

Converging Fast and Slow: Statistics vs Optimization

Raaz Dwivedi, Koulik Khamaru, Nhat Ho, Michael Jordan, Martin Wainwright, Bin Yu

Unifying Theory [1]

#steps = \mathbf{T} ,
#samples = \mathbf{N}

Optimization Rate

$$e^{-c\mathbf{T}}$$

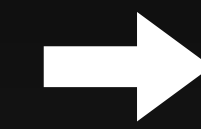
$$\mathbf{T}^{-a}$$

+

Stability

$$\mathbf{N}^{-1/2}$$

$$\mathbf{N}^{-b}$$

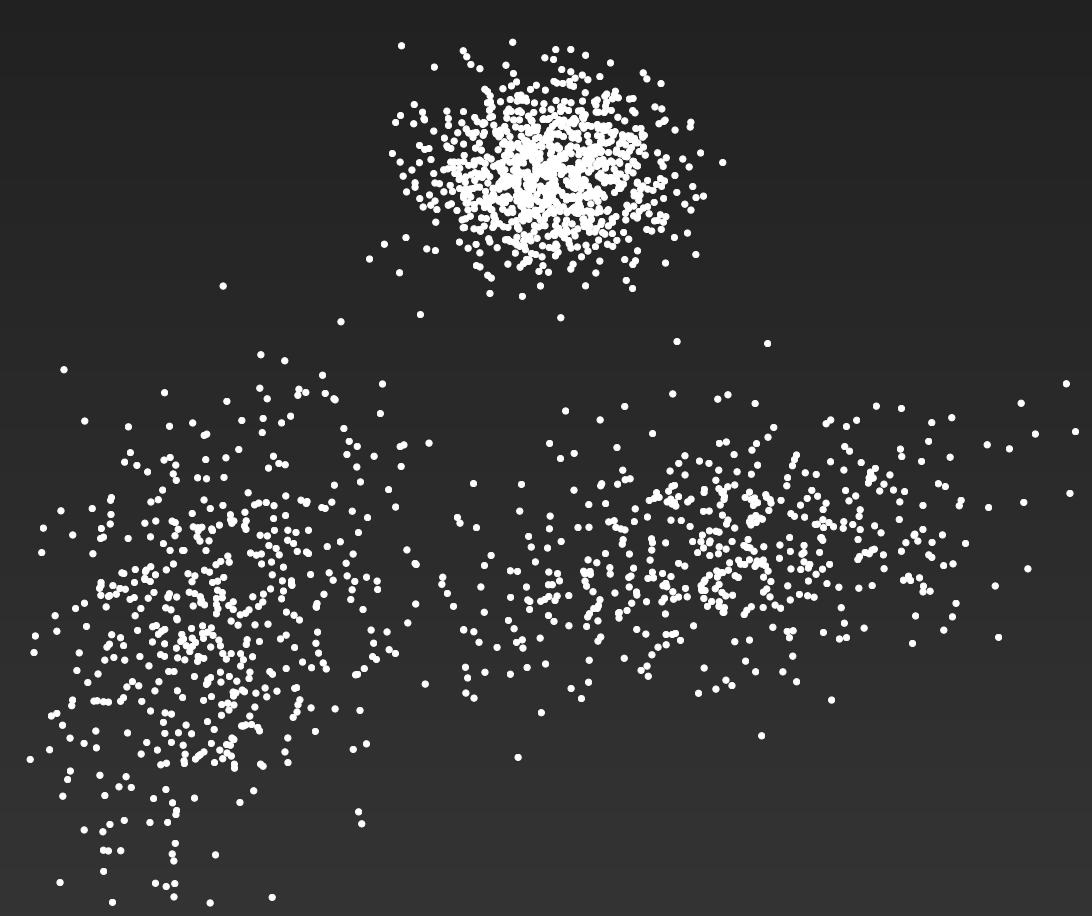


Final Statistical Error

$$\mathbf{N}^{-1/2} \text{ in } \log \mathbf{N} \text{ steps}$$

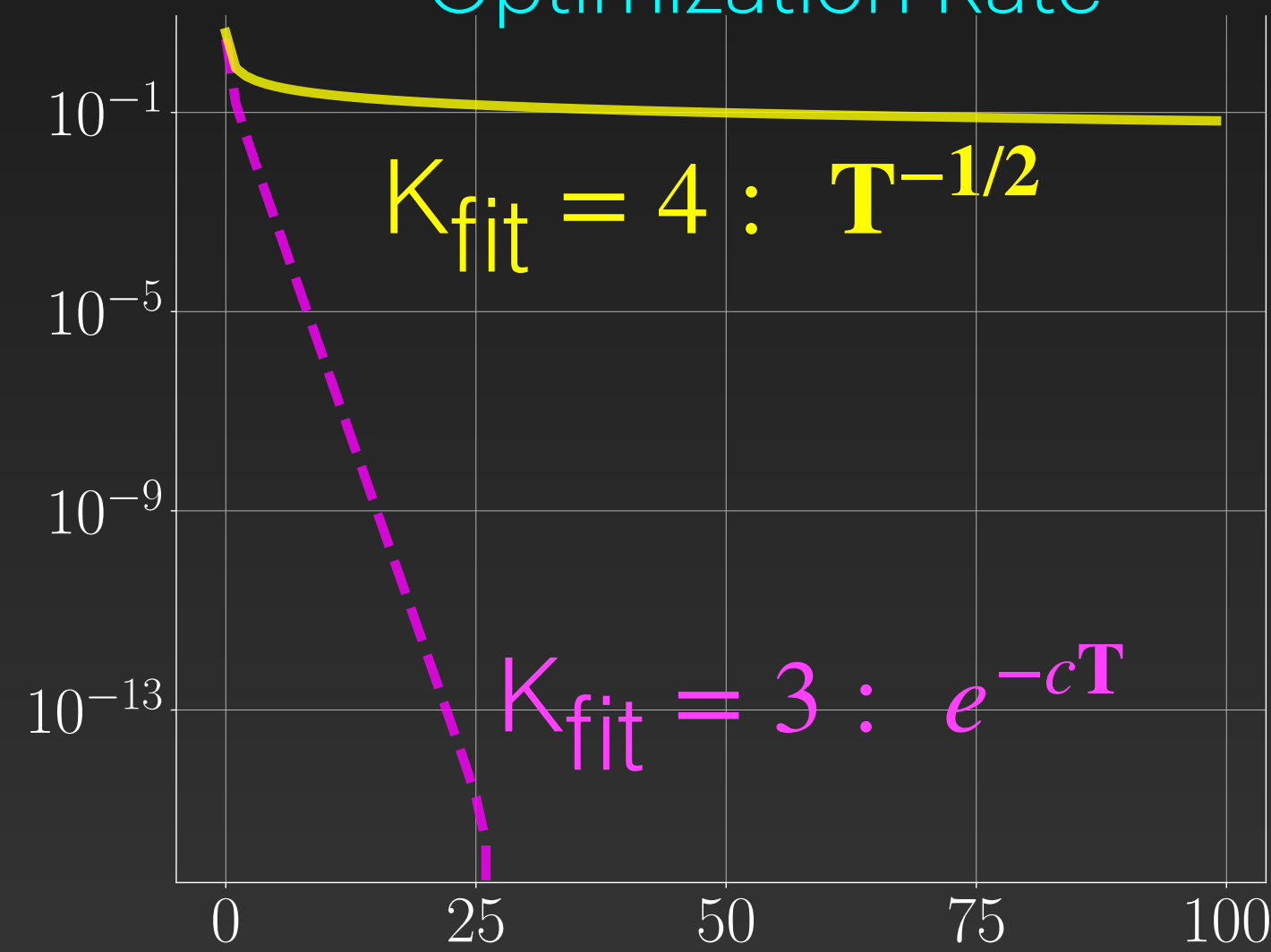
$$\mathbf{N}^{-ab} \text{ in } \mathbf{N}^b \text{ steps}$$

EM With Mixture Models [2]



$K_{\text{true}} = 3$

Optimization Rate



No. of steps \mathbf{T}

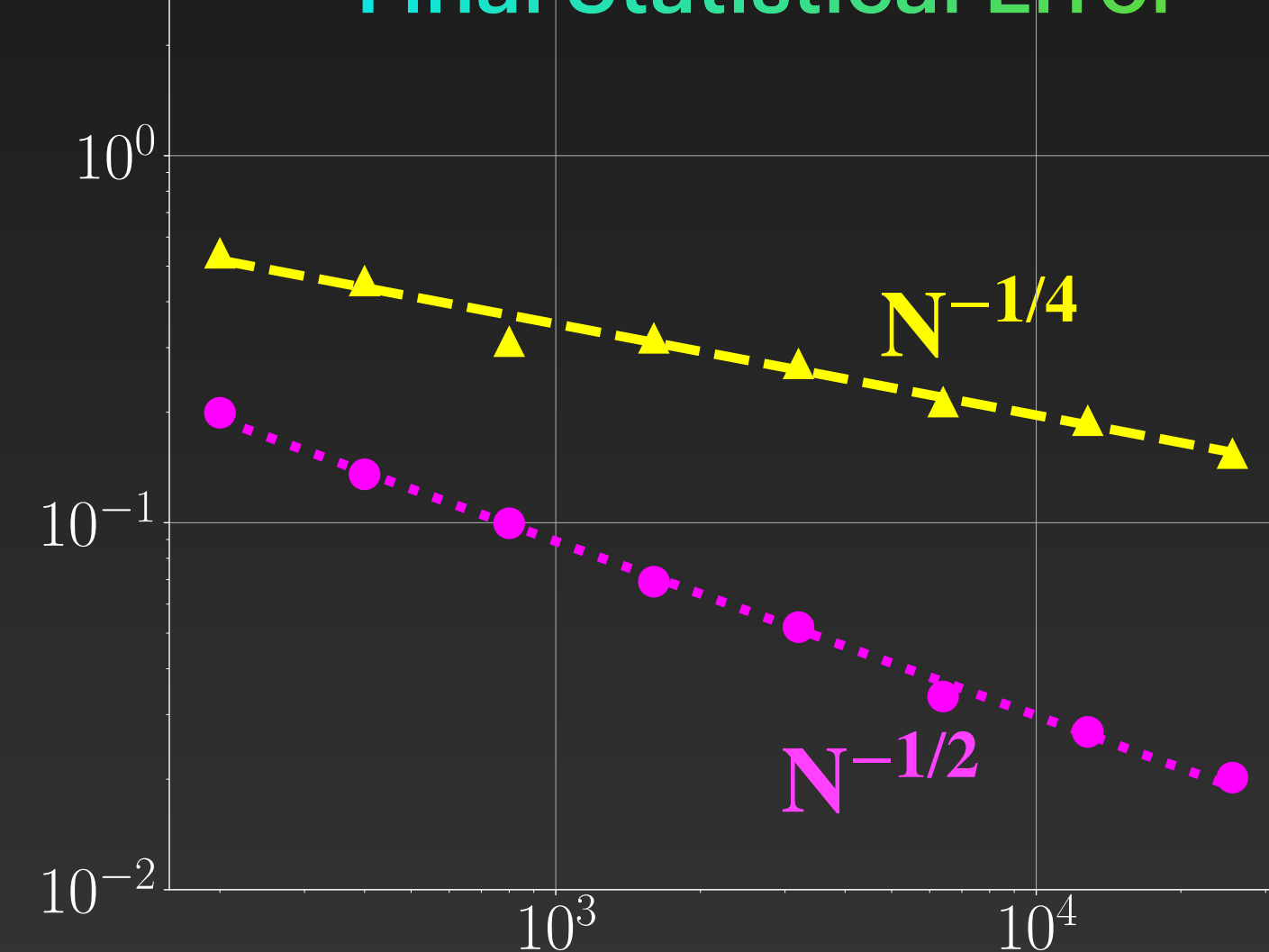
Stability

+

$$\mathbf{N}^{-1/2}$$



Final Statistical Error



No. of samples \mathbf{N}

[1] Instability, Computational Efficiency, and Statistical Accuracy, ArXiv

[2] Singularity, Misspecification, and the Convergence Rate of EM, Annals of Statistics