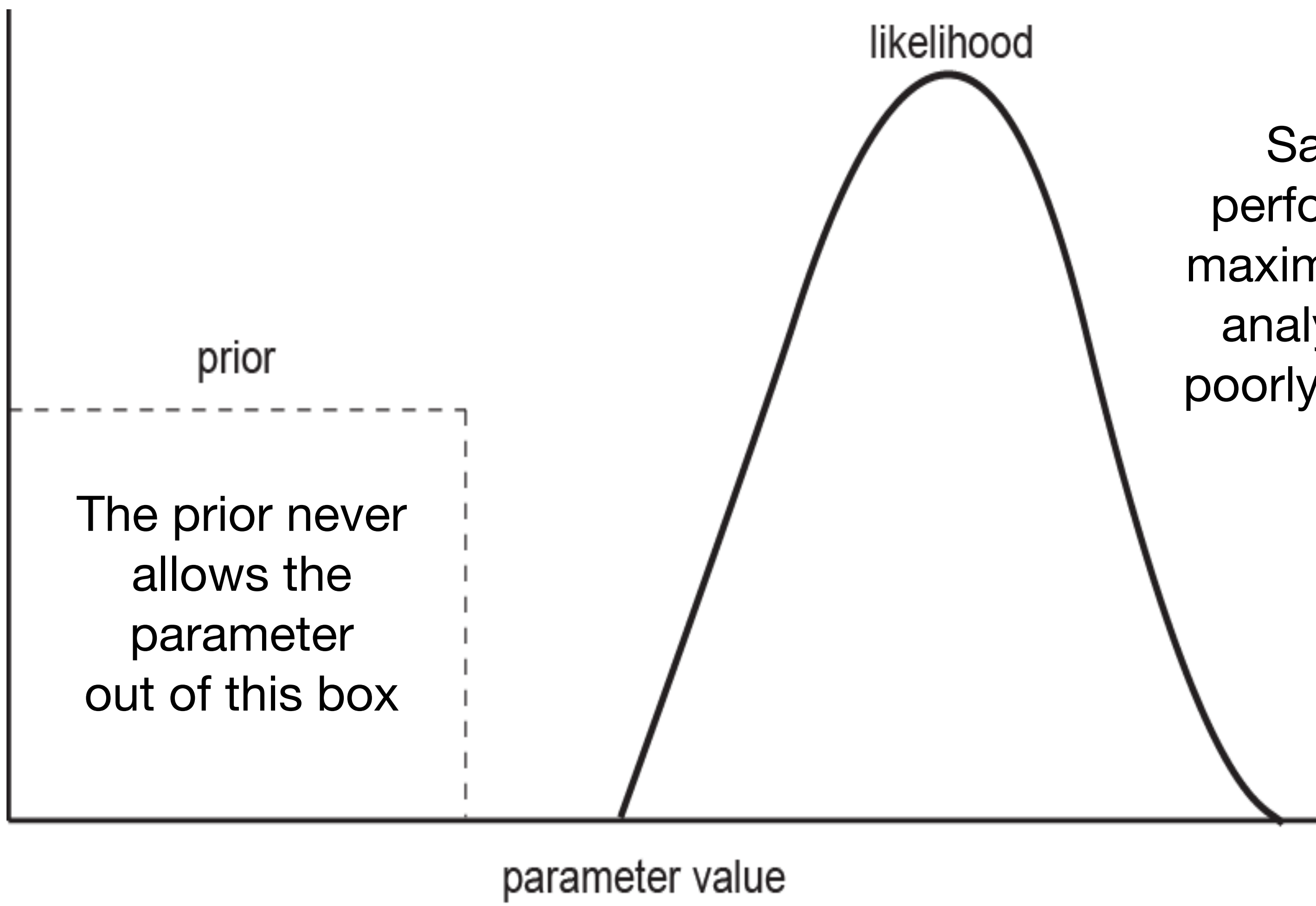
The choice of prior distributions can potentially turn a good model bad!



Same model performs well in a maximum likelihood analysis but very poorly in a Bayesian analysis

Bayes' rule

Likelihood

Prior probability
density

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

Posterior probability
density

Marginal probability
of the data

Marginal Likelihood

$$p(D) = \int_\theta p(D|\theta)\ p(\theta)\ d\theta$$

condition on model

Marginal
Likelihood

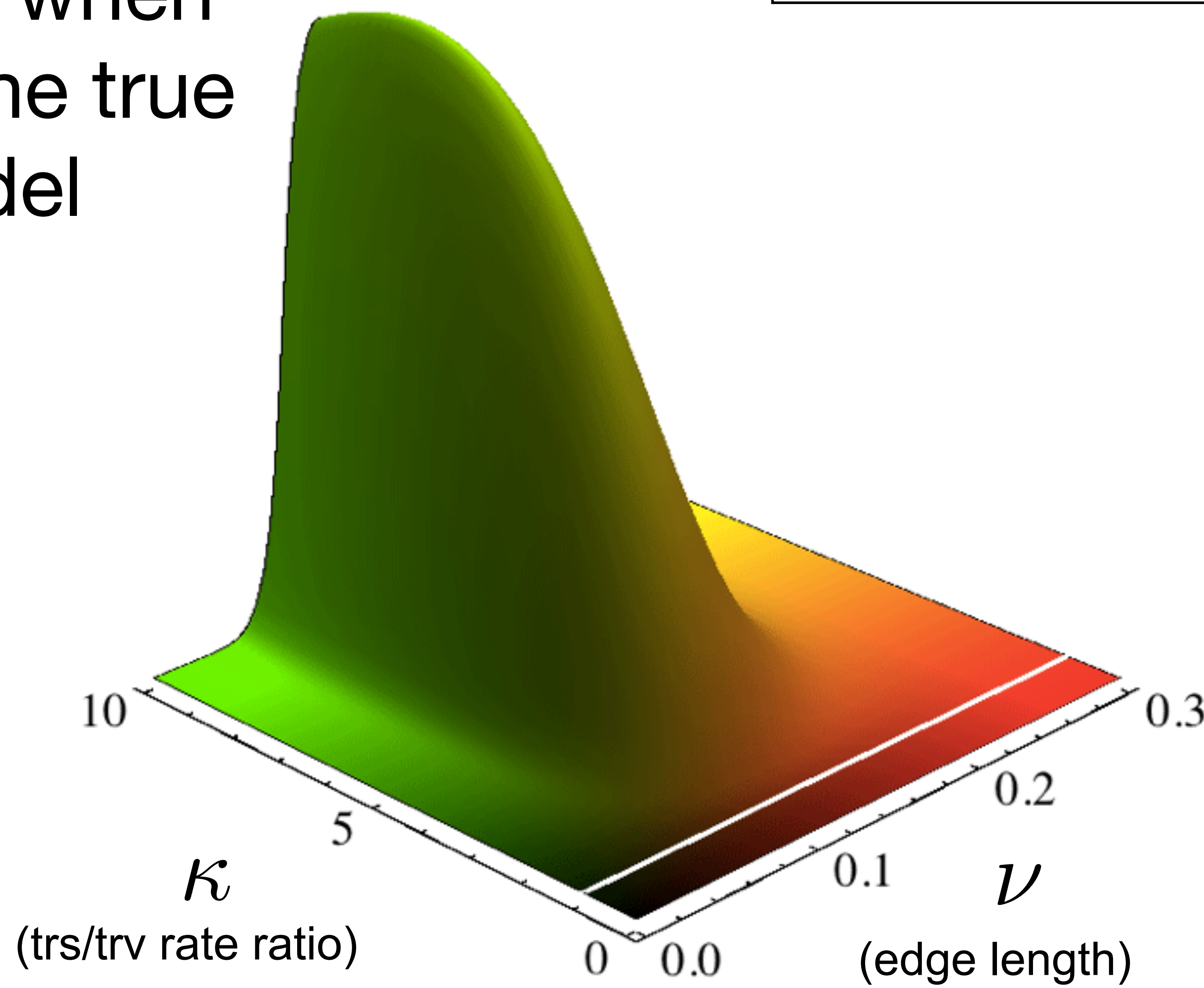
$$p(D) = \int_\theta p(D|\theta)\ p(\theta)\ d\theta$$

We always condition on model used
(but this is often not made explicit in notation used)

$$p(D|M) = \int_\theta p(D|\theta, M)\ p(\theta|M)\ d\theta$$

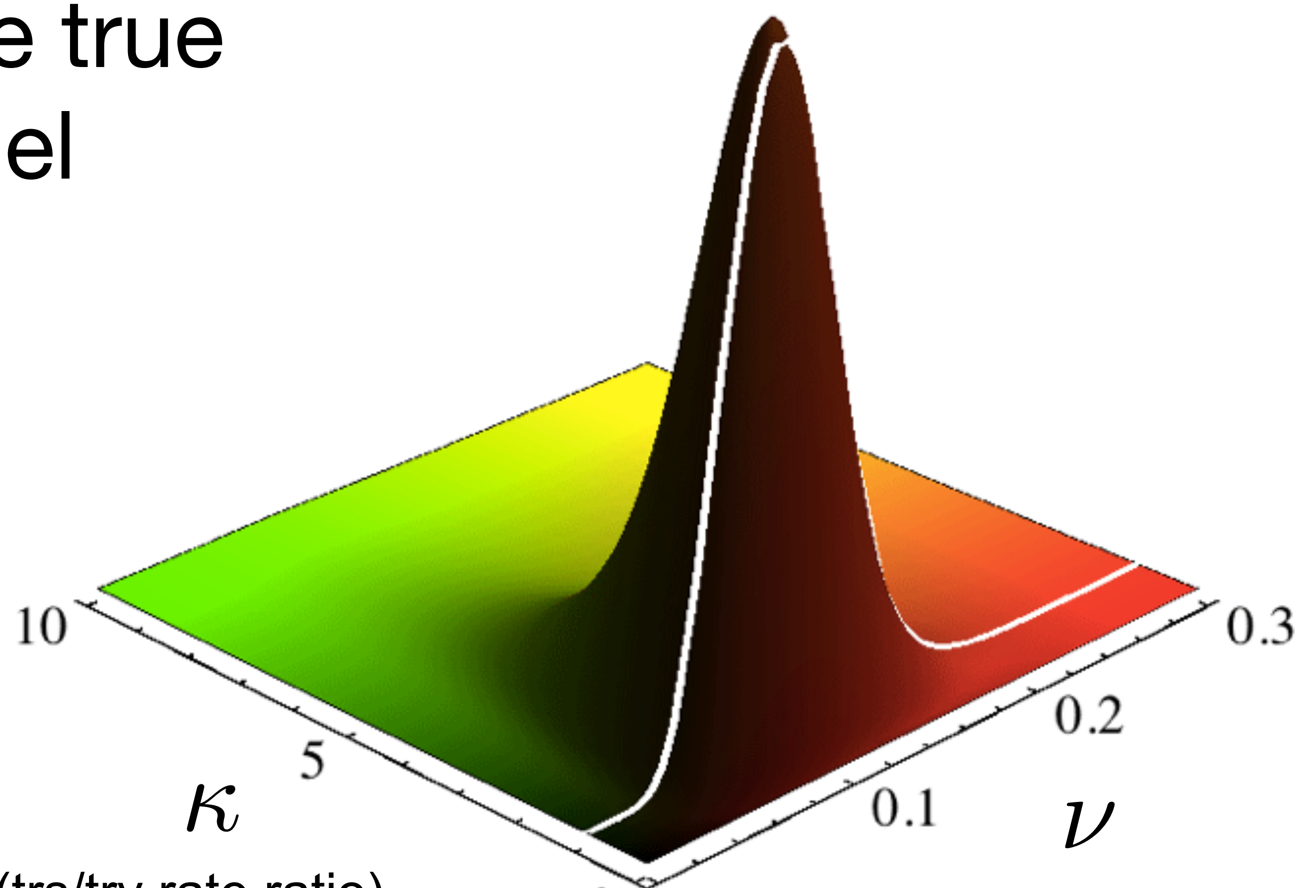
Likelihood surface when K80 is the true model

sequence length	= 500 sites
true edge length	= 0.15
true kappa	= **5.0**



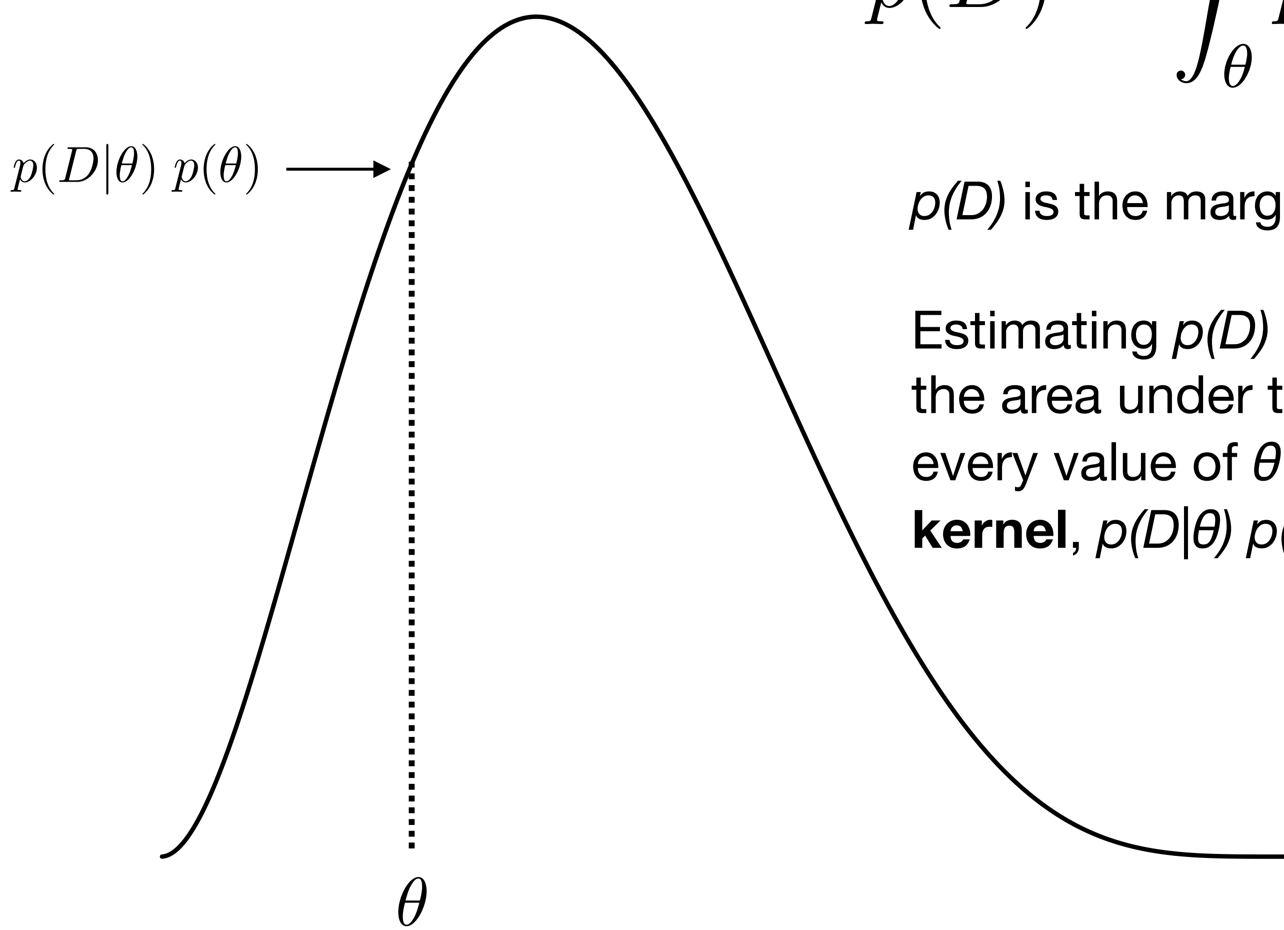
Likelihood surface when JC is the true model

sequence length	= 500 sites
true edge length	= 0.15
true kappa	= **1.0**



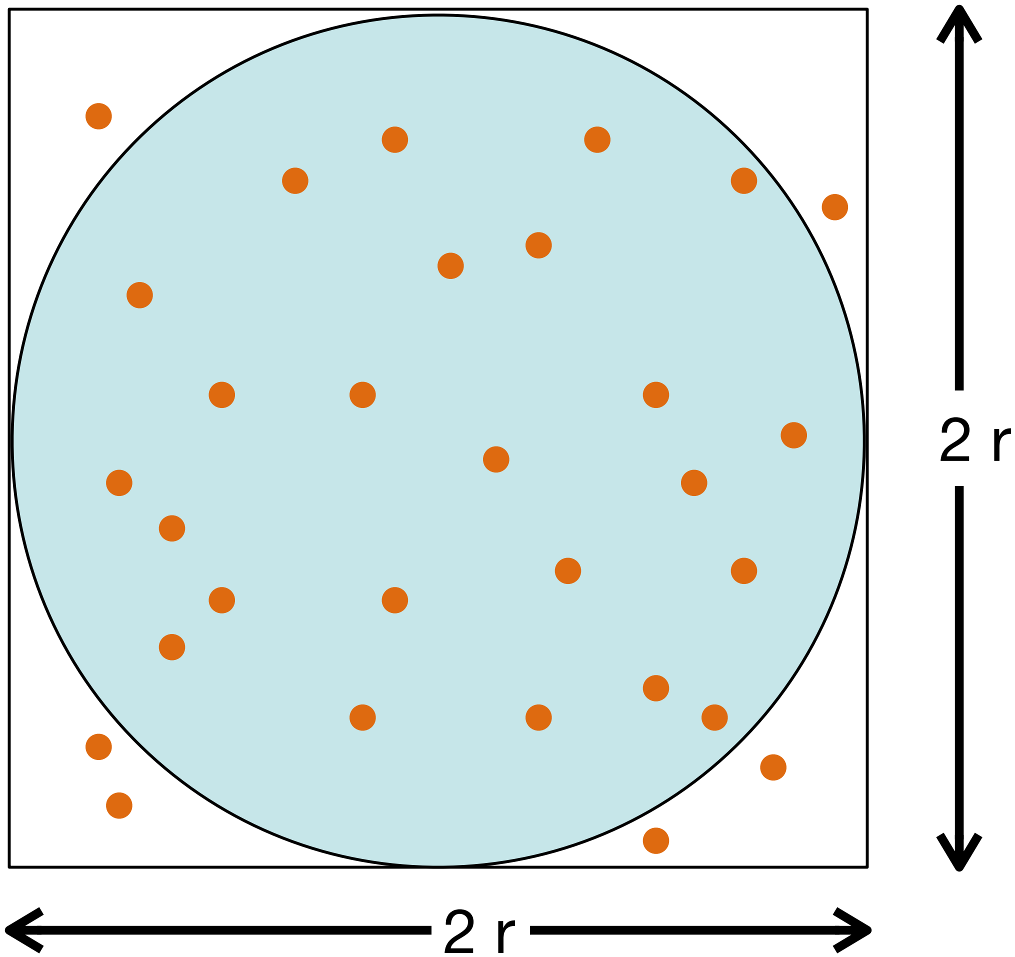
Estimating the marginal likelihood

$$p(D) = \int_\theta p(D|\theta)\ p(\theta)\ d\theta$$

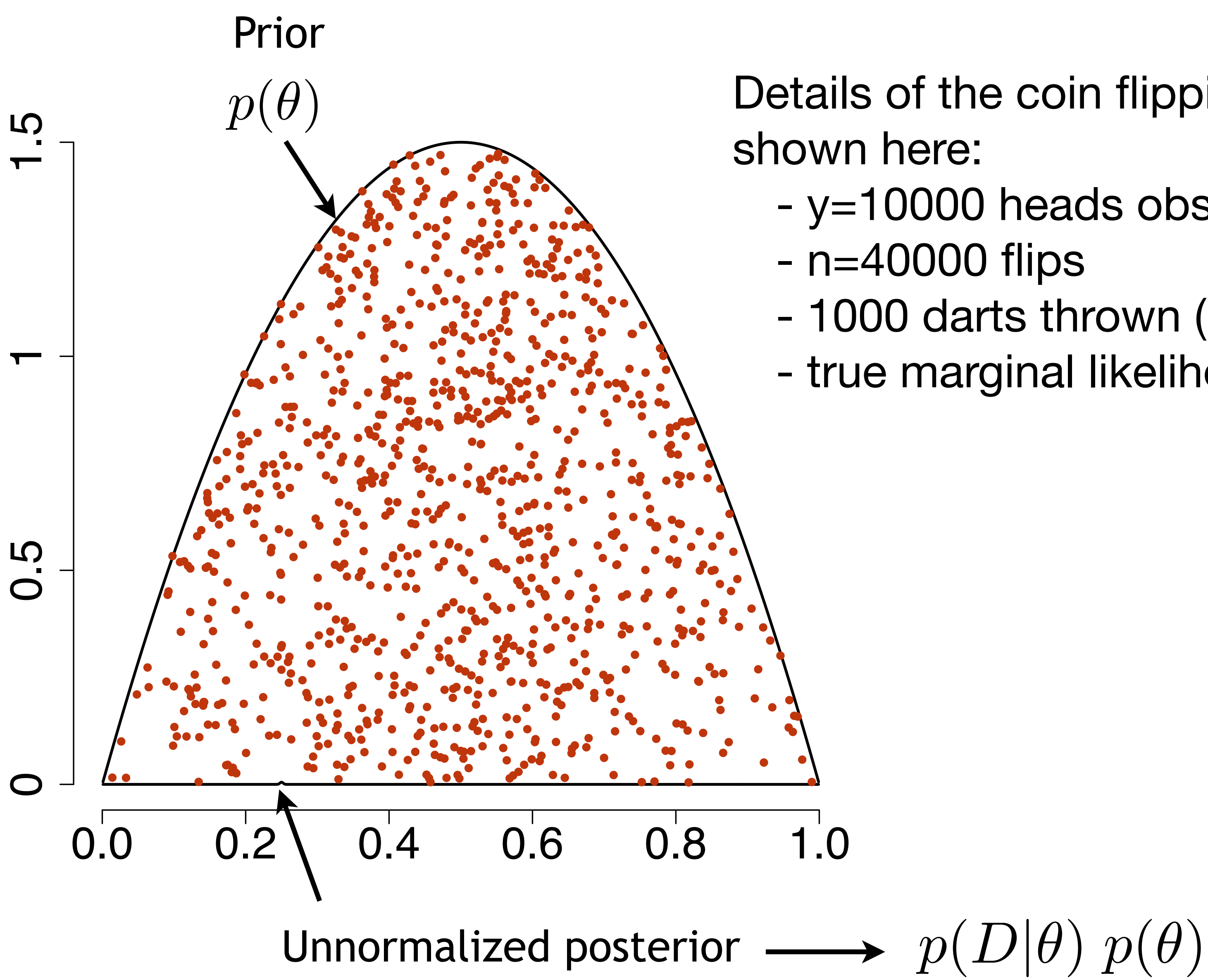


p(D) is the marginal likelihood

Estimating *p(D)* is equivalent to estimating the area under the curve whose height is, for every value of θ, equal to the **posterior kernel**, $p(D|\theta)\ p(\theta)$


2 r
2 r

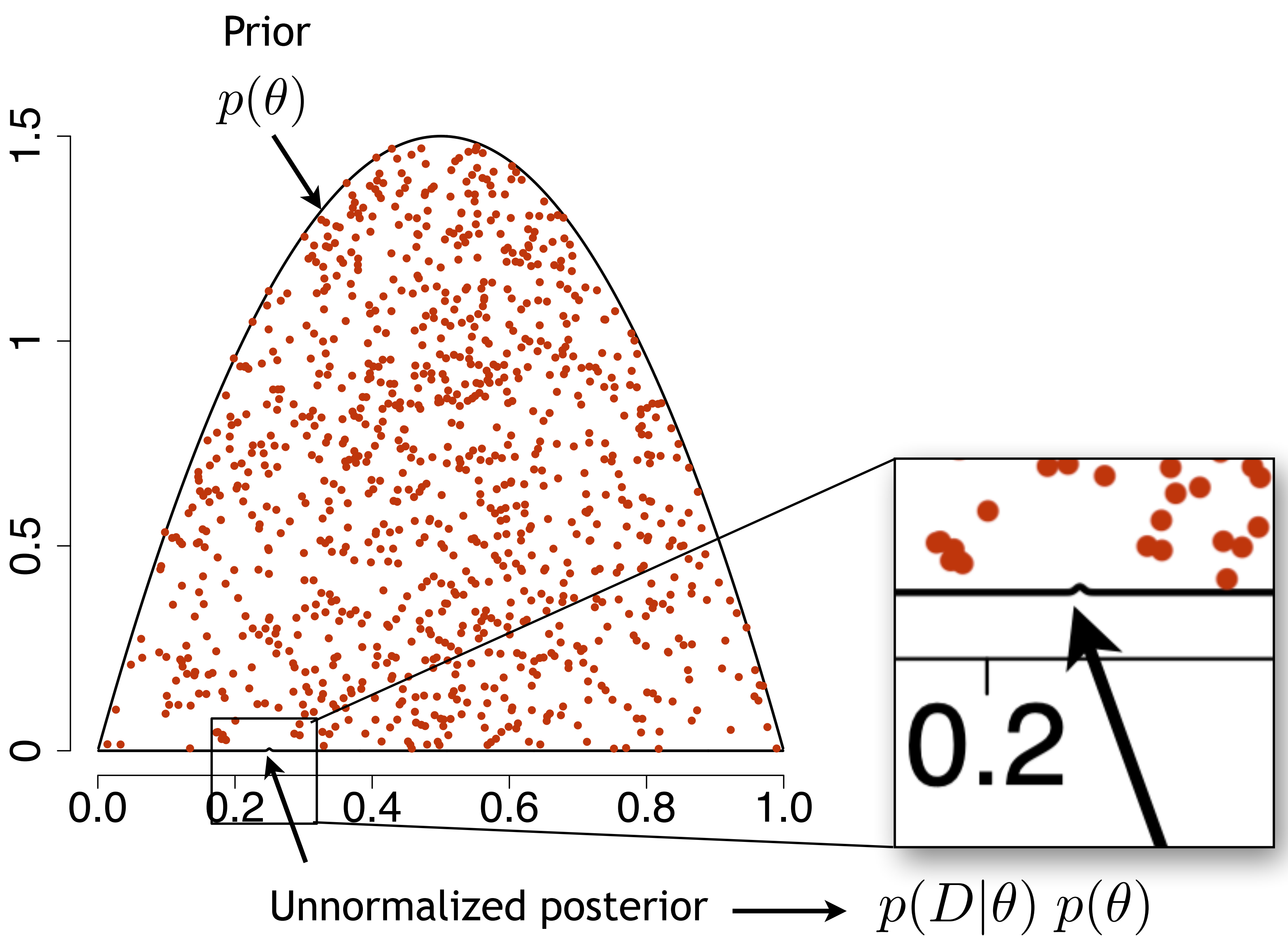
Estimating the marginal likelihood



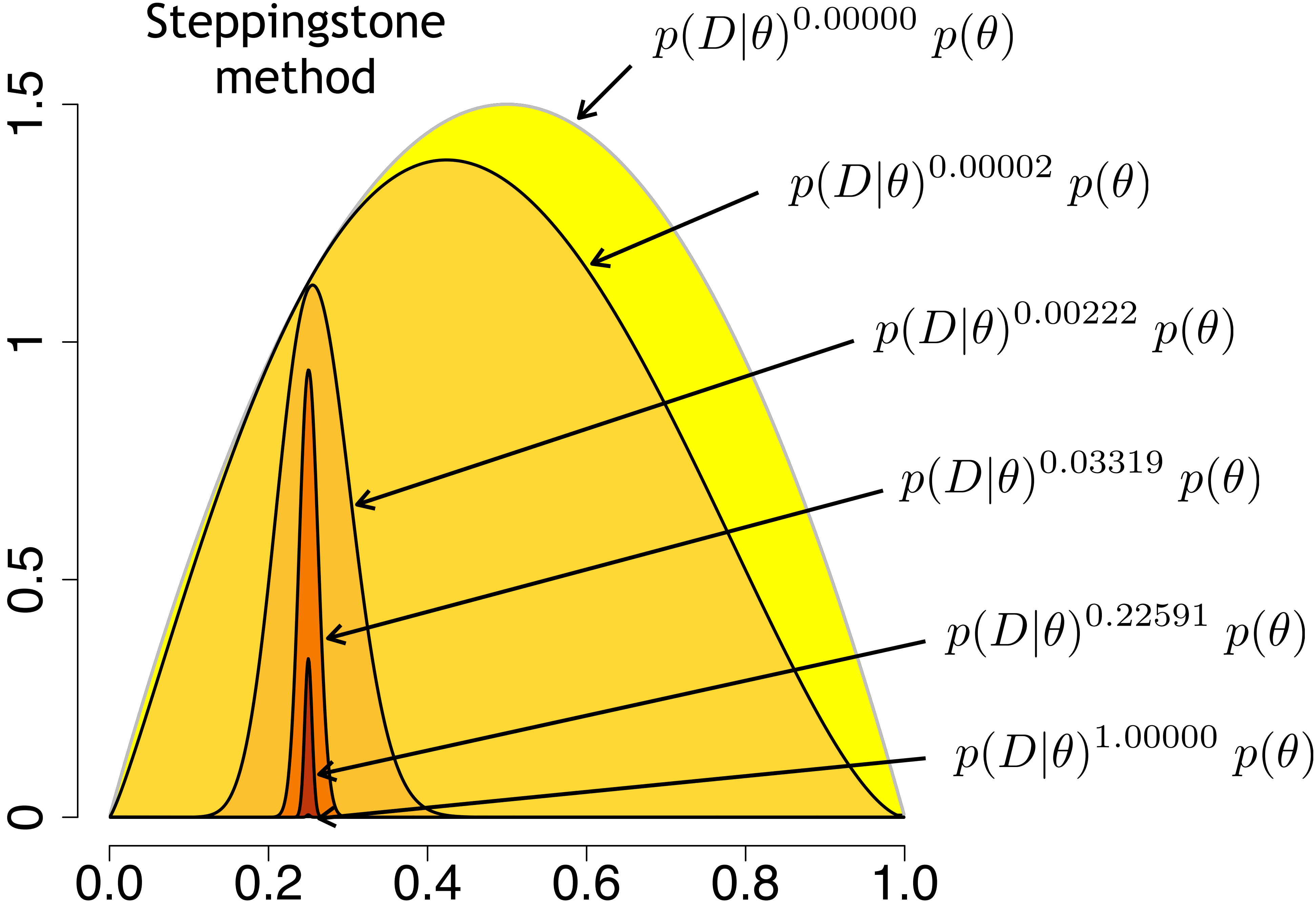
Details of the coin flipping experiment shown here:

- y=10000 heads observed
- n=40000 flips
- 1000 darts thrown (0 under posterior)
- true marginal likelihood 0.000025

Estimating the marginal likelihood

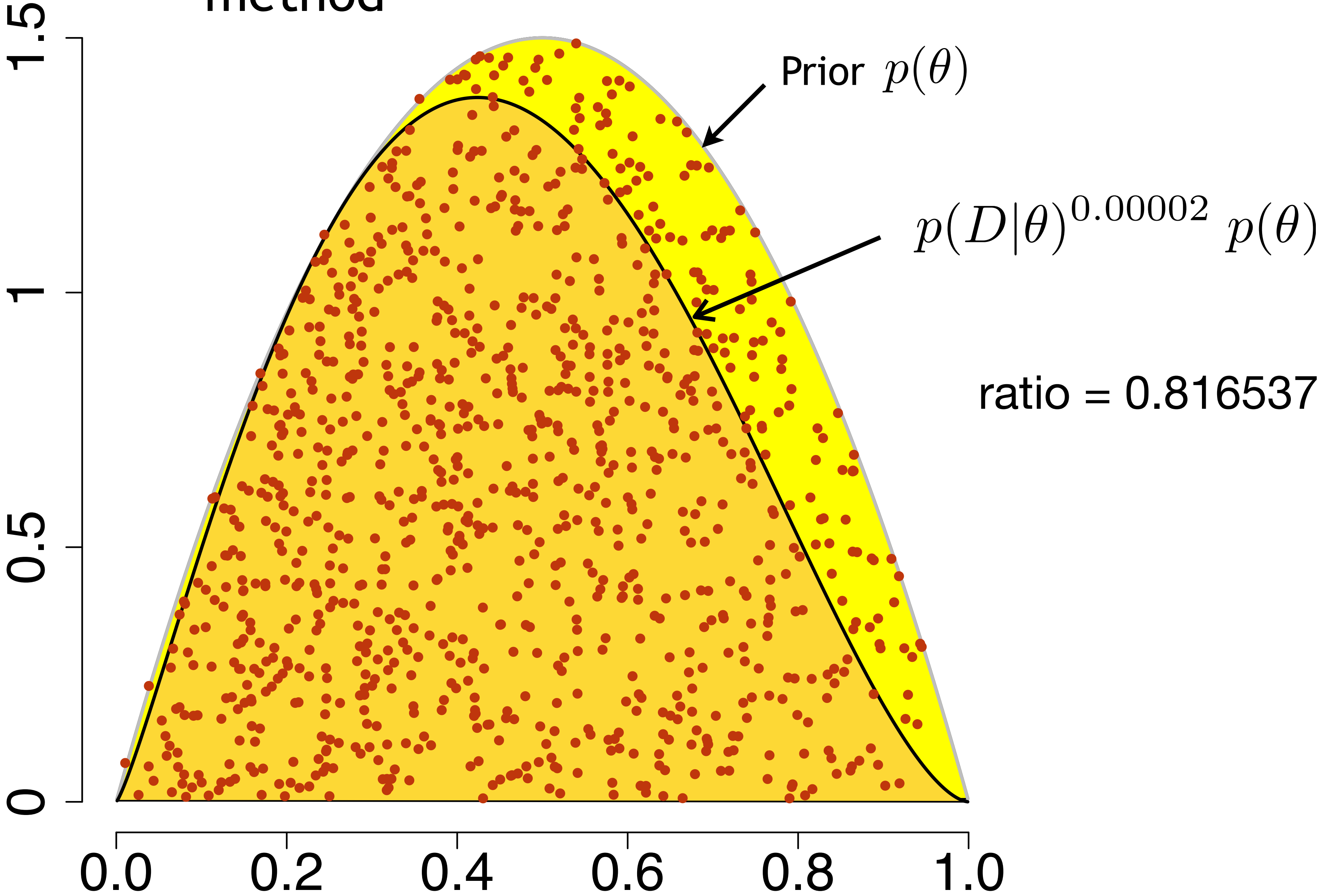


Steppingstone method

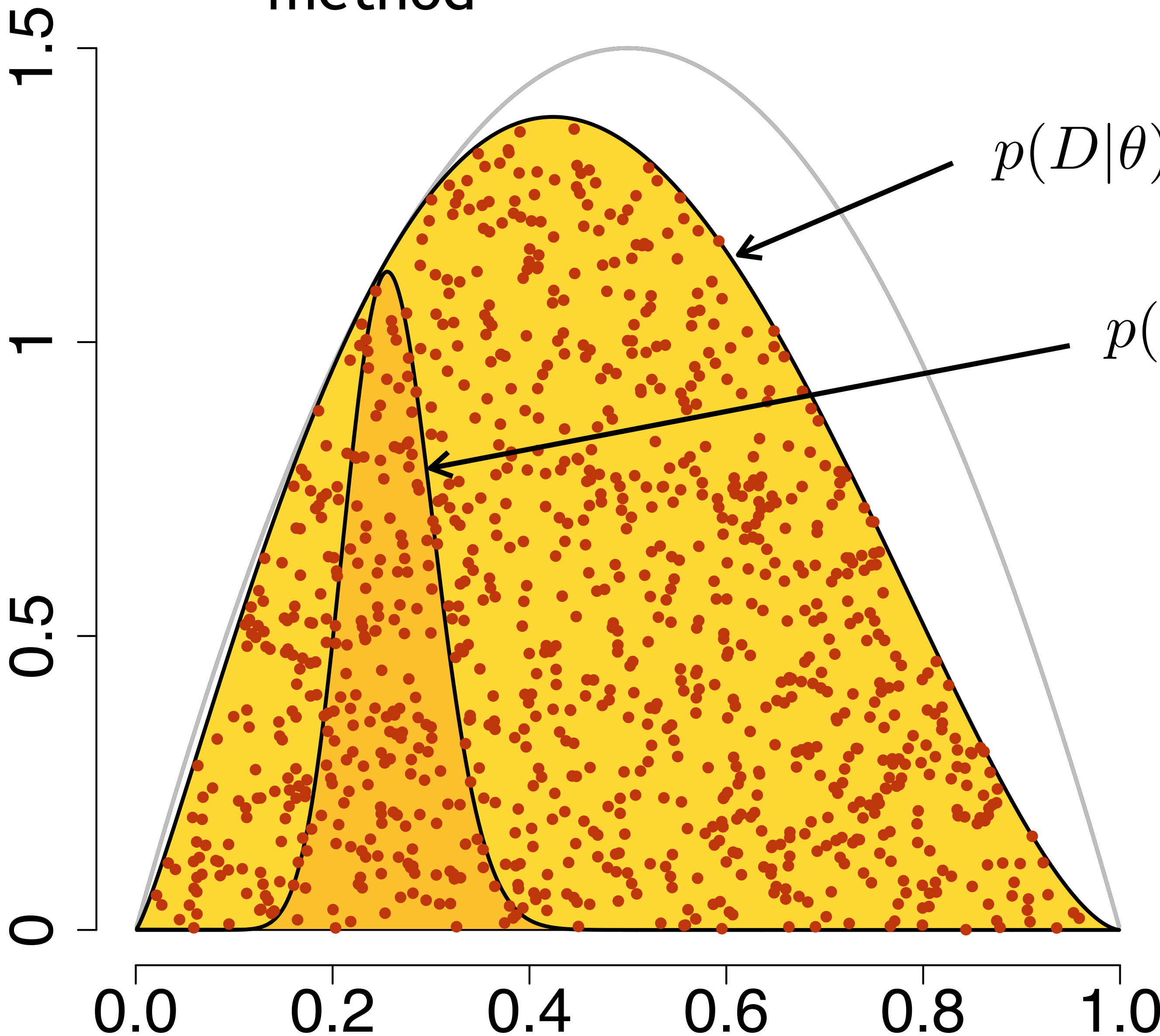


Xie et al. 2011

Steppingstone method



Steppingstone method

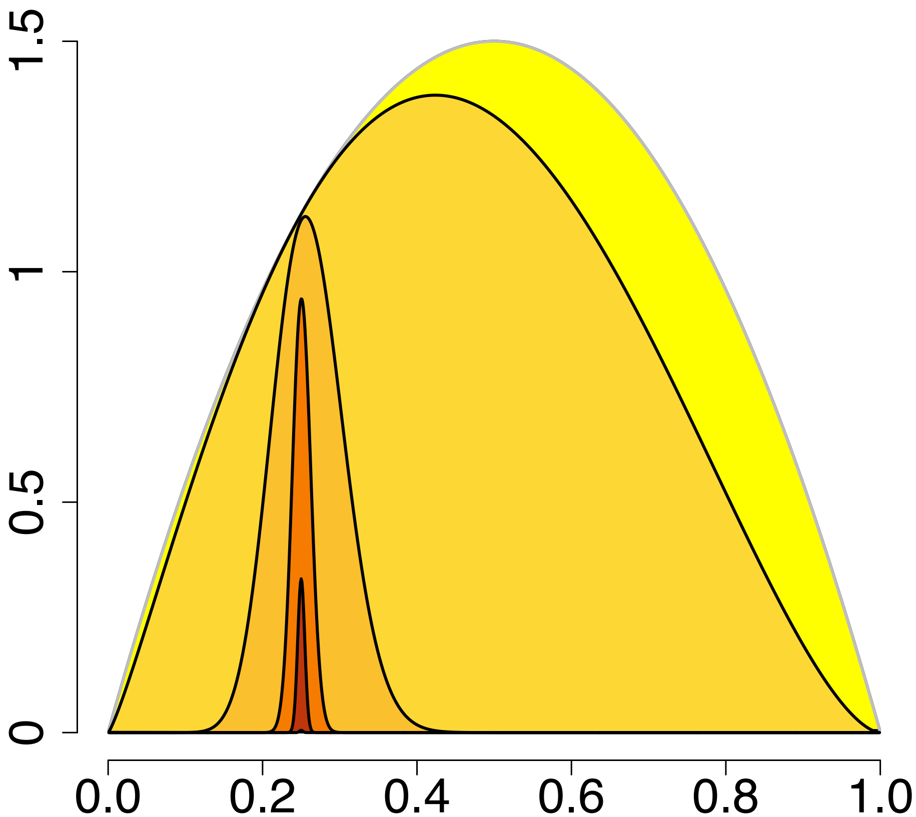


ratio = 0.167866

Steppingstone method

$$\frac{\beta_1}{\beta_0} = \left(\frac{\beta_{0.00002}}{\beta_0}\right)\left(\frac{\beta_{0.00222}}{\beta_{0.00002}}\right)\left(\frac{\beta_{0.03319}}{\beta_{0.00222}}\right)\left(\frac{\beta_{0.22591}}{\beta_{0.03319}}\right)\left(\frac{\beta_1}{\beta_{0.22591}}\right)$$

0.000061 = (0.816537) (0.167866) (0.289389) (0.172237) (0.008923)



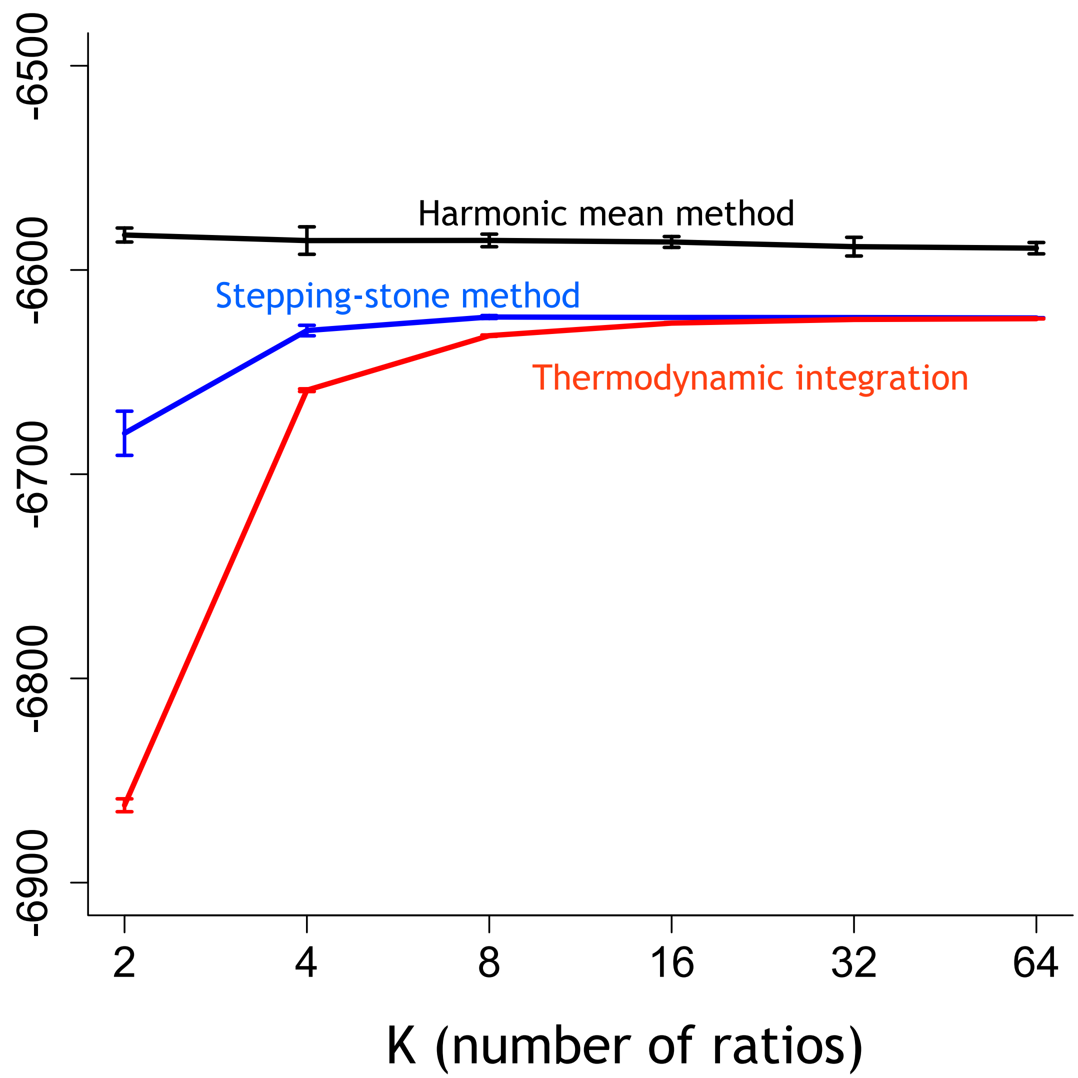
0.000025 = true value

How many “stepping stones” (i.e. ratios) are needed?

Newton and Raftery 1994

Xie et al. 2011; Fan et al. 2011

Lartillot and Philippe 2006

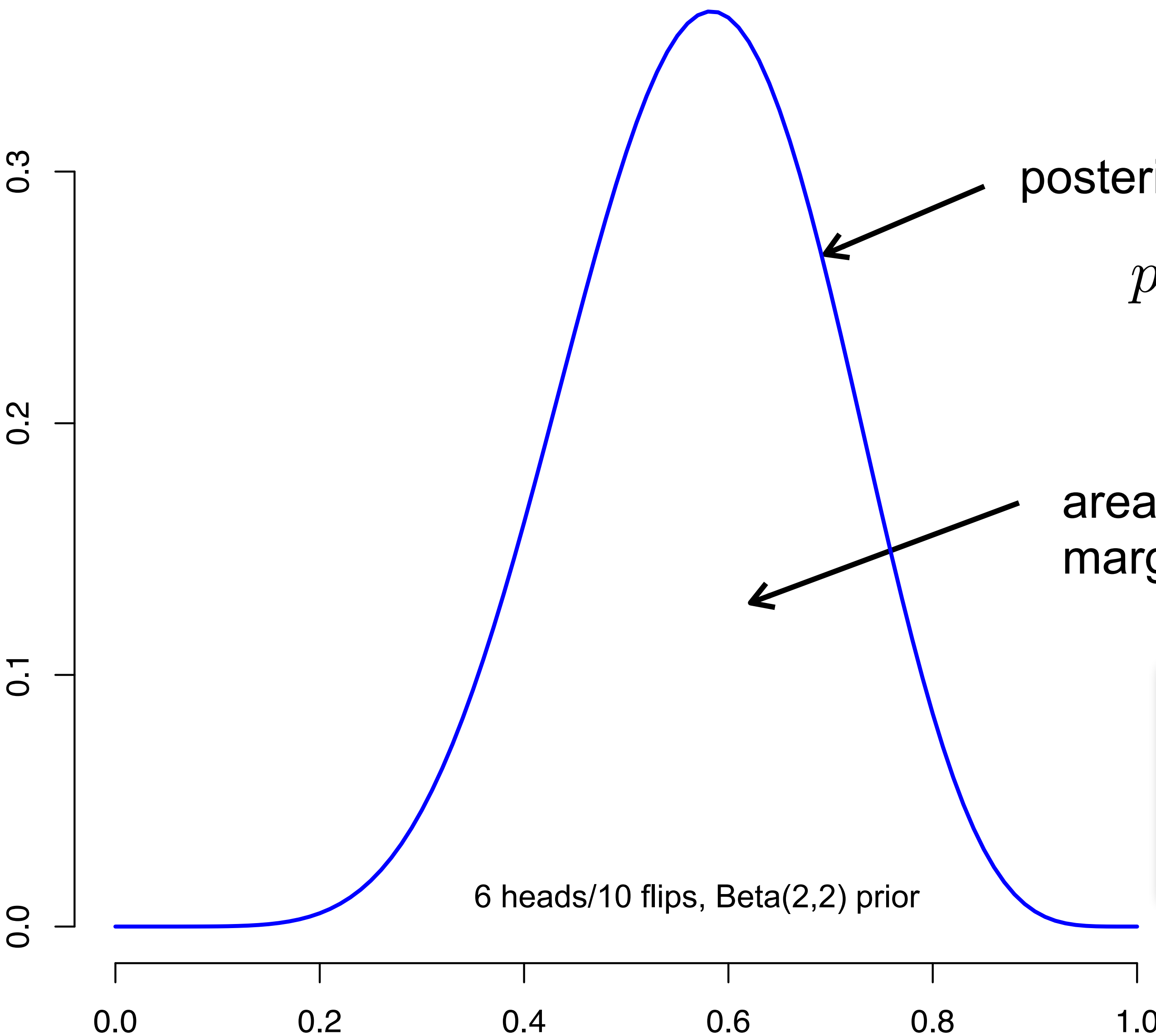


-6589

-6623

1000 samples/ratio

Bayesian Information Criterion (BIC)



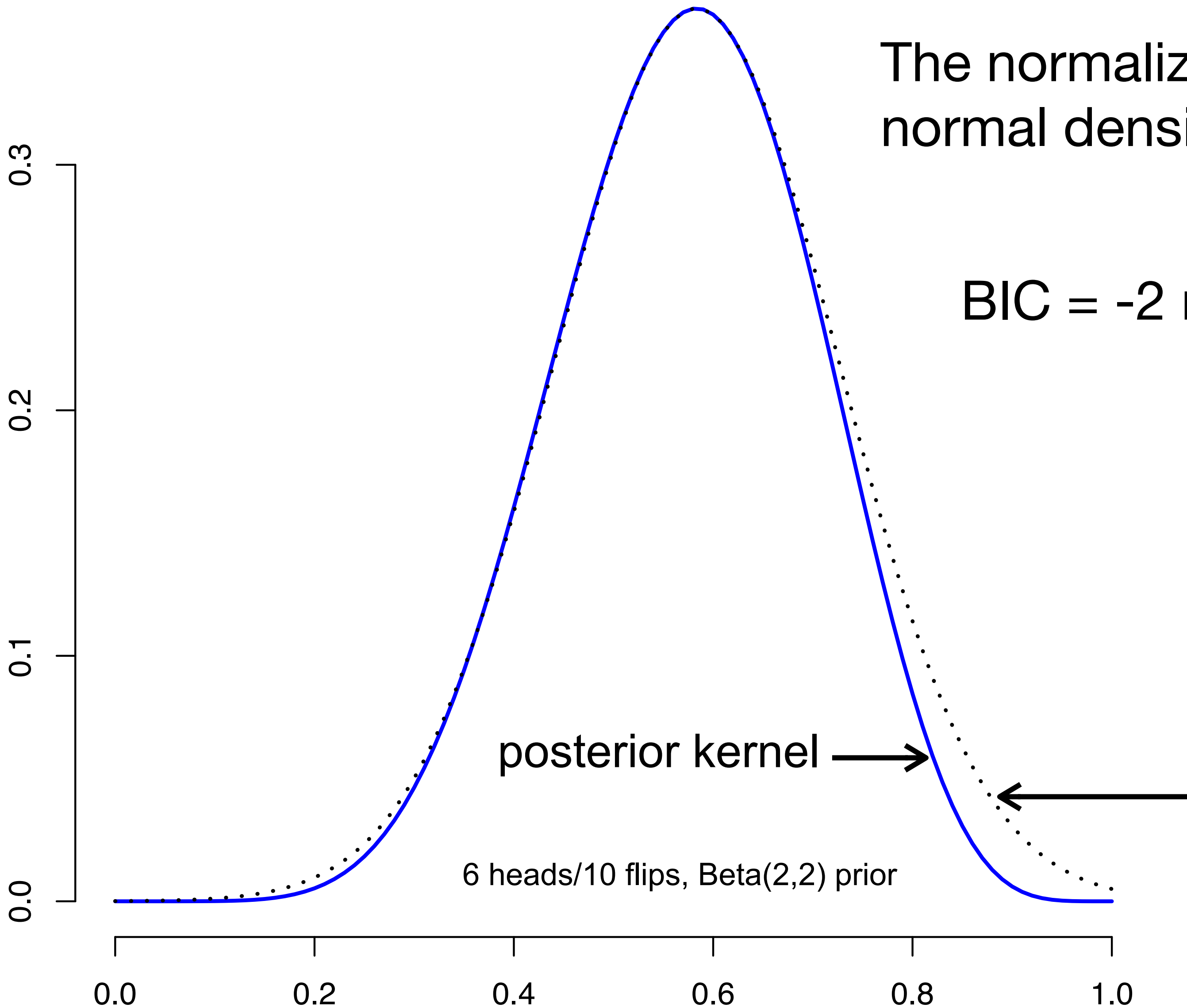
posterior kernel

$$p(y|\theta)p(\theta)$$

area underneath kernel equals marginal likelihood

Notice that the posterior kernel looks a lot like a normal density

BIC ≈ -log(marginal likelihood)



The normalizing constant of a normal density is easy to calculate

$$\mathrm{BIC} = -2\max(\log L) + \underbrace{K}_{\text{no. parameters}}\log(\underbrace{n}_{\text{no. sites}})$$

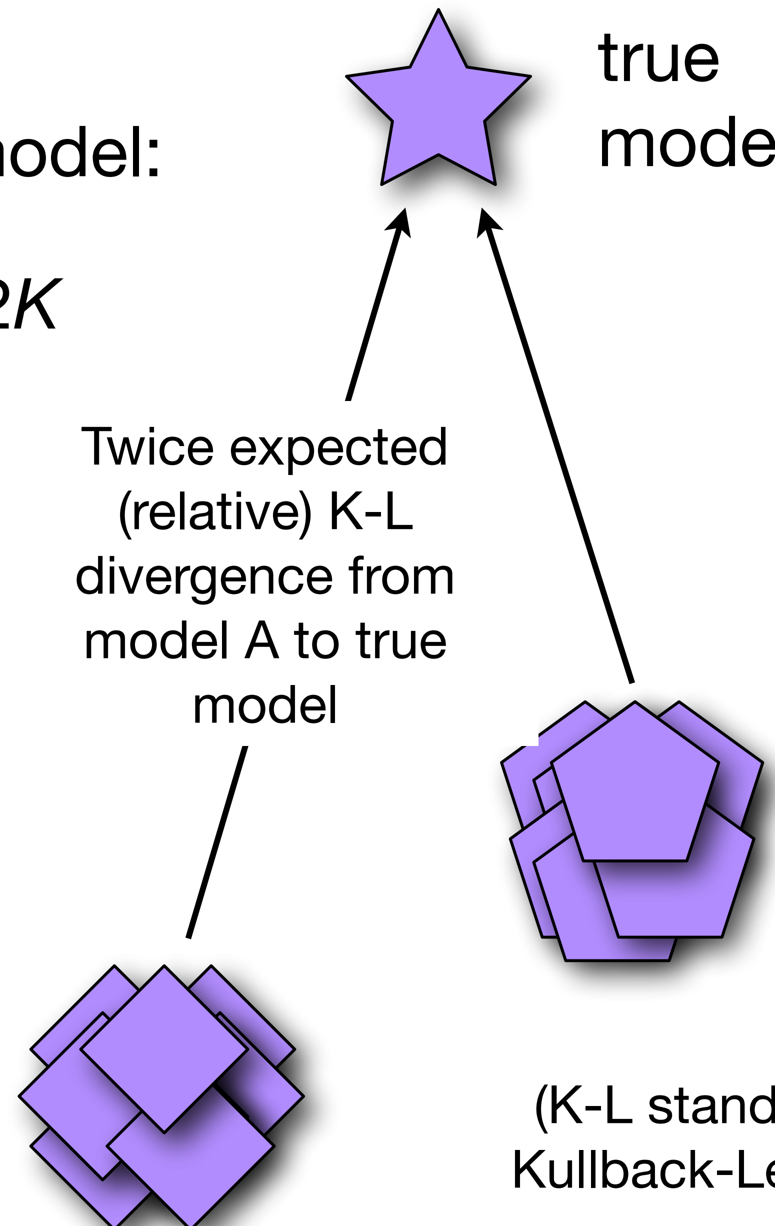
normal density scaled to match mode and curvature of posterior kernel at the peak

Akaike Information Criterion (AIC)

Calculate AIC for each model:

$$\text{AIC} = -2 \max(\log L) + 2K$$

Model with smallest AIC is best



(K-L stands for Kullback-Leibler)

Recall from likelihood lecture…

First 32 nucleotides of the ψη-globin gene of gorilla:

GAAGTCCTTGAGAAATAAACTGCACACACTGG

$$\log L = 12\log(\pi_A) + 7\log(\pi_C) + 7\log(\pi_G) + 6\log(\pi_T)$$

Find *maximum* logL under F81 (unconstrained) model:

$$\begin{aligned}\log L &= 12\log(\pi_A) + 7\log(\pi_C) + 7\log(\pi_G) + 6\log(\pi_T) \\ &= 12\log(0.375) + 7\log(0.219) + 7\log(0.219) + 6\log(0.187) \\ &= -43.1\end{aligned}$$

Find *maximum* logL under JC69 (constrained) model:

$$\begin{aligned}\log L &= 12\log(\pi_A) + 7\log(\pi_C) + 7\log(\pi_G) + 6\log(\pi_T) \\ &= 12\log(0.25) + 7\log(0.25) + 7\log(0.25) + 6\log(0.25) \\ &= -44.4\end{aligned}$$

F81 fits better (-43.1 > -44.4), but not significantly better (P = 0.457, chi-squared with 3 d.f.)

Akaike Information Criterion (AIC)

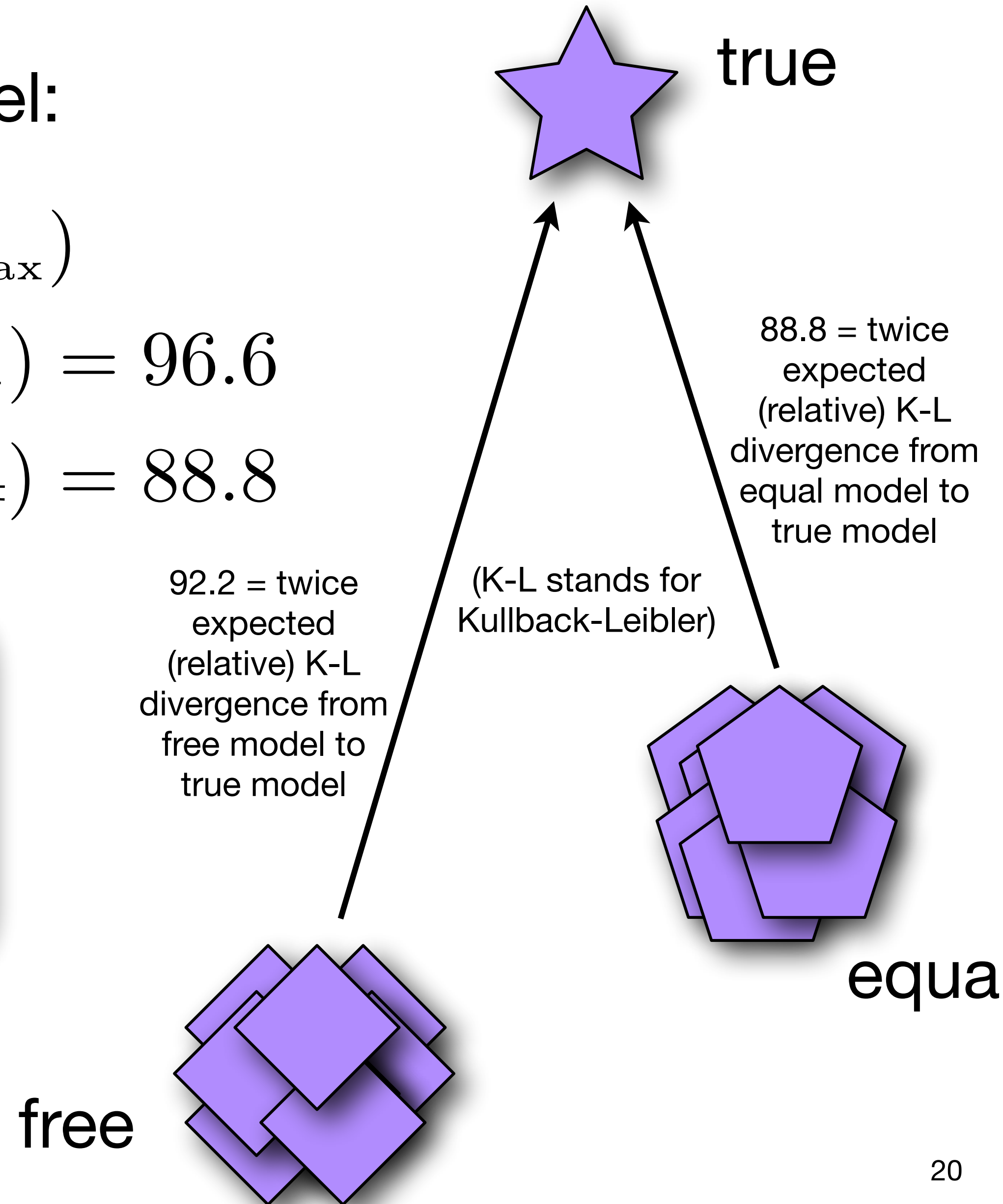
Calculate AIC for each model:

$$AIC = 2K - 2\log(L_{\mathrm{max}})$$

$$AIC_{\mathrm{free}} = 2(3) - 2(-43.1) = 96.6$$

$$AIC_{\mathrm{equal}} = 2(0) - 2(-44.4) = 88.8$$

The constrained model ("equal") is a better choice than the unconstrained model ("free") according to AIC



Bayesian Information Criterion (BIC)

Calculate BIC for each model:

$$BIC = K \log(n) - 2 \log(L_{\max})$$

$$BIC_{\text{free}} = 3 \log(32) - 2(-43.1) = 96.6$$

$$BIC_{\text{equal}} = 0 \log(32) - 2(-44.4) = 88.8$$

The constrained model ("equal") is a better choice than the unconstrained model ("free") according to BIC too