

Lucy L. Gao

Assistant Professor, 2020-current

Department of Statistics and Actuarial Science
University of Waterloo

lucy.gao@uwaterloo.ca

<http://lucylgao.com>

Education

PhD in Biostatistics, 2020

University of Washington
Advisor: [Dr. Daniela Witten](#)

BSc in Combined Mathematics & Statistics, 2015

University of Victoria

Publications and Preprints

Published or Accepted

1. **Gao, L.L.**, Witten, D. and Bien, J. (2021+) Testing for association in multi-view network data, to appear in *Biometrics*. [ASA Statistical Learning and Data Science Section Student Paper Award.]
2. Liu, P., **Gao, L.L.** and Zhou, J. (2021+) R-optimal designs for multi-response regression models with multi-factors, to appear in *Communications in Statistics - Theory and Methods*.
3. **Gao, L.L.**, Bien, J. and Witten, D. (2020) Are clusterings of multiple data views independent? *Biostatistics*, 21(4), 692-708. [ASA Biometrics Section Student Paper Award.]
4. **Gao, L.L.** and Zhou, J. (2020) Minimax D-optimal designs for multivariate regression models with multi-factors. *Journal of Statistical Planning and Inference*, 209, 160-173.
5. Hsu, E. K., Shaffer, M. L., **Gao, L.**, Sonnenday, C., Volk, M. L., Bucuvalas, J. and Lai, J. C. (2017) Analysis of liver offers to pediatric candidates on the transplant wait list. *Gastroenterology*, 153(4), 988-995.
6. **Gao, L. L.** and Zhou, J. (2017) D-optimal designs based on the second-order least squares estimator. *Statistical Papers*, 58(1), 77-94.
7. **Gao, L. L.** and Zhou, J. (2014) New optimal design criteria for regression models with asymmetric errors. *Journal of Statistical Planning and Inference*, 149, 140-151.

Preprints

1. Neufeld, A., **Gao, L.L.** and Witten, D. (2021+) Tree-values: selective inference for regression trees, submitted to *Journal of Machine Learning Research*. <https://arxiv.org/abs/2106.07816>.

2. Gao, L.L., Bien, J., and Witten, D. (2021+) Selective inference for hierarchical clustering, invited to resubmit by *Journal of the American Statistical Association: Theory and Methods*.
<https://arxiv.org/abs/2012.02936>.
-

Software

multiviewtest: Hypothesis Tests for Association Between Subgroups in Two Data Views

- R-Package, hosted on [CRAN](#).

clusterpval: Inference for Estimated Clusters in R

- R-Package, hosted on [Github](#).
 - Tutorials available at <http://www.lucylgao.com/clusterpval/index.html>.
-

Presentations

Invited

1. (October 2021) “Valid hypothesis testing after clustering” for the Department of Astronomy and Astrophysics at the University of Toronto’s Statistics and Machine Learning Journal Club, virtual.
2. (October 2021) “Selective inference on trees” for the HEC Montreal Department of Decision Sciences Seminar, virtual.
3. (September 2021) “Selective inference on trees” for the University of Toronto Department of Statistical Sciences Seminar, virtual.
4. (September 2021) “Valid inference after hierarchical clustering” for ICSA Applied Statistics Symposium 2021, virtual.
5. (July 2021) “Valid differential analysis for groups defined via clustering” for JOBIM2021, virtual.
6. (May 2021) “Valid inference after hierarchical clustering” for the Harvard-BCH Machine Learning and AI Journal Club, virtual.
7. (March 2021) “Hypothesis testing after hierarchical clustering” for the UCSF Department of Epidemiology and Biostatistics Seminar, virtual.
8. (November 2020) “Beyond sample-splitting: valid inference while ‘double-dipping’” for the University of Waterloo Student Seminar series, virtual.
9. (August 2020) “Selective inference for hierarchical clustering” for the International Seminar on Selective Inference, virtual.
10. (June 2020) “Statistical inference for multi-view clustering” for the Young Data Science Researcher Seminar at ETH Zurich, virtual.
11. (January 2020) “Statistical inference for multi-view clustering” for the University of Waterloo Department of Statistics and Actuarial Science Seminar, in Waterloo, O.N., Canada.
12. (December 2019) “Statistical inference for multi-view clustering” for the McGill University Department of Mathematics and Statistics Seminar, in Montreal, Q.C., Canada.

13. (September 2018) "Are clusterings of multiple data views independent?" for the University of Victoria Department of Statistics department seminar series, in Victoria, B.C., Canada.
14. (July 2018) "Are clusterings of multiple data views independent?" at the 2018 Joint Statistical Meetings, in Vancouver, B.C., Canada.

Contributed

1. (July 2020) "Testing for association in multi-view network data" at the 2020 Joint Statistical Meetings, virtual.
2. (June 2020) "Testing for association in multi-view network data" at the 2020 WNAR Annual Meeting of IBC (International Biometric Conferences), virtual.
3. (August 2019) "Are clusterings of multiple data views independent?" at the 2019 Joint Statistical Meetings, in Denver, Colorado.
4. (June 2019) "Testing for association in multi-view network data" at the 2019 WNAR Annual Meeting of IBC (International Biometric Conferences), in Portland, Oregon.
5. (December 2017) "Are clusterings of multiple data views independent?" at AT&T Graduate Student Symposium, in New York City, New York.
6. (September 2016) "Distributionally robust multinomial regression" at BIRS Robustness Theory and Methodology: Recent Advances and Future Directions workshop, in Banff, A.B., Canada.
7. (August 2015) "D-optimal designs based on the second-order least squares estimator" at the 2015 Joint Statistical Meetings, in Seattle, Washington.
8. (May 2014) "New optimal design criteria for regression models with asymmetric errors" at the Statistical Society of Canada (SSC) Student Conference, in Toronto, Ontario.

Research Funding as Principal Investigator

Internal

New Faculty Start-up Grant: \$60,000 CAD
University of Waterloo

2020-2025

External

NSERC Discovery Grants: \$90,000 CAD
Natural Sciences and Research Council of Canada

2021-2026

NSERC Discovery Launch Supplement: \$12,500 CAD
Natural Sciences and Research Council of Canada

2021-2022

Awards and Scholarships

- (2020) **University of Washington SPH Outstanding PhD Student**, awarded to a graduating Biostatistics PhD student in recognition of excellence in Coursework & Exam Performance, Thesis & Publications, Written & Oral Skills, and Leadership.
 - (2020) **ASA Statistical Learning and Data Science Section Student Paper Award**, for “*Testing for association in multi-view network data*”, \$1,000 USD.
 - (2019) **ASA Biometrics Section Student Paper Award**, for “*Are clusterings of multiple data views independent?*”, \$1,000 USD.
 - (2016-2019) **NSERC PGSD-3**, a doctoral scholarship from the Natural Sciences and Engineering Council of Canada, \$63,000 CAD.
 - (2015) **CIHR Summer Studentship Award**, an award providing funding for undergraduate research from the Canadian Institutes of Health Research, \$6,250 CAD.
 - (2014) **NSERC Undergraduate Student Research Award**, an award providing funding for undergraduate research from the Natural Sciences and Engineering Council of Canada, \$5,625 CAD.
 - (2013) **NSERC Undergraduate Student Research Award**, \$5,625 CAD.
-

Service

Internal

- (2021-2022) Clinical Research Ethics Board Member at the University of Waterloo
- (2020-2022) Department Seminar Committee Member at the University of Waterloo
- (2017-2018) Peer Mentoring Program Member at the University of Washington
- (2017-2018) Co-organizer for working groups at the University of Washington Biostatistics department (Witten Lab and SLAB Lab)

External

- Reviewer for *Biometrika*, *Journal of the American Statistical Association: Theory and Methods*, *Journal of Machine Learning Research*, *Statistical Science*, *Journal of Computational and Graphical Statistics*, *Journal of the Royal Statistical Society: Series C*, *Electronic Journal of Statistics*, *Biostatistics*, *Statistics in Medicine*, and *Statistical Papers*

**Research
Supervision**

[Graduated Master's Students](#)

1. Qiaoyu Liang
 - As of Fall 2021: PhD student in Statistics at the University of Toronto

[Ongoing PhD Students](#)

1. Andrew Kenig (2021-current)
 - Co-chair with Shoja'eddin Chenouri

Teaching

[Instructor at the University of Waterloo](#)

STAT 330: Mathematical Statistics Spring 2021

Two undergraduate sections with ~300 students. Lectures were delivered asynchronously online due to the COVID-19 pandemic. Average student course evaluation score: 4.5/5.

STAT 330: Mathematical Statistics Winter 2021

Two undergraduate sections with ~250 students. Lectures were delivered asynchronously online due to the COVID-19 pandemic. Average student course evaluation score: 4.2/5.

[Guest Lectures at the University of Washington](#)

STAT 435: Intro to Statistical Machine Learning Winter 2019

Undergraduate section with ~20 students. My guest lecture topic was unsupervised learning methods.

BIOST 546: ML for Biomedical & Public Health Winter 2019

Graduate section with ~30 students. My guest lecture topic was unsupervised learning methods.

BIOST 311: Regression Methods in Health Sciences Spring 2018

Undergraduate section with ~10 students. My guest lecture topic was regression methods for correlated data.

[Teaching Assistant at the University of Washington](#)

BIOST 310: Biostatistics in Health Sciences Spring 2018

Undergraduate section with ~125 students. I taught discussion/tutorial sections once a week for ~30 students, held office hours, and graded assignments. I also developed material for discussion/tutorial sections.

**Industry Work
Experience**

Data Scientist Intern Summer 2018
LinkedIn Co. Experimentation Science Team