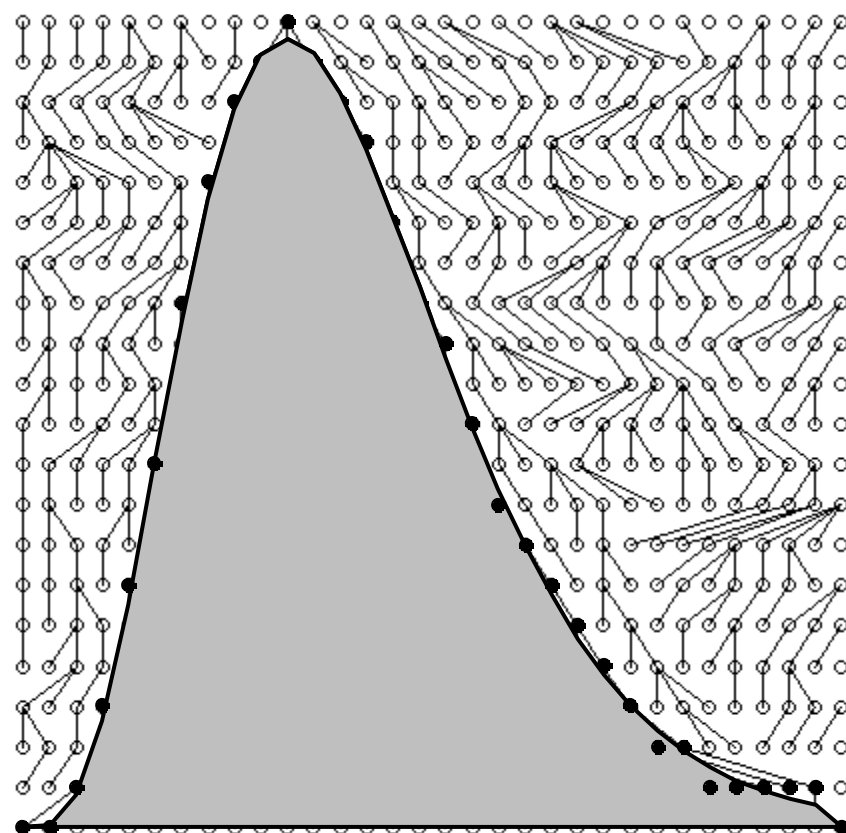


Complex scenarios and

# Bayesian Coalescence

Bayesian Coalescence



# Phylochronology

φυλιοχρονοιοδλ

- Reconstruct the history of related individuals
- Use a comparative hypothesis framework

Site 1



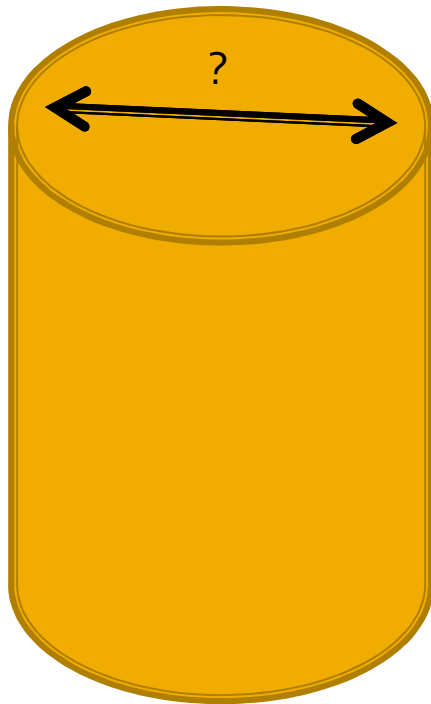
High genetic diversity  
(7 haplotypes)

Site 2

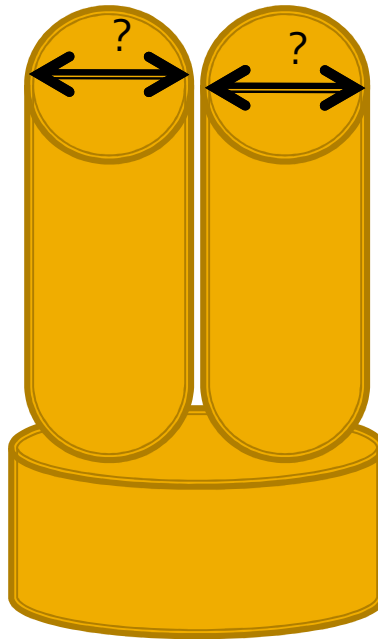


Low genetic diversity  
(3 haplotypes)

# Alternative Hypotheses



Hypothesis 1:  
One large  
population

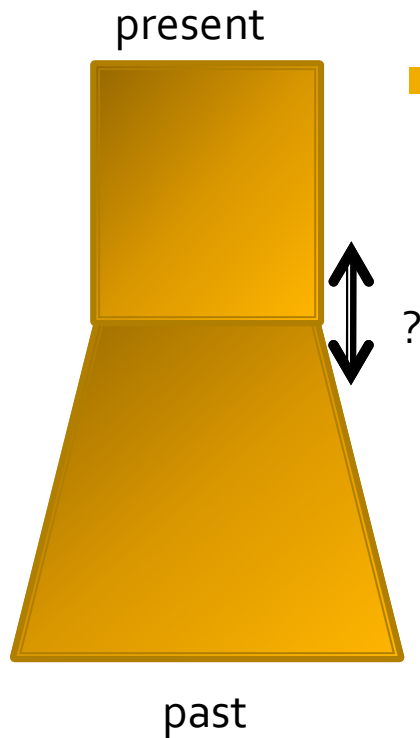


Hypothesis 2: Two  
long-separated  
populations

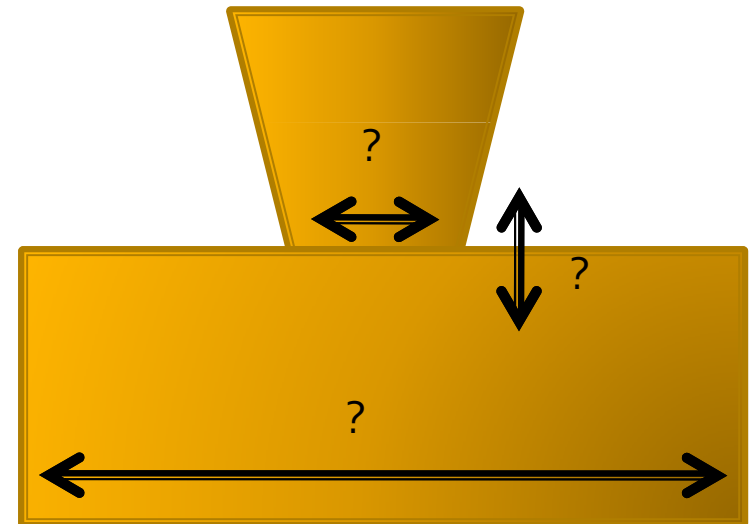
## Goals

- 1) Find best parameters of these models
- 2) Compare the likelihoods of the two models

# For example...



- When did bison start to decline? (~Shapiro *et al*, 2004)



- Did tuco-tucos go through a bottleneck? (Chan *et al*, 2006)

# Bayesian Black Boxes

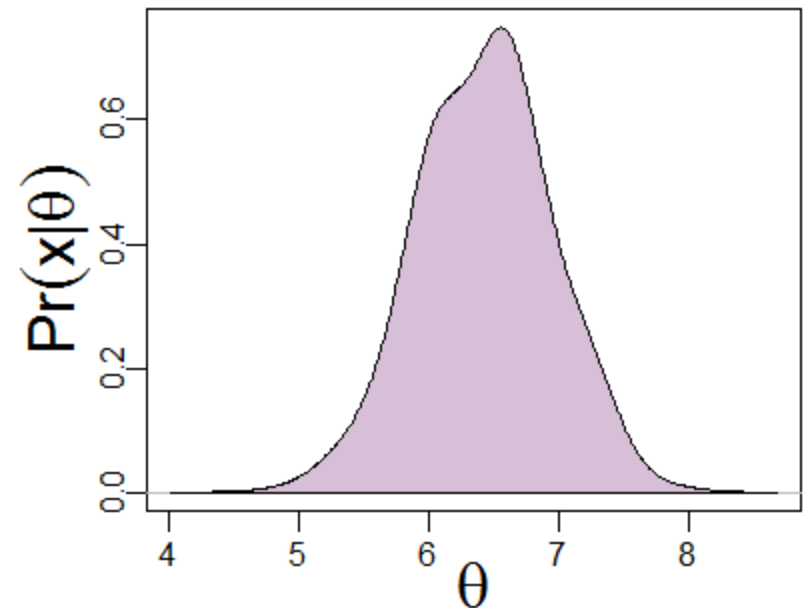
Program



Can  $\theta$  = some  
number? Yes or No

Frequentist

Bayesian



# Rejection Method (Pritchard *et al.* 1999)

ΚΕΙΜΕΝΟ ΜΕΤΡΟΦ (ΠΡΙΤΤΧΑΡΔ ΕΤ ΑΛ. 1999)

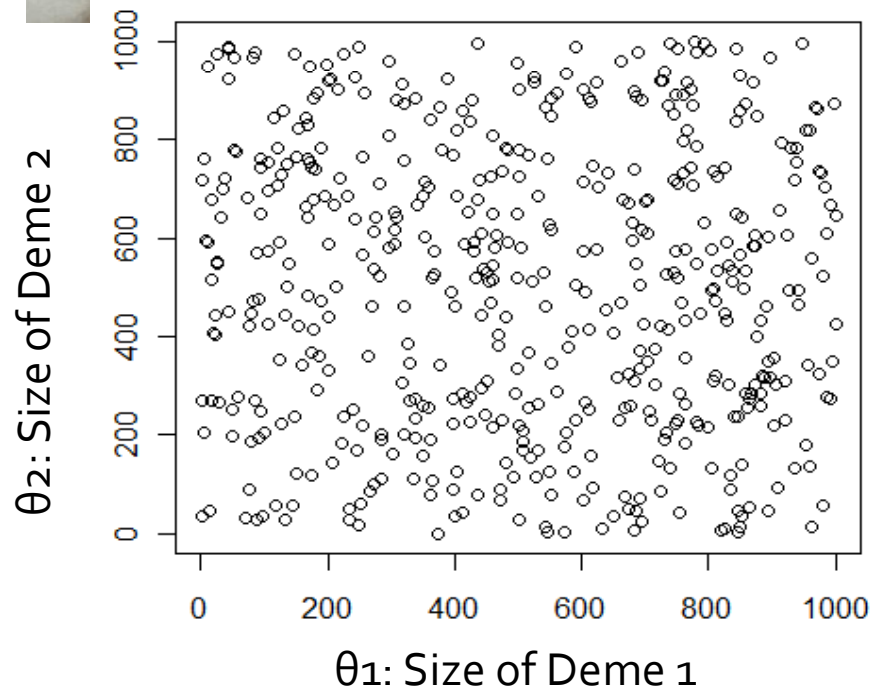
1. Set up your model
2. Try a lot of different  $\theta$  values
3. Calculate the difference between your actual data and the simulated data
4. Accept simulations where the difference is sufficiently small

# Rejection Method

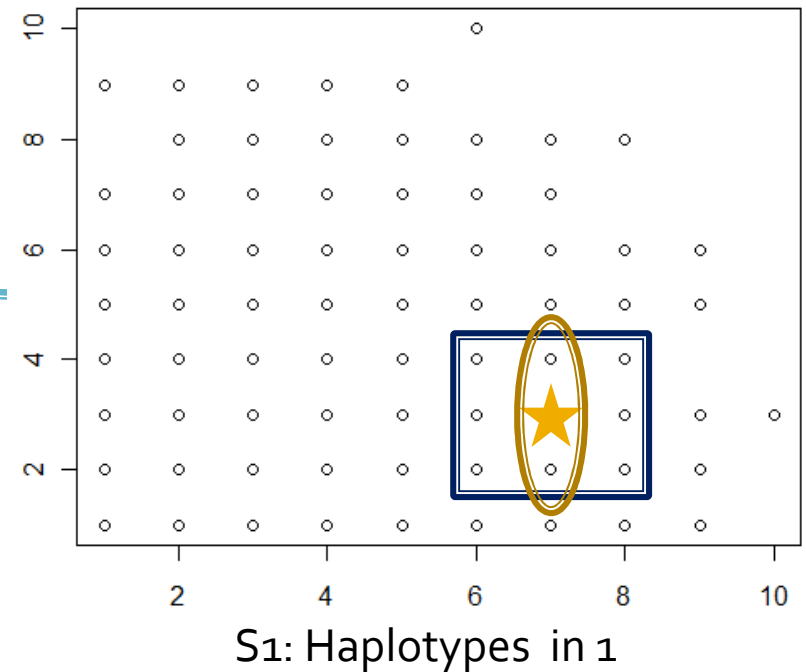
REJECTION METHOD



MODEL: Two populations size  $\theta_1$  and  $\theta_2$ , both  $\sim U(1,1000)$ ; split 500ya



S2: Haplotypes in 2

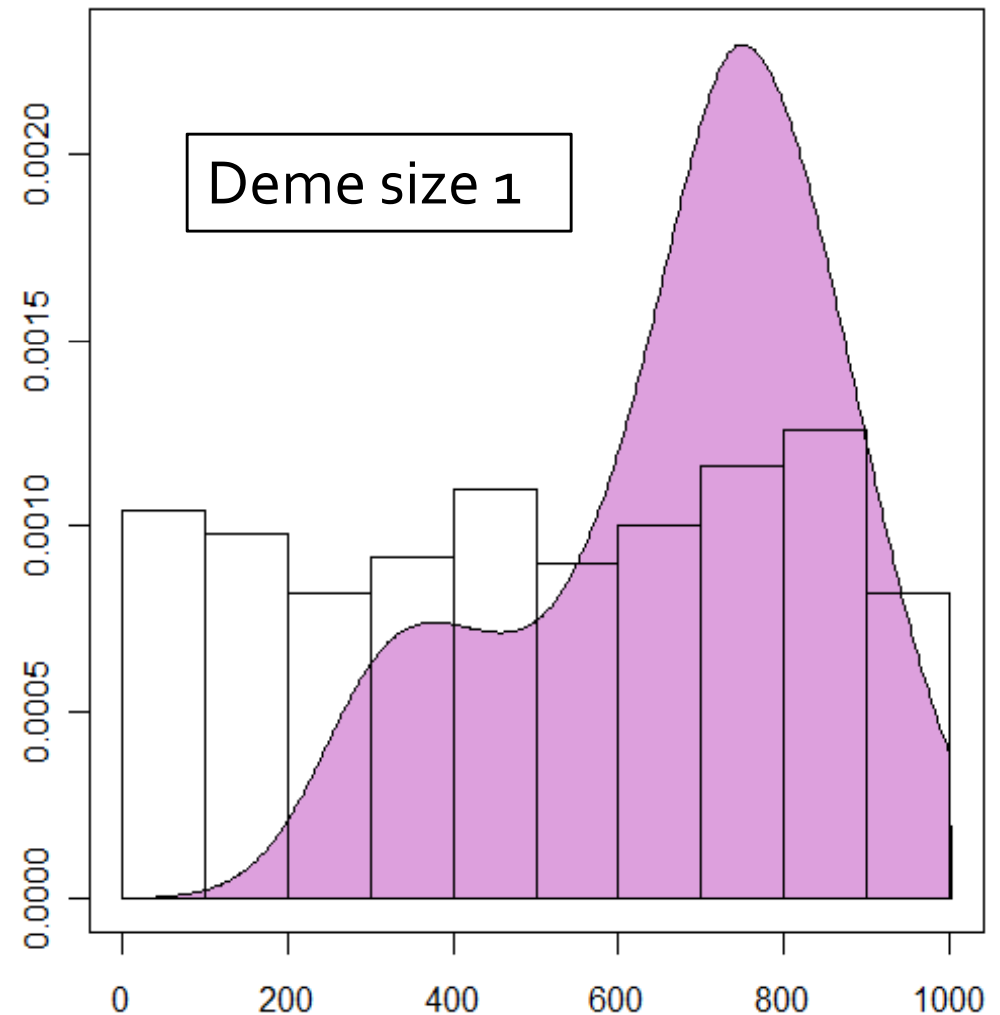




# Rejection method

rejection method

- Leaves you with a list of “acceptable” parameter values
- Either plot directly
- Or weight by “how” acceptable they were



# SIR (Rubin 1988)

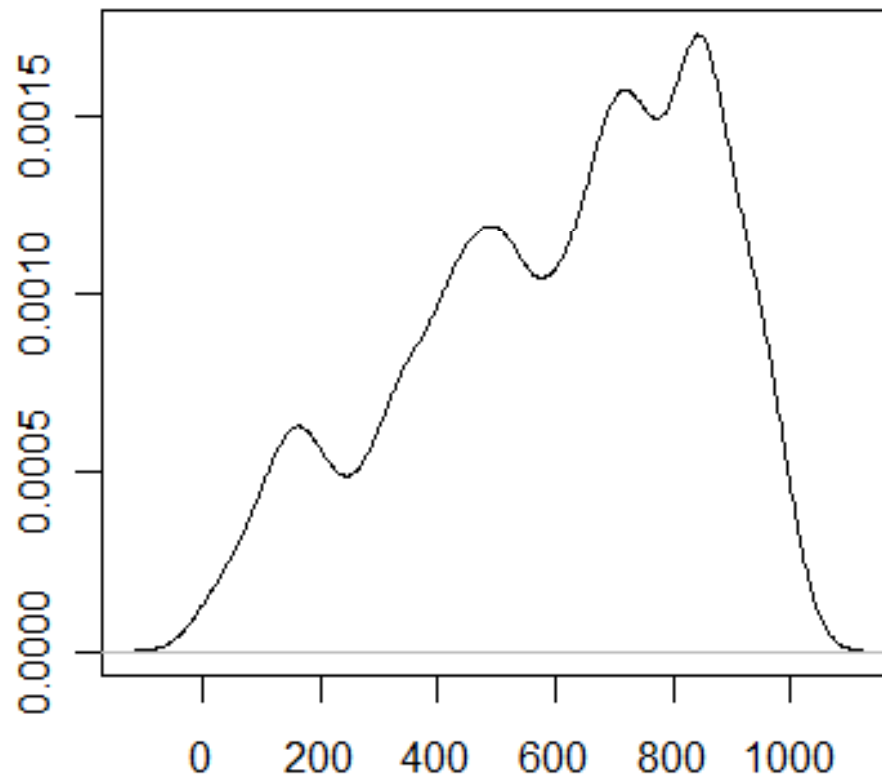
ΣΙΚ (ΚΟΡΙΝ ΤΘΡΡ)

- Sampling / Importance Resampling
  - Generate many values of  $\theta$  from priors:  $g(\theta)$
  - Get likelihood of your data for every  $\theta_i$ :  $f(\theta)$
  - Weight each point:  $\omega_i = \frac{f(\theta_i)}{g(\theta_i)}$
  - Resample  $\theta$  with weight  $\omega$

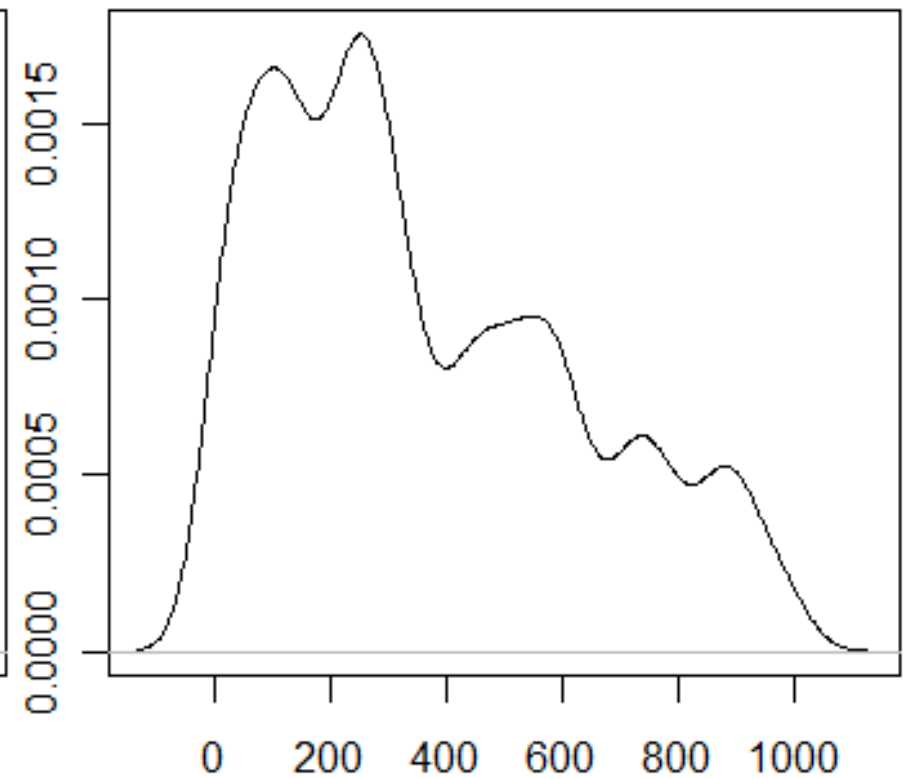
# Example

FX9wb16

Posterior 1

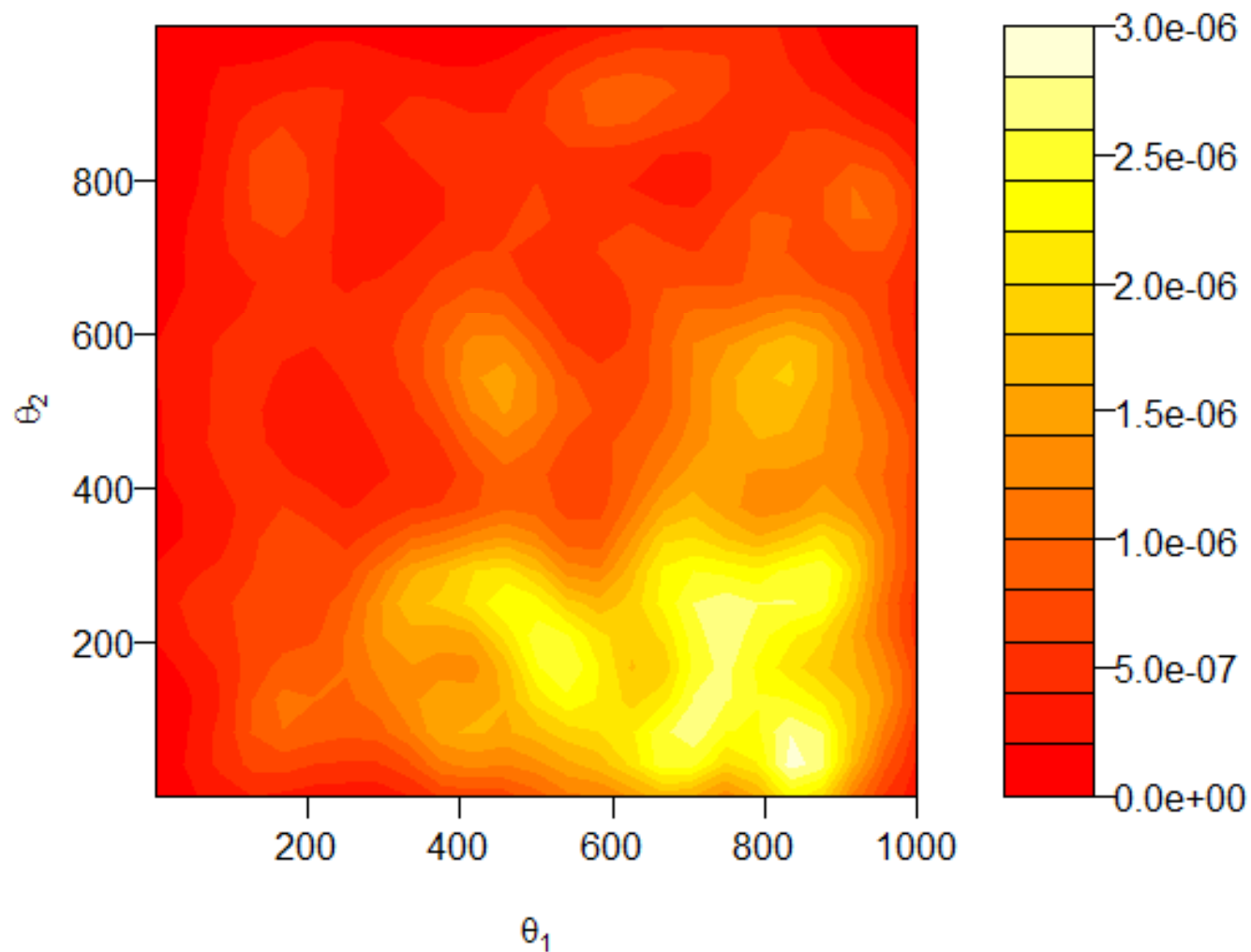


Posterior 2



# Example

FX9wb16



# Model Selection

- Open question as to what constitutes the best method
- Two main options

$$AIC = 2p - \text{Log}L(M, D)$$

$$BIC = 2p \log k - \text{Log}L(M, D)$$

# Likelihood

## Likelihood

- Run the program again, with priors drawn from posterior distributions
- Likelihoods are proportional to the number of acceptances of your data
  - Acceptance can be determined in several ways (eg Belle *et al* 2006). We will accept simulations within  $x\%$  of the range of values (board)
- Calculate  $2p \cdot \log(\% \text{ acceptances})$  for AIC, or  $2p \cdot \log(k) \cdot \log(\% \text{ acceptances})$  for BIC

# Your turn!

