

Intractable Posteriors

LSST Discovery Fellowship Program Day 5

Greg Gilbert | LSST Discovery Workshop | 24 May 2025

Modeling choices

Physical

What processes do you include?

What approximations do you make?

Statistical

Are data i.i.d.?

Is there correlated noise?

Do you account for data collection?

Model specification

Parameterization

Priors

Convergence criteria

Sampler

Grid search

Maximum likelihood

Markov Chain Monte Carlo

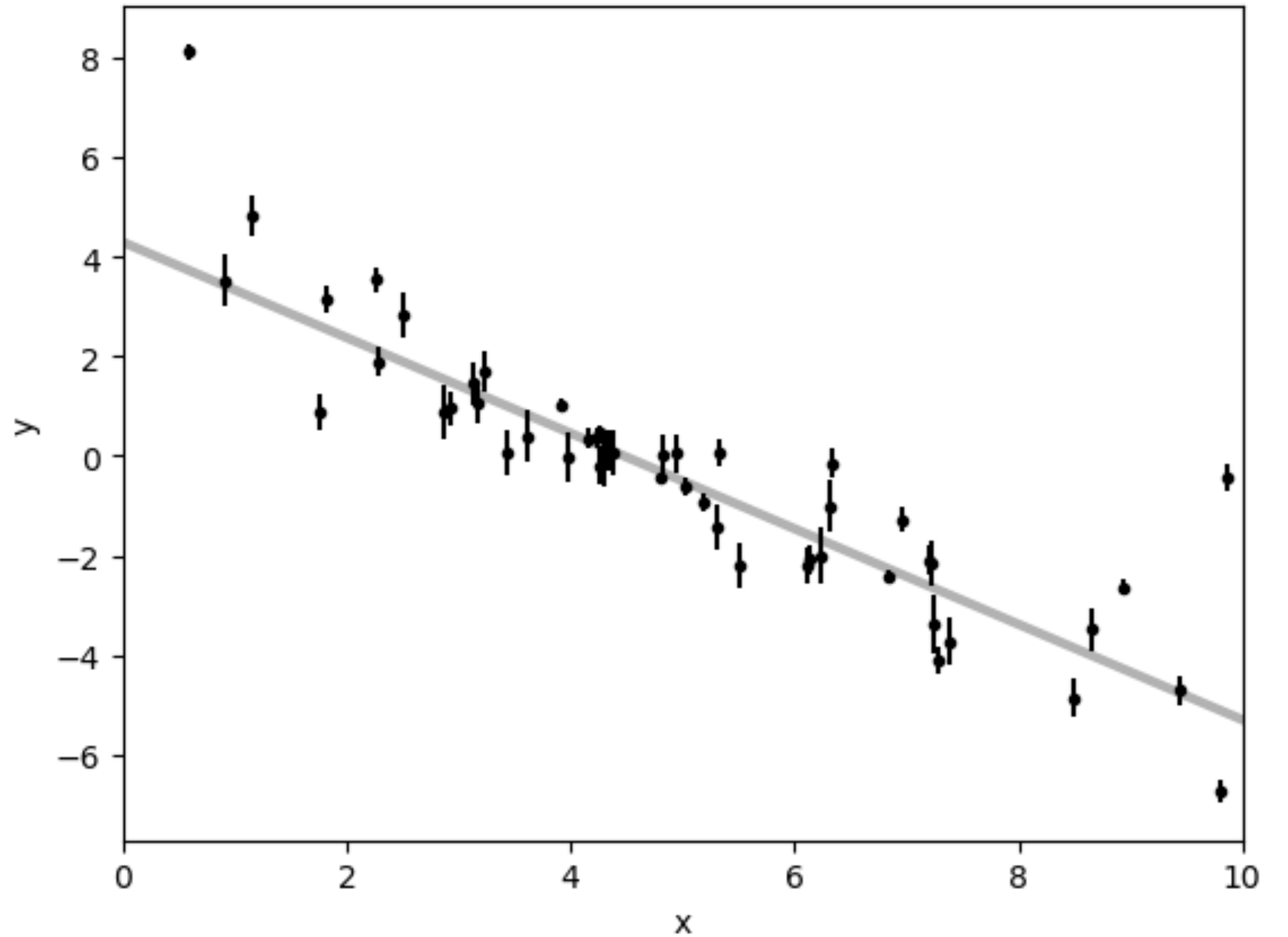
Nested Sampling

Linear Model

Let's parameterize the obvious way

$$y_{\text{mod}} = mx + b$$

*How can the model
specification be improved?*

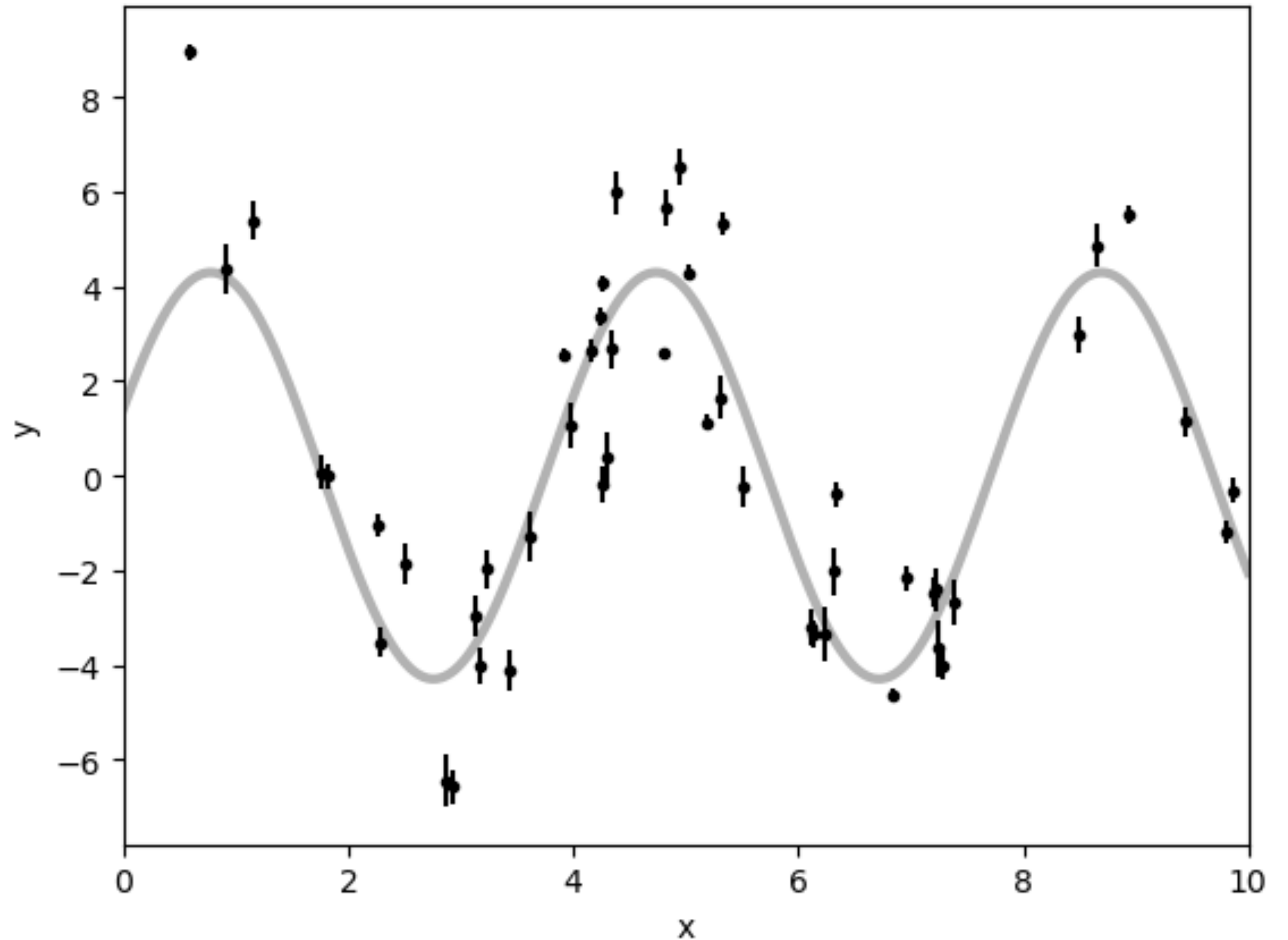


Sinusoid Model

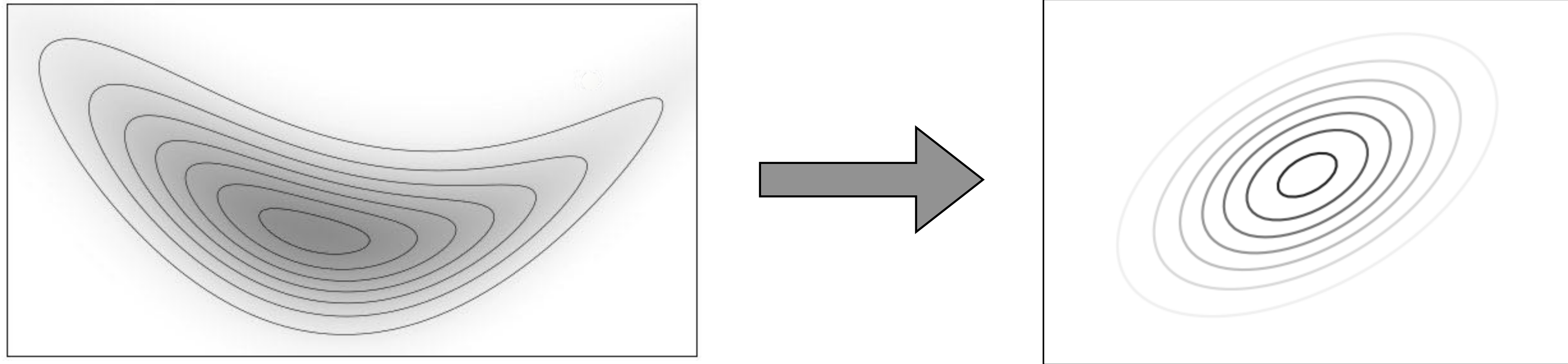
Let's parameterize the obvious way

$$y_{\text{mod}} = A \sin\left(\frac{2\pi t}{P} + \phi\right)$$

*How can the model
specification be improved?*



Transform to a (nearly) **Orthogonal Basis**

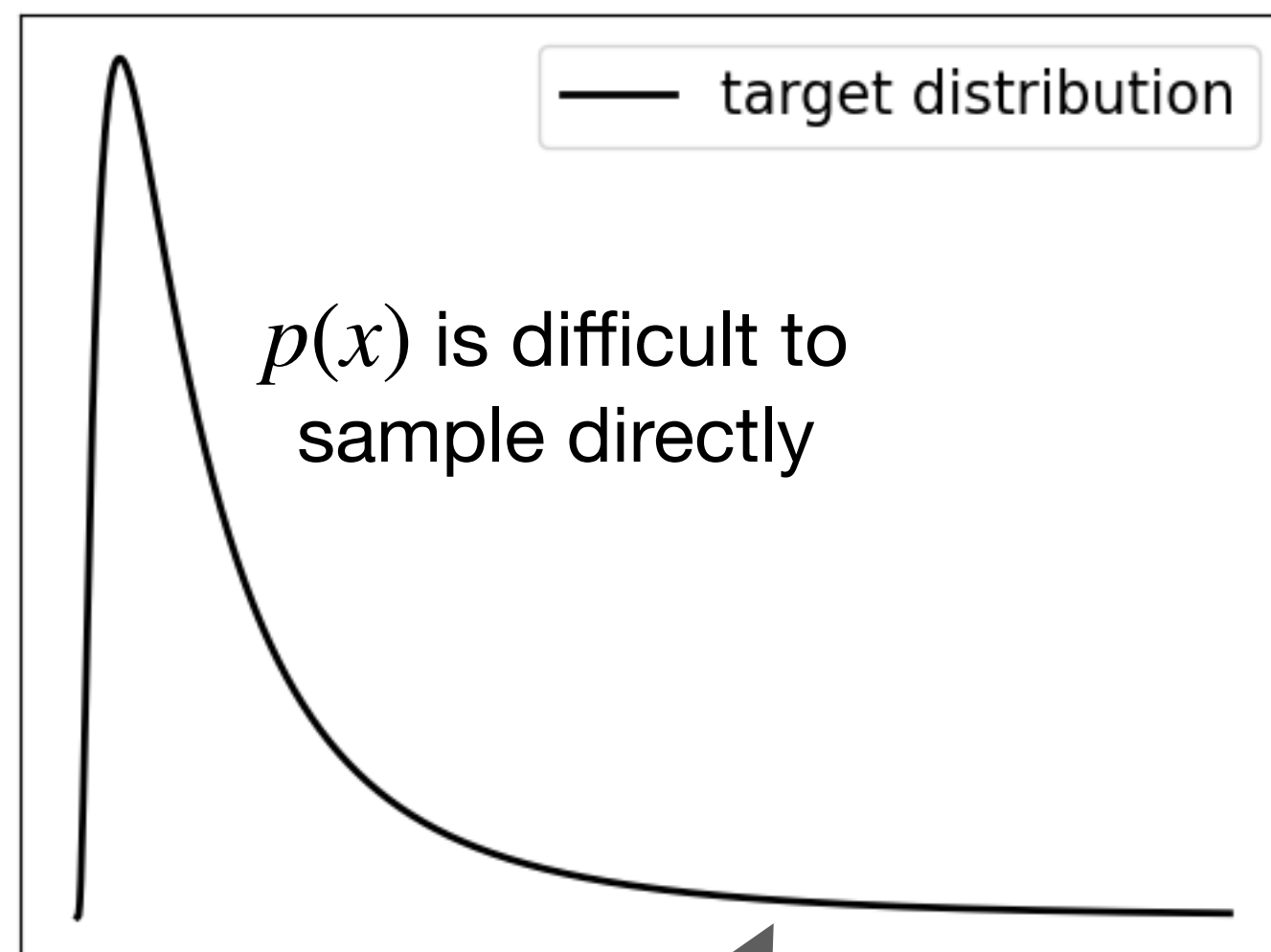


- Warps the posterior topology to be closer to Gaussian
- Reduces covariances and broadens funnels
- More efficient AND more accurate sampling results

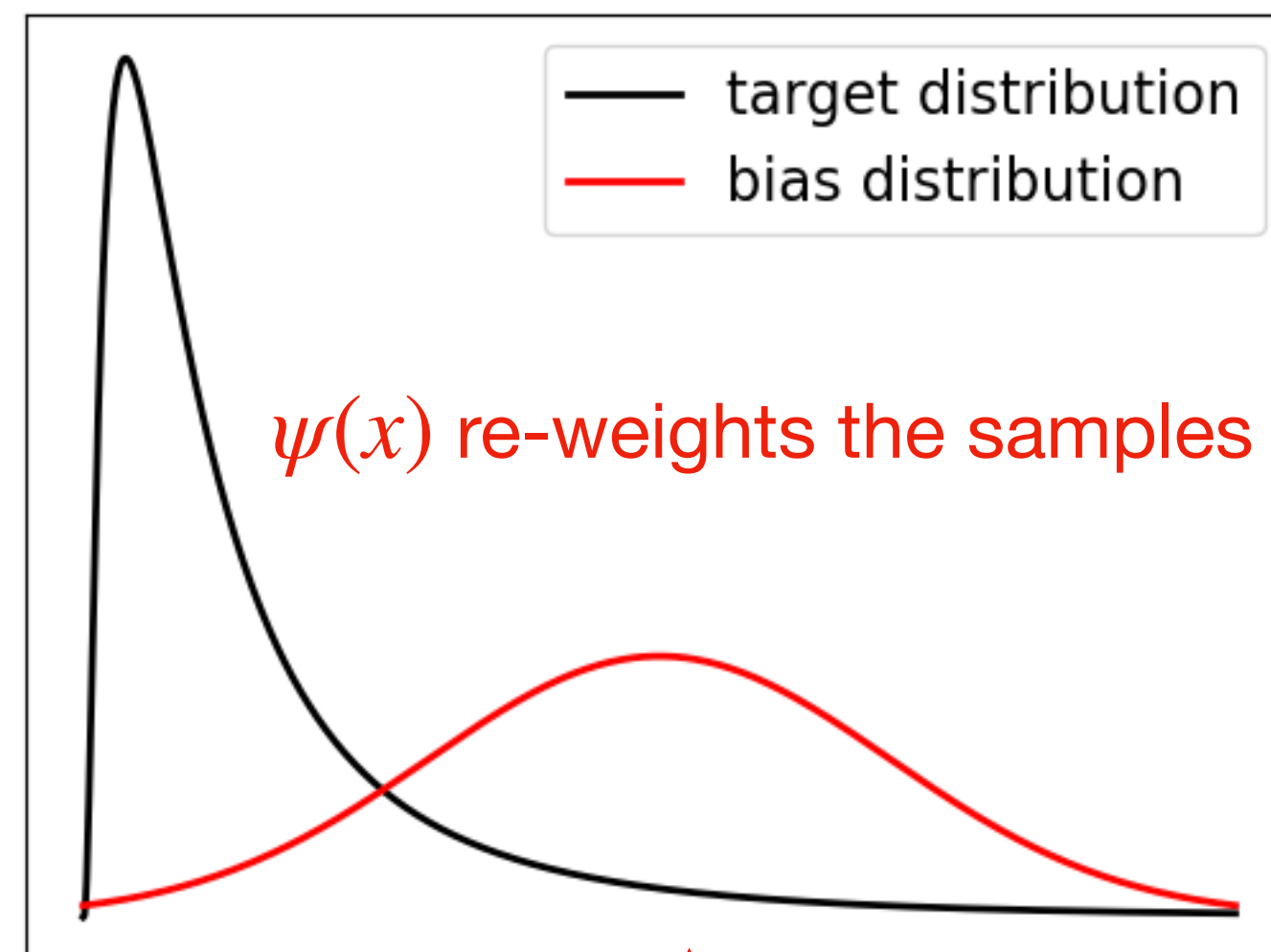
Warning: parameter transformations can introduce implicit priors

Don't forget to incorporate the appropriate Jacobian

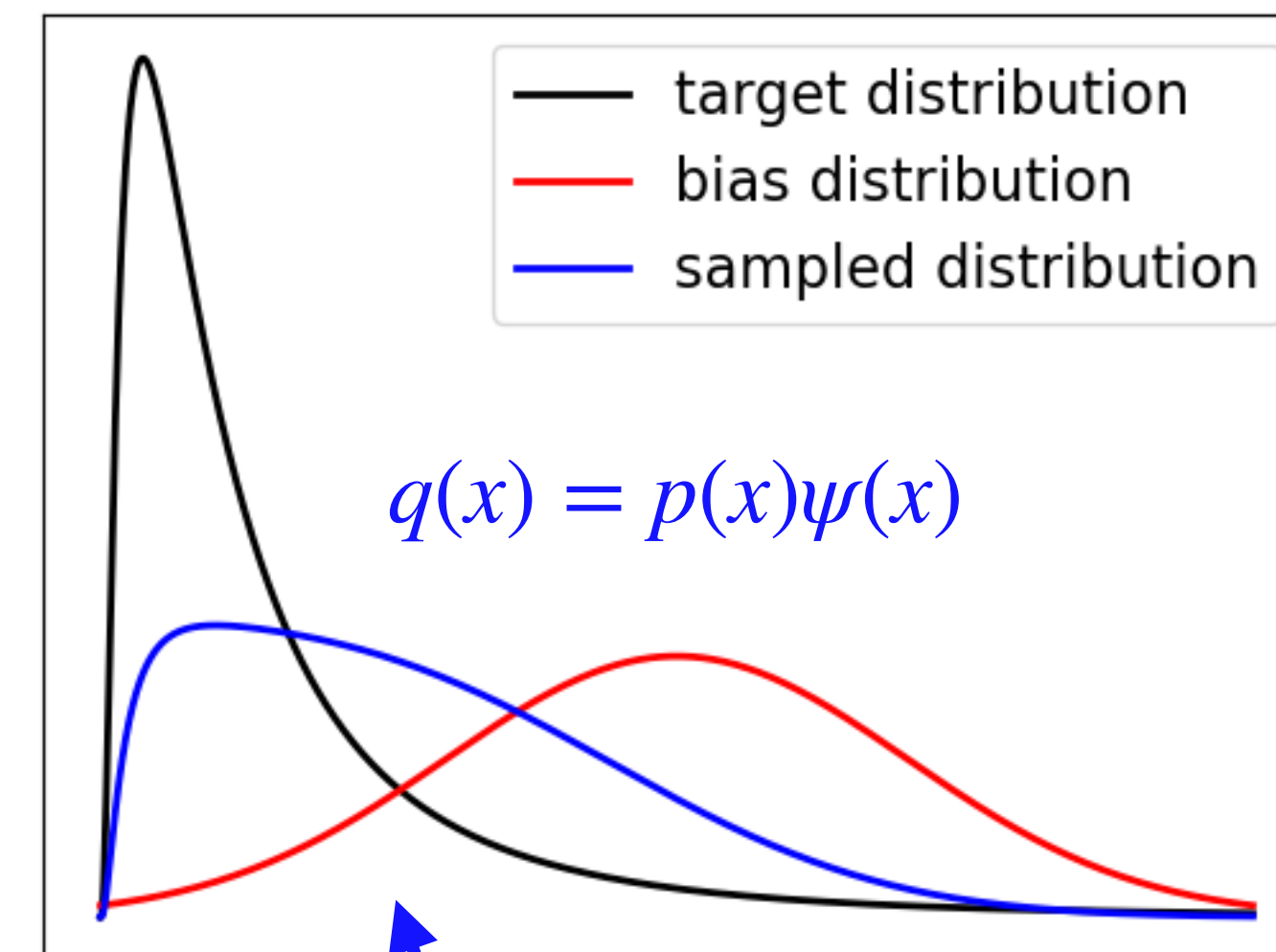
Importance Sampling



Samples in low-probability region strongly influence our posterior estimates of, e.g. $\{\mu_x, \sigma_x^2\}$



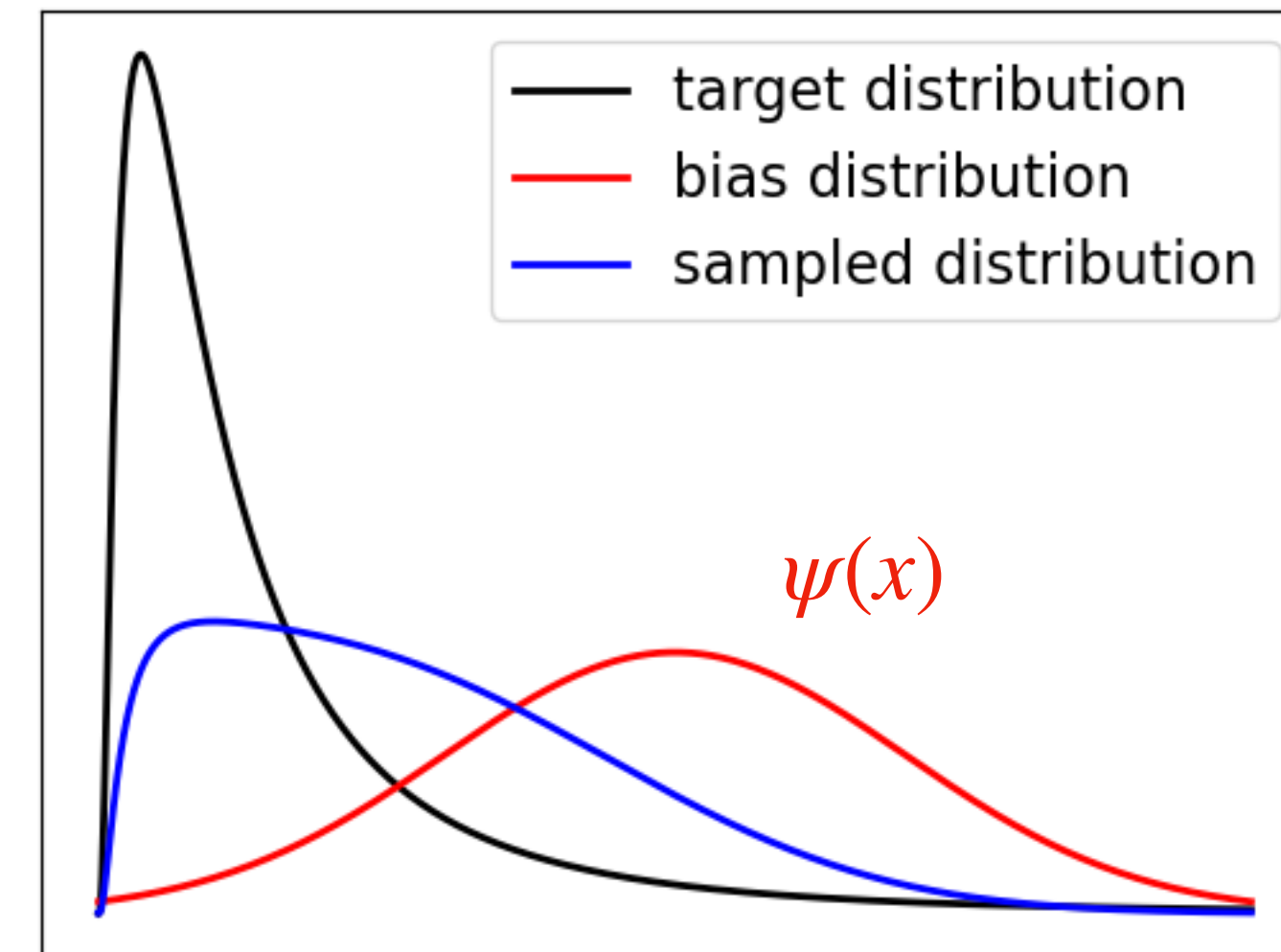
A user-specified bias function guides the sampler toward low-probability regions



The new distribution is easier to sample from

Bayesian Interpretation

$$P(\theta \mid D) = \frac{P(D \mid \theta) \boxed{P(\theta)}}{P(D)}$$



In Bayesian language, imposing a **bias function** is equivalent to imposing a **prior**

Because we know the prior exactly, we can re-weight our posterior samples to remove the effect of the bias

Some spicy opinions on priors

There is no such thing as an “uninformative” prior

Don’t get too hung up on picking a single “best” prior

Some less spicy advice

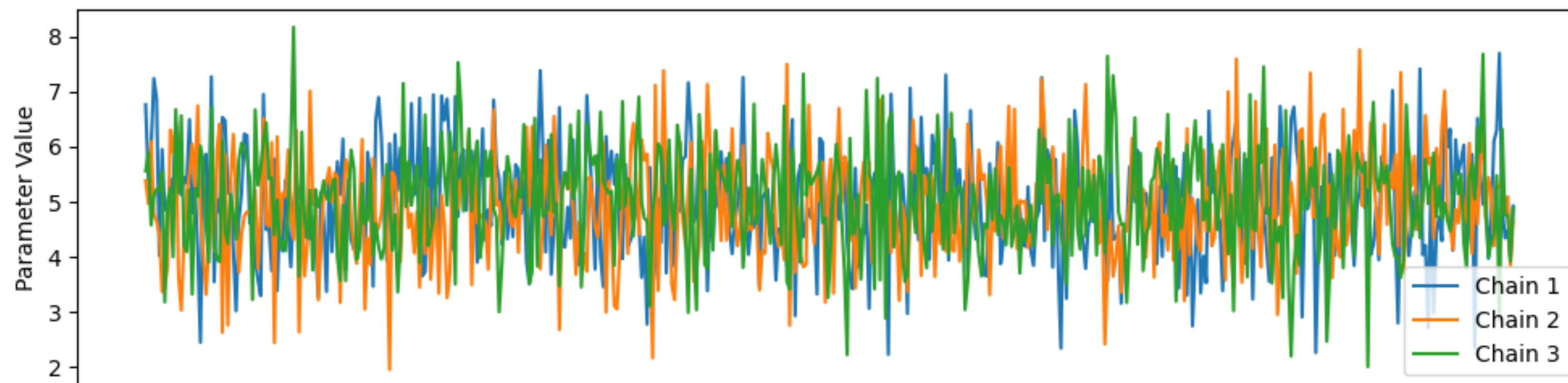
Always state your priors

Be mindful of parameter covariances

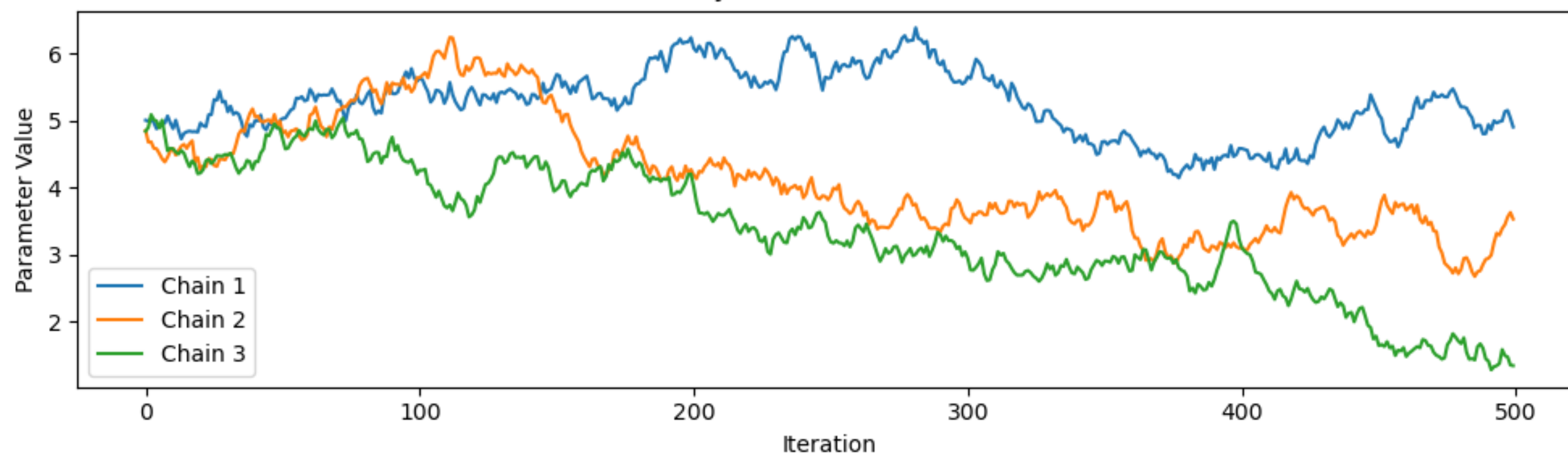
Test to ensure the full posterior is explored

How do we assess convergence?

Well-Mixed MCMC Chains



Poorly-Mixed MCMC Chains



Gelman-Rubin
statistic

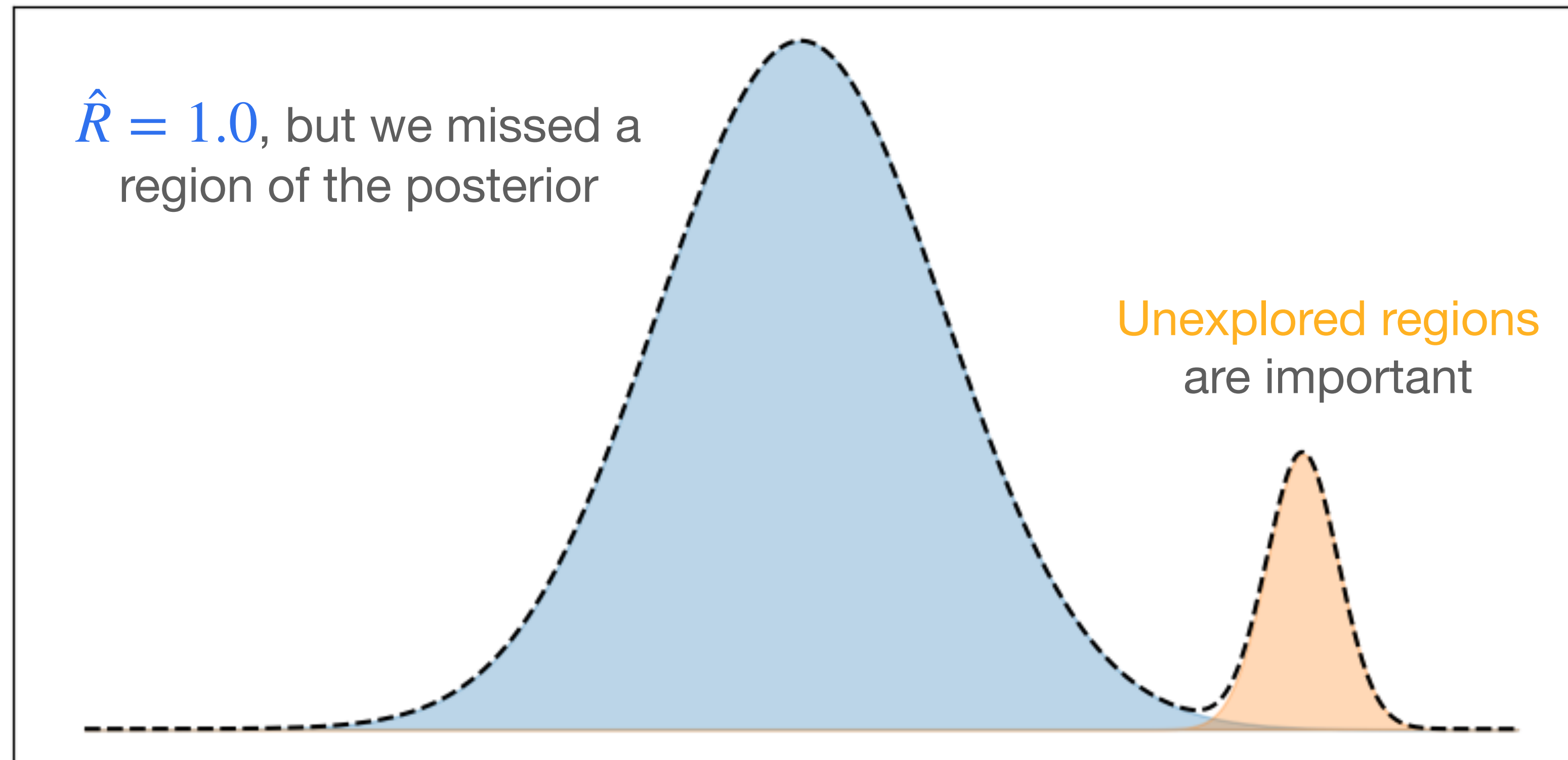
$$\hat{R} \sim \frac{W + B/L}{W}$$

W Within-chain variance

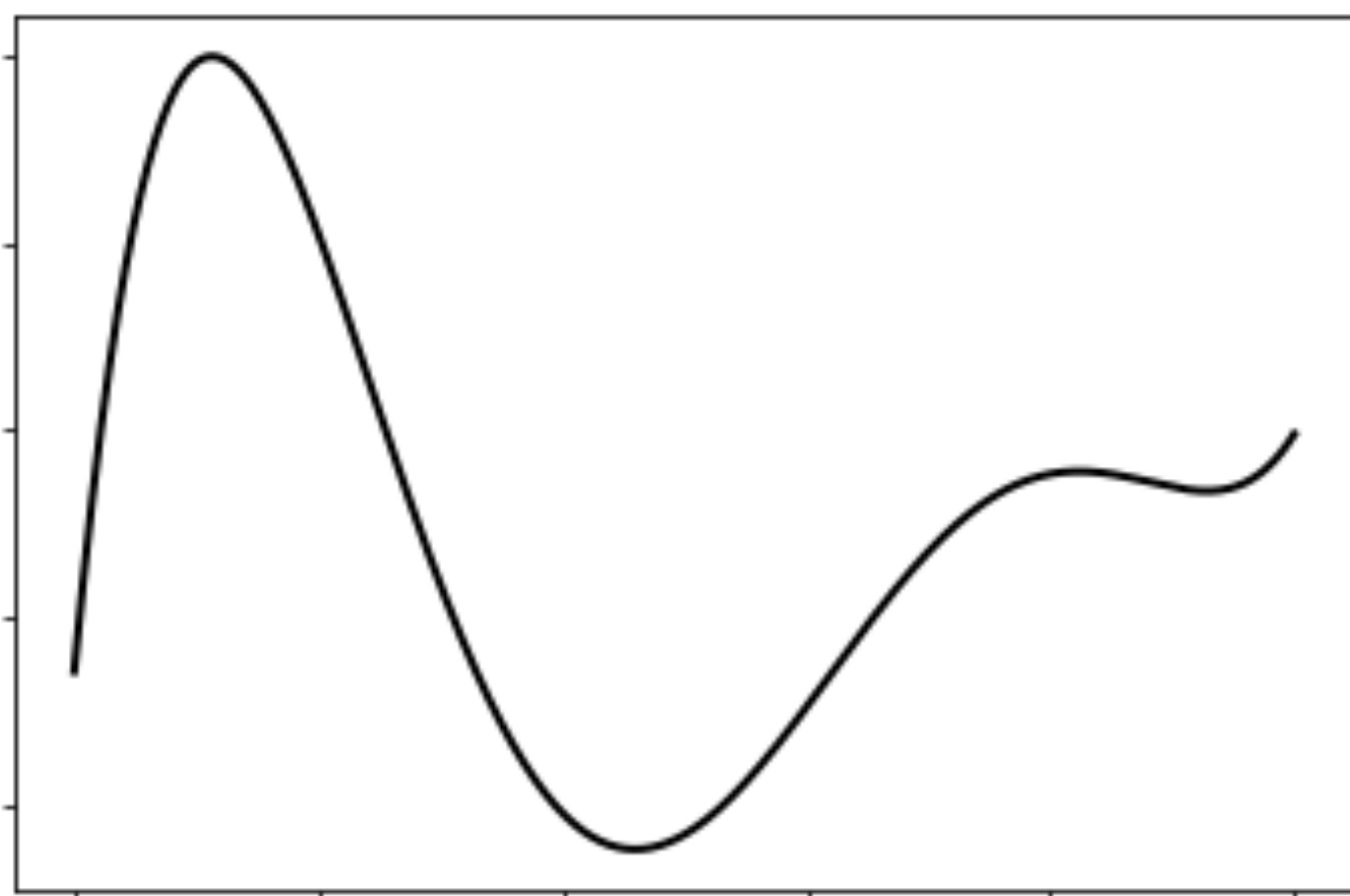
B Between-chain variance

L Length of chain

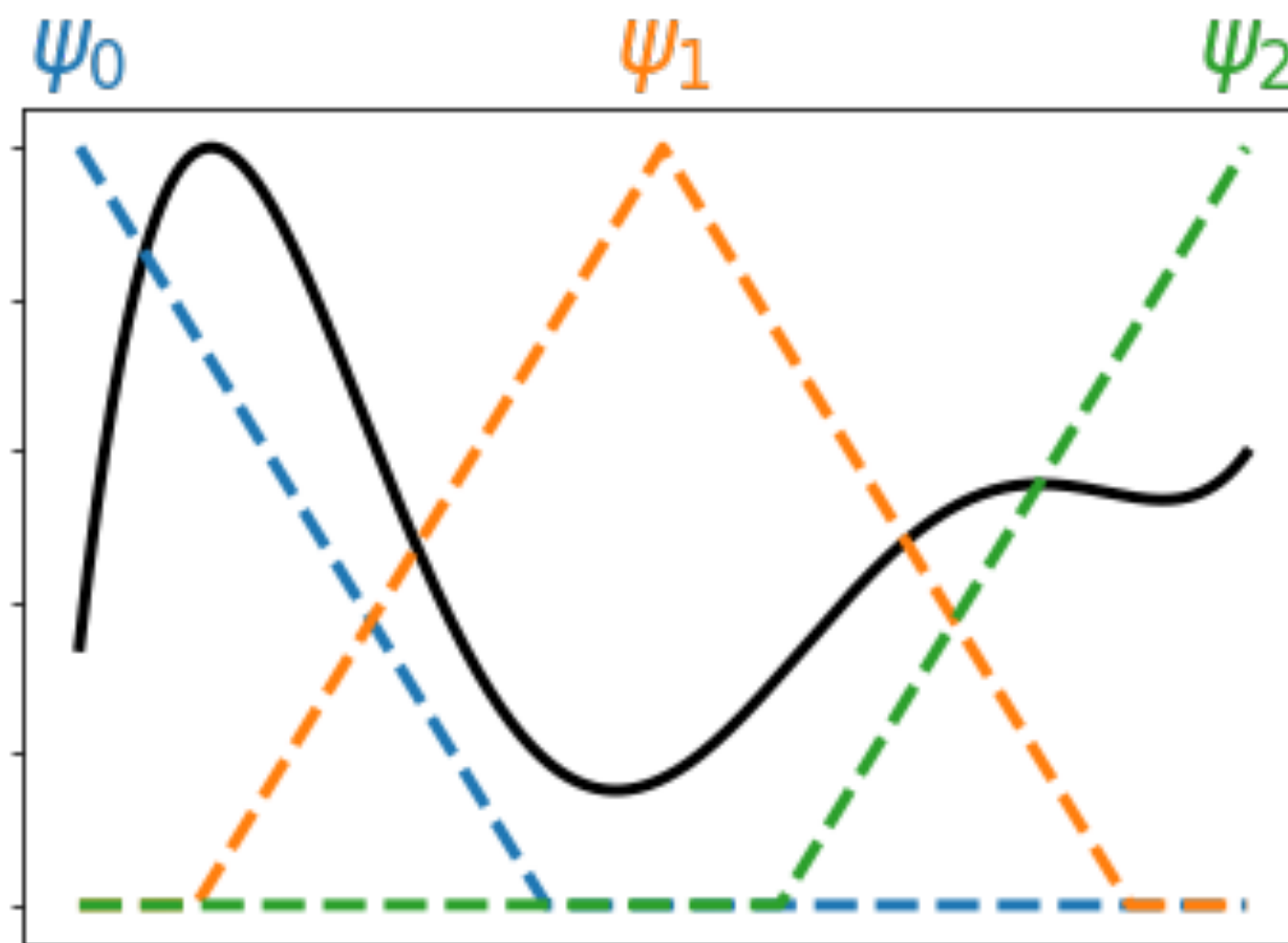
How do we assess convergence?



Umbrella Sampling

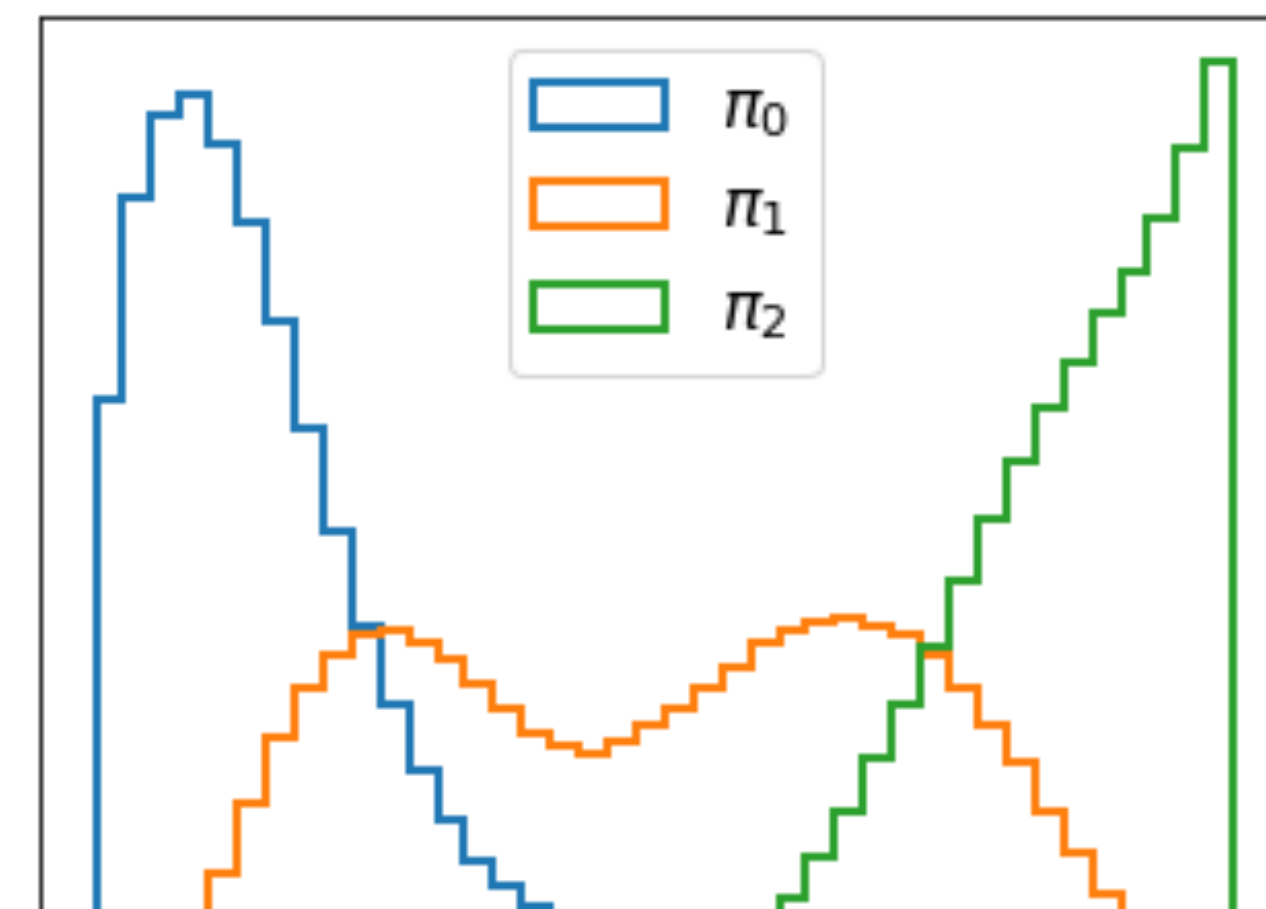


We have a complicated posterior with a narrow “bottleneck” that stymies most MCMC samplers



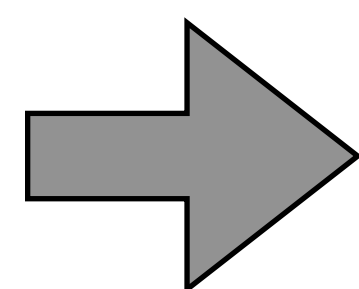
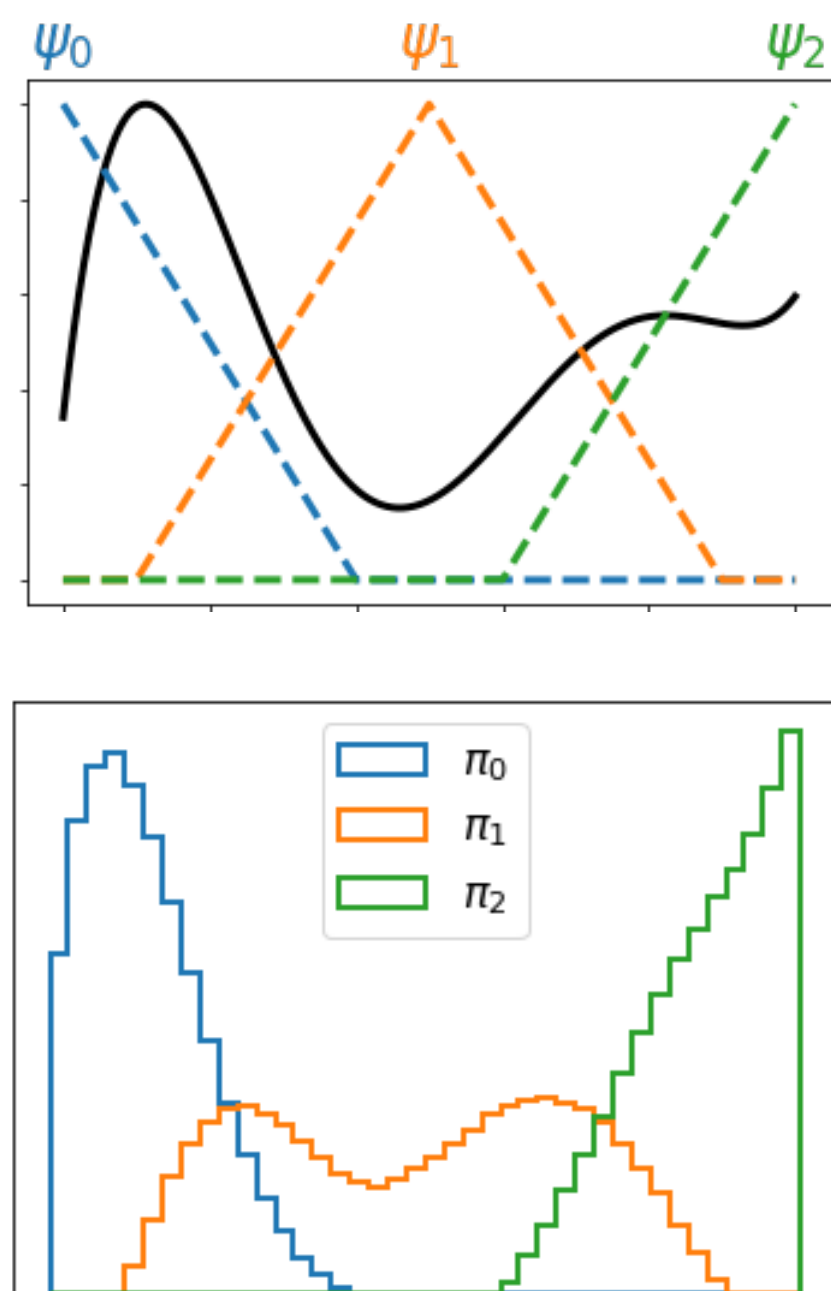
Split the posterior into three regions using bias functions

You'll need to run the sampler three times

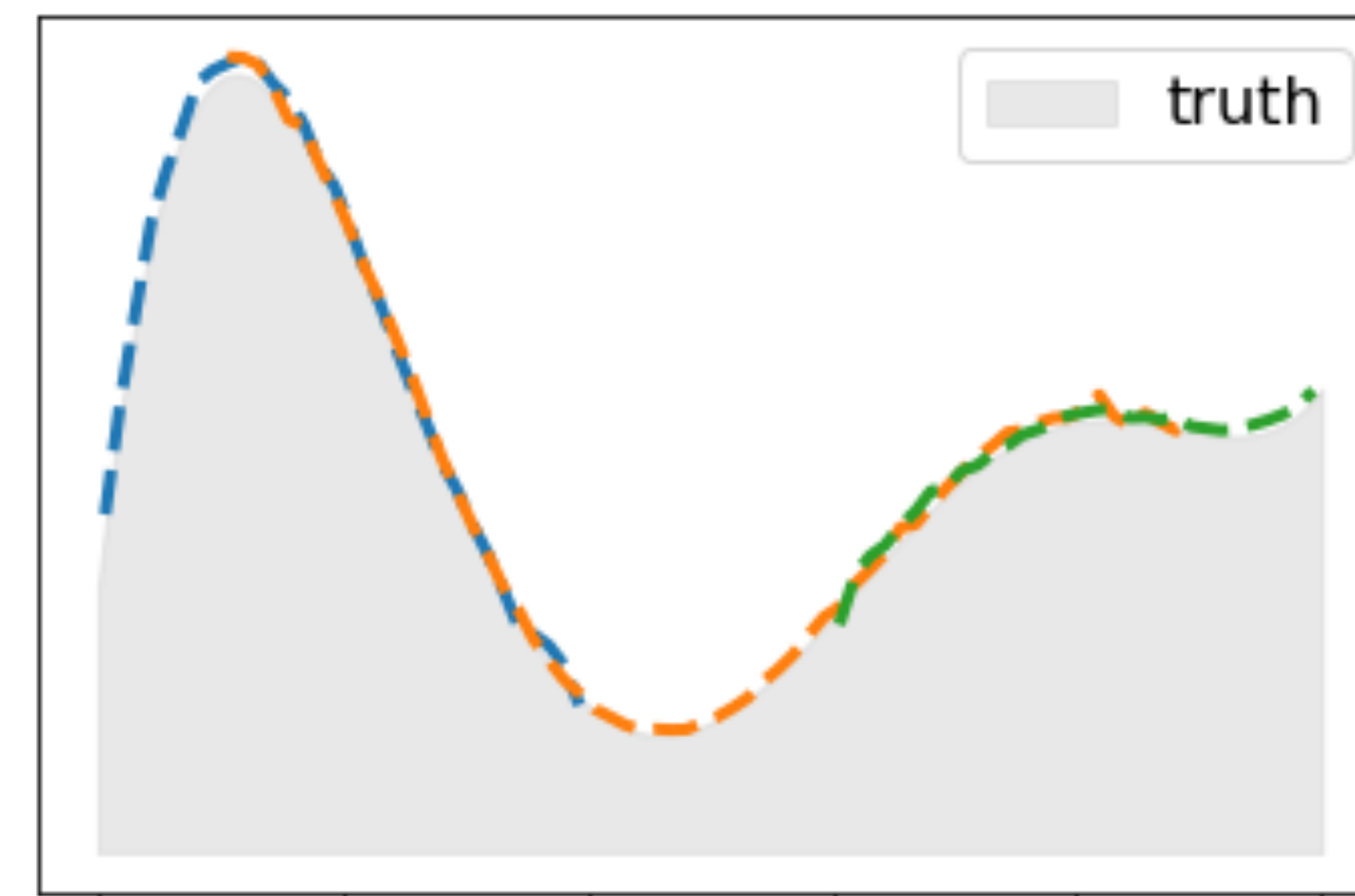
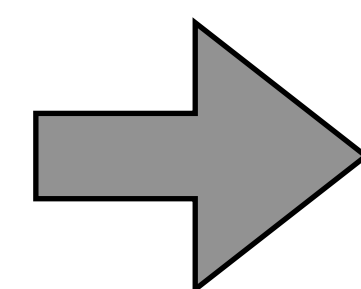


We now have three sets of (weighted) samples that need to be recombined

Umbrella Sampling



$$z_i = \int \psi_i(x) \pi(x) dx = \langle \psi_i \rangle_{\pi}$$



Estimating the overlap matrix F

$$z_i = \sum_{j=1}^N \left\langle \frac{\psi_j(x)}{\sum_{k=1}^N \psi_k(x)/z_k} \right\rangle_{\pi_i} \quad F_{ij} = \left\langle \frac{\psi_j/z_j}{\sum_{k=1}^N \psi_k/z_k} \right\rangle_{\pi_i} \quad F = \begin{bmatrix} 0.83 & 0.07 & 0 \\ 0.39 & 0.72 & 0.31 \\ 0 & 0.08 & 0.8 \end{bmatrix}$$

For a full walkthrough of the mathematics, see https://gjgilbert.github.io/tutorials/umbrella_sampling/

How do we identify unknown unknowns?

Domain expertise

Exploratory runs

Umbrella sampling

Your sampler alone will not save you!

Further reading

Model Parameterization

Betancourt & Girolami 2013, “Hamiltonian Monte Carlo for hierarchical models”, [arXiv:1312.0906](#)

Importance Sampling

Bayesian Data Analysis, by A. Gelman et al. Third Edition, Boca Raton, FL: Chapman & Hall 2014

Umbrella Sampling

Matthews et al., 2018, “Umbrella sampling: a powerful method to sample tails of distributions”, MNRAS, 480, 3

Gilbert 2022, “Accurate modeling of grazing transits using umbrella sampling”, AJ, 163, 3