator Scholarship for CROI 2020 Spring 2020
- Dean's List (Awarded to top 4% Students) Fall 2013 – Spring 2017
- Percy Lionel Davis Award for Excellence in Scholarship in Mathematics Spring 2017
- Senior Research Award, College of Chemistry Spring 2017
- Best Poster Presentation, Institute of International Studies Spring 2017
- Scholarship for Research Merit, Institute of International Studies Fall 2016
- Scholarship for Academic Excellence, Berkeley International Office Fall 2014
- Elected to join Phi Beta Kappa Fall 2014

SELECTED TEACHING EXPERIENCE

Teaching Assistant at University of Washington

- Categorical Data Analysis (BIOSTAT 536; graduate-level)
- Introductory Laboratory Based Biostatistics (UCONJ 510; graduate-level)
- Machine Learning for Biomedical and Public Health Data (BIOSTAT 546; graduate-level)
- Introductory Laboratory Based Biostatistics (UCONJ 510; graduate-level)

Mentor for Directed Reading Program at University of Washington

- Mentored undergraduate students on the topic of identification in missing data and causal inference.
- Student presentation can be found at `https://spa-drp.github.io/writeups/win2021/suh_slides.pdf`.

Guest Lectures at University of Washington

- Machine Learning for Biomedical and Public Health Data (BIOSTAT 546). Guest lectures on decision trees, support vector machines, and principal component analysis.

Teaching Assistant at University of California, Berkeley

- Introduction to Machine Learning (CS 189/289A; advanced undergraduate-level; Fall 2016 & Spring 2017)
- Discrete Mathematics and Probability (CS 70; undergraduate-level; Summer 2016 & Spring 2017)

SERVICE

Internal

- Student representative on the admission committee Spring 2022
- Student organizer of the general exam information session Spring 2022
- Panelist for the School of Public Health New Student Welcome Fall 2021
- Student representative on the curriculum committee Fall 2020 – Spring 2021
- Peer Mentor for Biostatistics Peer Mentoring Program Fall 2018 – Present
- Panelist for the Department Internship Workshop Fall 2018, Fall 2021
- Member of the SAUCE student committee Fall 2016

External

- Reviewer for *Biostatistics*, *Bioinformatics*, *Journal of Machine Learning Research*, CHIL 2022

SELECTED INDUSTRIAL EXPERIENCE

Data Scientist Intern June 2019 – September 2019
Waymo LLC (formely Google self-driving car project), Mountain View, CA

- Project: Modeling self-driving vehicles planning and reaction time using tools from causal inference and machine learning.

Applied Scientist Intern June 2018 – September 2018
A9.com (now Amazon Search), Palo Alto, CA

- Project: Using deep learning based language models for better summaries of the search queries on `amazon.com`.

TECHNICAL SKILLS

Programming language: `Python`, `R`, `c++`, *SQL, bash, Java*