

CONTACT INFORMATION

RAND Corporation
4570 Fifth Ave.
Pittsburgh, PA 15213

jsnoke@rand.org
www.joshuasnoke.com

EDUCATION

Ph.D., Statistics, Pennsylvania State University 08.2018
Graduate Minor: Social Data Analytics
Dissertation: *Statistical Data Privacy Methods for Increasing Research Opportunities*
Advisor: Aleksandra B. Slavković

B.S., Mathematics and Economics, Wheaton College 05.2013
Honors: Magna Cum Laude

RESEARCH APPOINTMENTS

Associate Statistician, RAND Corporation 08.2018 - Present
Research Adjunct 09.2016 - 08.2017
Summer Associate 06.2016 - 08.2016

Graduate Student Researcher, Pennsylvania State University
Research Assistant, Department of Statistics 01.2018 - 06.2018
05.2017 - 08.2017
01.2014 - 08.2014
Research Trainee, Big Data Social Science IGERT 09.2014 - 05.2016

Research Fellow, Administrative Data Research Centre - Scotland 05.2015 - 08.2015

PROFESSIONAL ACTIVITIES

Journal Referee

Journal of Survey Statistics and Methodology, IEEE Transactions on Information Forensics & Security, Statistical Science, Journal of Research on Educational Effectiveness

Professional Service

Human Subjects Protection Committee 06.2019 - Present
RAND Corporation

Committee on Privacy and Confidentiality 10.2018 - Present
American Statistical Association (ASA)

Subject Matter Expert for Differentially Private Synthetic Data Challenge 10.2018 - 05.2019
National Institute of Standards and Technology (NIST)

Conference Program Committees

Symposium on Data Science and Statistics 2020 10.2019 - 06.2020
ACM-IMS Foundations of Data Science 2020 01.2020 - 10.2020

PEER REVIEWED PUBLICATIONS

Snoke, J., Raab, G.M., Nowok, B., Dibben, C. and Slavković, A., (2018). General and specific

utility measures for synthetic data. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 181(3), pp.663-688. doi:10.1111/rssa.12358

Snoke, J., Brick, T.R., Slavković, A. and Hunter, M.D., (2018). Providing accurate models across private partitioned data: Secure maximum likelihood estimation. *The Annals of Applied Statistics*, 12(2), pp.877-914. doi:10.1214/18-AOAS1171.

Snoke, J. and Slavković, A., (2018). *pMSE* Mechanism: Differentially Private Synthetic Data with Maximal Distributional Similarity. In *Privacy in Statistical Databases*. pp.138-159. Springer. doi:10.1007/978-3-319-99771-1

Snoke, J., Brick, T. and Slavković, A., (2016). Accurate Estimation of Structural Equation Models with Remote Partitioned Data. In *Privacy in Statistical Databases*. pp.190-209. Springer. doi:10.1007/978-3-319-45381-1

NON-PEER REVIEWED ARTICLES

Snoke, J. and Bowen, Claire M.. Differential Privacy: What Is It? *Amstat News*. pp.26-28. March 2019. Issue #501. <https://magazine.amstat.org/blog/2019/03/01/differentialprivacy/>.

PAPERS IN PREP

Bowen, Claire M. and **Snoke, J.**. Comparative Study of Differentially Private Synthetic Data Algorithms and Evaluation Standards. Preprint available at <https://arxiv.org/abs/1911.12704>.

Orobia II, A.G., Linder, F. and **Snoke, J.**, Using Neural Generative Models to Release Synthetic Twitter Corpora with Reduced Stylometric Identifiability of Users. Preprint available at <https://arxiv.org/abs/1606.01151>.

FELLOWSHIPS & AWARDS

Fellowships

Dissertation Fellowship, U.S. Census Bureau 08.2016 - 07.2017

NSF Big Data Social Science IGERT Traineeship,
Pennsylvania State University 09.2014 - 08.2016

President's Award, Wheaton College 08.2009 - 05.2013

Research Awards

Joint Statistical Meetings ASA Student Paper Award 07.2016

General and Specific Utility Measures for Synthetic Data
Social Statistics, Government Statistics, and Survey Research Methods Section

Rao Prize Conference Student Poster Award 05.2015
Estimation of Structural Equation Models for Vertically Partitioned Data

TEACHING APPOINTMENTS

Pennsylvania State University

Graduate Teaching Assistant

Categorical Data Analysis 08.2017 - 12.2017

Applied Data Mining and Statistical Learning 08.2017 - 12.2017

Elementary Statistics 08.2013 - 12.2013

Wheaton College

Teaching Assistant

Intermediate Macroeconomics

Calculus II

08.2011 - 05.2012

01.2011 - 05.2011

PRESENTED WORK

Invited Talks

Simons Institute Program on Data Privacy: Foundations and Applications 03.2019
Statistical Perspectives on Differentially Private Synthetic Data

International Conference on Privacy in Statistical Databases 09.2018
pMSE Mechanism: Differentially Private Synthetic Data with Maximal Distributional Similarity

U.S. Census Bureau 06.2018
Statistical Data Privacy Methods for Increasing Research Opportunities

Joint Statistical Meetings 07.2017
Discussant. Differential Privacy in Statistical Agencies: Present and Future

International Statistical Institute World Statistics Congress 07.2017
Perspective on Utility for Synthetic Data Sets

International Conference on Privacy in Statistical Databases 09.2016
Accurate Estimation of Structural Equation Models with Remote Partitioned Data

Joint Statistical Meetings 08.2016
General and Specific Utility Measures for Synthetic Data

Guest Lectures

Data Privacy Methods, Pennsylvania State University 10.2014
Statistical Disclosure Control Methods

Contributed Talks

Joint Statistical Meetings 08.2019
pMSE Mechanism: A Statistical Approach to Differentially Private Synthetic Data

Joint Statistical Meetings 08.2018
Using Neural Generative Models to Release Synthetic Twitter Corpora with Reduced Stylometric Identifiability of Users

Joint Statistical Meetings 07.2017
Secure Multiparty Maximum Likelihood Estimation with Partitioned Databases

RAND Corporation 08.2016
Social Media and Events Marketing for Army Recruiting

FARR Institute International Conference 08.2015
Utility of Synthetic Microdata for Health Research

Workshop Talks

Stochastic Modeling and Computational Statistics Seminar 04.2017
Secure Multiparty Maximum Likelihood Estimation with Partitioned Databases

Isaac Newton Institute Synthetic Data Workshop 11.2016
Beyond Microdata: Approaches for Generating Synthetic Tweets

Poster Presentations

Department of Statistics 50th Anniversary Conference 05.2018
Differentially Private Synthetic Data with Maximal Distributional Similarity

Big Data Social Science IGERT Poster Session 04.2016
Statistical Disclosure Control for High Dimensional Data

Rao Prize Conference 05.2015
Big Data Social Science IGERT Poster Session 04.2015
Estimation of Structural Equation Models for Vertically Partitioned Data

COMPUTING SKILLS

Statistical Software

Proficient in R, SAS, STATA, Minitab

Other Tools

Proficient in Git, Python, Excel
Experience in Java, SQL

Development

Contributor to R package *synthpop*

ADDITIONAL PROFESSIONAL DEVELOPMENT

Neighborhood Learning Alliance

Summer Tutor 06.2013 - 08.2013

Boston University

Summer Institute for Training in Biostatistics 06.2012 - 07.2012

Via Science

Data Science Intern 08.2012 - 11.2012

LAST UPDATED

02.17.2020