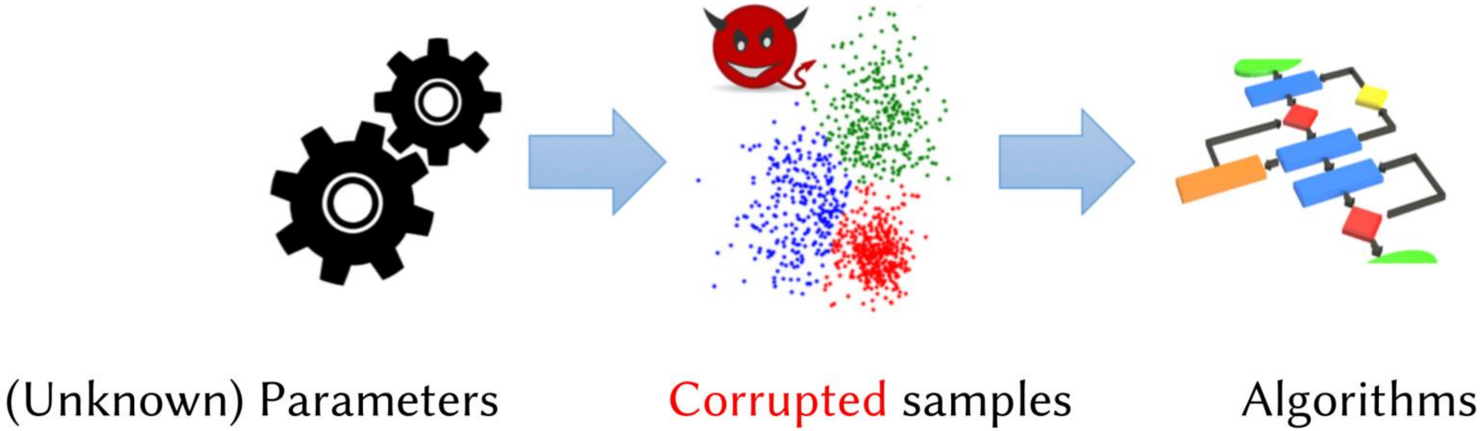
Robust Learning of Fixed-Structure Bayesian Network in Nearly-Linear Time

Yu Cheng
University of Illinois at Chicago

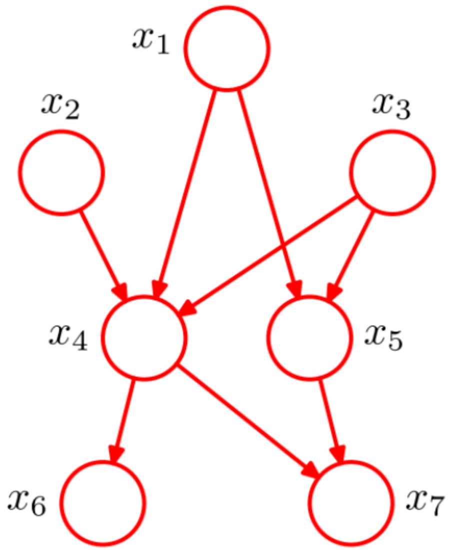
Honghao Lin
Shanghai University of Finance and Economics

Robust Parameter Estimation



Q: Can we design provably robust and efficient learning algorithms when a small fraction of the data is corrupted?

Bayesian Network



A Bayesian Network is a d-dimensional distribution supported on $\{0,1\}^d$, where $\Pr[x_i = 1 \mid x_1, \ldots, x_{i-1}]$ depends only on the values x_j where j is a parent of i.

Input: N samples $\{X_1, \ldots, X_N\}$ drawn from a Bayesian Network P on $\mathbb{R}^d$.

Goal: Learn a Bayesian Network Q such that $d_{TV}(P, Q)$ is small.

Corruption Model

- N samples are drawn from an unknown distribution P.

- Adversary replaces ϵN samples with arbitrary points (after inspecting P, the samples, and the algorithm).

- Stronger than Huber's contamination model.

  - Adaptive, and allows both additive and subtractive errors

Corruption Model

- Consider the simplest case: P is a binary product distribution with all $p_i = \frac{1}{2}$.

- The binary product distribution is a Bayesian Network with an empty dependence graph.

- Under the ϵ-corruption model, the empirical estimator may output a Q such that $||p - q||_2 = \epsilon\sqrt{d}$, where the error is dependent with the dimension d.

Previous Result

	$\mathrm{d_{TV}(P, Q)}(\delta)$	# of Samples (N)	Runtime
Filter[CDKS18]	$O(\epsilon\sqrt{\log(1/\epsilon)})$	$\tilde{O}\ (m/\delta\text{^}2)$	$\tilde{O}\ (Nd^2/\epsilon)$

- $m = \sum_i 2^{\lvert \mathrm{parent}(i) \rvert}$ is the size of the conditional probability table.
- Under the assumption: P is c-balanced($p_i \in [c, 1-c]$), the minimum probability of parental configuration of $P \geq \alpha$, where c and α are constants.
- The error and sample complexity are optimal up to polylog factor.

Yu Cheng et al. Robust Learning of Fixed-Structure Bayesian Networks. NeurIPS 2018

Question: Can we design a robust algorithm for learning Bayesian networks in the fixed structure setting that runs in nearly-linear time?

Our Result

	$\mathrm{d_{TV}(P, Q)}(\delta)$	# of Samples (N)	Runtime
Filter[CDKS18]	$O(\epsilon\sqrt{\log(1/\epsilon)})$	$\tilde{O}\ (m/\delta\hat{}2)$	$\tilde{O}\ (Nd^2/\epsilon)$
Our Result	$O(\epsilon\sqrt{\log(1/\epsilon)})$	$\tilde{O}\ (m/\delta\hat{}2)$	$\tilde{O}\ (Nd)$

- We give the first nearly-linear time algorithm for this problem with a dimension-independent error guarantee (Under the same assumption of α and c) .
- We show that robust learning of Bayes nets can be essentially reduced to robust mean estimation, and as a result, this allows us to take advantage of the recent (and future) advances in robust mean estimation.

Overview of Our Approach

- We map each sample $X \in \mathbb{R}^d$ to some $f(X, q) \in \mathbb{R}^m$, where q is our current estimation of the conditional probability table.

- If we learn a good estimation u of the mean of $f(X, q)$ to good accuracy, we can use u to recover p.

- We show that the second moment of this distribution depends on its first moment in a way that allow us to use robust mean estimation algorithm on $f(X, q)$.

Overview of Our Approach

- In each iteration round, we run the robust mean estimation algorithm on $f(X,q)$, then use the output to recover a new estimation of q for the next round.

- We show in each round we can achieve a small error or reduce the error by a constant factor.

Overview of Our Approach

- We give a robust mean estimation algorithm that runs in time nearly linear in the total number of nonzero entries in the input, based on the algorithm in [DHL19].

- The reason we need it: if we naively run the robust mean estimation algorithm on $f(X, q)$. The runtime will be $\tilde{O}\ (Nm)$, where m can be exp (d). But notice that $f(X, q)$ is d-sparse.

- The main computation bottleneck of [DHL19] is the matrix multiplicative weights update using JL Lemma, which we show can be done in nearly input-sparsity time.

Yihe Dong et al. Quantum Entropy Scoring for Fast Robust Mean Estimation and Improved Outlier Detection. NeurIPS 2019

Open Question

1. Remove the assumption of the balance condition and minimum probability of parental configuration.

2. Robustly Learning Bayes Nets when the graph structure is unknown.