Learning Exponential Families from Truncated Samples

Jane Lee

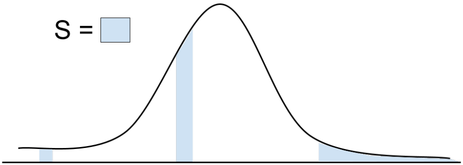
Joint work with Manolis Zampetakis and Andre Wibisono

Yale University

Background & Prior Work

Truncated Densities

Let $\rho^S := \rho(\cdot \mid \cdot \in S)$ be the conditional distribution of $x \sim \rho$ given that $x \in S$. That is, $\rho^S(x) = \frac{\rho(x) \cdot \mathbb{1}\{x \in S\}}{\rho(S)}$.



Prior Work

- A recent line of work provides the first efficient estimation algorithms for the parameters of a Gaussian distribution, linear regression with Gaussian noise, LDS, etc.
- These works rely on properties of Gaussian distributions.

Problem Statement

We are given truncated samples $\{x_i\}_{i=1}^n$, each $x_i \sim p_{\theta^*}^S$, where $p_{\theta^*}(S) = \alpha > 0$. Only accessing the truncation set S via a membership oracle, can one recover θ^* and thus p_{θ^*} **(computationally) efficiently**?

Our Result

We can recover θ^* efficiently from truncated samples from a high-dimensional exponential family distribution (under some assumptions).

Main Ingredients

In order to have an efficient procedure for which extrapolation is possible, we need to address these statistical and algorithmic challenges.

- We need to ensure the steps of a projected SGD (PSGD) procedure are efficient, and terminates in time polynomial in $(m, k, 1/\epsilon)$ (where $x \in \mathbb{R}^m$, $\theta \in \mathbb{R}^k$, ϵ is accuracy parameter).
- Strong convexity and smoothness of the truncated negative log-likelihood objective (in θ) depend on $p_\theta(S)$.
- Given that $p_{\theta^*}(S) = \alpha$, we can lower bound $p_\theta(S)$ in terms of $\|\theta - \theta^*\|$.
- We can design a procedure to find an initial parameter θ_0 so that $\|\theta_0 - \theta^*\|$ is small and project to a neighborhood around θ_0.

Implications

The current work has a few important implications:

- Our assumptions are met by exponential, Weibull, continuous Bernoulli, continuous Poisson, Gaussian distributions, and certain generalized linear models.
- Combined with ideas of a statistical Taylor theorem (prior work), we can learn log-concave distributions $\rho(x) = \exp(-f(x))$ by replacing $f(x) \approx \sum_i a_i t_i(x)$ by finite Taylor approximation.
- Given initial truncated examples, we can generate data from the non-truncated distribution (with small error).

Thank You