Feb 23

1. MAP vs MCMC
2. Hamiltonian MC
3. Assessing convergence
4. What can go wrong

Approximating the posterior

Recall: simple binomial model

5 trials
4 'successes'

$$4 \sim \text{Binom}(5, p)$$

$$p \sim \text{Beta}(1, 1)$$

Bayes' Rule

$$\Pr(p \mid n = 5, k = 4) = \frac{\Pr(k = 4 \mid n = 5, p)\Pr(p)}{\Pr(k = 4)}$$

Approximating the posterior

Recall: simple binomial model

5 trials
4 'successes'

$$4 \sim \text{Binom}(5, p)$$

$$p \sim \text{Beta}(1, 1)$$

Bayes' Rule

$$\Pr(p|n = 5, k = 4) = \frac{\Pr(k = 4|n = 5, p)\Pr(p)}{\Pr(k = 4)}$$

$$\propto \Pr(k = 4|n = 5, p)\Pr(p)$$

$$= \binom{5}{4} p^4 (1 - p)^1 \times 1$$

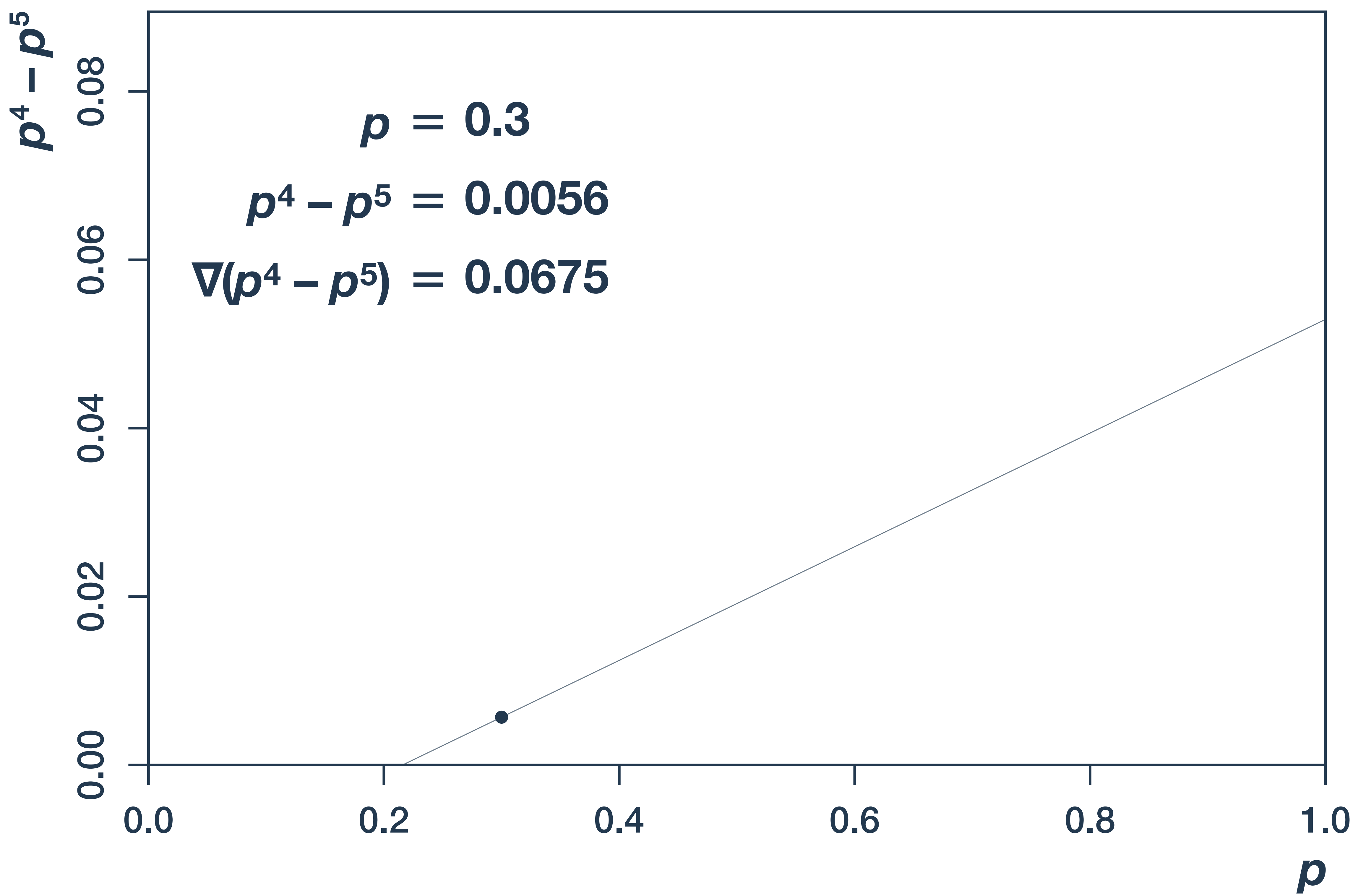
$$\propto \boxed{p^4 - p^5}$$

The posterior distribution for p is proportional to this

Maximum a posteriori

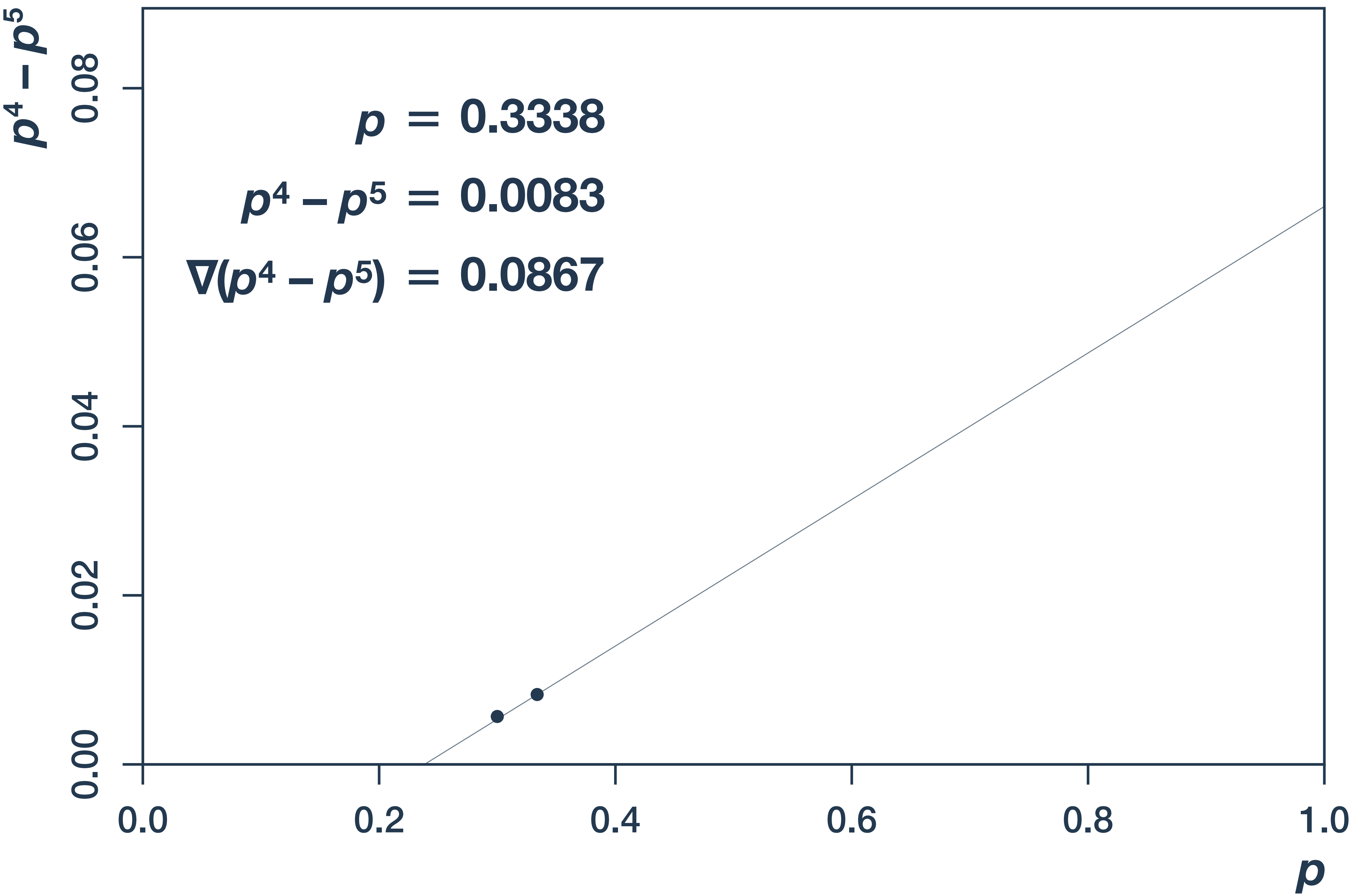
Maximum *a posteriori*

$\Pr(p|data) \propto p^4 - p^5$



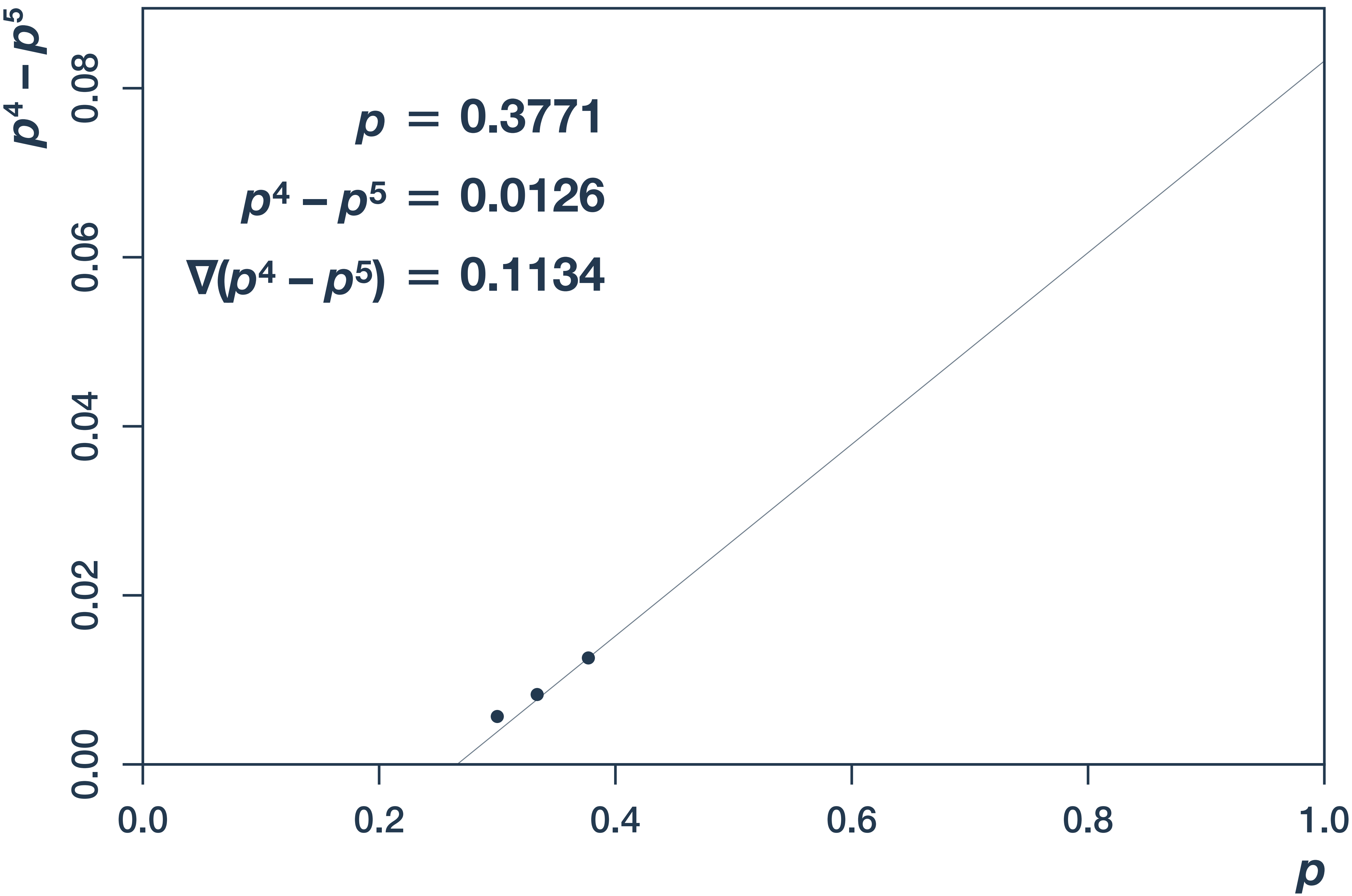
Maximum *a posteriori*

$$\Pr(p \mid data) \propto p^4 - p^5$$



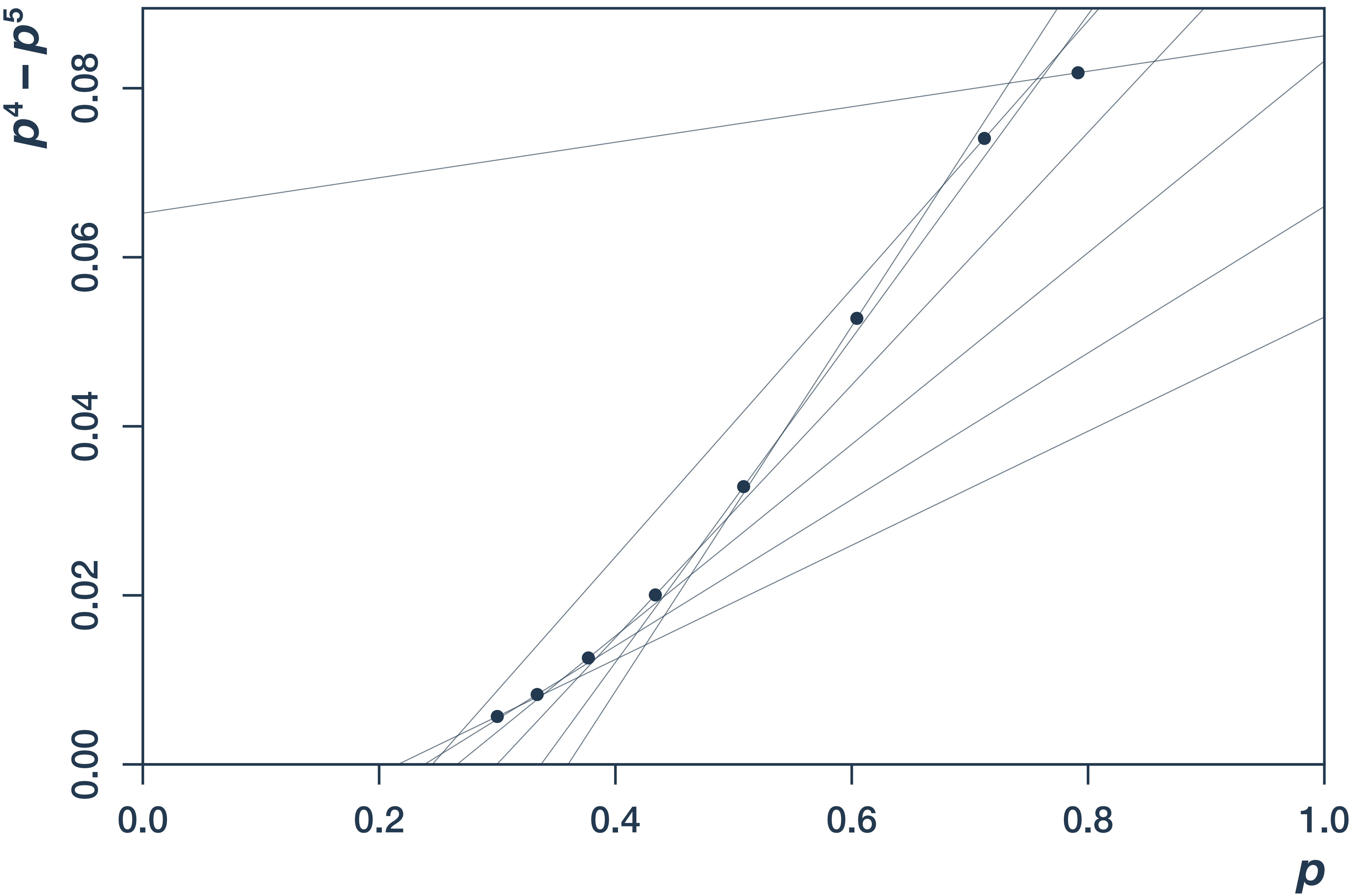
Maximum *a posteriori*

$$\Pr(p|data) \propto p^4 - p^5$$



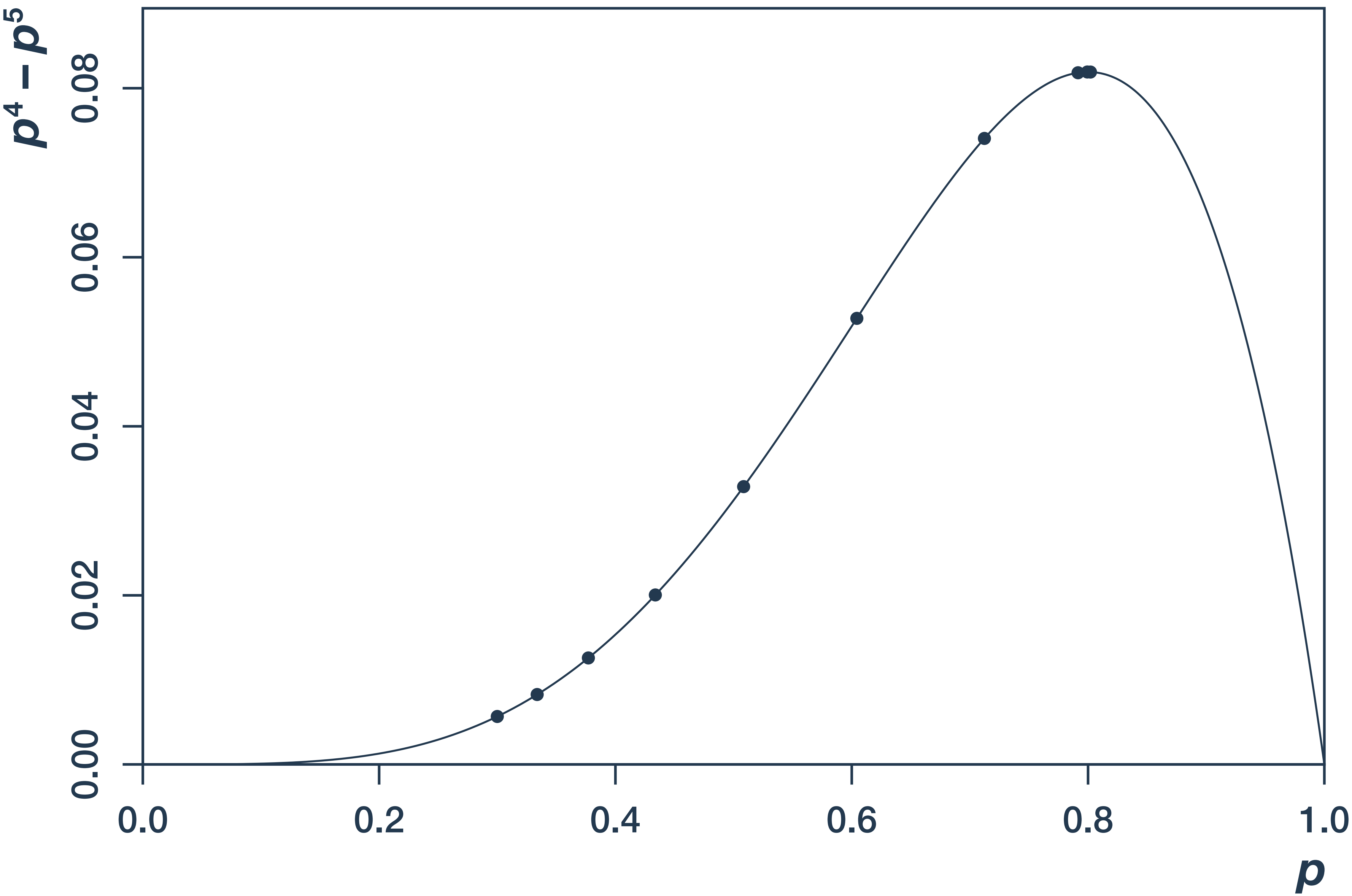
Maximum *a posteriori*

$$\Pr(p|data) \propto p^4 - p^5$$

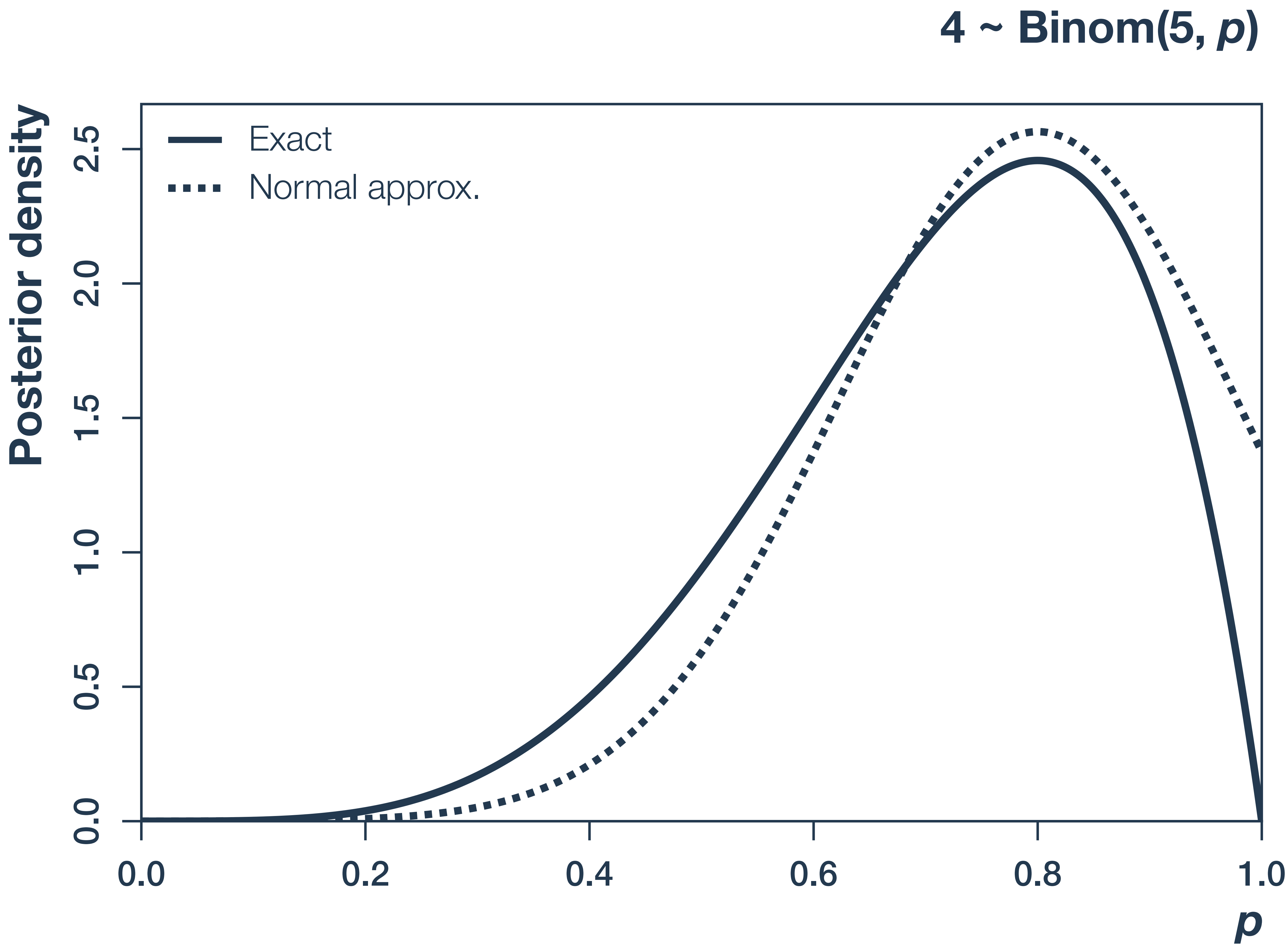


Maximum *a posteriori*

$$\Pr(p|data) \propto p^4 - p^5$$



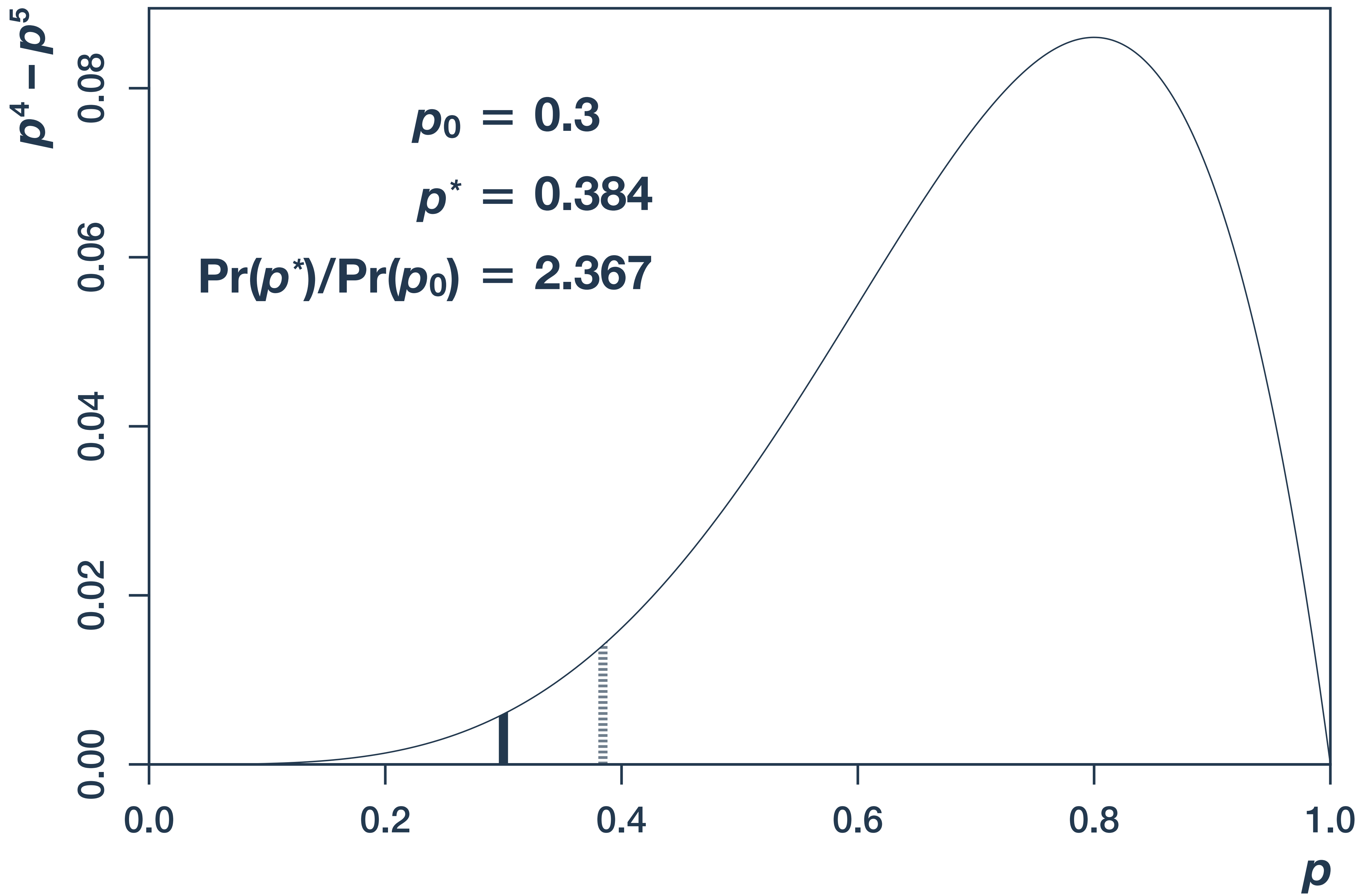
Normal approximation



Markov chain Monte Carlo

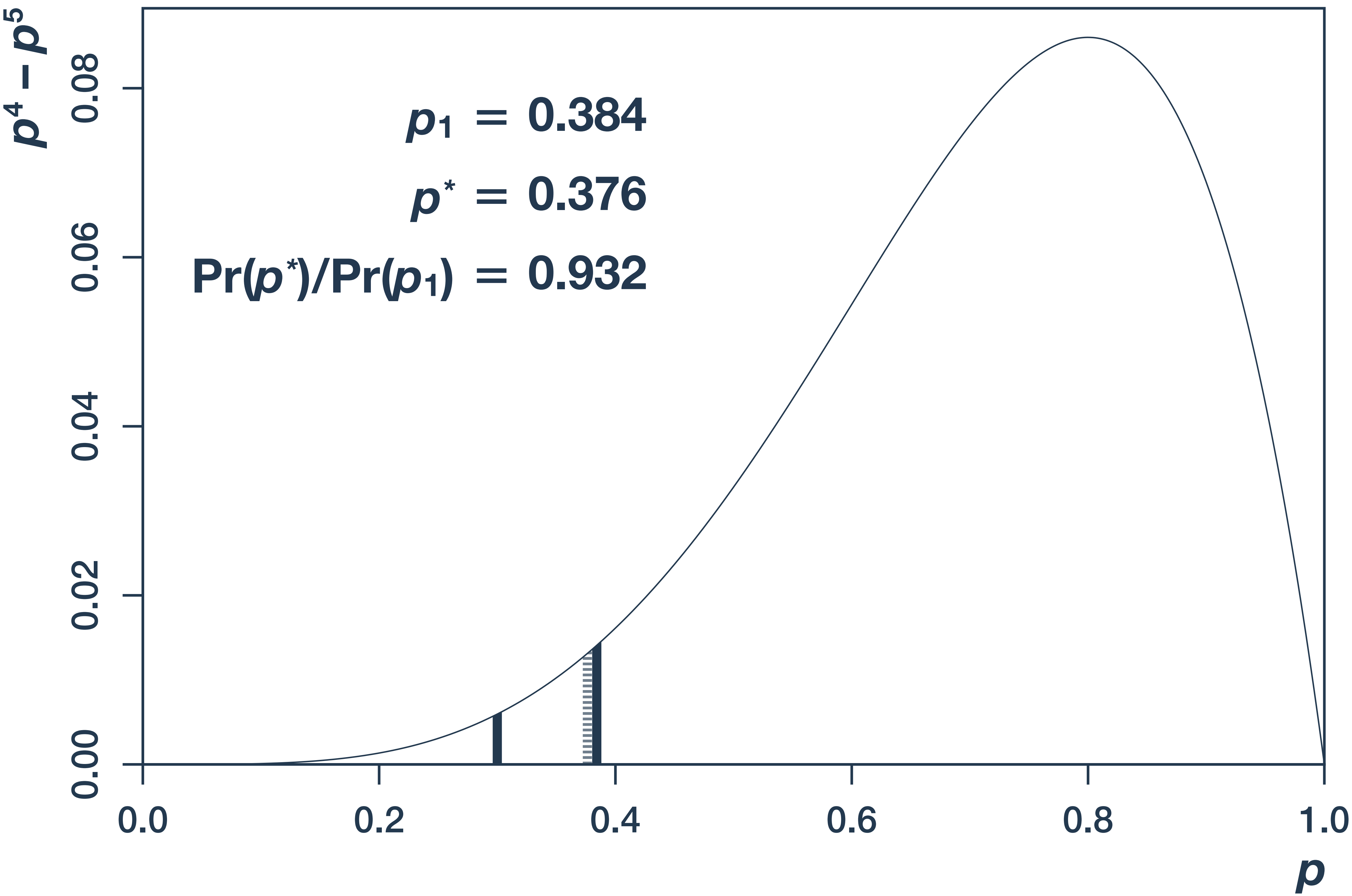
Markov chain Monte Carlo

$\Pr(p|data) \propto p^4 - p^5$



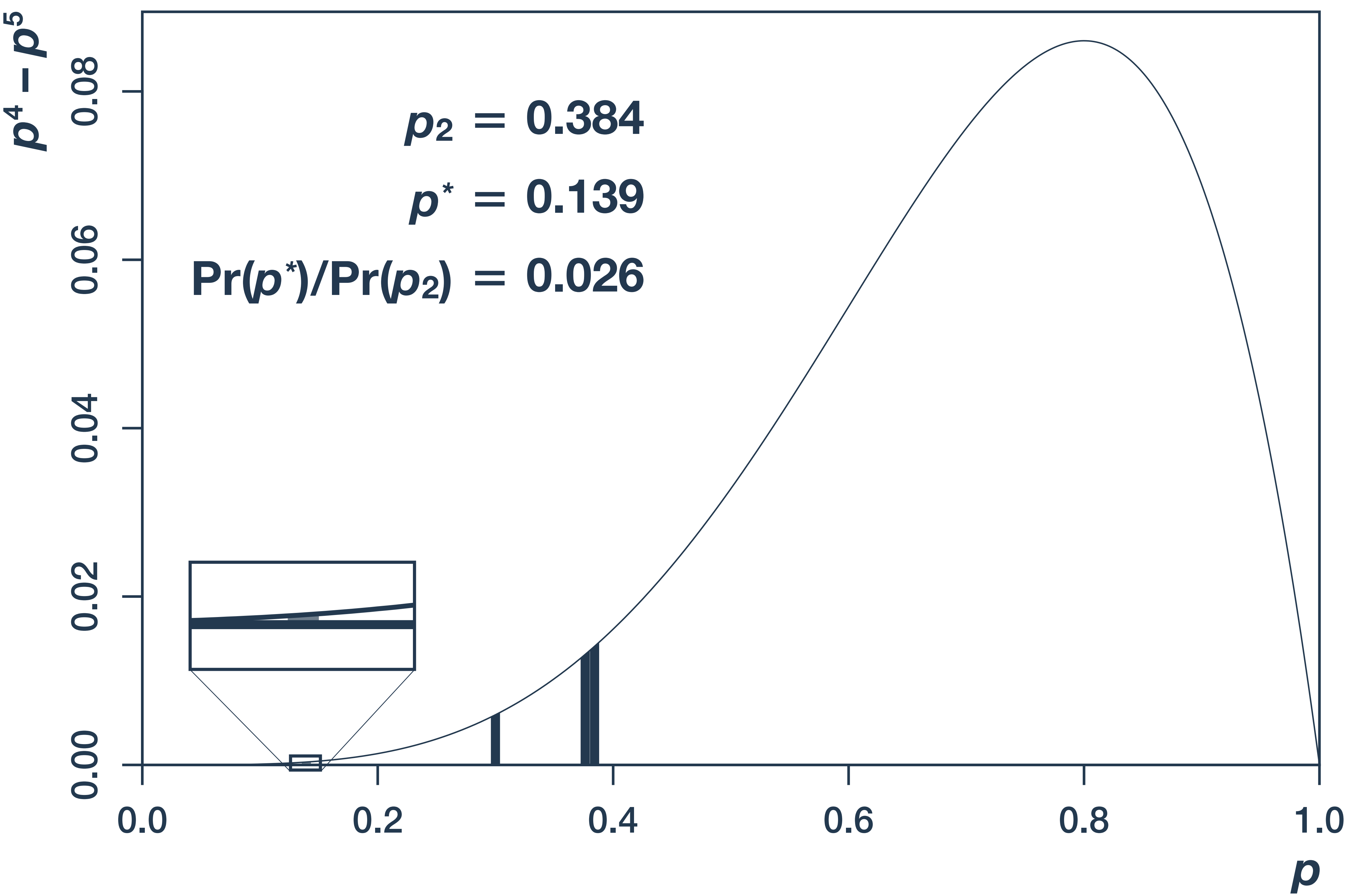
Markov chain Monte Carlo

$$\Pr(p|data) \propto p^4 - p^5$$



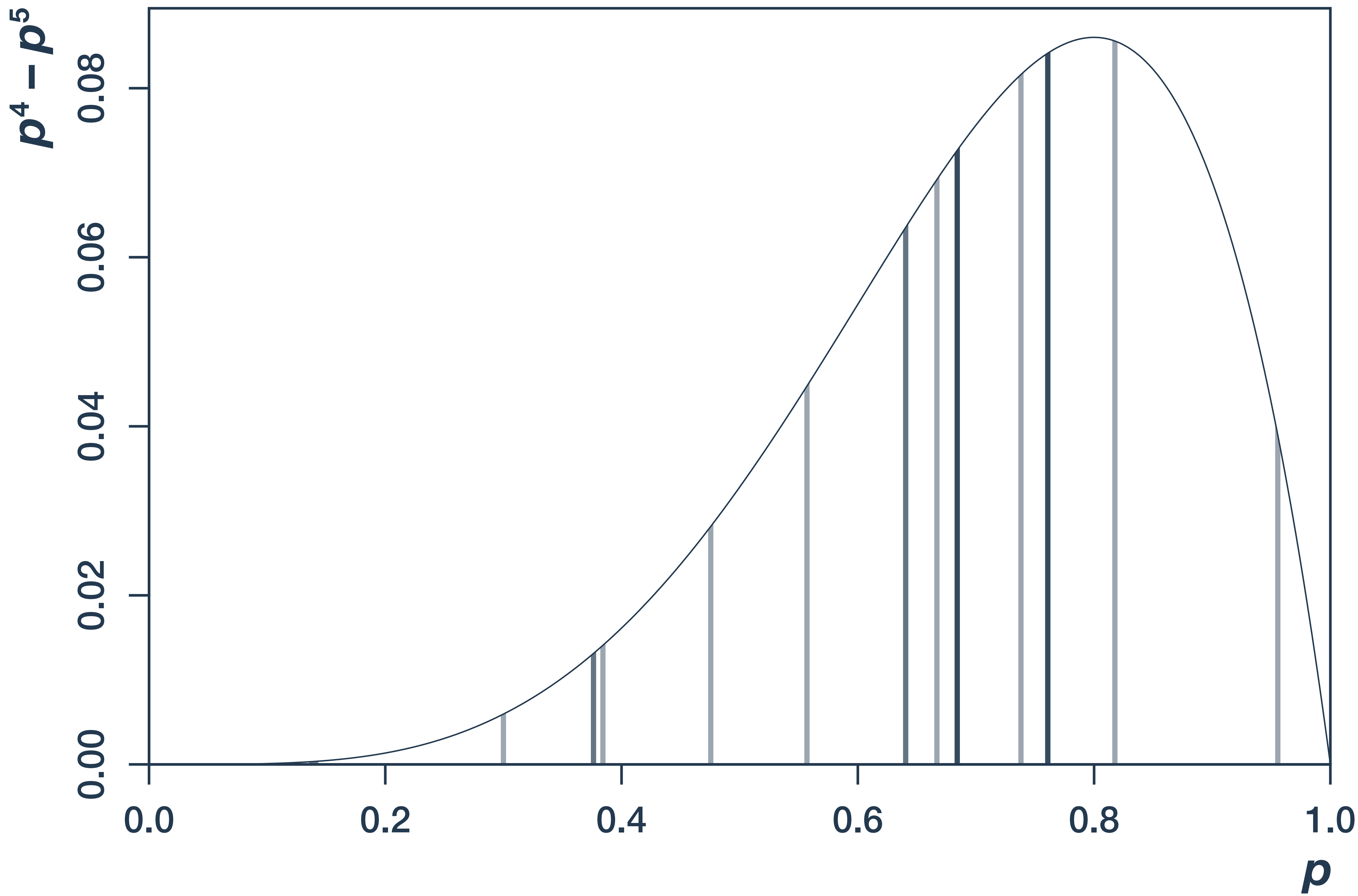
Markov chain Monte Carlo

$$\Pr(p|data) \propto p^4 - p^5$$

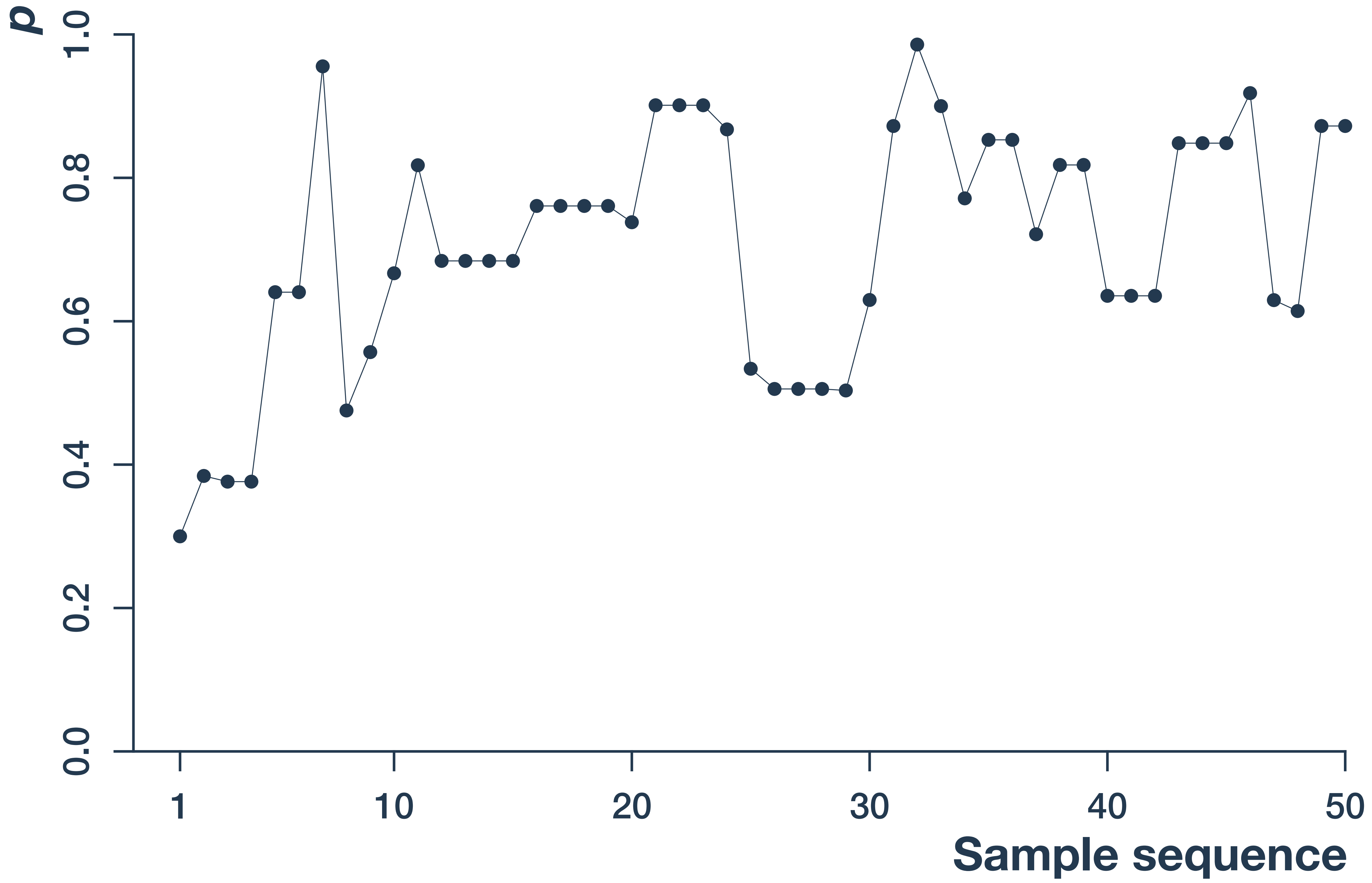


Markov chain Monte Carlo

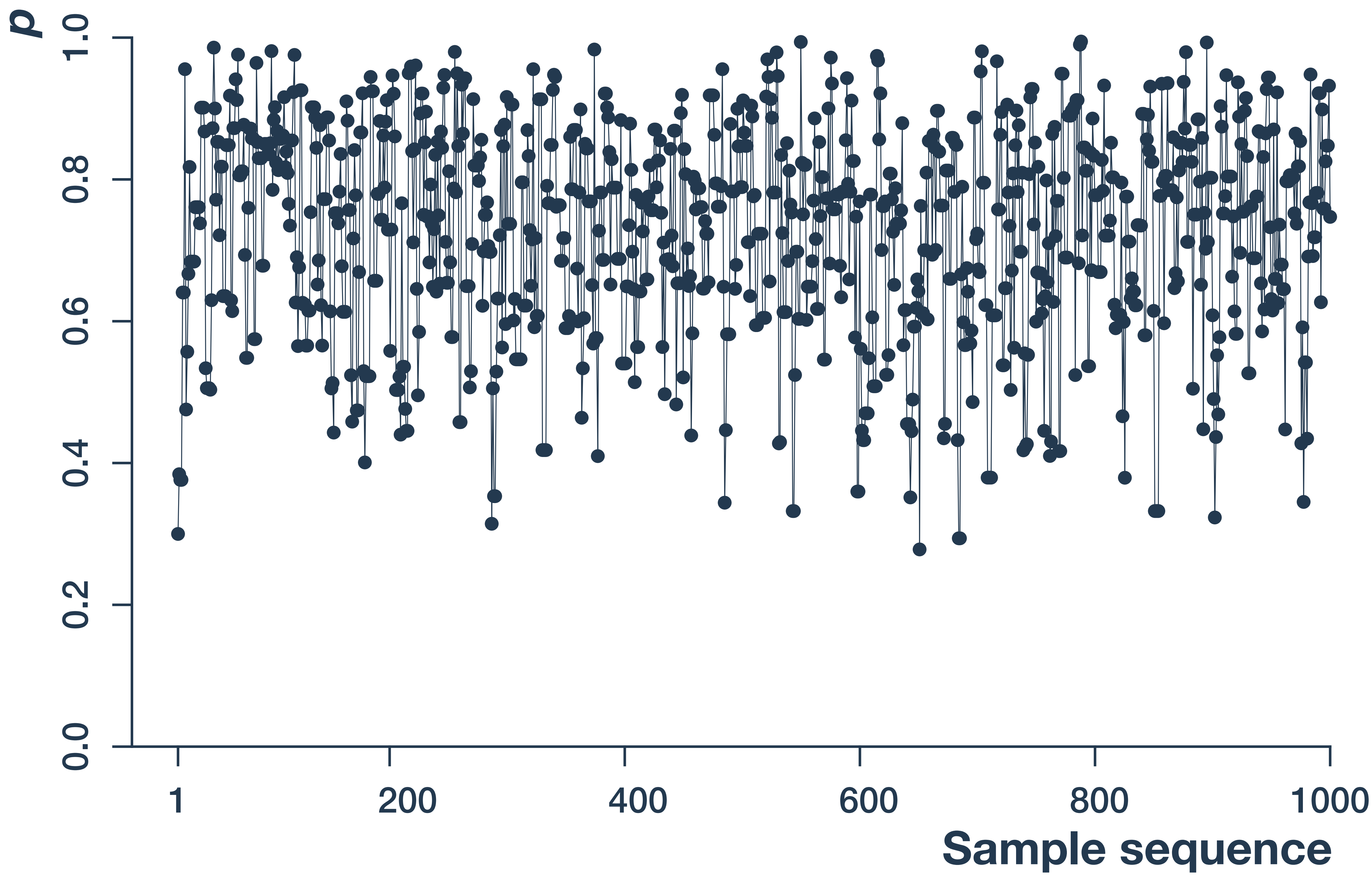
$\Pr(p|data) \propto p^4 - p^5$



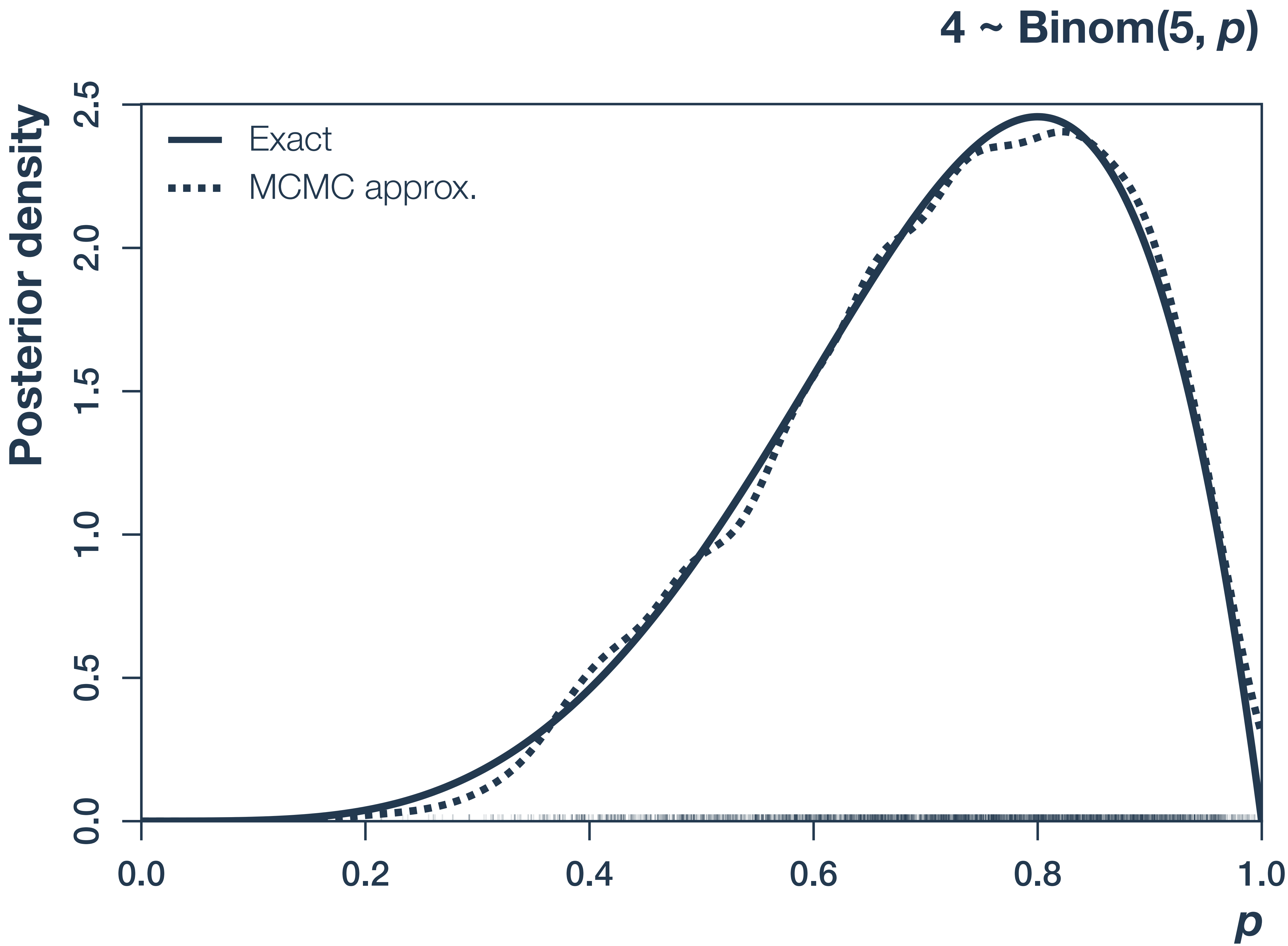
Markov chain Monte Carlo



Markov chain Monte Carlo



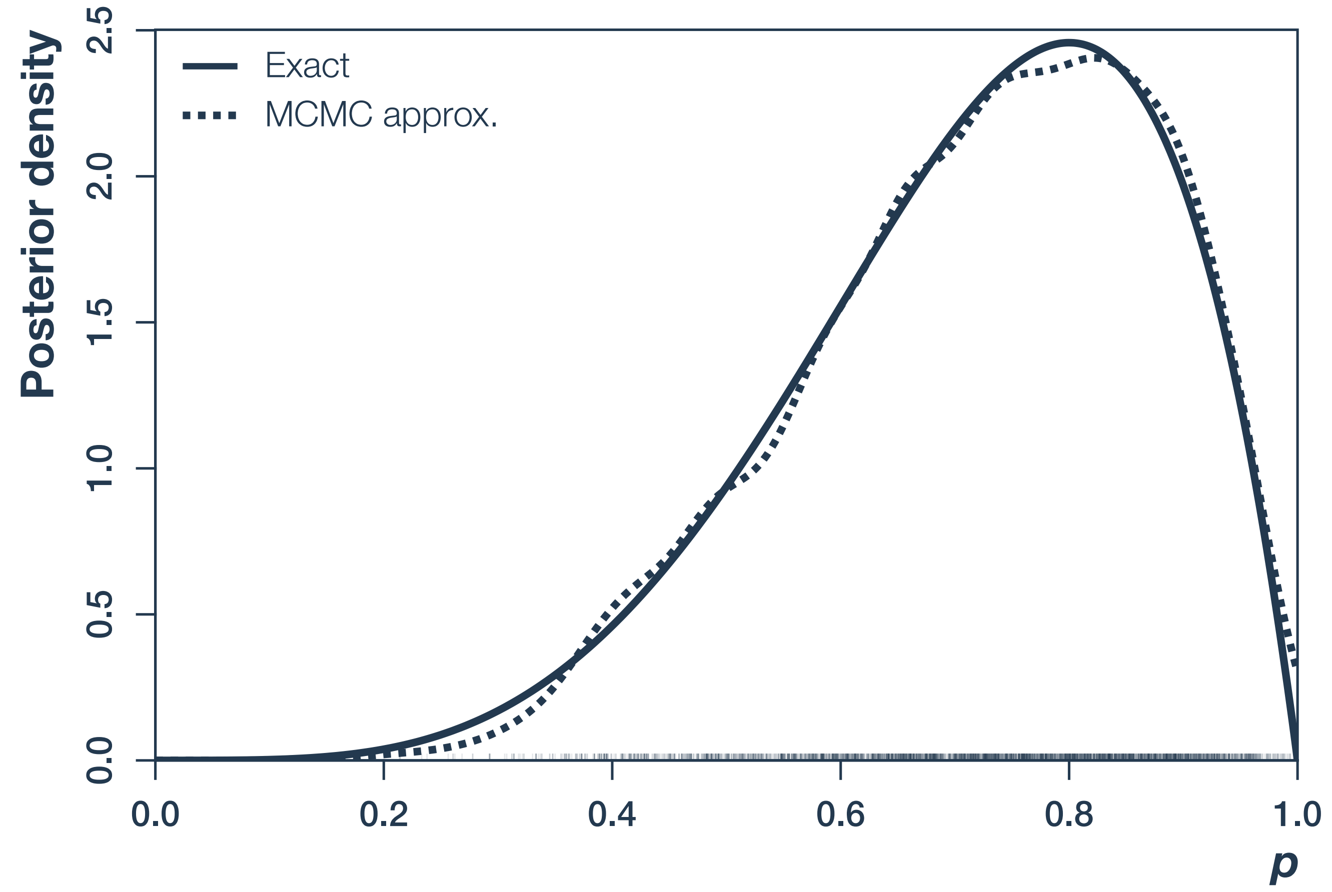
MCMC approximation



MCMC vs. MAP

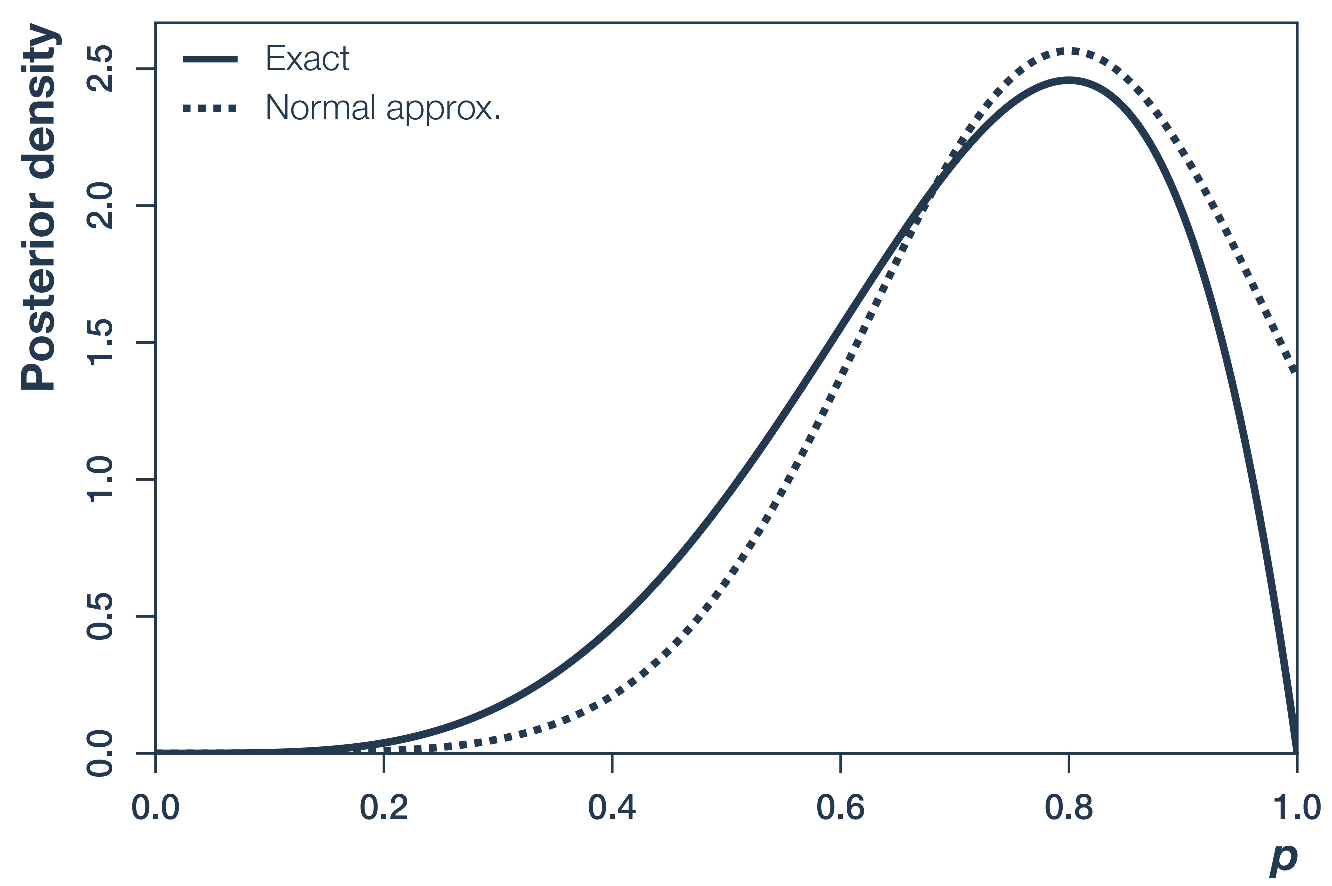
Markov chain Monte Carlo

Approximates posterior using a large approximate sample.



Maximum *a posteriori*

Approximates posterior by finding its mode and approximating with a normal distribution.



Hamiltonian Monte Carlo

Hamiltonian Monte Carlo

Simulate a physical system

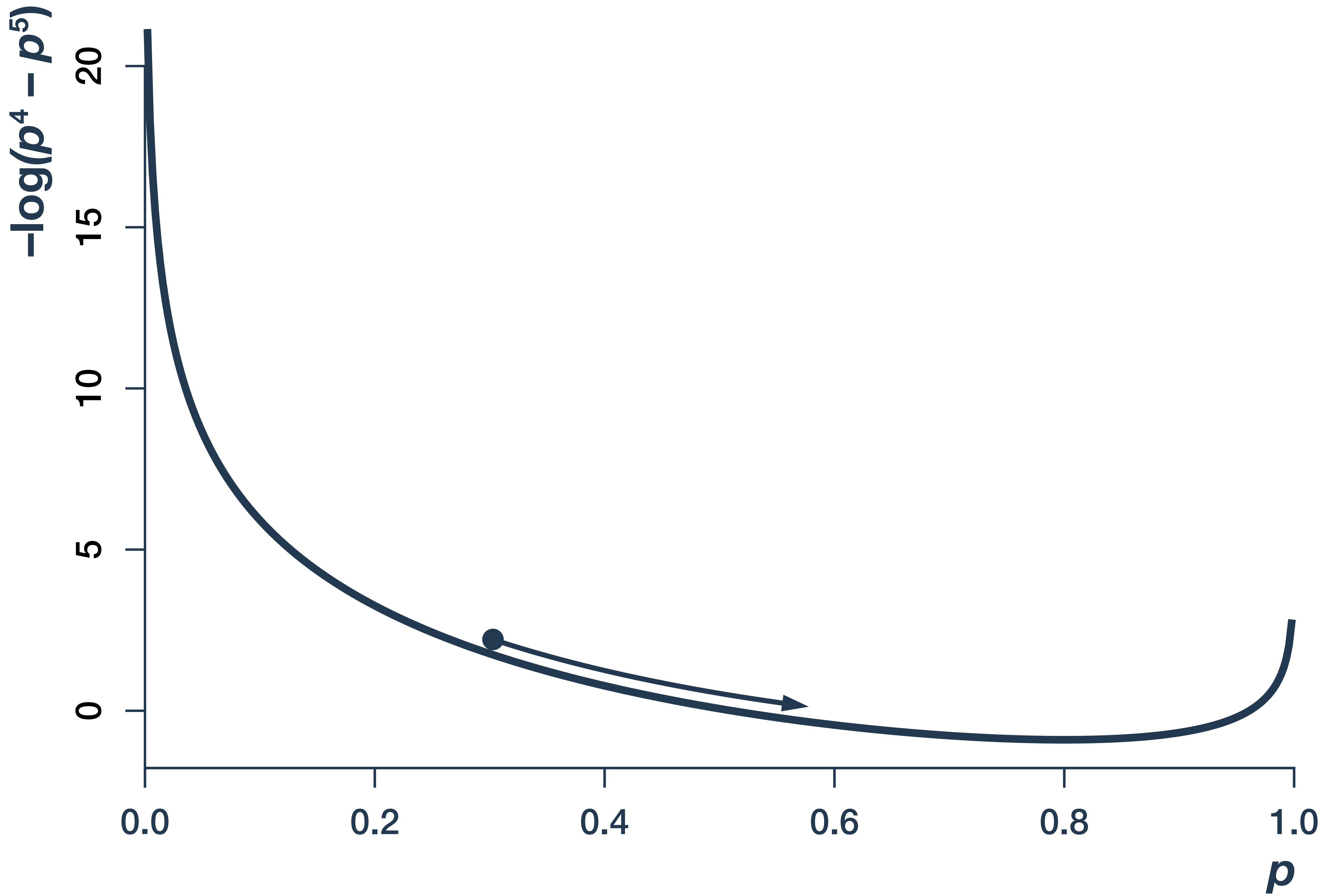
Energy at any point in the parameter space is proportional to the negative log likelihood of the posterior.

Random draws by perturbing a particle in that system

Place a particle in that system, give it a push in a random direction, and use Hamiltonian dynamics to simulate its motion.

Wherever the particle ends up after a fixed number of iterations is the next draw from the posterior.

Hamiltonian Monte Carlo



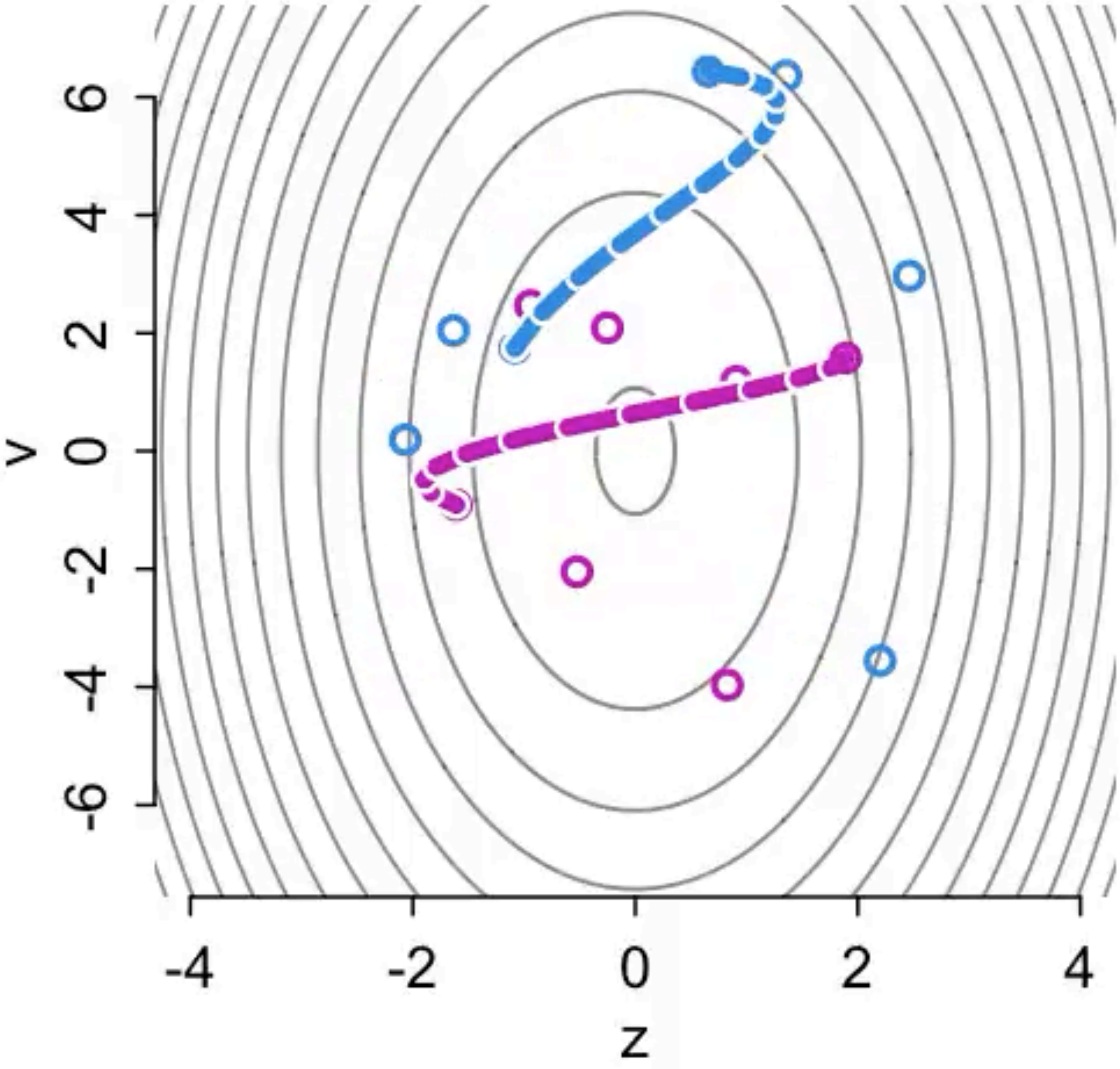
HMC vs. MCMC

Takes advantage of gradient	Gradient (slope) information helps HMC adjust to the shape of the posterior.
Reduces autocorrelation	HMC tends to explore the plausible areas of the parameter space much more quickly than ‘standard’ MCMC like Metropolis–Hastings. It is not likely to spend too much time in one small area.

No-U-Turn sampler (NUTS)

A version of HMC that automatically optimizes some of the meta-parameters of the algorithm.

Hamiltonian Monte Carlo

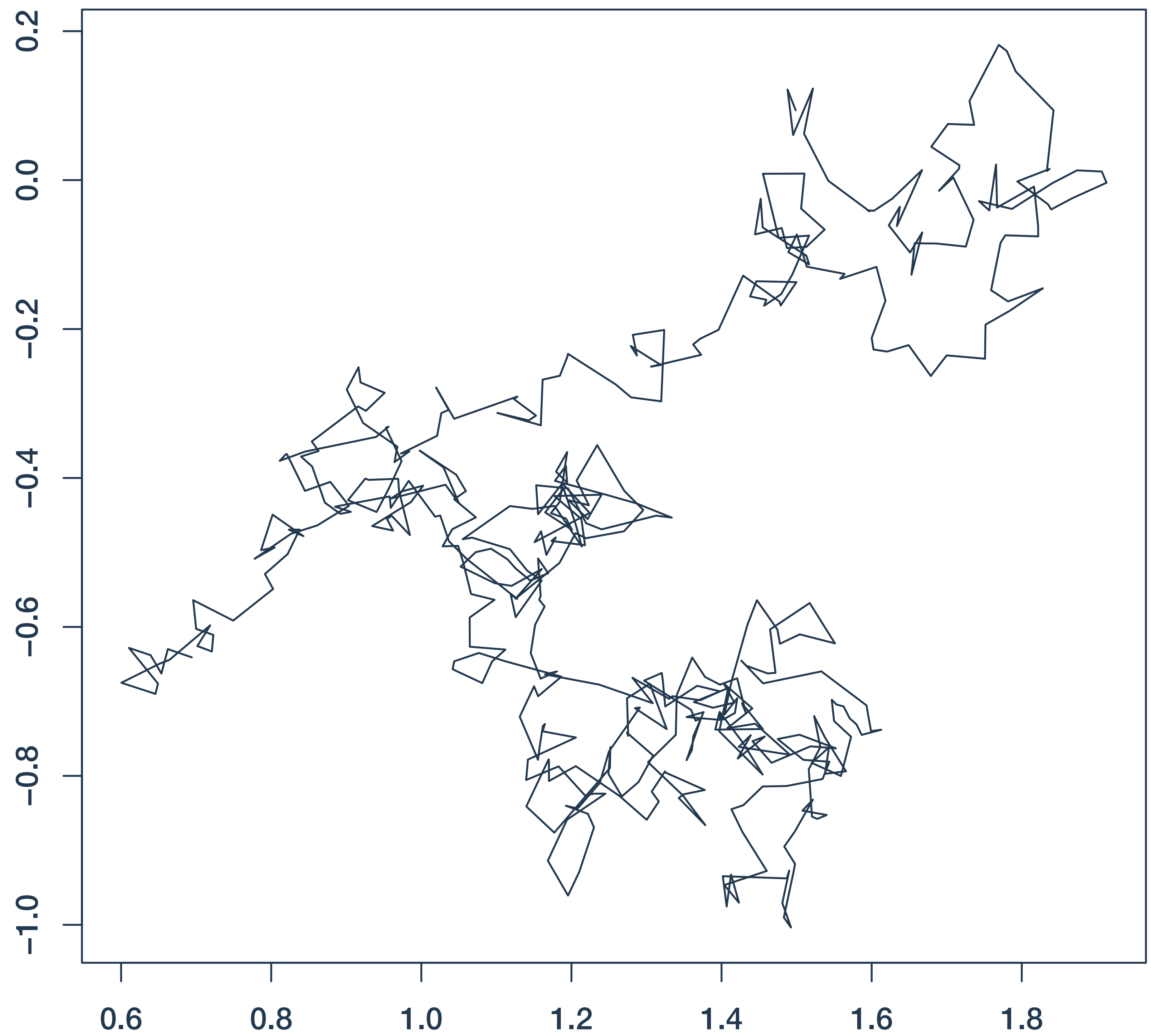


Animation credit: @rlmcelreath

What can go wrong?

Autocorrelation

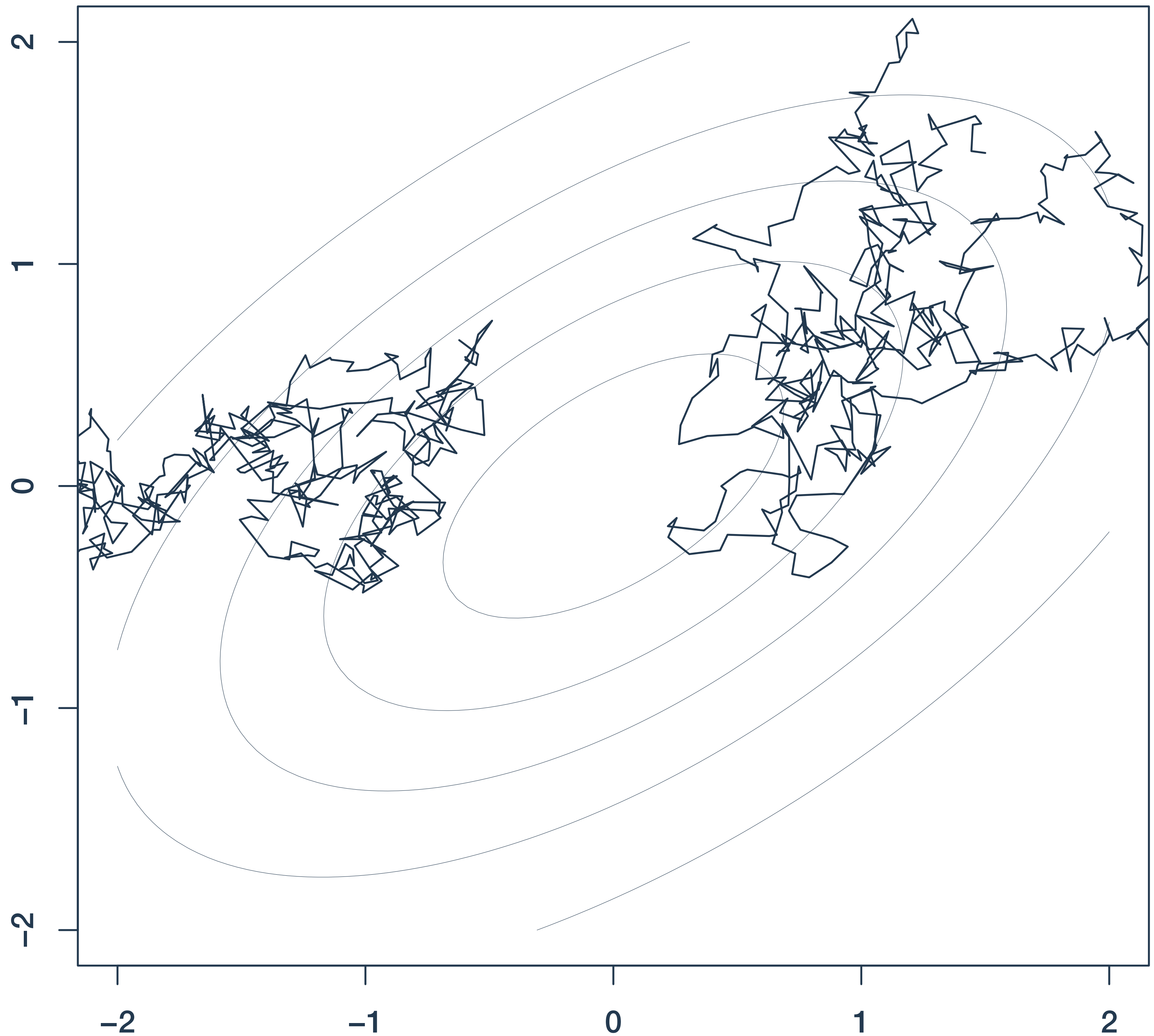
- Because each posterior sample depends on the previous sample, HMC usually displays some *autocorrelation*
- A sample of 1,000 autocorrelated samples will have less information than a sample of 1,000 independent samples
- Relevant quantity: *effective sample size (ESS)*



What can go wrong?

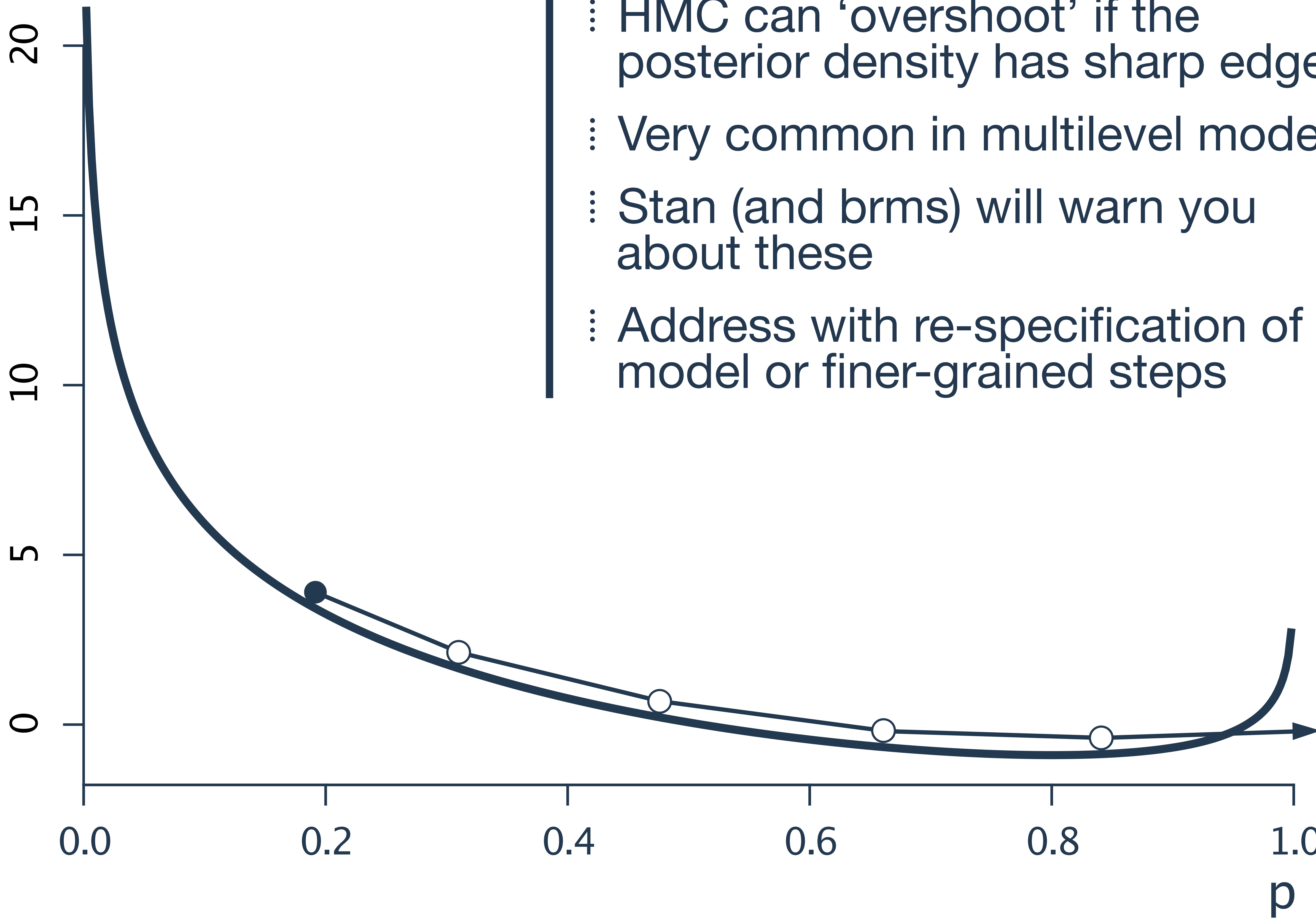
Non-convergence

- Sampling may have trouble ‘converging’ (representing the posterior)
- Many possible causes
  Bad model specification
  Insufficient iterations
  Badly tuned sampler
- Diagnose with $\hat{R}$ (“Rhat”) on multiple chains to check agreement



What can go wrong?

Divergent transitions



- HMC can ‘overshoot’ if the posterior density has sharp edges
- Very common in multilevel models
- Stan (and brms) will warn you about these
- Address with re-specification of model or finer-grained steps