

Anne-Sophie Charest

Abstract:

One way of ensuring and quantifying confidentiality protection is with differential privacy, a powerful criterion which provides a strict, measurable, guarantee of confidentiality. Several techniques have now been proposed to create completely synthetic datasets which satisfy differential privacy. We consider specifically synthetic data created using the multiple imputation framework, and in particular the beta-binomial synthesizer for count data. Our interest is in methods for users to analyze such synthetic datasets. We will show that inferences from usual statistical methods are not necessarily valid when the datasets are created to achieve differential privacy. We then propose a solution to analyze differentially-private synthetic datasets by directly modeling the data generation process in a Bayesian framework and discuss its limitations. This talk is based on a JPC paper, which can be found at <http://repository.cmu.edu/jpc/vol2/iss2/3/>.