BOYAN DUAN

Email: boyand@google.com ⋄ Website: https://duanby.github.io/

EMPLOYMENT AND EDUCATION

Google, CA, USA   *June 2021 - Now*
Data Scientist
Use causal inference methods to design and suggest experiments and new ways to establish causality, assess attribution and answer strategic questions using data. Build and prototype analysis pipelines iteratively to provide insights at scale.

Carnegie Mellon University, Pittsburgh, USA   *September 2016 - June 2021*
Ph.D. in Statistics
Department of Statistics & Data Science

University of Science and Technology of China, Hefei   *September 2012 - June 2016*
Bachelor of Statistics
School of the Gifted Young

University of Birmingham, UK   *June 2015 - August 2015*
Intern in Computer Science

RESEARCH INTERESTS

My Ph.D. thesis focus on advancing the boundaries of interactive human-in-the-loop inference, advised by Aaditya Ramdas and Larry Wasserman. In practice, analysts tend to try several methods or models on the same dataset until the results are satisfying, but this violates the validity of standard testing methods (e.g., invalid FDR control). Interactive testing is a recent idea that emerged in response to the growing practical needs of allowing human interaction in the process of data analysis.

We have developed interactive tests for various problems, such as multiple testing with FWER control, and identifying individuals with positive effects with FDR control. We view our work as steps towards enabling experts to work together with statistical models and machine learning algorithms in order to discover scientific insights with rigorous guarantees.

Keywords:
Human-in-the-loop interactive testing, Causal inference, Martingale, Adaptive sequential testing

PAPERS AND SUBMISSIONS

Interactive Martingale Tests for the Global Null arXiv (EJS, 2020)
Boyan Duan, Aaditya Ramdas, Sivaraman Balakrishnan, Larry Wasserman
Presented at the International Conference on Multiple Comparison Procedures (MCP) in Dec., 2019.

Familywise error rate control by interactive unmasking arXiv (ICML, 2020)
Boyan Duan, Aaditya Ramdas, Larry Wasserman

Interactive rank testing by betting
arXiv (Conference on Causal Learning and Reasoning, 2022)
Boyan Duan, Aaditya Ramdas, Larry Wasserman

Data fission: splitting a single data point
arXiv (Journal of the American Statistical Association, 2023)
James Leiner, Boyan Duan, Larry Wasserman, Aaditya Ramdas

Interactive identification of individuals with positive treatment effect while controlling false discoveries arXiv (In submission)
Boyan Duan, Larry Wasserman, Aaditya Ramdas
Presented at the International Seminar on Selective Inference in Mar. 2021; and the International Chinese Statistical Association in Sep. 2021.

PROGRAMMING LANGUAGE SKILLS

R, Python, MATLAB, SQL

TEACHING ASSISTANT

In-class tutorials, office hours, and grading for the following classes:
36-218 Probability Theory for Computer Scientists (Head TA)
36-708 Statistical Methods in Machine Learning (PhD Level Class)
36-705 Intermediate Statistics (PhD Level Class)
46-927 Statistical Machine Learning II
36-410 Introduction to Probability Modeling
36-401 Modern Regression
36-217 Probability Theory and Random Processes
36-225 Introduction to Probability Theory

PAST PROJECTS

Perception of security over time in the Democratic Republic of Congo
with Robin Mejia, Anjali Mazumder, Patrick Vinck, Phuong Pham
Over the past decades, there has been continuous armed conflict and economic and political instability in the Democratic Republic of the Congo (DRC). Despite the effort made by the Congolese government to rebuild the country and the ongoing United Nations peacekeeping mission, there has been little improvement in terms of peace and justice. In discussions, Vinck noted that conflicts in Congo are often described as ethnic conflicts, despite a lack of formal study to assess this description. To examine the evolution of the conflict, we studied the changing pattern of the populations perception of security and ethnic relations and uncover regions and subpopulations with different trends.

The automatic generation of semantic environment maps from robot sensor data
with Lars Kunze
During my undergraduate internship, I participated in an international research project on robotics for long-term autonomy in mobile robots involving six European universities, called STRANDS. We acquired robot sensor data using the Robot Operating System (ROS), and used simple machine learning algorithms to interpret sensor data from the robots, such as identifying objects, recognizing the space to move freely, classifying the types of rooms, etc.