Agenda

Monte-carlo methods

1. Administrative
2. MAP/ML
3. MCMC
4. Hamiltonian MC
5. What can go wrong
6. ***Hands on:***
   Convergence in `brms` *and* `lme4`

Worksheet 4

- Forthcoming!
- Hopefully later today

Maximizing probability

Recall: simple binomial model

5 trials; 4 'successes'

$$4 \sim \text{Binom}(5, p)$$
$$p \sim \text{Beta}(1, 1)$$

This is the same as an intercept-only logistic distribution with a logit-transformed uniform for the prior on α

Bayes' rule

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

EXPLORING THE POSTERIOR

Recall: simple binomial model
5 trials; 4 'successes'

$$4 \sim \mathrm{Binom}(5, p)$$
$$p \sim \mathrm{Beta}(1, 1)$$

This is the same as an intercept-only logistic distribution with a logit-transformed uniform for the prior on α

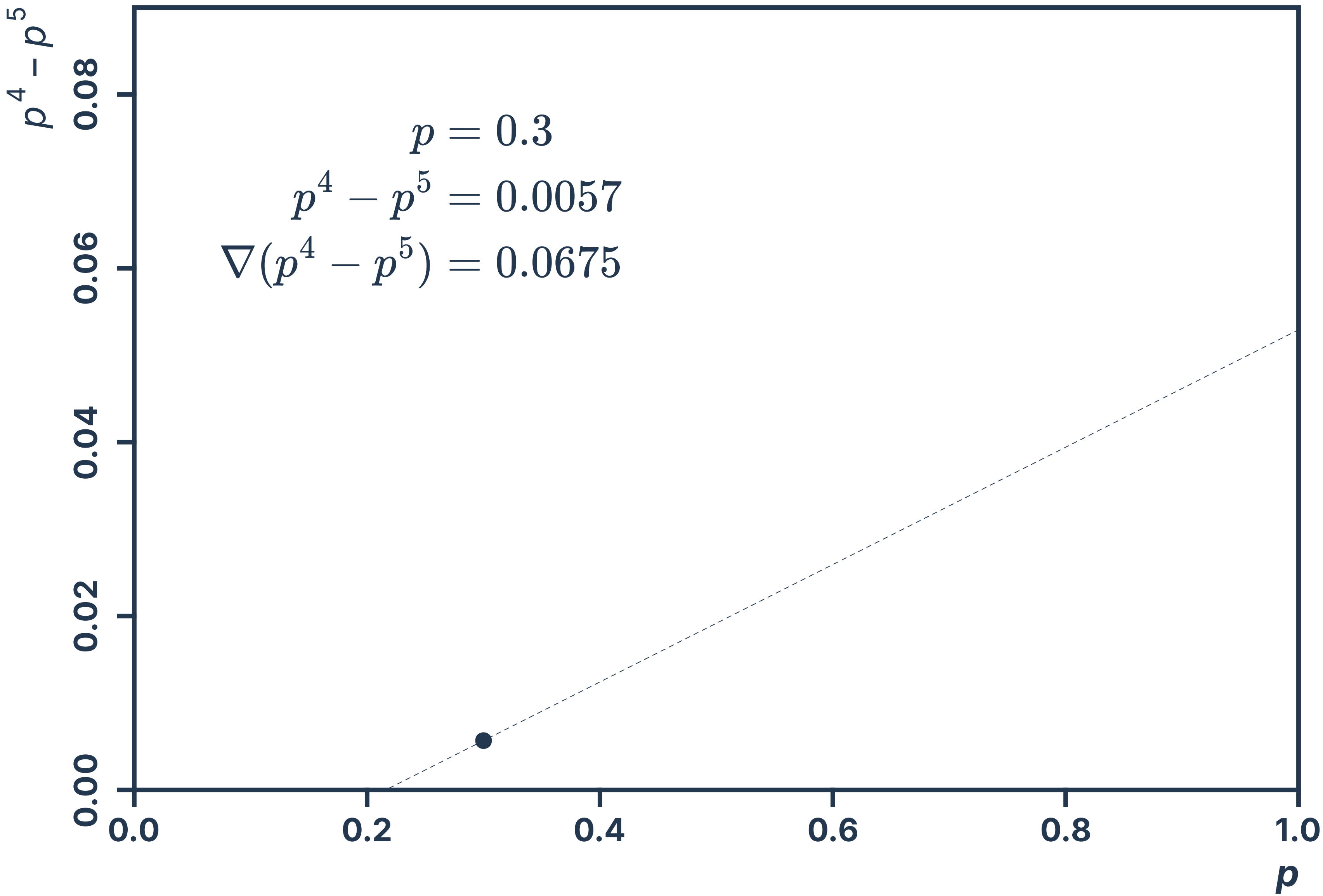
Bayes' rule

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

the posterior distribution of p is proportional to this

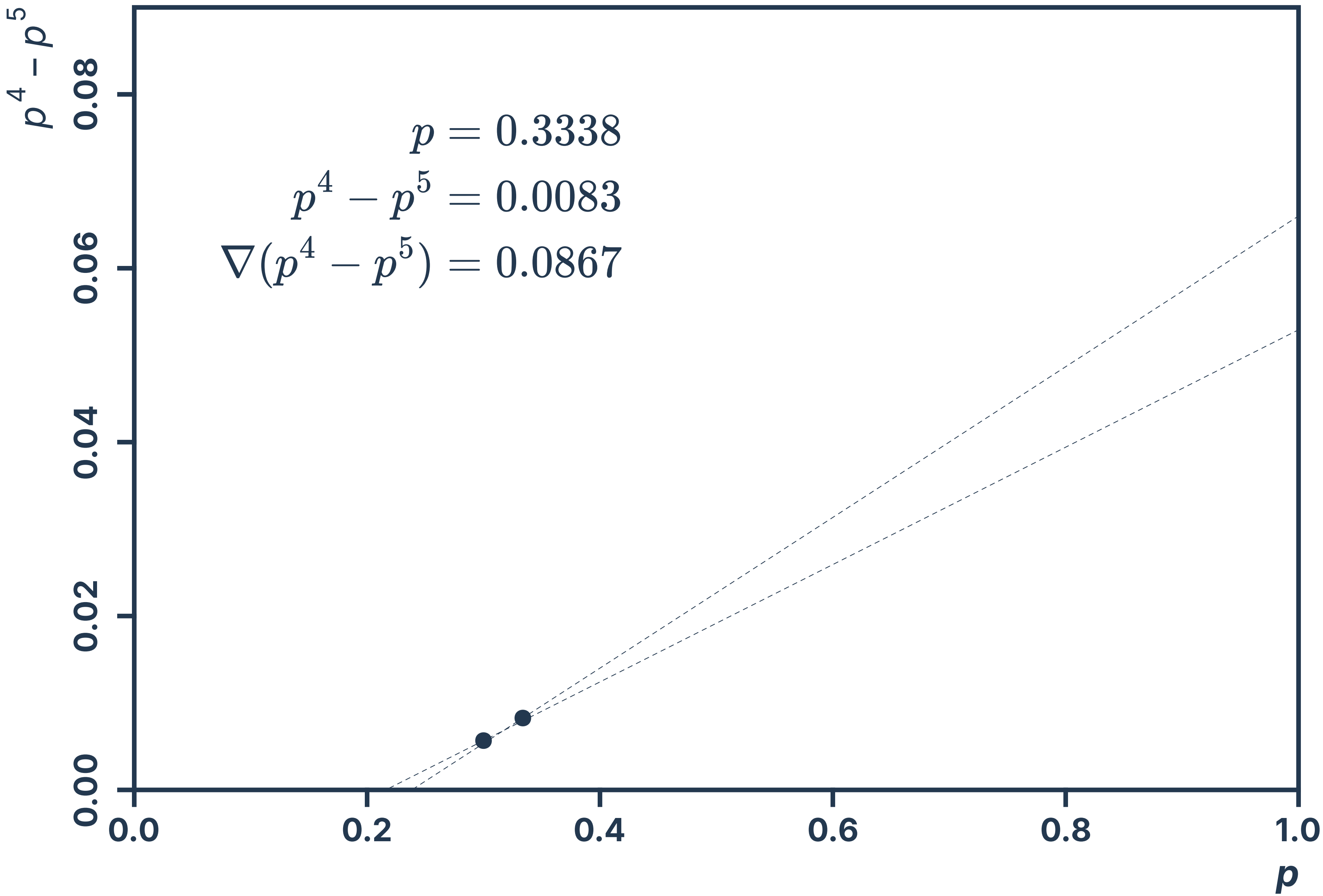
HILL CLIMBING (MAP/ML)

$$\text{Prob}(p|\text{data}, \text{model}) \propto p^4 - p^5$$



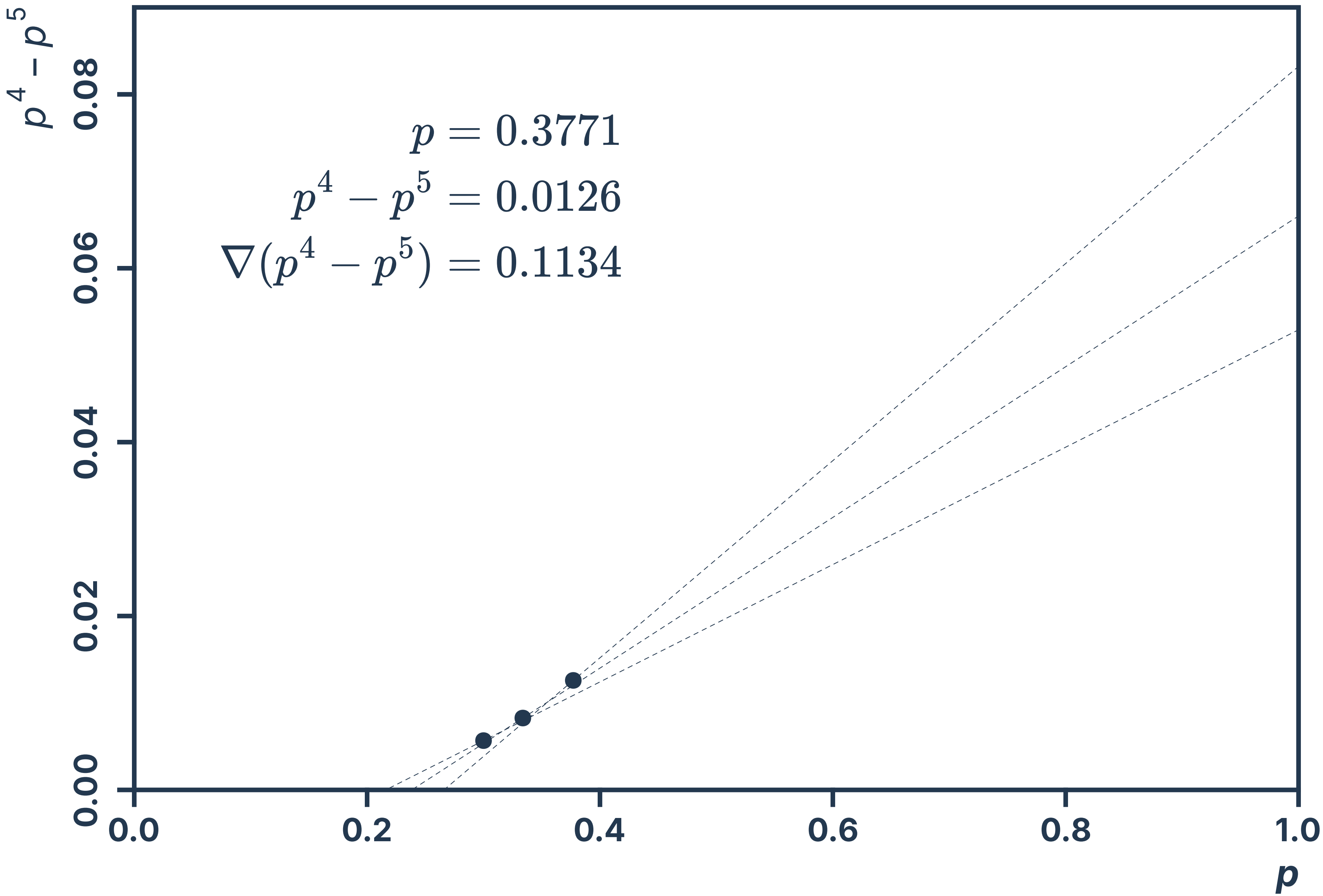
HILL CLIMBING (MAP/ML)

$$\mathrm{Prob}(p|\mathrm{data}, \mathrm{model}) \propto p^4 - p^5$$



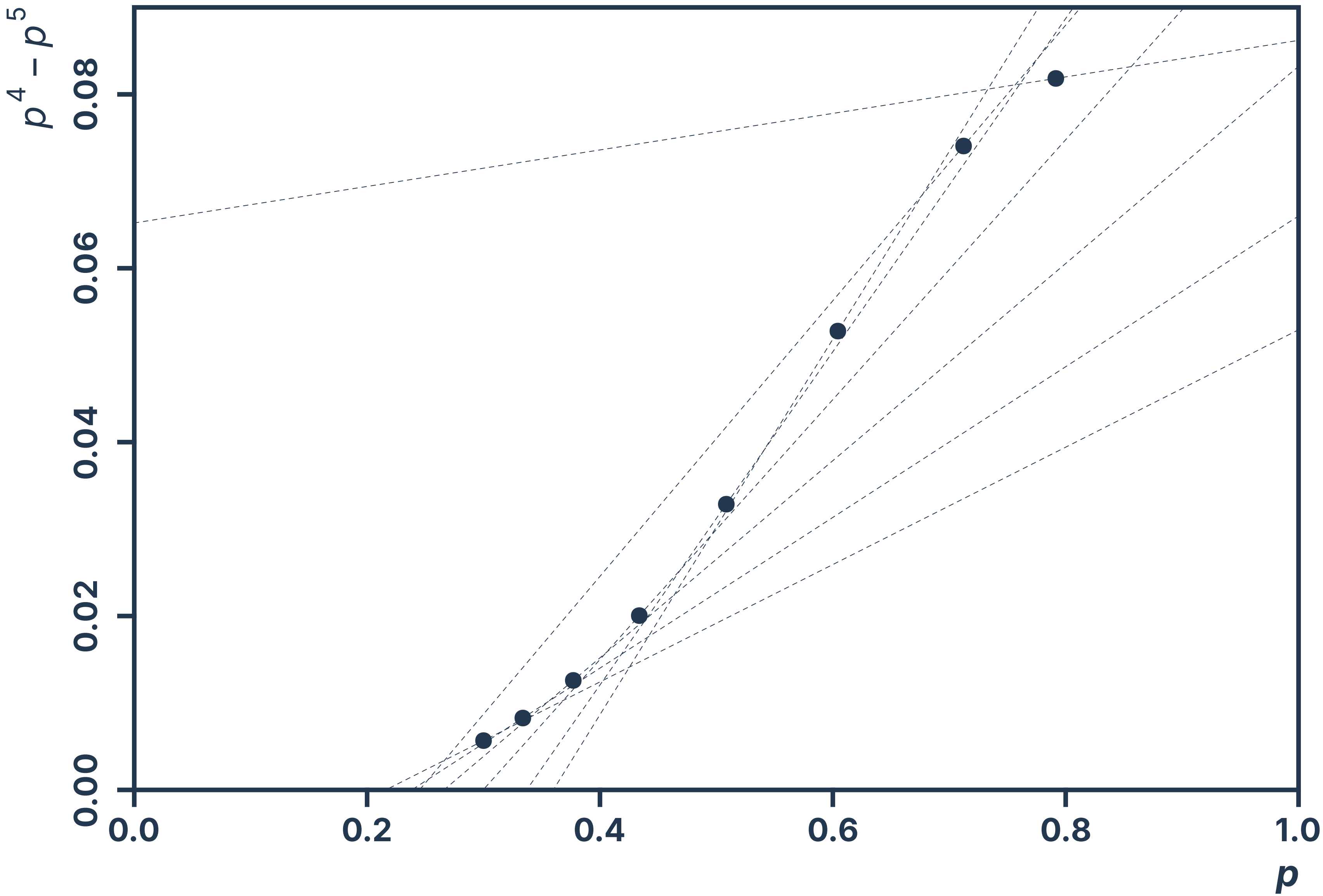
HILL CLIMBING (MAP/ML)

$\text{Prob}(p|\text{data}, \text{model}) \propto p^4 - p^5$

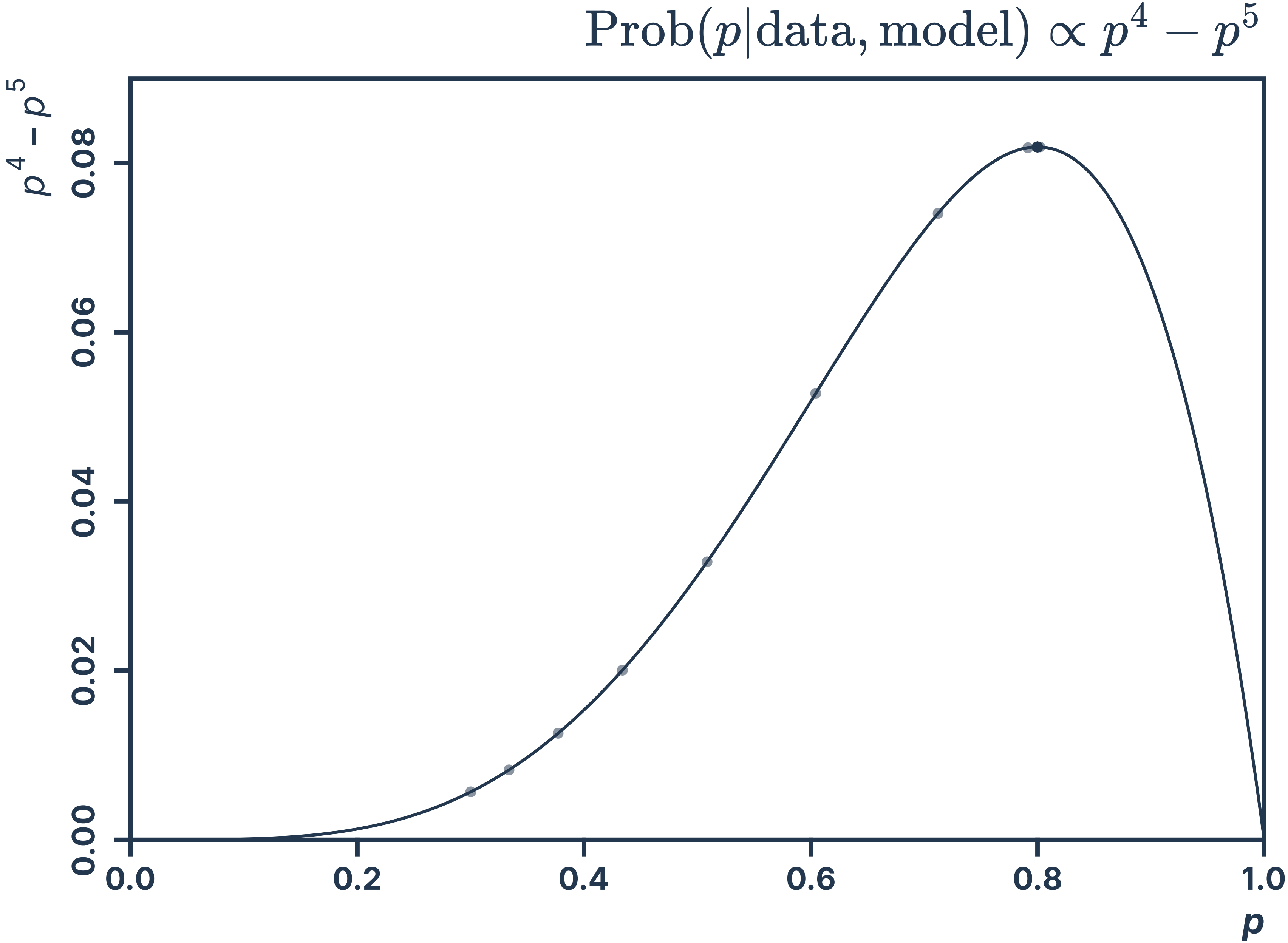


HILL CLIMBING (MAP/ML)

$$\mathrm{Prob}(p|\mathrm{data},\mathrm{model}) \propto p^4 - p^5$$

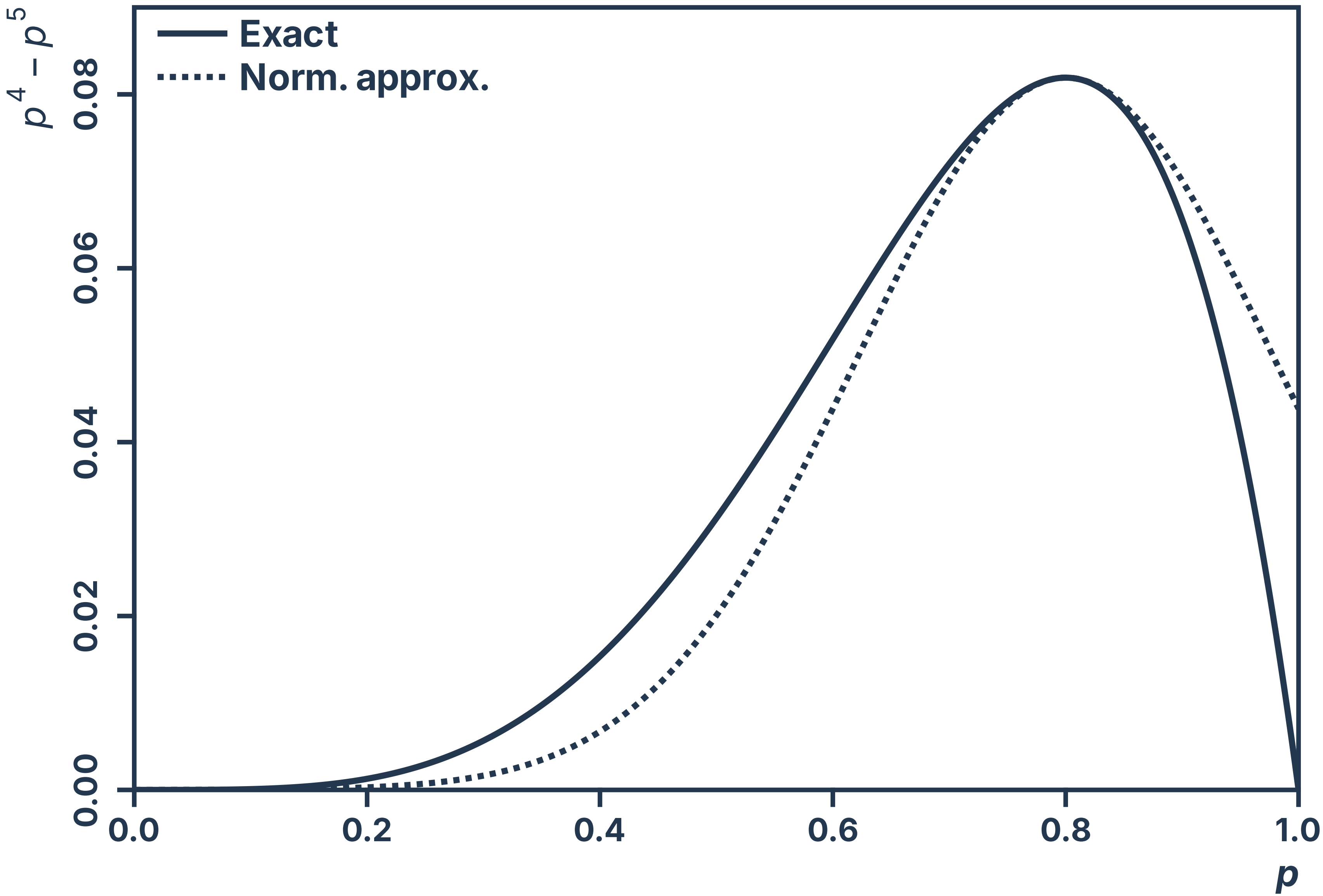


HILL CLIMBING (MAP/ML)



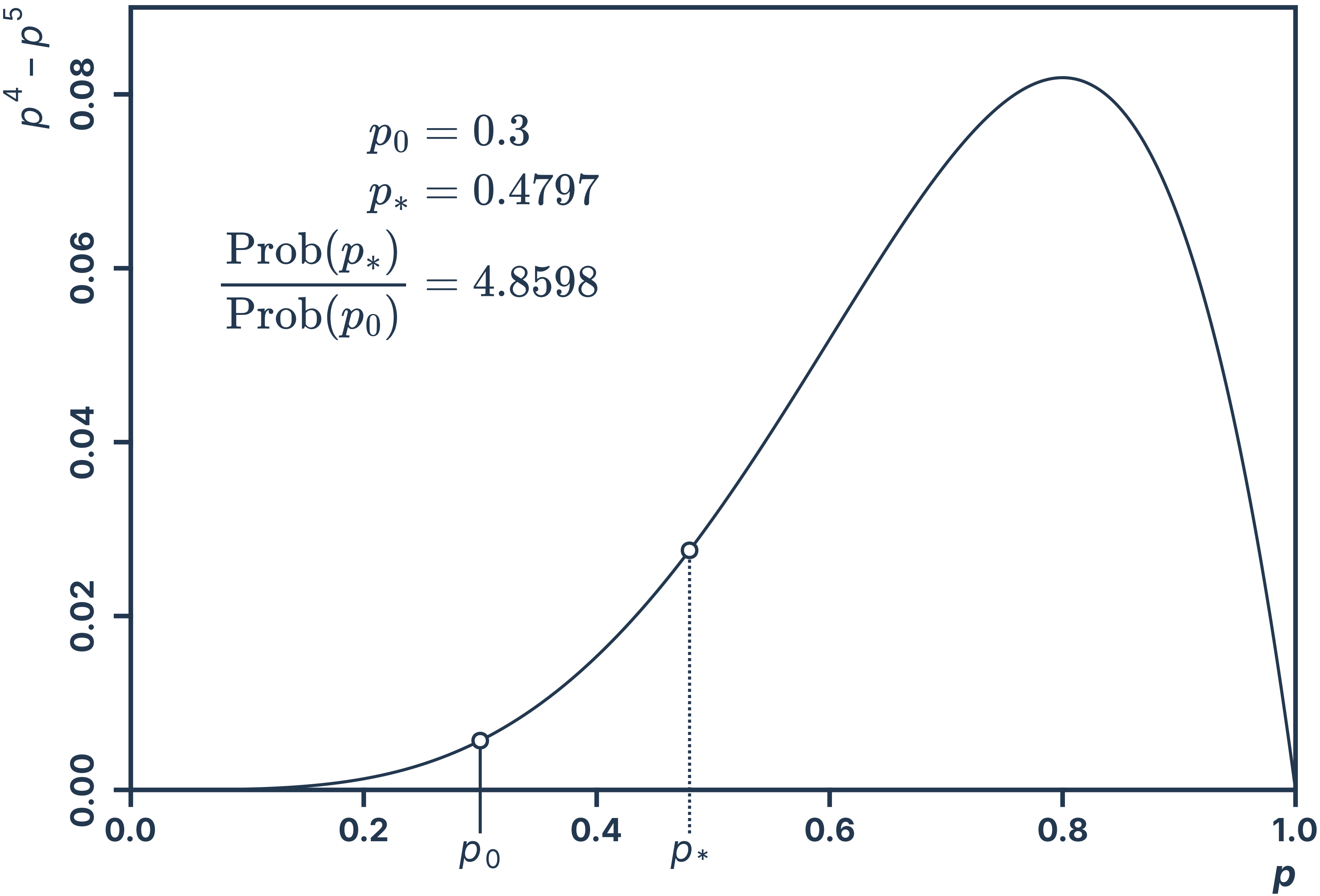
HILL CLIMBING (MAP/ML)

$$\mathrm{Prob}(p|\mathrm{data},\mathrm{model}) \propto p^4 - p^5$$

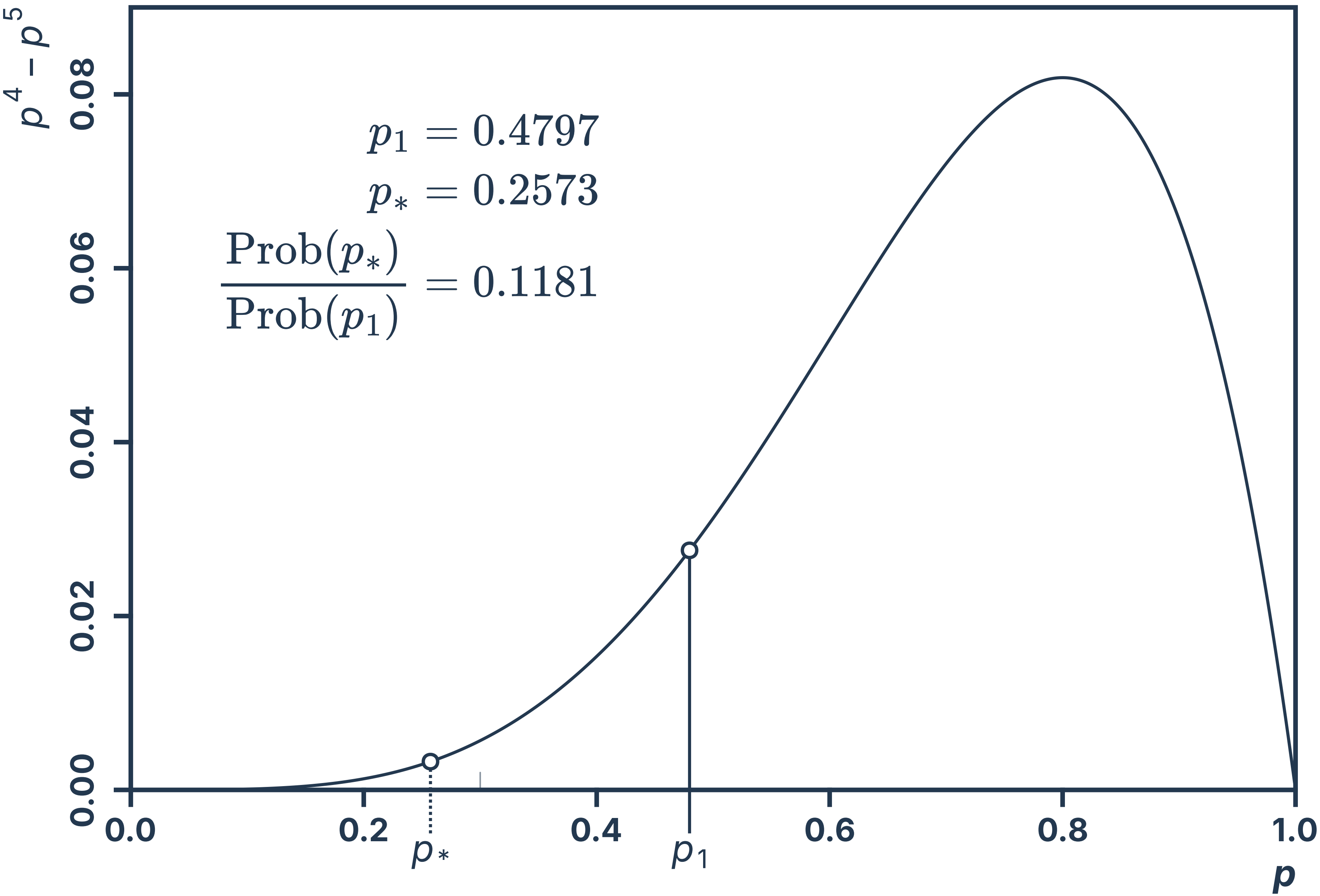


Markov chain Monte Carlo

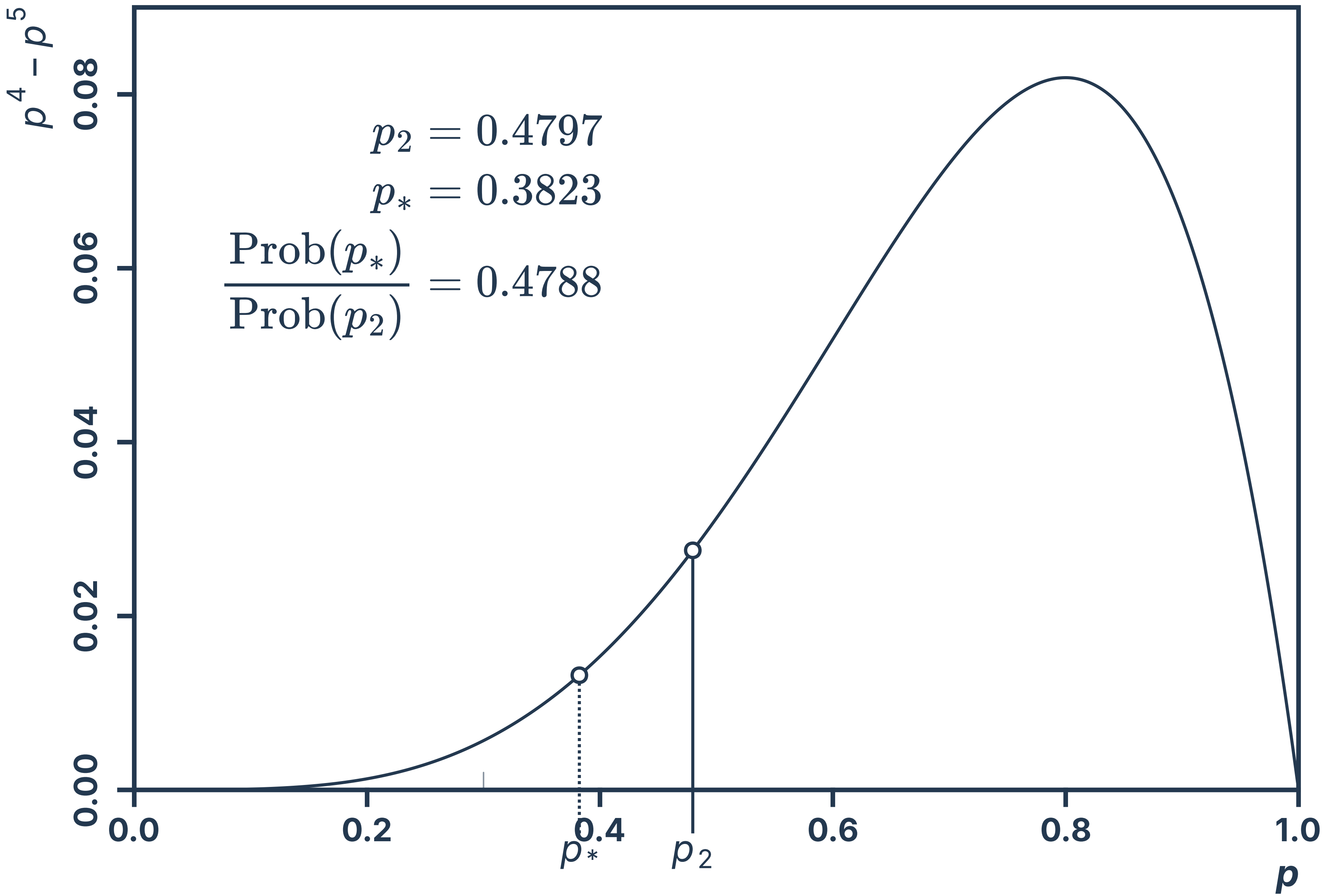
$$\mathrm{Prob}(p|\mathrm{data},\mathrm{model}) \propto p^4 - p^5$$



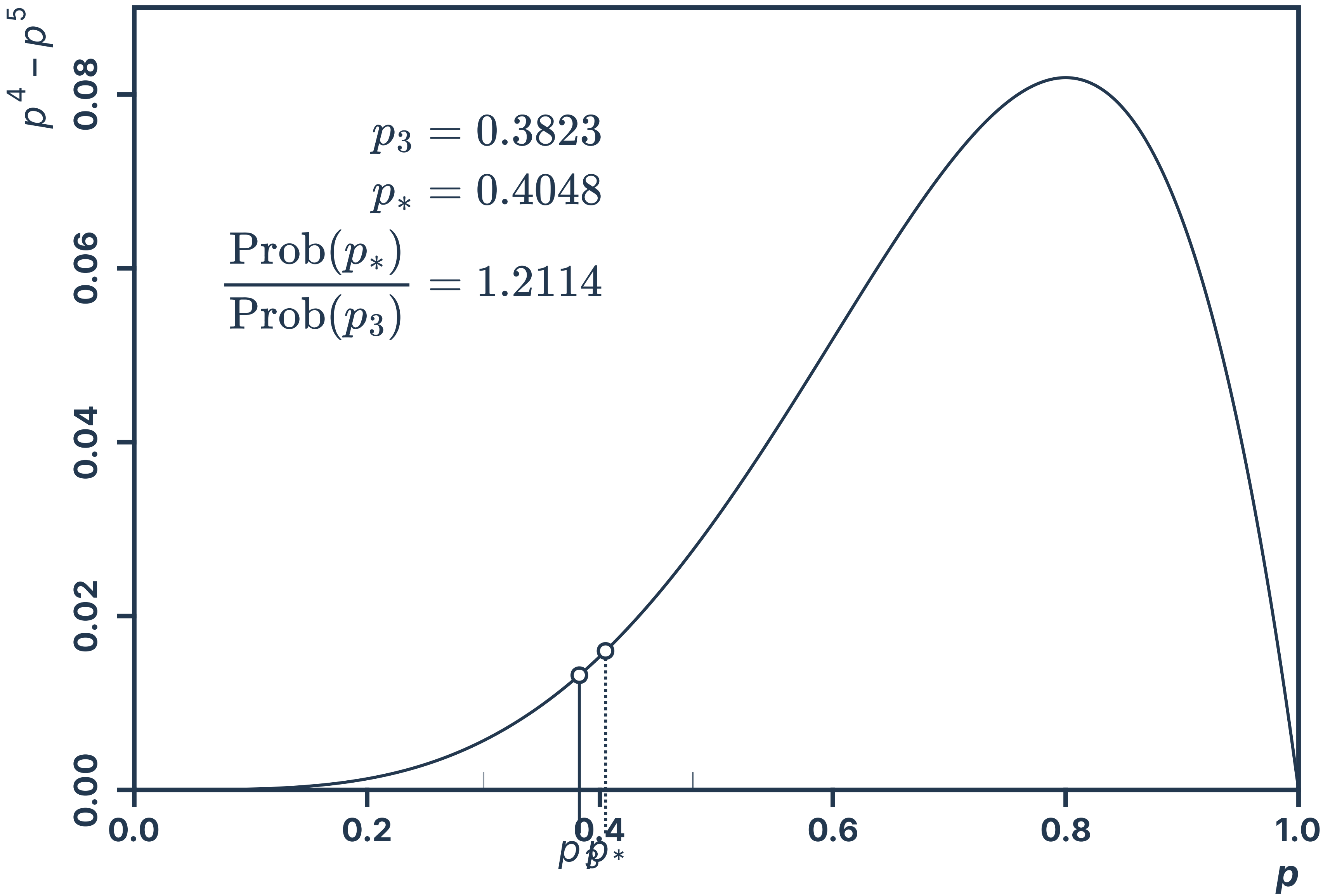
$$\mathrm{Prob}(p|\mathrm{data},\mathrm{model}) \propto p^4 - p^5$$



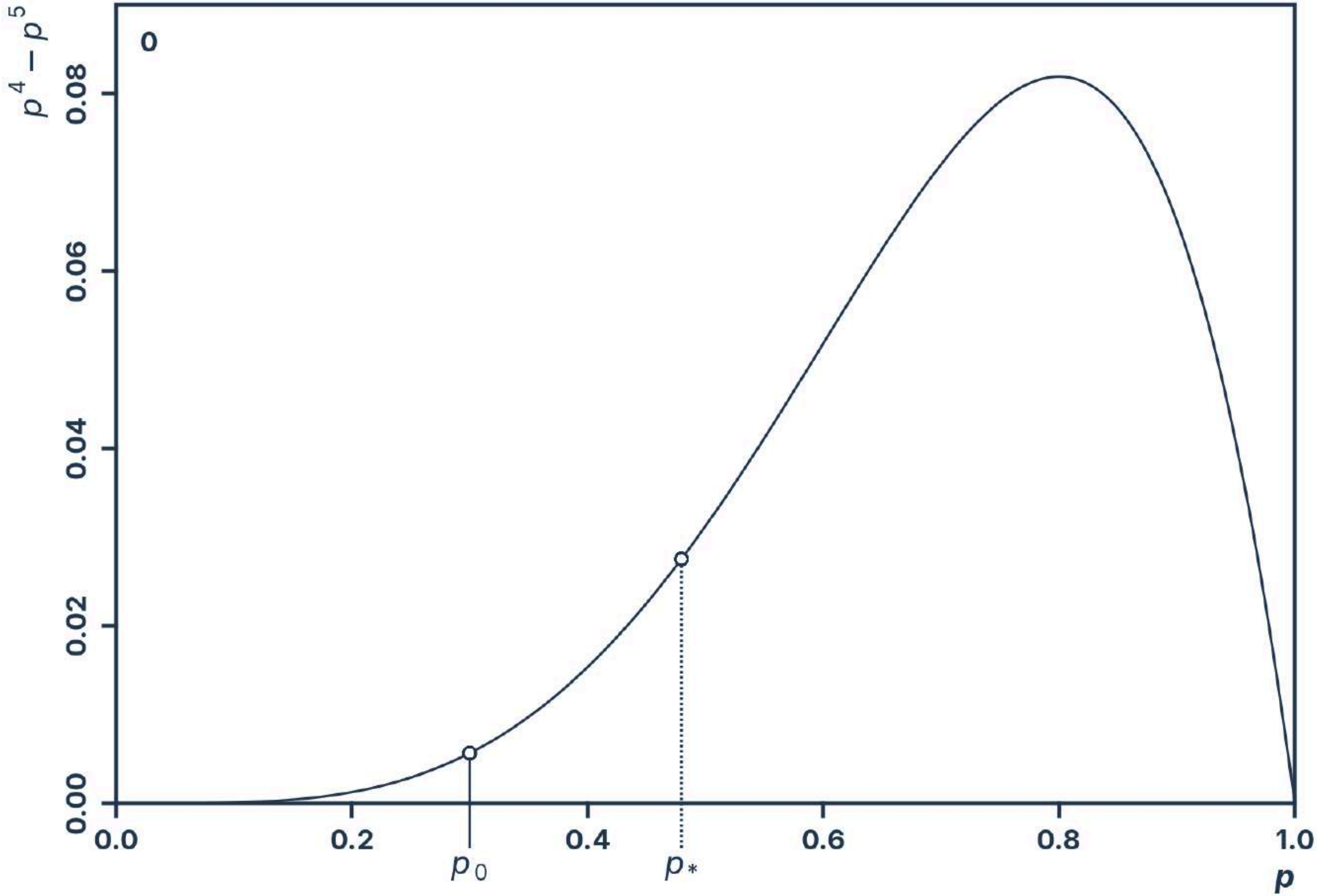
$$\text{Prob}(p|\text{data, model}) \propto p^4 - p^5$$

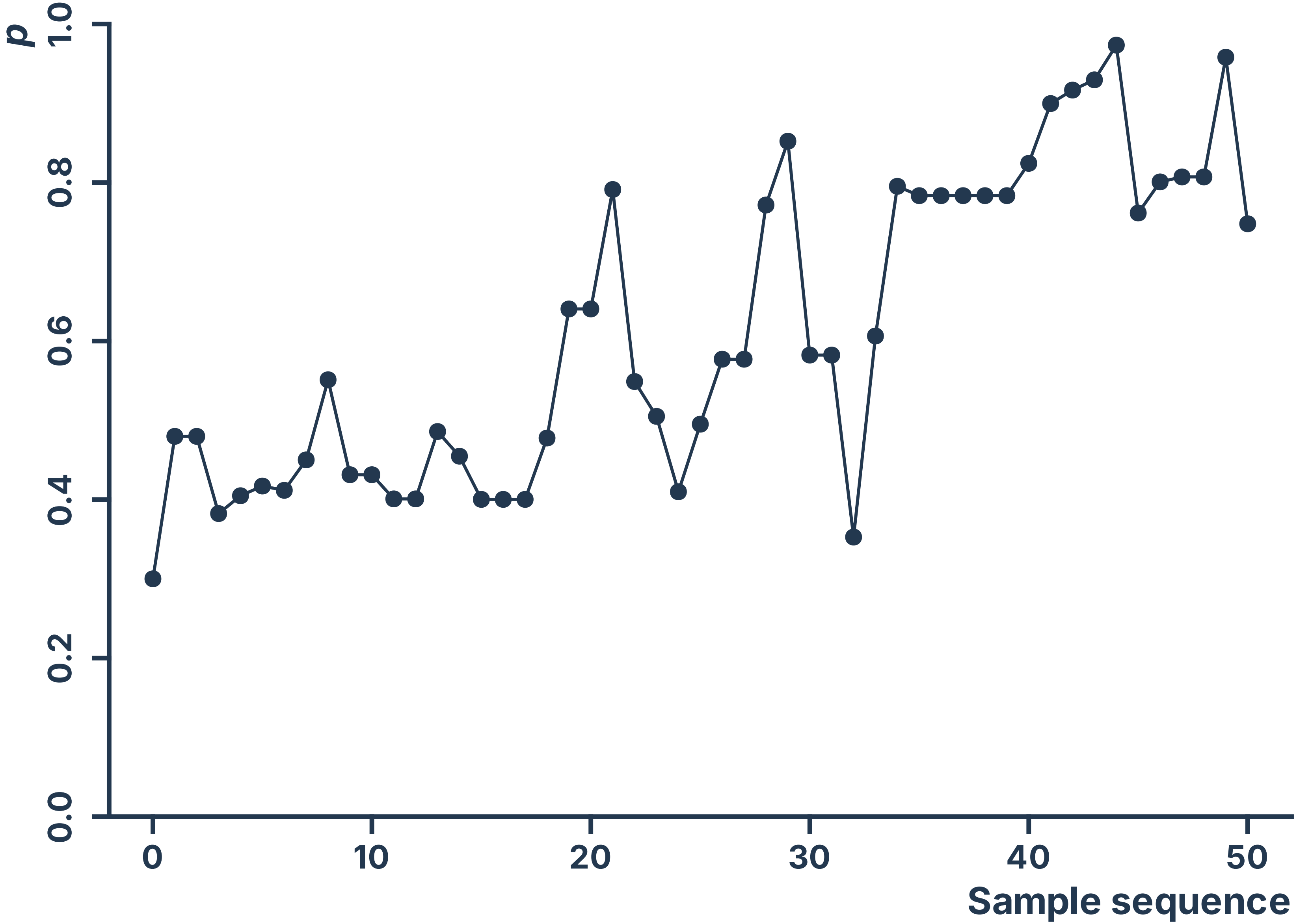


$$\text{Prob}(p|\text{data, model}) \propto p^4 - p^5$$

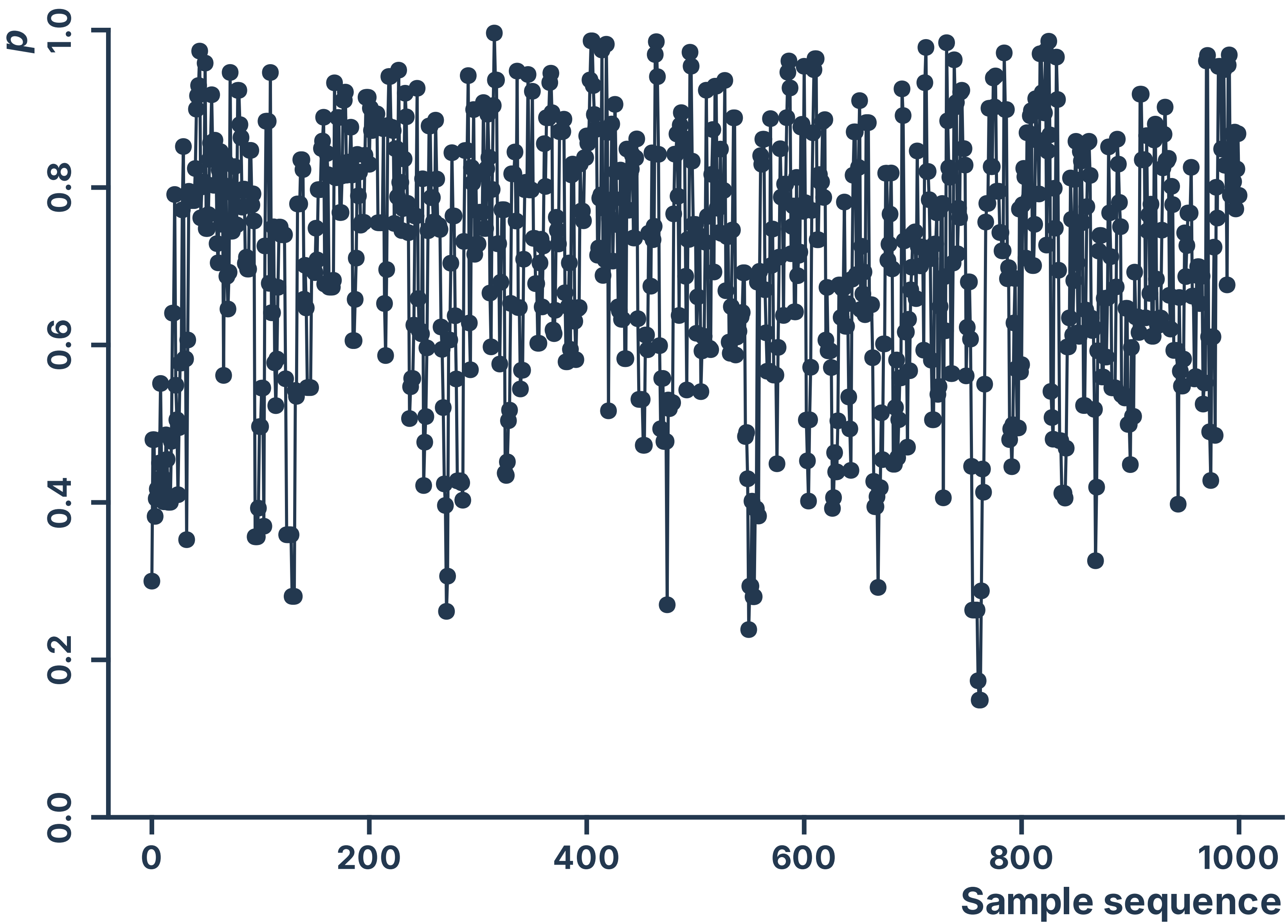


$$\text{Prob}(p|\text{data}, \text{model}) \propto p^4 - p^5$$

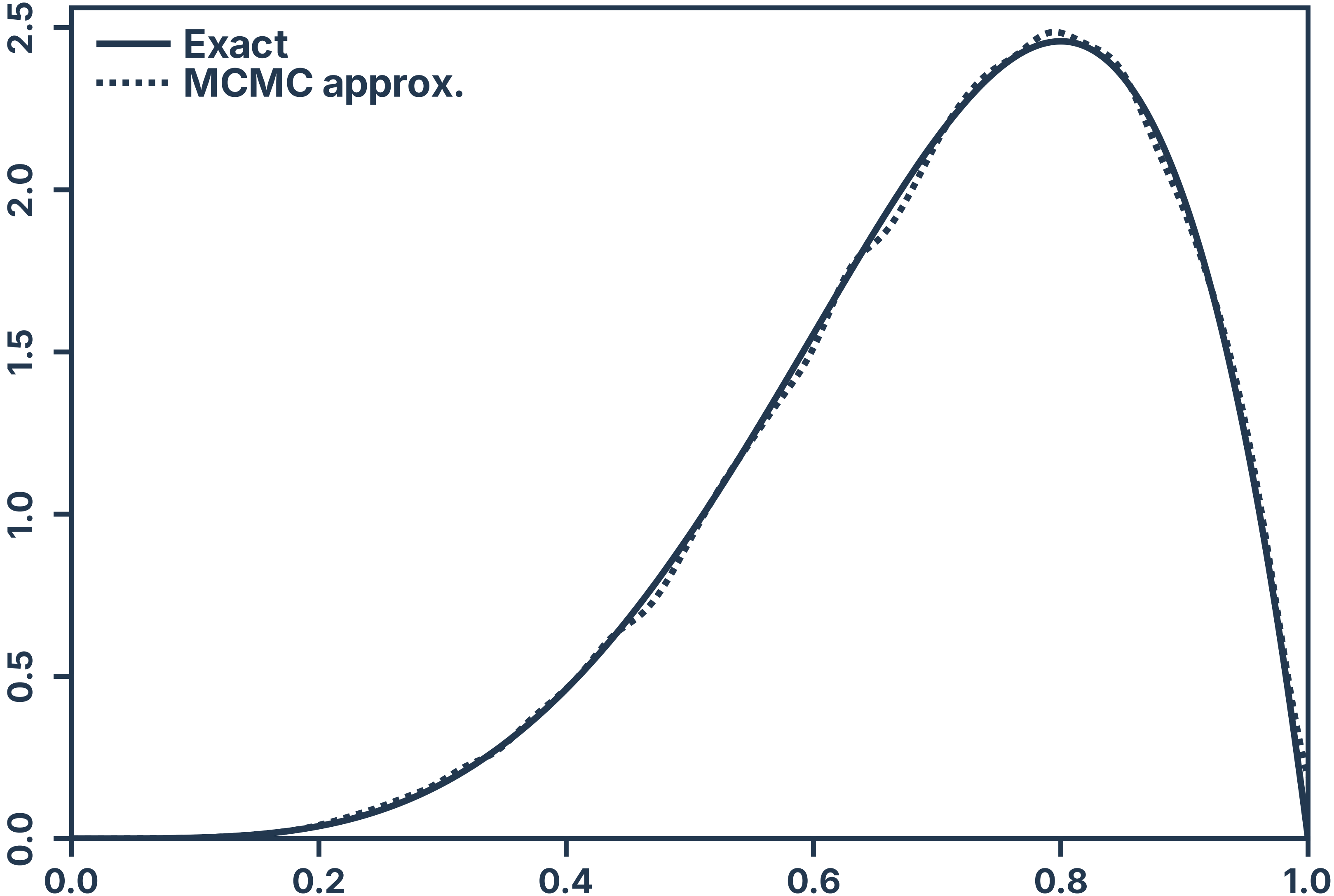



p
1.0
0.8
0.6
0.4
0.2
0.0
0
10
20
30
40
50
Sample sequence

MARKOV CHAIN MONTE CARLO



$\text{Prob}(p|\text{data}, \text{model}) \propto p^4 - p^5$



Density based on 50,000 posterior observations

Hamil-
tonian
Monte
Carlo

Simulate a physical system

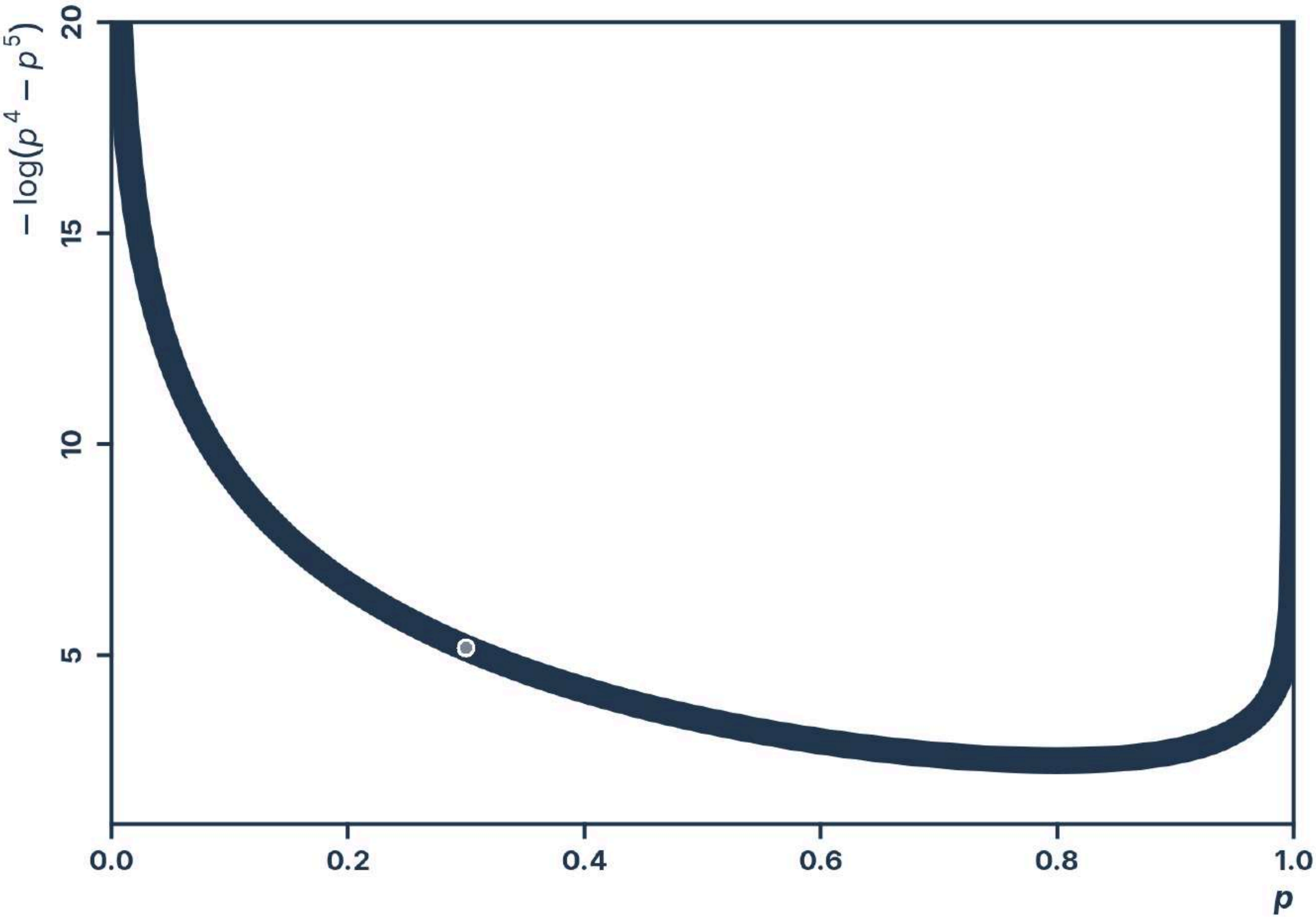
"Energy" at any point in the parameter space is proportional to the negative log posterior probability:

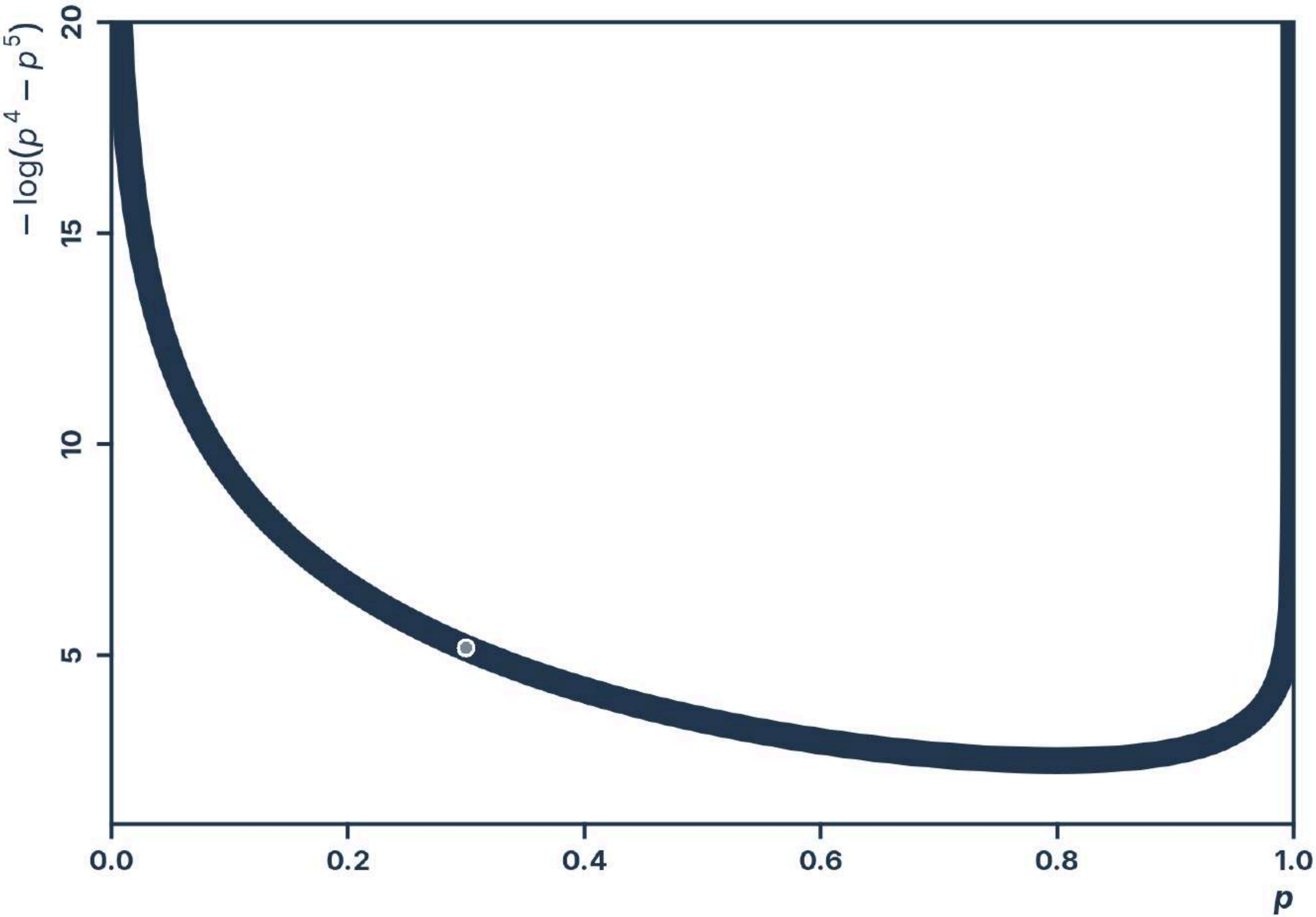
$$U_\theta \propto -\log\left(\text{Prob}(\theta|\text{data}, \text{model})\right)$$

Get random draws by 'perturbing' a particle in that field

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 amount of time is the next candidate draw from the posterior.


$-\log(p^4 - p^5)$
20
15
10
5
0.0
0.2
0.4
0.6
0.8
1.0
p



$-\log(p^4 - p^5)$
20
15
10
5
0.0
0.2
0.4
0.6
0.8
1.0
p

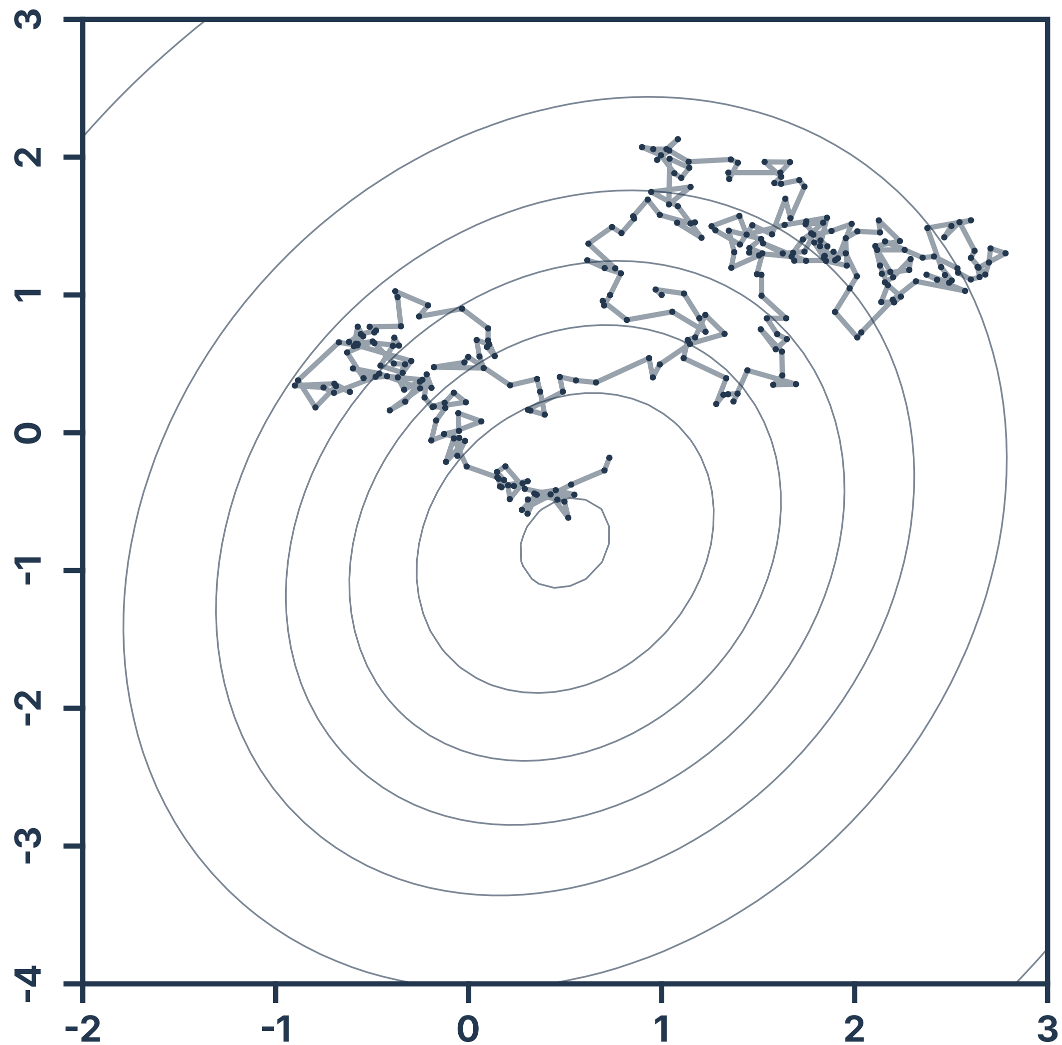
HMC VS. MCMC

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.

What can possibly go wrong?

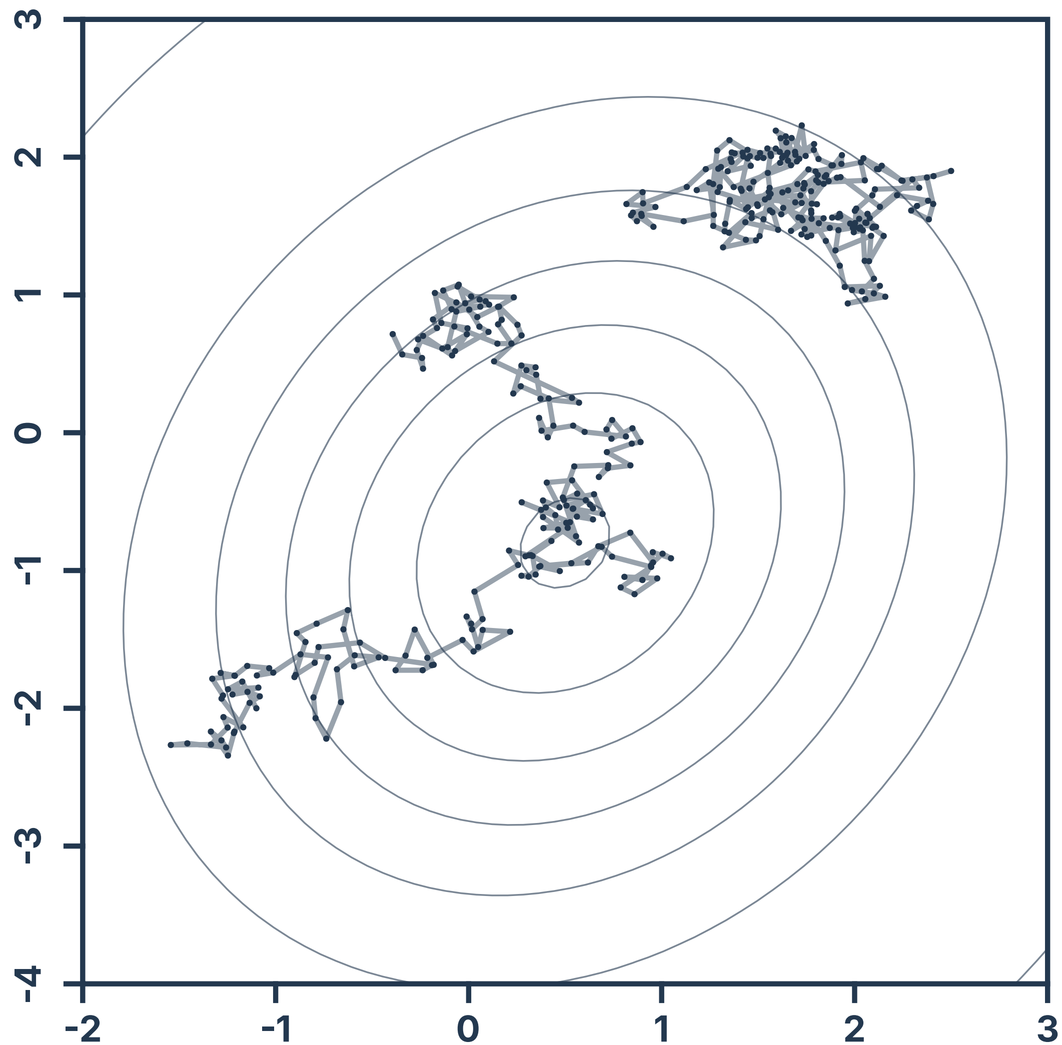
Autocorrelation

- 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)***



Non-convergence

- Sampling may have trouble “converging” (properly 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 between multiple chains



Divergent transistions

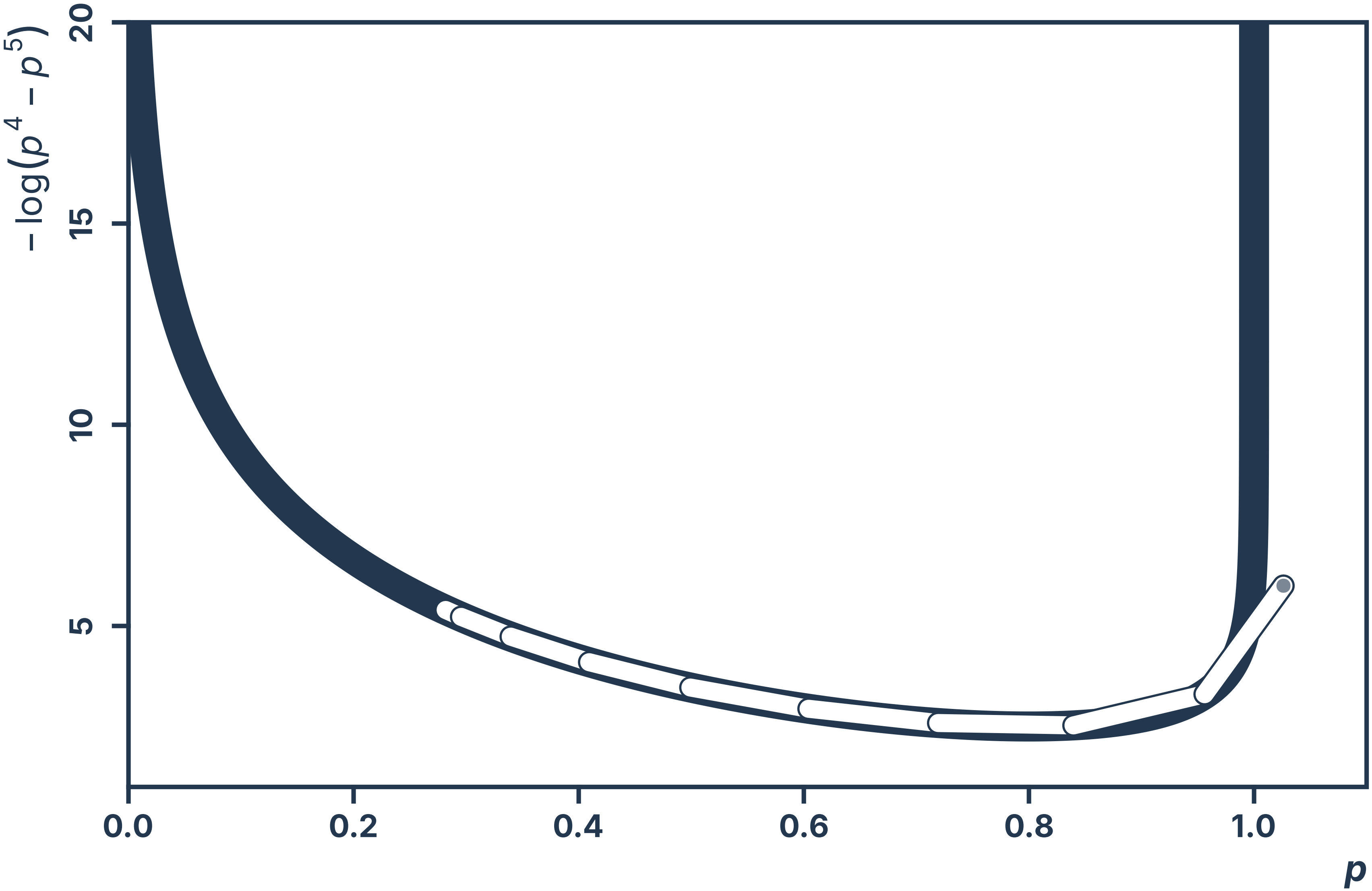
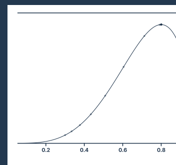
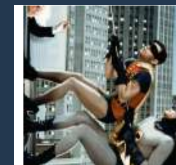


Image credit

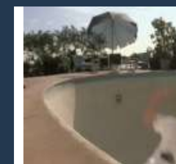


Figures by Peter McMahan (source code)



Still from Batman (1966)

Still from The Legend of the Drunken Master (1994)



Clip from Gleaming the Cube (1989)

Clip from Raiders of the Lost Ark (1981)