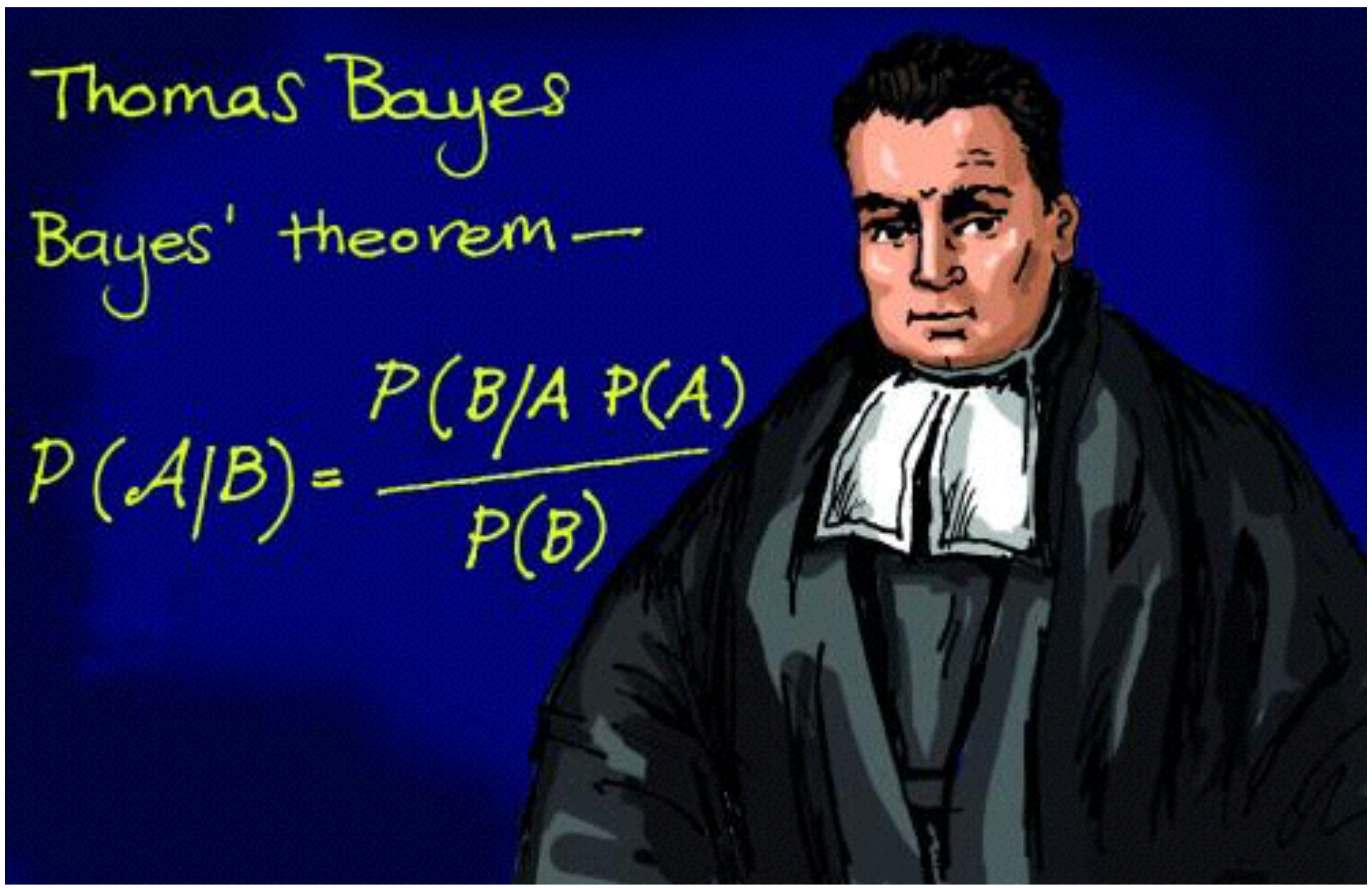
A Brief Overview of Bayesian Inference



http://www.pharmaceutical-journal.com/opinion/blogs/bayess-theorem-falls-foul-of-the-law/11089087.blog

Some Facts About Probability

$$P(A \cap B) = P(A|B)P(B) = P(B|A)P(A)$$

$$P(A|B)P(B) = P(B|A)P(A)$$

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Bayes Theorem

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

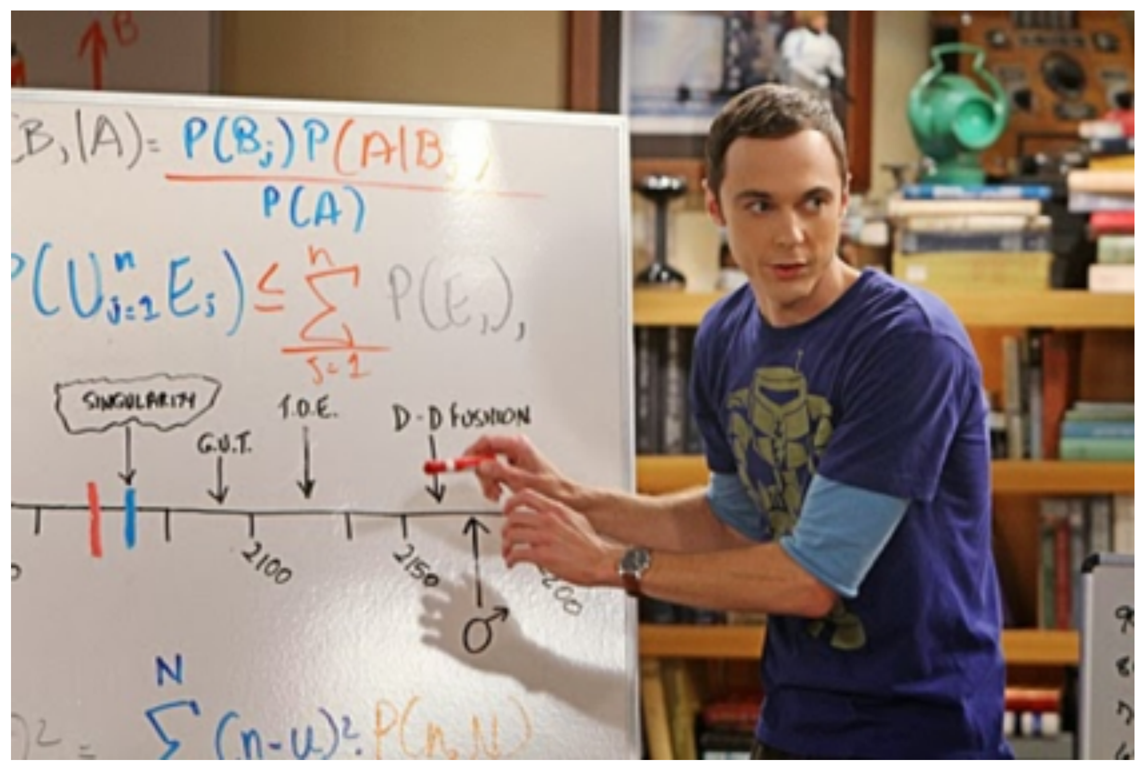
T. Bayes.

https://en.wikipedia.org/wiki/Thomas_Bayes

Bayes Theorem

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

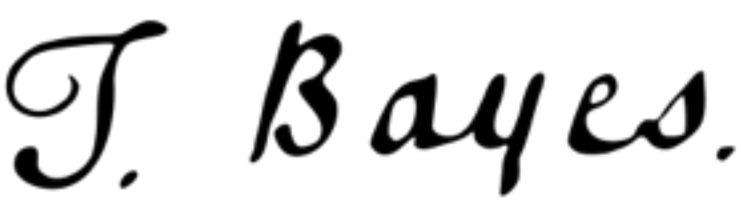
T. Bayes.

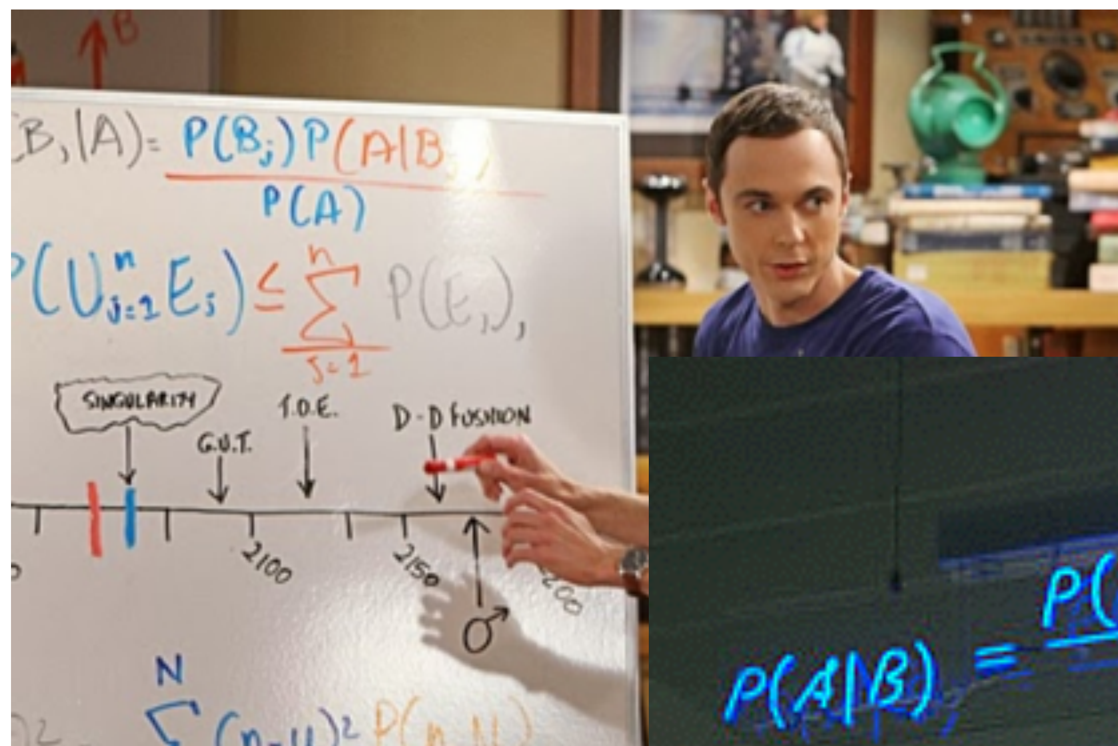


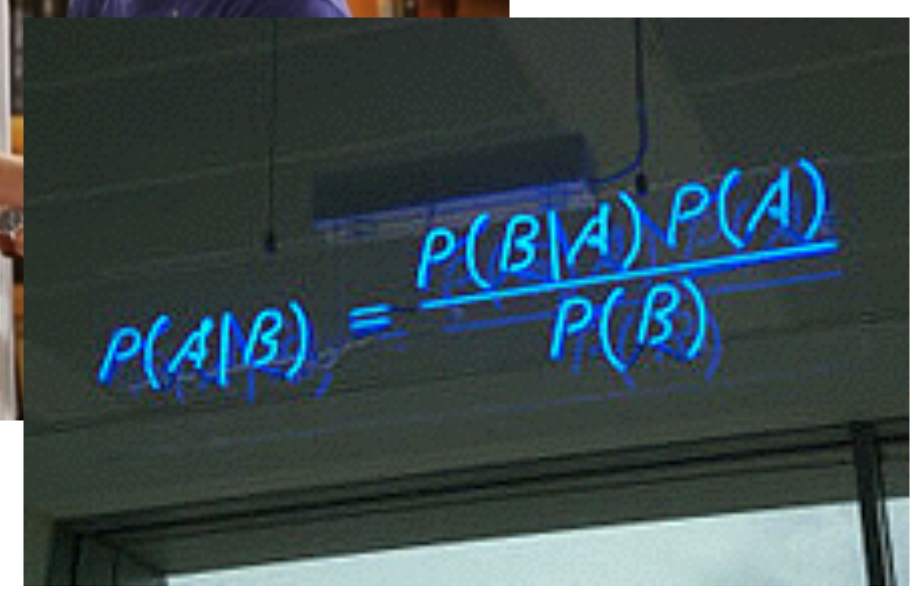
https://en.wikipedia.org/wiki/Thomas_Bayes

Bayes Theorem

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$



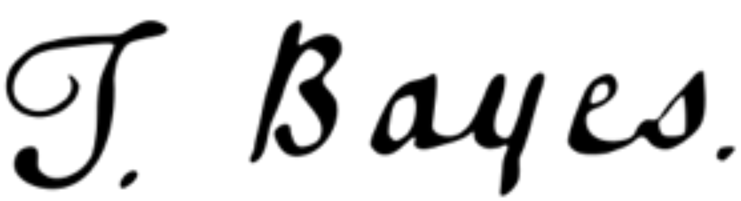


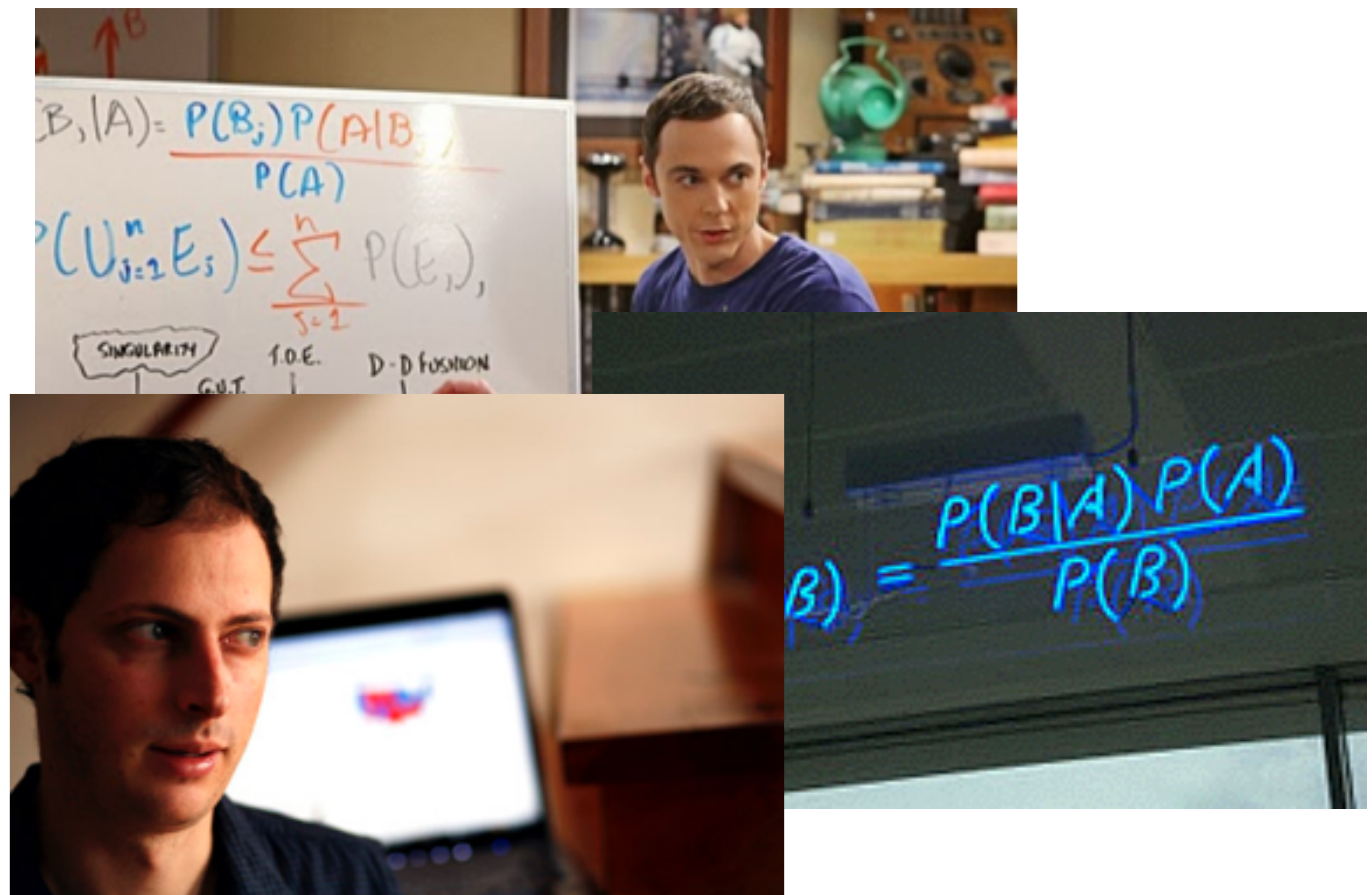


https://en.wikipedia.org/wiki/Thomas_Bayes

Bayes Theorem

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

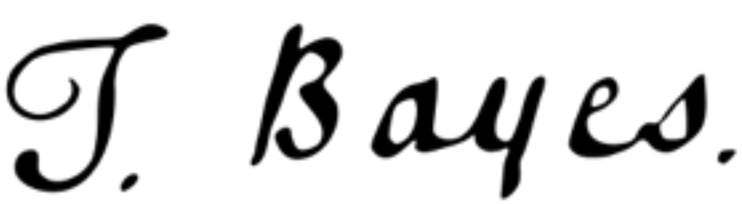


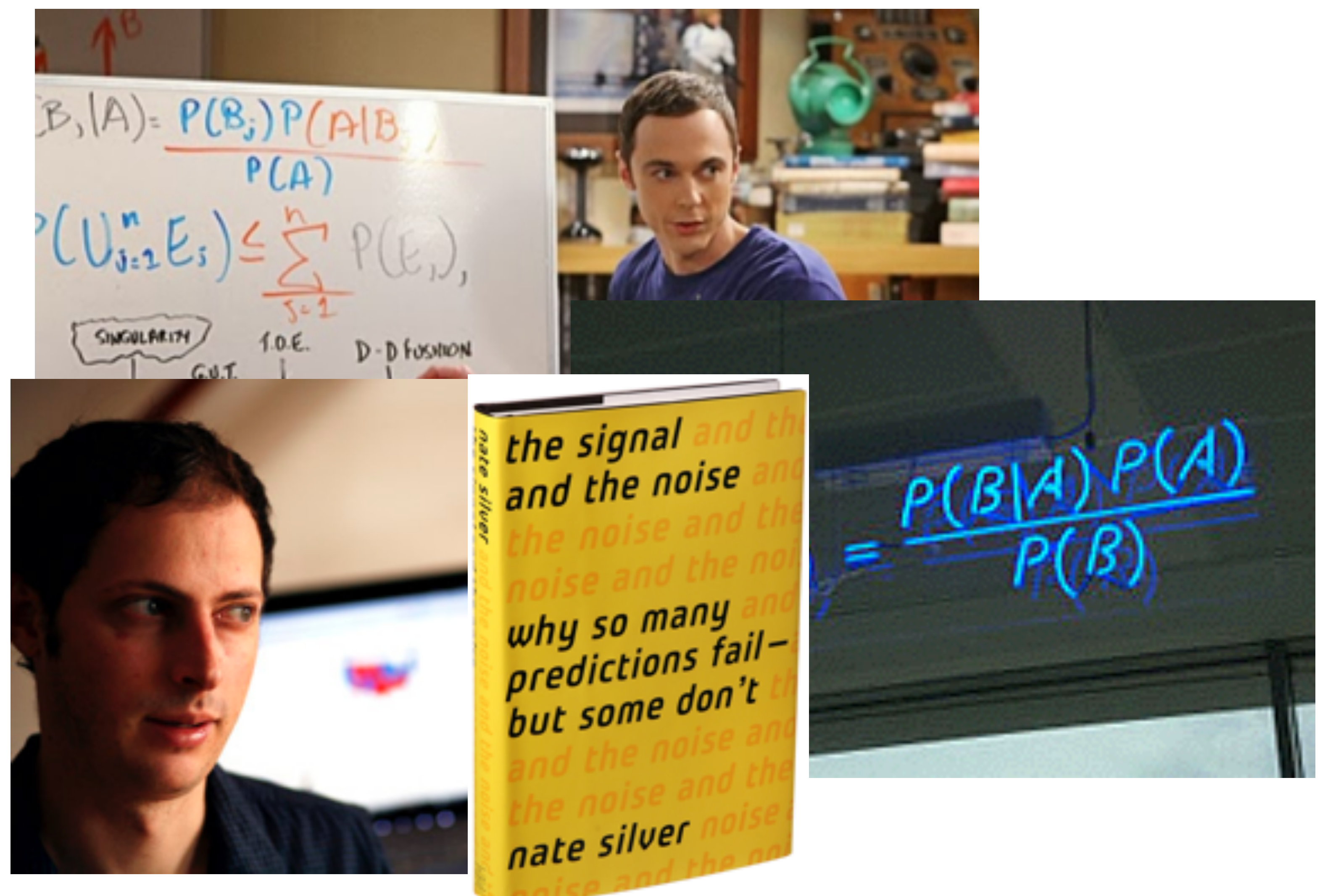


https://en.wikipedia.org/wiki/Thomas_Bayes

Bayes Theorem

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$





https://en.wikipedia.org/wiki/Thomas_Bayes

Bayes Theorem

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

T. Bayes.

Rev. Thomas Bayes
1701-1761

"An Essay towards solving a Problem in the Doctrine of Chances" published in 1763 (Richard Price)

Binomial with a uniform prior on *p*

https://en.wikipedia.org/wiki/Thomas_Bayes

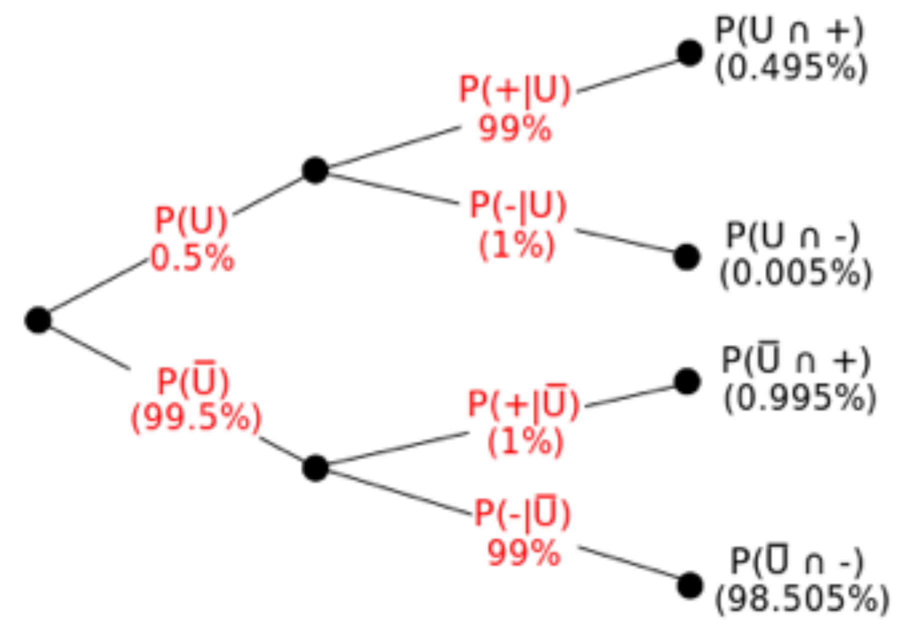
Bayes Theorem

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Also used with frequentist probabilities!

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

Suppose a drug test is 99% sensitive and 99% specific. That is, the test will produce 99% true positive results for drug users and 99% true negative results for non-drug users. Suppose that 0.5% of people are users of the drug. If a randomly selected individual tests positive, what is the probability that he is a user?



$$\begin{aligned} P(\text{User} \mid +) &= \frac{P(+ \mid \text{User})P(\text{User})}{P(+ \mid \text{User})P(\text{User}) + P(+ \mid \text{Non-user})P(\text{Non-user})} \\ &= \frac{0.99 \times 0.005}{0.99 \times 0.005 + 0.01 \times 0.995} \\ &\approx 33.2\% \end{aligned}$$

https://en.wikipedia.org/wiki/Bayes'_theorem

Bayes Theorem

Likelihood

Posterior

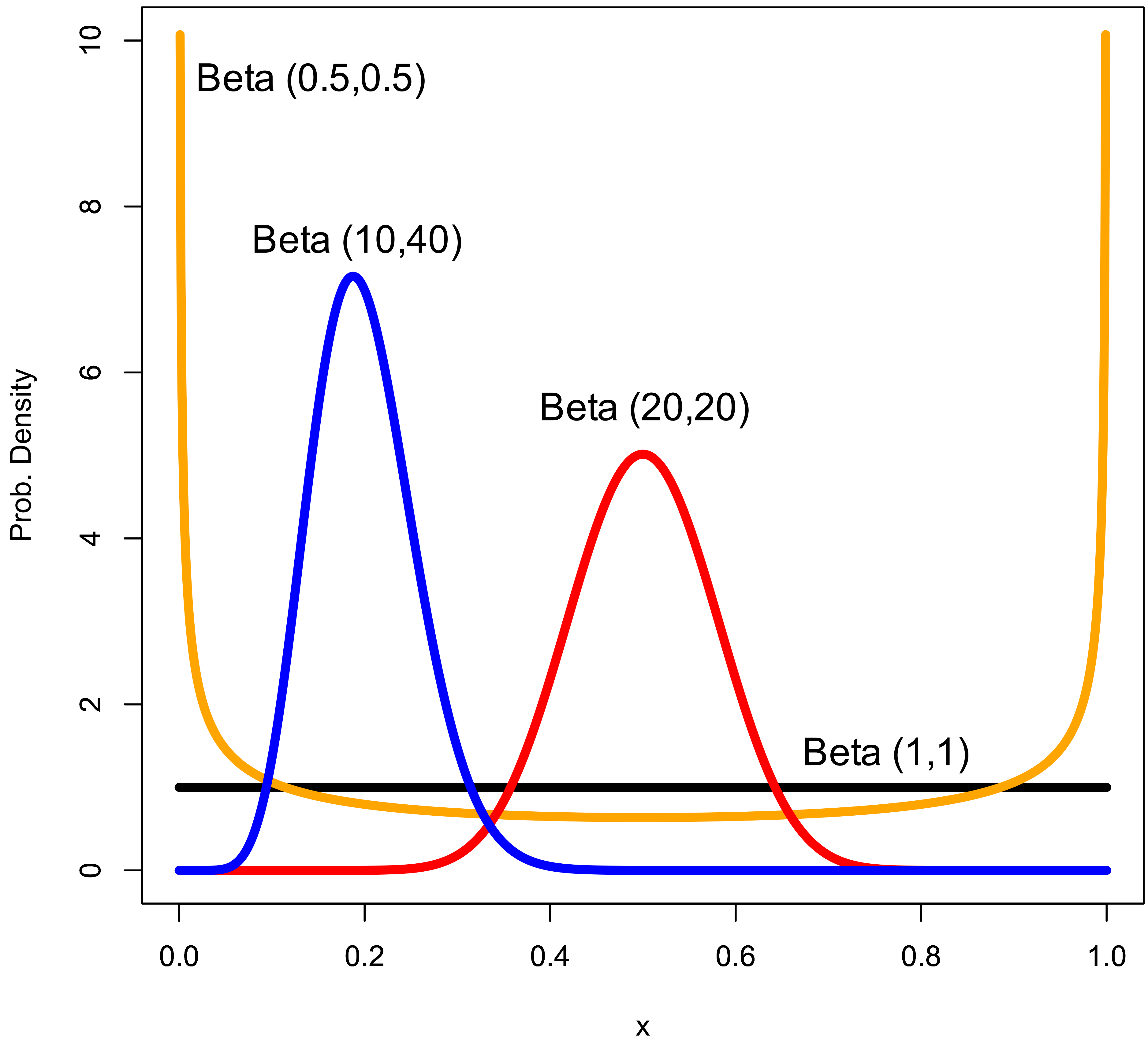
Prior

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

Normalizing Constant
(Marginal Likelihood)

Beta distributions are a flexible class of possible priors for continuous numbers between 0 and 1.

Beta Distributions



Odds Ratios

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

Odds Ratios

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

Odds Ratios

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

Odds Ratios

Posterior Odds = Prior Odds × Bayes Factor

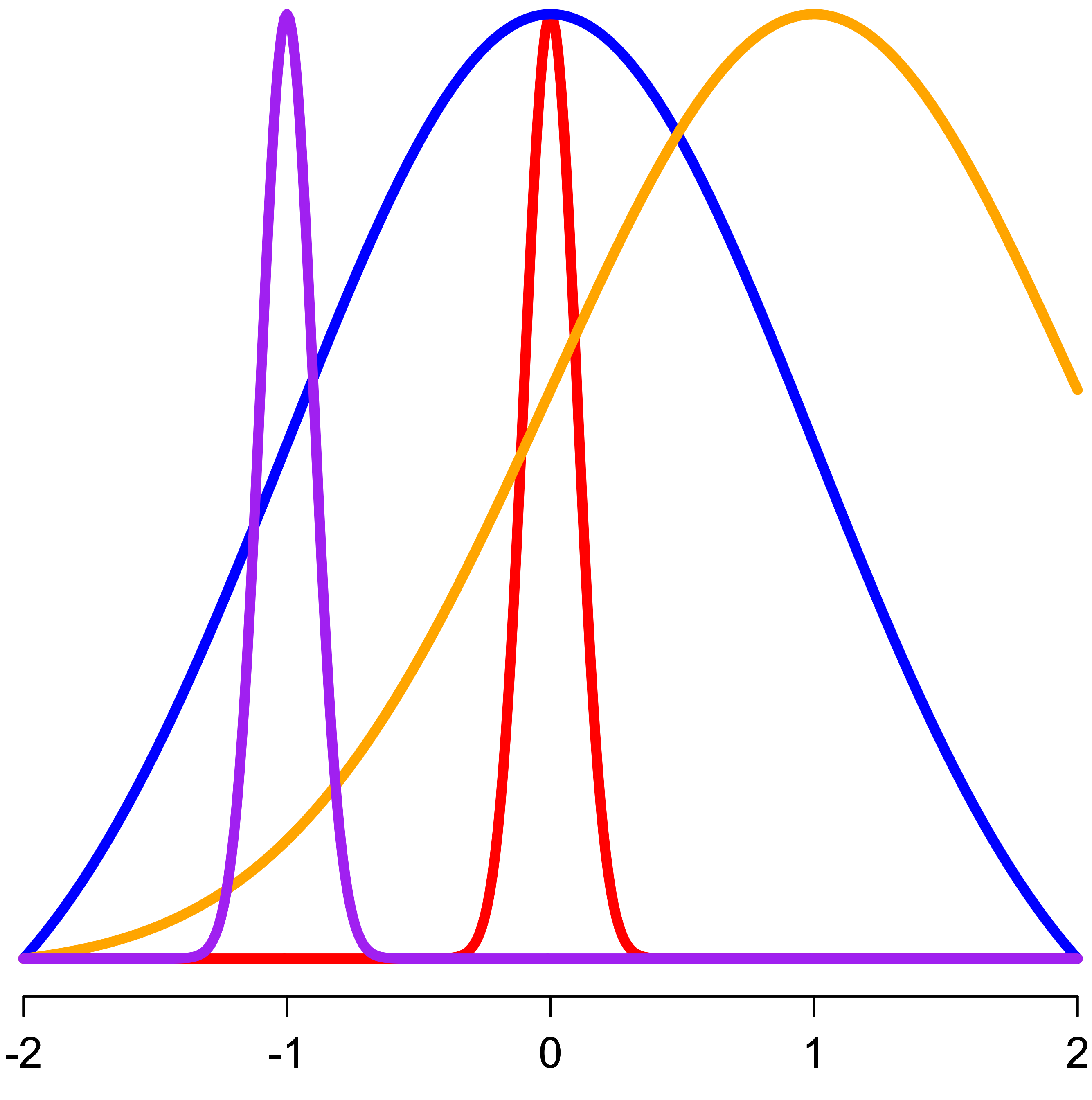
$$\boxed{\frac{P(H_1|D)}{P(H_2|D)}} = \boxed{\frac{P(H_1)}{P(H_2)}}\boxed{\frac{P(D|H_1)}{P(D|H_2)}}$$

Odds Ratios

Prior Odds · Bayes Factor · Posterior Odds

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

Estimating Parameters of a Normal Distribution



Mean (μ)

Standard Deviation (σ)

Marginalizing

Joint Probabilities

	σ_1	σ_2	σ_3	
μ_1	0.10	0.07	0.12	0.29
μ_2	0.05	0.22	0.06	0.33
μ_3	0.05	0.19	0.14	0.38
	0.20	0.48	0.32	

Marginal Probabilities

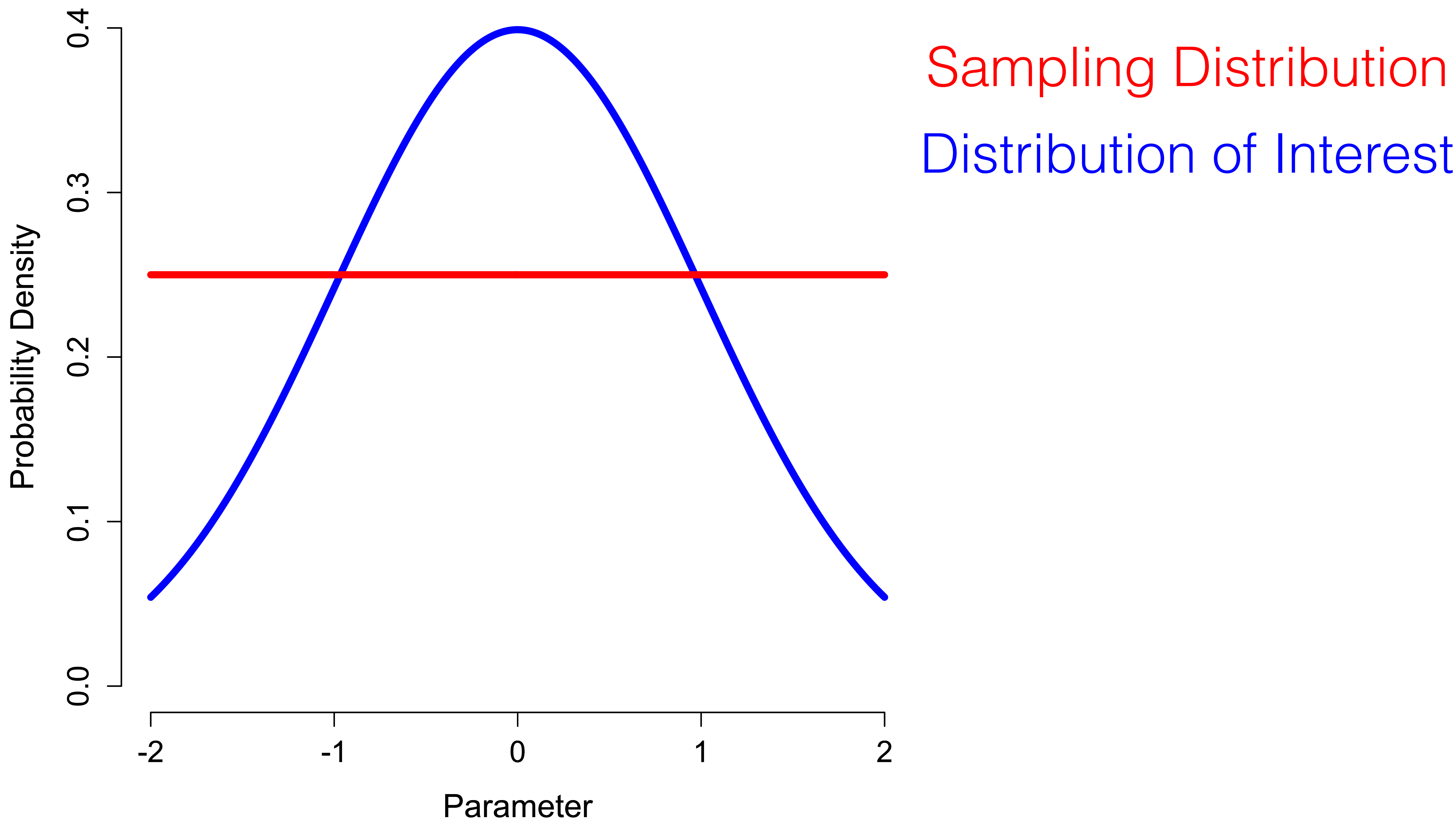
Monte Carlo Methods

An entire class of methods to draw samples from or estimate moments (mean, variance, etc.) from distributions when closed-form solutions are not available. Rely on drawing (pseudo-)random numbers.

(Monaco, not Las Vegas)

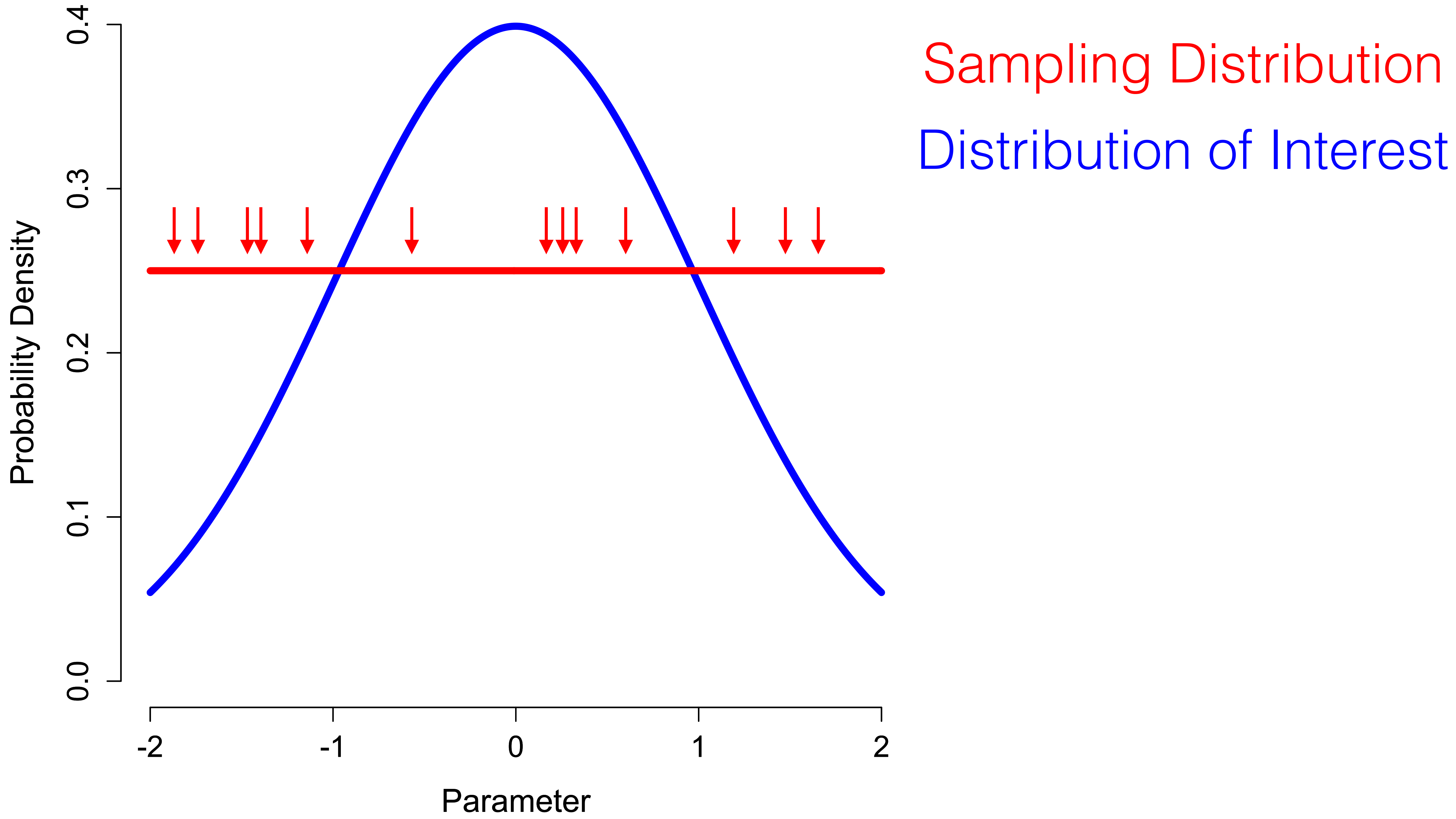
Monte Carlo Methods

Importance Sampling



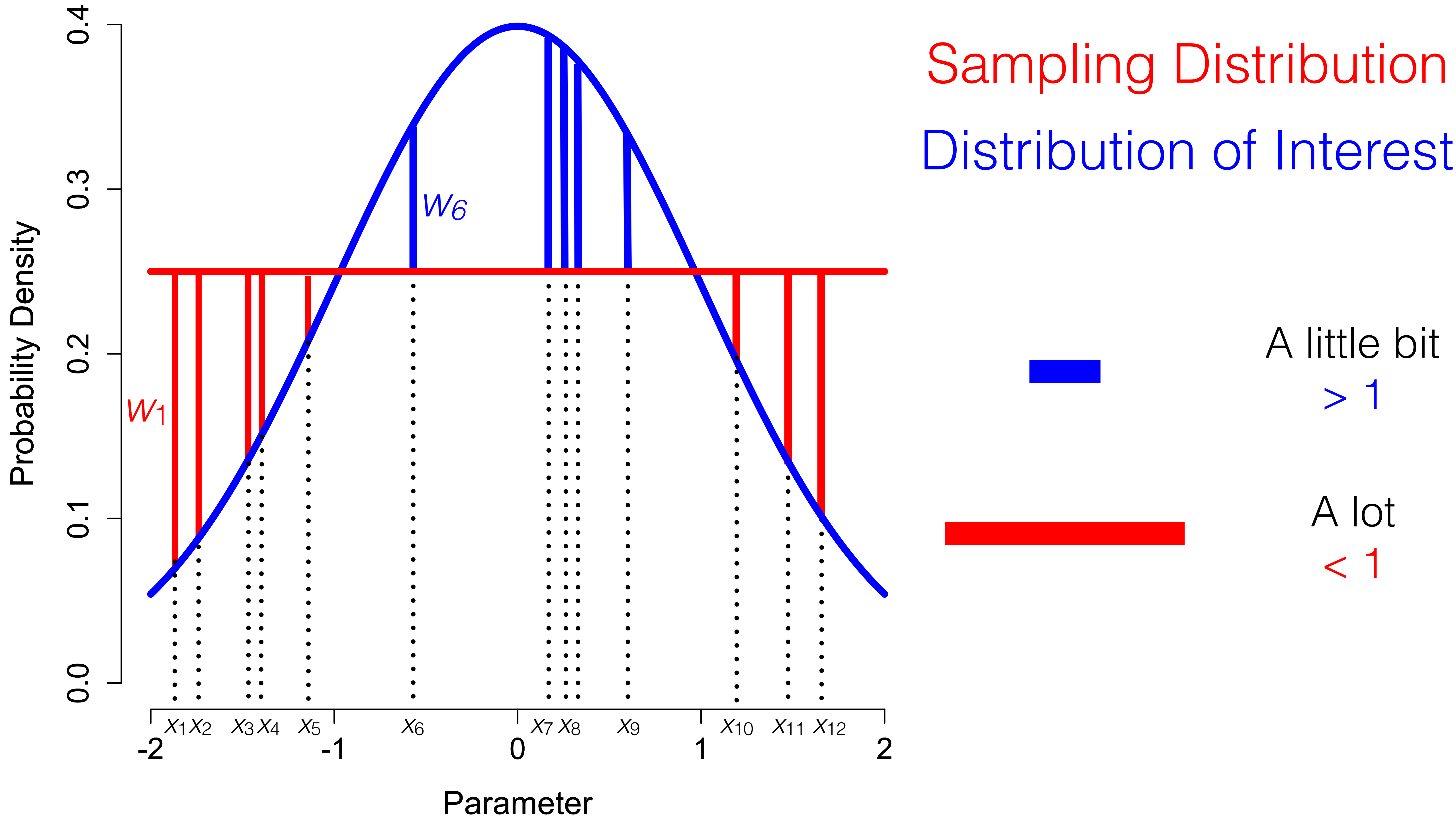
Monte Carlo Methods

Importance Sampling



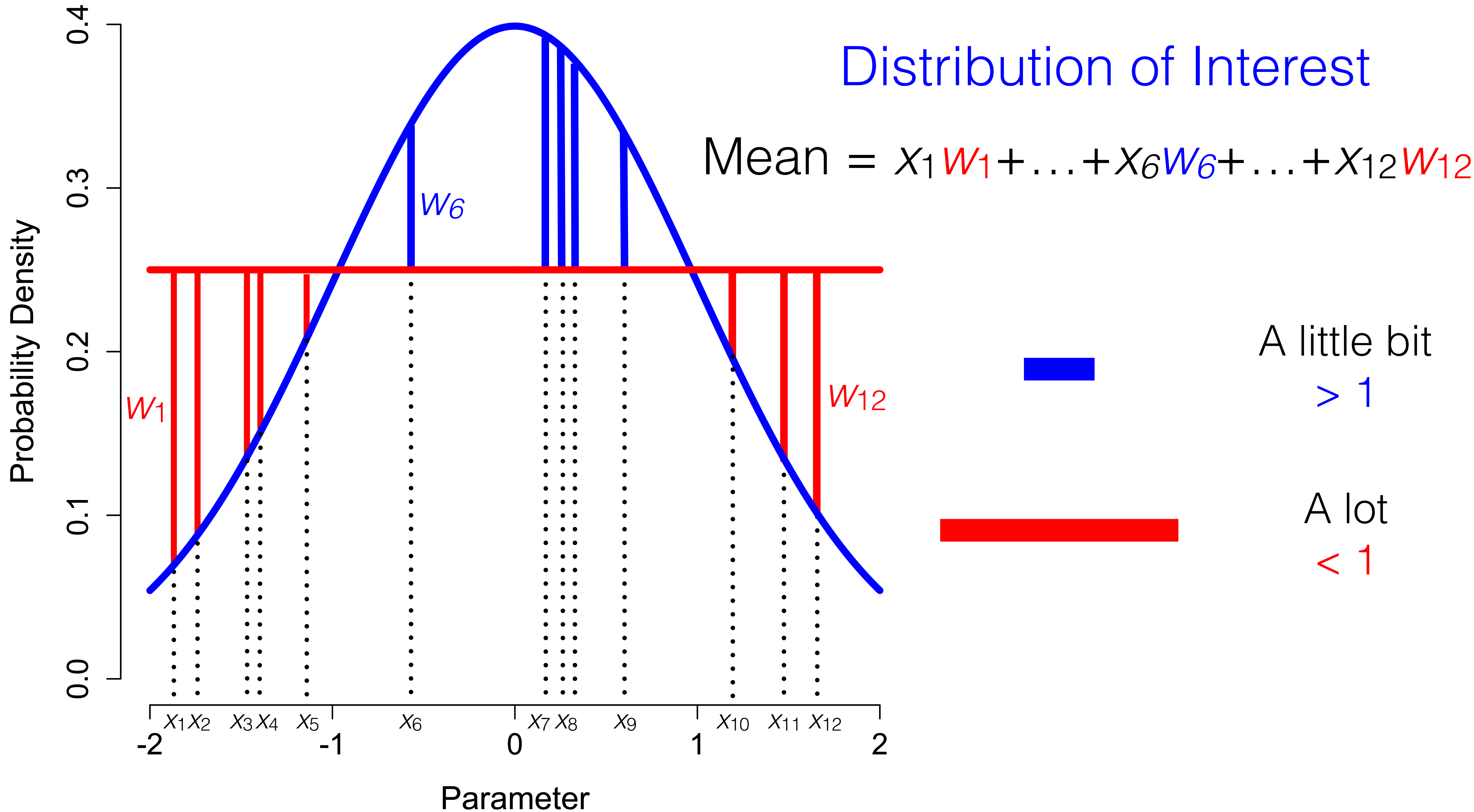
Monte Carlo Methods

Importance Sampling



Monte Carlo Methods

Importance Sampling

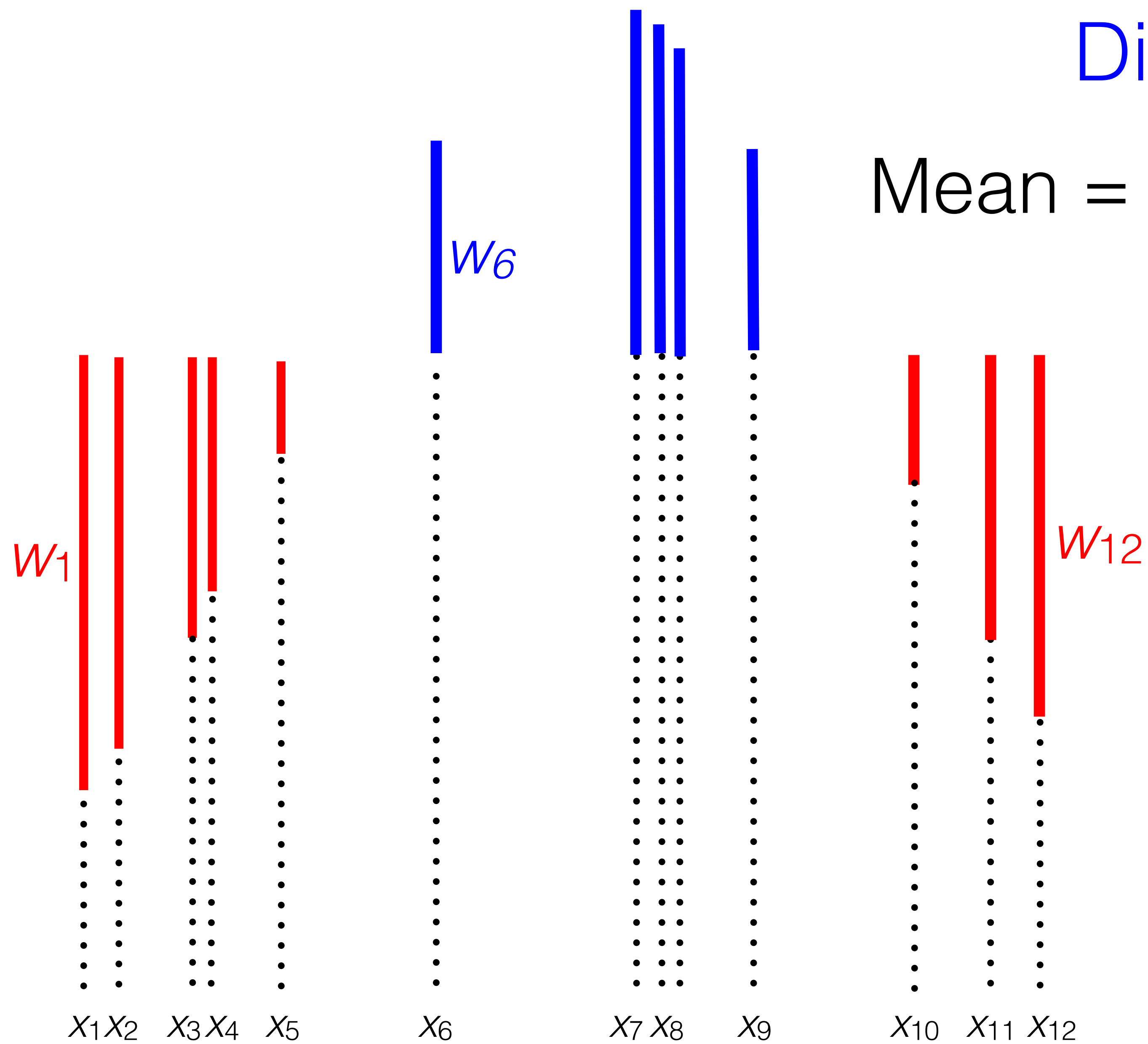


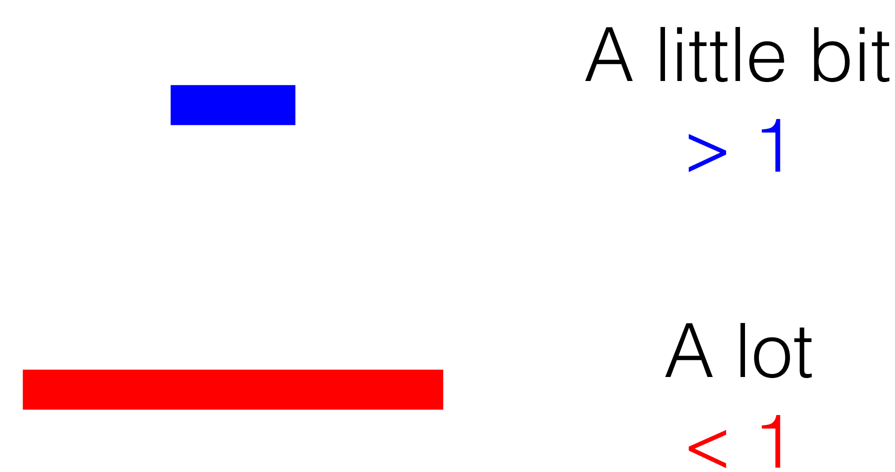
Monte Carlo Methods

Importance Sampling

Distribution of Interest

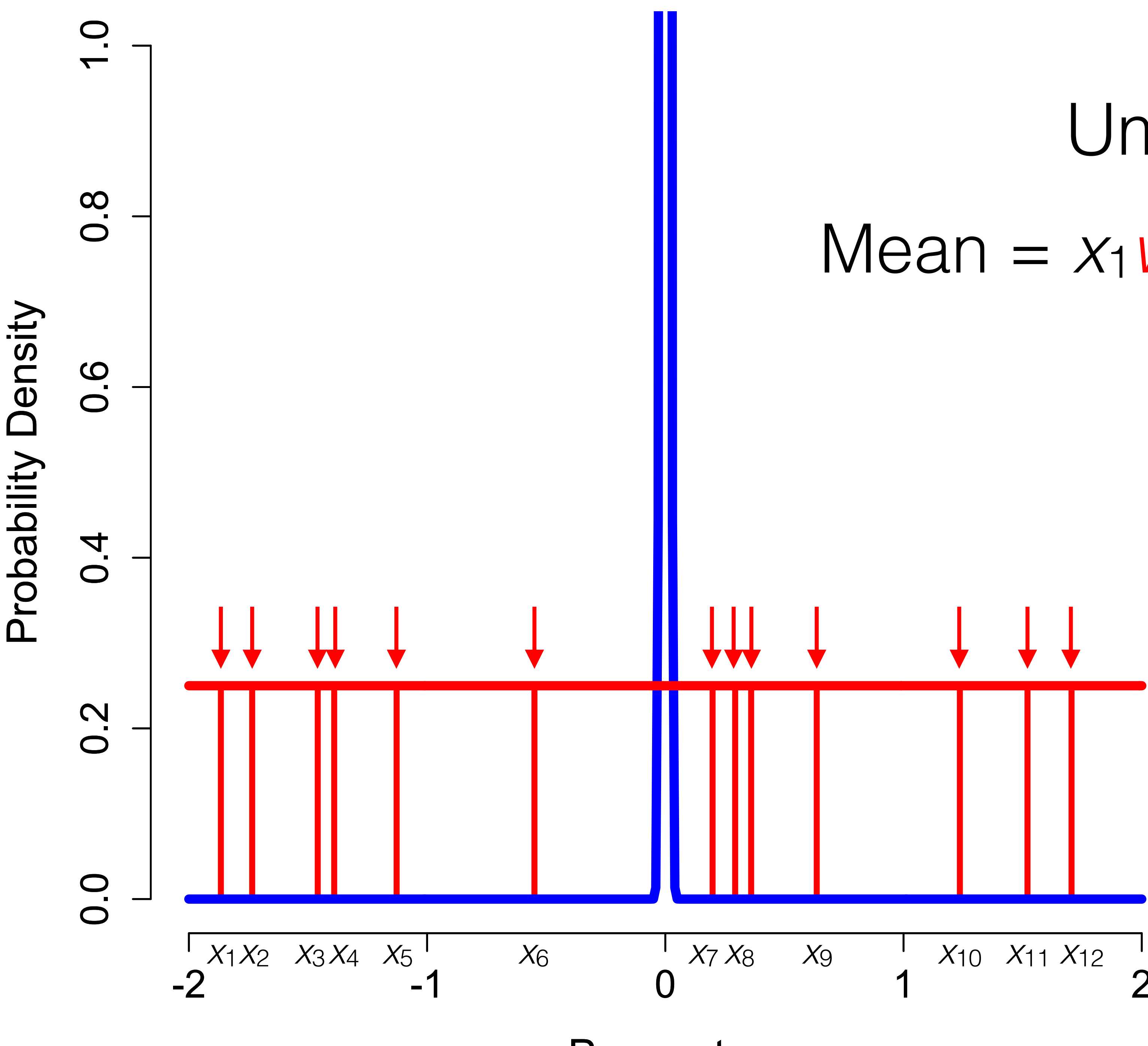
Mean = $x_1 w_1 + \ldots + x_6 w_6 + \ldots + x_{12} w_{12}$





Monte Carlo Methods

Importance Sampling



Ummmm..........

Mean = $x_1 w_1 + \ldots + x_6 w_6 + \ldots + x_{12} w_{12}$

A little bit

> 1

A lot

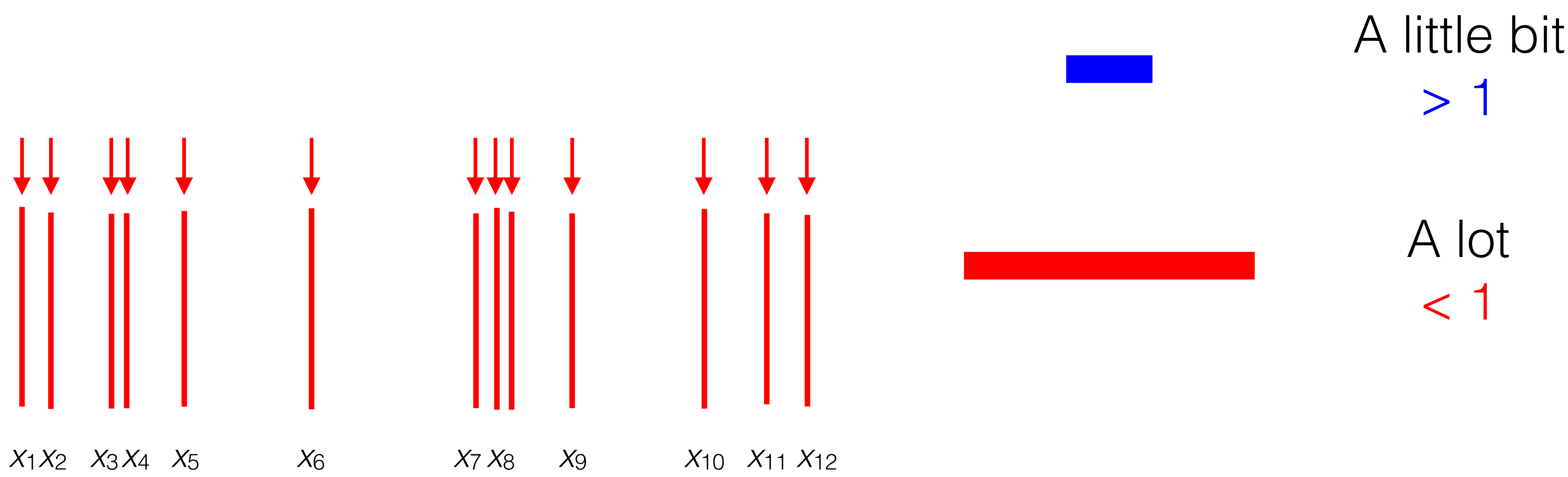
< 1

Monte Carlo Methods

Importance Sampling

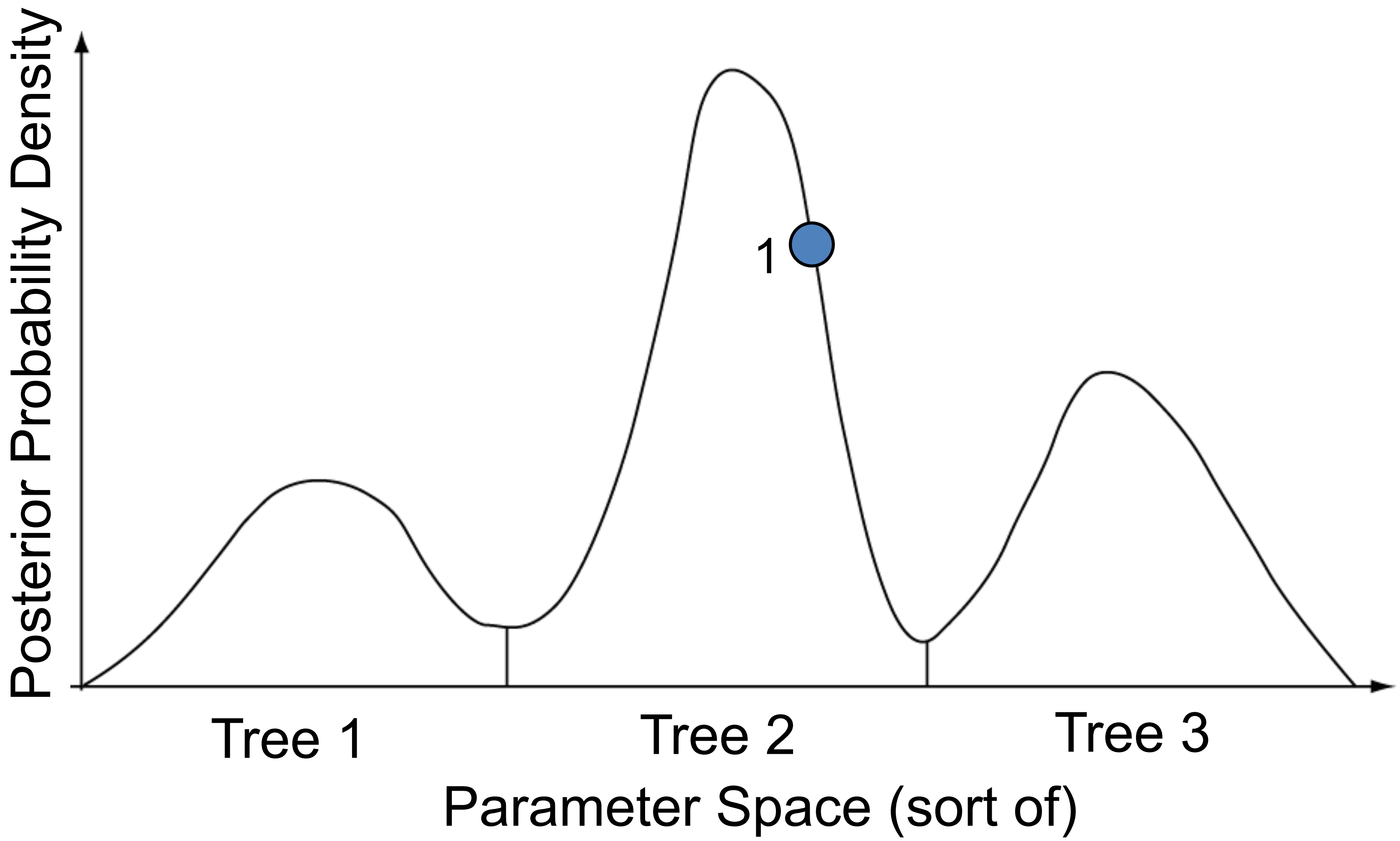
Does this look like our
Distribution of Interest?

Mean = $x_1 w_1 + \ldots + x_6 w_6 + \ldots + x_{12} w_{12}$



Markov chain Monte Carlo

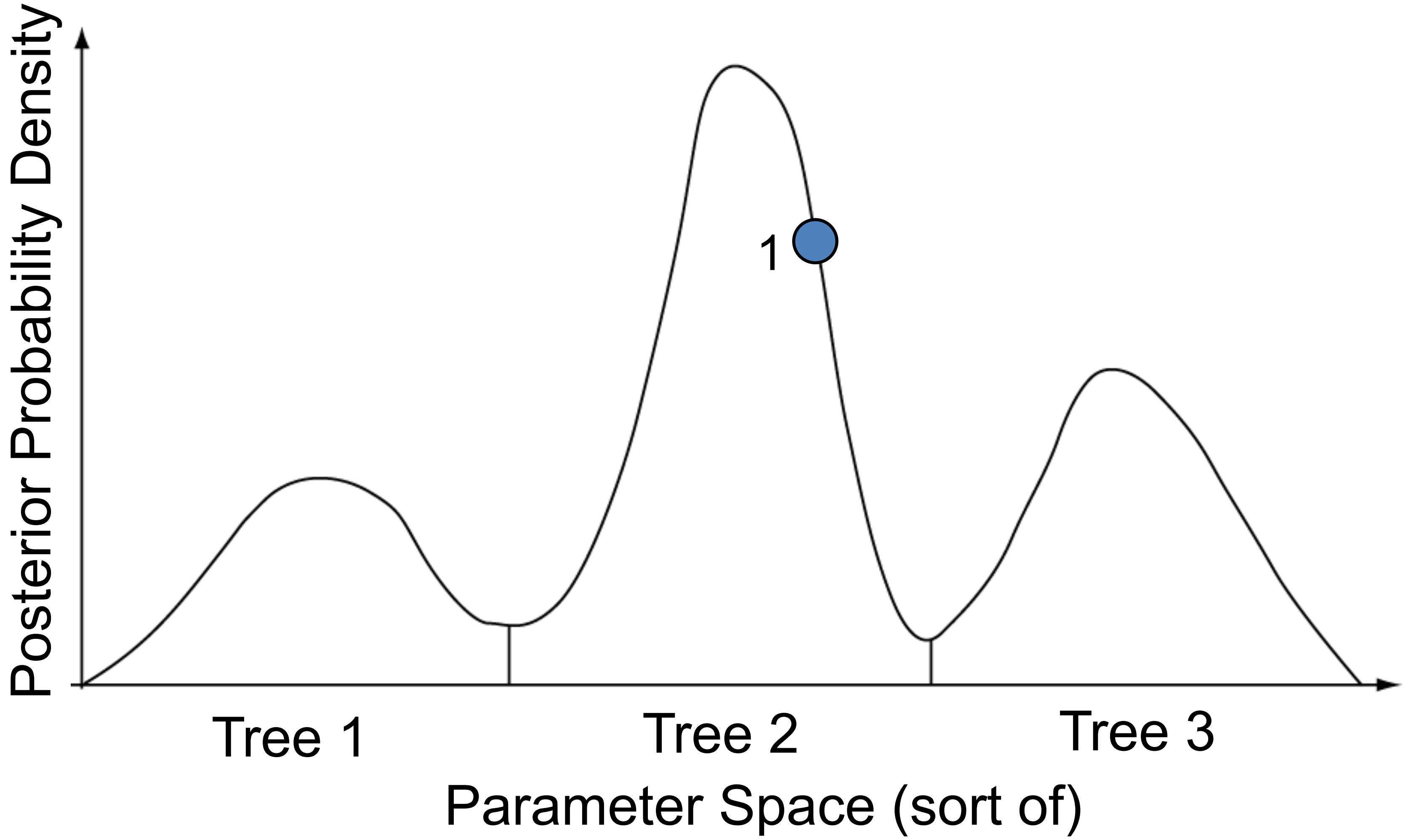
1. Start at an arbitrary point



This slide "borrowed" from F. Ronquist

Markov chain Monte Carlo

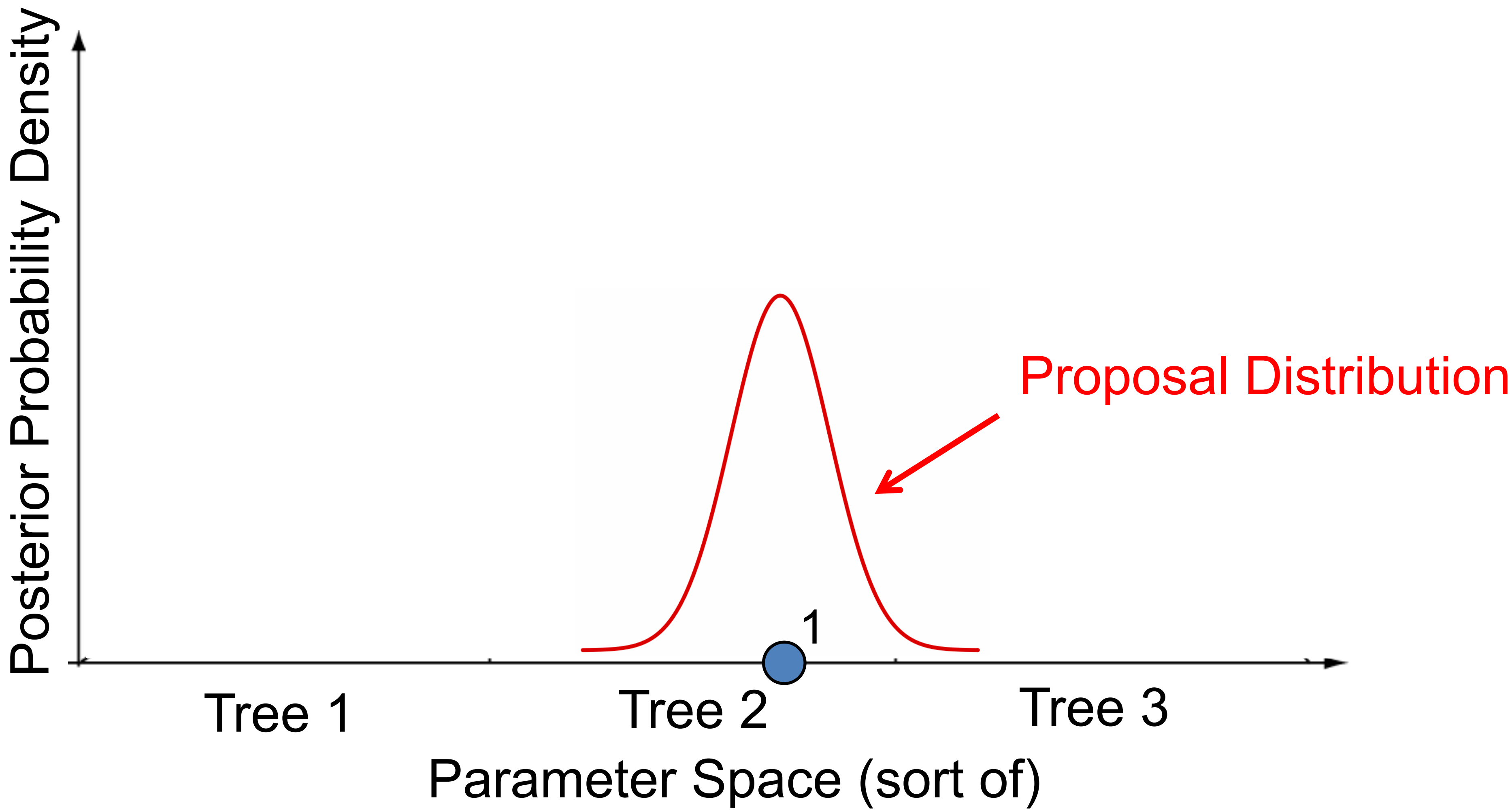
1. Start at an arbitrary point
2. Make a small random move



This slide "borrowed" from F. Ronquist

Markov chain Monte Carlo

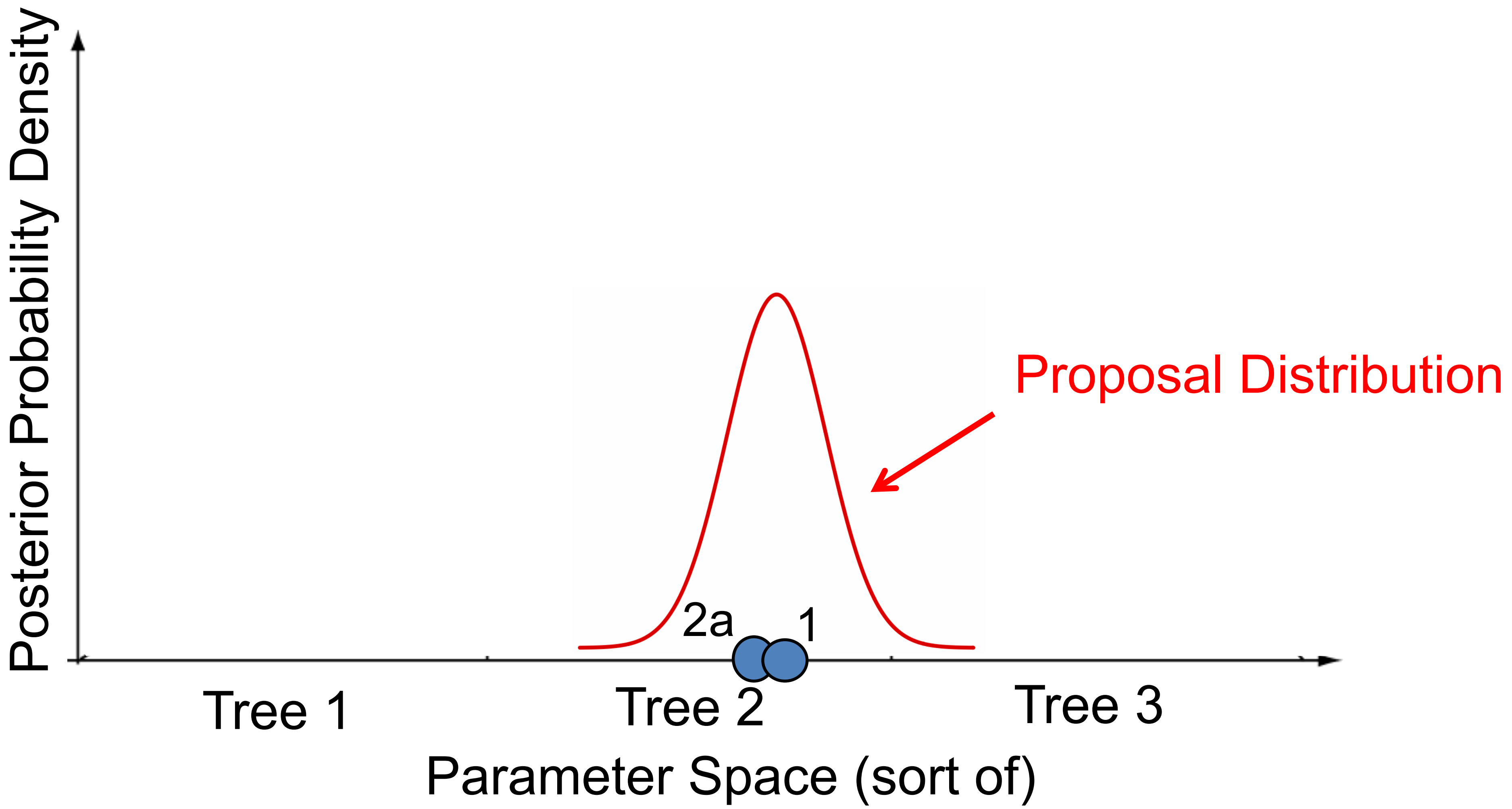
1. Start at an arbitrary point
2. Make a small random move



This slide "borrowed" from F. Ronquist

Markov chain Monte Carlo

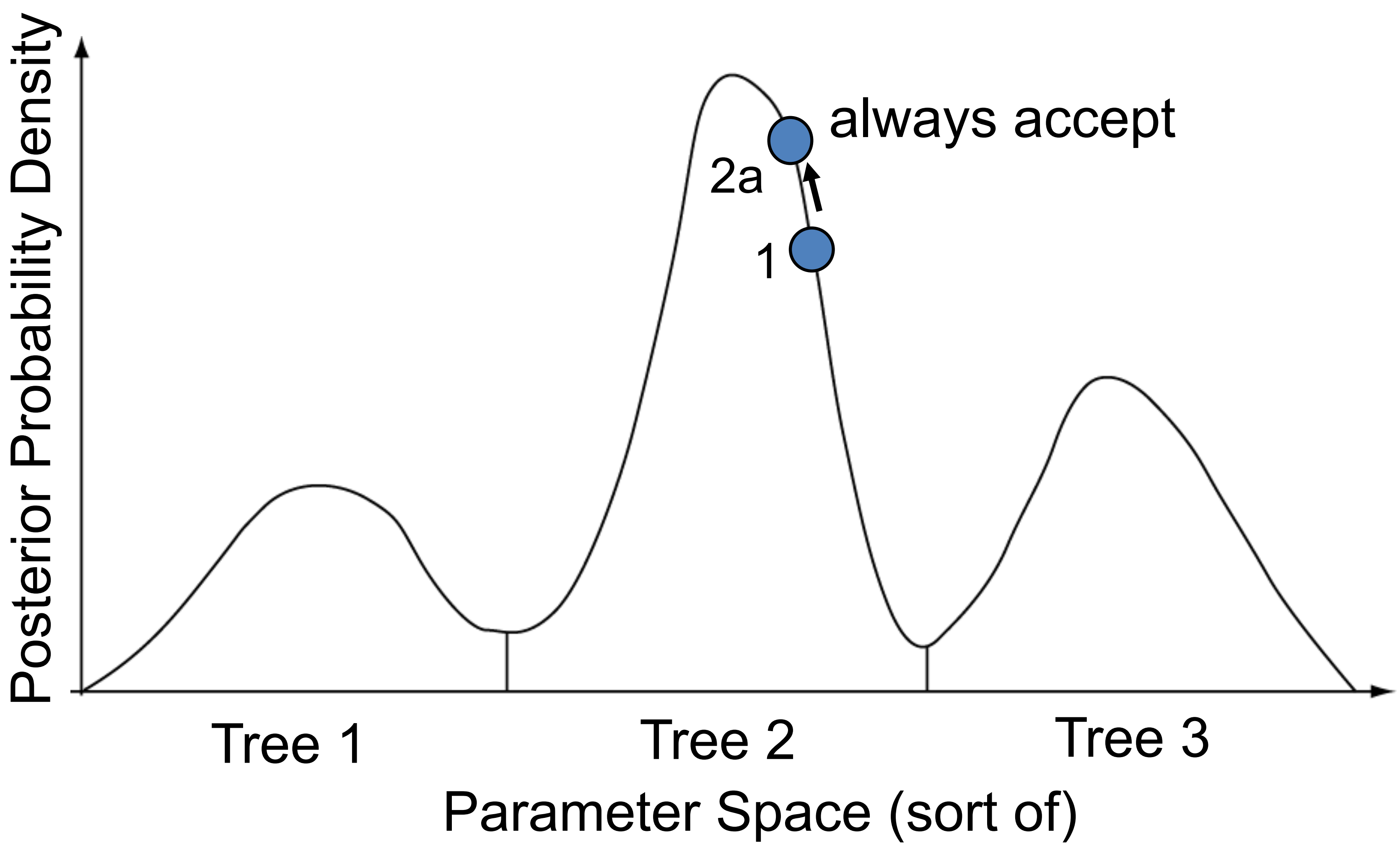
1. Start at an arbitrary point
2. Make a small random move



This slide "borrowed" from F. Ronquist

Markov chain Monte Carlo

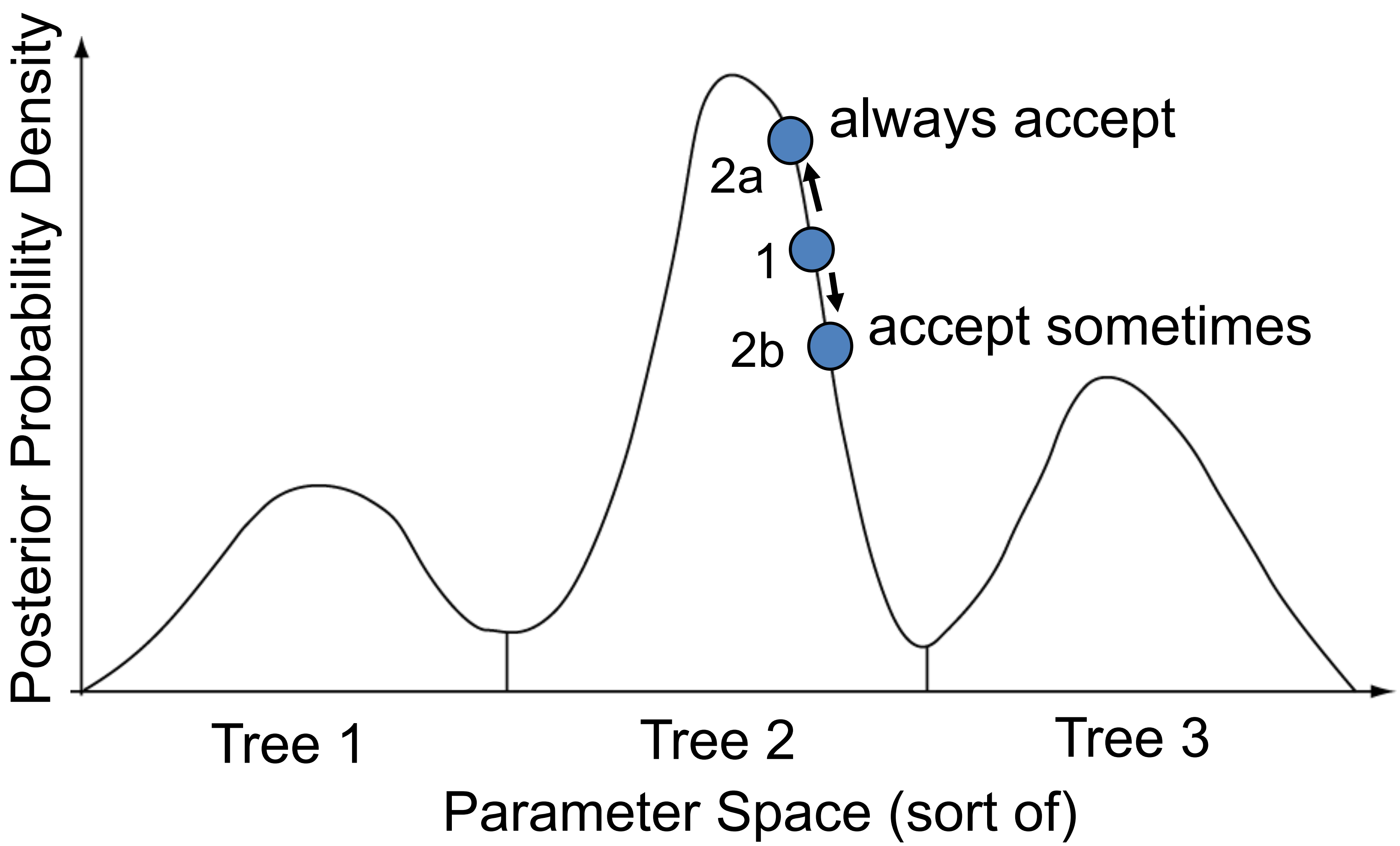
1. Start at an arbitrary point
2. Make a small random move
3. Calculate posterior density ratio (*r*) of new state to old state:
   a) r > 1 -> new state accepted
   b) r < 1 -> new state accepted with probability *r*. If new state not accepted, stay in the old state



This slide "borrowed" from F. Ronquist

Markov chain Monte Carlo

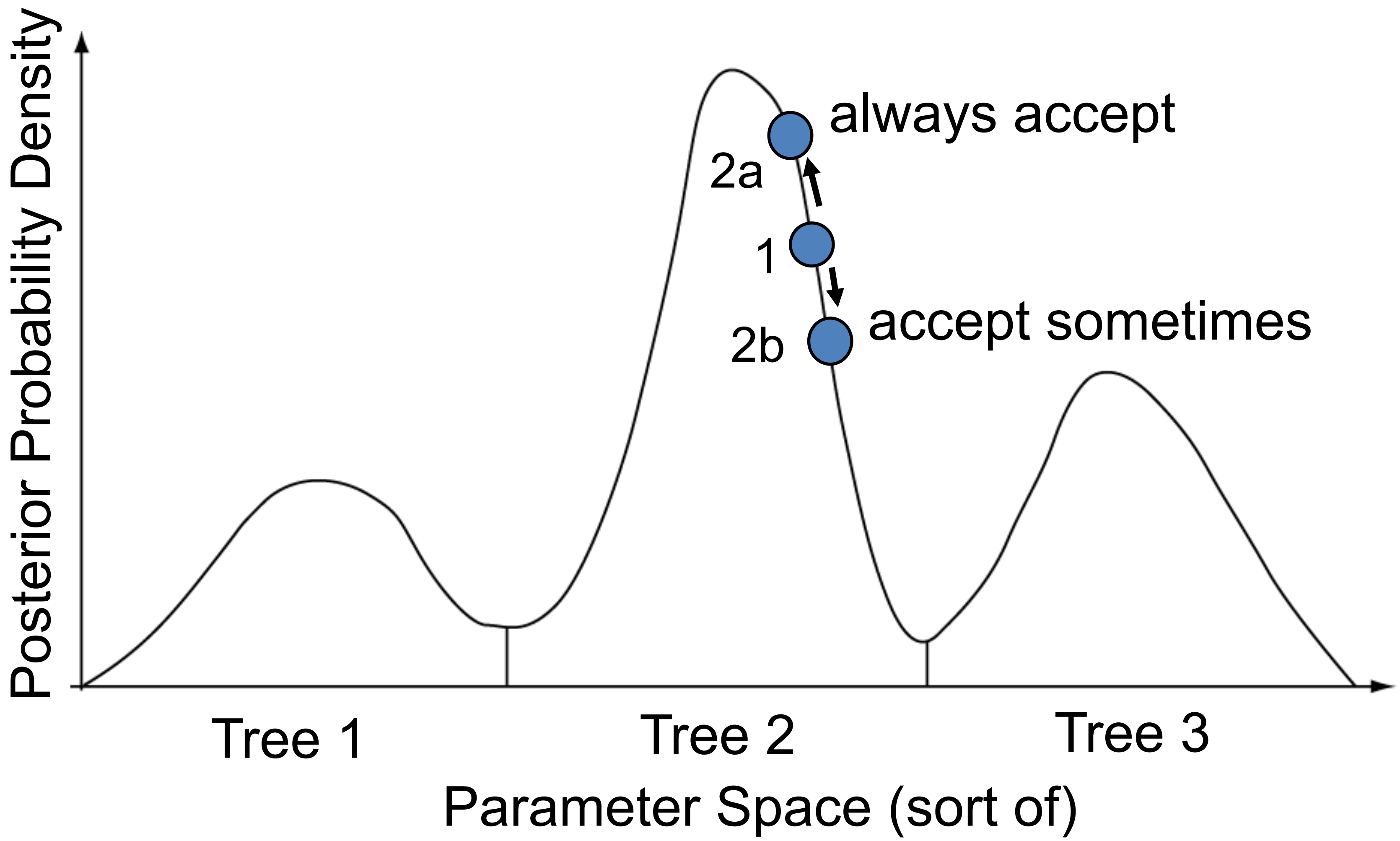
1. Start at an arbitrary point
2. Make a small random move
3. Calculate posterior density ratio (*r*) of new state to old state:
   a) r > 1 -> new state accepted
   b) r < 1 -> new state accepted with probability *r*. If new state not accepted, stay in the old state



This slide "borrowed" from F. Ronquist

Markov chain Monte Carlo

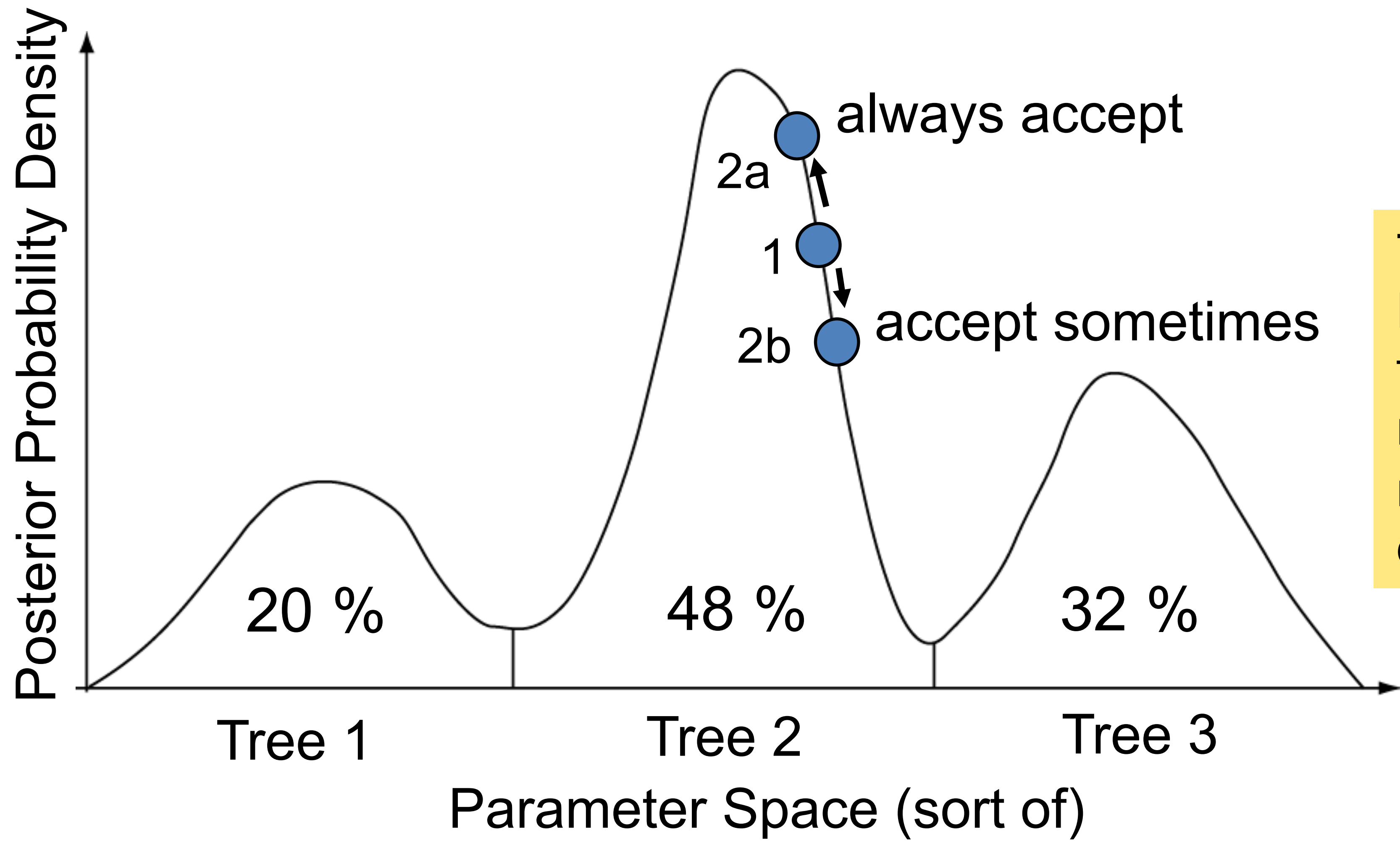
1. Start at an arbitrary point
2. Make a small random move
3. Calculate posterior density ratio (*r*) of new state to old state:
   a) r > 1 -> new state accepted
   b) r < 1 -> new state accepted with probability *r*. If new state not accepted, stay in the old state
4. Go to step 2 a BUNCH (x 10,000's – x 10,000,000's)



This slide "borrowed" from F. Ronquist

Markov chain Monte Carlo

1. Start at an arbitrary point
2. Make a small random move
3. Calculate posterior density ratio (*r*) of new state to old state:
   a) r > 1 -> new state accepted
   b) r < 1 -> new state accepted with probability *r*. If new state not accepted, stay in the old state
4. Go to step 2 a BUNCH (x 10,000's – x 10,000,000's)



The proportion of time the MCMC procedure samples from a particular parameter region is an estimate of that region's posterior probability density

This slide "borrowed" from F. Ronquist

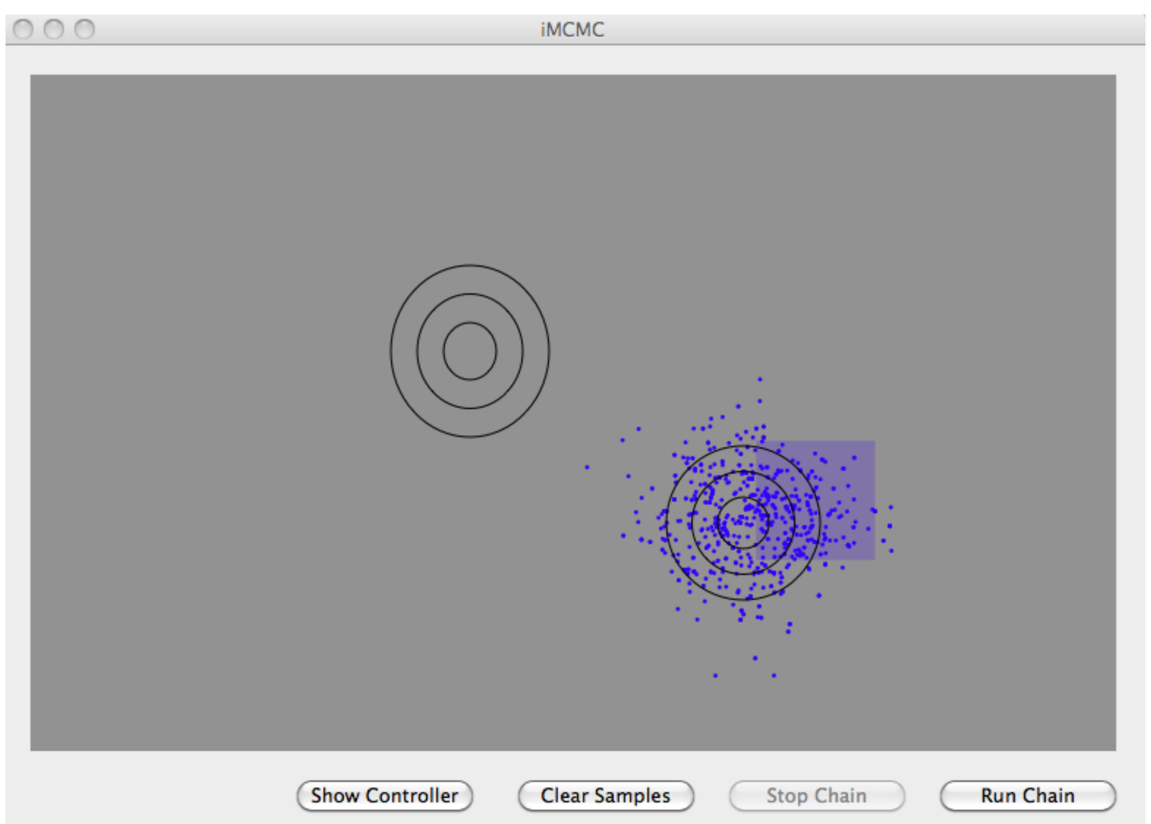
Toy MCMC Software

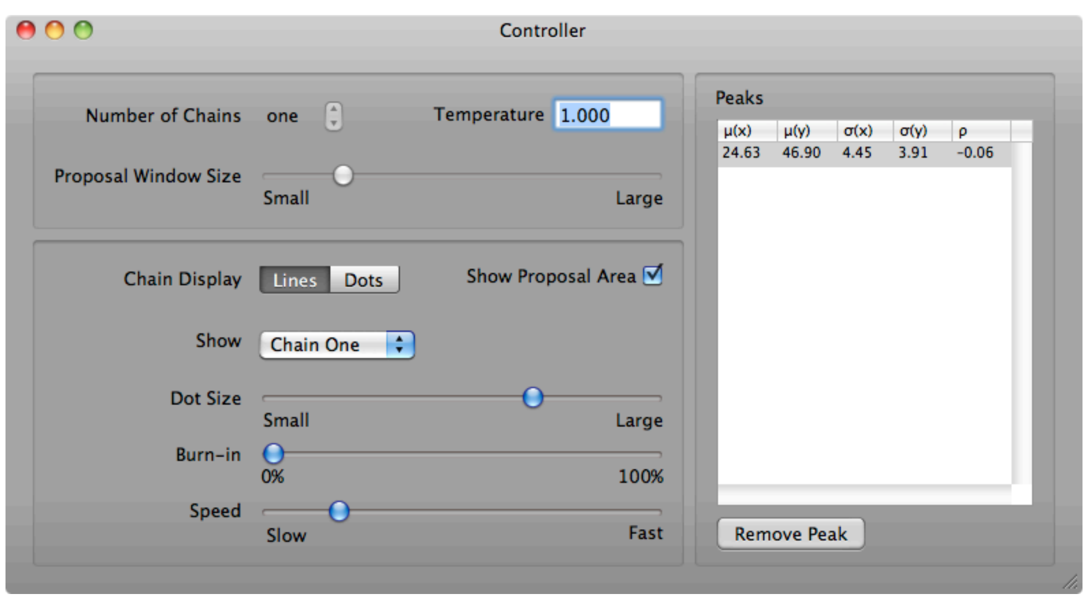
MCMC Robot (Paul Lewis; PC & iOS)

http://www.mcmcrobot.org

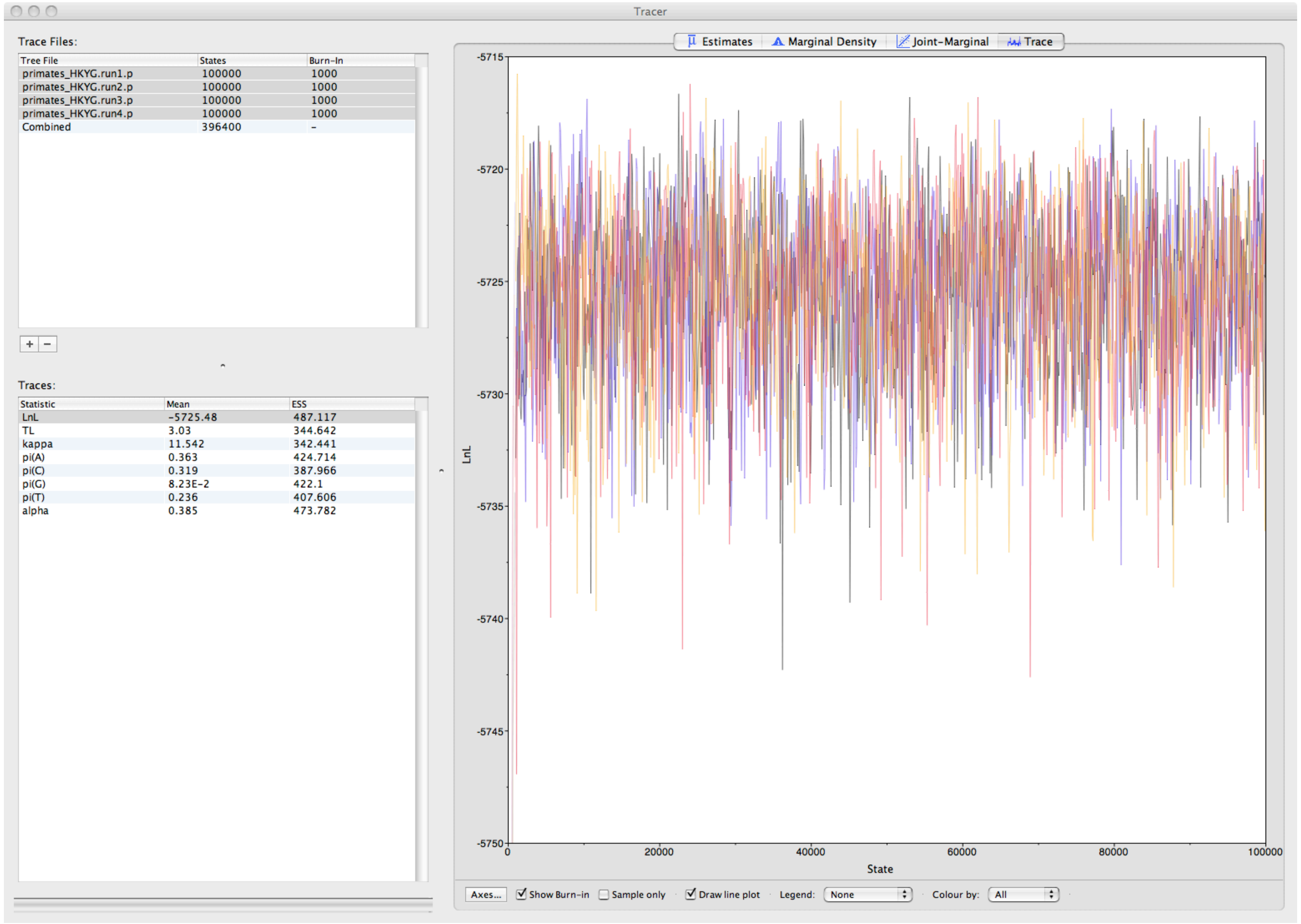
iMCMC (John Huelsenbeck; Mac)

http://cteg.berkeley.edu/software/huelsenbeck/McmcApp.zip





Convergence of Scalars - Tracer

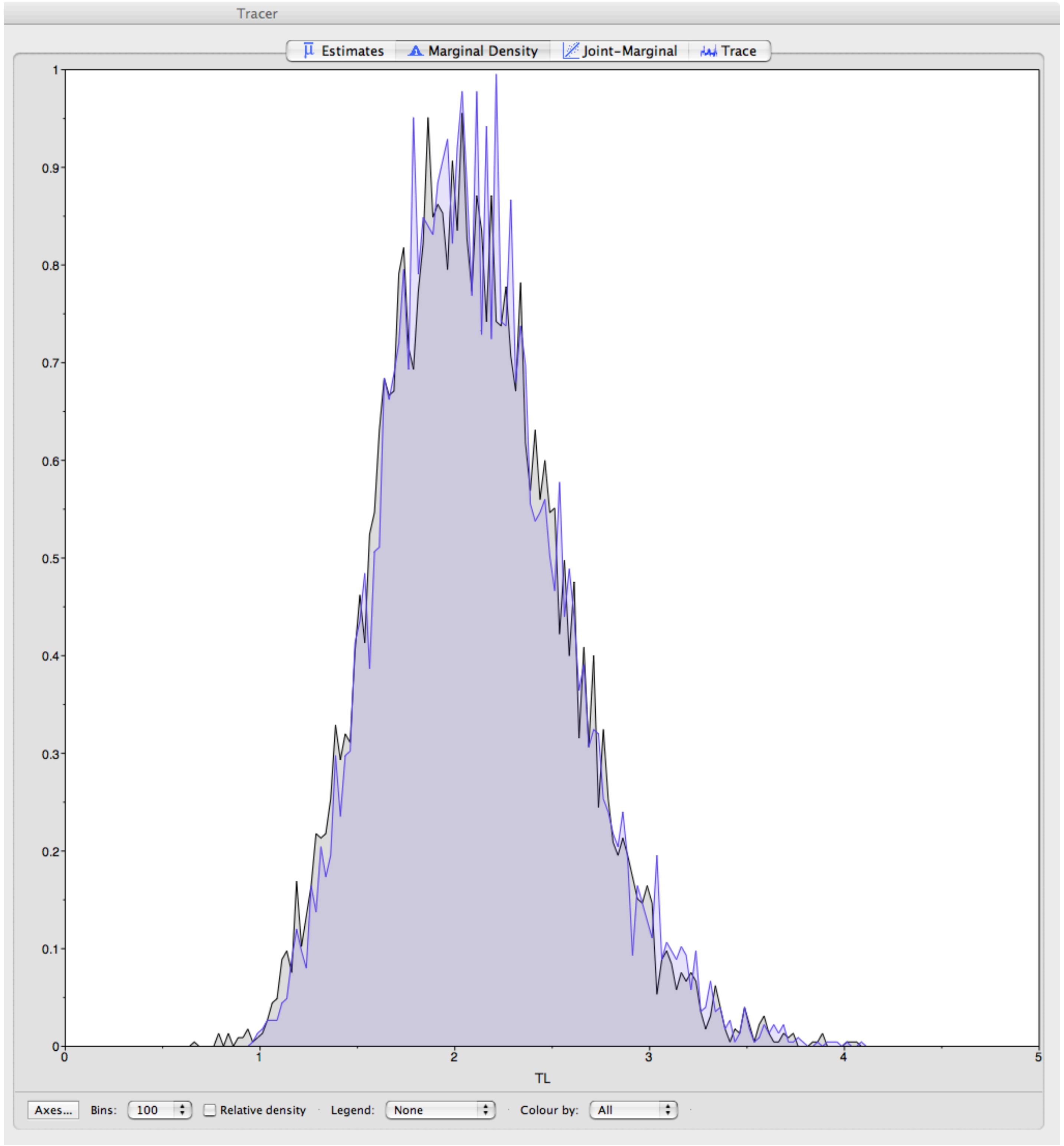


Running on Empty

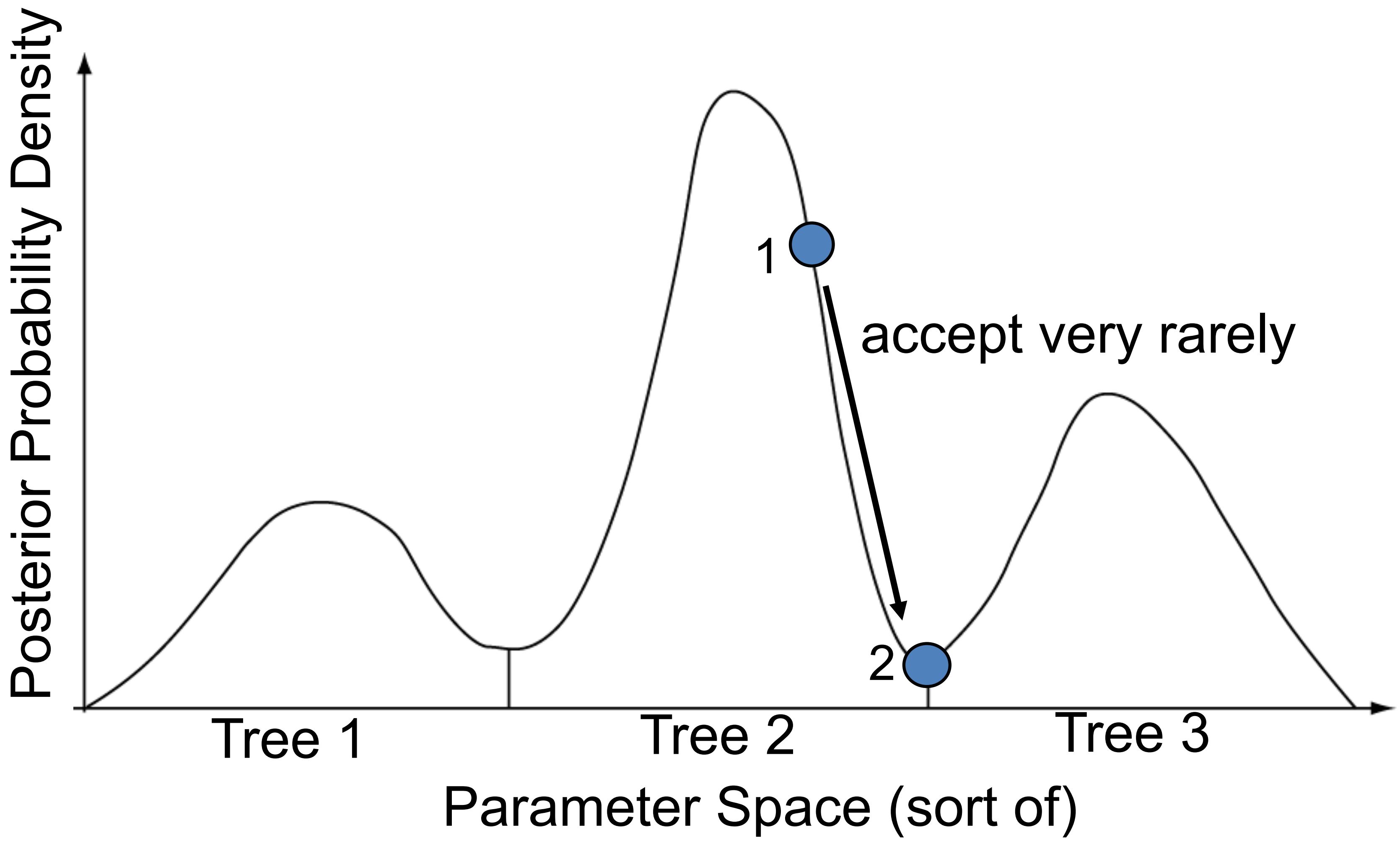
```
#NEXUS
begin data;
dimensions ntax=12 nchar=5;
format datatype=dna interleave=no gap=- missing=?;
matrix
Tarsius_syrichta    ?????
Lemur_catta         ?????
Homo_sapiens        ?????
Pan                 ?????
Gorilla             ?????
Pongo               ?????
Hylobates           ?????
Macaca_fuscata      ?????
M_mulatta           ?????
M_fascicularis      ?????
M_sylvanus          ?????
Saimiri_sciureus    ?????
;
end;
```

Or now in MrBayes:

mcmc data=no



Metropolis Coupling

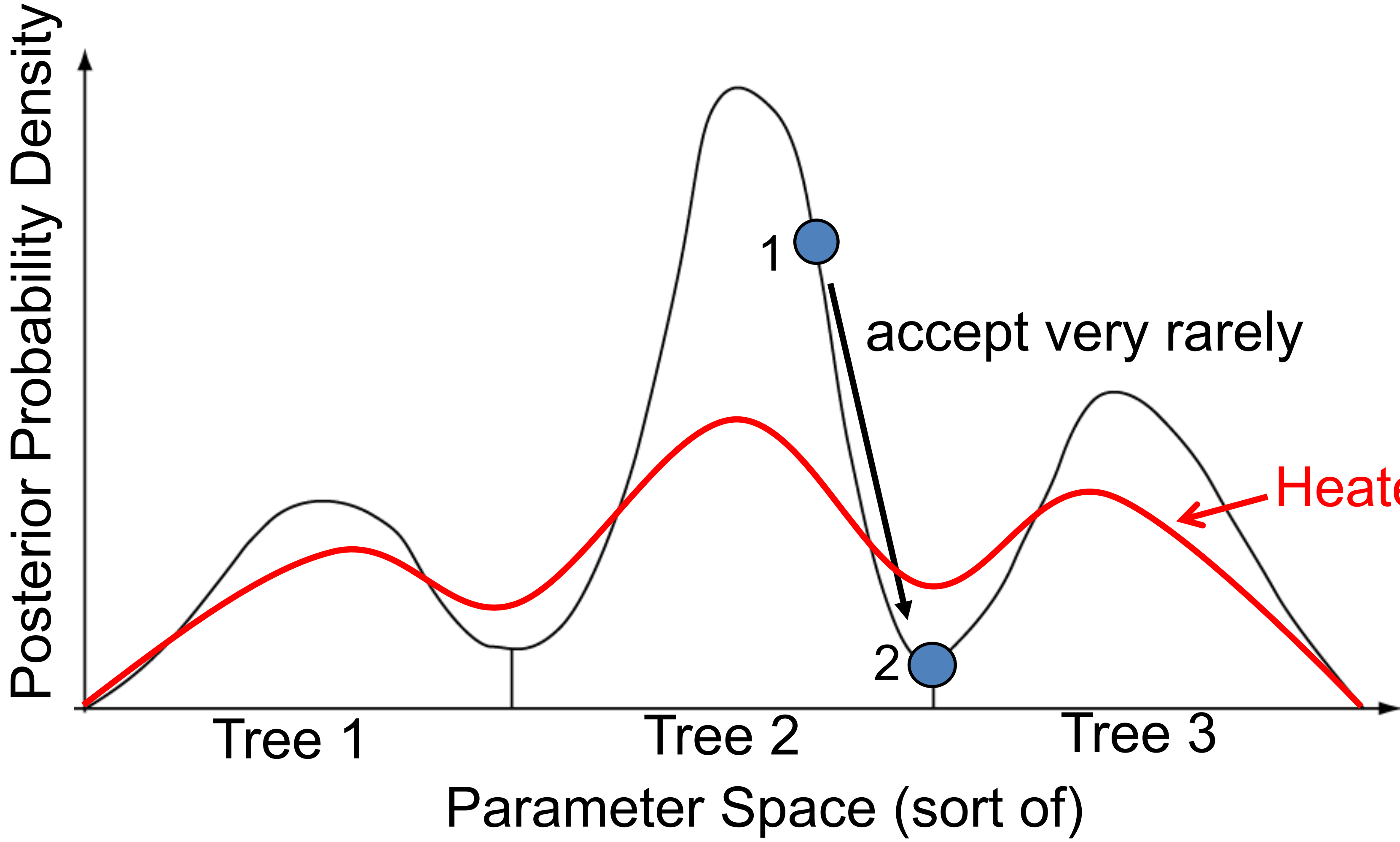


This slide "borrowed" from F. Ronquist

Metropolis Coupling

- Same rules as regular MCMC, but now there are multiple chains with different ‘temperatures’.
- ‘Heated’ chains sample a ‘melted’ version of the posterior
- Only difference is that heated chains raise the ratio of posterior densities to (1-temp) when deciding whether to accept a move.

$r^{(1\text{-temp})}$ approaches 1 (flat surface) as temp. increases

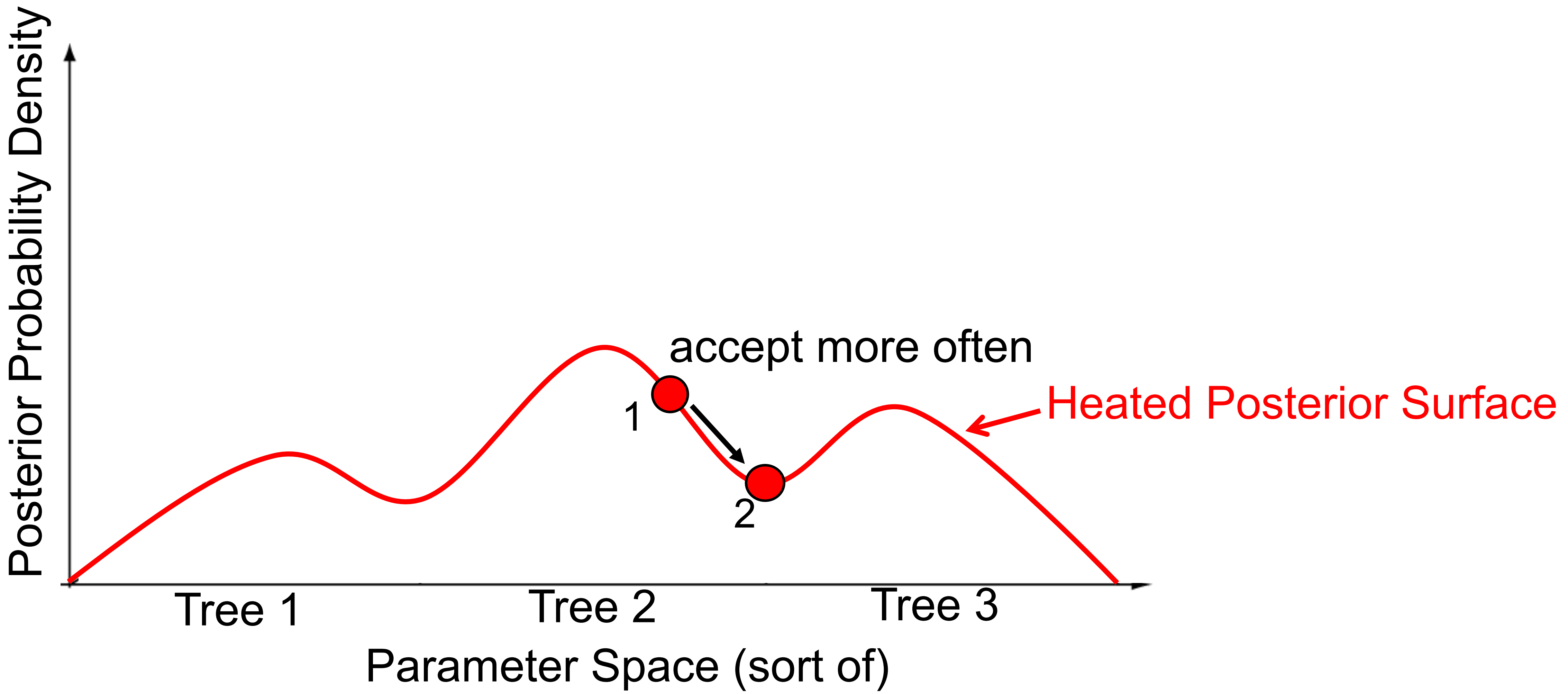


This slide “borrowed” from F. Ronquist

Metropolis Coupling

- Same rules as regular MCMC, but now there are multiple chains with different 'temperatures'.
- 'Heated' chains sample a 'melted' version of the posterior
- Only difference is that heated chains raise the ratio of posterior densities to (1-temp) when deciding whether to accept a move.

$r^{(1\text{-temp})}$ approaches 1 (flat surface) as temp. increases



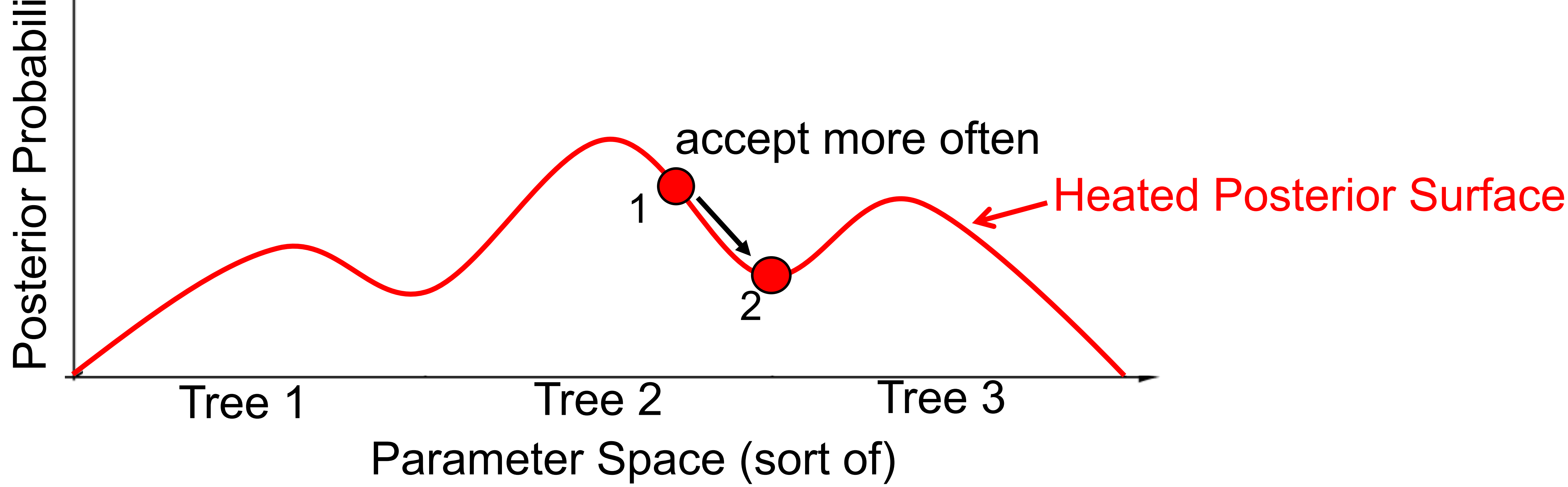
This slide "borrowed" from F. Ronquist

Metropolis Coupling

- Same rules as regular MCMC, but now there are multiple chains with different 'temperatures'.
- 'Heated' chains sample a 'melted' version of the posterior
- Only difference is that heated chains raise the ratio of posterior densities to (1-temp) when deciding whether to accept a move.

$r^{(1\text{-temp})}$ approaches 1 (flat surface) as temp. increases

- Samples only recorded from the 'cold' chain
- Heated chains are 'scout's. Occasionally propose to trade places with the cold chain.



This slide "borrowed" from F. Ronquist