Theoretical guarantees for EM under misspecified Gaussian mixture models

Raaz Dwivedi*, Nhat Ho*, Koulik Khamaru*, Martin Wainwright and Michael Jordan

{raaz.rsk, minhnhat, koulik, wainwrig}@berkeley.edu, jordan@cs.berkeley.edu

THE VOLEON GROUP | Berkeley EECS Electrical Engineering and Computer Sciences | University of California, Berkeley Department of Statistics

Objective

Goal: Understand parameter estimation for mixture models when **the number of mixtures is not correctly-specified**

Model set-up:

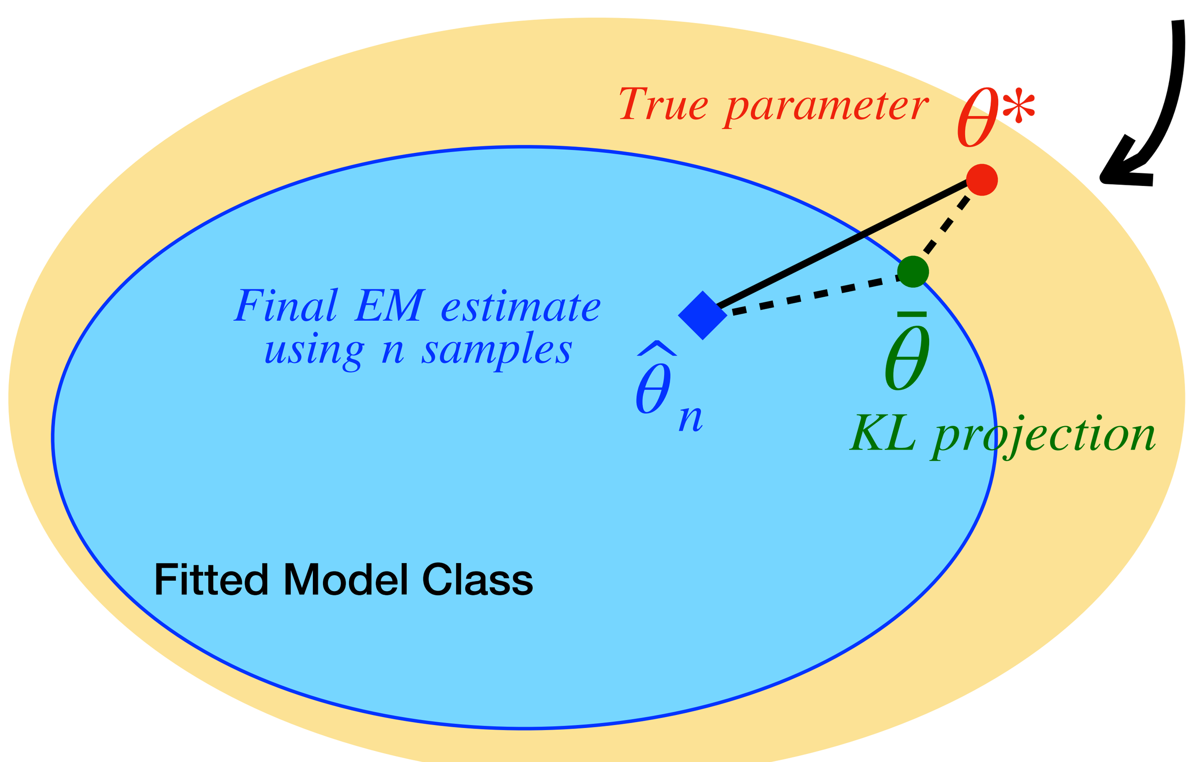
$$\text{True Model:}\quad \mathbb{P}_{\theta^*} = \sum_{i=1}^{\mathbf{k}^*} \pi_i \mathcal{N}(\theta_i^*, \sigma^2)$$

$$\text{Fitted Model:}\quad \mathbb{P}_{\theta} = \sum_{i=1}^{\mathbf{k}} \pi_i \mathcal{N}(\theta_i, \sigma^2)$$

Algorithm: Run Expectation-Maximization (EM) to estimate the parameters of model $\mathbb{P}_\theta$ given n i.i.d. samples from $\mathbb{P}_{\theta^*}$

EM on Gaussian mixture models

	Correctly-specified k=k* [1]	Over-specified k>k* [3]	Under-specified k<k* (This poster, [2])
Bias	Zero bias	Zero bias	Non-zero bias
Convergence rate of EM iterates	Fast rate e^{-cT}	Slow rate $\frac{c}{\sqrt{T}}$	Fast rate $e^{-c'T}$
Statistical Error of EM estimate	Parametric $n^{-1/2}$	Non-parametric $n^{-1/4}$	Parametric $n^{-1/2}$



Under-specified mixtures: Simple settings

Case 1: Three-Gaussian mixture with two close components

$$\mathbb{P}_{\theta^*} = \frac{1}{2}\mathcal{N}(-\theta^*, 1) + \frac{1}{4}\mathcal{N}(\theta^*(1+\rho), 1) + \frac{1}{4}\mathcal{N}(\theta^*(1-\rho), 1)$$

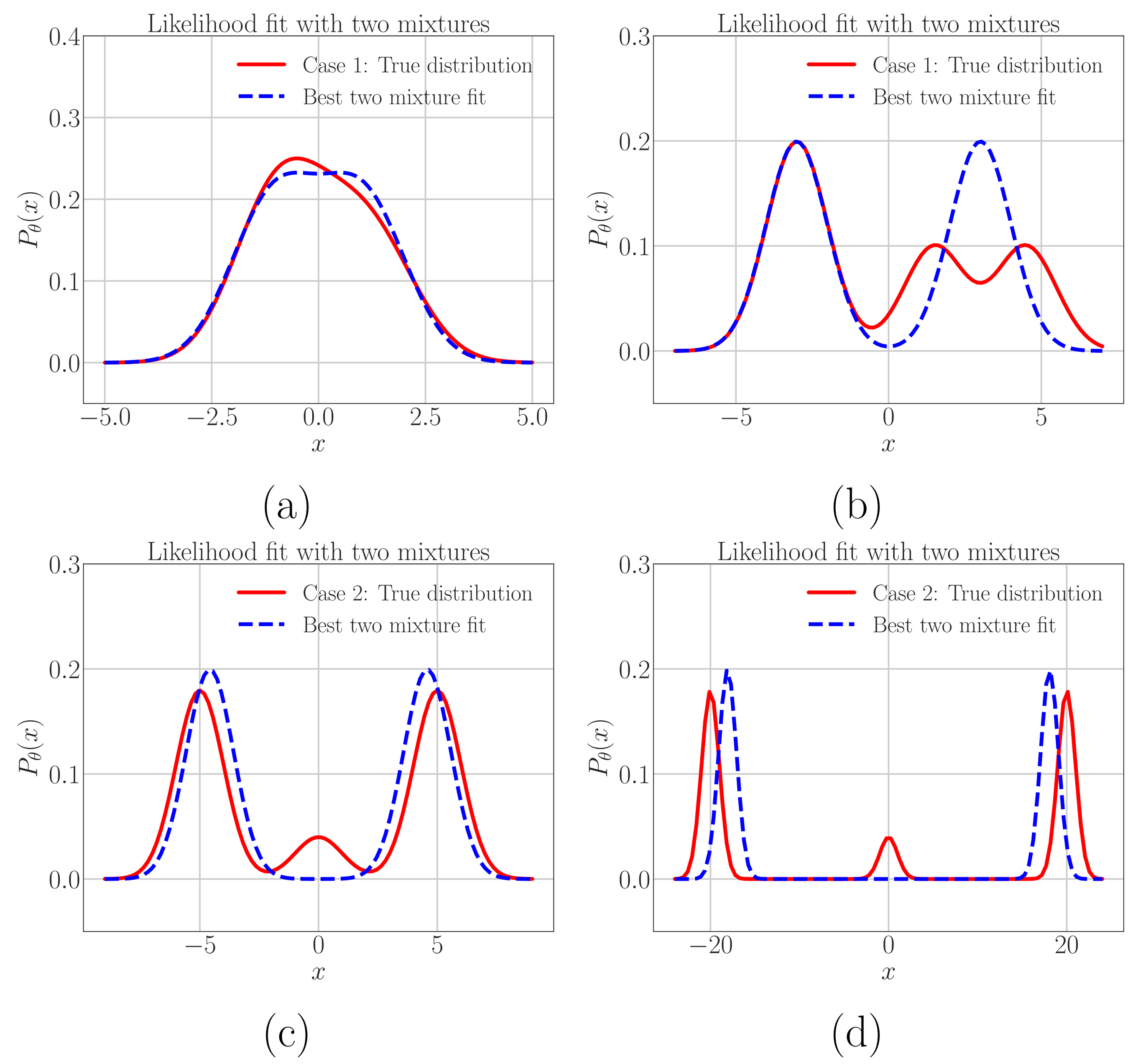
Case 2: Three-Gaussian mixture with one small component

$$\mathbb{P}_{\theta^*} = \frac{1-\omega}{2}\mathcal{N}(-\theta^*, 1) + \frac{1-\omega}{2}\mathcal{N}(\theta^*, 1) + \omega\mathcal{N}(0, 1).$$

where ω and ρ are *small positive scalars.*

The model fit: Using EM, fit a two Gaussian mixture

$$\mathbb{P}_{\theta} = \frac{1}{2}\mathcal{N}(-\theta, 1) + \frac{1}{2}\mathcal{N}(\theta, 1).$$



Quantities of interest:

1. Algorithmic rate of convergence of EM
2. Final statistical error $|\widehat{\theta}_n - \theta^*|$ where $\widehat{\theta}_n$ is the final EM estimate

$$\underbrace{|\widehat{\theta}_n - \theta^*|}_{\textit{Estimation error}} \leq \underbrace{|\widehat{\theta}_n - \overline{\theta}|}_{\textit{Statistical error}} + \underbrace{|\theta^* - \overline{\theta}|}_{\textit{Bias}}$$

Theoretical Results

1. How fast do the EM iterates converge to $\widehat{\theta}_n$?
   —***Exponentially fast,*** $\log n$ ***steps at most!***
2. What is the scaling of the error $|\widehat{\theta}_n - \overline{\theta}|$ with sample size n?
   —***The usual*** $n^{-1/2}$ ***scaling!***
3. How large is the bias term $|\theta^* - \overline{\theta}|$?
   —$\mathcal{O}(\rho^{1/4})$ ***for case 1, and*** $\mathcal{O}(\omega^{1/8})$ ***for case 2.***
   (Upper bounds)

Numerical Experiments

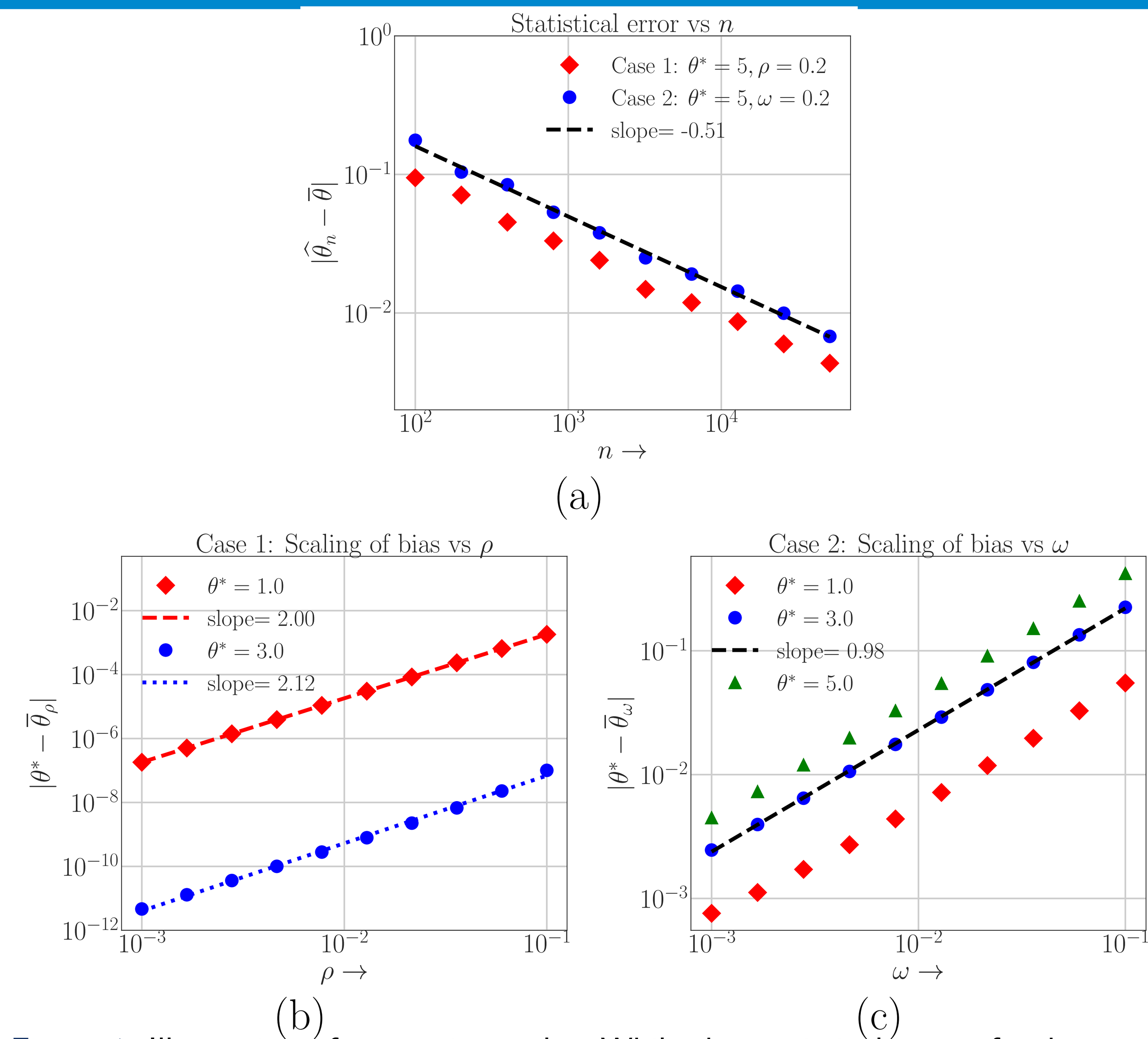


Figure 1: Illustration of our main results. While the statistical error of order $n^{-1/2}$ (panel (a)) matches our theoretical predictions, in panels (b) and (c), we observe that the biases for cases 1 and 2 have a scaling of $\mathcal{O}(\rho^2)$ and $\mathcal{O}(\omega)$ which suggest the potential looseness of our theoretical bounds for the biases.

References

[1] S.Balakrishnan, M. J. Wainwright and B. Yu, *Statistical Guarantees for the EM Algorithm: From Population to Sample-based Analysis*, AoS (2017).

[2] R. Dwivedi*, N. Ho*, K. Khamaru*, M. J. Wainwright and M. I. Jordan, *Theoretical guarantees for EM under misspecified Gaussian mixture models*, NIPS 2018.

[3] R. Dwivedi*, N. Ho*, K. Khamaru*, M. J. Wainwright, M. I. Jordan and Bin Yu, *Singularity, misspecification and the convergence rate of EM*, arXiv:1810.00828