Entropy testing and its application to testing Bayesian networks

Clément Canonne and Joy Qiping Yang

University of Sydney

Introduction -- distribution testing

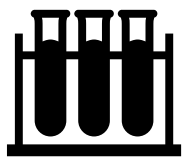
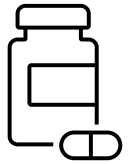
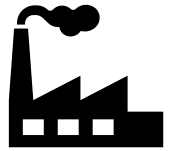
- Someone gives you a coin.
- Question: is it fair or far from being fair by ε?
- Pr[head] = Pr[tail] = 1/2?

- Flip a lot and check the frequency.
  - Want to be **confident**: 99% correct prob.
  - How many times until you stop? -> **Sample complexity**!
- Each flip costs (a lot if you are lazy).
- Each sample really costs a lot (complex, independent experiments)!

Motivation for distribution testing

- **Fundamental** statistical problem.

- Hypothesis testing – quality control, medical trial etc.

- Checking assumptions for learning algorithms.

Some highlights on our result

- Sample **near-optimal** algorithm for testing **entropy difference**.
  - Entropy test as subtests for high-dim Bayes nets!

Testing Bayesian networks:

- **Fast**: $Poly(n)$ if d is constant.
  - In contrast to **learning Bayesian network**: believed to be NP-hard!

- **No extra assumption** needed!
  - Prior works assume heavy/various assumptions.

Thank you!

- Please come visit our poster if you are interested!