Keynote Speaker

Susan Murphy, Ph.D., is the Mallinckrodt Professor of Statistics and of Computer Science, Radcliffe Alumnae Professor at the Radcliffe Institute, Harvard University. Her research focuses on improving sequential, individualized, decision making in health, in particular on clinical trial design and data analysis to inform the development of mobile health treatment policies. Susan is a Fellow of the Institute of Mathematical Statistics, a Fellow of the College on Problems in Drug Dependence, a former editor of the Annals of Statistics, a member of the US National Academy of Medicine and a 2013 MacArthur Fellow.

Title: Inference for longitudinal data after adaptive sampling

Abstract: Adaptive sampling methods, such as reinforcement learning (RL) and bandit algorithms, are increasingly used for the real-time personalization of interventions in digital applications like mobile health and education. As a result, there is a need to be able to use the resulting adaptively collected user data to address a variety of inferential questions, including questions about time-varying causal effects. However, current methods for statistical inference on such data (a) make strong assumptions regarding the environment dynamics, e.g., assume the longitudinal data follows a Markovian process, or (b) require data to be collected with one adaptive sampling algorithm per user, which excludes algorithms that learn to select actions using data collected from multiple users. These are major obstacles preventing the use of adaptive sampling algorithms more widely in practice. In this work, we proved statistical inference for the common Z-estimator based on adaptively sampled data. The inference is valid even when observations are non-stationary and highly dependent over time, and (b) allow the online adaptive sampling algorithm to learn using the data of all users. Furthermore, our inference method is robust to miss-specification of the reward models used by the adaptive sampling algorithm. This work is motivated by our work in designing the Oralytics oral health clinical trial in which an RL adaptive sampling algorithm will be used to select treatments, yet valid statistical inference is essential for conducting primary data analyses after the trial is over.