Finite-Sample Symmetric Mean Estimation with Fisher Information Rate

Shivam Gupta [1] Jasper C.H. Lee [2] Eric Price [1]

[1]The University of Texas at Austin [2]University of Wisconsin–Madison

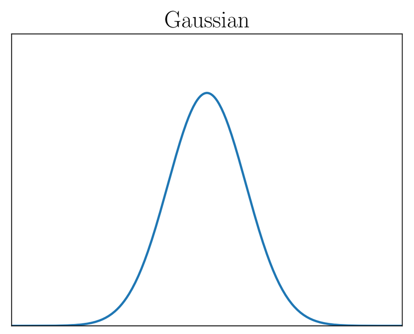
Mean Estimation

- Given n samples from an unknown variance-σ^2 distribution, the empirical mean has variance $\frac{\sigma^2}{n}$ and asymptotically converges to a Gaussian
- In the *finite-sample* setting, **[Catoni; 2012], [Lee, Valiant; 2022]** show estimator $\hat{\mu}$ of the mean μ, such that with probability $1-\delta$,

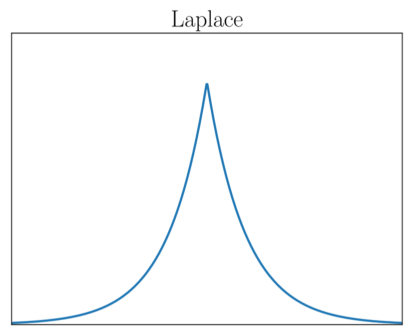
$$|\hat{\mu} - \mu| \le \sigma\sqrt{\frac{2\log\frac{2}{\delta}}{n}}(1+o(1))$$

- **Natural Question:** Is it possible to do better?

Location Estimation

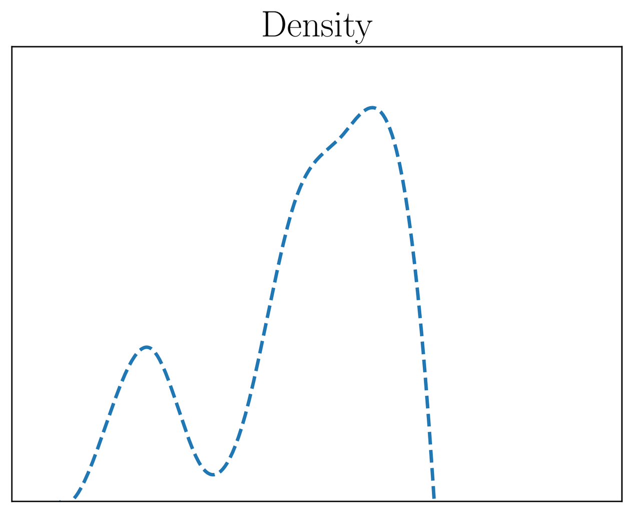


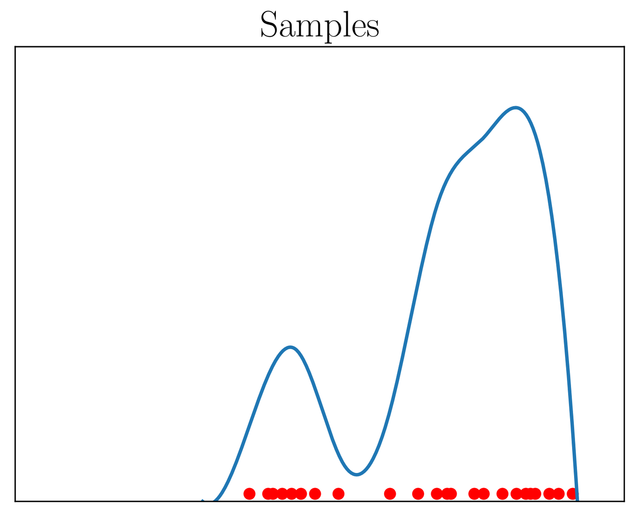
- Given n samples from a **Gaussian** with variance σ^2, optimal estimator is the empirical mean, which has $1-\delta$ confidence radius $\sigma\sqrt{\frac{2\log\frac{1}{\delta}}{n}}$



- For the **Laplace** distribution, the median achieves error $\sigma\sqrt{\frac{\log\frac{1}{\delta}}{n}}$, a factor $\sqrt{2}$ savings over the above

Given a density f (up to shift), and n samples $X_1, \dots, X_n$, what is the best estimator of the mean? **Mean Estimation with known density**.

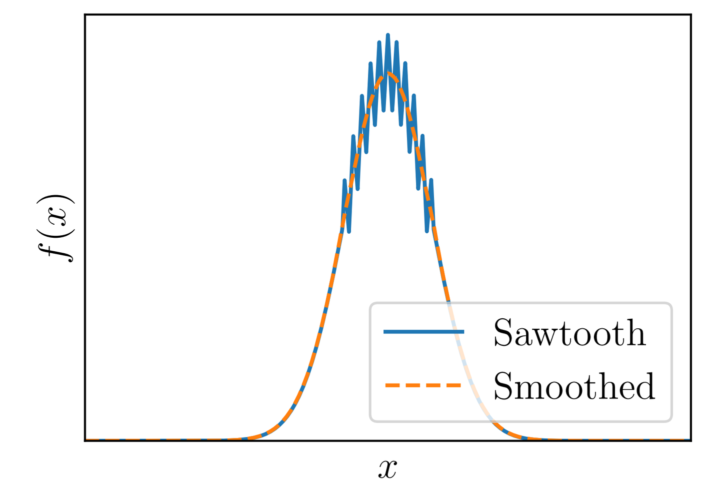


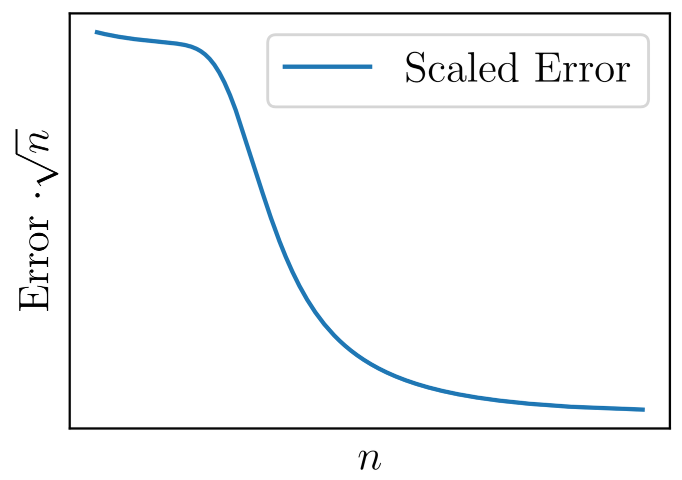


- Fit samples to density, aka **Maximum Likelihood Estimation (MLE)**
- Asymptotic convergence to Gaussian with variance $\mathcal{I}^{-1}/n$, where $\mathcal{I}$ is the **Fisher Information**
- Tight: Cramér-Rao bound says must have variance at least $\mathcal{I}^{-1}/n$

Finite-Sample Setting – Smoothed MLE

- When density known, might expect $|\hat{\mu} - \mu| \le \sqrt{\frac{2\log\frac{2}{\delta}}{n\mathcal{I}}}$. Unfortunately, impossible!
- **Solution:** Smoothing **[Gupta, Lee, Price, Valiant; NeurIPS 2022]**



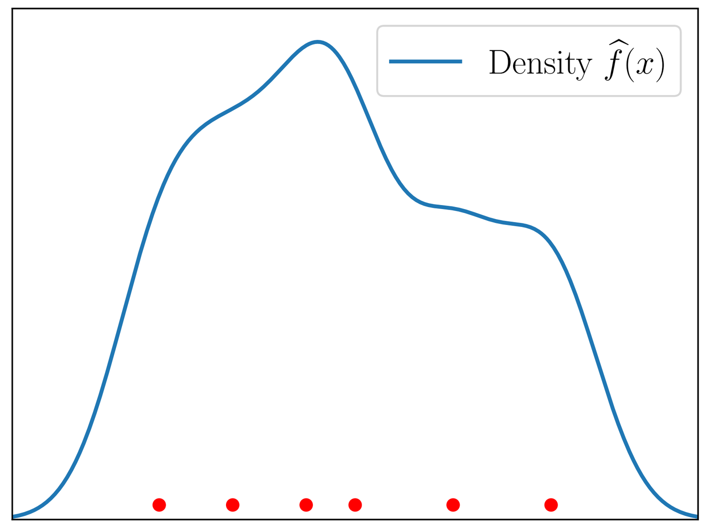


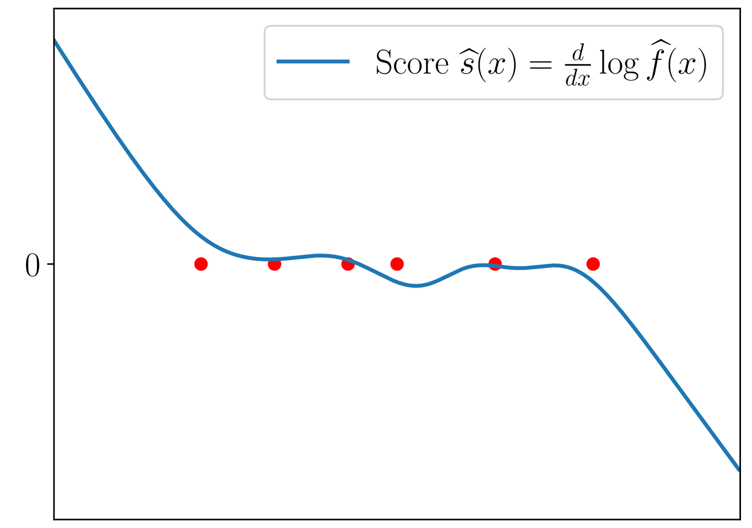
Smooth samples and distribution with a radius $r \approx \sigma/n^{1/6}$ Gaussian, then run (variant of) MLE. With probability $1-\delta$,

$$|\hat{\mu} - \mu| \le \sqrt{\frac{2\log\frac{2}{\delta}}{n\mathcal{I}_r}}(1+o(1))$$

Symmetric Mean Estimation

- If we don't know the density, is it still possible to get Fisher Information rate? In general, no. **[Dang, Lee, Song, Valiant; 2023]**. However, if the distribution is **symmetric**, yes!
- **Idea:** Estimate density using the Kernel Density Estimate (KDE) on the first (say) $n^{1/100}$ samples

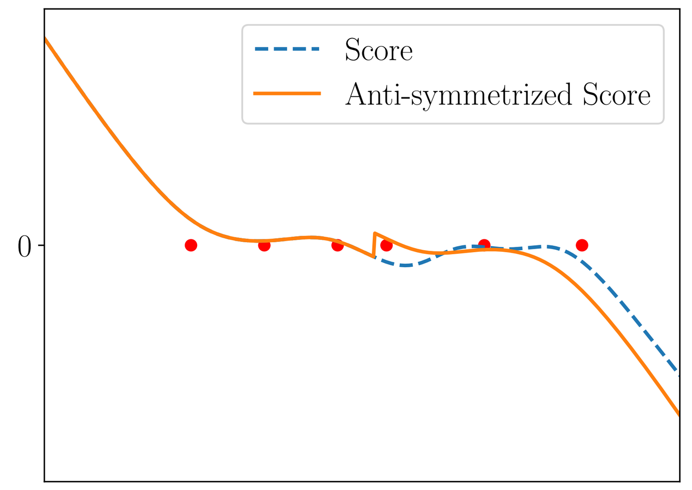




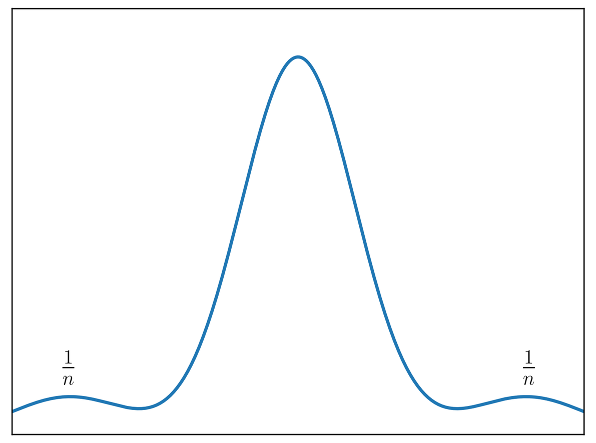
- Naive algorithm: Run (smoothed) MLE/Find zero of estimated score close to initial estimate of mean. Two issues:
  1. Bias
  2. Inability to estimate score well in "atypical" regions

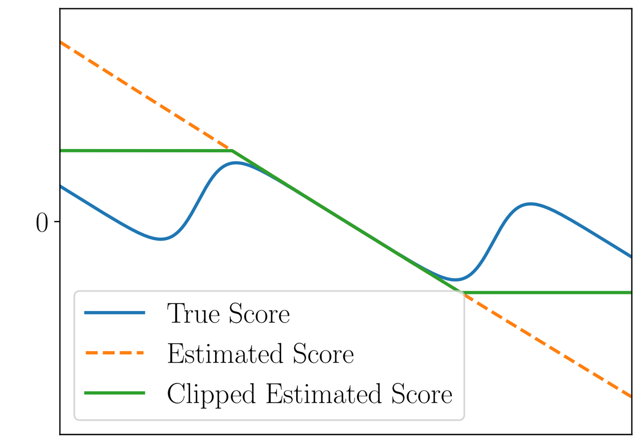
Correcting the KDE

1. Bias
   - Performing MLE wrt a different distribution introduces bias
   - For a *symmetric* distribution, MLE with respect to *any* (possibly different) symmetric distribution is an unbiased estimator
   - **Idea:** Anti-symmetrize the KDE score.



2. Atypical regions
   - Consider the distribution $\frac{n-2}{n}\mathcal{N}(0,1) + \frac{1}{n}\mathcal{N}(c,1) + \frac{1}{n}\mathcal{N}(-c,1)$.





   - Since we use only the first $n^{1/100}$ samples to compute the KDE, we only see samples from the central Gaussian with high probability
   - This score just corresponds to the empirical mean, which can be arbitrarily bad for the remaining $n \cdot (1-o(1))$ samples
   - **Solution:** Clipping

Summary

- Use the first (say) $n^{1/100}$ samples to compute the KDE
- Anti-symmetrize and clip the KDE score appropriately
- Run (variant of) smoothed MLE using the anti-symmetrized and clipped KDE score on remaining samples

Let $\mathcal{I}_r$ be the r-smoothed Fisher information. For large enough r decaying polynomially in n, with probability $1-\delta$, our estimator $\hat{\mu}$ satisfies

$$|\hat{\mu} - \mu| \le (1+o(1))\sqrt{\frac{2\log\frac{2}{\delta}}{n\mathcal{I}_r}}$$

shivamgupta@utexas.edu