

Department of Mathematics and Statistics
 Box 27, 124 Raymond Ave
 Vassar College, Poughkeepsie, NY 12604

Phone: (845) 518 6499
 Email: jihu@vassar.edu
 Web: <http://pages.vassar.edu/jihu/>

RESEARCH INTERESTS

Bayesian education, Bayesian methods, blended/hybrid learning, data confidentiality, differential privacy, disclosure risk, missing data, multiple imputation, non-parametric Bayes, synthetic data

EDUCATION

Ph.D. in Statistical Science, *Duke University* 2015
 Advisor: Jerome P. Reiter
 Title: “Dirichlet Process mixture models for nested categorical data”

M.S. in Statistical Science, *Duke University* 2013

B.S. in Computing Mathematics, *City University of Hong Kong* 2011
 First class honor, minor in Finance

POSITIONS

Assistant Professor 2015 - present
Department of Mathematics and Statistics, Vassar College

Faculty Fellow 2020 - present
The National Center for Science and Engineering Statistics, National Science Foundation

Statistical Consultant 2020 - present
resonate

ASA/NSF/BLS Fellow 2017 - 2018
The U.S. Bureau of Labor Statistics

CIRJE Visitor Spring 2018
Faculty of Economics, University of Tokyo

Research Assistant 2012 - 2015
Department of Statistical Science, Duke University

Visiting Scholar Summer 2014
Institute for Employment Research (IAB), Germany

Summer Intern Summer 2013
IIS Statistical Methodology Group, Novartis

TEACHING EXPERIENCE**SEMESTER-LONG COURSES**

Instructor, *Department of Mathematics and Statistics, Vassar College* 2015 - present

- . MATH 126 - Calculus IIA: Integration Theory
- . MATH 127 - Calculus IIB: Sequences and Series
- . MATH 141 - Introduction to Statistical Reasoning
- . MATH 144 - Foundations of Data Science
- . MATH 220 - Multivariable Calculus
- . MATH 240 - Introduction to Statistics
- . MATH 241 - Probability
- . MATH 301 - Data Confidentiality
- . MATH 341 - Statistical Inference
- . MATH 347 - Bayesian Statistics
- . INTD 183 - Intro to Critical Data Science (LACOL)

Instructor, *Liberal Arts Collaborative for Digital Innovation (LACOL)* 2017 - present

Bayesian Statistics, Course Sharing Project

Project info: <https://lacol.net/category/collaborations/course-sharing/>

Course material: YouTube playlists - Fall 2017, Spring 2019, Fall 2019

STUDENT RESEARCH

Supervisor of independent studies, *Vassar College*

- . Bayesian Estimation of Future Realized Volatility Fall 2019
- . Bayesian Inferences with Python Spring 2020
- . Bayesian Methods for Sparse Data Spring 2020
- . Bayesian Non-Parametric Models Spring 2017
- . Bayesian Time Series Fall 2017 & Spring 2019
- . Bayesian Variable and Model Selection Fall 2019
- . Identification Risks of Partial Synthetic Data Fall 2019 & Spring 2020
- . LACOL Course Developer Spring 2020
- . Python for Data Science Spring 2020
- . Topics of Data Science Fall 2019
- . Tree-Based Methods for Synthetic Data Fall 2017

SHORT COURSES AND WORKSHOPS

Co-Instructor, *Joint Statistical Meetings (virtual)* November 2020

- . Short Course - Bayesian Thinking: Fundamentals, Computation, and Hierarchical Modeling

Instructor, *NCSES Lecture Series* June 2020

- . Workshop - Data Privacy: Bayesian Data Synthesis and Differential Privacy

Co-Instructor, *Liberal Arts Collaborative for Digital Innovation* March 2020

- . Workshop - Liberal Arts Remote Teaching

Instructor, *Blended Learning in the Liberal Arts Conference* May 2019

- . Workshop - Teaching a Shared/Hybrid/Online course using Zoom

Instructor, *U.S. Bureau of Labor Statistics* October 2018
 . Short course - Introduction to Bayesian Inference in R

Instructor, *CIRJE, University of Tokyo* January 2018
 . Short course - The Dirichlet Process and DP Mixture Models

SUMMER COURSES

Co-Instructor, *Liberal Arts Collaborative for Digital Innovation (LACOL)* 2018 - present
 Introduction to Data Science, Course Sharing Project
 Project info: <https://lacol.net/critical-ds/>

Instructor, *Department of Statistical Science, Duke University* Summer 2014
 STA 101 - Data Analysis and Statistical Inference

CERTIFICATES

Certificate of College Teaching, *Duke Graduate School* 2015

PUBLICATIONS & MANUSCRIPTS

PUBLICATIONS & MANUSCRIPTS OF SCHOLARLY RESEARCH

* indicates an undergraduate student co-author

1. Hornby, R.* and **Hu, J.**, “Identification risk evaluation of continuous synthesized variables”, submitted. arXiv link
2. **Hu, J.**, Savitsky, T. D. and Williams, M. R., “Re-weighting of vector-weighted mechanisms for utility maximization under differential privacy”, submitted. arXiv link
3. Savitsky, T. D., Williams, M. R. and **Hu, J.**, “Bayesian Pseudo Posterior Mechanism under Differential Privacy”. arXiv link
4. **Hu, J.**, Savitsky, T. D. and Williams, M. R., “Risk-efficient Bayesian data synthesis for privacy protection”, submitted. arXiv link
5. **Hu, J.**, Akande, O., and Wang, Q., “Data imputation and data synthesis with the R package NPBayesImputeCat”, submitted. arXiv link
6. **Hu, J.** and Savitsky, T. D., “Bayesian data synthesis and disclosure risk quantification: an application to the Consumer Expenditure Surveys”, submitted. arXiv link
7. **Hu, J.**, Savitsky, T. D. and Williams, M. R. (2020+), “Risk-weighted data synthesizers for microdata dissemination”, minor revision submitted to *CHANCE Special Issue on Data Privacy (invited)*.

8. Drechsler, J. and **Hu, J.** (forthcoming), “Synthesizing geocodes to facilitate access to detailed geographical information in large-scale administrative data”, to appear at *Journal of Survey Statistics and Methodology*. arXiv link
9. Ros, K.* , Olsson, H.* and **Hu, J.** (2020), “Two-phase data synthesis for income: an application to the NHIS”, *Privacy in Statistical Databases e-Proceedings*. link
10. **Hu, J.** (2019), “Bayesian estimation of attribute and identification disclosure risks in synthetic data”, *Transactions on Data Privacy*, 12:1, 61-89.
11. **Hu, J.** and Hoshino, N. (2018), “The Quasi-Multinomial synthesizer for categorical data”, *Privacy in Statistical Databases (Lecture Notes in Computer Science 11126)*, ed. J. Domingo-Ferrer and F. Montes, Springer, 75-91.
12. Manrique-Vallier, D. and **Hu, J.** (2018), “Bayesian non-parametric generation of synthetic multivariate categorical data in the presence of structural zeros”, *Journal of the Royal Statistical Society, Series A (Statistics in Society)*, 181(3), 635-647.
13. **Hu, J.**, Reiter, J.P. and Wang, Q. (2018), “Dirichlet Process mixture models for modeling and generating synthetic versions of nested categorical data”, *Bayesian Analysis*, 13(1), 183-200.
14. **Hu, J.** and Drechsler, J. (2015), “Generating synthetic geocoding information for public release”, In: *S. A. Europäische Kommission (Hrsg.), NTTTS -Conferences on New Techniques and Technologies for Statistics*, 56-59.
15. **Hu, J.**, Reiter, J.P. and Wang, Q. (2014), “Disclosure risk evaluation for fully synthetic categorical data”, *Privacy in Statistical Databases (Lecture Notes in Computer Science 8744)*, ed. J. Domingo-Ferrer, Springer, 185-199.
16. **Hu, J.**, Mitra, R. and Reiter, J.P. (2013), “Are independent parameter draws necessary for multiple imputation?” *The American Statistician*, 67(3), 143-149.
17. **Hu, J.** and Reiter, J.P. (2013), “Non-parametric Bayesian model for generating synthetic household data ”, *Joint UNECE/Eurostat Work Session on Statistical Data Confidentiality 2013*, http://www.unece.org/fileadmin/DAM/stats/documents/ece/ces/ge.46/2013/Topic_2_Duke_University.pdf

PUBLICATIONS & MANUSCRIPTS OF STATISTICS EDUCATION

1. **Hu, J.** and Guo, F., “Bayesian Modeling - A Primer”, submitted.
2. Albert, J. and **Hu, J.** (2020+), “Bayesian Computing in the Undergraduate Statistics Curriculum”, submitted. arXiv link
3. **Hu, J.** (forthcoming), “A Bayesian Statistics Course for Undergraduates: Bayesian Thinking, Computing, and Research”, *Journal of Statistics Education*. arXiv link
4. Albert, J., Cetinkaya-Rundel, M., and **Hu, J.** (2020), “Online statistics teaching and learning”, *Teaching and Learning Mathematics Online*, ed. J. P. Howard and J. F. Beyers, Chapman and Hall/CRC, Boca Raton, Florida.
5. Garcia, S. R., **Hu, J.**, and Miller, S. J. (2020), “Upper level mathematics and statistics courses share across campuses”, *Teaching and Learning Mathematics Online*, ed. J. P. Howard and J. F. Beyers, Chapman and Hall/CRC, Boca Raton, Florida.

6. **Hu, J.** (2019), “Teaching upper level undergraduate statistics through a shared/hybrid model”, *Chance Magazine Statistics Education Column, Taking a Chance in the Classroom*, 32(2), 31-36.

BOOK PROJECTS

1. Albert, J. and **Hu, J.** (2019), “Probability and Bayesian Modeling”, *Texts in Statistical Science*, CRC Press.

SOFTWARE

1. Hornby, R. and **Hu, J.**, “IdentificationRiskCalculation: Calculating the Identification Risk in Partially Synthetic Microdata”, GitHub link.
2. Wang, Q., Akande, O., **Hu, J.**, Reiter, J. P. and Barrientos, A., “NestedCategBayesImpute: Modeling and Generating Synthetic Versions of Nested Categorical Data in the Presence of Impossible Combinations”, CRAN link.
3. Wang, Q., Manrique-Vallier, D., Reiter, J. P., and **Hu, J.**, “NPBayesImputeCat: Non-Parametric Bayesian Multiple Imputation for Categorical Data”, CRAN link.

RESEARCH & TEACHING SUPPORT

RESEARCH SUPPORT

“Modern Disclosure Limitation and Privacy Protection Methods for Survey Data”

National Center for Science and Engineering Statistics (NSF) Faculty Fellowship

Role: Principle Investigator. Period: 6/1/2020 - 12/31/2020. Amount: \$31,608.13

“Disclosure Risks Evaluation of Public Use Microdata”

Research Committee Award, Vassar College

Role: Principal Investigator. Period: 5/6/2019 - 12/31/2020. Amount: \$3,871.00

“Synthetic Consumer Expenditure Survey Data at the Bureau of Labor Statistics”

ASA/NSF/BLS Research Fellowship

Role: Principal Investigator. Period: 2/1/2018 - 4/30/2020. Amount: \$88,531.00

Research Committee Award, Vassar College

Role: Principal Investigator. Period: 5/5/2017 - 12/31/2018. Amount: \$2,940.45

TEACHING SUPPORT**“Foundations of Data Science Course”**

Grand Challenge Climate Change, Vassar College

Role: Co-Principal Investigator. Period: 6/1/2019 - 6/30/2021. Amount: \$11,000.00

“LACOL Digitally Shared Course, Bayesian Statistics”

Liberal Arts Collaborative for Digital Innovation (LACOL) funding

Role: Principal Investigator. Period: 10/3/2018 - 12/31/2019. Amount: \$20,000.00

Endowment for Strategic Faculty Support, Vassar College

Role: Principal Investigator. Period: 4/18/2017 - 12/31/2017. Amount: \$1,287.75

“LACOL Digitally Shared Course, Introduction to Data Science”

Liberal Arts Collaborative for Digital Innovation (LACOL) funding

Role: Co-Principal Investigator. Period: 1/1/2019 - 9/30/2020. Amount: \$107,950.00

“LACOL Digitally Shared Course, Data Confidentiality”

Liberal Arts Collaborative for Digital Innovation (LACOL) funding

Role: Principal Investigator. Period: 1/1/2020 - 6/30/2020. Amount: \$4,500.00

“LACOL Digitally Shared Course, Bayesian Inference with Python”

Liberal Arts Collaborative for Digital Innovation (LACOL) funding

Role: Principal Investigator. Period: 1/1/2020 - 6/30/2020. Amount: \$3,000.00

PRESENTATIONS**PRESENTATIONS AT STATISTICAL MEETINGS**

“Bayesian Pseudo Posterior Synthesis for Data Privacy Protection” June 2021
The Sixth International Conference on Establishment Statistics (invited), New Orleans, LA, USA

“Two-phase data synthesis for income: an application to the NHIS” September 2020
Privacy in Statistical Databases 2020, Zoom

“Thinking beyond the p-values: advancing Bayesian education for the undergraduates” August 2020
Joint Statistical Meetings (invited, organizer & chair), Zoom

“Risk-efficient Bayesian Data Synthesis for Privacy Protection” August 2020
Joint Statistical Meetings (invited talk), Zoom

“Bayesian Pseudo Posterior Synthesis for Data Privacy Protection” August 2019
Joint Statistical Meetings (topic-contributed talk), Denver, CO, USA

“Teaching Upper-Level Statistics Courses Through a Shared/Hybrid Model” May 2019

Symposium on Data Science and Statistics, Bellevue, WA, USA

“The Quasi-Multinomial Synthesizer for Categorical Data” September 2018
Privacy in Statistical Databases 2018, Valencia, Spain

“Synthetic Consumer Expenditure Survey Data at U. S. Bureau of Labor Statistics” June 2018
International Conference on Applied Mathematics (ICAM 2018), Hong Kong

“Teaching Upper-Level Statistics Courses Through a Shared/Hybrid Model” October 2017
Women in Statistics and Data Science, La Jolla, CA, USA

“Introducing Bayesian Statistics at Courses of Various Levels” August 2017
Joint Statistical Meetings (topic-contributed, organizer & discussant), Baltimore, MD, USA

“Strategies to facilitate access to geocoding information based on synthetic data” June 2017
The 1st International Conference on Econometrics and Statistics (invited), Hong Kong

“Strategies to facilitate access to geocoding information based on synthetic data” May 2017
Third International Conference on Engineering and Computational Mathematics (invited), Hong Kong

“Graduate student internships” & “Making sense of the transition to faculty life” October 2016
Women in Statistics and Data Science (panelist), Charlotte, NC, USA

“Professional Opportunities at Smaller Colleges and Universities” August 2016
Joint Statistical Meetings (topic-contributed panel), Chicago, IL, USA

“Dirichlet Process Mixture Models for Nested Unordered Categorical Data” June 2016
Frontiers of Statistics and Data Sciences (invited), Hong Kong

“Dirichlet Process Mixture Models for Nested Unordered Categorical Data” August 2015
Joint Statistical Meetings (topic-contributed talk), Seattle, WA, USA

“Nested Dirichlet Process Models for Household Data Synthesis” October 2014
Advances in Interdisciplinary Statistics and Combinatorics, Greensboro, NC, USA

“Disclosure Risk Evaluation for Fully Synthetic Categorical Data” September 2014
Privacy in Statistical Databases 2014, Eivissa, Spain

“Nested Dirichlet Process Models for Household Data Synthesis” August 2014
Joint Statistical Meetings (contributed talk), Boston, MA, USA

“Nested Dirichlet Process Models for Household Data Synthesis” October 2013
UNECE/Eurostat Work Session on Statistical Data Confidentiality (invited), Ottawa, Canada

“Nonparametric Bayesian Models for Generating Synthetic Household Data” August 2013
Joint Statistical Meetings (topic-contributed talk), Montreal, Canada

PRESENTATIONS AT FEDEARL STATISTICAL AGENCIES

- “A Friendly Discussion on Differential Privacy” September 2020
National Center for Health Statistics (invited, co-presenter), USA
- “Access to Restricted Data Files for Select Statistical Research Projects” July 2019
Consumer Expenditure Surveys (CE) Microdata Users’ Workshop 2019 (panelist), Washington D.C., USA
- “Using CE Microdata in Undergraduate Statistics Courses” July 2019
Consumer Expenditure Surveys (CE) Microdata Users’ Workshop 2019, Washington D.C., USA
- “Bayesian Pseudo Posterior Synthesis for Data Privacy Protection” March 2019
USDA National Agricultural Statistical Service, Washington D.C., USA
- “Bayesian Pseudo Posterior Synthesis for Data Privacy Protection” March 2019
Center for Statistical Research and Methodology Seminar, U.S. Census Bureau, Washington D.C., USA
- “Exploring Synthetic Consumer Expenditure Surveys Data” December 2018
Disclosure Review Board Seminar (invited), U.S. Bureau of Labor Statistics, Washington D.C., USA
- “The synthetic data approach to data confidentiality” December 2016
U.S. Bureau of Labor Statistics (invited), Washington DC, USA

PRESENTATIONS AT UNIVERSITIES AND ORGANIZATIONS

- “Risk-efficient Bayesian Data Synthesis for Privacy Protection” February 2020
Seminar, University of Massachusetts Amherst, Amherst, MA, USA
- “Generating Synthetic Family Income for the Consumer Expenditure Surveys” February 2020
Seminar, Smith College, Northampton, MA, USA
- “Risk-efficient Bayesian Data Synthesis for Privacy Protection” January 2020
Westat, Rockville, MD, USA
- “Bayesian Pseudo Posterior Synthesis for Data Privacy Protection” May 2019
Epidemiology and Biostatistics Seminar, Drexel University, PA, USA
- “Synthetic Consumer Expenditure Survey Data at U. S. Bureau of Labor Statistics” October 2018
Statistics Seminar, Binghamton University, NY, USA
- “Teaching Upper-Level Statistics Courses Through a Shared/Hybrid Model” October 2018
Statistics Seminar, Binghamton University, NY, USA
- “The synthetic data approach to data confidentiality” January 2018
Workshop on Statistical Disclosure Control, Kanazawa University, Japan
- “Dirichlet Process Mixture Models for Nested Unordered Categorical Data” January 2018

The Applied Statistics Workshop, University of Tokyo, Japan

“Teaching an Undergraduate Bayesian Statistics Course” December 2017
Statistical Science Department Seminar, Duke University, NC, USA

“Dirichlet Process Mixture Models for Nested Unordered Categorical Data” October 2017
Department of Mathematics and Statistics, Bowling Green State University, OH, USA

“The synthetic data approach to data confidentiality” February 2017
AALAC Workshop on Data Ethics, Pomona College, Claremont, CA, USA

“Strategies to facilitate access to geocoding information based on synthetic data” October 2016
NYU Center for Urban Science + Progress (invited), New York, NY, USA

“Why and How to Make Fake Data” August 2016
Faculty Research Presentations, Vassar College, Poughkeepsie, NY, USA

“Generating Synthetic Household Data for the Decennial Census” June 2015
Department of Mathematics (invited), City University of Hong Kong

“Missing Not at Random in SAS” August 2013
Norvatis Statistical Science VC, East Hanover, NJ, USA

PRESENTATIONS AT EDUCATION MEETINGS

“Experience Sharing Online Classes across Liberal Arts Colleges” October 2020
AMS Committee on Education Virtual Mini-conference (invited)

“Summer Data Science Online: Building Bridges through Collaboration in the Liberal Arts” June 2020
OLC Innovate, Zoom

“LACOL 2020 Summer Data Science Panel” June 2020
LACOL 2020 Virtual Workshop, Zoom

“Summer Data Science (Online) by and for the Liberal Arts” June 2019
National Workshop on Data Science Education 2019 (lightning talk), UC Berkeley, CA, USA

“Teaching Upper-Level Statistics Courses Through a Shared/Hybrid Model” May 2018
Blended Learning in the Liberal Arts 2018, Bryn Mawr College, PA, USA

“Where’s the remote? Exploring upper level math/stats hybrid course sharing for the liberal arts” January 2018
EDUCAUSE Learning Initiative (ELI) Annual Meeting, New Orleans, LA, USA

COLLEGE SERVICE

Committee Member

Science / Math Brainstorming Group (covid-19), Vassar College 2020

<i>Fellowship Committee, Vassar College</i>	2019 - present
<i>Committee on Committees, Vassar College</i>	2016 - 2018
<i>Thompson-Bartlett Fellowship Selection Committee, Vassar College</i>	2017
<i>Graduate Consultative Committee, Department of Statistical Science, Duke University</i>	2012 - 2013

Departmental Service

<i>Administrative assistant search</i>	2019
<i>Statistics faculty search</i>	2019
<i>Applied Mathematics faculty search</i>	2016 - 2017
<i>Major advisor</i>	2016 - present

Pre-major Academic Advisor

2016 - present

*Vassar College***Co-organizer**

<i>DataFest @ Vassar</i> http://pages.vassar.edu/datafest/	annually since 2016
<i>Women of Color in STEM Workshop, Women of Color Conference, CHAS</i>	2017
<i>Exploring Online Bridging Program, Vassar College</i>	2017

College Program Involvement

<i>HHMI Grand Challenges</i>	2019 - present
------------------------------	----------------

Faculty Host

2016 - present

*International Friendship Program, Vassar College***Faculty Panelist/Speaker**

<i>Gender Inclusivity in STEM dinner</i>	2019
<i>Journeys, Campus Life and Diversity Office, Vassar College</i>	2017
<i>Hidden Figures Strong House Screening, Vassar College</i>	2017
<i>Story Sharing Session, Chinese Student Community (CSC), Vassar College</i>	2017
<i>Discover America, Office of International Services, Vassar College</i>	2017
<i>Vassar Science Scholars Program, Vassar College</i>	2017
<i>Talking about Teaching session on international students, Vassar College</i>	2017
<i>Strong STEM Brunch Series (Computer Science and Math), Vassar College</i>	2015

PROFESSIONAL SERVICE**Editorial Board**

<i>Transactions on Data Privacy</i>	2019 - present
-------------------------------------	----------------

Referee

Annals of Applied Statistics, Bayesian Analysis, CRC Press, Epidemiology, Journal of Computational and Graphical Statistics, Journal of Official Statistics, Journal of Survey Statistics and Methodology, Journal of Statistics Education, METRON, National Science Foundation, Pearson, Princeton University Press, Revista CEA, Statistical Analysis and Data Mining, The American Statistician

Treasurer

2017 - present

Education Research and Practice Section, International Society for Bayesian Analysis (ISBA)

Volunteer 2016
Statistics Education Booth, Joint Statistical Meetings 2016

AWARDS

Faculty Advisor to Honorable Mention Winner Spring 2020
The Undergraduate Research Project Competition (USRESP), CAUSE & ASA

Faculty Advisor to Student 1st-place Winner Spring 2019
The Undergraduate Class Project Competition (USCLAP), intermediate statistics, CAUSE & ASA

Faculty Advisor to Student 1st-place Winner Fall 2016
The Undergraduate Class Project Competition (USCLAP), intermediate statistics, CAUSE & ASA

Travel Award
CBMS Regional Conference on Spatial Statistics 2017
Women in Statistics and Data Science 2016 & 2017

ASA Student Paper Competition Award, Joint Statistical Meetings 2015
Survey Research Methods, Government Statistics, and Social Statistics Sections

Ph.D. First Year Fellowship 2011 - 2012
Department of Statistical Science, Duke University

Scholarship for Mainland China Students 2007-2011
City University of Hong Kong