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

Asymptotic Mean Estimation

- Given n samples from a distribution, want to estimate mean μ.

	Estimator	Converges to	Notes
Unknown Distribution	Empirical Mean	$\mathcal{N}(\mu, \frac{\sigma^2}{n})$	Central Limit Theorem
Known Distribution	MLE	$\mathcal{N}(\mu, \frac{1}{n\mathcal{I}})$	$\mathcal{I}$ is the *Fisher Information*
Unknown *Symmetric* Distribution	KDE + MLE	$\mathcal{N}(\mu, \frac{1}{n\mathcal{I}})$	[Stone; 1975]

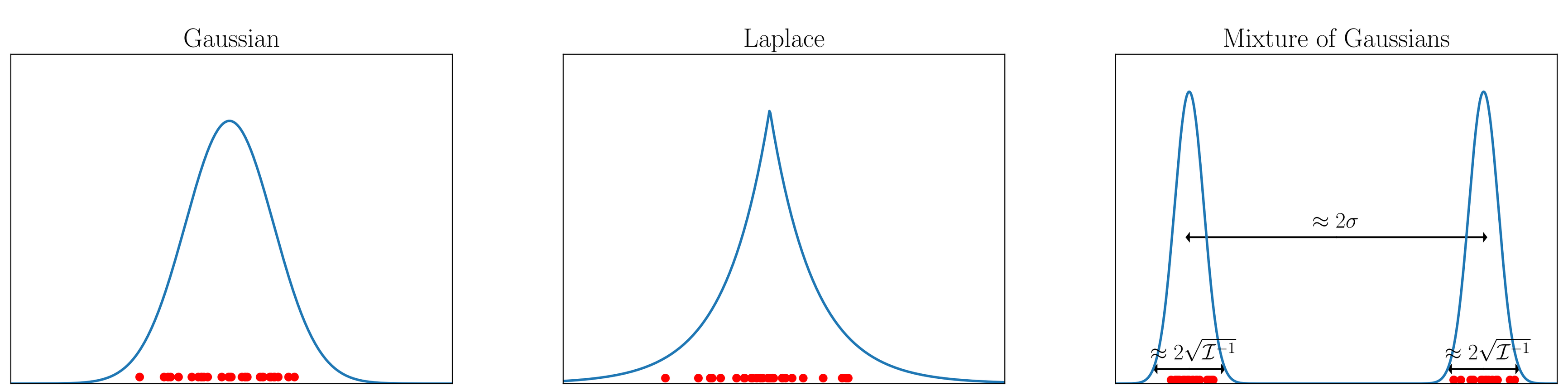
Table 1. Classical Asymptotic Results

- In ***finite-sample*** setting, when distribution is **unknown**, [Catoni; 2012], [Lee, Valiant; 2022] show estimator $\widehat{\mu}$ such that with probability $1-\delta$,

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

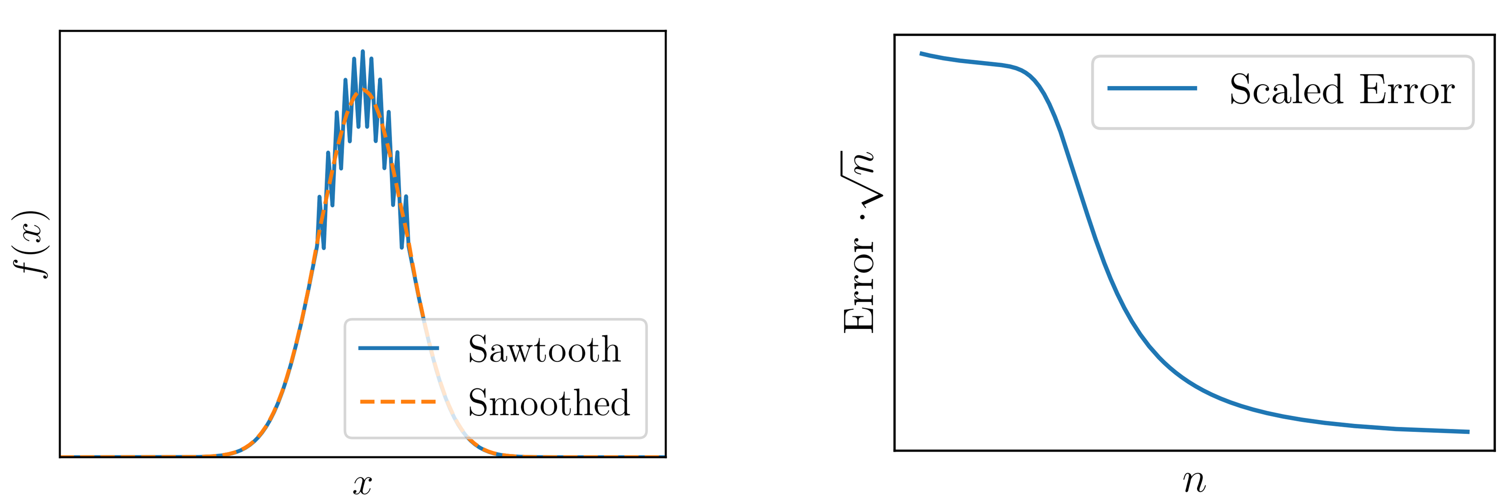
- **Natural Question:** What if distribution is **known/symmetric**?

Location Estimation (Known Distribution Case)



$\frac{1}{\mathcal{I}} = \sigma^2$ $\qquad$ $\frac{1}{\mathcal{I}} = \frac{\sigma^2}{2}$ $\qquad$ $\frac{1}{\mathcal{I}} \ll \sigma^2$

- **Finite-Sample Setting:** Might expect $|\widehat{\mu} - \mu| \le \sqrt{\frac{2 \log \frac{2}{\delta}}{n\mathcal{I}}}$. Unfortunately, impossible!
- **Solution:** Smoothing [Gupta, Lee, Price, Valiant; NeurIPS 2022]



Smooth with radius $r \approx \sigma/n^{1/6}$ Gaussian, then run MLE. With prob. $1-\delta$,

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

Finite-Sample Mean Estimation

	Error	Notes
Unknown Distribution	$\sqrt{\frac{2\sigma^2 \log \frac{2}{\delta}}{n}}$	[Cat12, LV22]
Known Distribution	$\sqrt{\frac{2 \log \frac{2}{\delta}}{n\mathcal{I}_r}}$	[GLPV22], $\mathcal{I}_r$ is the *smoothed* Fisher Information
Unknown *Symmetric* Distribution	$\sqrt{\frac{2 \log \frac{2}{\delta}}{n\mathcal{I}_r}}$	This paper

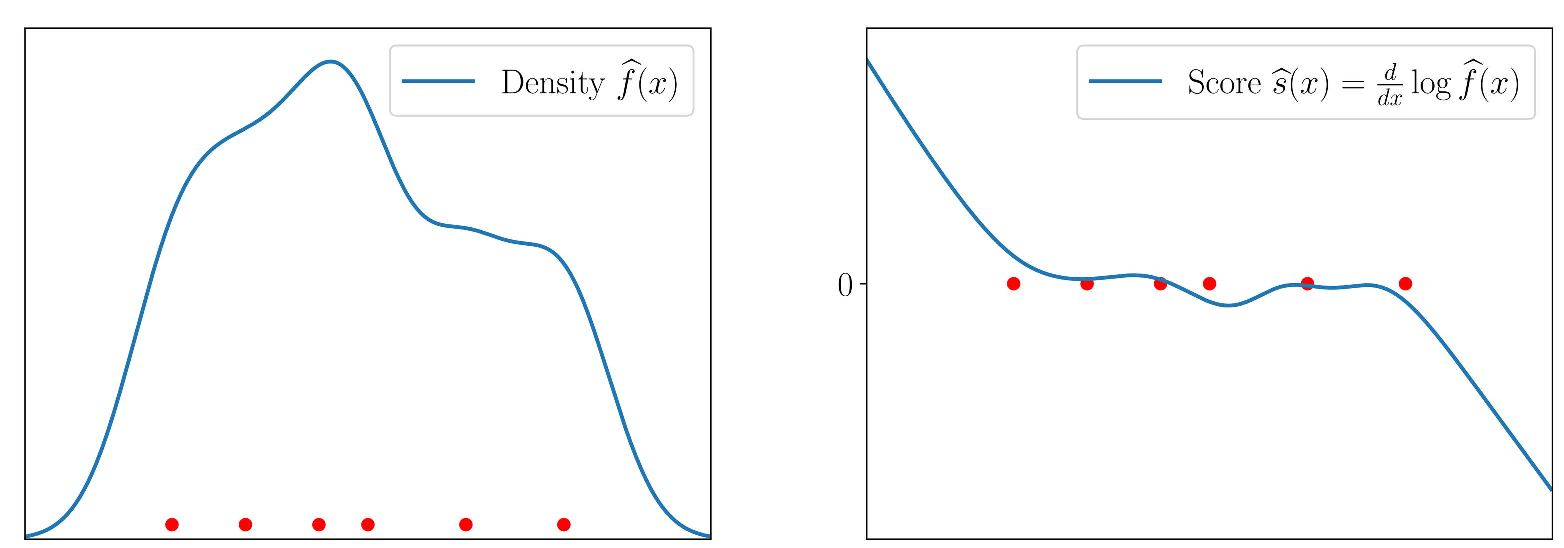
Table 2. Finite-Sample Results

Main Theorem (Informal): For $r \approx \sigma/n^{1/13}$, our estimator $\widehat{\mu}$ given n samples from a *symmetric* distribution satisfies, with probability $1-\delta$,

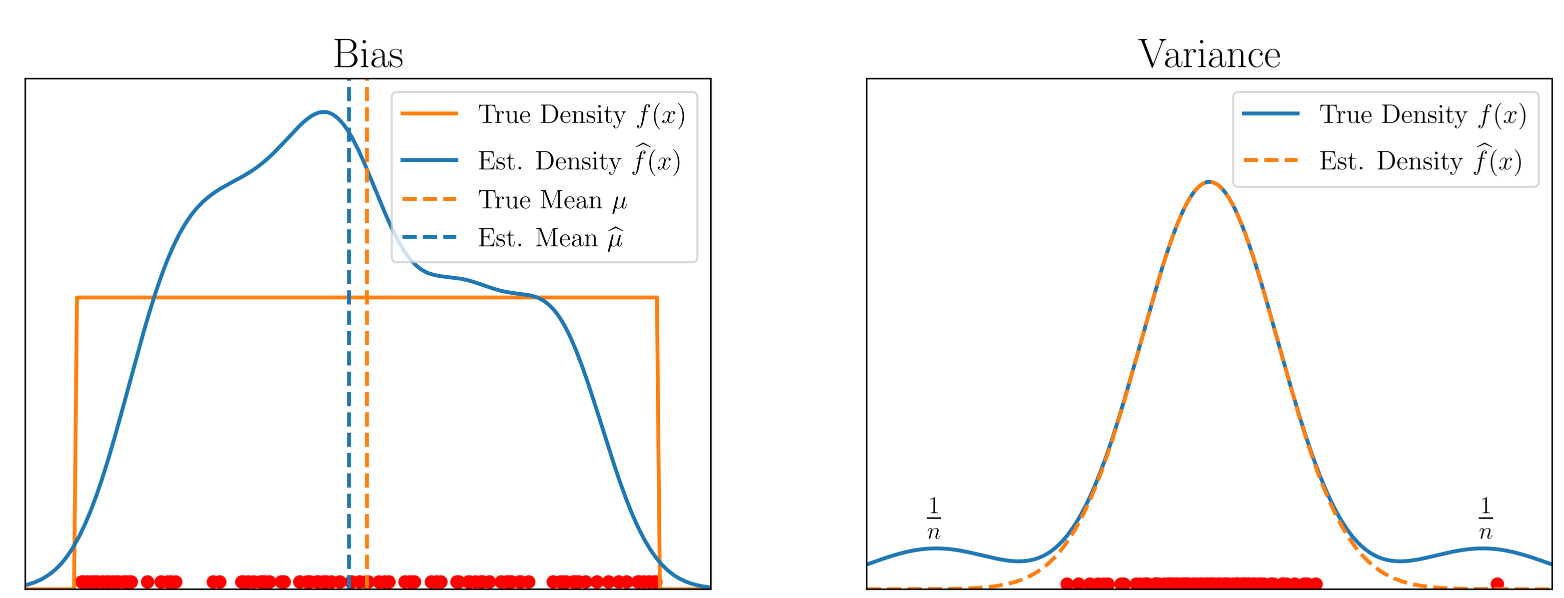
$$|\widehat{\mu} - \mu| \le (1 + o(1))\sqrt{\frac{2 \log \frac{2}{\delta}}{n\mathcal{I}_r}}$$

The $o(1)$ depends on δ, n, but is independent of the distribution.

- **Idea:** Estimate density using the Kernel Density Estimate (KDE) on the first (say) $n^{1/100}$ samples

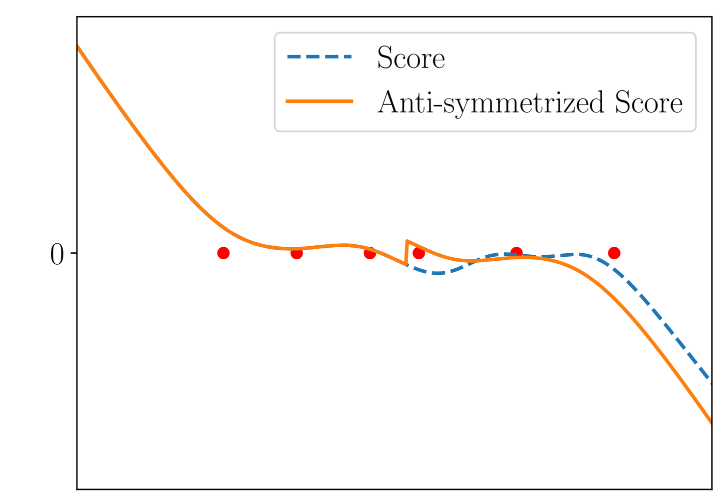


- Naive algorithm: Run (smoothed) MLE on the KDE/Find zero of KDE score using remaining samples. Two issues:
  1. Bias
  2. Variance



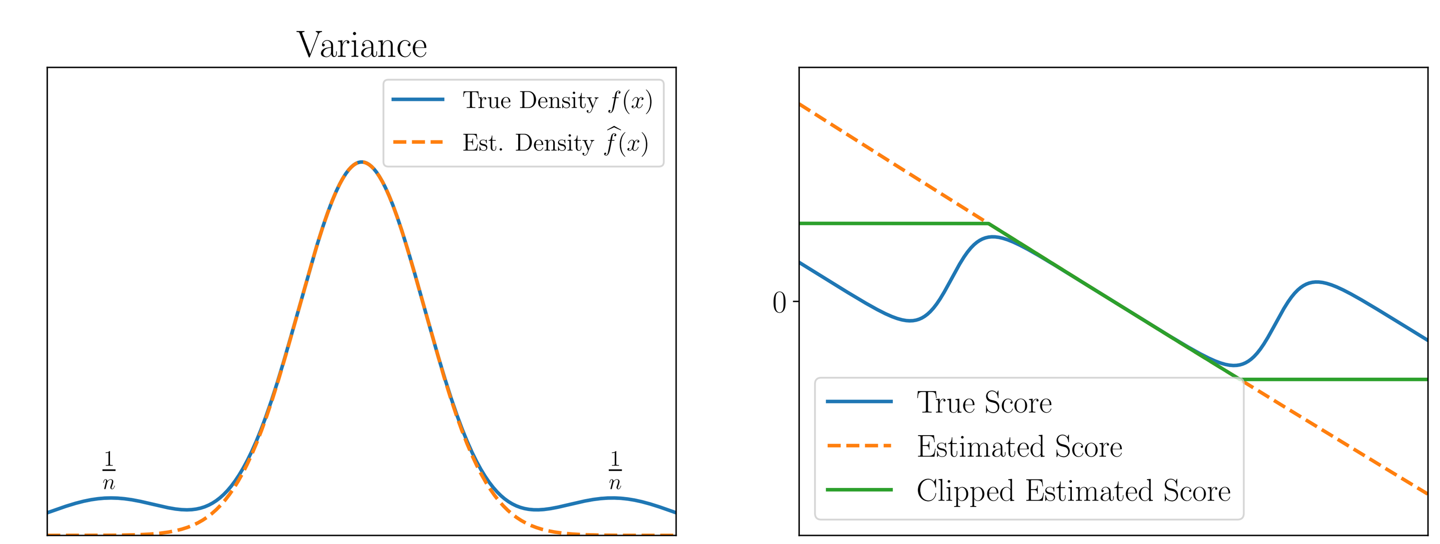
Correcting the KDE – Bias

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



Correcting the KDE – Variance

- In example, true score is close to 0 near bumps, while estimated score is large
- **Solution:** Clip the score



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

Main Theorem (Informal): For $r \approx \sigma/n^{1/13}$, our estimator $\widehat{\mu}$ given n samples from a *symmetric* distribution satisfies, with probability $1-\delta$,

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

shivamgupta@utexas.edu