Oracle Efficient Truncated Statistics

Presented by Konstantinos Karatapanis

Joint work with Vasilis Kontonis and Christos Tzamos

Archimedes, Athena Research Center, Greece
National Technical University of Athens
University of Texas at Austin
University of Athens

ICLR 2025

Background & Motivation

- **Truncated Densities:** For an exponential family

$$p_\theta(x) = h(x)\exp\big(\theta^\top T(x) - A(\theta)\big),$$

the truncated density on a survival set S is

$$p_\theta^S(x) = \frac{p_\theta(x)\mathbf{1}_S(x)}{p_\theta(S)} \quad \text{with} \quad p_\theta(S) = \int_S p_\theta(x)\,dx.$$

- **Historical Context:** Truncated statistics have a long history—from early statistical analyses to modern inference problems.
- **Recent Work:** Lee et al. have recently proposed efficient algorithms for learning from truncated samples. Our work expands on theirs by relaxing several assumptions and improving the oracle complexity.

Problem Statement & Main Result

- **Problem Statement:** Given n i.i.d. samples from the truncated distribution $p^S_{\theta^*}$, with survival mass α, can we efficiently recover the true parameter θ^*?

Problem Statement & Main Result

- **Problem Statement:** Given n i.i.d. samples from the truncated distribution $p^S_{\theta^*}$, with survival mass α, can we efficiently recover the true parameter θ^*?

- **Main Result:** There exists an algorithm that uses few samples and corresponding oracle calls to yield an estimator $\hat{\theta}$, that can be computed efficiently satisfying

$$\mathrm{KL}(p^S_{\theta^*} \parallel p^S_{\hat{\theta}}) \leq \varepsilon,$$

with high probability.

Optimization Framework

- **Parameter Recovery:** We recover the true parameter as the minimizer of the truncated negative log-likelihood:

$$\theta^* = \arg\min_{\theta \in \Theta} \mathcal{L}_S(\theta),$$

where $\mathcal{L}_S(\theta) = -\mathbb{E}_{x \sim p^S_{\theta^*}} \left[\log p^S_\theta(x) \right].$

Optimization Framework

- **Parameter Recovery:** We recover the true parameter as the minimizer of the truncated negative log-likelihood:

$$\theta^* = \arg\min_{\theta \in \Theta} \mathcal{L}_S(\theta),$$

where $\mathcal{L}_S(\theta) = -\mathbb{E}_{x \sim p^S_{\theta^*}} \left[\log p^S_\theta(x) \right]$.

- **Gradient and Hessian:** The gradient of $\mathcal{L}_S$ is given by

$$\nabla \mathcal{L}_S(\theta) = \mathbb{E}_{x \sim p^S_\theta}[T(x)] - \mathbb{E}_{x \sim p^S_{\theta^*}}[T(x)],$$

and its Hessian is $\nabla^2 \mathcal{L}_S(\theta) = \mathrm{Cov}_{x \sim p^S_\theta}[T(x)]$.

Optimization Framework

- **Parameter Recovery:** We recover the true parameter as the minimizer of the truncated negative log-likelihood:

$$\theta^* = \arg\min_{\theta \in \Theta} \mathcal{L}_S(\theta),$$

where $\mathcal{L}_S(\theta) = -\mathbb{E}_{x \sim p^S_{\theta^*}} \left[\log p^S_\theta(x) \right]$.

- **Gradient and Hessian:** The gradient of $\mathcal{L}_S$ is given by

$$\nabla \mathcal{L}_S(\theta) = \mathbb{E}_{x \sim p^S_\theta}[T(x)] - \mathbb{E}_{x \sim p^S_{\theta^*}}[T(x)],$$

and its Hessian is $\nabla^2 \mathcal{L}_S(\theta) = \mathrm{Cov}_{x \sim p^S_\theta}[T(x)]$.

- **Dependence on Survival Mass:** The smoothness and variance properties of these quantities—and consequently the convergence behavior of our SGD algorithm—depend critically on the survival mass $p_\theta(S)$. Efficient gradient estimation by rejection sampling is achievable only when $p_\theta(S)$ is large enough.

General Proof Technique

1. **Initialization:** We start at the non-truncated minimizer

$$\theta_0 = \arg\min_{\theta \in \Theta} \mathcal{L}(\theta),$$

   which can be well approximated by the empirical mean of $T(x)$ from the truncated samples, i.e., $\theta_0 \approx \mathbb{E}_{x \sim p^S_{\theta^*}}[T(x)]$. This initialization ensures a high survival mass.

General Proof Technique

1. **Initialization:** We start at the non-truncated minimizer

$$\theta_0 = \arg\min_{\theta \in \Theta} \mathcal{L}(\theta),$$

   which can be well approximated by the empirical mean of $T(x)$ from the truncated samples, i.e., $\theta_0 \approx \mathbb{E}_{x \sim p^S_{\theta^*}}[T(x)]$. This initialization ensures a high survival mass.

2. **Expanding the Domain:** From θ_0, we define a convex domain

$$D = \{\theta \in \Theta : \mathcal{L}(\theta) - \mathcal{L}(\theta_0) \leq \Delta\},$$

   with $\Delta = \log(1/\alpha)$ in our presentation. In D, we guarantee that $\theta^* \in D$ and $p_\theta(S) \geq \Omega(\alpha^2)$.

General Proof Technique

1. **Initialization:** We start at the non-truncated minimizer

$$\theta_0 = \arg\min_{\theta \in \Theta} \mathcal{L}(\theta),$$

   which can be well approximated by the empirical mean of $T(x)$ from the truncated samples, i.e., $\theta_0 \approx \mathbb{E}_{x \sim p^S_{\theta^*}}[T(x)]$. This initialization ensures a high survival mass.

2. **Expanding the Domain:** From θ_0, we define a convex domain

$$D = \{\theta \in \Theta : \mathcal{L}(\theta) - \mathcal{L}(\theta_0) \leq \Delta\},$$

   with $\Delta = \log(1/\alpha)$ in our presentation. In D, we guarantee that $\theta^* \in D$ and $p_\theta(S) \geq \Omega(\alpha^2)$.

3. **Refined Domain Expansion** A more careful, multi-stage (nested) optimization—i.e., gradually increasing Δ from 0 to $\log(1/\alpha)$—can improve the survival mass guarantee to

$$p_\theta(S) \geq \Omega(\alpha).$$

Thank you!

Questions?

k.karatapanis@athenarc.gr