Lucy L. Gao

Assistant Professor, Department of Statistics and Actuarial Science

***Email*: lucy dot gao at uwaterloo dot ca** **_Website_: lucylgao dot com**

Education

SEPTEMBER 2015 - JUNE 2020

University of Washington, Seattle WA - *Ph.D. Biostatistics*

SEPTEMBER 2011 - JUNE 2015

University of Victoria, Victoria BC - *B.Sc. Honours Mathematics and Statistics*

Work

JULY 2020 - NOW

University of Waterloo, Waterloo ON - *Assistant Professor, Department of Statistics and Actuarial Science*

JULY 2018 - SEPTEMBER 2018

LinkedIn Co, Sunnyvale CA - *Data Scientist Intern, Experimentation Science*

Publications

Statistical Methodology

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*. **[Won a 2020 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*, https://doi.org/10.1080/03610926.2020.1748655.
3. **Gao, L. L.**, Bien, J. and Witten, D. (2020) Are clusterings of multiple data views independent? *Biostatistics*, 21(4), 692-708. **[Won a 2019 ASA Biometrics Section Travel 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. **Gao, L. L.** and Zhou, J. (2017) D-optimal designs based on the second-order least squares estimator. *Statistical Papers*, 58(1), 77-94.
6. **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. **[A video based on this paper was a finalist in the SSC Statistics on Reels video competition.]**

Preprints:

1. Neufeld, A., **Gao, L.L.** and Witten, D. (2021+) Tree-values: selective inference for regression trees. Preprint available at https://arxiv.org/pdf/2106.07816.
2. **Gao, L.L.**, Bien, J., and Witten, D. (2021+) Selective inference for hierarchical clustering. Preprint available at https://arxiv.org/abs/2012.02936.

Statistical Applications

1. 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. **[Received an editorial in *Gastroenterology*.]**

Presentations

Invited Presentations

1. (July 2021) "Valid Differential Analysis for Groups Defined via Clustering" for JOBIM2021, virtually via Zoom.
2. (May 2021) "Valid inference after hierarchical clustering" for the Harvard-BCH Machine Learning and AI Journal Club, virtually via Zoom.
3. (March 2021) "Hypothesis testing after hierarchical clustering" for the UCSF Department of Epidemiology and Biostatistics Seminar, virtually via Zoom.
4. (November 2020) "Beyond sample-splitting: valid inference while 'double-dipping'" for the University of Waterloo Student Seminar series, virtually via Zoom.
5. (August 2020) "Selective inference for hierarchical clustering" for the International Seminar on Selective Inference, virtually via Zoom.
6. (June 2020) "Statistical inference for multi-view clustering" for the Young Data Science Researcher Seminar at ETH Zurich, virtually via Zoom.
7. (January 2020) "Statistical inference for multi-view clustering" for the University of Waterloo Department of Statistics and Actuarial Science department seminar series, in Waterloo, O.N., Canada.
8. (December 2019) "Statistical inference for multi-view clustering" for the McGill University Department of Mathematics and Statistics department seminar series, in Montreal, Q.C., Canada.
9. (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.
10. (July 2018) "Are clusterings of multiple data views independent?" at the 2018 Joint Statistical Meetings, in Vancouver, B.C., Canada.

Contributed Presentations

1. (July 2020) "Testing for association in multi-view network data" at the 2020 Joint Statistical Meetings, virtually via Zoom.

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

Funding

- NSERC Discovery Grants (2021-2026), valued at a total of $102,500 CAD

Awards and Scholarships

- (2020) **University of Washington SPH Outstanding PhD Student**
- (2020) **ASA Statistical Learning and Data Science Section Student Paper Award**, valued at *$1,000 USD*
- (2019) **ASA Biometrics Section Travel Award**, valued at *$1,000 USD*
- (2016-2019) **NSERC PGSD-3**, a doctoral scholarship from the Natural Sciences and Engineering Council of Canada, valued at *$21,000 CAD/year for 3 years*
- (2015) **CIHR Summer Studentship Award**, an undergraduate research award from the Canadian Institutes of Health Research, valued at *$6,250 CAD*
- (2014) **NSERC Undergraduate Student Research Award**, valued at *$5,625 CAD*
- (2013) **NSERC Undergraduate Student Research Award**, valued at *$5,625 CAD*

Teaching Experience

JAN 2021-AUG 2021, University of Waterloo

Instructor for STAT 330: Mathematical Statistics

MARCH 2019, University of Washington, Seattle

Guest Lecturer for STAT 435: Introduction to Statistical Machine Learning

- STAT 435 is a class targeted at undergraduate students majoring in statistics.
- Gave a guest lecture on unsupervised learning methods.

MARCH 2019, University of Washington, Seattle

Guest Lecturer for STAT 546: Machine Learning for Biomedical and Public Health

- STAT 546 is a class targeted at graduate students in the School of Public Health.

- Gave a guest lecture on unsupervised learning methods.

June 2018, University of Washington, Seattle

Guest Lecturer for BIOST 311: Regression Methods in the Health Sciences

- BIOST 311 is a class targeted at undergraduate students in the School of Public Health.
- Gave a guest lecture introducing regression methods for correlated data.

April 2018-June 2018, University of Washington, Seattle

Teaching Assistant for BIOST 310: Biostatistics in the Health Sciences

- BIOST 310 is a class targeted at undergraduate students in the School of Public Health.
- Taught discussion/tutorial sections once a week, held office hours, and graded assignments.
- Developed material for discussion/tutorial sections.

Service

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

(2017-2018) UW Peer Mentoring Program Member

(2017-2018) Co-organizer for UW Biostatistics working groups (Witten Lab and SLAB Lab)

Software

multiviewtest, R Package on CRAN

clusterpval, R Package on Github. See http://www.lucylgao.com/clusterpval/index for details.