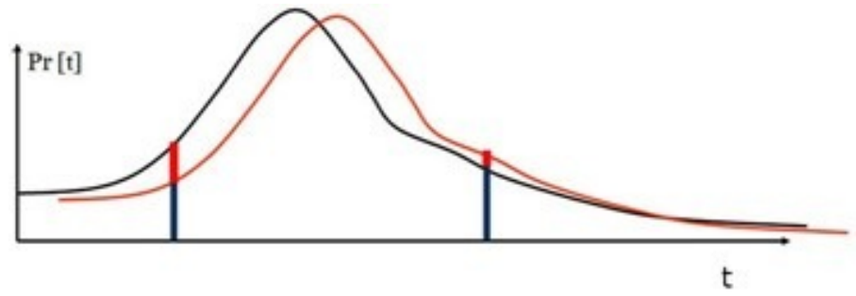
Data Privacy, Mechanism Design and Learning

Steven Wu
University of Pennsylvania

ITCS 2016 Graduating Bits

Differential Privacy [DMNS06]



- Basic query release problem: release aggregate statistics on sensitive data (e.g. medical records)

- Fast and practical query release algorithm [GGHR**W**'14]

- Adapt the notion of DP to other domains: mechanism design; searching for targeted population [KR**W**Y'16]

Data Analyst

queries

(approximate) answers that preserve privacy

Private Algorithm

Sensitive Dataset

Privacy as tool in mechanism design

- Privacy as a notion of algorithmic stability:
  misreporting one agent's data doesn't change the output distribution of all the other agents by much
- A powerful tool to design truthful mediator that implements optimal outcome [KPRU'14]
- Need to compute some kind of equilibrium under the constraint of (joint) differential privacy
  - Allocation problem [HHRR**W**'14][HHR**W**'16]
  - Stable matching [KMR**W**'15]
  - Traffic routing [RRU**W**'15]
  - Aggregative games [CRK**W**'15]

Connection with learning theory

- Learning theory and differential privacy are concerned with being able to discover distributional information about data-sets

- Privacy implies Generalization [DFHPRR'15]: *insensitivity to individual data points is desired so as to make learning algorithms robust to over-fitting*

- Application to fundamental adaptive data analysis task: *post-selection inference (POSI) in variable/model selection (e.g. stepwise regression)*