

# Model Selection

---

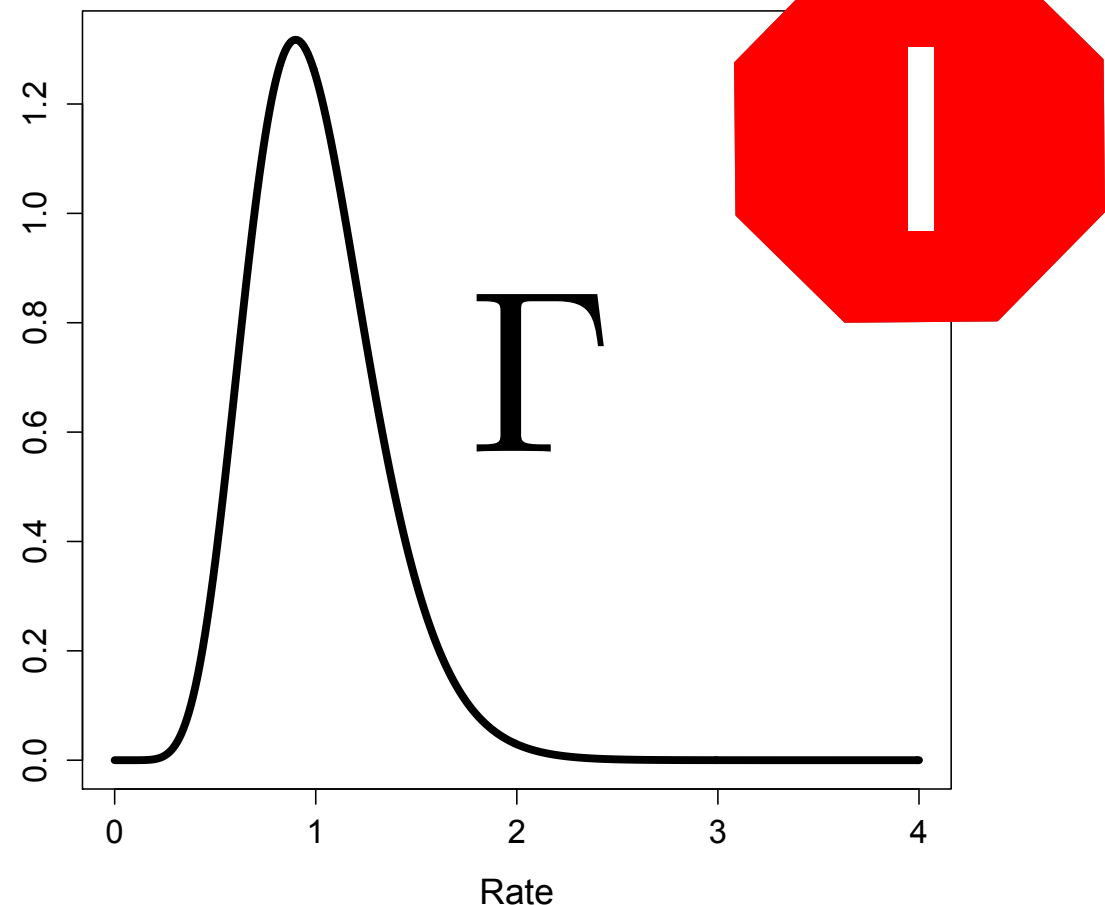
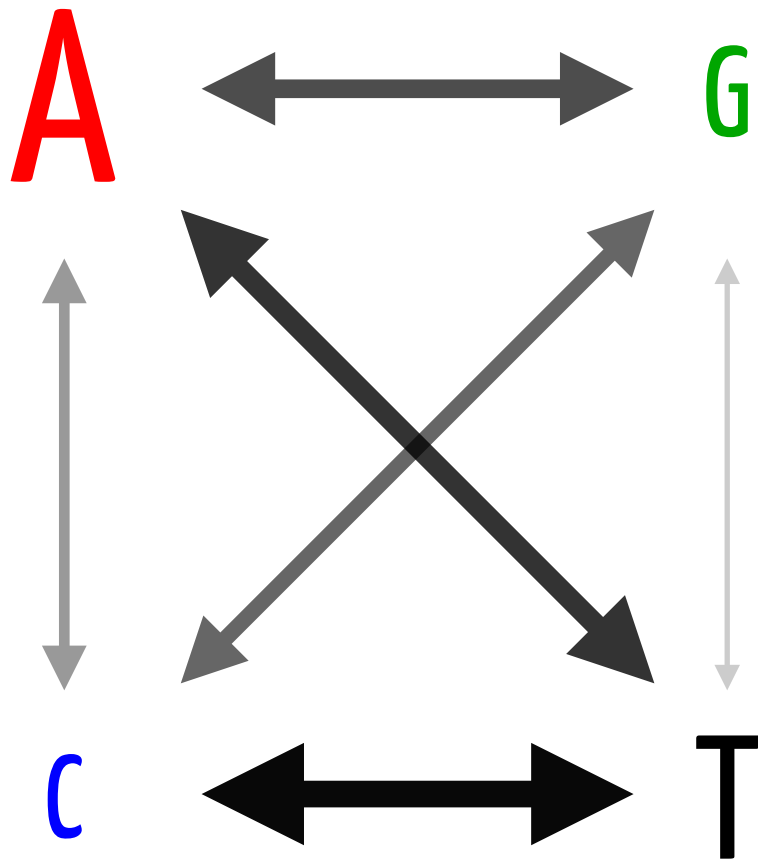
# How do we use models?

---

- Select the best available model (model selection)
- Critically evaluate the fit of this model (model adequacy)
- Accept, refine, or reject (the art)

# Should we always use GTR+I+ $\Gamma$ ?

- GTR+I+ $\Gamma$  seems pretty complicated!
- 10 parameters to describe change in 4 nucleotides
- Surely that's enough to capture evolutionary dynamics.



# Just remember...

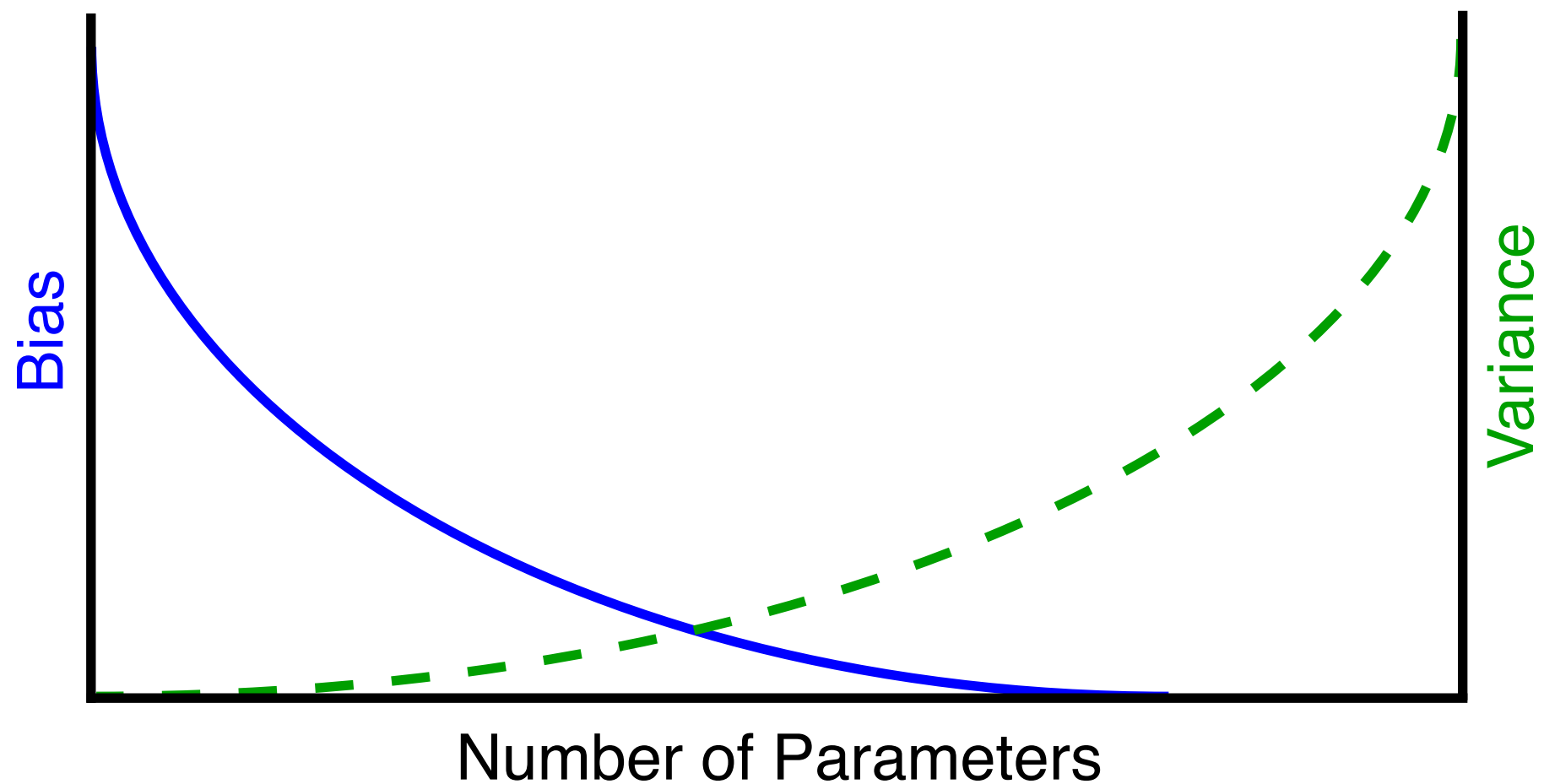
---

- Nearly all of our models (or at least the ones we usually consider) still assume a lot of things:
  - Independence of sites
  - Constant site rates across the tree
  - Constant base frequencies across the tree
  - Consistent evolutionary dynamics across the tree

# Model Selection

---

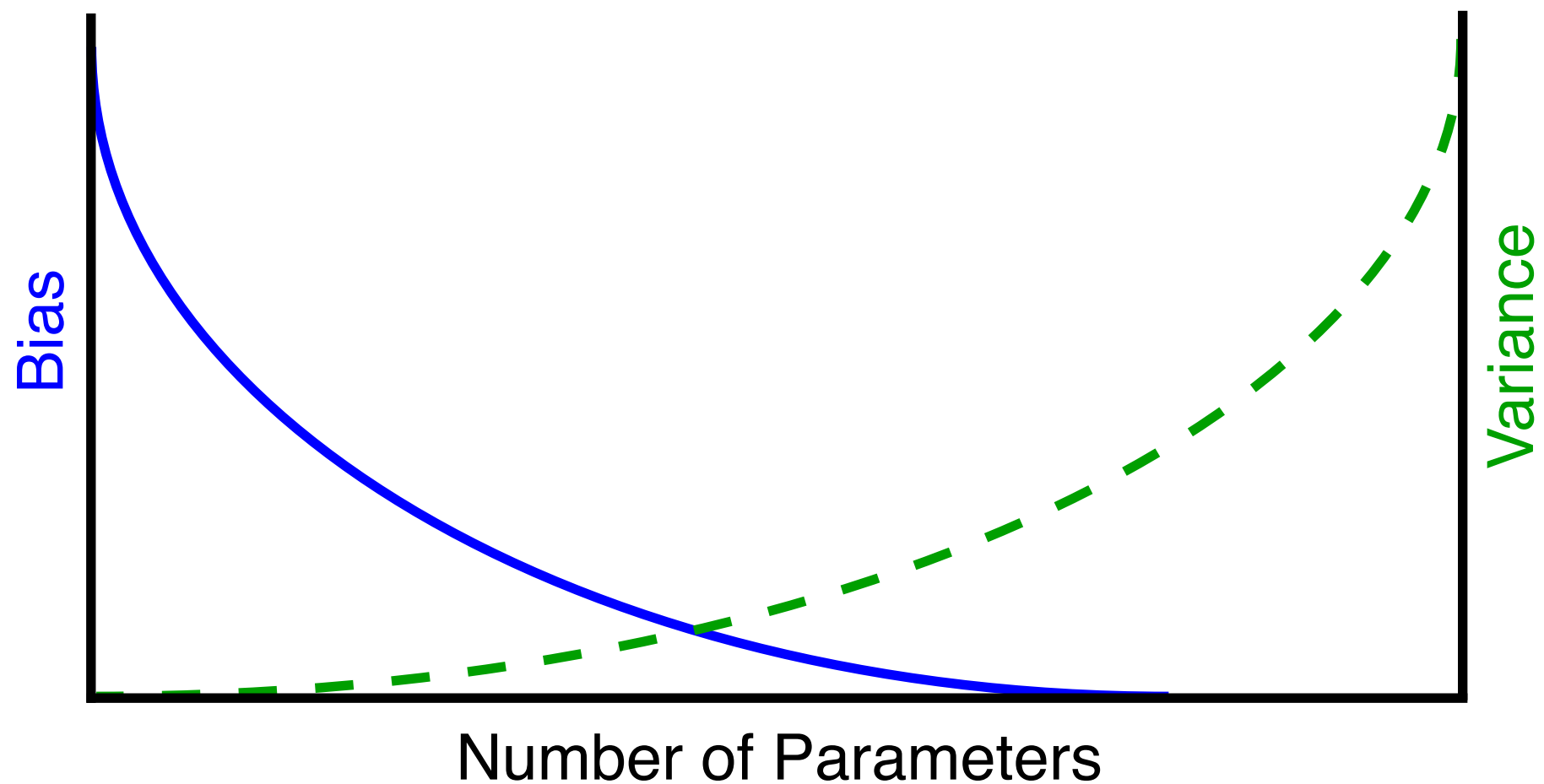
## The Fundamental Tradeoff



# Model Selection

---

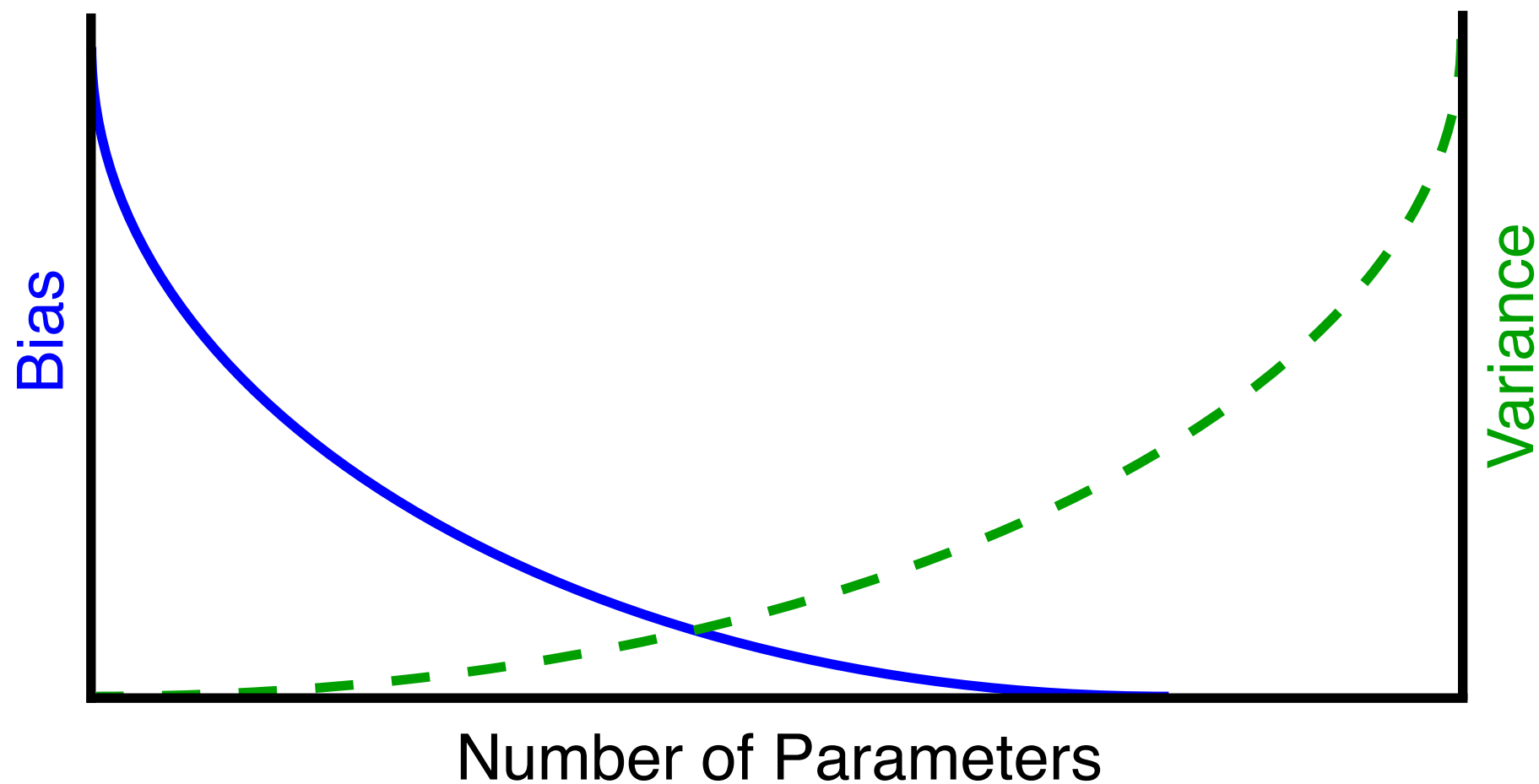
Model too simple!  
We're misinterpreting the data.



# Model Selection

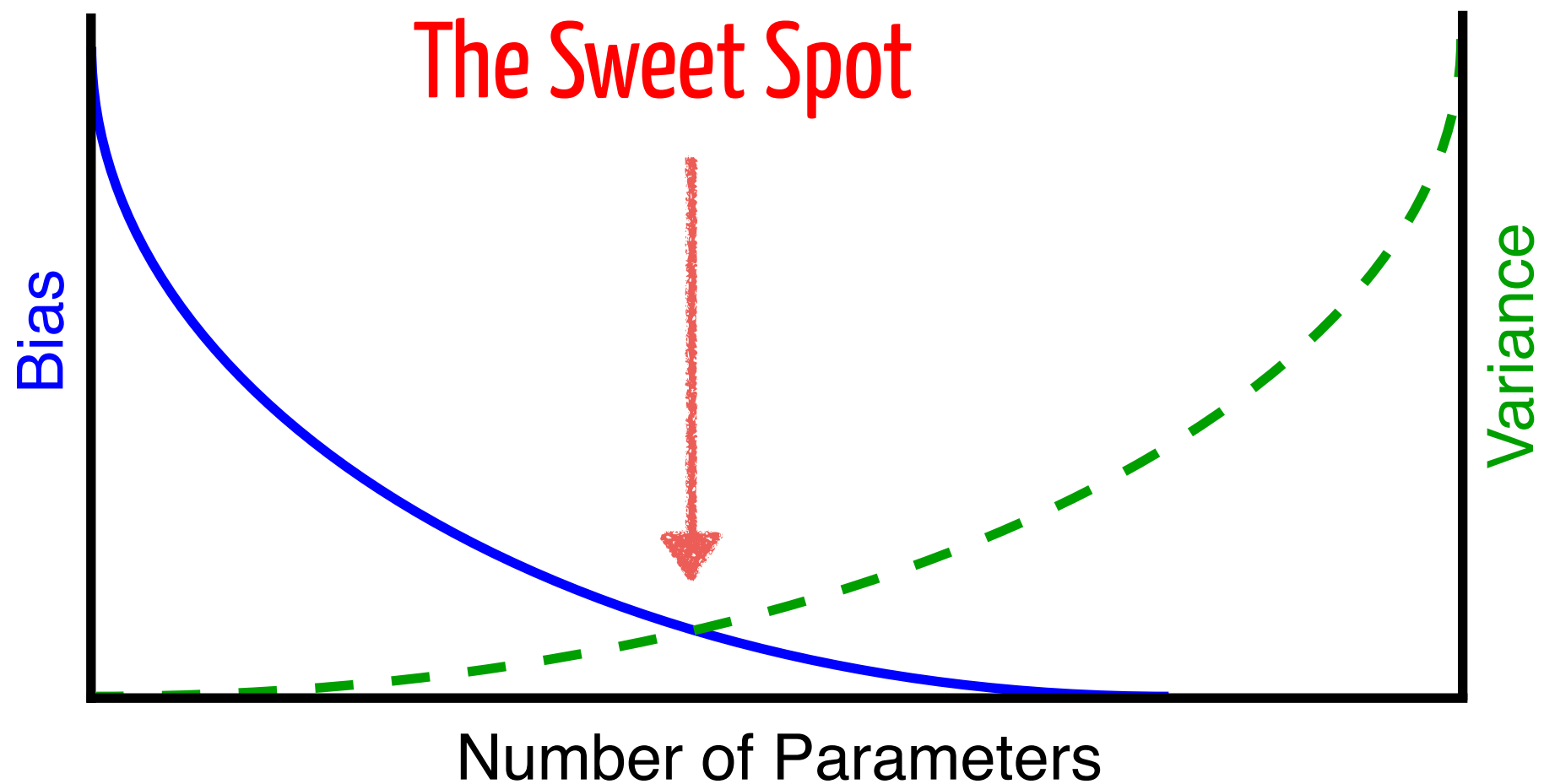
---

Model too complicated!  
We don't have enough information.



# Model Selection

---



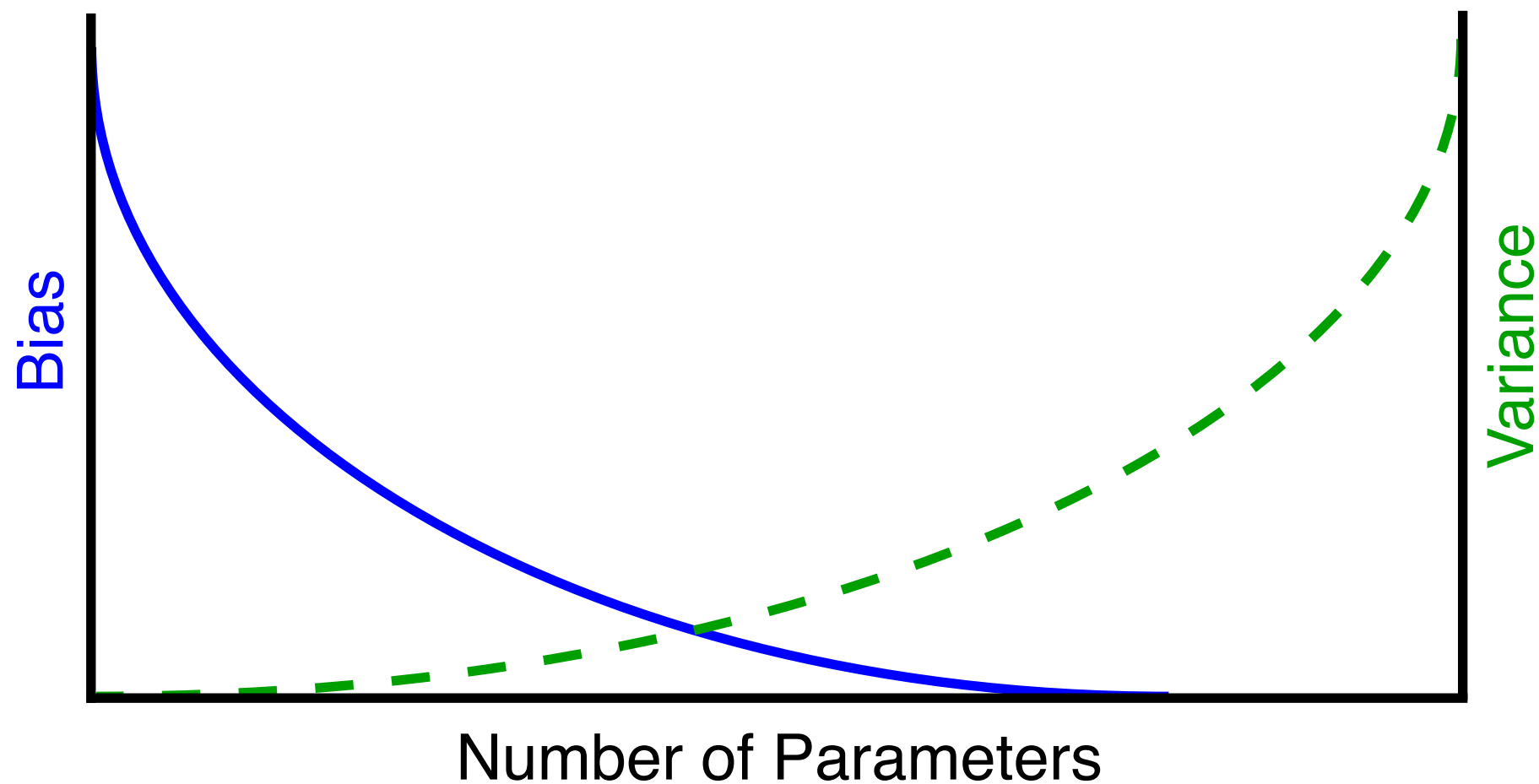


# Model Selection

---


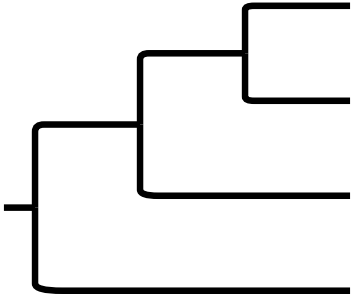
**Bias** and **Variance** can be traded off in different ways.

This leads to **multiple criteria** for model selection.



# The Likelihood Function

---

$$P(\text{  | \theta, \text{  })$$

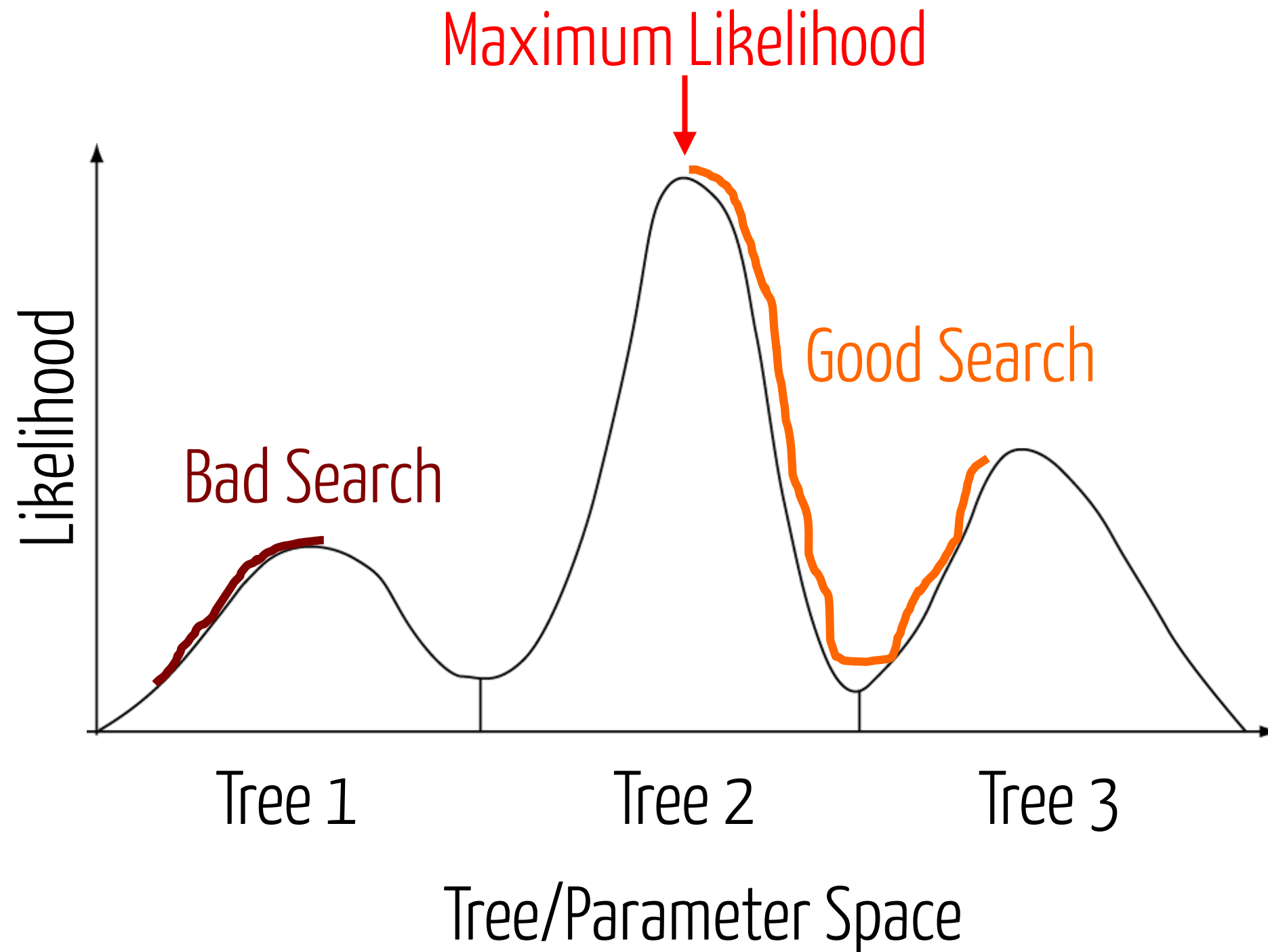
Read as “**the probability of the sequence data given a tree and model**”.

The quantity by which the data provide information.

Compares how well different trees and models predict the observed data or as a “**measure of relative surprise**”.

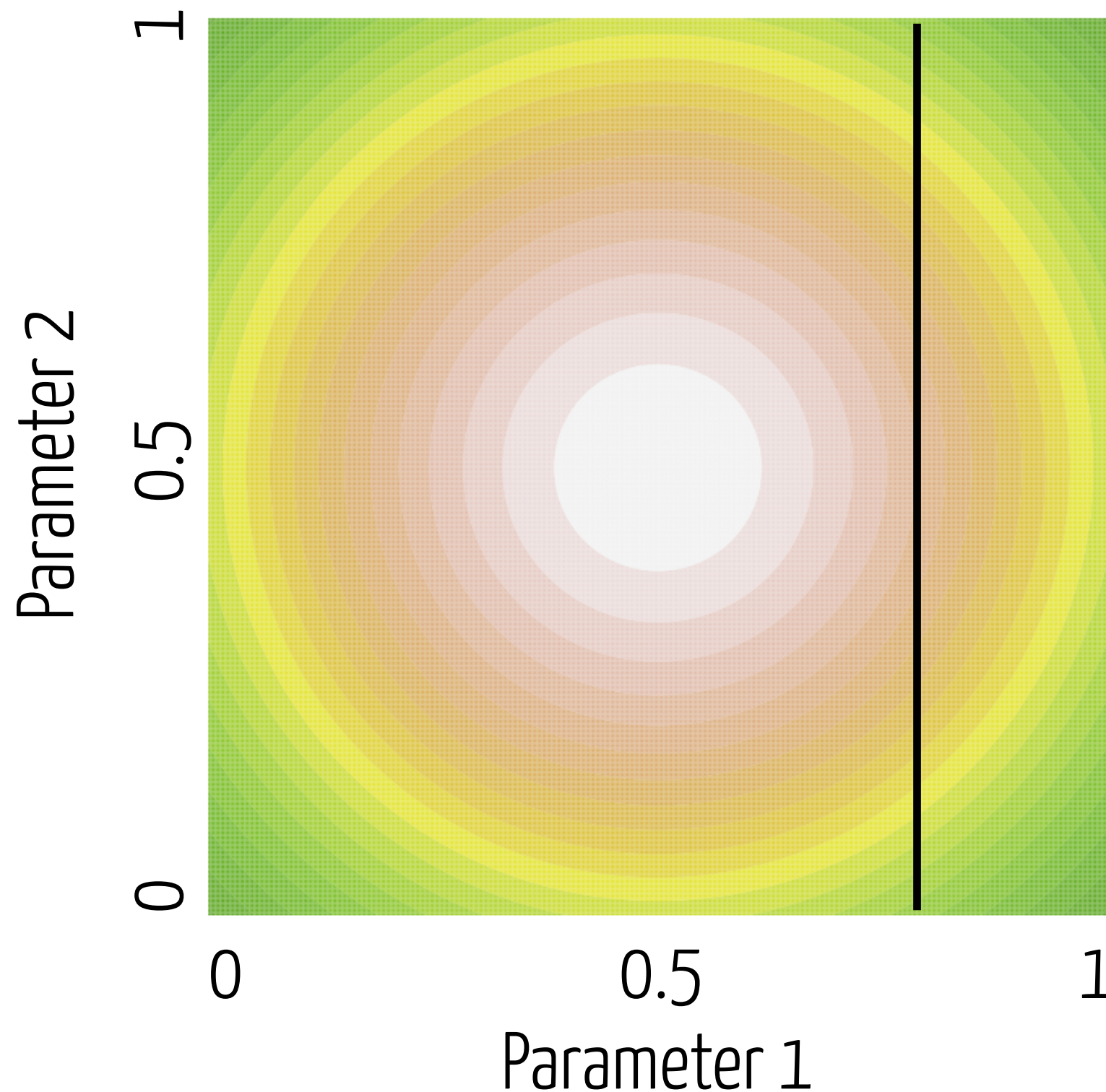
# Maximum Likelihood

---



# More Parameters = Better Likelihood

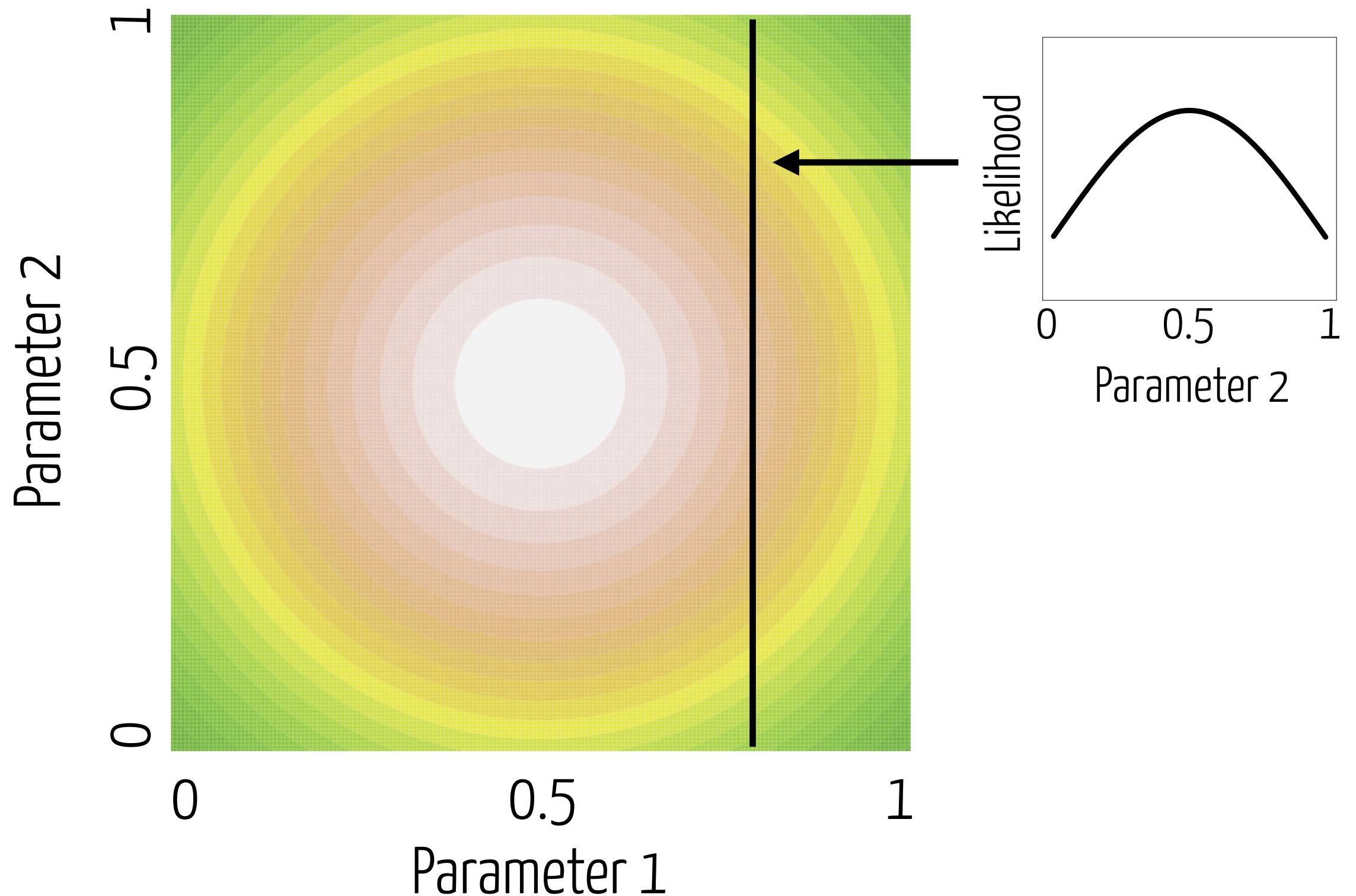
---





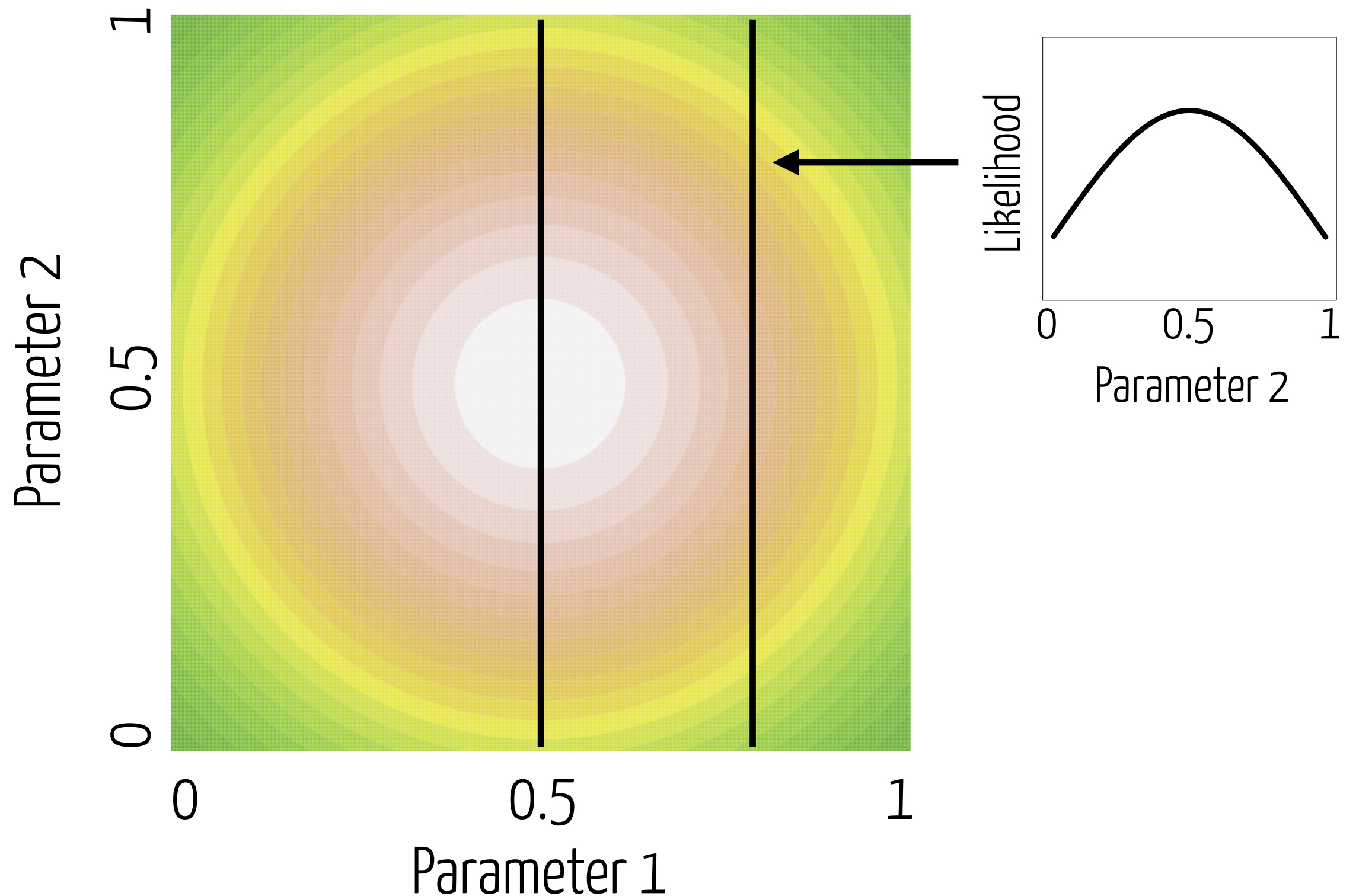
# More Parameters = Better Likelihood

---



# More Parameters = Better Likelihood

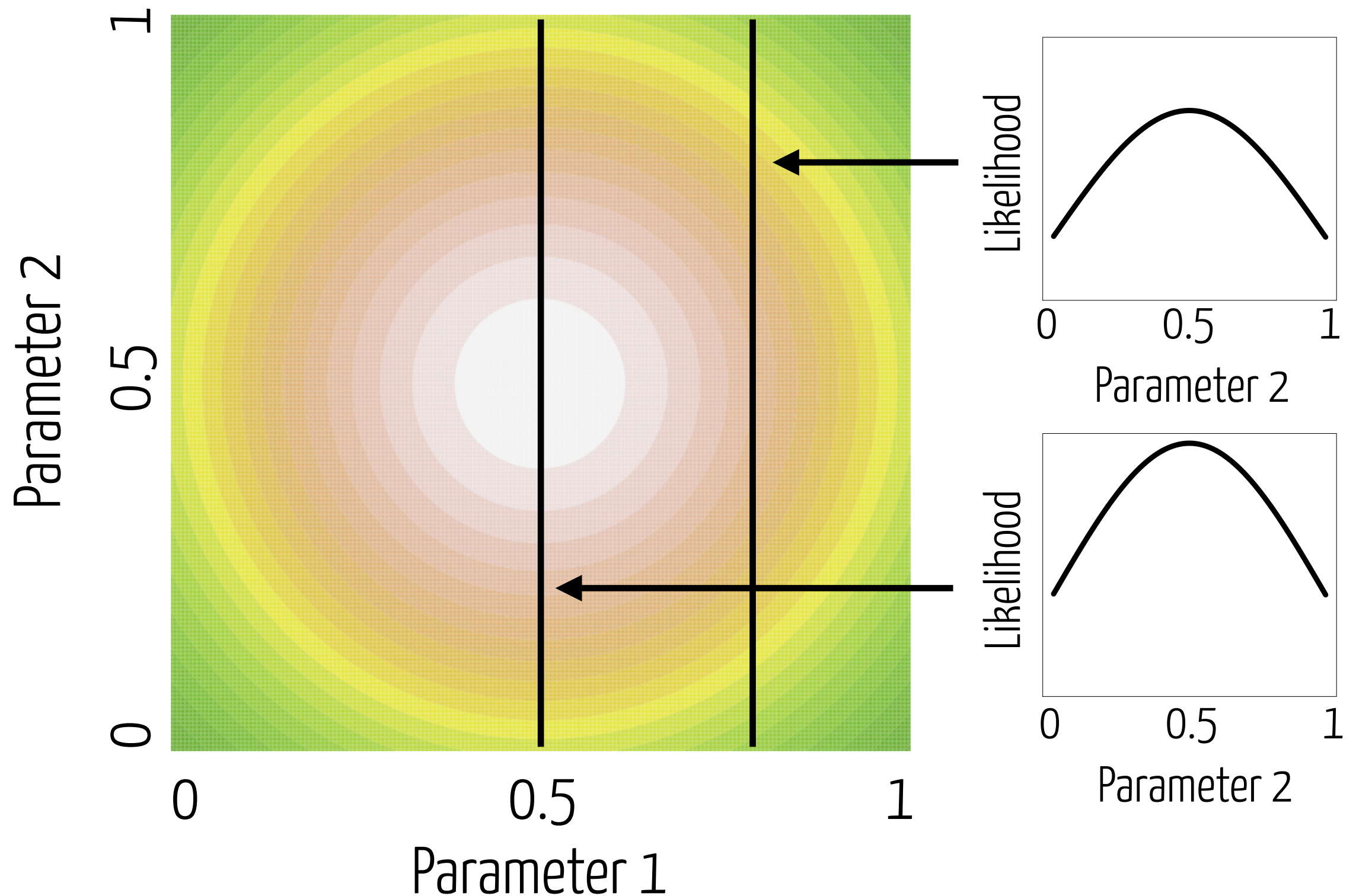
---





# More Parameters = Better Likelihood

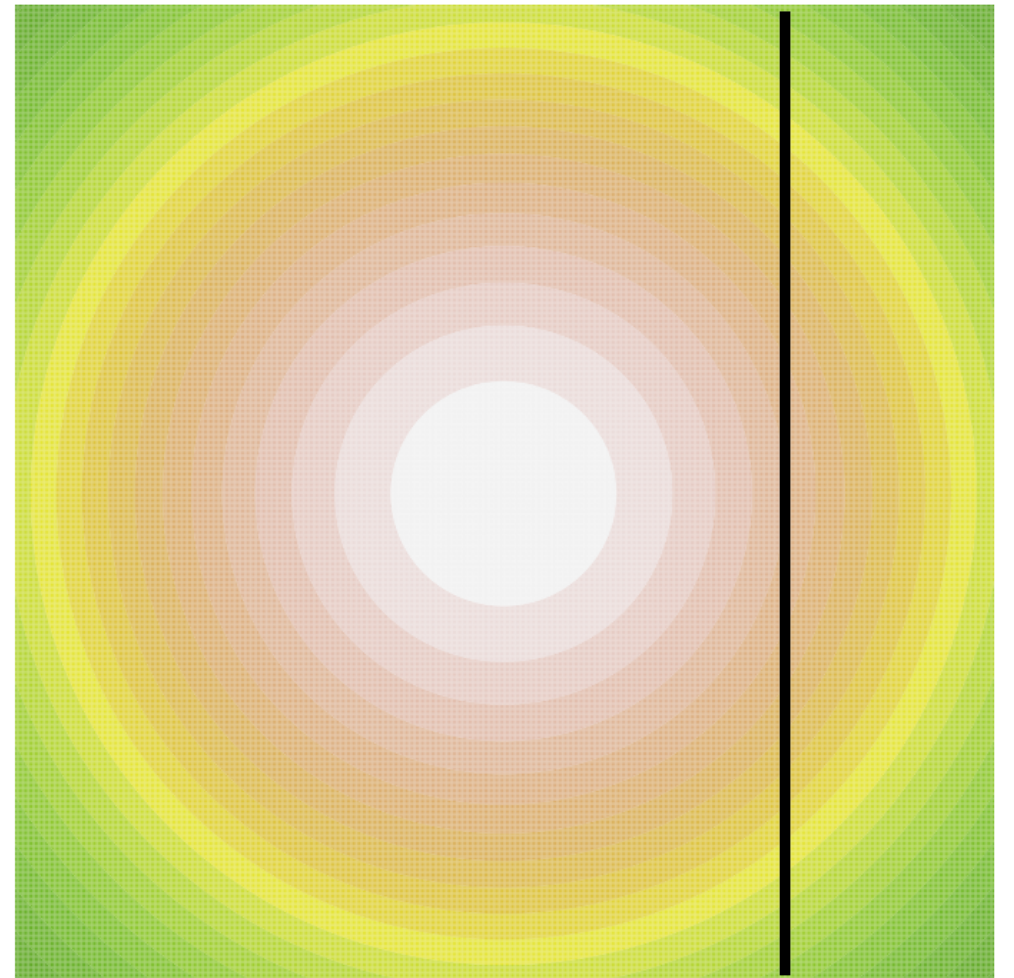
---



# ML-based Model Selection

---

If the **more complex** model always gives a likelihood that is **at least as good** as a simpler model, even if the simpler one is true, we need ways to assess **whether it's enough better to warrant our attention**.





# ML-based Model Selection

---

If the **more complex** model always gives a likelihood that is **at least as good** as a simpler model, even if the simpler one is true, we need ways to assess **whether it's enough better to warrant our attention**.

- Akaike's Information Criterion (**AIC**)
- Bayesian Information Criterion (**BIC**)
- Likelihood Ratio Test (**LRT**)

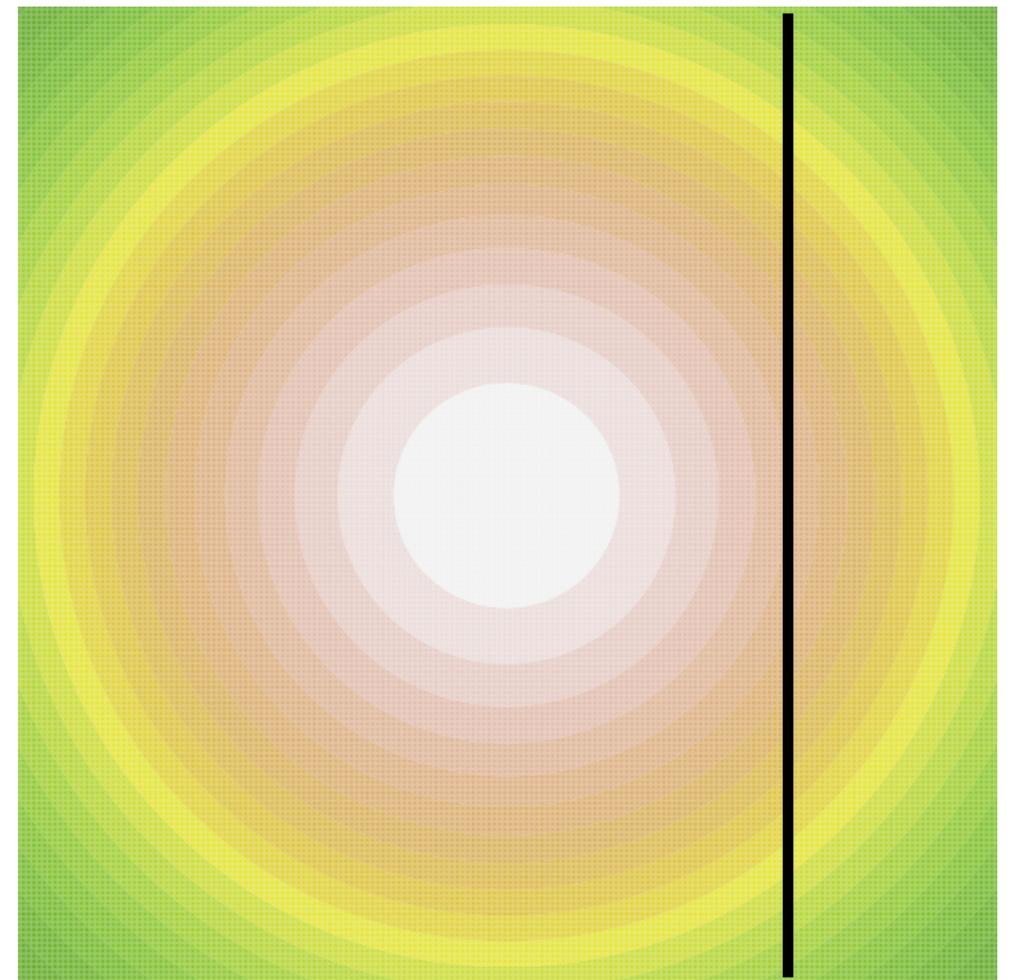


# ML-based Model Selection

---

If the **more complex** model always gives a likelihood that is **at least as good** as a simpler model, even if the simpler one is true, we need ways to assess **whether it's enough better to warrant our attention**.

- Akaike's Information Criterion (**AIC**)
- Bayesian Information Criterion (**BIC**)
- Likelihood Ratio Test (**LRT**)



Different penalties for extra parameters.

# ML-based Model Selection

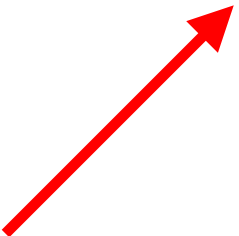
---

## Akaike's Information Criterion (**AIC**)

Minimum AIC preferred.

$$\text{AIC} = 2k - 2\ln(\hat{L})$$

Penalty for more  
parameters ( $k$ ).



Likelihood term becomes  
more negative when  $\hat{L}$  worse.



# ML-based Model Selection

---

$$\text{AIC} = 2k - 2\ln(\hat{L})$$

$$\text{BIC} = \ln(n)k - 2\ln(\hat{L})$$



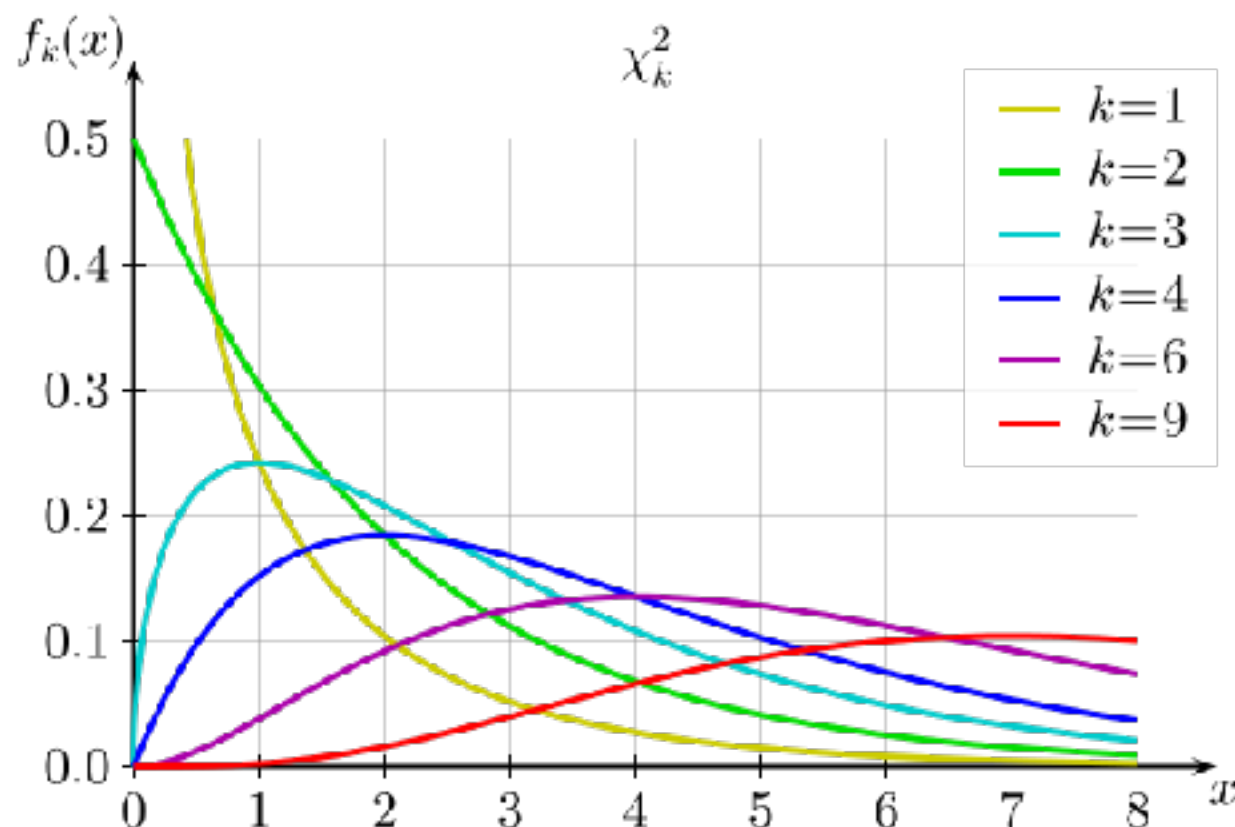
# ML-based Model Selection

---

LRT

Hypothesis test

Pairwise



Difference in  
free parameters

If the simpler model is true, twice the difference in log-likelihoods between the true and more complex model will follow a **Chi-squared** distribution with d.f. = the difference in complexity between the models.

Only for **nested models** (simple = restriction of complex)

# Bayesian Model Selection

---

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

# Bayesian Model Selection

---

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

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

# Bayesian Model Selection

---

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

$$P(\theta|D, M) = \frac{P(D|\theta, M)P(\theta|M)}{\int P(D|\theta, M)P(\theta|M)d\theta}$$

Marginal Likelihood

Probability of the data given the model, considering uncertainty in model parameters.



# Bayesian Model Selection

---

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

$$P(\theta|D, M) = \frac{P(D|\theta, M)P(\theta|M)}{\int P(D|\theta, M)P(\theta|M)d\theta}$$

Marginal Likelihood

Essentially, the **weighted average likelihood**, weighted by the prior probability of different parameter values.

# Marginal Likelihood Example

Evolutionary Distance

Sp. A ————— Sp. B

Compare **JC** and **K80** models

**v**: edge length  
estimated in both models

**k**: transition-transversion ratio  
estimated in K80 and fixed at 1 for JC

# Marginal Likelihood Example

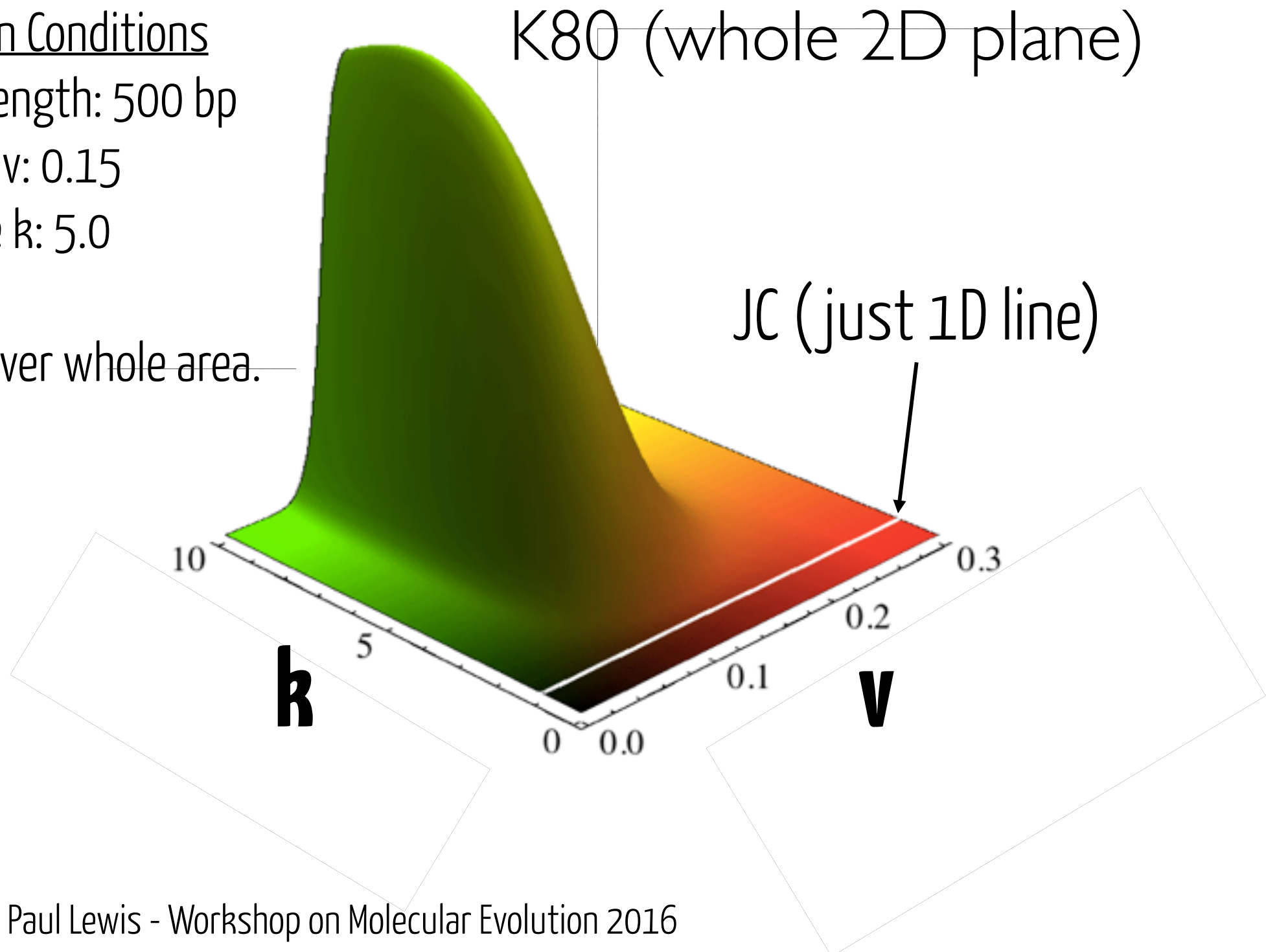
## Simulation Conditions

Sequence length: 500 bp

True  $v$ : 0.15

True  $k$ : 5.0

Prior is flat over whole area.



# Marginal Likelihood Example

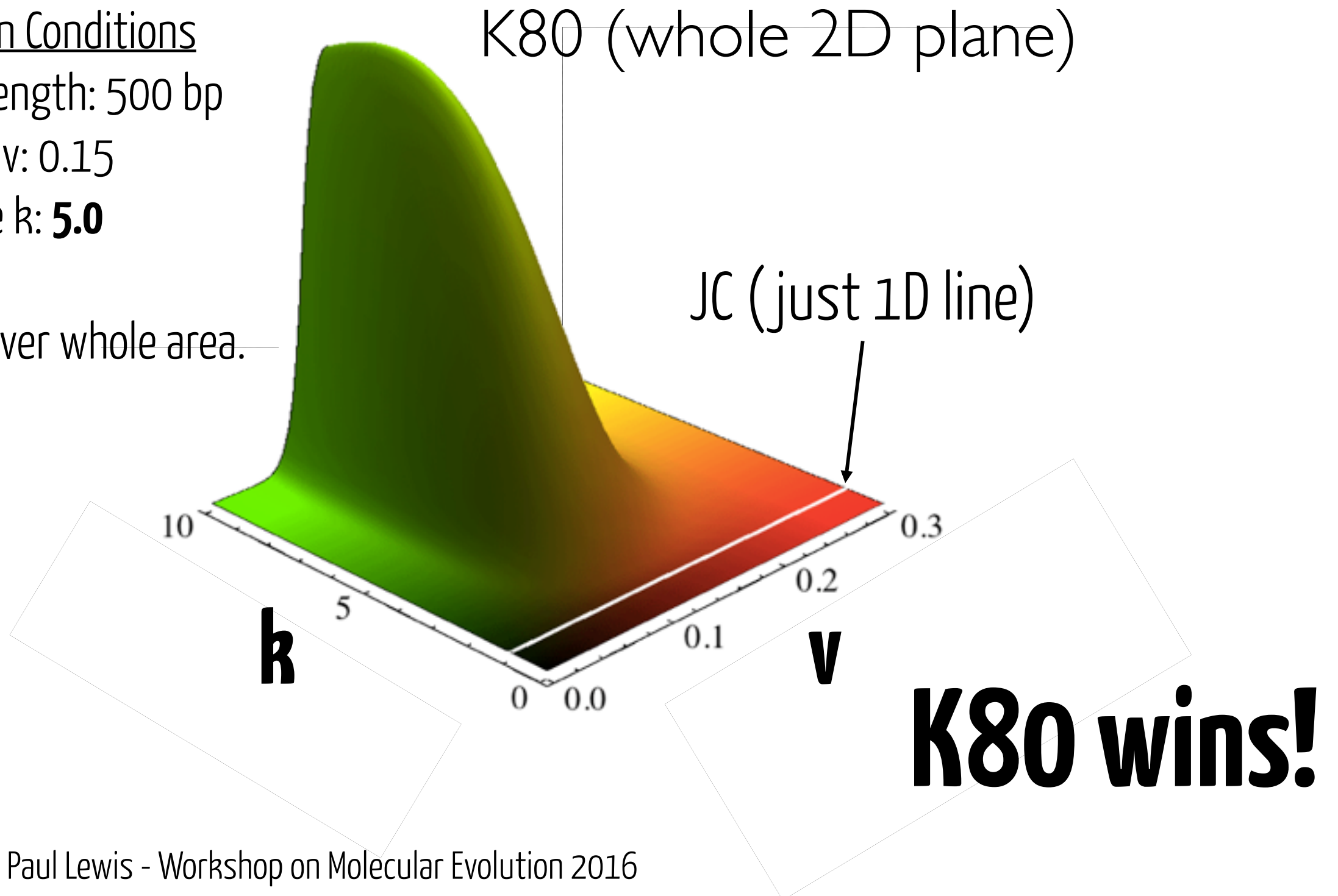
## Simulation Conditions

Sequence length: 500 bp

True  $v$ : 0.15

True  $k$ : **5.0**

Prior is flat over whole area.



# Marginal Likelihood Example

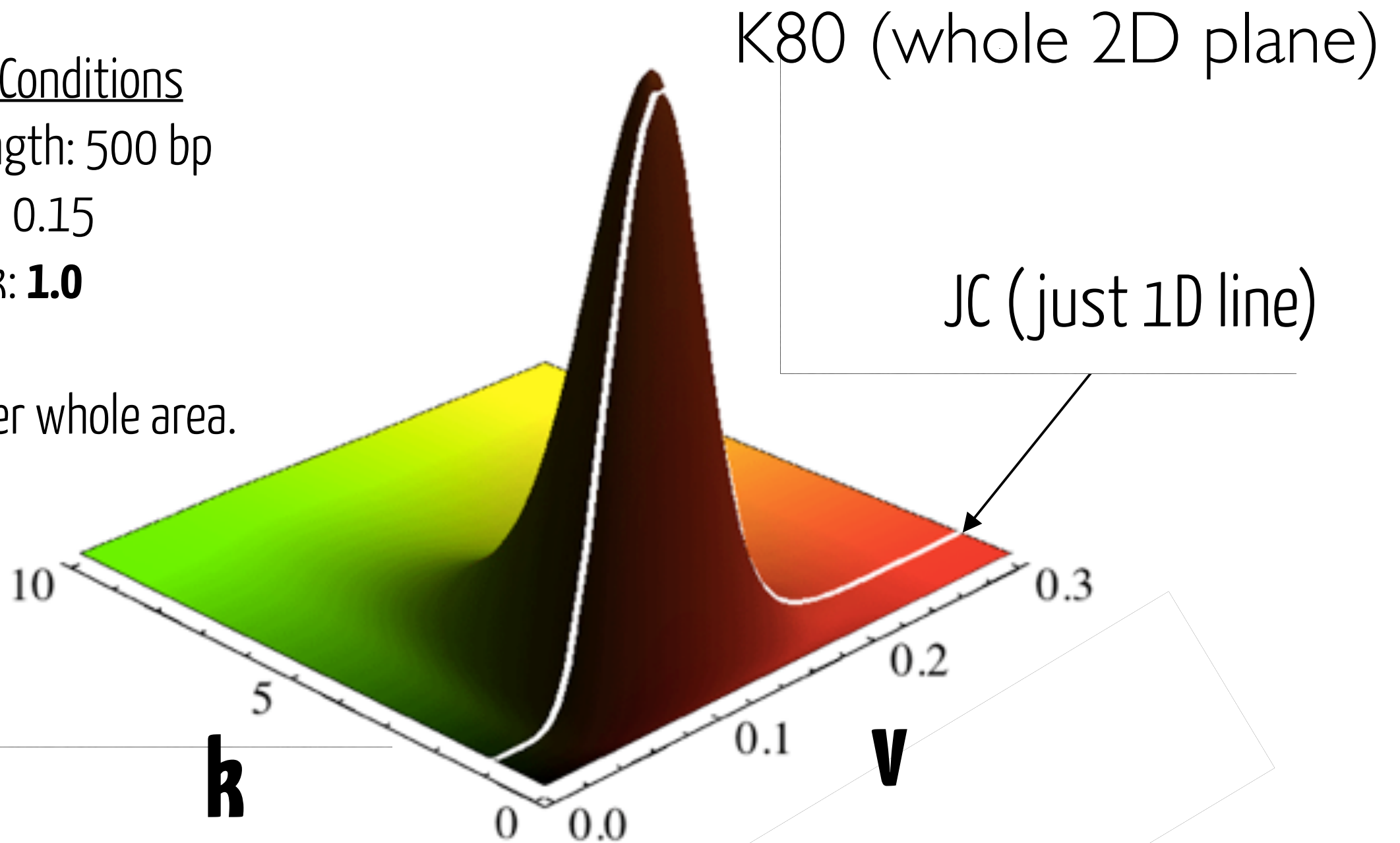
## Simulation Conditions

Sequence length: 500 bp

True  $v$ : 0.15

True  $k$ : **1.0**

Prior is flat over whole area.



# Marginal Likelihood Example

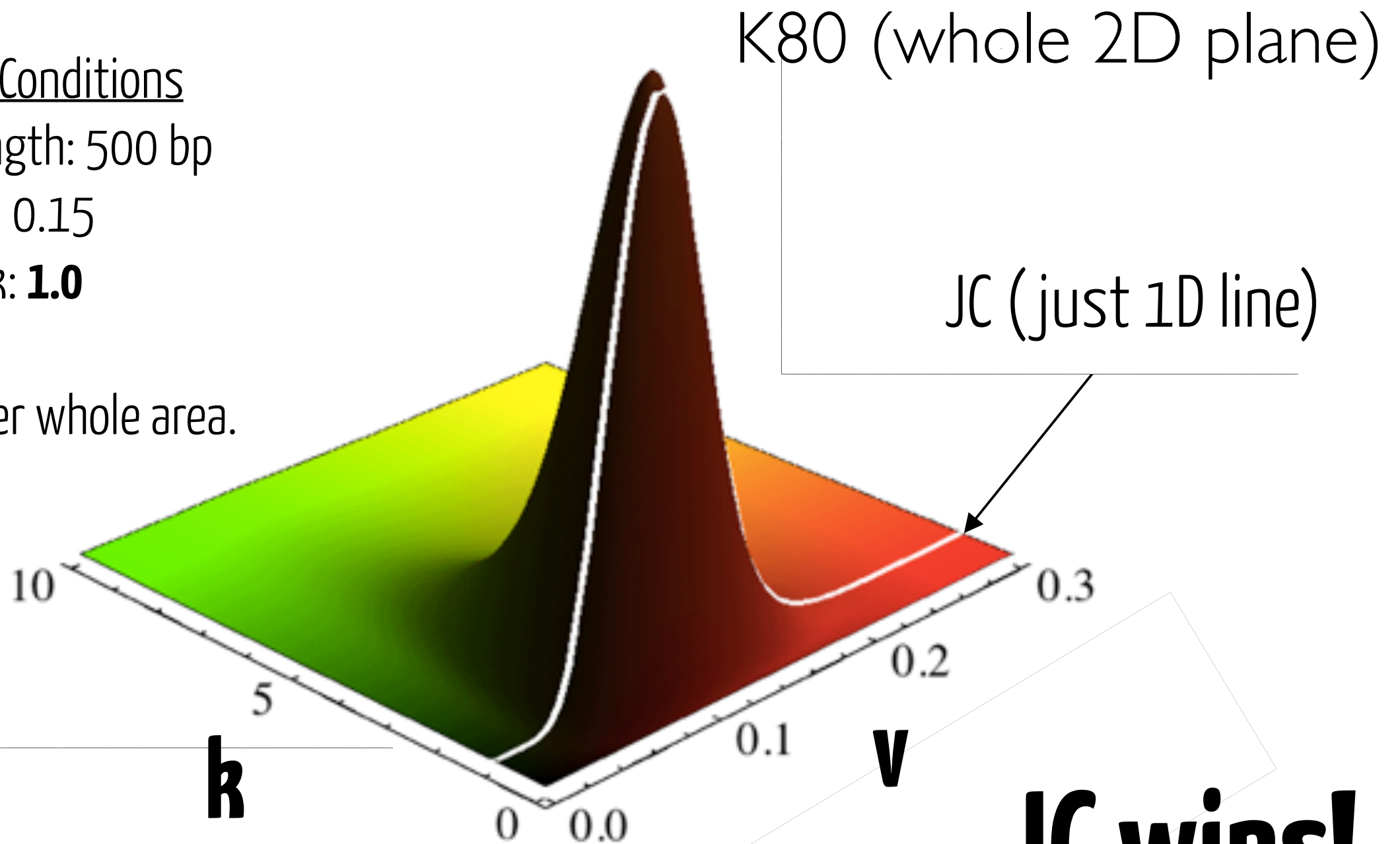
## Simulation Conditions

Sequence length: 500 bp

True  $v$ : 0.15

True  $k$ : **1.0**

Prior is flat over whole area.



**JC wins!**

# Marginal Likelihood Example

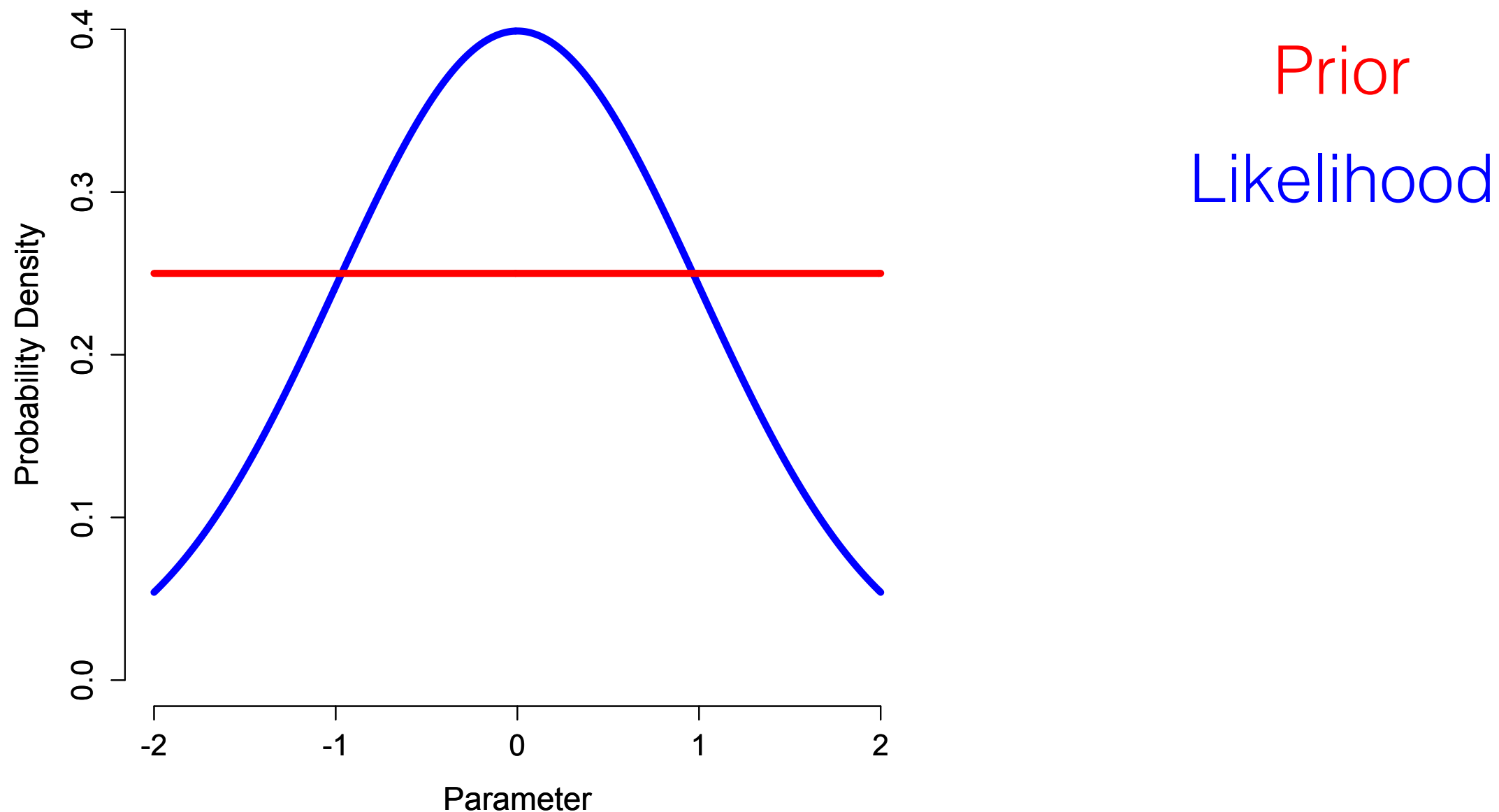
Important contrast with ML-based model selection: by averaging, rather than maximizing, marginal likelihoods automatically account for extra parameters.

**More complicated models can have lower marginal likelihoods.**

# Calculating Marginal Likelihoods

---

## Easy Approach 1 - Sample from the prior

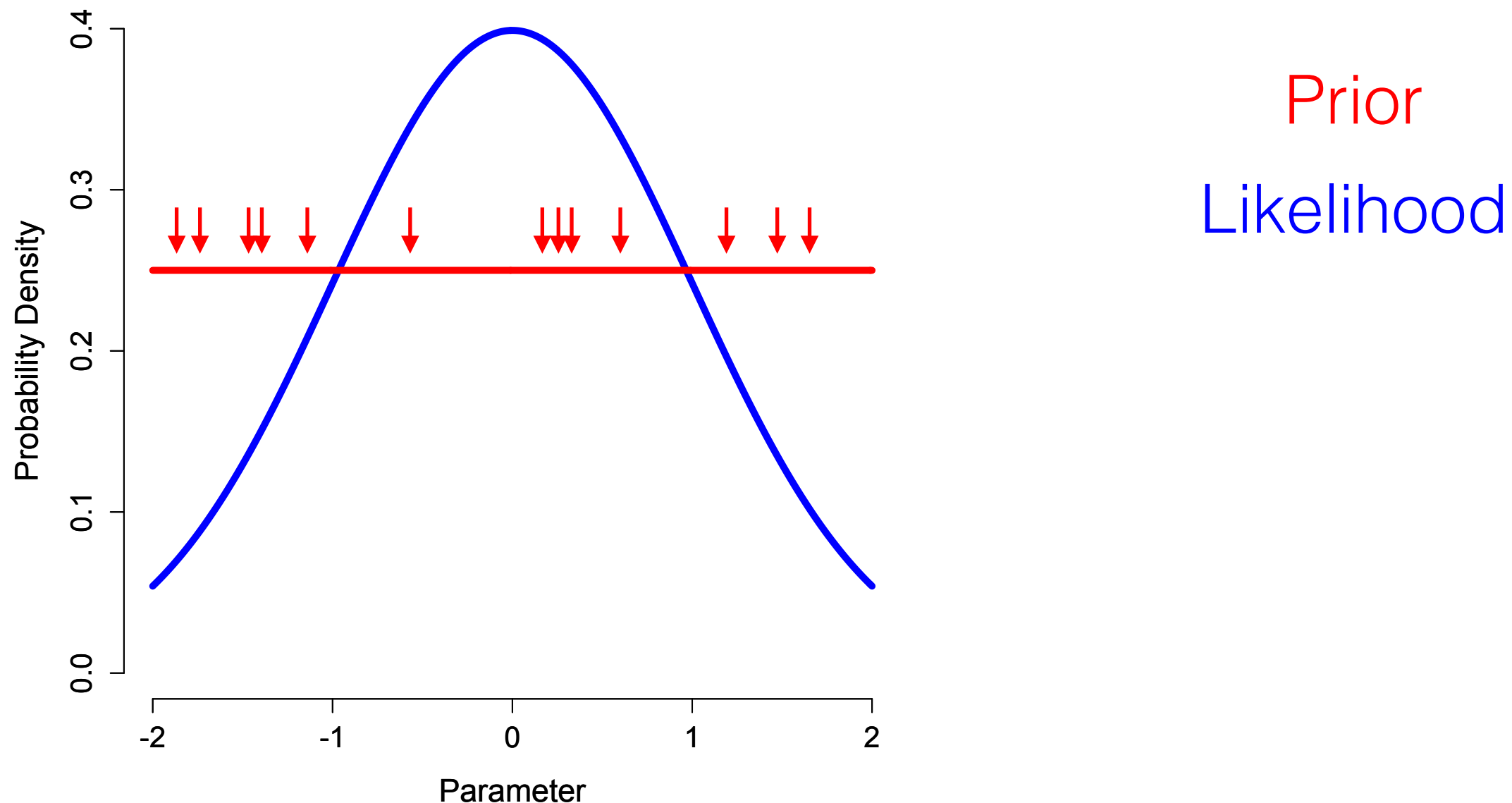




# Calculating Marginal Likelihoods

---

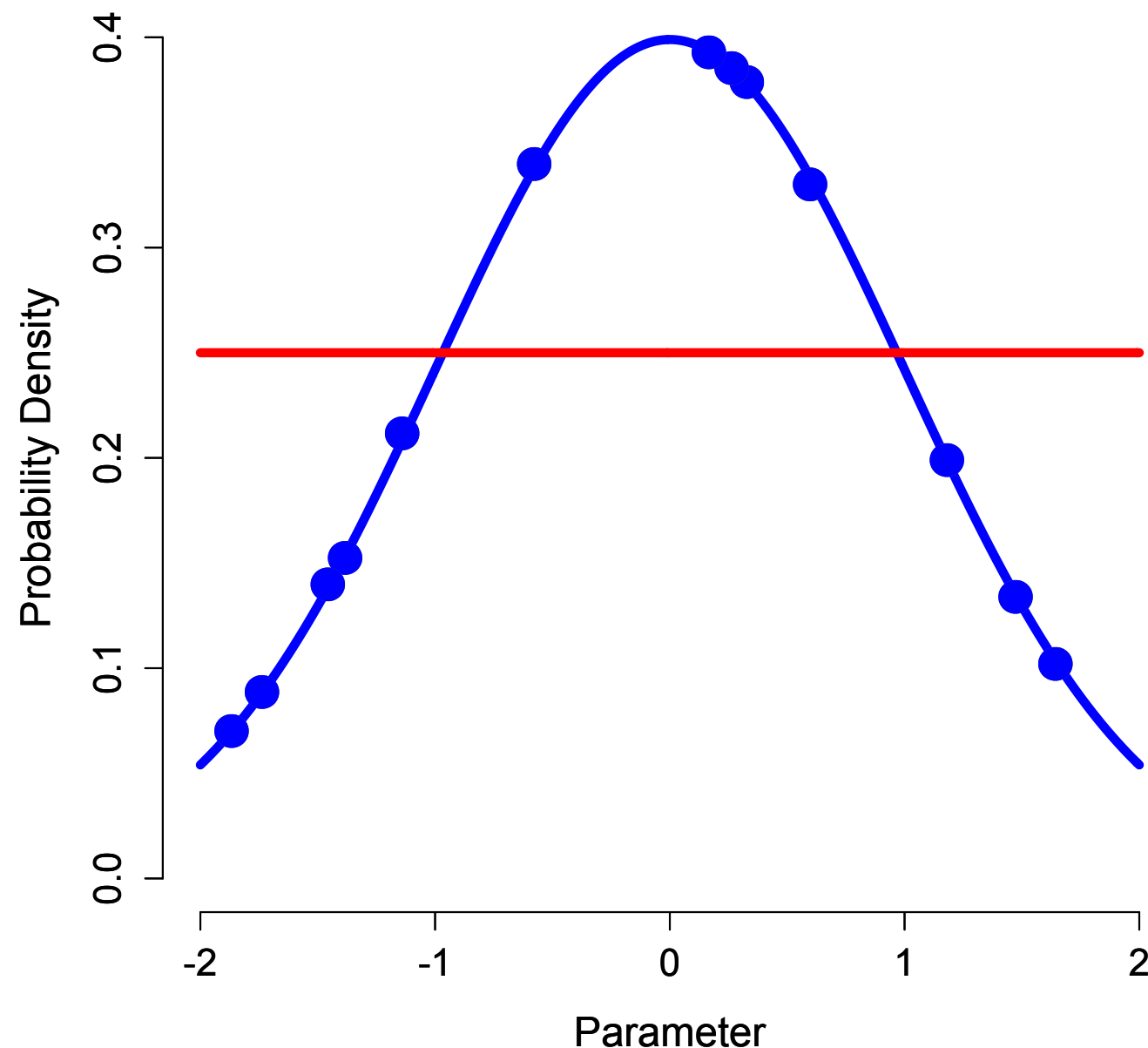
## Easy Approach 1 - Sample from the prior



# Calculating Marginal Likelihoods

---

## Easy Approach 1 - Sample from the prior



Prior

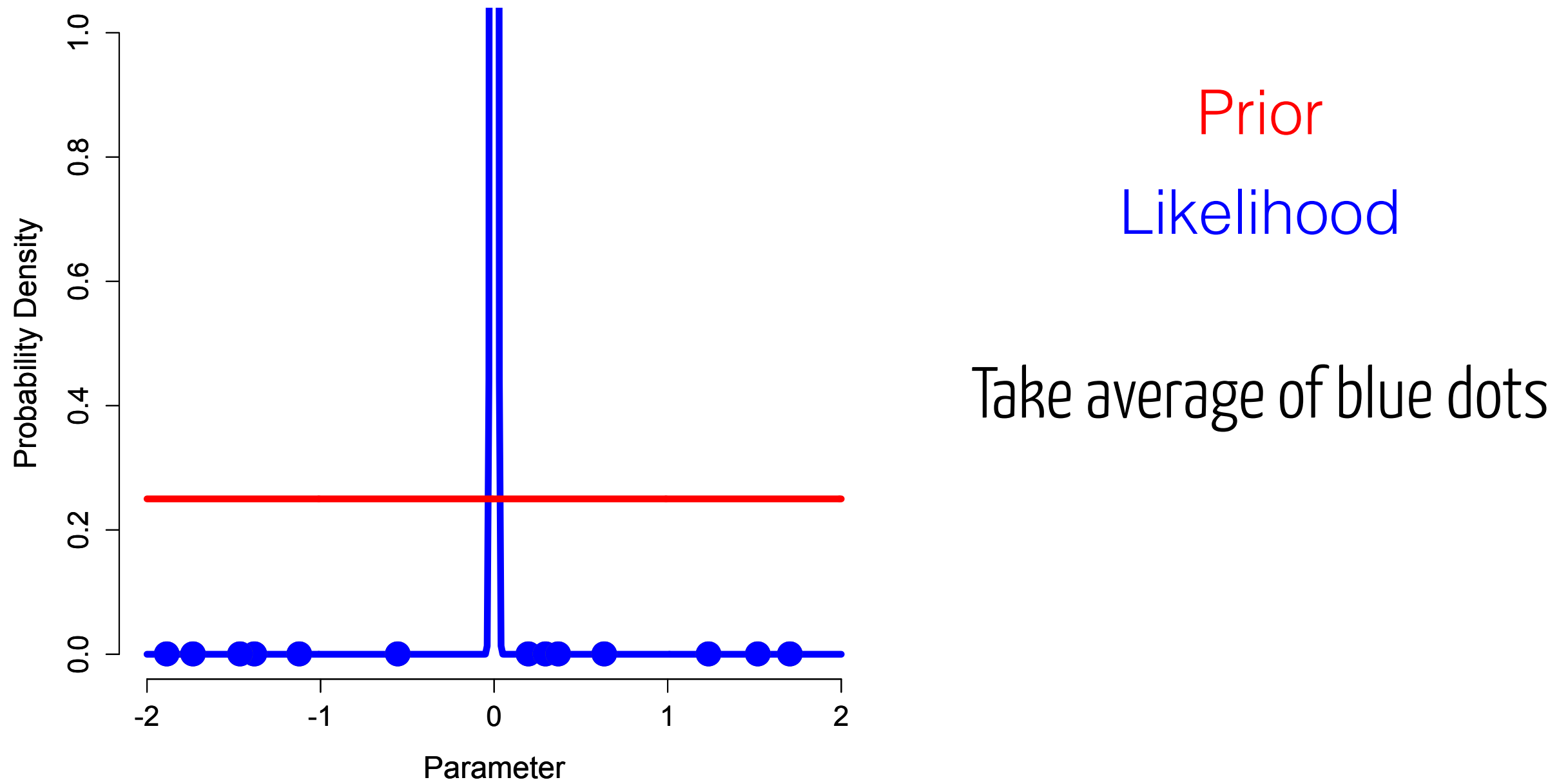
Likelihood

Take average of blue dots

# Calculating Marginal Likelihoods

---

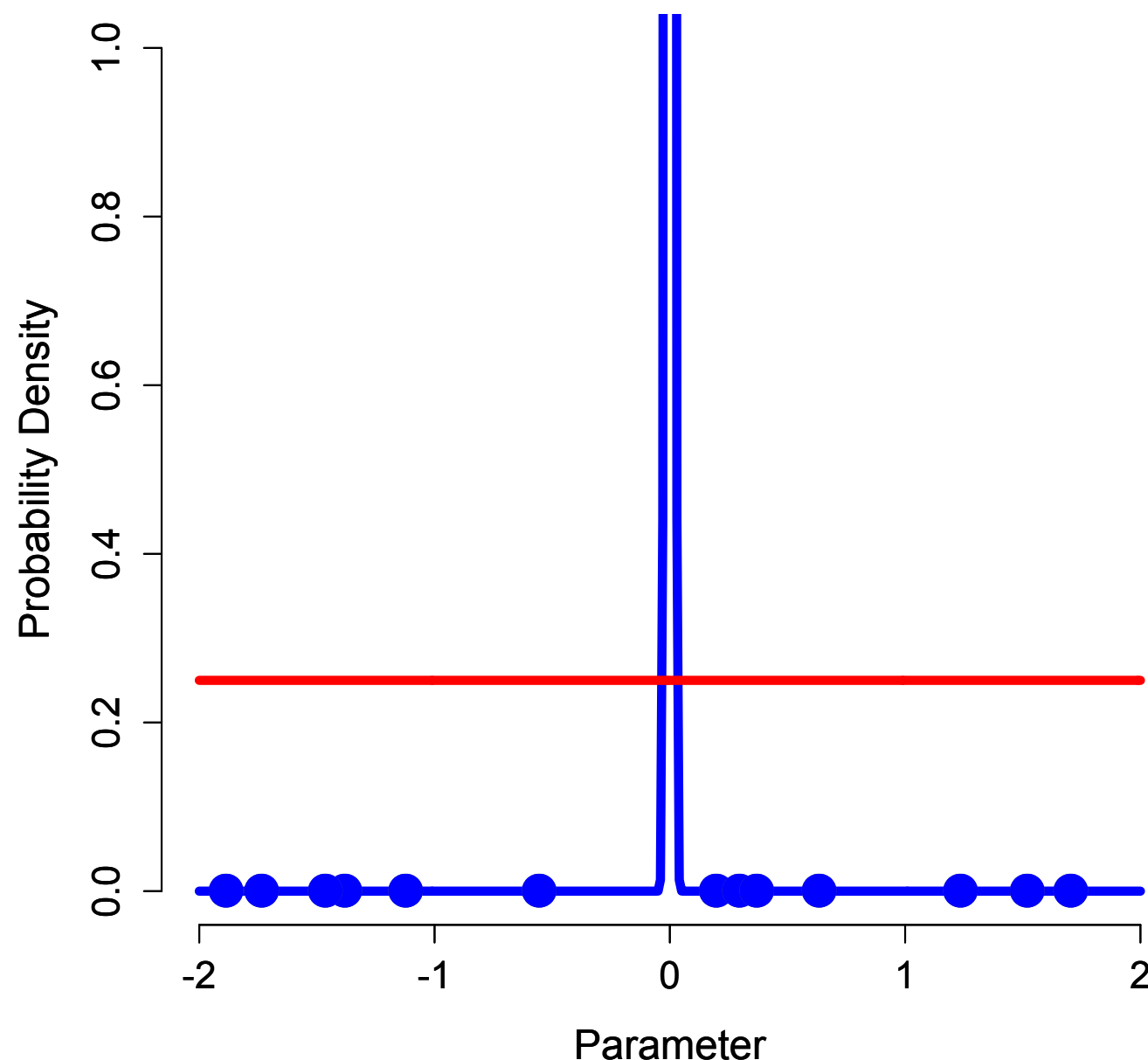
## Easy Approach 1 - Sample from the prior



# Calculating Marginal Likelihoods

---

## Easy Approach 1 - Sample from the prior



Prior

Likelihood

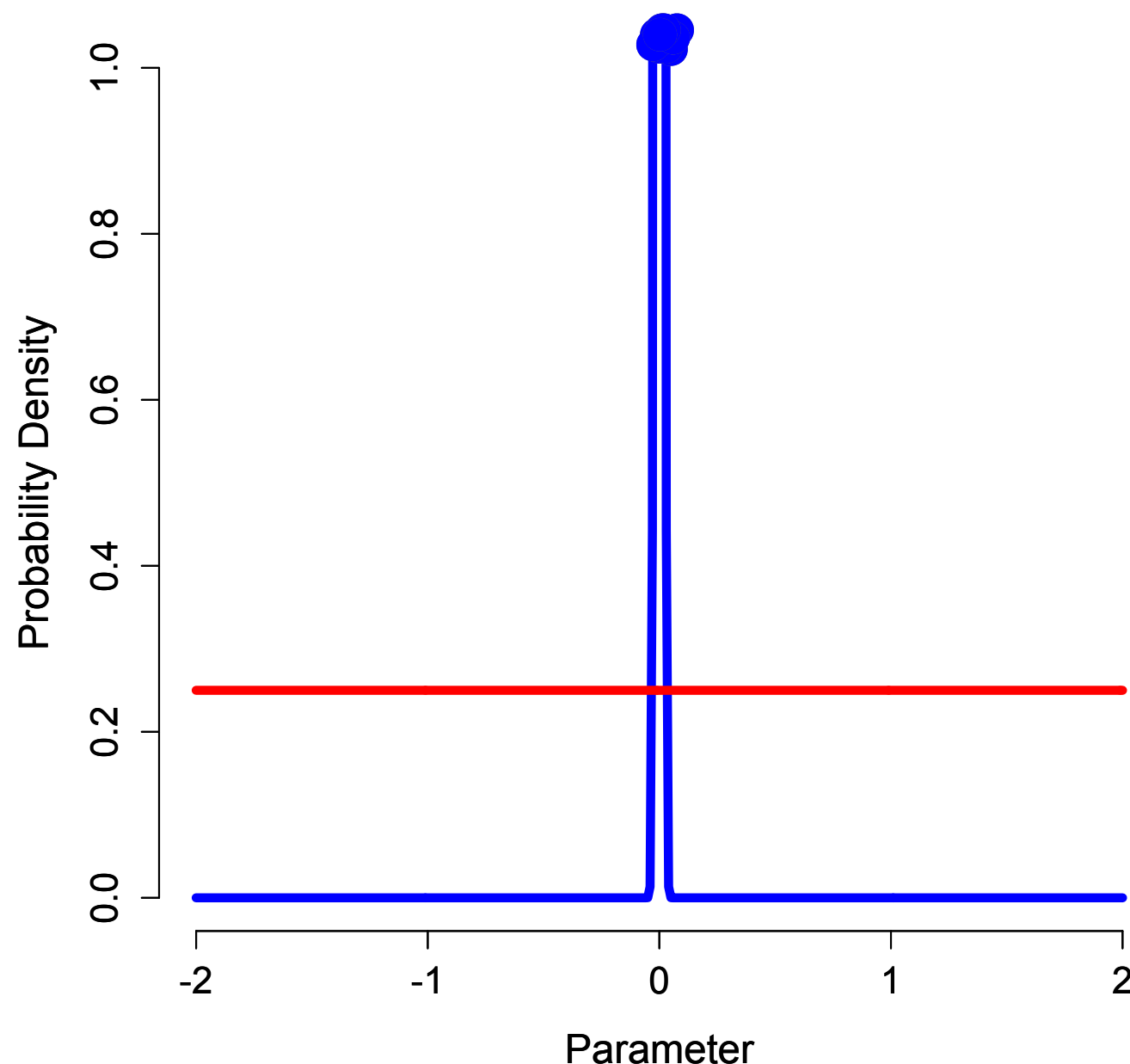
Take average of blue dots??!

We'd like to make sure we're sampling high likelihood parts of space with reasonable frequency.

# Calculating Marginal Likelihoods

---

## Less-Naive Approach 2- Sample from the posterior



Prior

Posterior  
(~Likelihood)

Since we're supposed to be integrating across the prior, we need to correct for the fact that our samples are from the posterior.

# Calculating Marginal Likelihoods

Less-Naive Approach 2- Sample from the posterior

The Harmonic Mean Method

$$\frac{1}{ML} = \frac{1}{L_1} + \frac{1}{L_2} + \dots + \frac{1}{L_3}$$

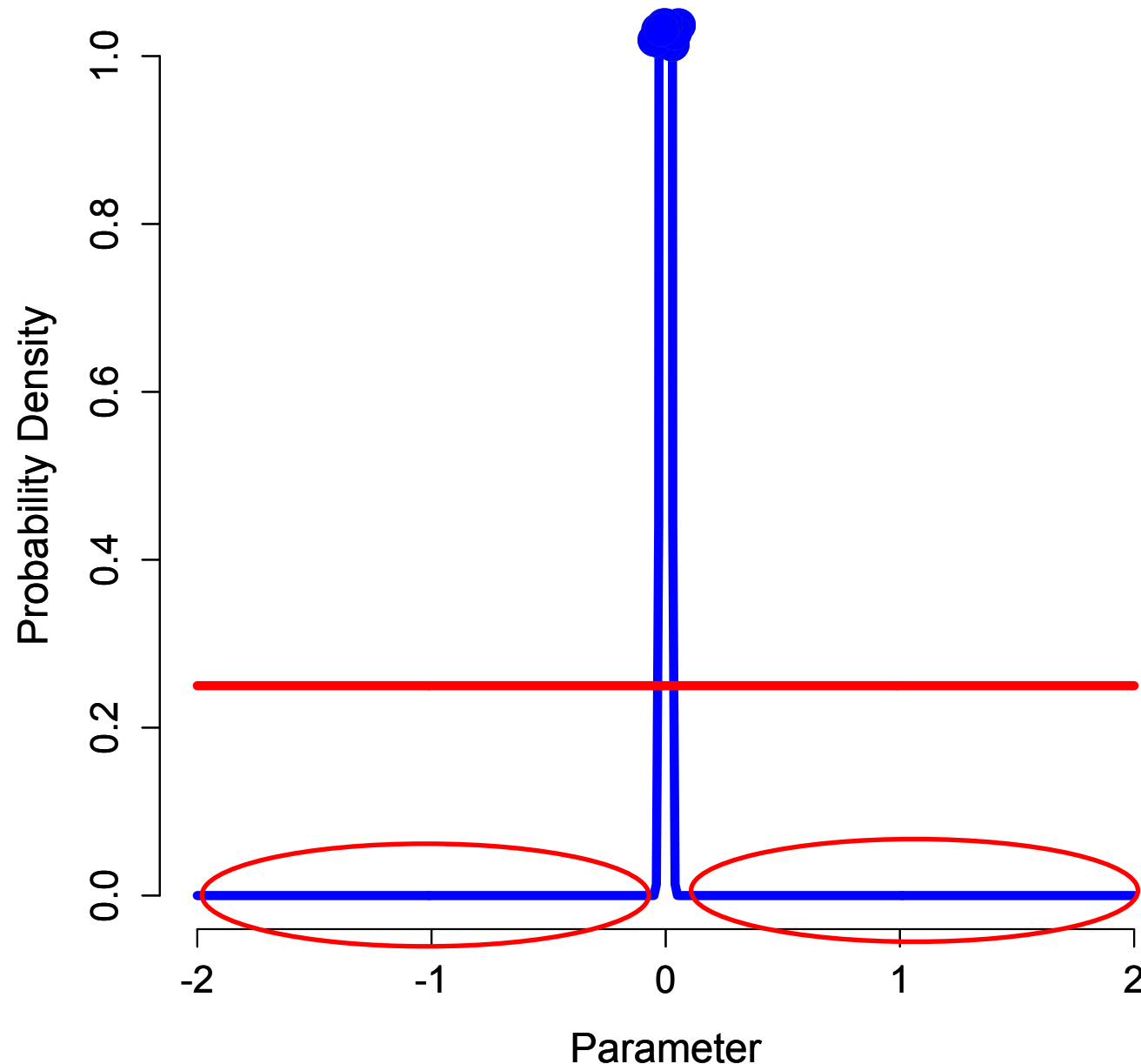
What's an important property of harmonic means?

Anyone remember discussing bottlenecks in pop gen?

# Calculating Marginal Likelihoods

---

## Less-Naive Approach 2- Sample from the posterior



The reverse problem to our first naive approach!

Rarely sampled low likelihoods have a big influence on estimates.

Very unstable.

# Calculating Marginal Likelihoods

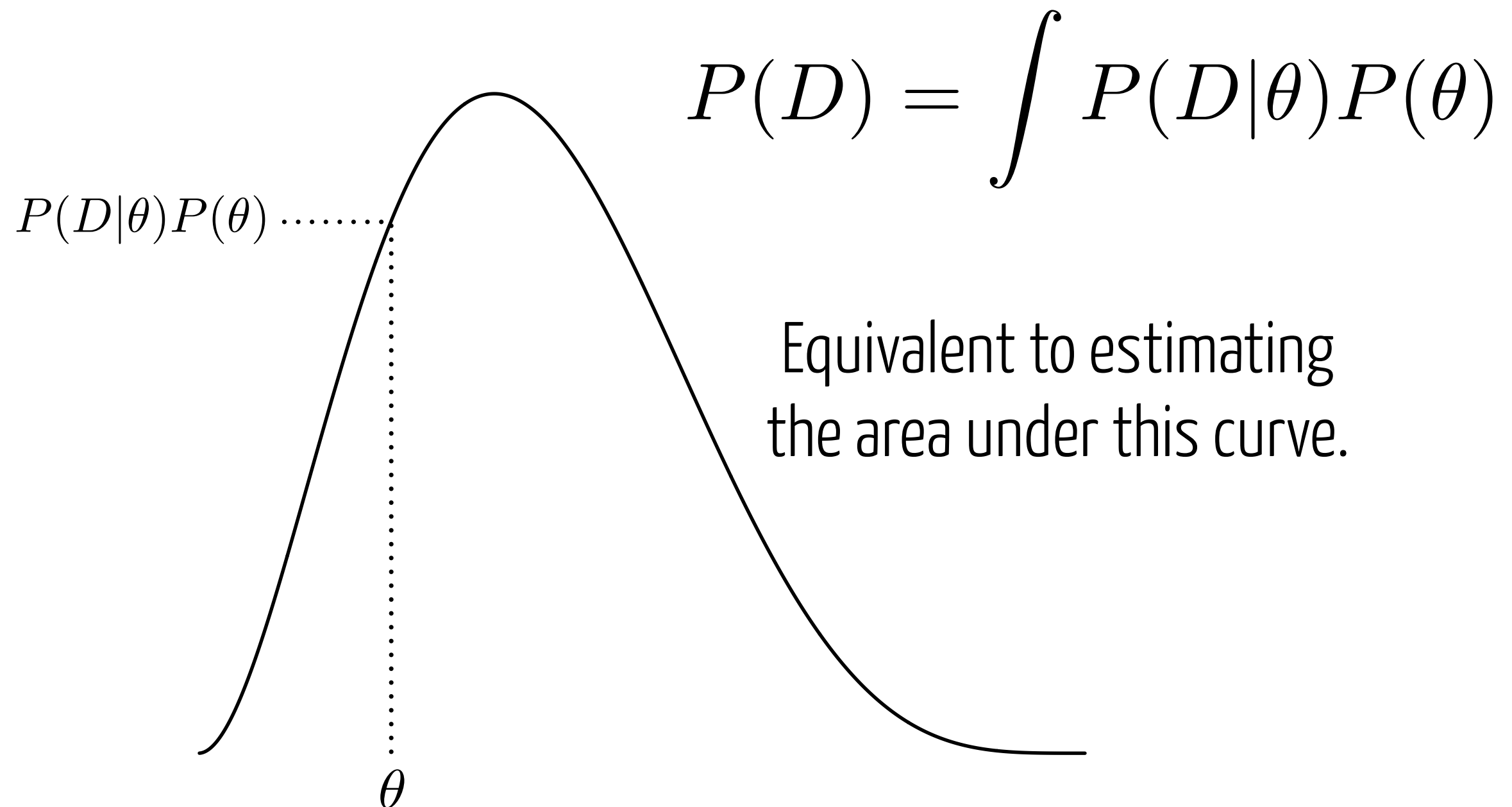
Approach 3\* - Sample from a series of distributions

Steppingstone or path sampling

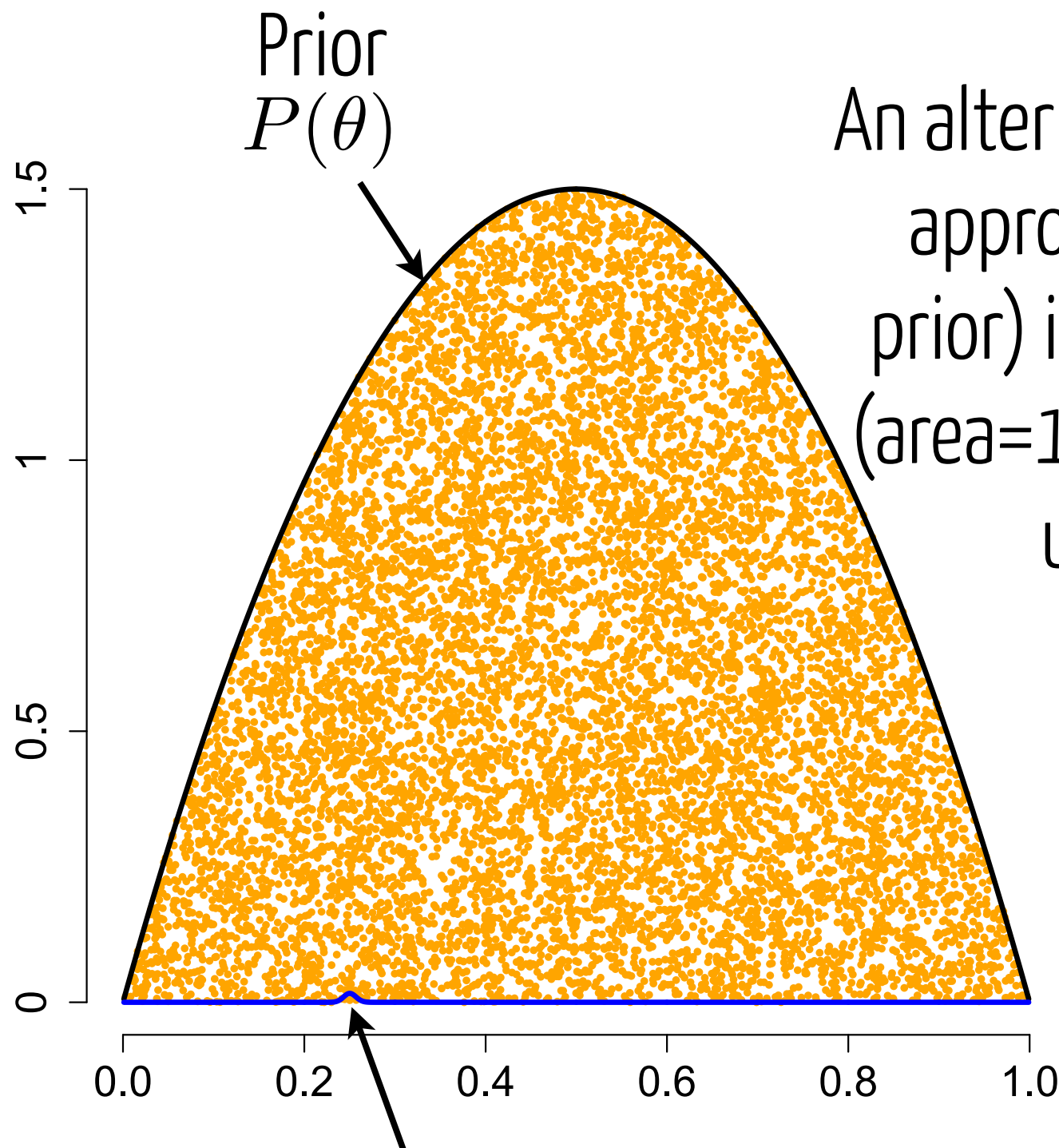


# Steppingstone Sampling

---



# Steppingstone Sampling



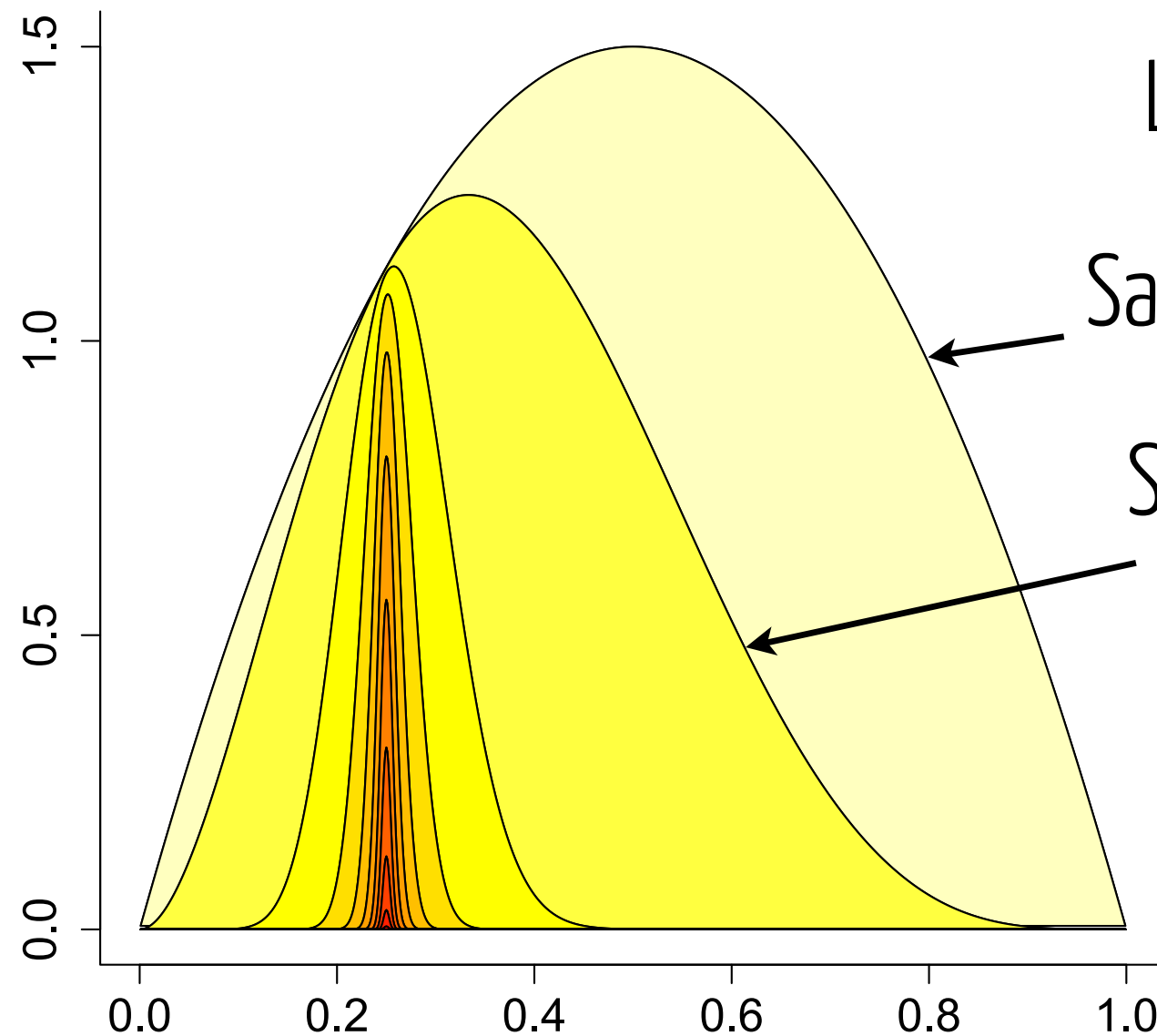
An alternative way to think about our first approach (sampling directly from the prior) is to sample points from the prior (area=1.0), then ask what proportion fall under the curve of interest.

**Unfortunately, not many!**  
**As before, this is unstable.**

Unnormalized Posterior  $\longrightarrow P(D|\theta)P(\theta)$

# Steppingstone Sampling

$P(D)$



Let's try it in **steps!**

Sample from this distribution

See what fraction of samples are under this curve

That fraction is an estimate of this ratio

$$\frac{c_{1.0}}{c_{0.0}} = \left( \frac{c_{1.0}}{c_{0.9}} \right) \left( \frac{c_{0.9}}{c_{0.8}} \right) \left( \frac{c_{0.8}}{c_{0.7}} \right) \left( \frac{c_{0.7}}{c_{0.6}} \right) \left( \frac{c_{0.6}}{c_{0.5}} \right) \left( \frac{c_{0.5}}{c_{0.4}} \right) \left( \frac{c_{0.4}}{c_{0.3}} \right) \left( \frac{c_{0.3}}{c_{0.2}} \right) \left( \frac{c_{0.2}}{c_{0.1}} \right) \left( \frac{c_{0.1}}{c_{0.0}} \right)$$

# Power Posteriors

---

$$\begin{array}{ccc}
 P(D) \swarrow & & \\
 \frac{c_{1.0}}{c_{0.0}} = \left( \frac{c_{1.0}}{\cancel{c_{0.9}}} \right) \left( \frac{\cancel{c_{0.9}}}{\cancel{c_{0.8}}} \right) \left( \frac{\cancel{c_{0.8}}}{\cancel{c_{0.7}}} \right) \left( \frac{\cancel{c_{0.7}}}{\cancel{c_{0.6}}} \right) \left( \frac{\cancel{c_{0.6}}}{\cancel{c_{0.5}}} \right) \left( \frac{\cancel{c_{0.5}}}{\cancel{c_{0.4}}} \right) \left( \frac{\cancel{c_{0.4}}}{\cancel{c_{0.3}}} \right) \left( \frac{\cancel{c_{0.3}}}{\cancel{c_{0.2}}} \right) \left( \frac{\cancel{c_{0.2}}}{\cancel{c_{0.1}}} \right) \left( \frac{\cancel{c_{0.1}}}{c_{0.0}} \right) & & \\
 1 \nearrow & & \\
 \text{Posterior} & & \text{Prior} \\
 \beta = 1 & & \beta = 0
 \end{array}$$

$$P(D|\theta)_{\beta} \propto P(D|\theta)^{\beta} P(\theta)$$

# Power Posteriors

---

$$\frac{c_{1.0}}{c_{0.0}} = \text{Stable estimate of marginal likelihood!}$$



But it requires a **specific type of analysis**,  
independent of standard MCMC.

# Bayes Factors

---

The diagram illustrates Bayes' theorem with color-coded components and arrows:

- Posterior** (red text) points to  $P(H|D)$  with a red arrow.
- Likelihood** (blue text) points to  $P(D|H)$  with a blue arrow.
- Prior** (green text) points to  $P(H)$  with a green arrow.
- Normalizing Constant (Marginal Likelihood)** (orange text) points to  $P(D)$  with an orange arrow.

$$P(H|D) = \frac{P(D|H)P(H)}{P(D)}$$

# Bayes Factors

---

$$\frac{P(H_1|D)}{P(H_2|D)} = \frac{\frac{P(H_1)P(D|H_1)}{P(D)}}{\frac{P(H_2)P(D|H_2)}{P(D)}}$$

# Bayes Factors

---

$$\frac{P(H_1|D)}{P(H_2|D)} = \frac{\frac{P(H_1)P(D|H_1)}{\cancel{P(D)}}}{\frac{P(H_2)P(D|H_2)}{\cancel{P(D)}}}$$



# Bayes Factors

---

$$\frac{P(H_1|D)}{P(H_2|D)} = \frac{P(H_1)P(D|H_1)}{P(H_2)P(D|H_2)}$$

# Bayes Factors

---

Posterior Odds

$$\frac{P(H_1|D)}{P(H_2|D)}$$

Prior Odds

$$\frac{P(H_1)}{P(H_2)}$$

Bayes Factor

$$\frac{P(D|H_1)}{P(D|H_2)}$$

=

# Bayes Factors

---

Prior Odds

Bayes Factor

Posterior Odds

$$\frac{P(H_1)}{P(H_2)} \times \frac{P(D|H_1)}{P(D|H_2)} = \frac{P(H_1|D)}{P(H_2|D)}$$

# Bayes Factors

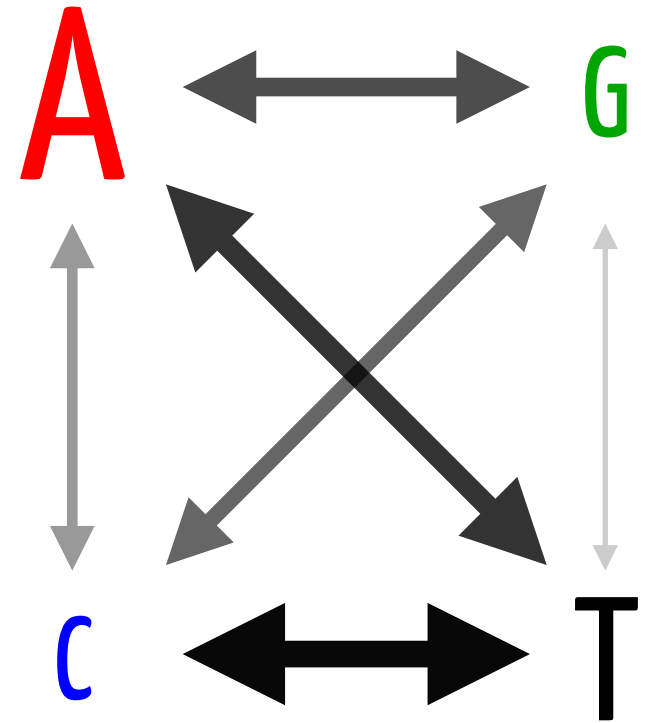
---

<b><math>2\ln(\text{BF})</math></b>	<b>BF</b>	<b>Strength of evidence</b>
0-2	1-3	Barely worth mentioning.
2-6	3-20	Positive
6-10	20-150	Strong
>10	>150	Very Strong

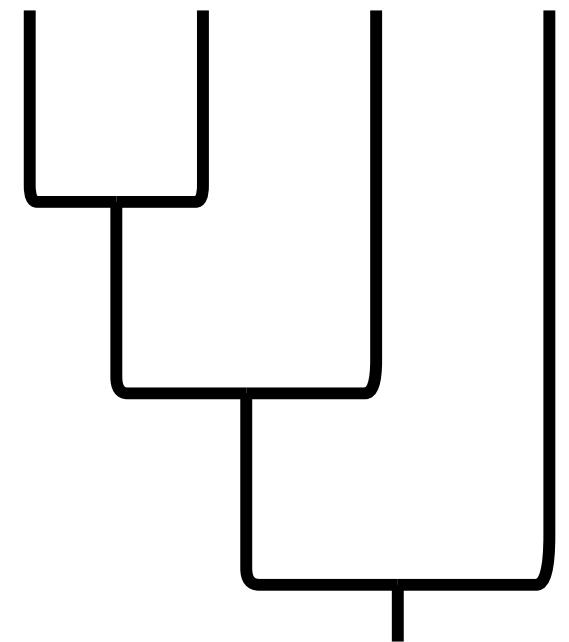
# Bayes Factors

---

For now, we're going to use these to **compare different models of sequence evolution** as our hypotheses.



However, BFs can also be used for other hypotheses, like **topological relationships**.



# Or...don't choose a model!

## Reversible Jump MCMC

Instead of picking a model, include MCMC moves that jump between. Integrate out uncertainty about which model is best.  
This is a Bayesian form of **model averaging**.

We already do this for trees. Can also do this for models.

# Or...don't choose a model!

## Reversible Jump MCMC

Instead of picking a model, include MCMC moves that jump between. Integrate out uncertainty about which model is best.  
This is a Bayesian form of **model averaging**.

We already do this for trees. Can also do this for models.

\*Disclaimer: Setting up proper reversible jump moves can often be **very challenging**.