Agenda

1. **Sidebar on Bayesian estimation: MAP vs. MCMC (vs. HMC)**
2. **Random slopes**
3. **Multivariate normal distribution**
4. **Jointly distributed random effects**

Approximating the posterior

Recall simple binomial model

5 trials
4 'successes'

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

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

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'

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

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

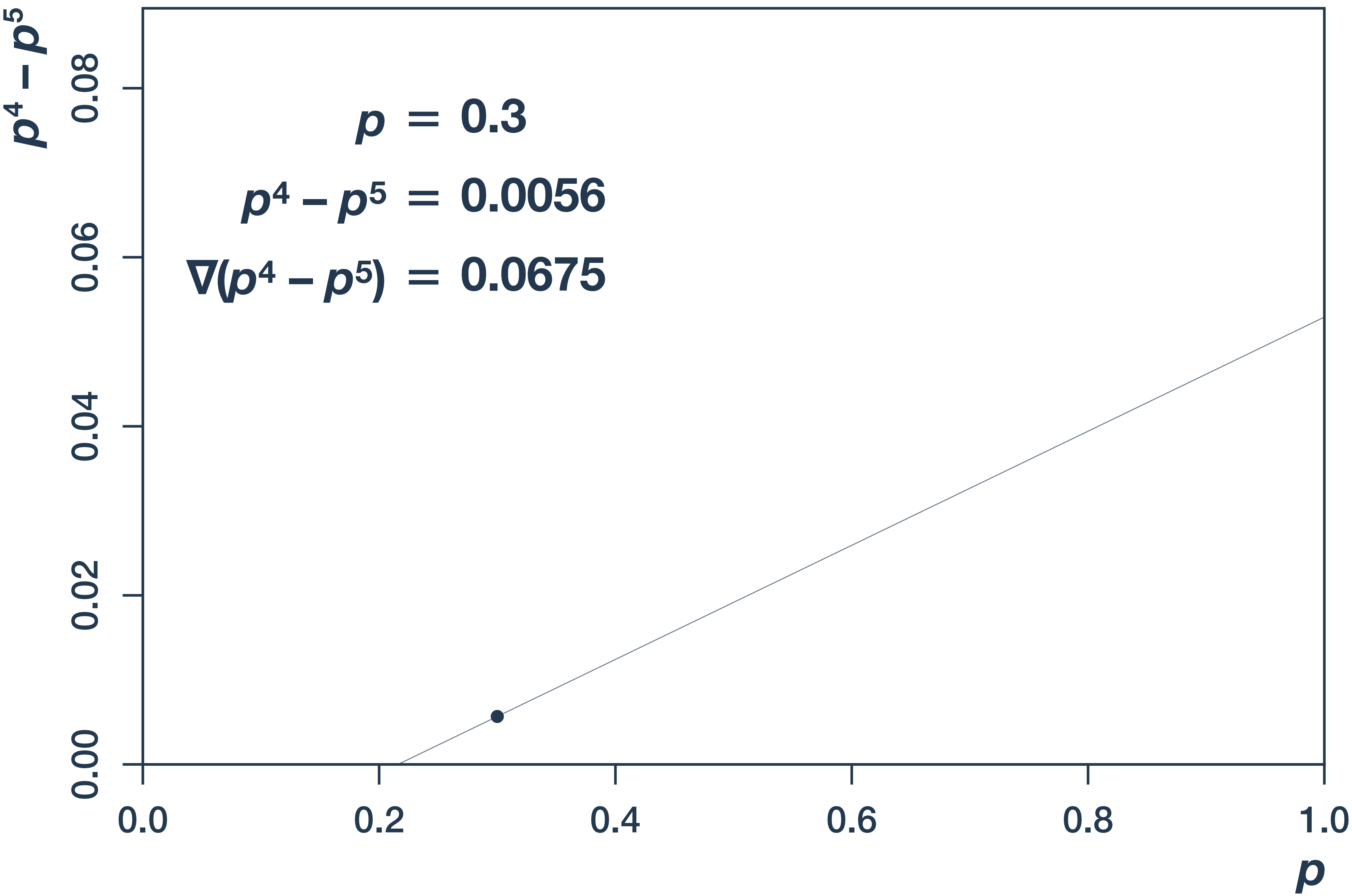
Bayes' Rule

$$\begin{aligned}
\Pr(p|n = 5, k = 4) &= \frac{\Pr(k = 4|n = 5, p)\Pr(p)}{\Pr(k = 4)} \\
&\propto \Pr(k = 4|n = 5, p)\Pr(p) \\
&= \binom{5}{4} p^4 (1 - p)^1 \times 1 \\
&\propto \boxed{p^4 - p^5}
\end{aligned}$$

The posterior distribution for p is proportional to this

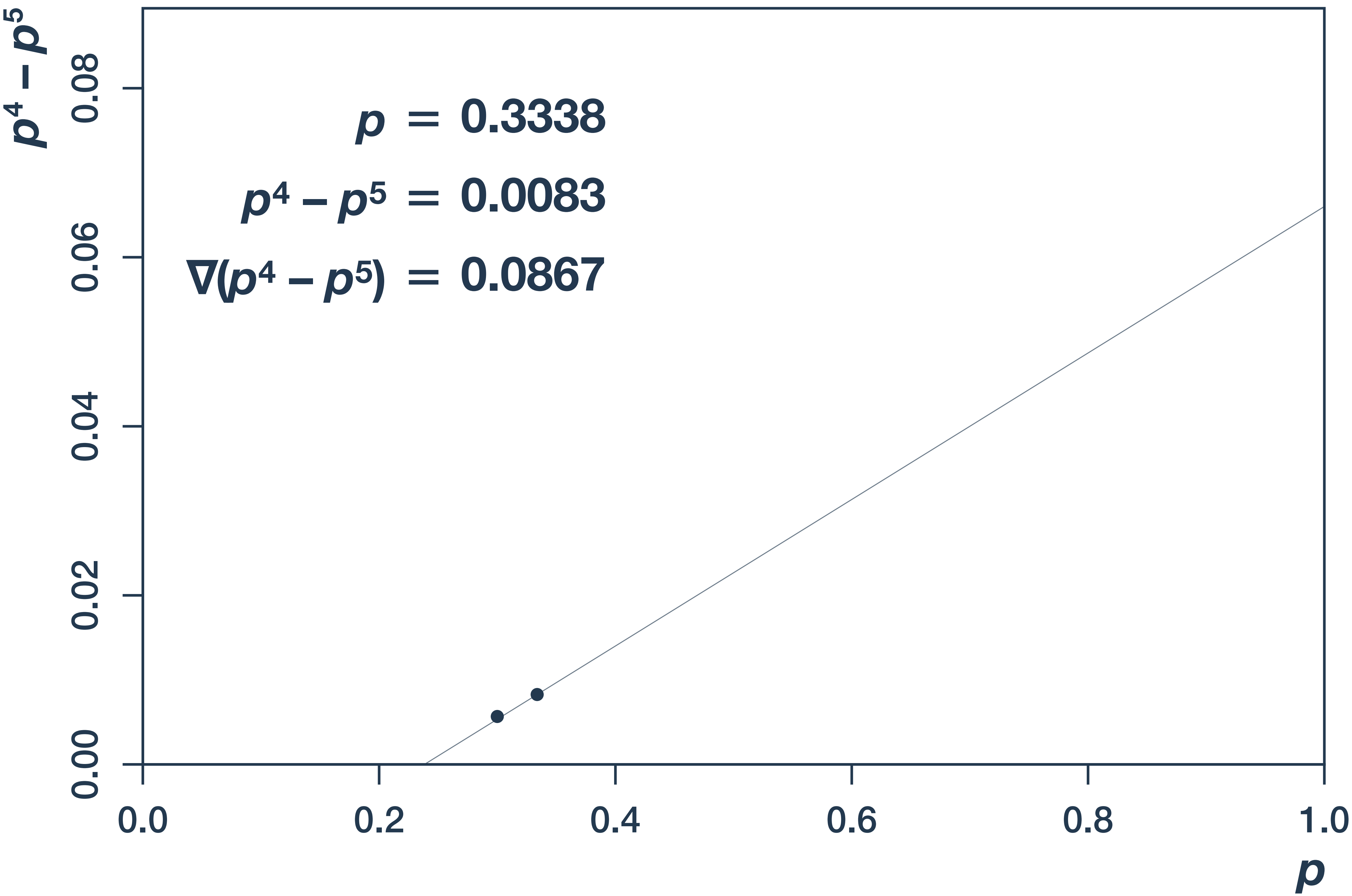
Maximum *a posteriori*

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



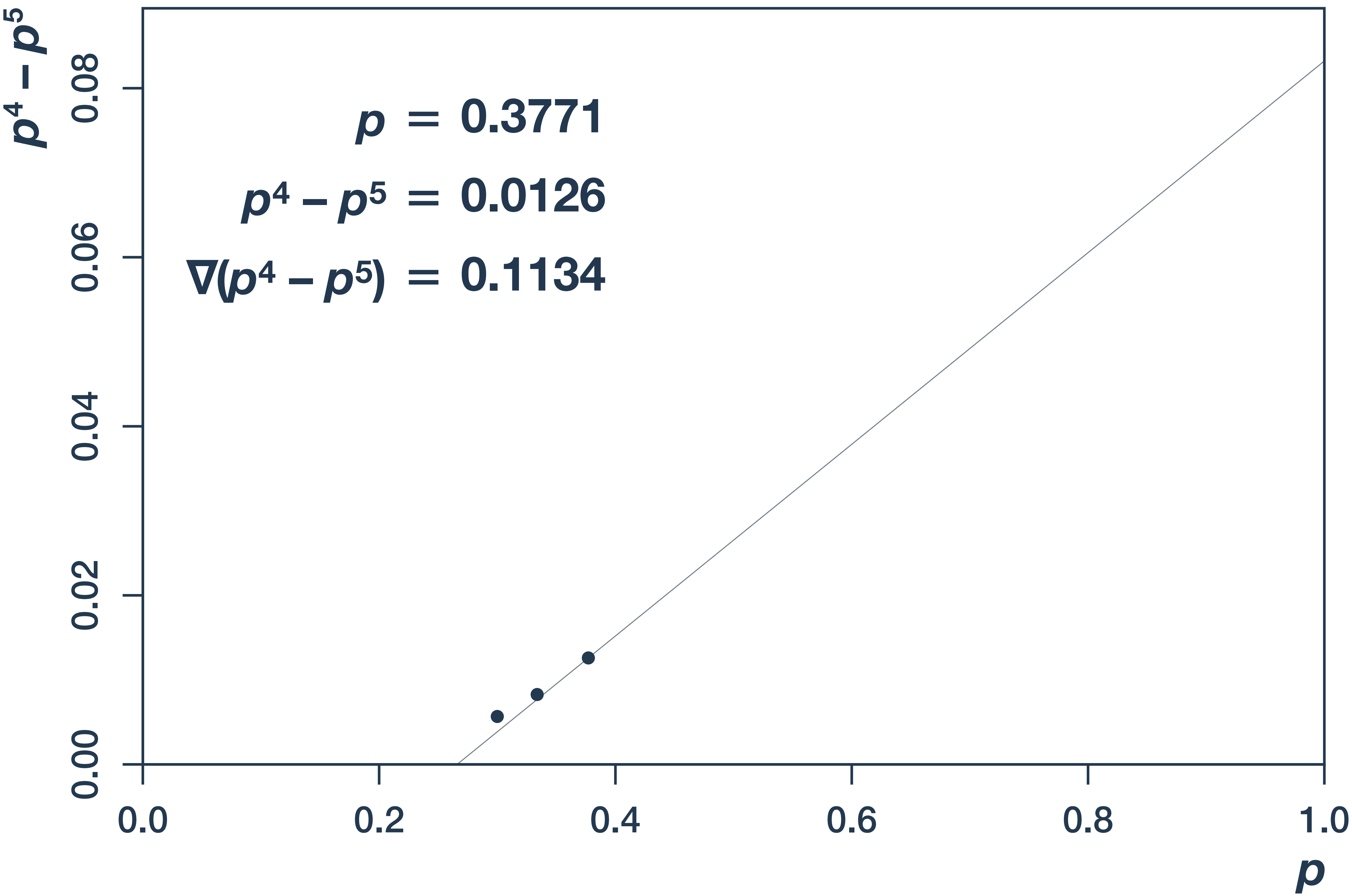
Maximum *a posteriori*

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



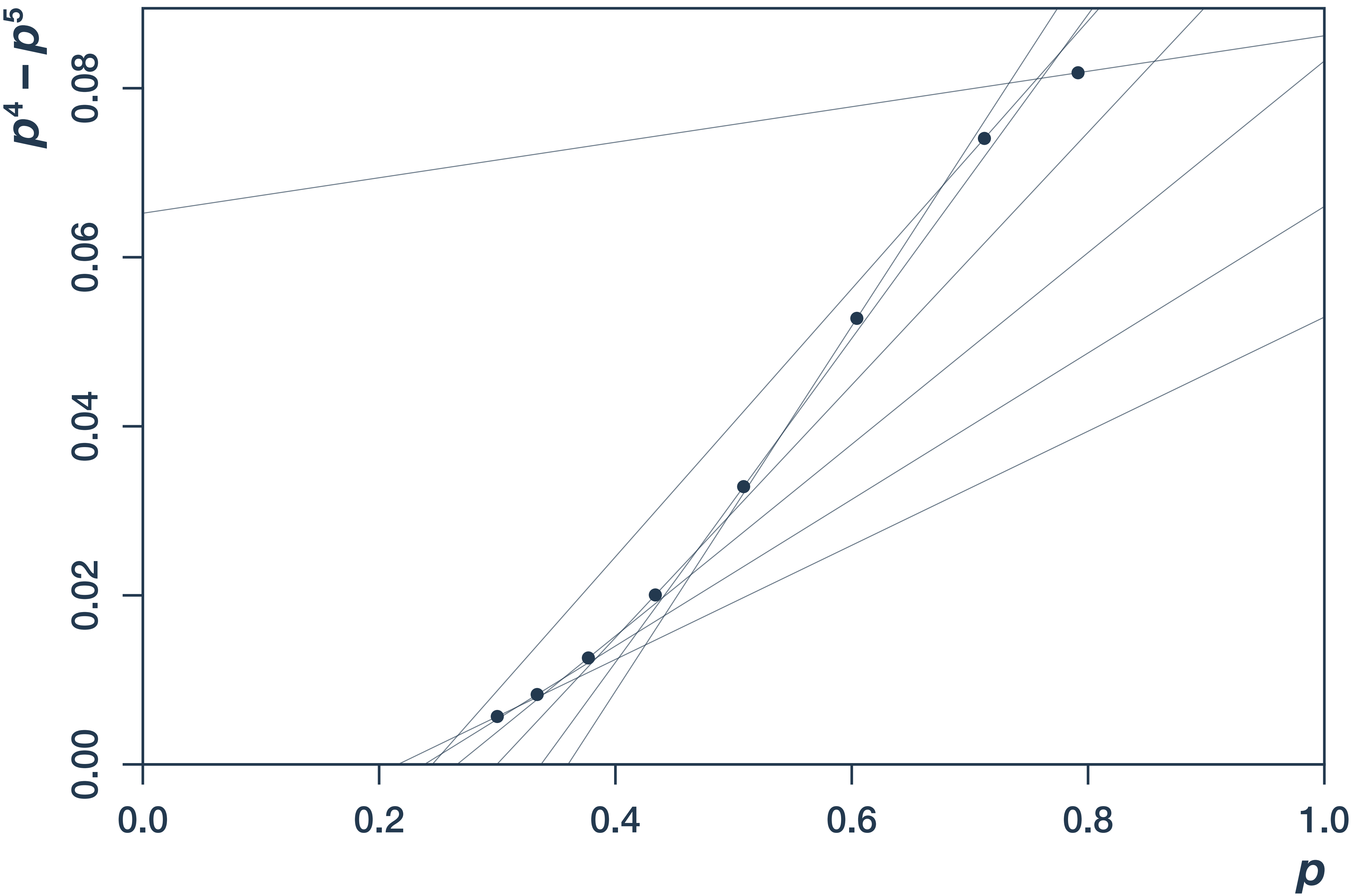
Maximum *a posteriori*

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



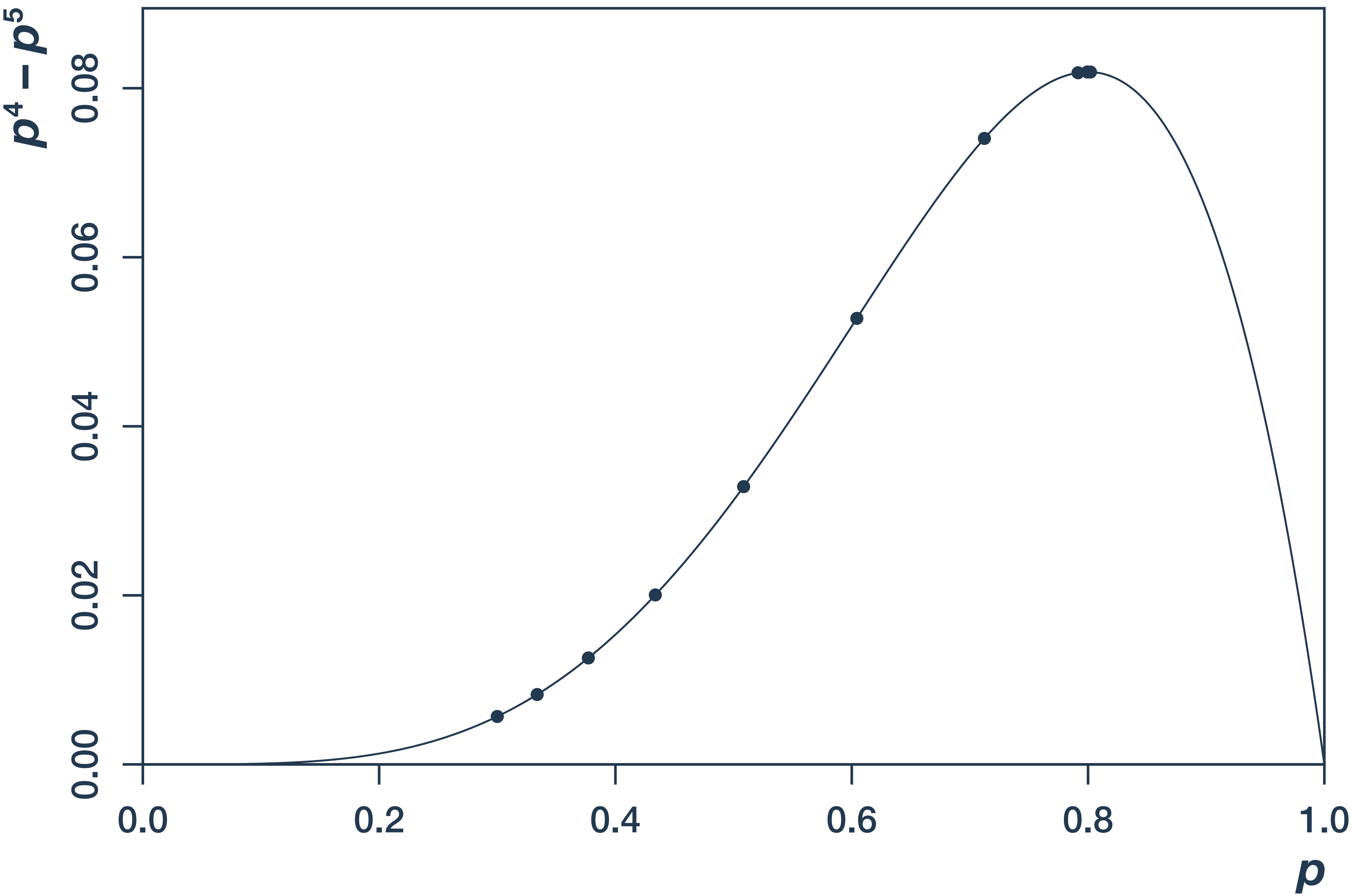
Maximum *a posteriori*

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

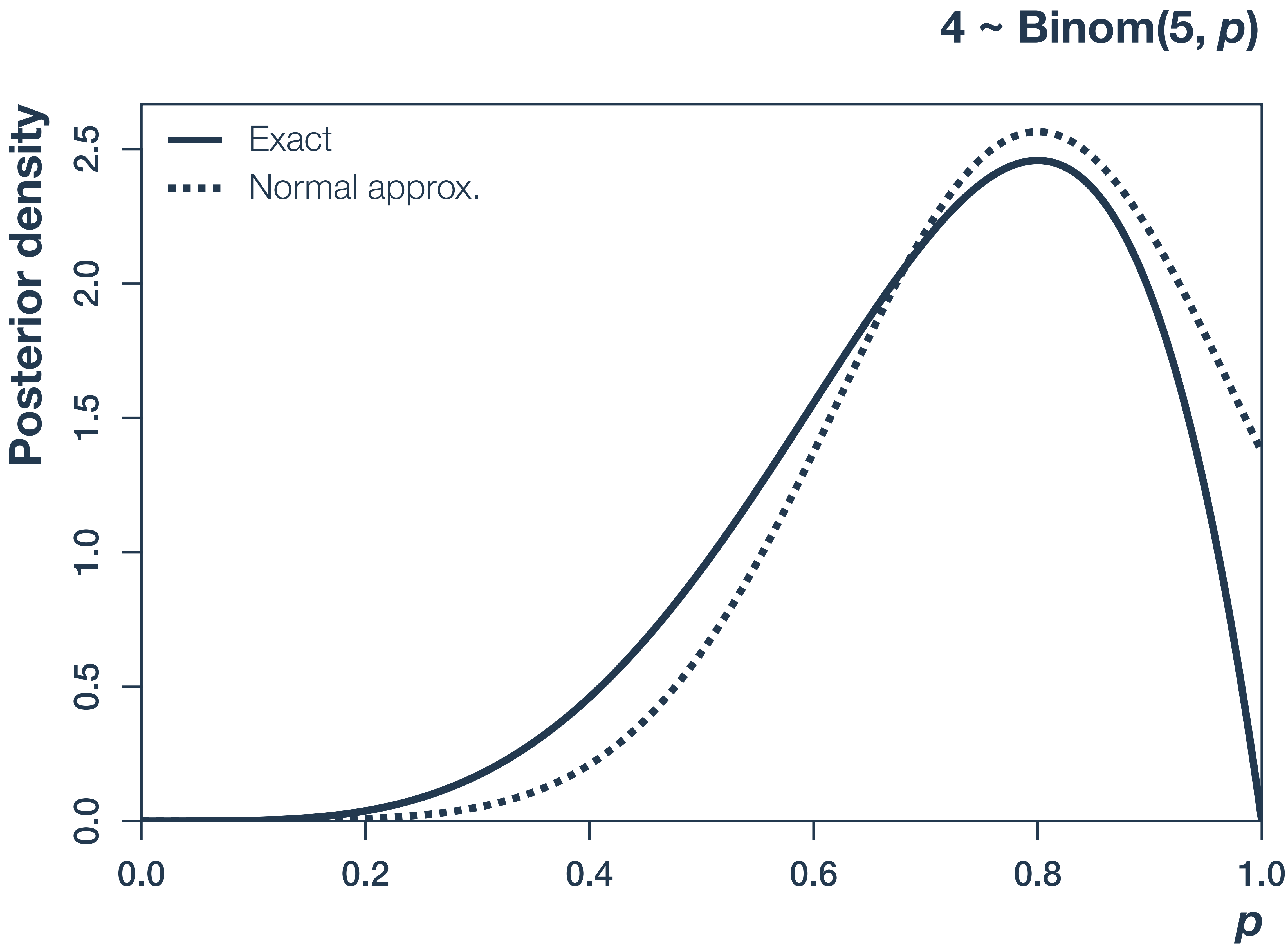


Maximum *a posteriori*

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

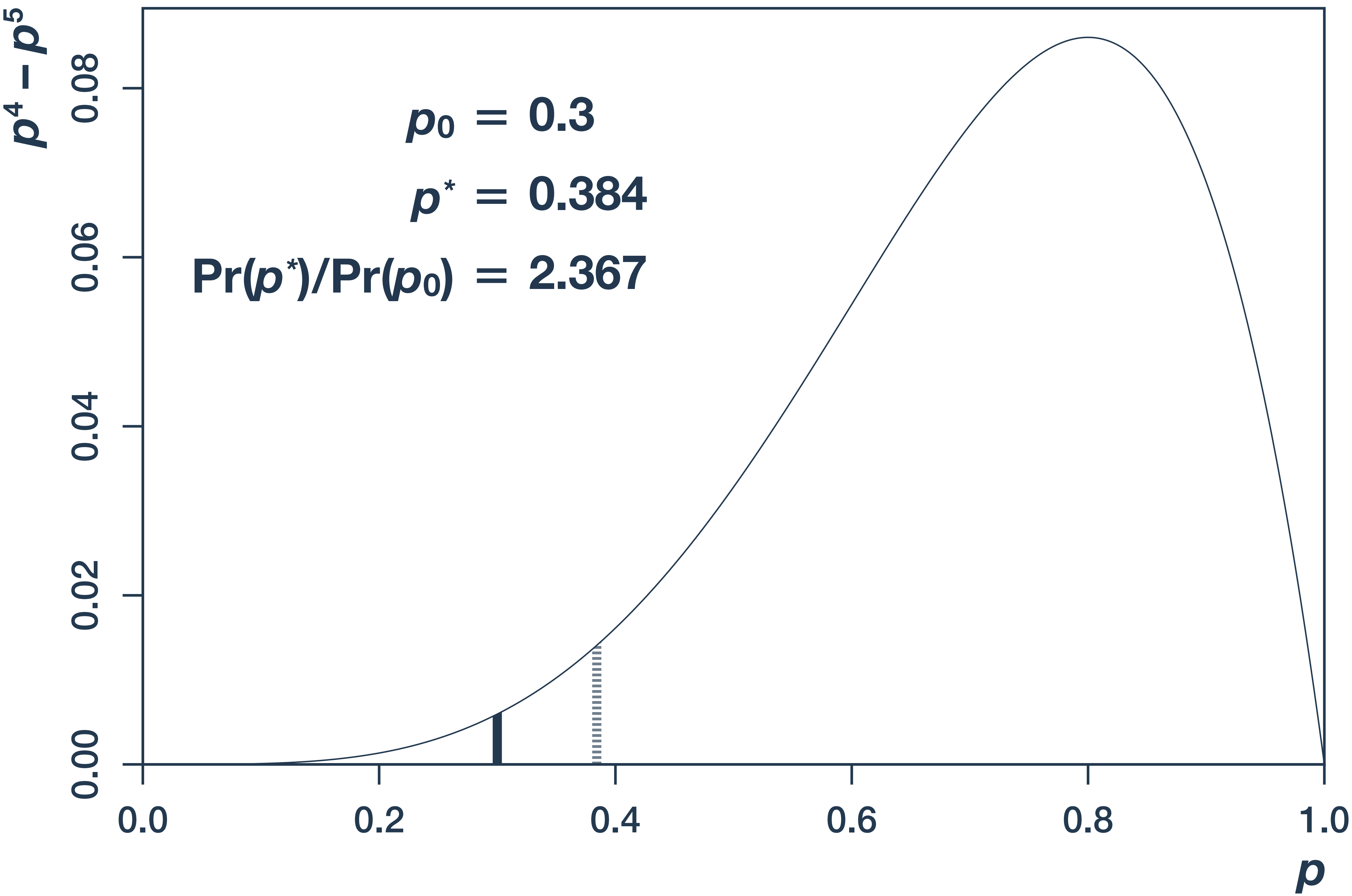


Normal approximation



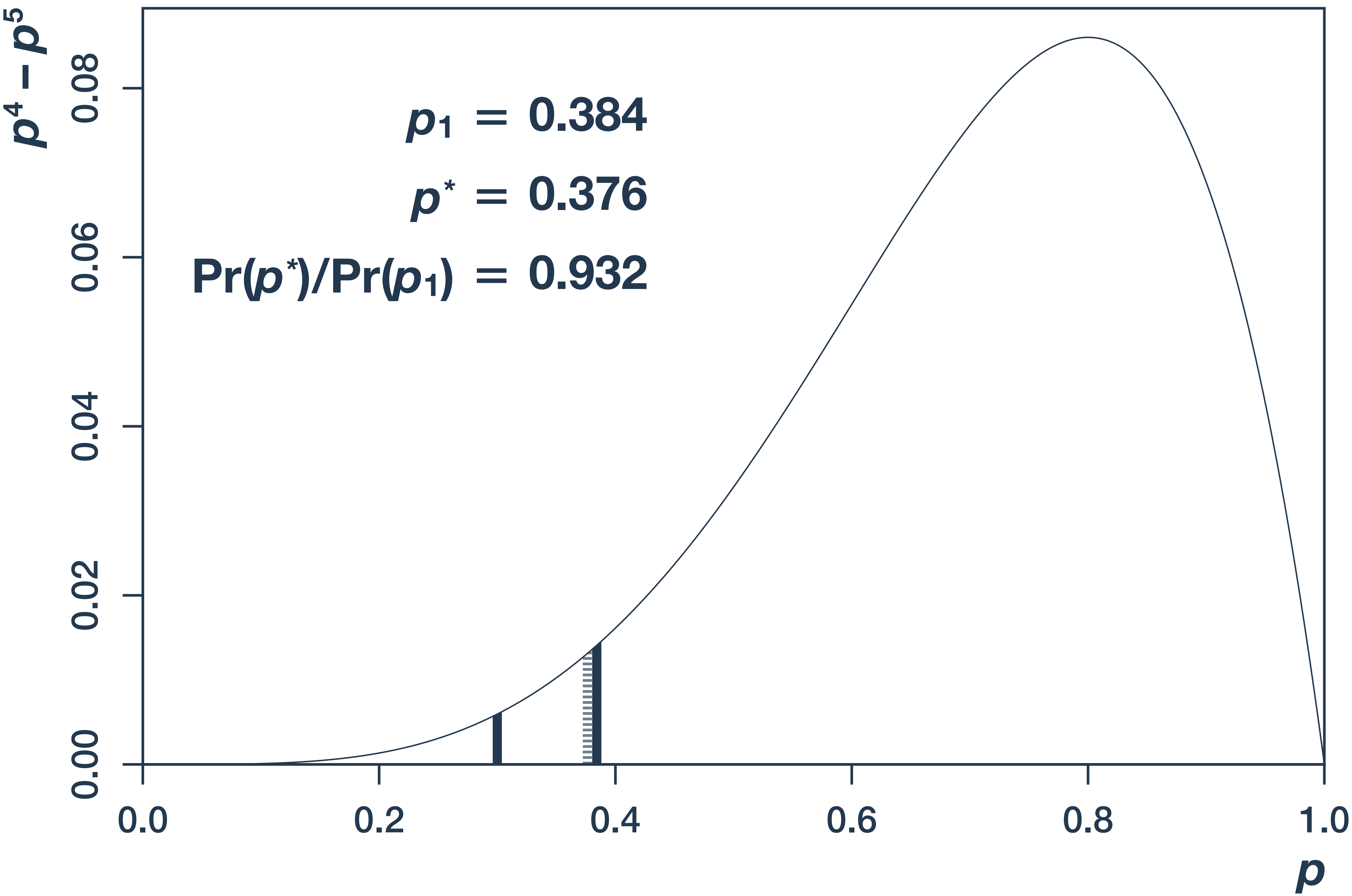
Markov chain Monte Carlo

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



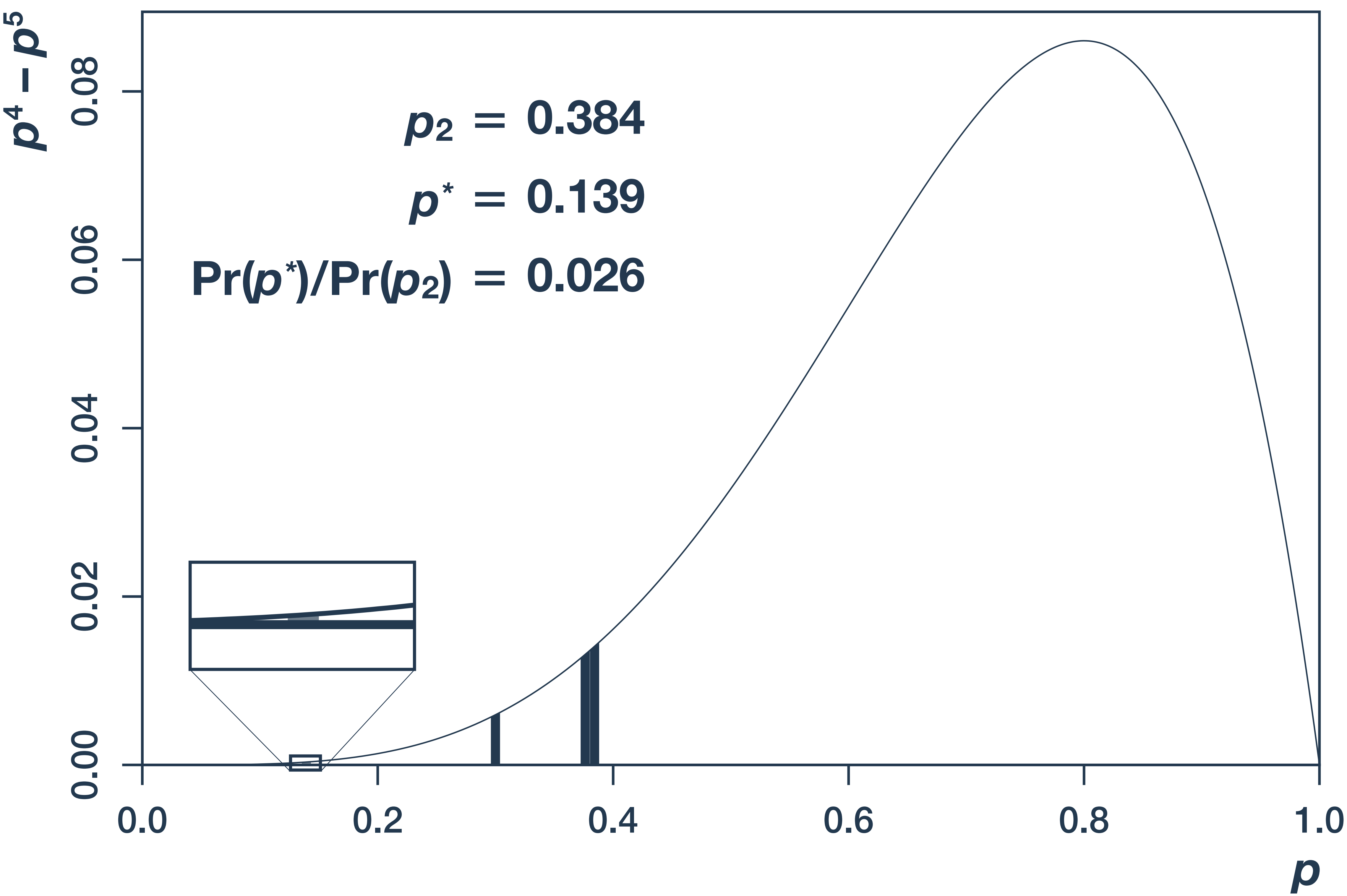
Markov chain Monte Carlo

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



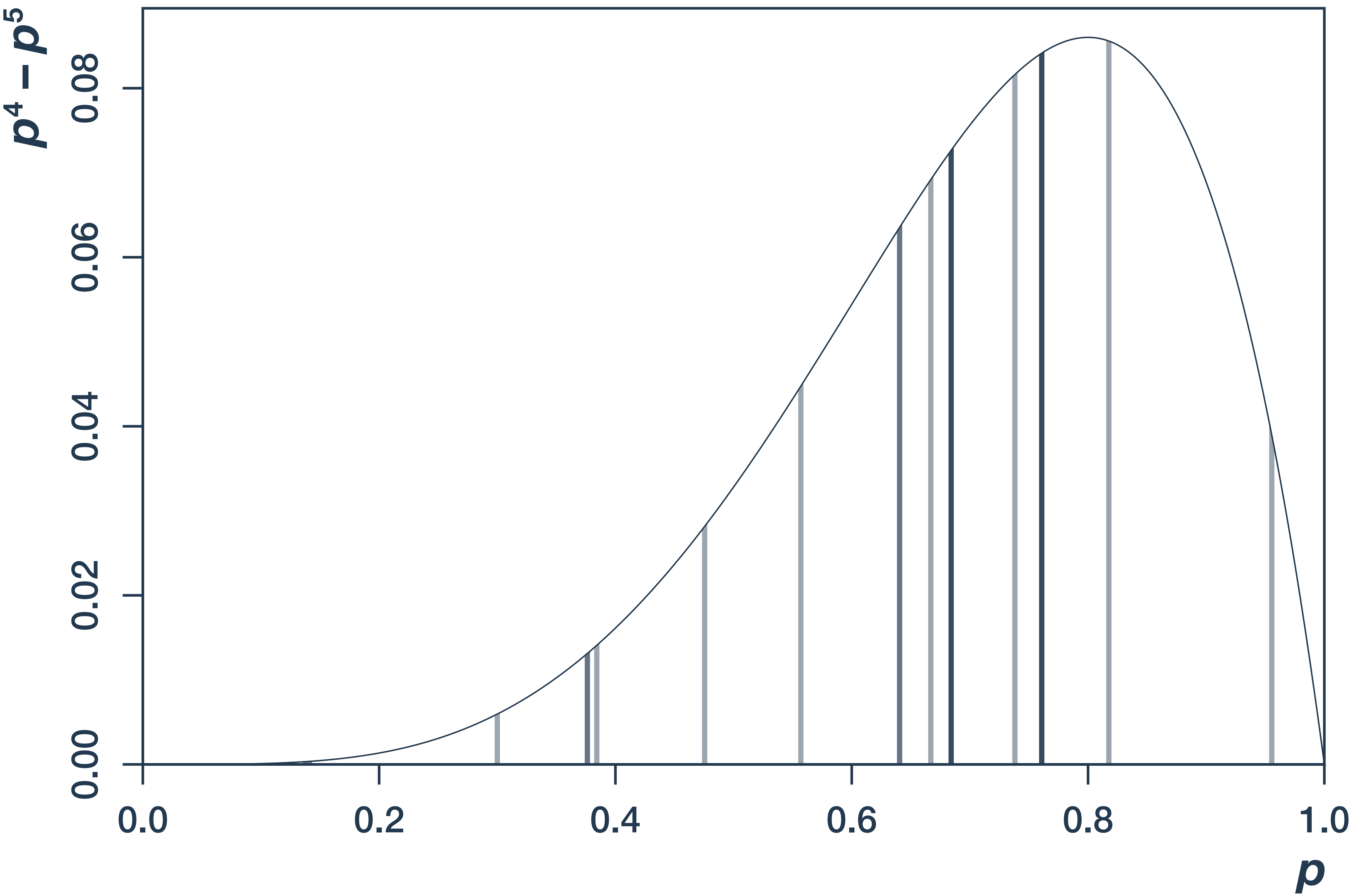
Markov chain Monte Carlo

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

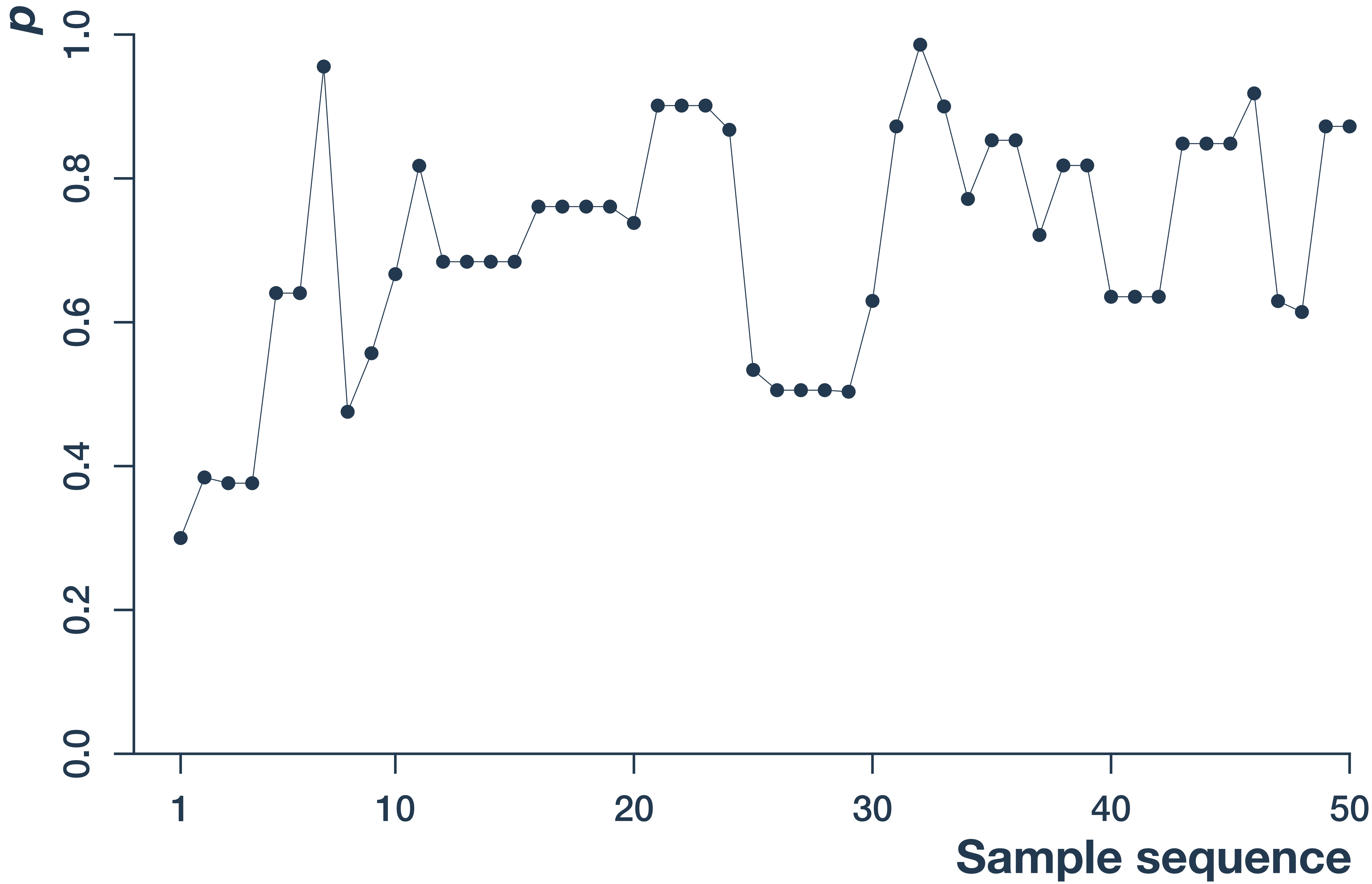


Markov chain Monte Carlo

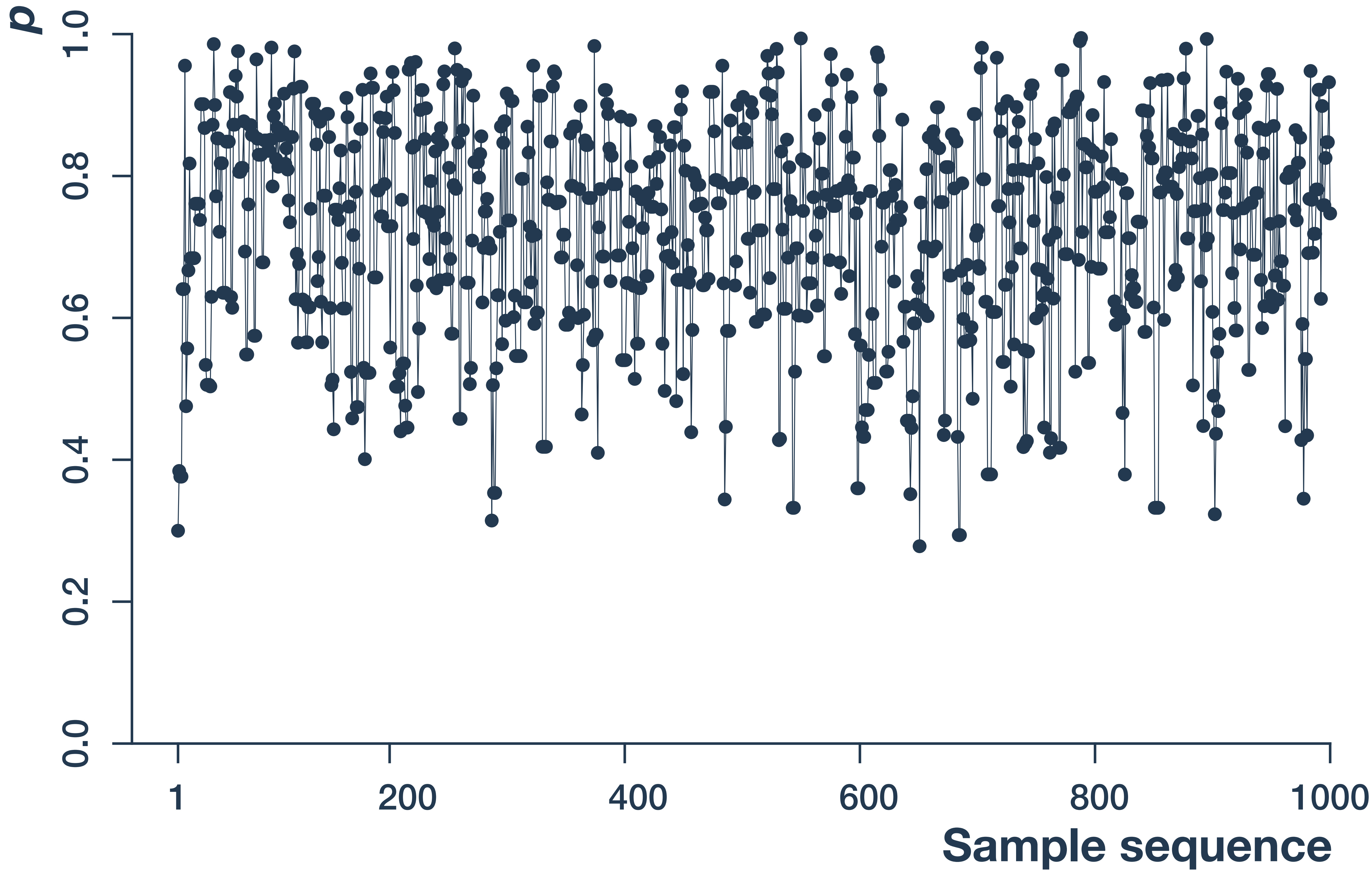
$$\Pr(p \mid data) \propto p^4 - p^5$$



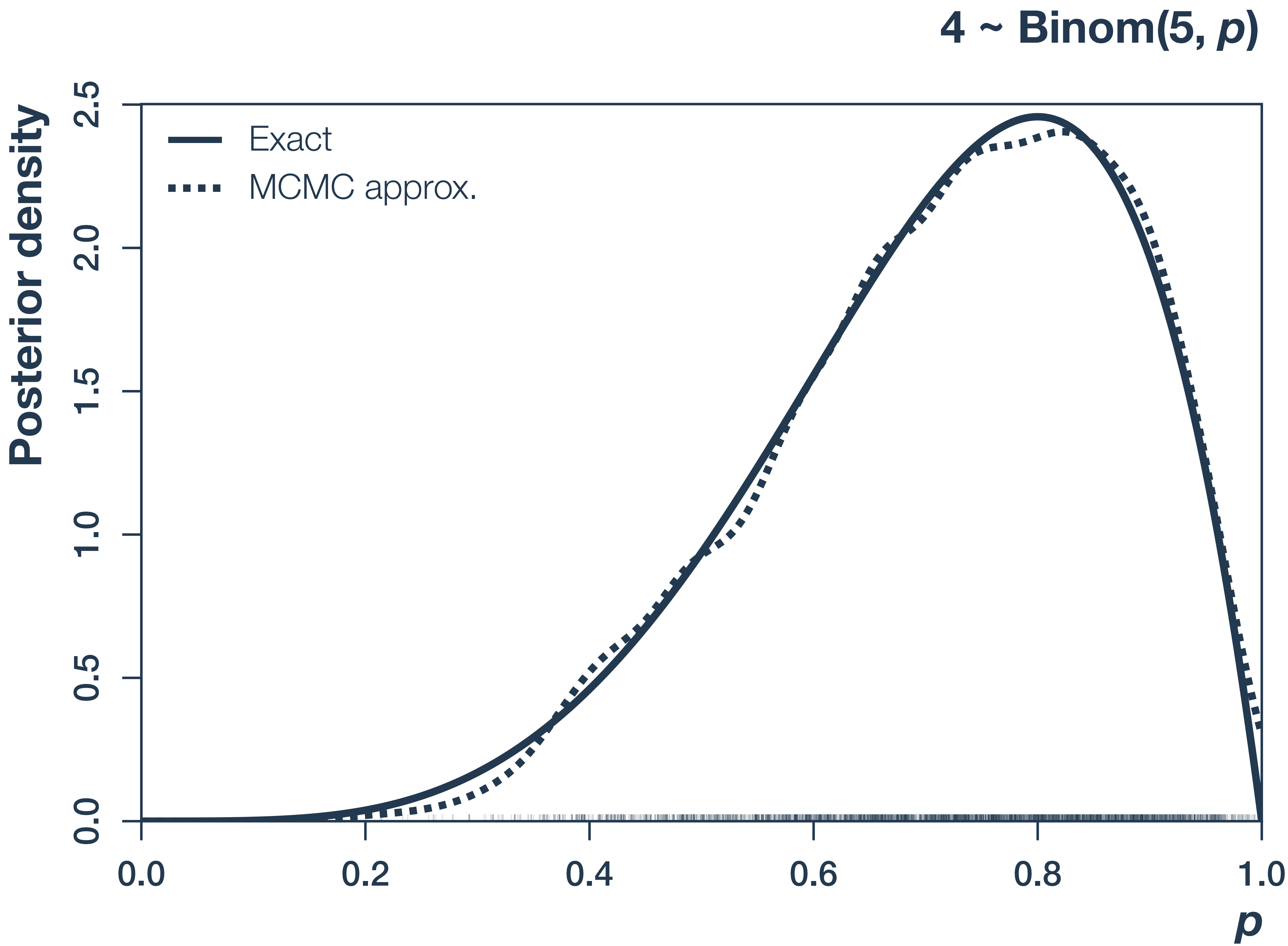
Markov chain Monte Carlo



Markov chain Monte Carlo



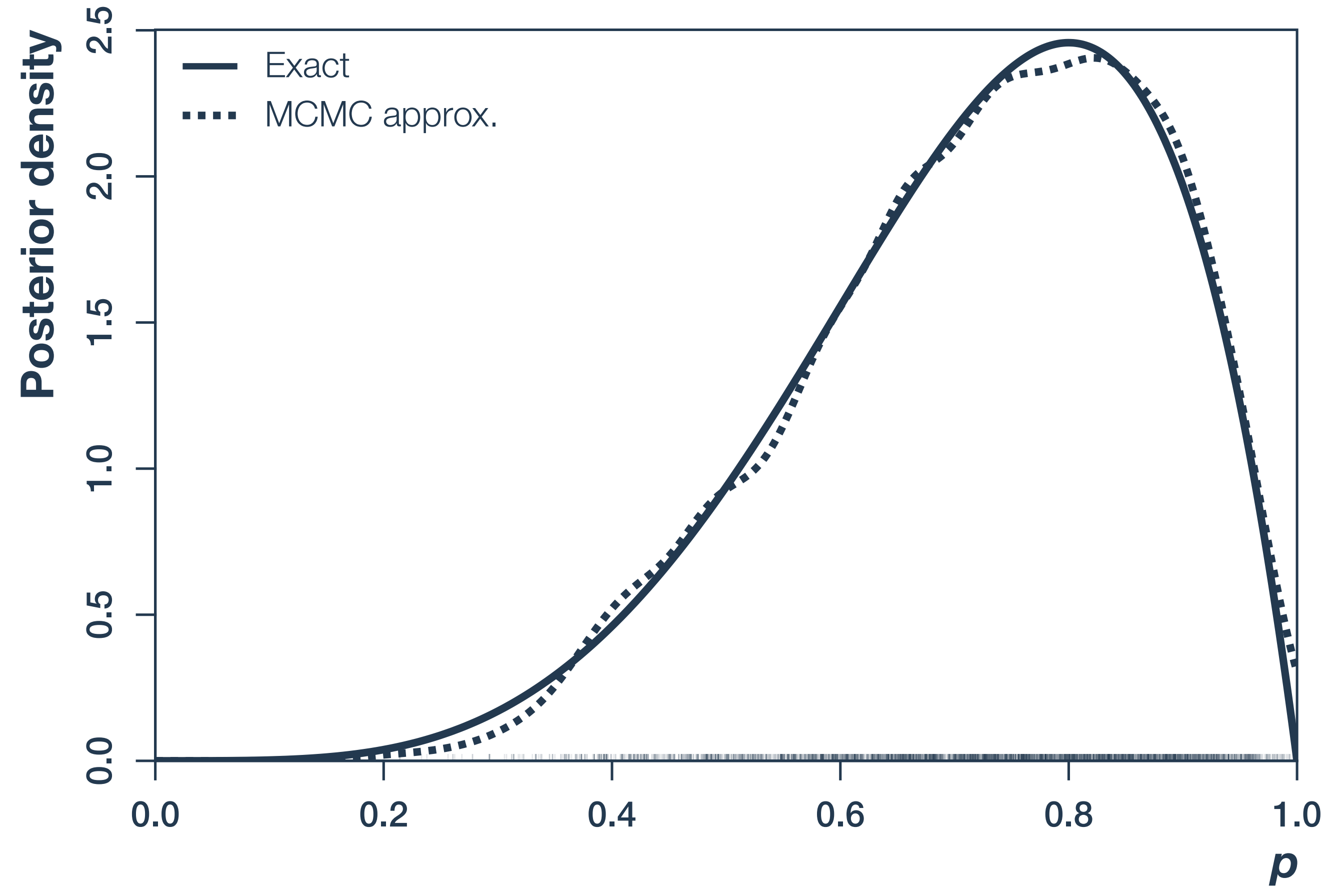
MCMC approximation



MCMC vs. MAP

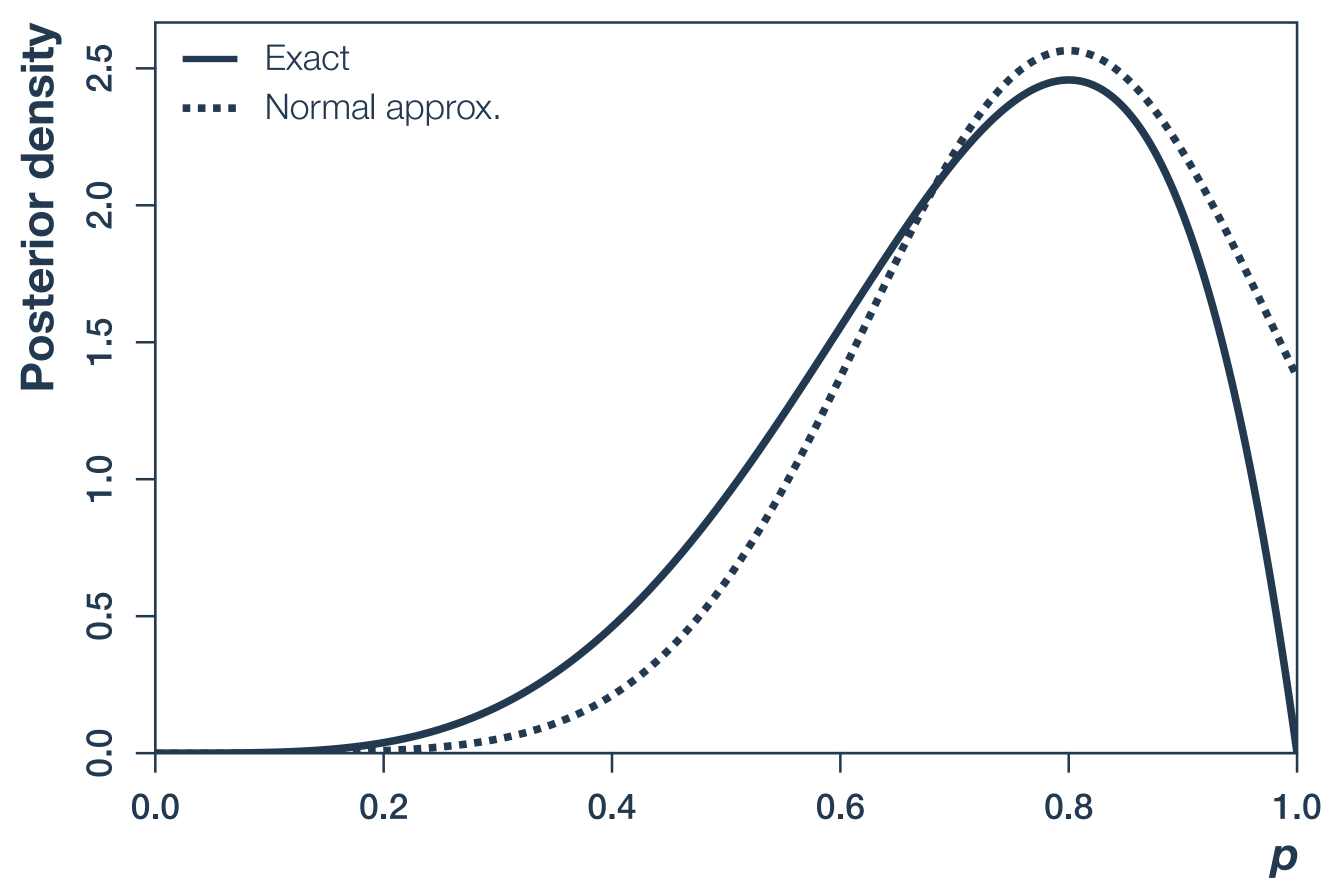
Markov chain Monte Carlo

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

Simulate a physical system

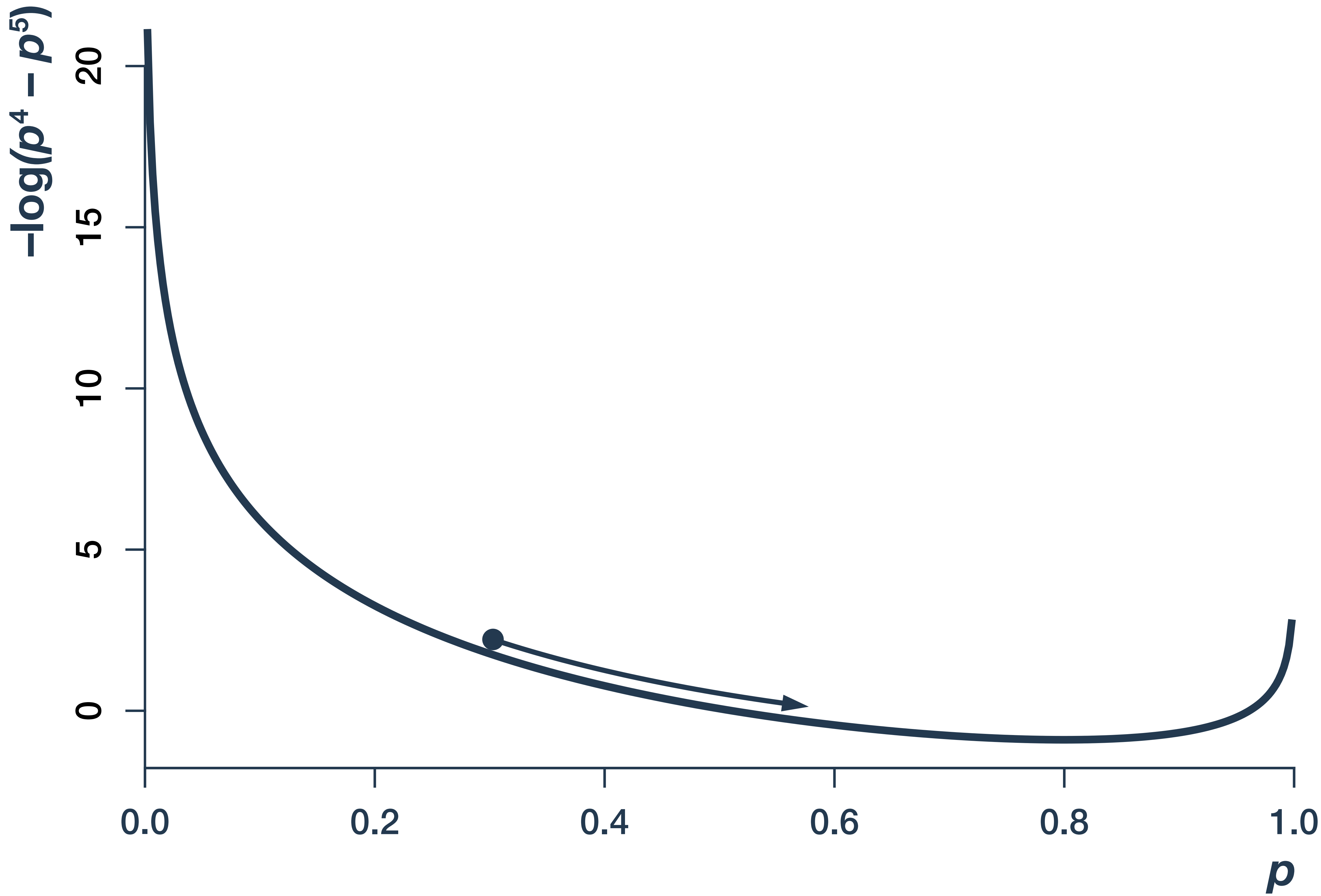
Energy at any point in the parameter space is proportional to the negative log likelihood of the posterior.

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. 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.

Random slopes

Independent random coefficients

$$S_{ik} \sim \text{Norm}(\mu_{ik}, \sigma)$$

$$\mu_{ik} = \beta_{0k} + \beta_{1k} Age_i$$

$$\beta_{0k} = \gamma_{00} + \eta_{0k}$$

$$\beta_{1k} = \gamma_{10} + \eta_{1k}$$

$$\boxed{\eta_{0k} \sim \text{Norm}(0, \phi_0)}$$

$$\boxed{\eta_{1k} \sim \text{Norm}(0, \phi_1)}$$

η_{0k} and η_{1k} are *independent*: knowing one tells us nothing about the other.

(Fixed priors omitted)

Independence of random effects is rarely a realistic assumption.

Multivariate normal distribution

Joint distribution over y_0 and y_1

Variables may not be independent

$$\begin{bmatrix} y_0 \\ y_1 \end{bmatrix} \sim \text{MVNorm}\left(\boxed{\begin{bmatrix} \mu_0 \\ \mu_1 \end{bmatrix}}, \boxed{\begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix}} \right)$$

μ_0 and μ_1 are mean of y_0 and y_1, respectively.

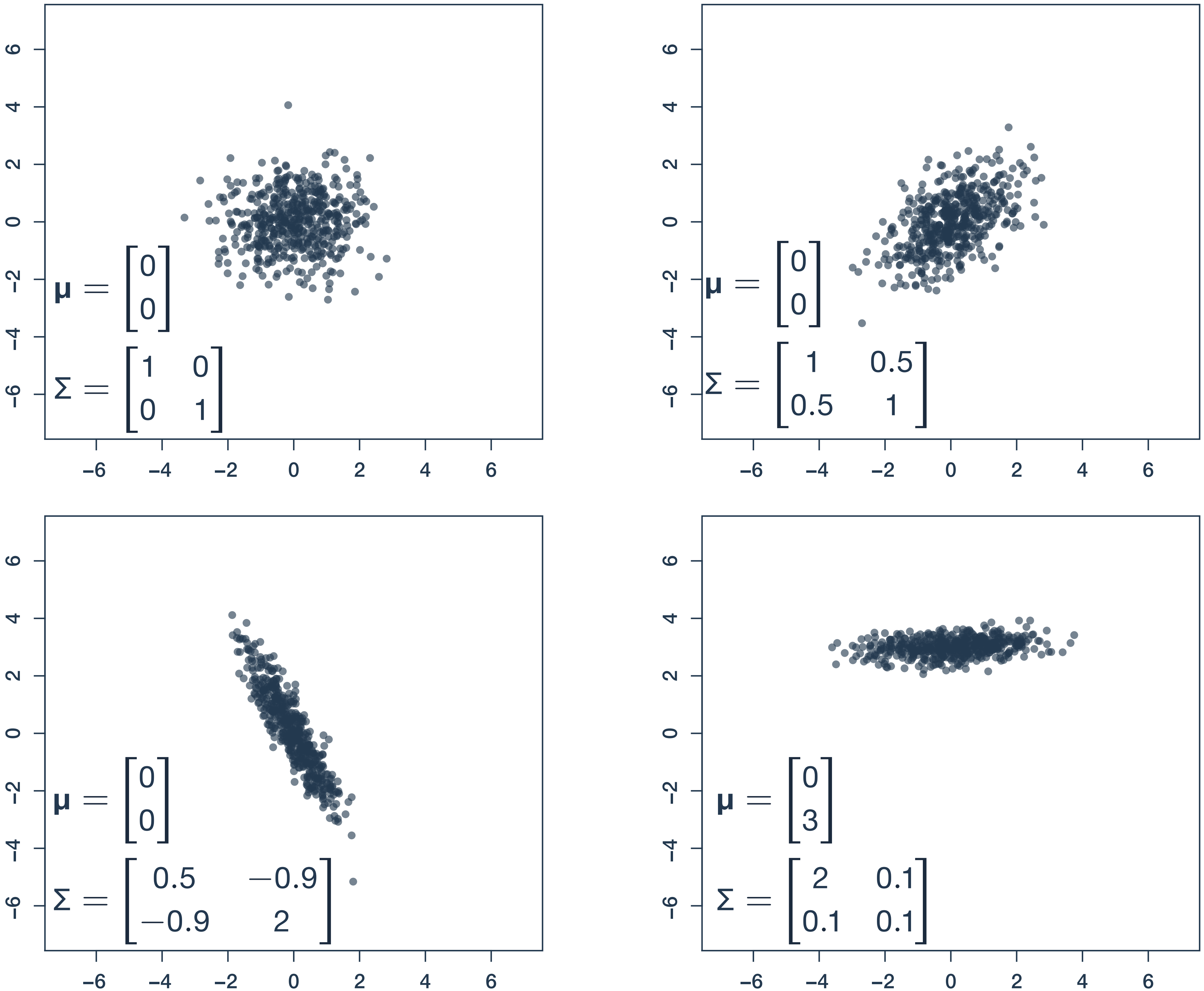
The *covariance matrix* describes the way that y_0 and y_1 inform another.

Covariance matrix

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{12} & \sigma_2^2 \end{bmatrix} = \begin{bmatrix} \sigma_1^2 & \boxed{\sigma_1 \sigma_2 \rho_{12}} \\ \boxed{\sigma_1 \sigma_2 \rho_{12}} & \sigma_2^2 \end{bmatrix}$$

Off-diagonal elements depend on the correlation between y_0 and y_1 (ρ_{12}).

Multivariate normal distribution



Multivariate normal distribution

$$\begin{bmatrix} y_0 \\ y_1 \\ \vdots \\ y_k \end{bmatrix} \sim \text{MVNorm}\left(\begin{bmatrix} \mu_0 \\ \mu_1 \\ \vdots \\ \mu_k \end{bmatrix}, \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1k} \\ \sigma_{12} & \sigma_2^2 & \cdots & \sigma_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1k} & \sigma_{2k} & \cdots & \sigma_k^2 \end{bmatrix} \right)$$

Jointly distributed random effects

Independent	Joint
$S_{ik} \sim \text{Norm}(\mu_{ik}, \sigma)$	$S_{ik} \sim \text{Norm}(\mu_{ik}, \sigma)$
$\mu_{ik} = \beta_{0k} + \beta_{1k} Age_i$	$\mu_{ik} = \beta_{0k} + \beta_{1k} Age_i$
$\beta_{0k} = \gamma_{00} + \eta_{0k}$	$\beta_{0k} = \gamma_{00} + \eta_{0k}$
$\beta_{1k} = \gamma_{10} + \eta_{1k}$	$\beta_{1k} = \gamma_{10} + \eta_{1k}$
$\eta_{0k} \sim \text{Norm}(0, \phi_0)$ $\eta_{1k} \sim \text{Norm}(0, \phi_1)$	$\begin{bmatrix} \eta_{0k} \\ \eta_{1k} \end{bmatrix} \sim \text{MVNorm}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \Phi \right)$

(Fixed priors omitted)

Estimates

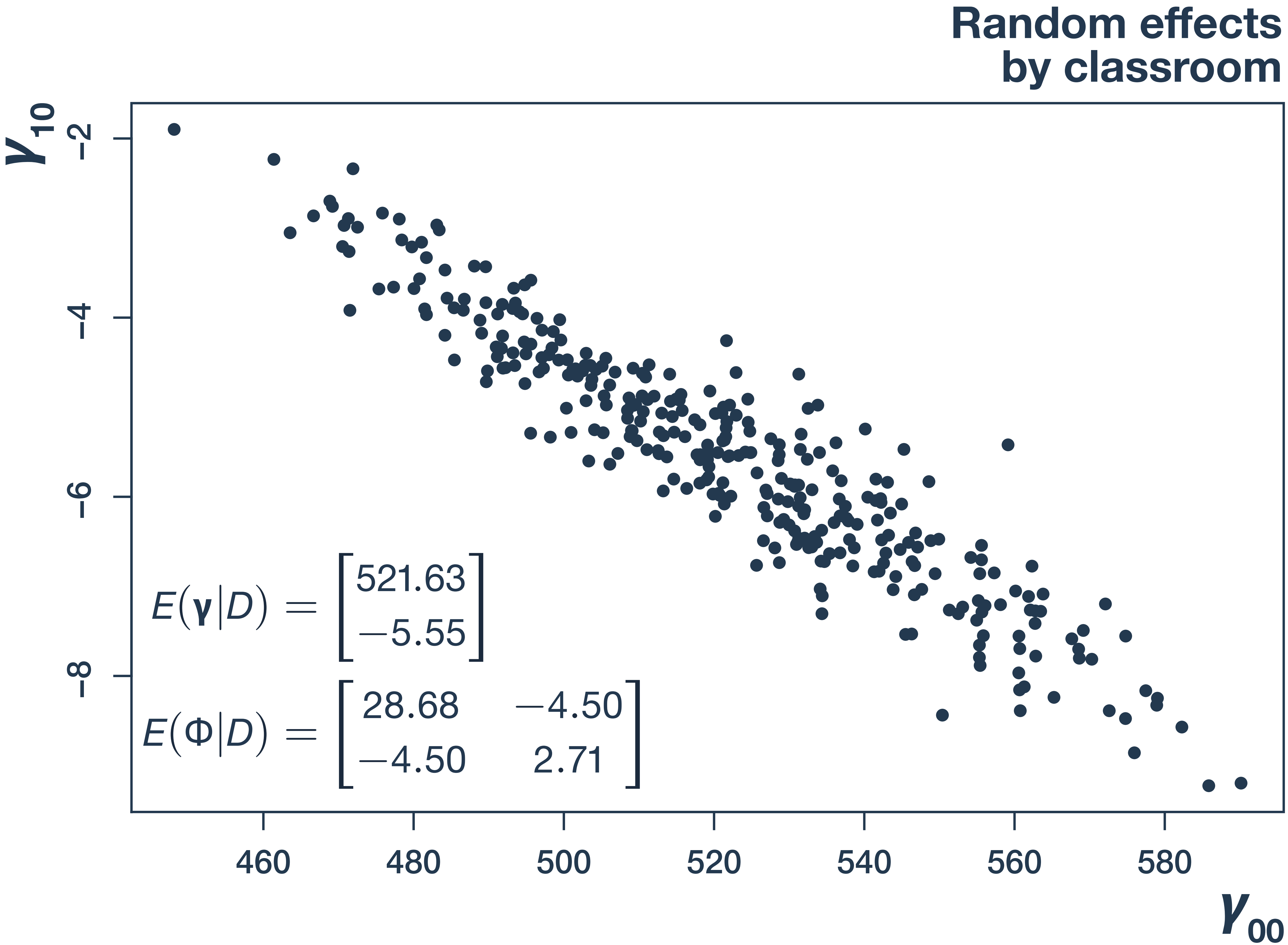
$$S_{ik} \sim \text{Norm}(\mu_{ik}, \sigma)$$
$$\mu_{ik} = \beta_{0k} + \beta_{1k} Age_i$$

$$\beta_{0k} = \gamma_{00} + \eta_{0k}$$
$$\beta_{1k} = \gamma_{10} + \eta_{1k}$$

$$\begin{bmatrix} \eta_{0k} \\ \eta_{1k} \end{bmatrix} \sim \text{MVNorm}\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \Phi \right)$$

$$E(\gamma_{00} | D) = 521.63$$
$$E(\gamma_{10} | D) = -5.55$$
$$E(\sigma | D) = 46.96$$
$$E(\Phi | D) = \begin{bmatrix} 28.68 & -4.50 \\ -4.50 & 2.71 \end{bmatrix}$$

Class-level estimates



Class-level predictions