Journal Watch
IEEE Transactions on Information Theory
November 2019 and December 2019

Lekshmi Ramesh

Indian Institute of Science
Bangalore

December 14, 2019

- Distribution testing: Given samples from an unknown distribuion p on $\mathcal{X}^n$, decide whether p has a particular property or is far from having it

- Distribution testing: Given samples from an unknown distribuion p on $\mathcal{X}^n$, decide whether p has a particular property or is far from having it
- Canonical examples:

- Distribution testing: Given samples from an unknown distribuion p on $\mathcal{X}^n$, decide whether p has a particular property or is far from having it
- Canonical examples:
  - Identity testing
    Disinguish w.p. at least 2/3 between $p = q$ and $d_{TV}(p, q) > \varepsilon$ for a fixed known q

- Distribution testing: Given samples from an unknown distribuion p on $\mathcal{X}^n$, decide whether p has a particular property or is far from having it
- Canonical examples:
  - Identity testing
    Disinguish w.p. at least 2/3 between $p = q$ and $d_{TV}(p, q) > \varepsilon$ for a fixed known q
  - Independence testing

- Distribution testing: Given samples from an unknown distribuion p on $\mathcal{X}^n$, decide whether p has a particular property or is far from having it
- Canonical examples:
  - Identity testing
    Disinguish w.p. at least 2/3 between $p = q$ and $d_{TV}(p, q) > \varepsilon$ for a fixed known q
  - Independence testing
- Univariate case requires $\Theta(\sqrt{k}/\varepsilon^2)$ samples , multivariate case requires $\Theta(k^{\frac{n}{2}}/\varepsilon^2)$ samples

- Distribution testing: Given samples from an unknown distribuion p on $\mathcal{X}^n$, decide whether p has a particular property or is far from having it
- Canonical examples:
  - Identity testing
    Disinguish w.p. at least 2/3 between $p = q$ and $d_{TV}(p, q) > \varepsilon$ for a fixed known q
  - Independence testing
- Univariate case requires $\Theta(\sqrt{k}/\varepsilon^2)$ samples , multivariate case requires $\Theta(k^{\frac{n}{2}}/\varepsilon^2)$ samples
- This work: Distribution testing for Ising models

- Ising model: a distribution on $\{-1, +1\}^n$, has three parameters

- Ising model: a distribution on $\{-1, +1\}^n$, has three parameters
    - A graph $G = (V, E)$

- Ising model: a distribution on $\{-1, +1\}^n$, has three parameters
  - A graph $G = (V, E)$
  - Edge parameters θ_{uv}

- Ising model: a distribution on $\{-1,+1\}^n$, has three parameters
  - A graph $G = (V, E)$
  - Edge parameters θ_{uv}
  - Node parameters θ_v

$$p(x) \propto \exp\left(\sum_{v \in V} \theta_v x_v + \sum_{(u,v) \in E} \theta_{uv} x_u x_v\right)$$

- Ising model: a distribution on $\{-1, +1\}^n$, has three parameters
  - A graph $G = (V, E)$
  - Edge parameters θ_{uv}
  - Node parameters θ_v

$$p(x) \propto \exp\left(\sum_{v \in V} \theta_v x_v + \sum_{(u,v) \in E} \theta_{uv} x_u x_v\right)$$

- Main result: sample complexity bounds for Ising model identity testing and independence testing

- Ising model: a distribution on $\{-1, +1\}^n$, has three parameters
  - A graph $G = (V, E)$
  - Edge parameters θ_{uv}
  - Node parameters θ_v

$$p(x) \propto \exp\left(\sum_{v \in V} \theta_v x_v + \sum_{(u,v) \in E} \theta_{uv} x_u x_v\right)$$

- Main result: sample complexity bounds for Ising model identity testing and independence testing
- For tree structured models, independence testing requires $O(n/\varepsilon)$ samples

- Recovery of sparse, binary signals using biased measurement matrices

- Recovery of sparse, binary signals using biased measurement matrices
- Measurement matrix entries usually assumed to be drawn from centered distributions

- Recovery of sparse, binary signals using biased measurement matrices
- Measurement matrix entries usually assumed to be drawn from centered distributions
- For recovering binary signals, biased entries can lead to better performance

- Recovery of sparse, binary signals using biased measurement matrices
- Measurement matrix entries usually assumed to be drawn from centered distributions
- For recovering binary signals, biased entries can lead to better performance
- Basis pursuit with box constraints

$$\min_x \|x\|_1 \text{ subject to } Ax = b, x \in [0,1]^N$$

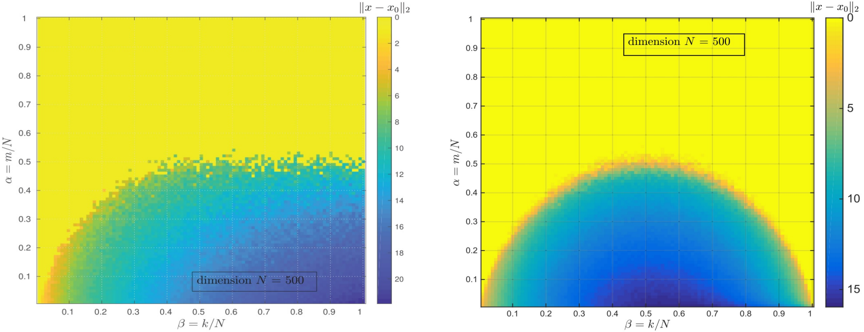

Figure 1: Recovery performance (a) Gaussian measurement matrix, (b) Bernoulli measurement matrix.

- Main result: Analysis of box constrained BP for biased A

$$A = \mu \mathbb{1} + D,$$

where μ is the bias and D has centered, subgaussian entries

- Main result: Analysis of box constrained BP for biased A

$$A = \mu \mathbb{1} + D,$$

  where μ is the bias and D has centered, subgaussian entries

- For x_0 binary and k-sparse and $y = Ax_0$, x_0 is the unique solution to box constrained BP whp if

$$m \geq \max\left\{\frac{R^2}{\mu^2}, \min\{k, N-k\}\right\} \log N$$

- Main result: Analysis of box constrained BP for biased A

$$A = \mu \mathbb{1} + D,$$

  where μ is the bias and D has centered, subgaussian entries

- For x_0 binary and k-sparse and $y = Ax_0$, x_0 is the unique solution to box constrained BP whp if

$$m \geq \max\left\{\frac{R^2}{\mu^2}, \min\{k, N-k\}\right\} \log N$$

- Under the same assumption, x_0 can be recovered by solving

$$\min_x \|Ax - b\|_2 \text{ subject to } x \in [0,1]^N$$

- The privacy-utility tradeoff

- The privacy-utility tradeoff
    - User shares data with analyst and receives some utility

- The privacy-utility tradeoff
  - User shares data with analyst and receives some utility
  - Involves privacy risk since analyst can make additional inferences

Privacy With Estimation Guarantees

Hao Wang, Lisa Vo, Flavio P. Calmon, Muriel Médard, Ken R. Duffy, and Mayank Varia

- The privacy-utility tradeoff
  - User shares data with analyst and receives some utility
  - Involves privacy risk since analyst can make additional inferences
  - Apply privacy preserving mechanism to data before sharing, while guaranteeing utility

Privacy With Estimation Guarantees

Hao Wang, Lisa Vo, Flavio P. Calmon, Muriel Médard, Ken R. Duffy, and Mayank Varia

- The privacy-utility tradeoff
  - User shares data with analyst and receives some utility
  - Involves privacy risk since analyst can make additional inferences
  - Apply privacy preserving mechanism to data before sharing, while guaranteeing utility
- Examples: Users sharing movie ratings, medical records shared for learning patterns

- Formulation

- Formulation
  - A hidden variable S (deemed private by user)

- Formulation
  - A hidden variable S (deemed private by user)
  - A useful variable X that depends on S

- Formulation
  - A hidden variable S (deemed private by user)
  - A useful variable X that depends on S
  - A privacy preserving mapping is applied to X to obtain Y

- Formulation
  - A hidden variable S (deemed private by user)
  - A useful variable X that depends on S
  - A privacy preserving mapping is applied to X to obtain Y
  - Y is disclosed to analyst

- Formulation
    - A hidden variable S (deemed private by user)
    - A useful variable X that depends on S
    - A privacy preserving mapping is applied to X to obtain Y
    - Y is disclosed to analyst
- Main result: A χ^2 privacy utility function that captures how well analyst can reconstruct functions of X while restricting his ability to reconstruct functions of S

- Formulation
  - A hidden variable S (deemed private by user)
  - A useful variable X that depends on S
  - A privacy preserving mapping is applied to X to obtain Y
  - Y is disclosed to analyst
- Main result: A χ^2 privacy utility function that captures how well analyst can reconstruct functions of X while restricting his ability to reconstruct functions of S
- For finite $\mathcal{X}$ and $\mathcal{Y}$, $\chi^2(X,Y)$ related to singular values of the matrix $D_X^{-\frac{1}{2}} \mathrm{P}_{XY} D_Y^{-\frac{1}{2}}$ where

  $\mathrm{P}_{XY} \in \mathbb{R}^{|\mathcal{X}| \times |\mathcal{Y}|}$ has entries $\mathrm{P}_{XY}(i,j)$,
  D_X and D_Y are diagonal with marginals of X and Y on diagonal

Other interesting papers

- Snapshot Compressed Sensing: Performance Bounds and Algorithms. *S. Jalali and X. Yuan*
- Confidence Region of Singular Subspaces for Low-Rank Matrix Regression. *D. Xia*
- Statistical Mechanics of MAP Estimation: General Replica Ansatz. *A. Bereyhi, R. R. Müller, and H. Schulz-Baldes*
- Analysis of Approximate Message Passing With Non-Separable Denoisers and Markov Random Field Priors. *Y. Ma, C. Rush, and D. Baron*