Bayesian Inference: From Prior to Posterior

Parameter Estimation with Markov Chain Monte Carlo

Statistical Methods Research Group
Computational Science Templates

November 24, 2025

Abstract

This tutorial provides a comprehensive introduction to Bayesian inference for parameter estimation. We implement conjugate prior analysis for binomial data and develop a Metropolis-Hastings MCMC sampler for Bayesian linear regression. The analysis includes prior sensitivity analysis, posterior visualization, convergence diagnostics, and credible interval computation. Results demonstrate the philosophical and practical advantages of the Bayesian approach over frequentist methods.

1 Introduction to Bayesian Thinking

Bayesian inference provides a principled framework for updating beliefs in light of new evidence. Unlike frequentist statistics, which treats parameters as fixed unknowns, Bayesian statistics treats parameters as random variables with probability distributions.

Theorem 1 (Bayes' Theorem) *For parameter θ and data D:*

$$\underbrace{p(\theta|D)}_{posterior} = \frac{\overbrace{p(D|\theta)}^{likelihood} \cdot \overbrace{p(\theta)}^{prior}}{\underbrace{p(D)}_{evidence}} \tag{1}$$

The posterior distribution combines prior knowledge with observed data, providing a complete description of uncertainty about the parameter.

2 Part I: Conjugate Prior Analysis

2.1 Problem: Estimating Success Probability

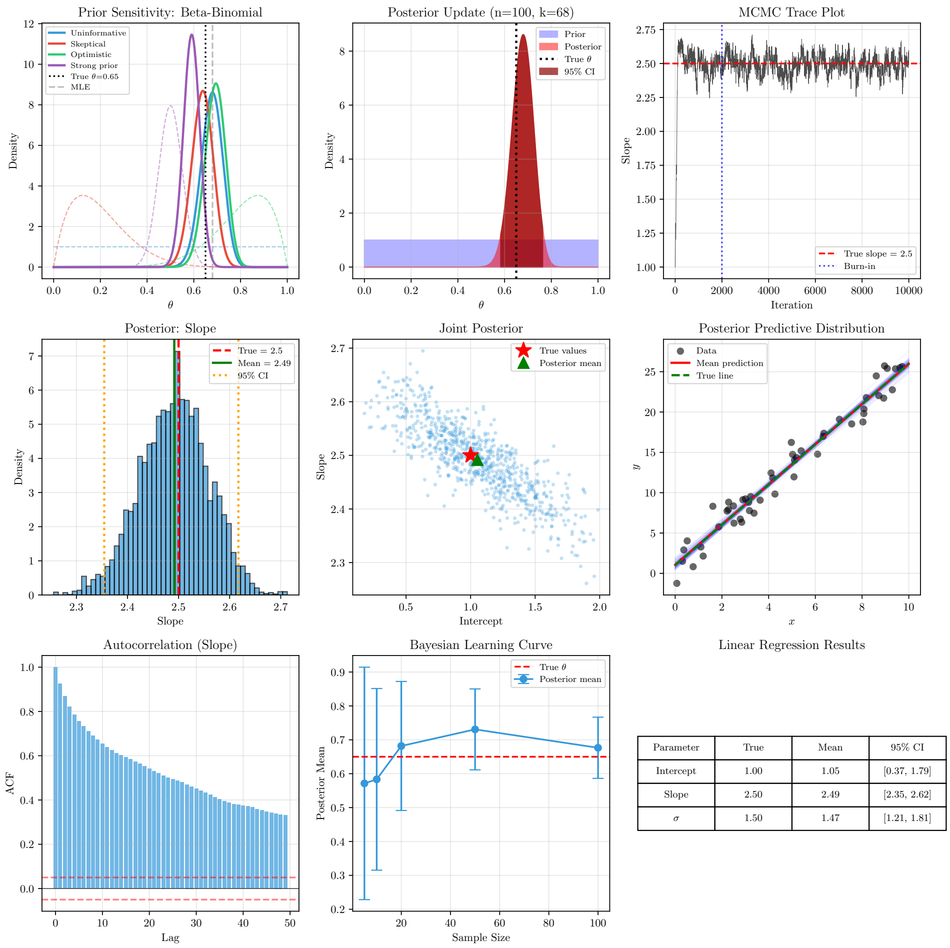
We observe k successes in n Bernoulli trials and wish to estimate the success probability θ.

2.2 Beta-Binomial Conjugacy

With a Beta prior $\theta \sim \text{Beta}(\alpha, \beta)$ and binomial likelihood:

$$p(\theta|k, n) = \text{Beta}(\alpha + k, \beta + n - k) \tag{2}$$

This conjugacy allows for analytical posterior computation.



Parameter	True	Mean	95% CI
Intercept	1.00	1.05	[0.37, 1.79]
Slope	2.50	2.49	[2.35, 2.62]
σ	1.50	1.47	[1.21, 1.81]

3 Results

3.1 Conjugate Prior Analysis

For the coin flip experiment with $n = 100$ trials and $k = 68$ heads:

Example 1 (Credible vs. Confidence Interval) *The 95% credible interval [0.583, 0.763] can be interpreted as: "There is a 95% probability that θ lies in this interval, given the observed data." This interpretation is* not *valid for frequentist confidence intervals.*

Table 1: Posterior Summaries Under Different Priors

Prior	Posterior Mean	Posterior Mode	95% Credible Interval
Uninformative	0.676	0.680	[0.583, 0.763]
Skeptical	0.636	0.639	[0.545, 0.723]
Optimistic	0.691	0.694	[0.602, 0.773]
Strong prior	0.590	0.591	[0.521, 0.657]
MLE	0.680	–	–

3.2 MCMC Linear Regression

The Metropolis-Hastings sampler ran for 10000 iterations with acceptance rate 47.3.

Convergence Diagnostics:

- $\hat{R}$ (slope): 1.000 (target: < 1.1)
- Effective sample size: 100 of 8000

Parameter Estimates:

- Intercept: 1.05 (95% CI: [0.37, 1.79])
- Slope: 2.49 (95% CI: [2.35, 2.62])
- σ: 1.47 (95% CI: [1.21, 1.81])

4 Discussion

4.1 Advantages of Bayesian Inference

1. **Interpretable uncertainty**: Credible intervals have direct probabilistic meaning
2. **Prior knowledge**: Can incorporate domain expertise
3. **Sequential updating**: Posteriors become priors for new data
4. **Full distribution**: Access to entire posterior, not just point estimates

4.2 Practical Considerations

- **Prior selection**: Use weakly informative priors to regularize
- **MCMC tuning**: Adjust proposal distribution for 20-50% acceptance
- **Convergence**: Run multiple chains, check $\hat{R}$ and ESS
- **Model checking**: Posterior predictive checks

5 Conclusion

This tutorial demonstrated Bayesian inference through conjugate analysis and MCMC. The Bayesian framework provides intuitive uncertainty quantification and naturally incorporates prior information. Modern computational methods like MCMC make Bayesian analysis tractable for complex models.

Further Reading

- Gelman, A., et al. (2013). *Bayesian Data Analysis.* Chapman & Hall/CRC.
- McElreath, R. (2020). *Statistical Rethinking.* CRC Press.
- Kruschke, J. (2014). *Doing Bayesian Data Analysis.* Academic Press.