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: Provide a theoretical analysis of Expectation-Maximization algorithm (EM) when the number of components in a mixture model is not well-specified.

Given: n i.i.d. samples from a k^*-Gaussian mixture model (GMM) with known variance and unknown means

$$X_1, X_2, \dots, X_n \overset{\text{i.i.d.}}{\sim} P_{\theta^*} = \pi_i \sum_{i=1}^{k^*} \mathcal{N}(\theta_i^*, \sigma^2)$$

Method: Using EM, fit a Gaussian mixture model in which the number of components is less than k^*.

Fitting 3-mixture models with 2-mixtures

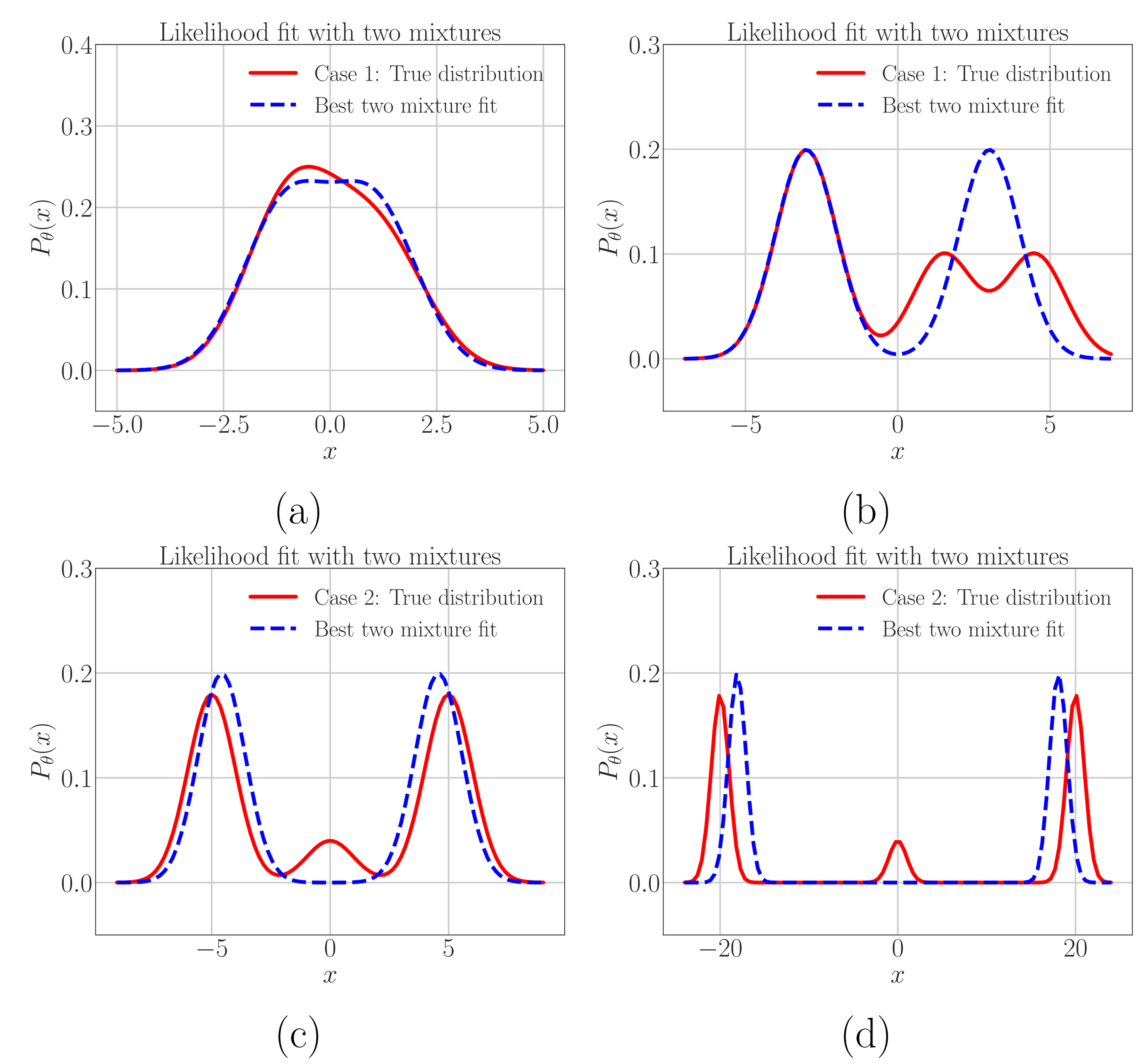


Figure 1: Fitting a three Gaussian mixture (3-GMM) with a two mixture model. Panels (a) and (b): The true 3-GMM has two components very close to each other. Panels (c) and (d): The true 3-GMM has one component with small mixture weight. In both cases, the best two mixture model approximation serves as a good fit. Also a visual inspection of the samples may suggest that they are generated from a two mixture model leading to a misspecification while estimating parameters.

Mathematical set-up

Case 1: Three Gaussian mixture with two close components

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

where ρ is a *small positive scalar*.

Case 2: Three Gaussian mixture with one small component

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

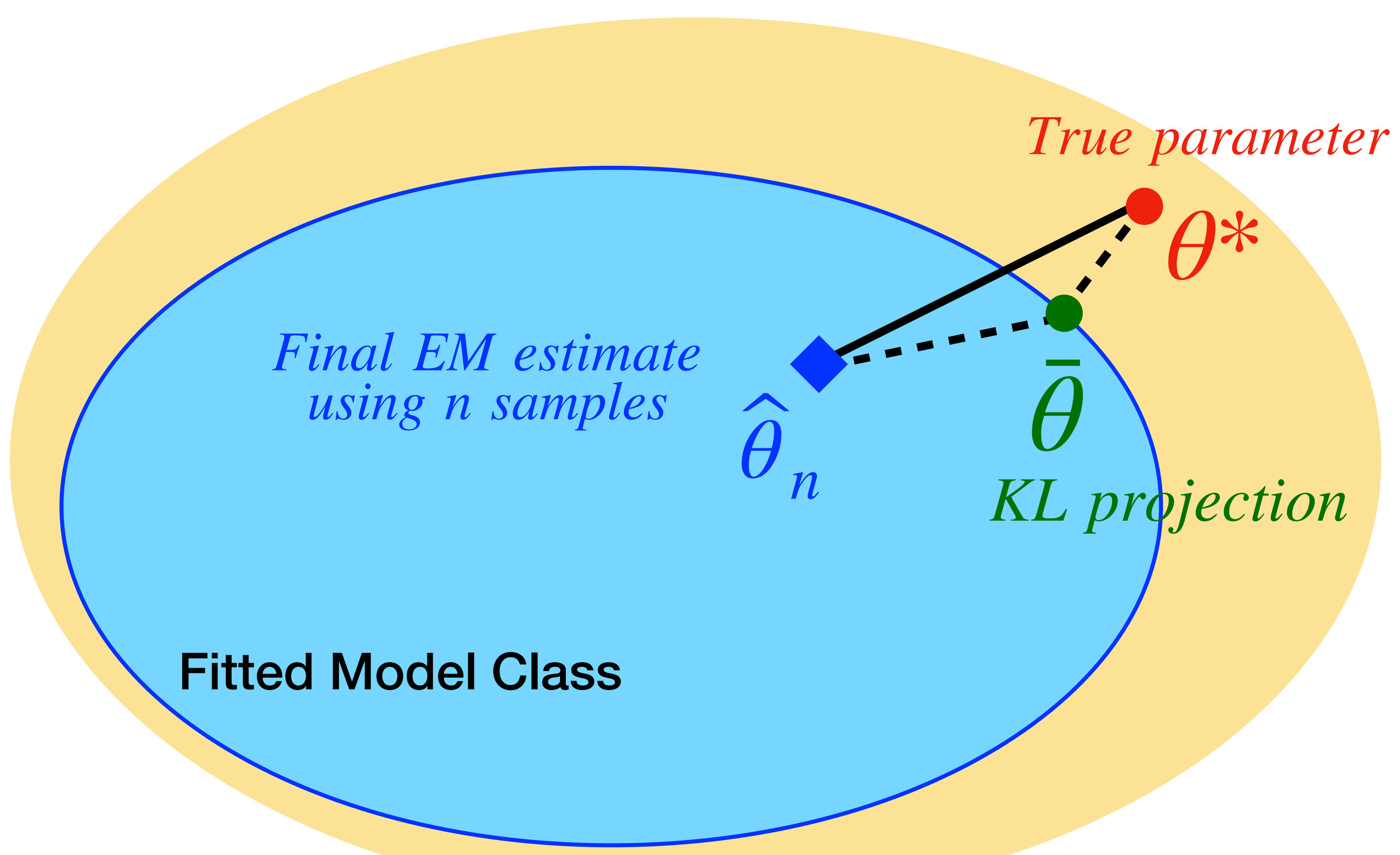
where ω is a *small positive scalar*.

Natural fitted model: Using EM, fit a two Gaussian mixture

$$P_\theta = \frac{1}{2}\mathcal{N}(-\theta, 1) + \frac{1}{2}\mathcal{N}(\theta, 1)$$

Quantity of interest: $|\widehat{\theta}_n - \theta^*|$ where $\widehat{\theta}_n$ is the final EM estimate.

Two aspects of parameter estimation error



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

Main Results

1. What is the scaling of the error $|\widehat{\theta}_n - \bar{\theta}|$ with sample size n?
   —***The usual $n^{-1/2}$ scaling!***
2. How fast do the EM iterates converge to $\widehat{\theta}_n$?
   —***Exponentially fast, $\log n$ steps at most!***
3. How large is the bias term $|\theta^* - \bar{\theta}|$?
   —$\mathcal{O}(\rho^{1/4})$ ***for case 1, and*** $\mathcal{O}(\omega^{1/8})$ ***for case 2.***
4. Any caveats?
   —The bias bounds are probably not tight.

Numerical Experiments

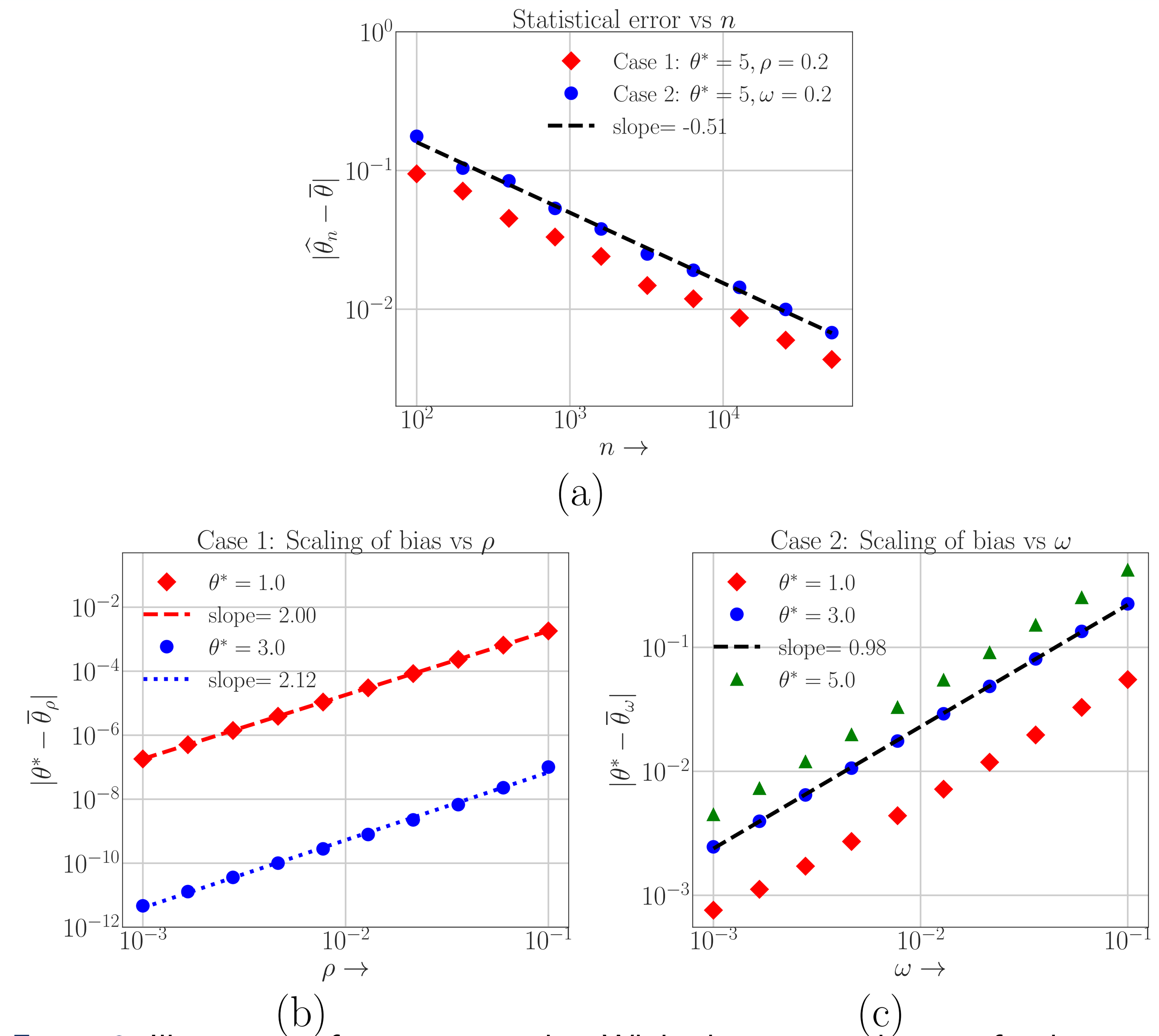


Figure 2: Illustration of our main results. While the statistical error of order $n^{-1/2}$ (panel (a)) matches our theoretical predictions, in panels (b) and (c), we observe that the biases for cases 1 and 2 have a scaling of $\mathcal{O}(\rho^2)$ and $\mathcal{O}(\omega)$ which suggest the potential 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, M. I. Jordan *Theoretical guarantees for EM under misspecified Gaussian mixture models*, NIPS 2018.