Bayesian Statistics: A Practical Primer

Key Concepts and Intuitions

Your Name

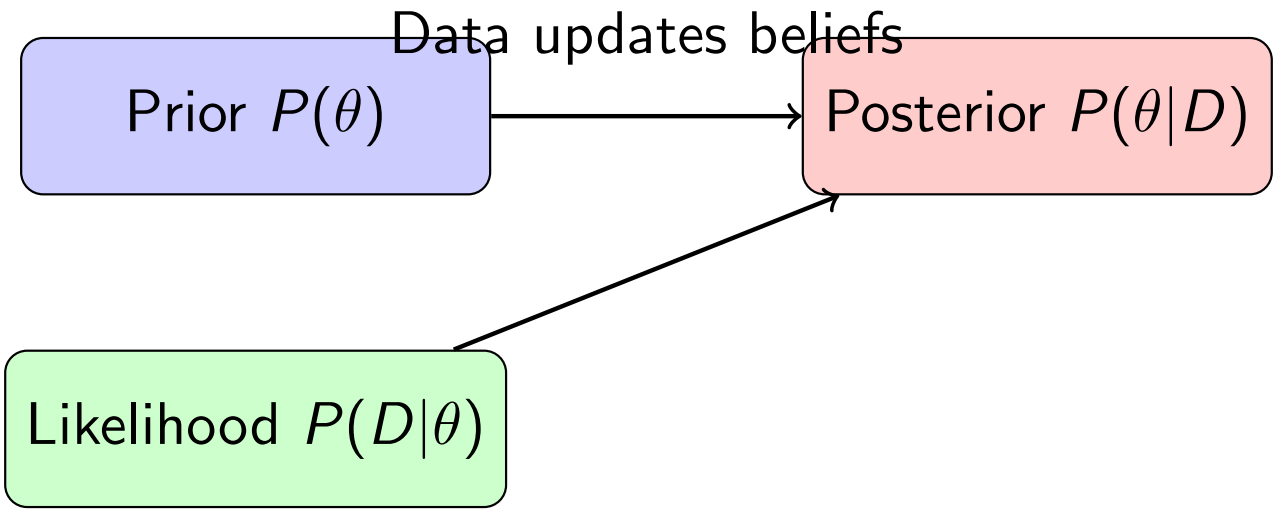
June 6, 2026

What Makes a Model "Bayesian"?

- **Parameters as random variables**
  - Not fixed unknown constants
  - Have probability distributions
- **Prior distributions**
  - Formalize what we know before seeing data
  - Can be informative or weakly informative
- **Posterior distributions**
  - Update prior beliefs with observed data
  - Represent our knowledge after seeing data
- **Uncertainty quantification**
  - Full distributions, not just point estimates
  - Natural framework for decision-making

Bayes' Rule: The Core Idea

$$P(\theta|D) = \frac{P(D|\theta) \times P(\theta)}{P(D)}$$



- **Prior:** What we believe before seeing data
- **Likelihood:** How the data relates to parameters
- **Posterior:** Updated beliefs after seeing data
- **Evidence:** $P(D)$ normalizes the posterior

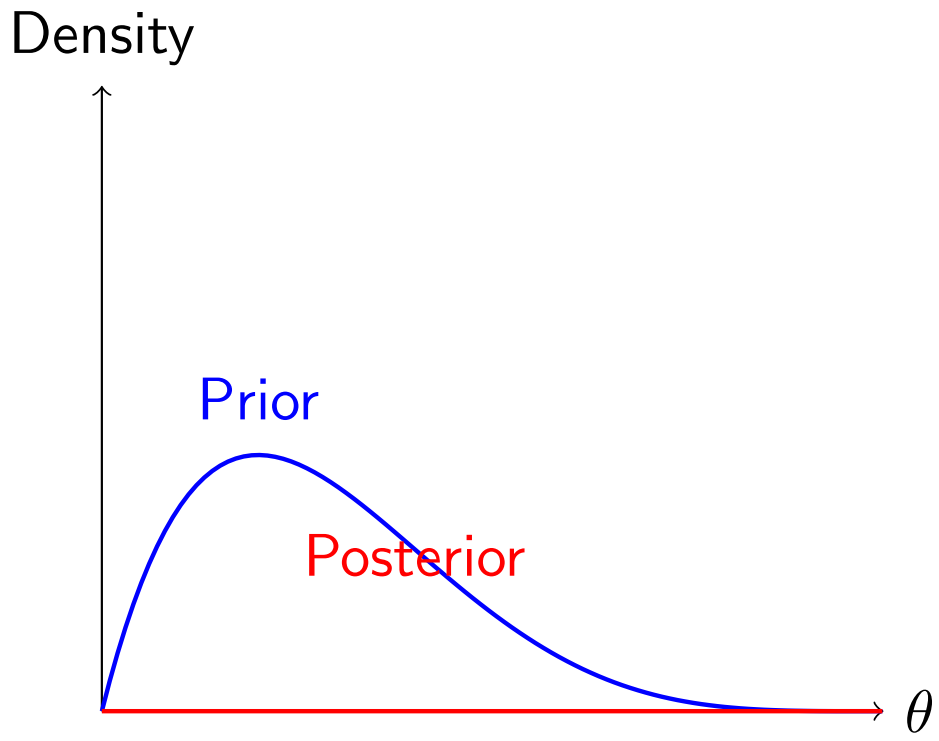
The Beta-Binomial: A Simple Example

Model:

$$\theta \sim \text{Beta}(\alpha, \beta)$$
$$y \sim \text{Binomial}(n, \theta)$$
$$\theta|y \sim \text{Beta}(\alpha + y, \beta + n - y)$$

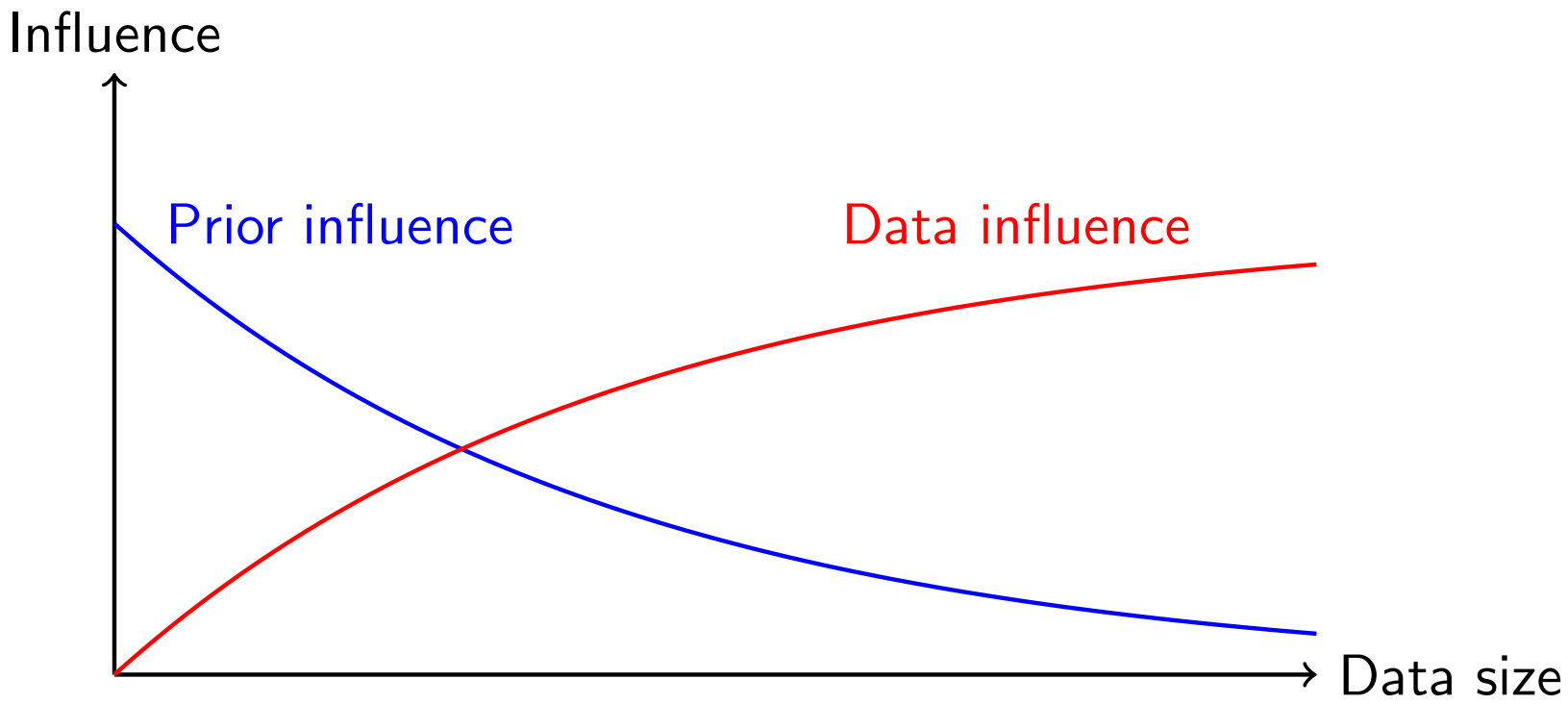
Example:

- Prior: Beta$(2, 8)$ (expect $\sim$20%)
- Data: 7 successes in 20 trials
- Posterior: Beta$(9, 21)$



Key Insights from Conjugate Models

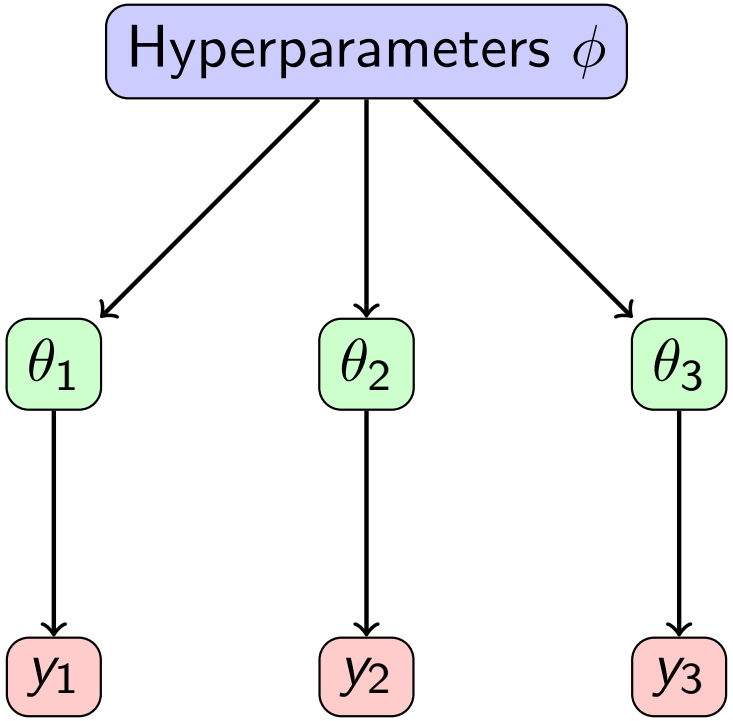
- Prior parameters act as "pseudo-counts"
- Posterior combines prior information with data
- With more data, likelihood dominates prior
- Conjugate priors give analytical solutions
- Posterior predictive distribution accounts for parameter uncertainty



The Bayesian Workflow

1. **Specify the model**
   - Likelihood: How data is generated given parameters
   - Prior: Initial beliefs about parameters
2. **Compute the posterior**
   - Analytical solutions (conjugate models)
   - Numerical methods (MCMC, variational inference)
3. **Check the model**
   - Prior/posterior predictive checks
   - Sensitivity analysis
4. **Use the model**
   - Inference about parameters
   - Predictions for new data
   - Decision-making under uncertainty

Hierarchical Models: Sharing Information



Key idea: Parameters come from a common distribution

Benefits:

- Partial pooling of information
- Robust to small sample sizes
- Reduces overfitting
- Models variation between groups

Example: School effects, hospital outcomes, sports analytics

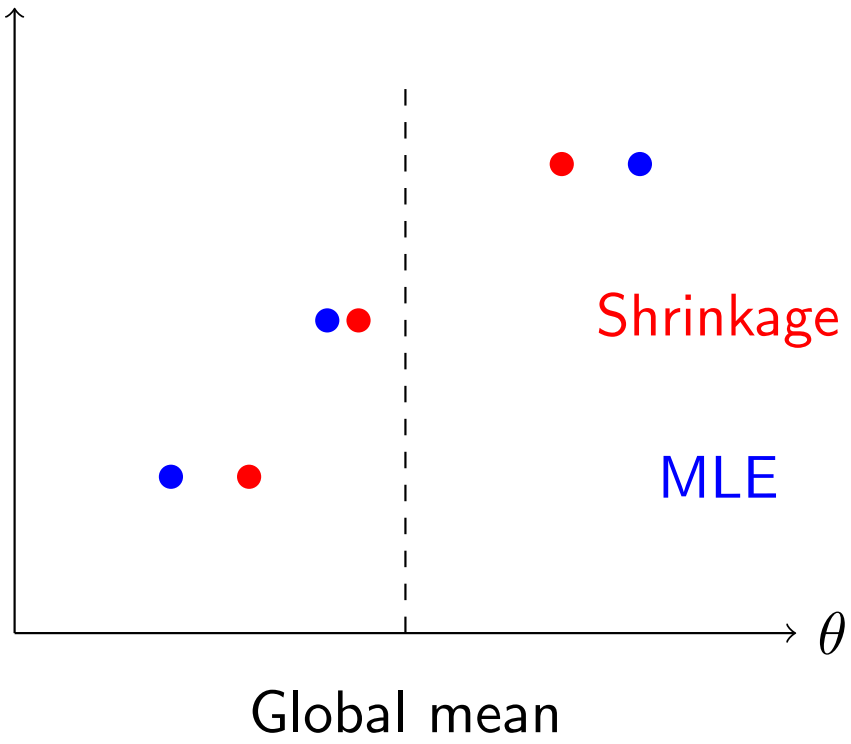
Shrinkage: The Bayesian Perspective

Problem: Estimating multiple means
MLE (no shrinkage):

$$\hat{\theta}_i = y_i$$

Bayesian (with shrinkage):

$$\hat{\theta}_i = \frac{\frac{1}{\sigma_i^2} y_i + \frac{1}{\tau^2} \mu}{\frac{1}{\sigma_i^2} + \frac{1}{\tau^2}}$$



Key insights:

- Extreme estimates pulled toward the global mean
- Amount of shrinkage depends on uncertainty
- Less shrinkage with more data
- Improves overall accuracy (lower MSE)

The Bayesian Approach to Regularization

Frequentist view:

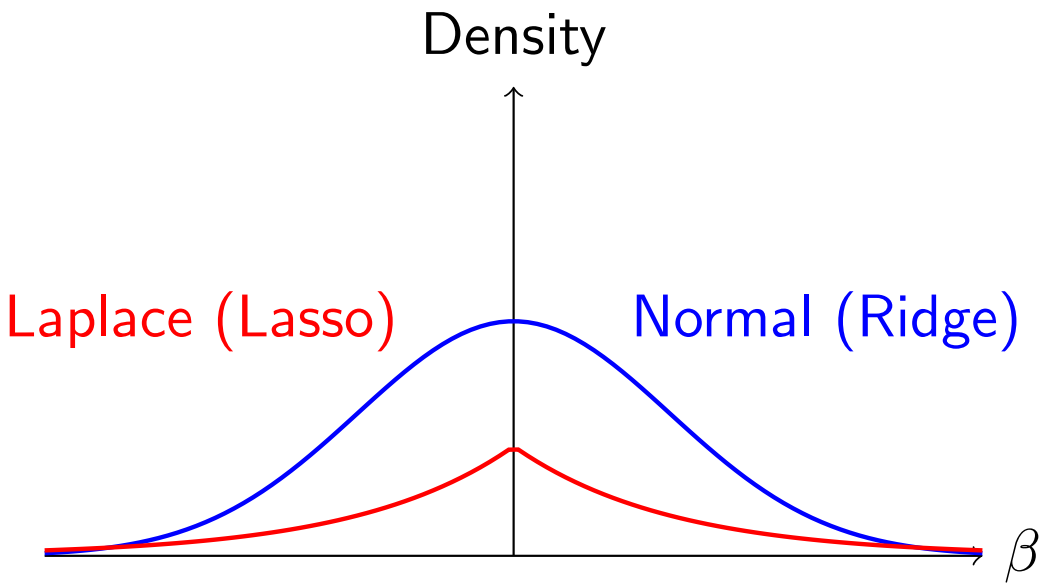
$$\hat{\beta}_{\text{ridge}} = \arg\min_{\beta} \|y - X\beta\|^2 + \lambda\|\beta\|^2$$

$$\hat{\beta}_{\text{lasso}} = \arg\min_{\beta} \|y - X\beta\|^2 + \lambda\|\beta\|_1$$

Bayesian view:

$$\beta_{\text{ridge}} \sim \mathcal{N}(0, \tau^2 I)$$

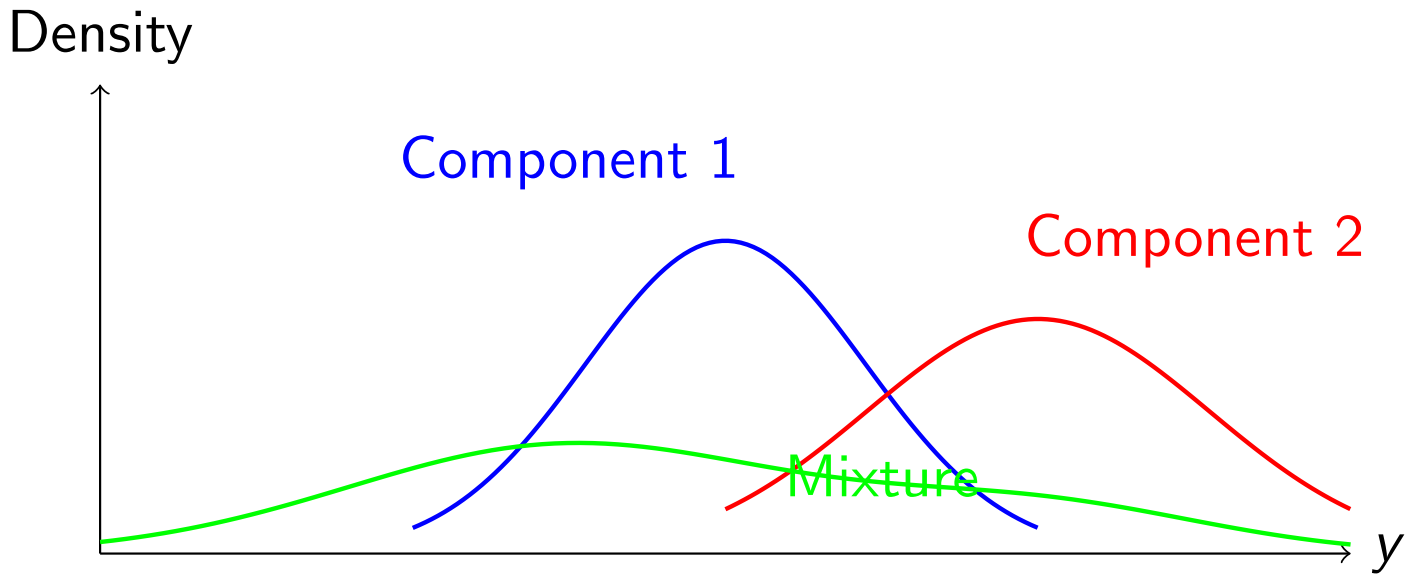
$$\beta_{\text{lasso}} \sim \text{Laplace}(0, b)$$



Advantages: Uncertainty quantification, adaptive regularization, model selection

Mixture Models: Discovering Latent Structure

$$p(y) = \sum_{k=1}^{K} \pi_k p(y|\theta_k)$$



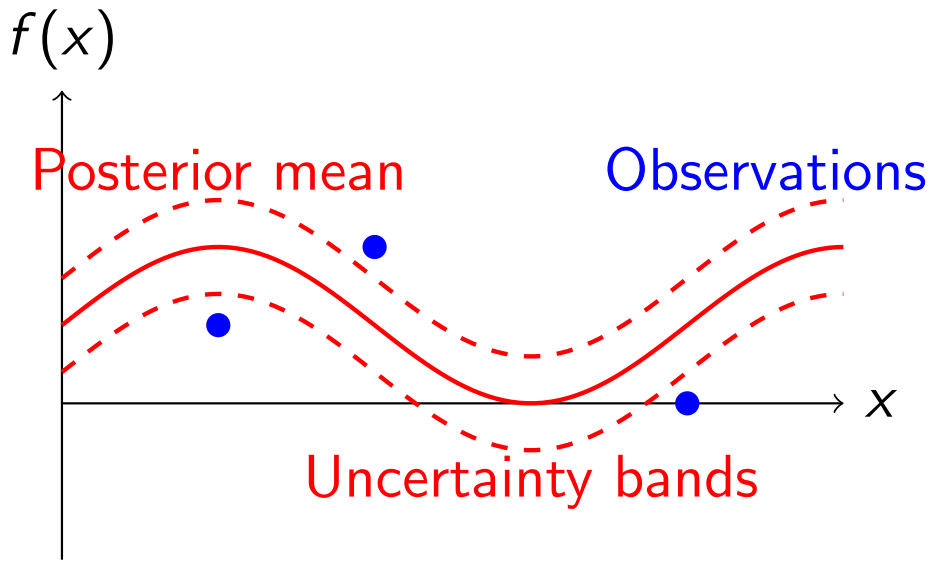
Bayesian approach:

- Place priors on mixture weights π_k and component parameters θ_k
- Infer number of components K (e.g., using Dirichlet process)
- Obtain full posterior over all parameters

Applications: Clustering, density estimation, classification

Gaussian Processes: Flexible Function Modeling

$$f(x) \sim \mathcal{GP}(m(x), k(x, x'))$$
$$y_i = f(x_i) + \epsilon_i$$



Key properties:

- Non-parametric - flexibility increases with data
- Kernel encodes function properties (smoothness, periodicity)
- Provides uncertainty bands
- Handles irregular sampling naturally

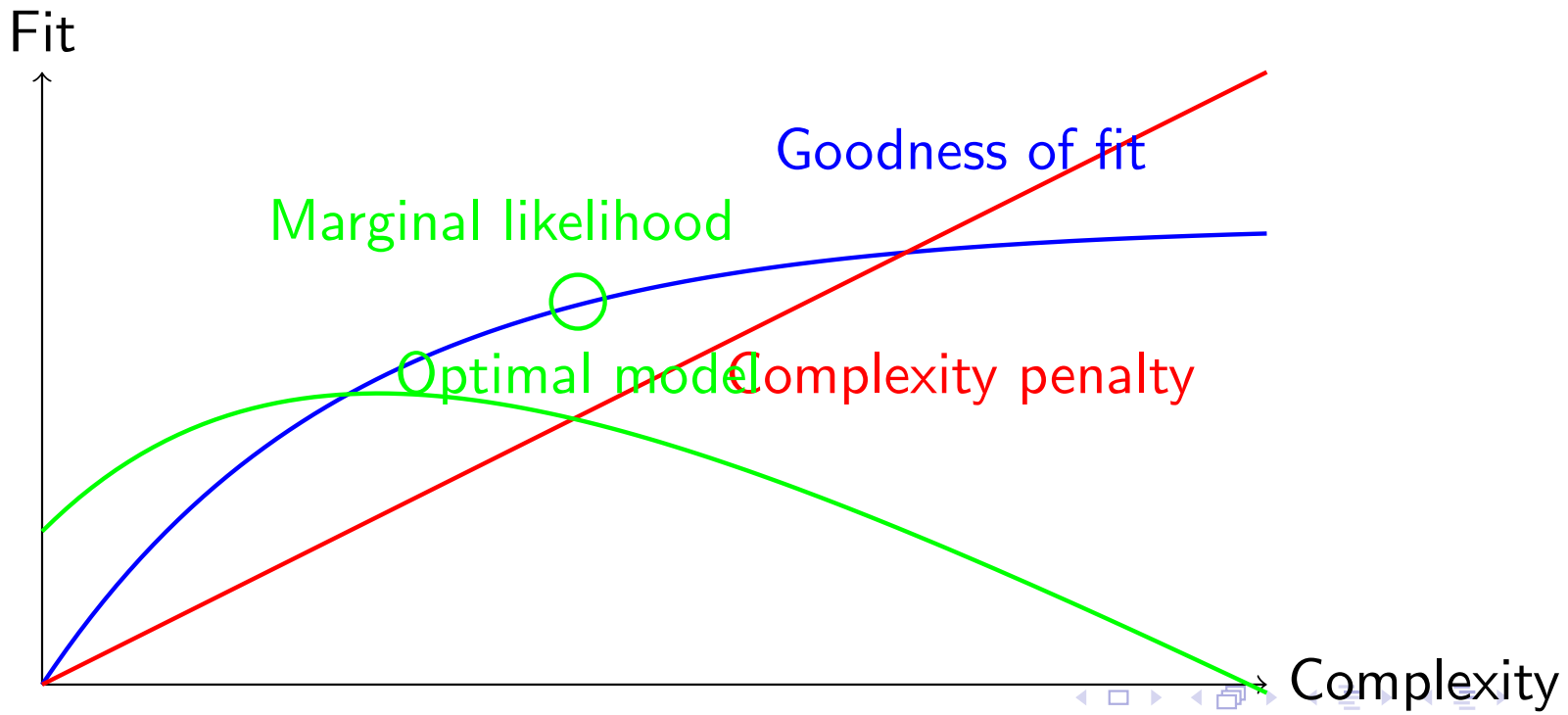
Bayesian Model Comparison

Bayes factors:

$$\mathrm{BF}_{12} = \frac{P(D|M_1)}{P(D|M_2)} = \frac{\int P(D|\theta_1, M_1)P(\theta_1|M_1)d\theta_1}{\int P(D|\theta_2, M_2)P(\theta_2|M_2)d\theta_2}$$

Posterior model probabilities:

$$P(M_k|D) = \frac{P(D|M_k)P(M_k)}{\sum_j P(D|M_j)P(M_j)}$$



Bayesian Computation: From Theory to Practice

Analytical solutions:

- Conjugate models (Beta-Binomial, Normal-Normal)
- Limited to simple models

Markov Chain Monte Carlo (MCMC):

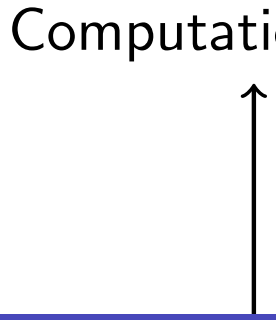
- Gibbs sampling, Metropolis-Hastings
- Hamiltonian Monte Carlo (HMC)
- Flexible but can be slow

Variational inference:

- Approximates posterior with simpler distribution
- Faster but less accurate than MCMC
- Good for large-scale problems

Software tools:

- Stan, PyMC, JAGS
- TensorFlow Probability, Pyro
- R packages (brms, rstanarm)



Practical Lessons in Bayesian Statistics

1. **Start simple, then add complexity**
   - Begin with simple models you understand
   - Add complexity only when needed
2. **Prior specification matters**
   - Use weakly informative priors when possible
   - Check sensitivity to prior choices
3. **Model checking is essential**
   - Prior/posterior predictive checks
   - Residual analysis
   - Cross-validation
4. **Computational diagnostics are crucial**
   - Check MCMC convergence (R-hat, ESS)
   - Beware of divergences and other warnings
5. **Communicate uncertainty honestly**
   - Show full posterior distributions
   - Avoid overconfident conclusions

Why Bayesian Methods Matter

Principled uncertainty quantification

- Full posterior distributions
- Credible intervals with direct probability interpretation

Incorporating prior knowledge

- Formalize domain expertise
- Stabilize inference with small data

Hierarchical modeling

- Share information across groups
- Balance between pooled and separate analyses

Handling missing data and latent variables

- Natural framework for incomplete information
- Integrates over uncertainty in unobserved variables

Questions?

Thank you!

Questions and Discussion