

Course overview

- Tuesday lecture
 - Those not presenting turn in short review of a paper using the method being discussed
- Thursday computer lab
 - Turn in short write up from previous computer lab.
 - Some methods take some computational time, so bring something else along.

Maximum Likelihood, MCMC, and Bayes Theorem

Pop Quiz

Maximum Likelihood, MCMC, and Bayes Theorem

**Maximum likelihood
methods find the parameters
most likely to produce the
data observed given a
specific model.**

Maximum likelihood methods find the parameters most likely to produce the data observed given a specific model.

The likelihood (L) is the probability of the data given the hypothesis (or parameter value).

$$L = P(\text{data} | \text{hypothesis})$$

We will use ML for variety of calculations

- What is the ML estimate of d_N/d_S for a codon alignment?
- What parameters optimize conserved versus neutral regions?
- What number of populations maximizes the likelihood?

What is maximum likelihood?

Comparison to probability theory:

Probability of # heads in 5 coin tosses

Heads	Prob.
0	.03
1	.16
2	.31
3	.31
4	.16
5	.03

$$P(x) = (n! / (n-x)!) p^x q^{n-x}$$

Same calculation for coins with different bias.

	Bias of coin towards Heads				
Heads	.1	.3	.5	.7	.9
0	.59	.17	.03	.00	.00
1	.33	.36	.16	.03	.00
2	.07	.31	.31	.13	.01
3	.01	.13	.31	.31	.07
4	.00	.03	.16	.36	.33
5	.00	.00	.03	.17	.59

Same calculation for coins with different bias.

	Bias of coin towards Heads				
Heads	.1	.3	.5	.7	.9
0	.59	.17	.03	.00	.00
1	.33	.36	.16	.03	.00
2	.07	.31	.31	.13	.01
3	.01	.13	.31	.31	.07
4	.00	.03	.16	.36	.33
5	.00	.00	.03	.17	.59

Probabilities sum to 1

Same calculation for coins with different bias.

Observe 1 head in 5 coin tosses, what is the Maximum likelihood estimate for coin bias?

	Bias of coin towards Heads					
Heads	.1	.3	.5	.7	.9	← Want to determine bias from observed # heads
0	.59	.17	.03	.00	.00	
1	.33	.36	.16	.03	.00	
2	.07	.31	.31	.13	.01	
3	.01	.13	.31	.31	.07	
4	.00	.03	.16	.36	.33	
5	.00	.00	.03	.17	.59	

Probabilities sum to 1

The likelihood (L) is the probability of the data given the hypothesis (or parameter value).

$$L = P(\text{data} | \text{hypothesis})$$

Same calculation for coins with different bias.

	Bias of coin towards Heads					
Heads	.1	.3	.5	.7	.9	← Want to
0	.59	.17	.03	.00	.00	determine
1	.33	.36	.16	.03	.00	bias from
2	.07	.31	.31	.13	.01	observed
3	.01	.13	.31	.31	.07	# heads
4	.00	.03	.16	.36	.33	
5	.00	.00	.03	.17	.59	

Likelihoods do not sum to 1. Maximum is parameter that best fits the observed data.

The likelihood (L) is the probability of the data given the hypothesis (or parameter value).

$$L = P(\text{data} | \text{hypothesis})$$

Same calculation for coins with different bias.

	Bias of coin towards Heads					
Heads	.1	.3	.5	.7	.9	← Want to
0	.59	.17	.03	.00	.00	determine
1	.33	.36	.16	.03	.00	bias from
2	.07	.31	.31	.13	.01	observed
3	.01	.13	.31	.31	.07	# heads
4	.00	.03	.16	.36	.33	L = P(data bias)
5	.00	.00	.03	.17	.59	

**Likelihoods usually represented as $\ln(L)$,
So looking for least negative value.**

$$\ln(0.33) = -1.1$$

$$\ln(0.36) = -1.02$$

$$\ln(0.16) = -1.8$$

One use of maximum likelihood for phylogeny inference.

The likelihood (L) of a tree is the probability of the data given the tree and model (hypothesis or parameter value).

$$L = P(\text{data}|\text{tree})$$

Problem is that there are LOTS of possible trees (hypotheses).

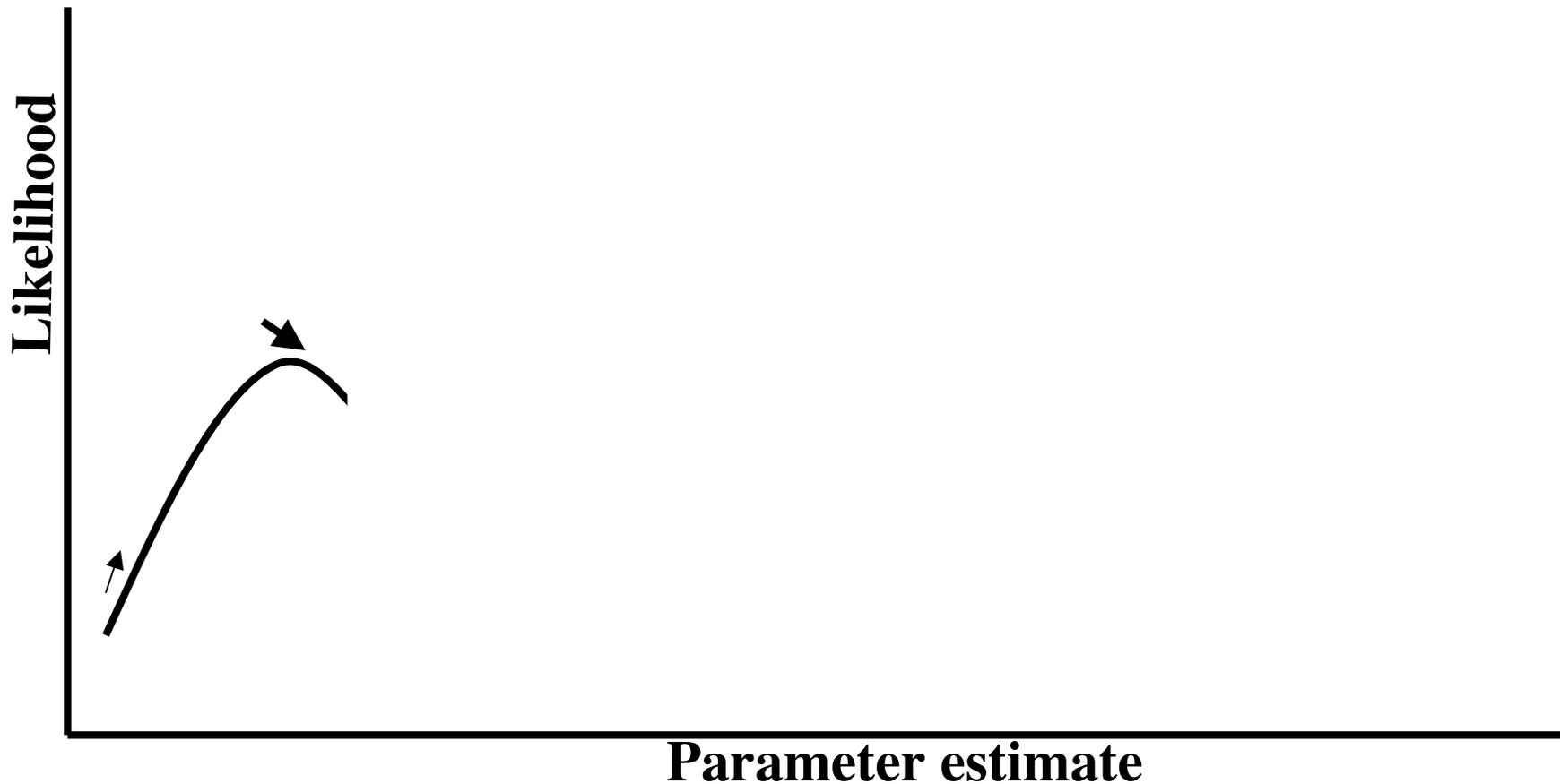
s	# Unrooted	#Rooted
3	1	3
4	3	15
5	15	105
6	105	945
7	945	10395
8	10395	135135
9	135135	2027025
10	2027025	34459425
20	2.2×10^{20}	8.2×10^{21}
50	2.8×10^{74}	2.7×10^{76}
100	1.7×10^{182}	3.4×10^{184}
1000	1.9×10^{2860}	3.8×10^{2863}

How do we calculate/estimate
likelihoods?

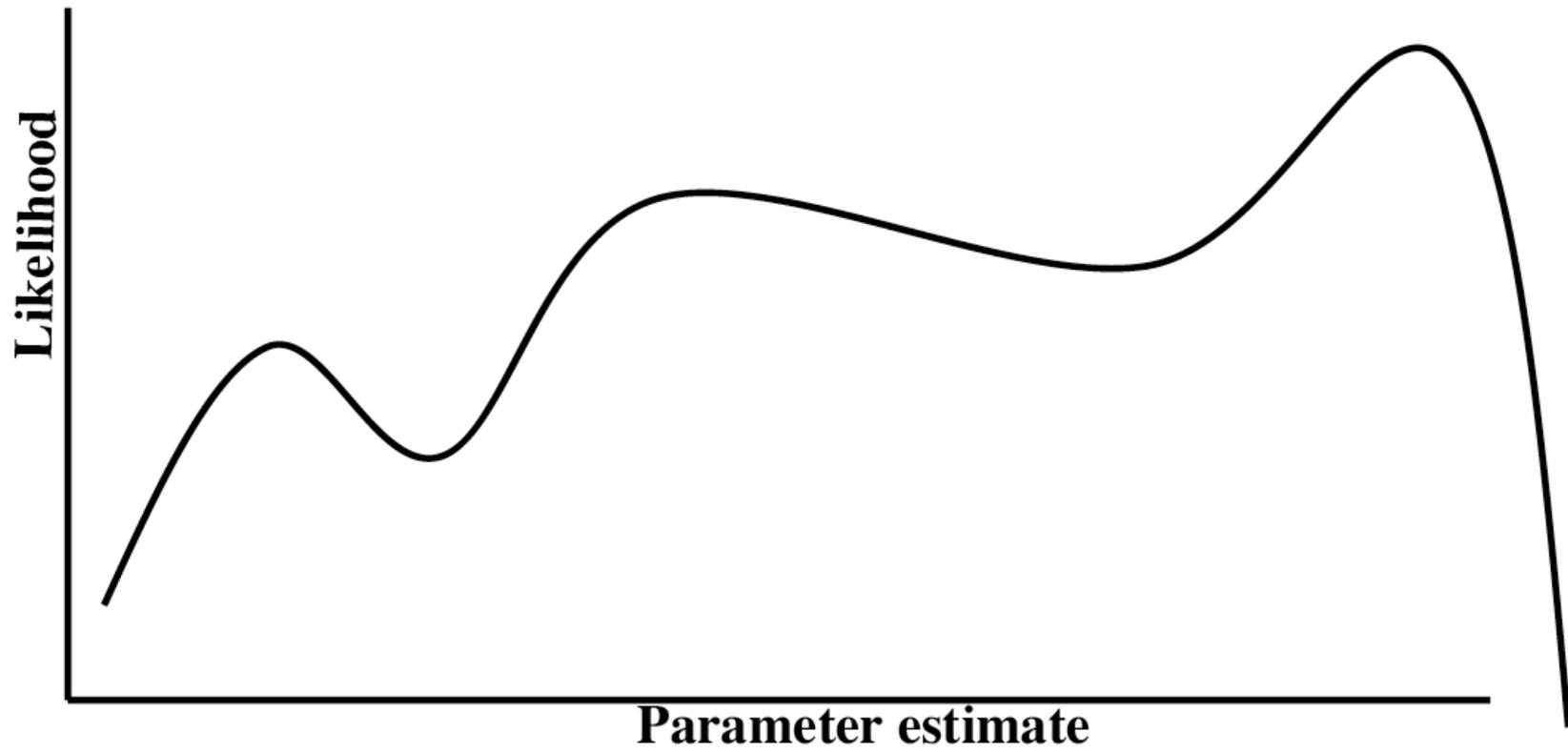
One way to get the likelihood is to estimate them using Markov Chain Monte Carlo (MCMC) methods.

-analogy to walking up hill.

Parameter estimation is made by changing values, estimating likelihood, and repeating until the function has been maximized.



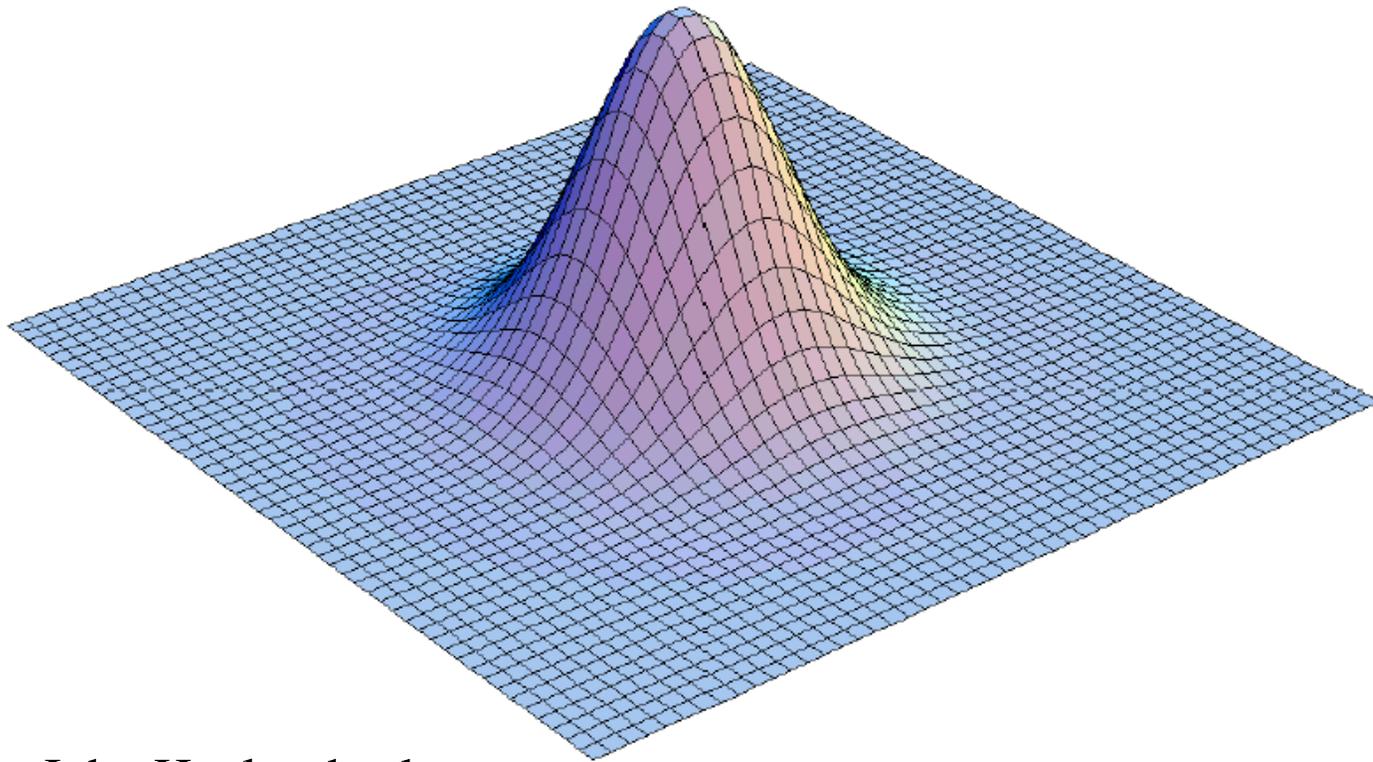
Problem of multiple peaks and valleys



Markov chain Monte Carlo

Metropolis, N., A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller. 1953. Equations of state calculations by fast computing Machines. *J. Chem. Phys.* 21:1087–1091.

Hastings, W. K. 1970. Monte Carlo sampling methods using Markov chains and their applications. *Biometrika* 57:97–109.



-Graphs from John Huelsenbeck

Markov Chain Monte Carlo

- Start with proposed state

Markov Chain Monte Carlo

- Start with proposed state
- Perturb old state and calculate probability of new state

Markov Chain Monte Carlo

- Start with proposed state
- Perturb old state and calculate probability of new state
- Test if new state is better than old state, accept if ratio of new to old is greater than a randomly drawn number between 0 and 1.

Markov Chain Monte Carlo

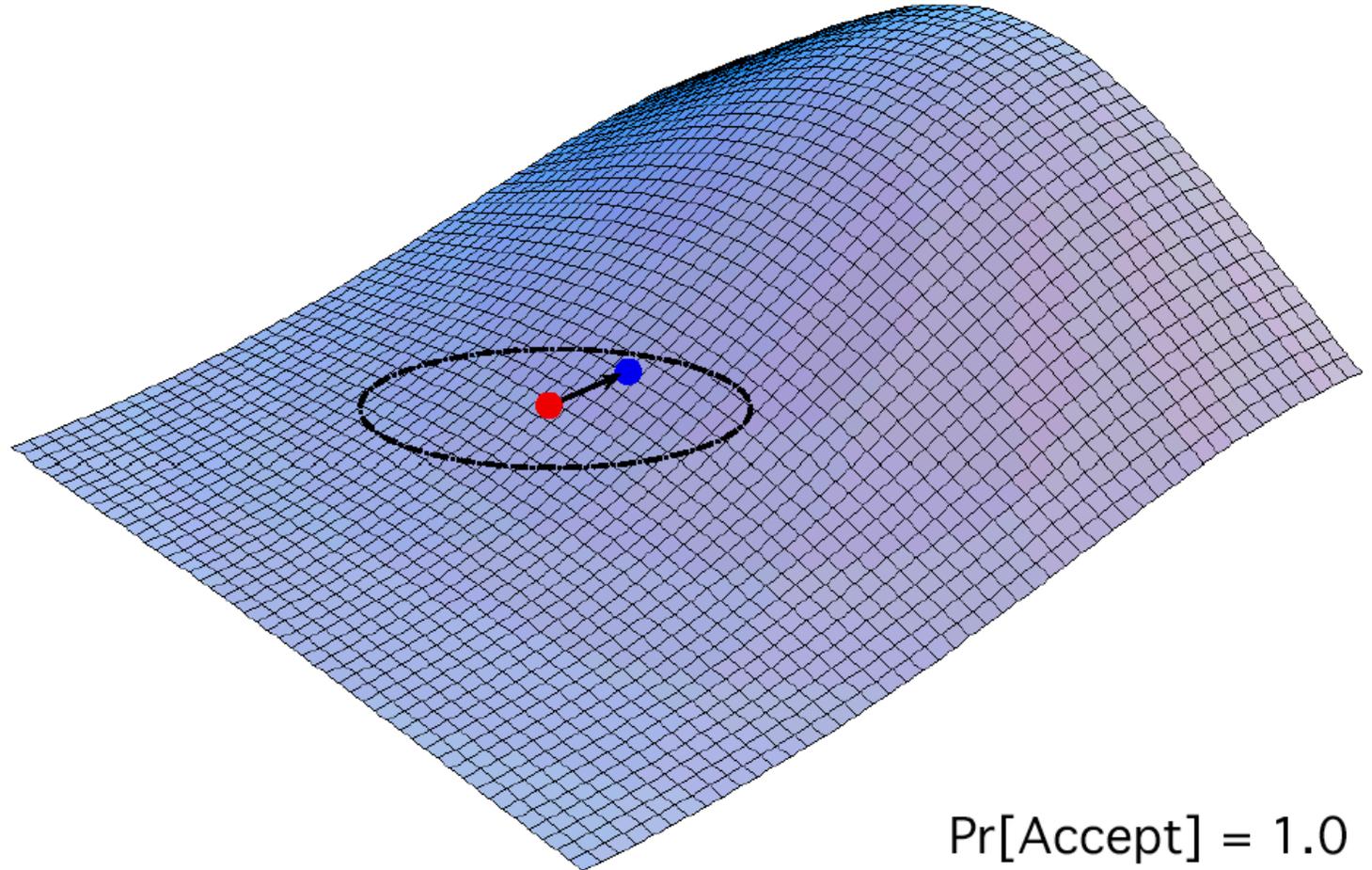
- Start with proposed state
- Perturb old state and calculate probability of new state
- Test if new state is better than old state, accept if ratio of new to old is greater than a randomly drawn number between 0 and 1.
- Move to new state if accepted, if not stay at old state

Markov Chain Monte Carlo

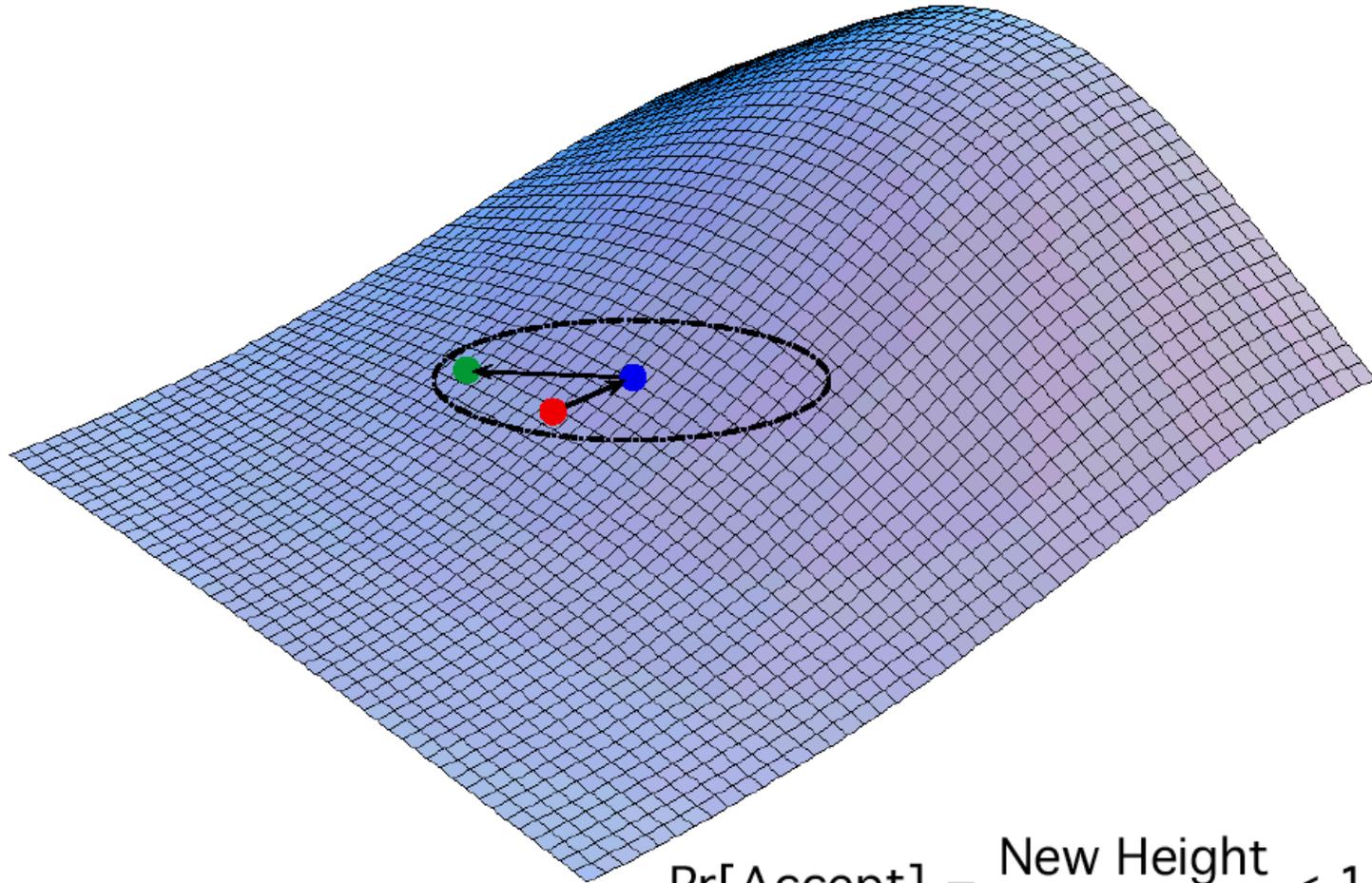
- Start with proposed state
- Perturb old state and calculate probability of new state
- Test if new state is better than old state, accept if ratio of new to old is greater than a randomly drawn number between 0 and 1.
- Move to new state if accepted, if not stay at old state
- Start over

Caveats: The proposal mechanism is at the discretion of the programmer, but must satisfy a few basic requirements: all states must be reachable, the chain must be aperiodic, and the mechanism must be stochastic.

Circle represents amount of potential proposed change.

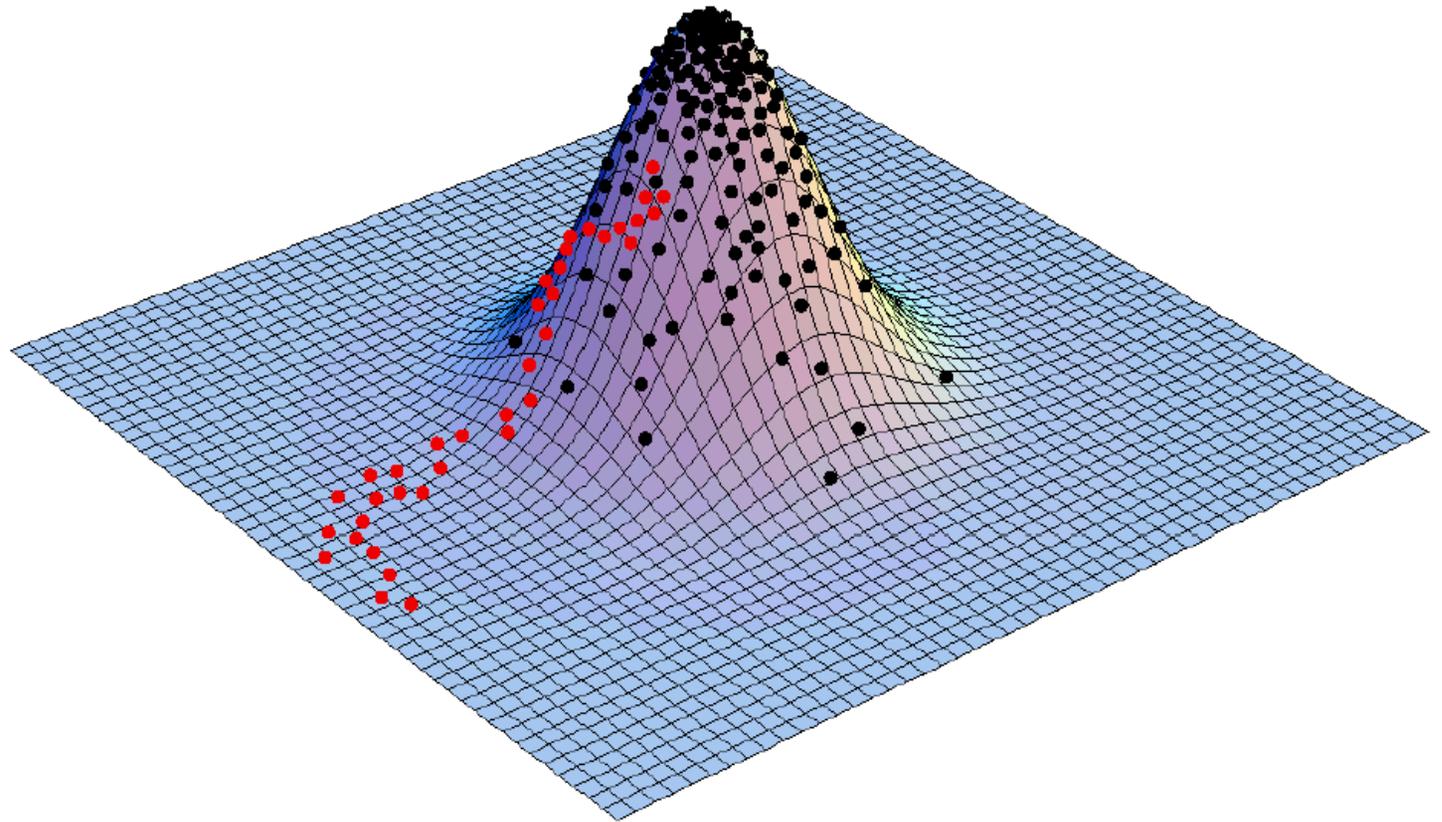


Pr[Accept] = 1.0



$$\Pr[\text{Accept}] = \frac{\text{New Height}}{\text{Old Height}} < 1.0$$

Repeat steps until you find the peak.



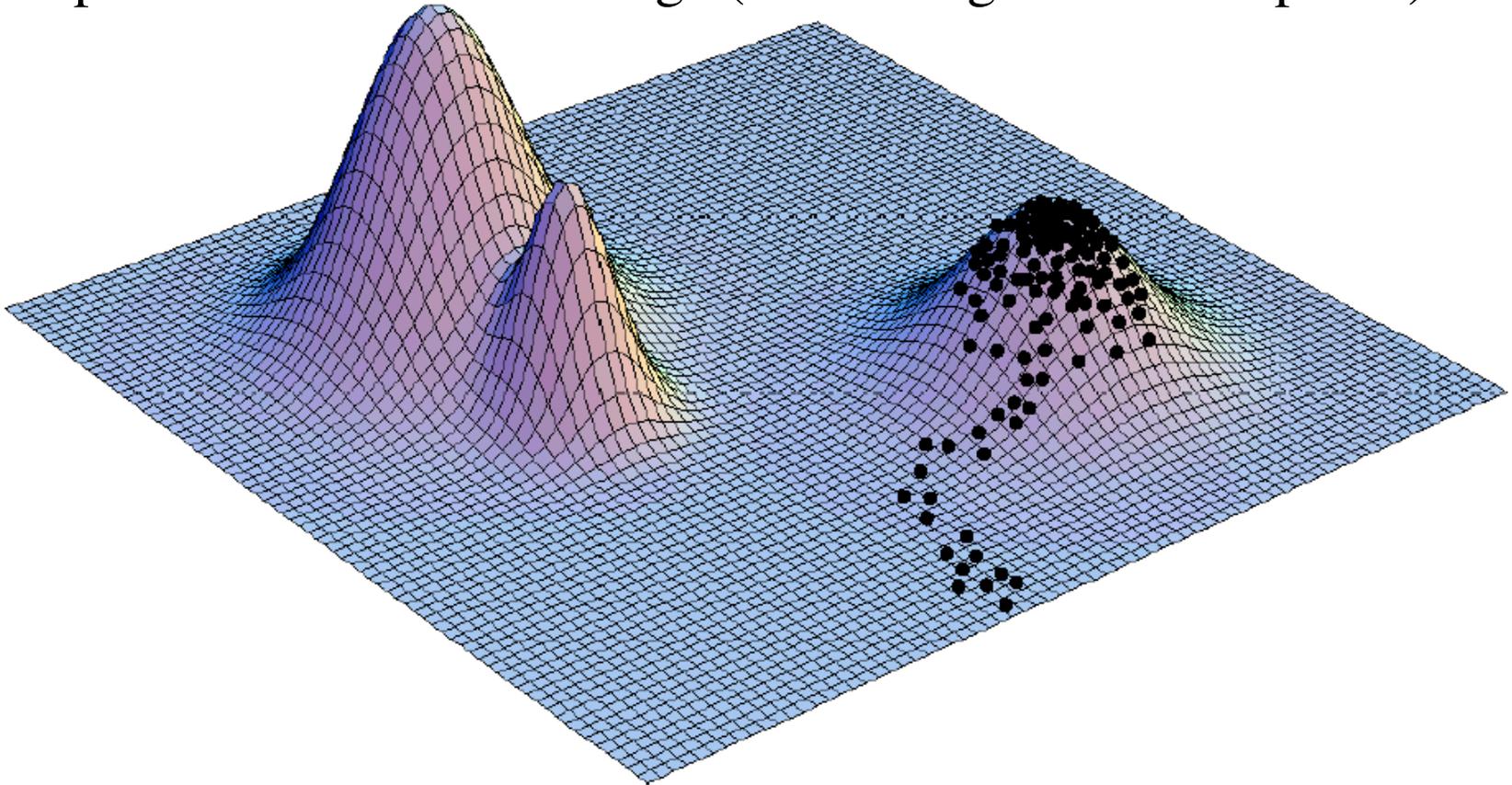
What is the “answer”

- Peak = maximum likelihood
- Mean
- Mode
- Median
- Credible set (ie with confidence interval)

How do you know if you reached the “peak” (maximum likelihood)?

Convergence = tested all of likelihood surface and found maximum

- example which did not converge (chain caught in local optima)



Convergence = tested all of likelihood surface and found maximum

Check convergence by

starting different initial estimates

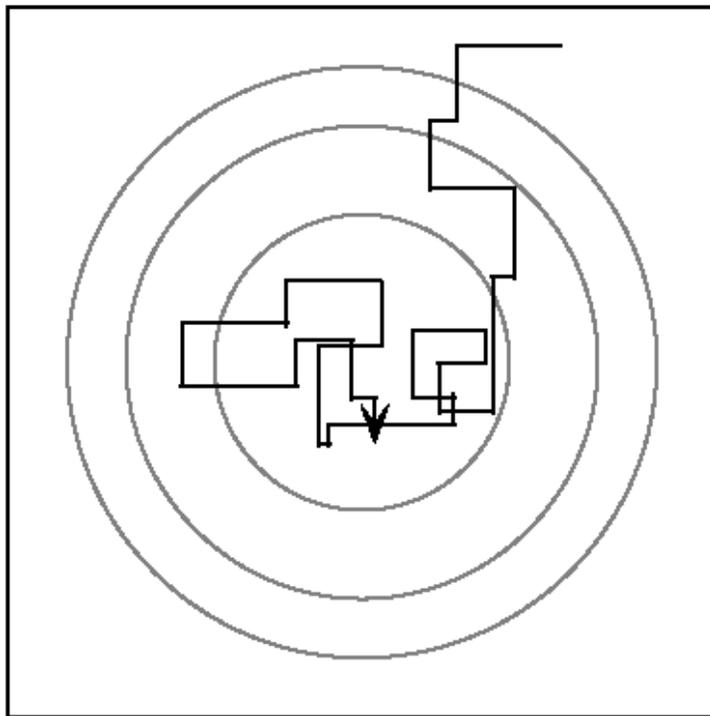
increase the amount values altered for parameter optimization

rerun the analysis several times

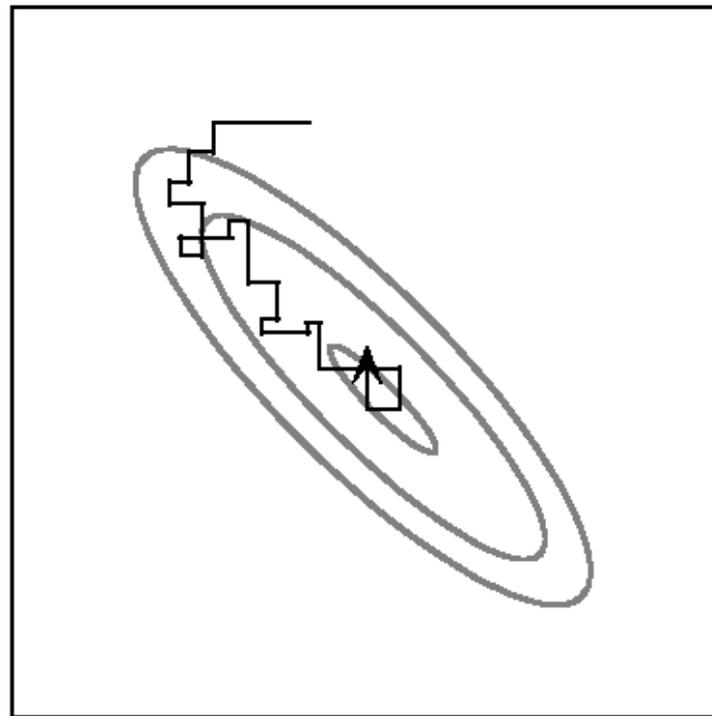
run for very long time

MC-Robot demo:

A

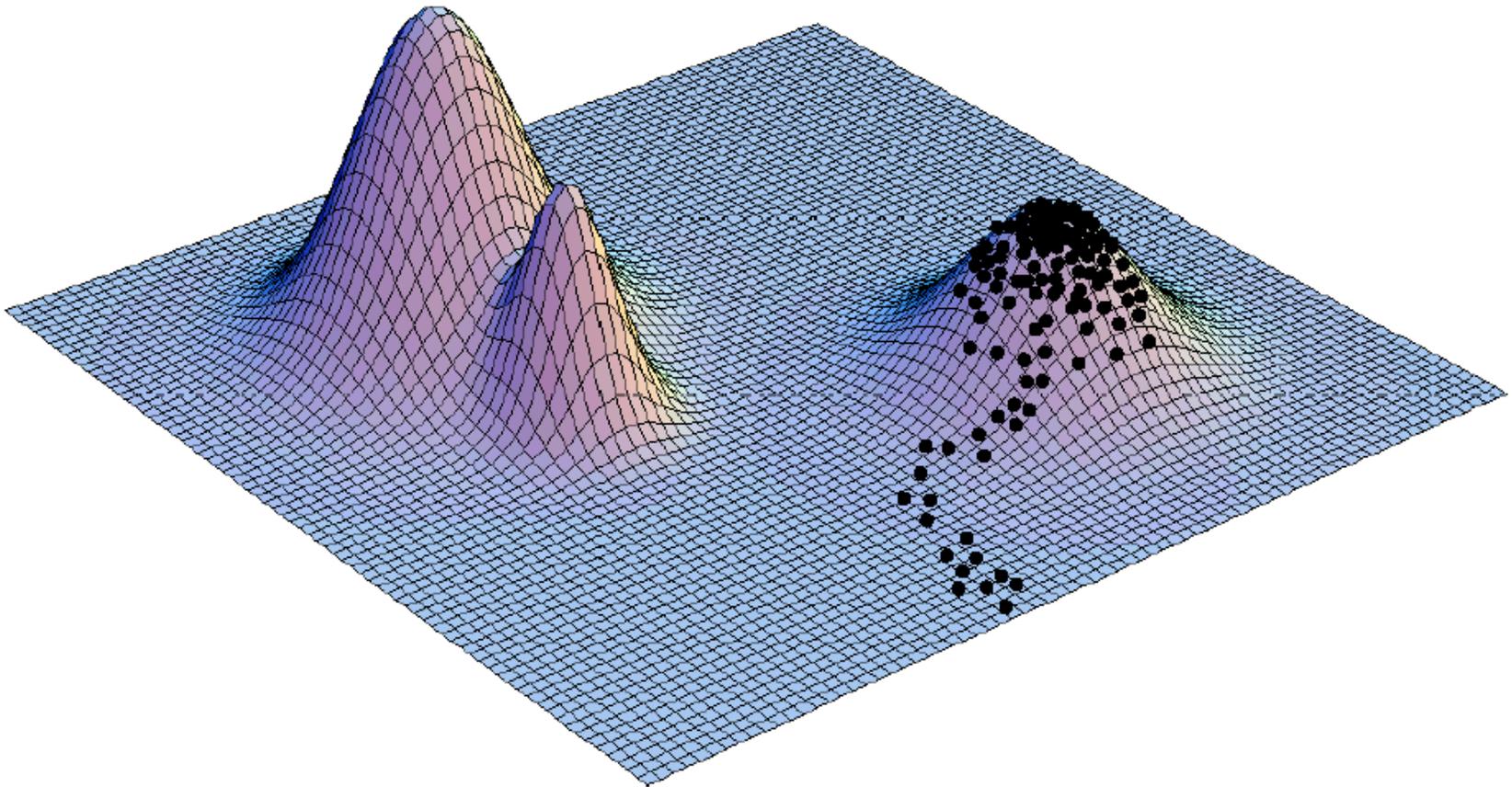


B



θ_1

Any other methods to explore parameter space?



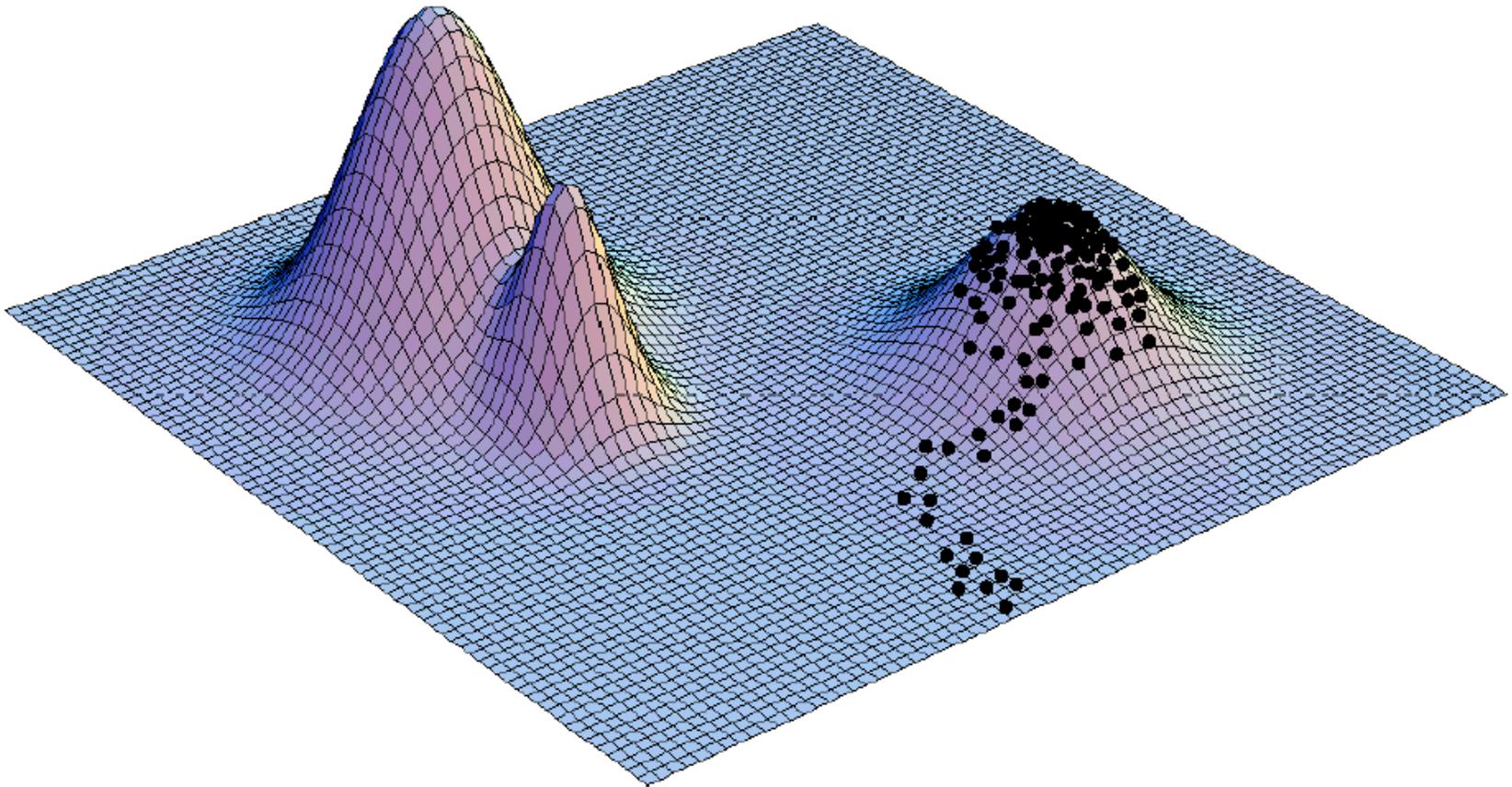
Metropolis-coupled Markov chain Monte Carlo (MCMCMC)

- Run n chains, $n-1$ of which are heated
- Every generation, each chain proposes and accepts/rejects moves independently
- A swap of the states of two independently chosen chains is attempted.
- States are sampled from the cold chain

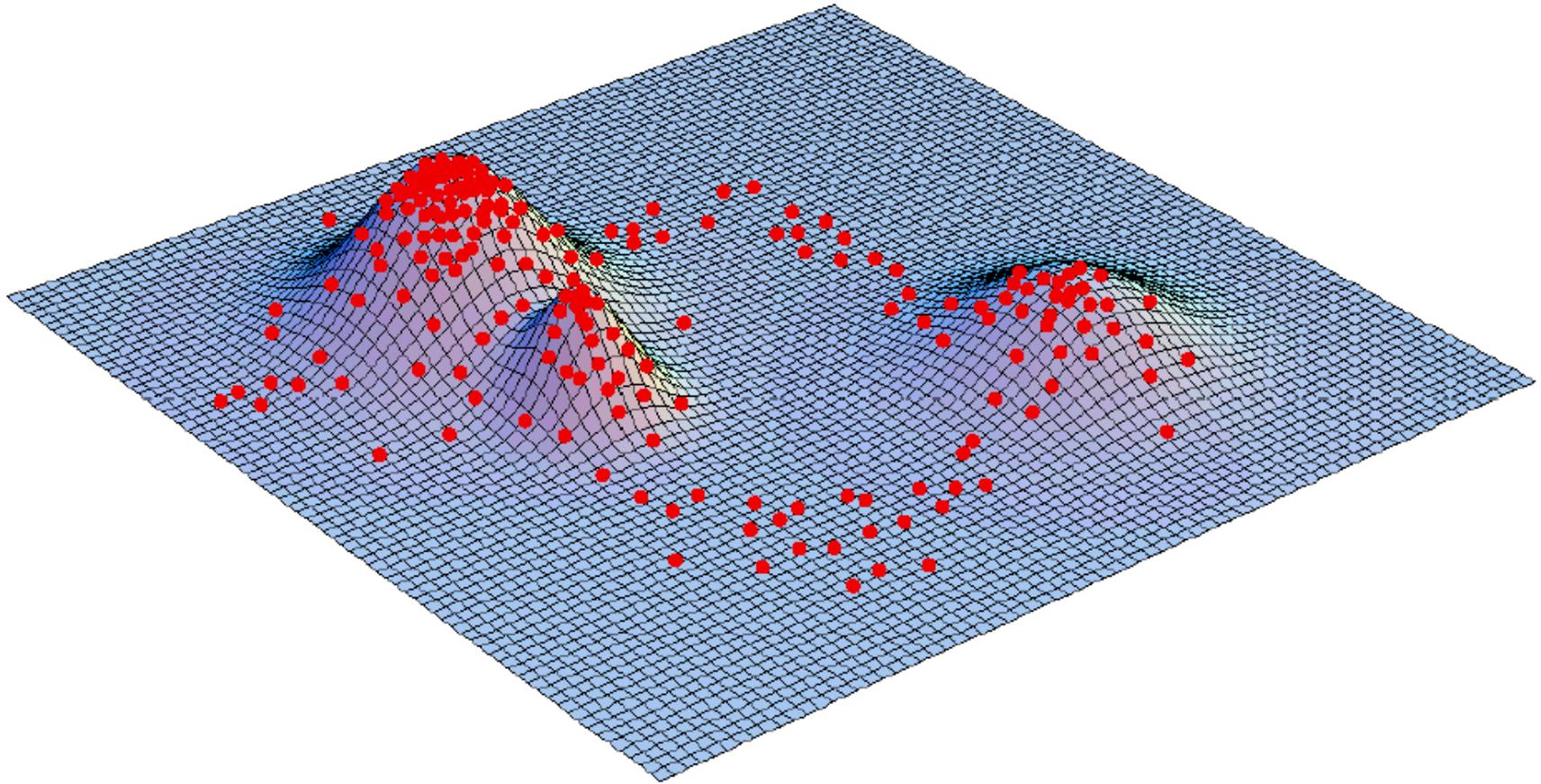
Heated chains

- Heating refers to powering the acceptance - rejection formula (new height/old height).
 - If greater than 1, always accept
 - If < 1 and cold chain accept with $p = 0.25$, twice as hot chain would accept with $p^{1/2} = 0.5$

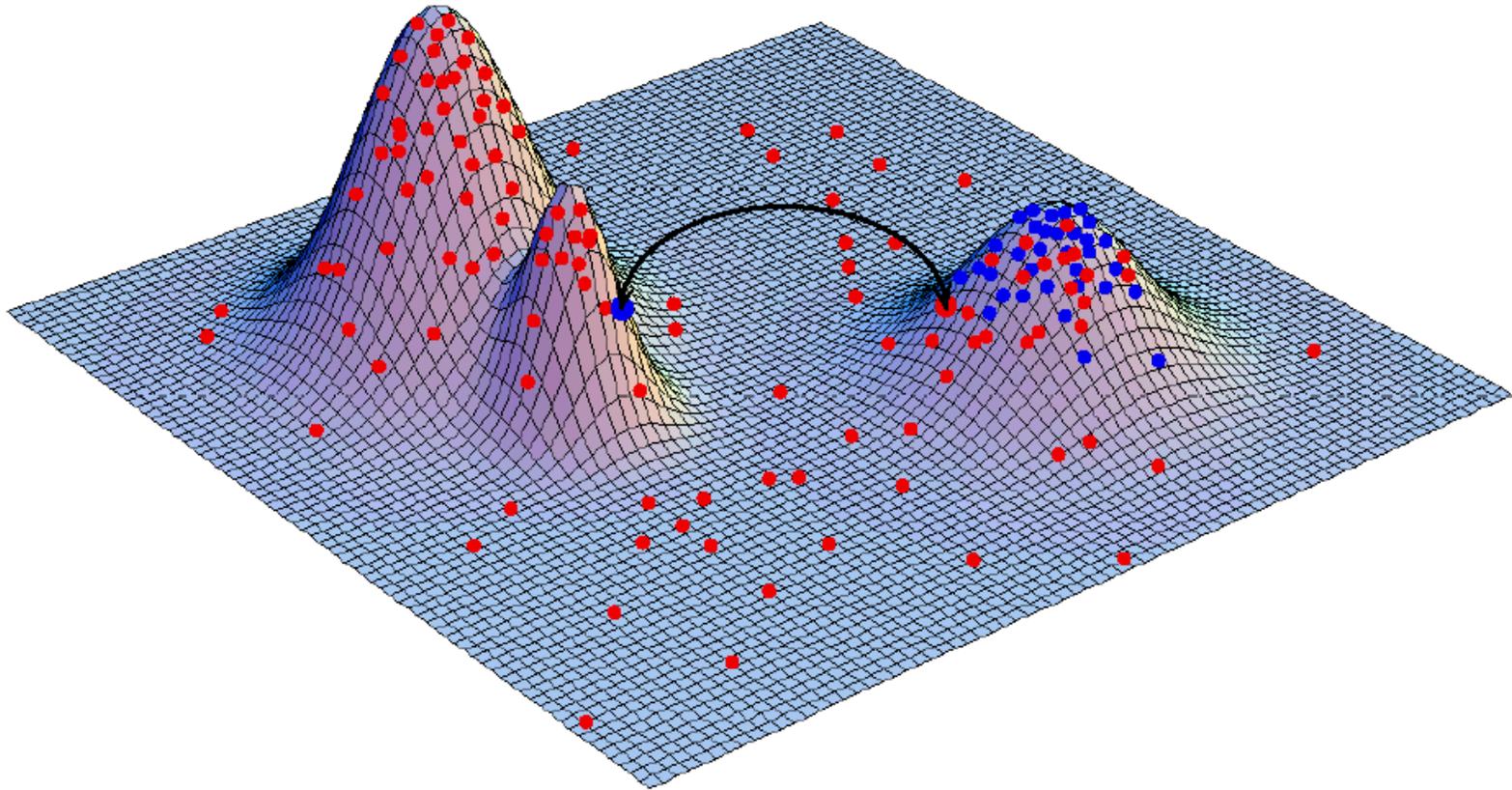
Cold chain

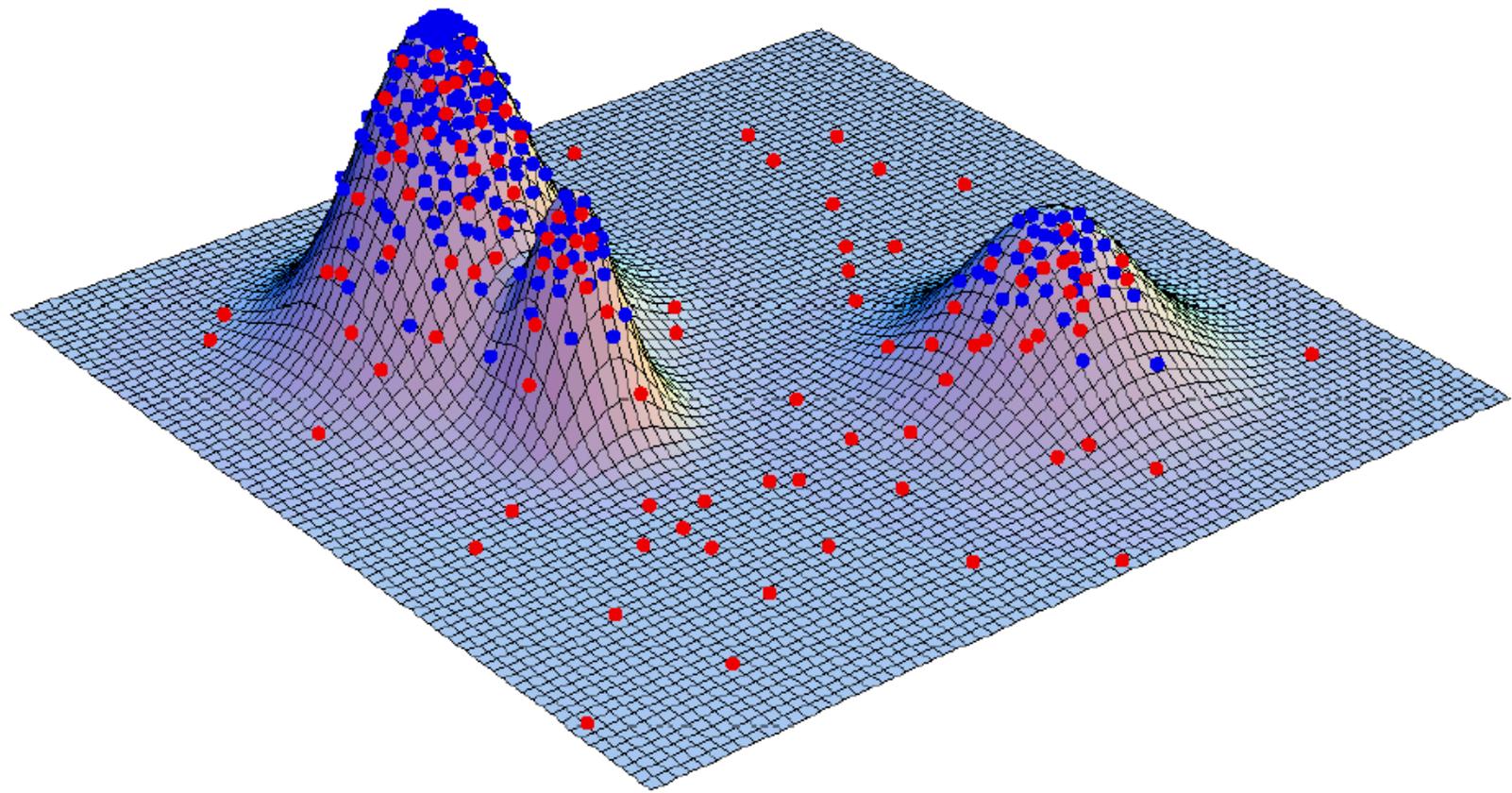


Hot Chain - notice peaks lower



Swap between chains





MC-Robot demo

Now what to do with likelihoods

- Compare models
 - Likelihood ratio test (LRT)
 - Is the d_N/d_S ratio for gene A significantly different from 1?
 - **Does a selection model fit better than a neutral model**
- Compute posterior probabilities
 - Bayes Theorem
 - Determine confidence

Compare two nested models (LRT)

- For each additional parameter added, the likelihood should always improve
 - How do you determine if improvement is significant?

Compare two nested models

- Free estimate (H_1). Estimate of d_N/d_S (ω) from alignment.
 - 1 parameter estimated
 - $\ln L = -100$
- Null model (H_0). Fix d_N/d_S (ω) to be 1.
 - 0 parameters estimated (all fixed)
 - $\ln L = -90$

Compare two nested models

- Free estimate (H_1). Estimate of d_N/d_S (ω) from alignment.
 - 1 parameter estimated
 - $\ln L = -100$
- Null model (H_0). Fix d_N/d_S (ω) to be 1.
 - 0 parameters estimated (all fixed)
 - $\ln L = -90$

LRT statistic is $-2[\ln L(H_1) - \ln L(H_0)]$

Degrees of freedom = difference # parameters

Compare to χ^2 distribution

Chi-square table:

Significance level

DF	0.9950	0.9750	0.9000	0.5000	0.1000	0.0500	0.0100	0.0010
1	0.0000	0.0010	0.0158	0.4549	2.7055	3.8415	6.6349	10.8276
2	0.0100	0.0506	0.2107	1.3863	4.6052	5.9915	9.2103	13.8155
3	0.0717	0.2158	0.5844	2.3660	6.2514	7.8147	11.3449	16.2662
4	0.2070	0.4844	1.0636	3.3567	7.7794	9.4877	13.2767	18.4668
5	0.4117	0.8312	1.6103	4.3515	9.2364	11.0705	15.0863	20.5150

Is it really so simple?

Reliabilities of Parsimony-based and Likelihood-based Methods for Detecting Positive Selection at Single Amino Acid Sites

Yoshiyuki Suzuki and Masatoshi Nei

Institute of Molecular Evolutionary Genetics, Department of Biology, The Pennsylvania State University

M7 is the Null model (H_0).

M8 is the alternate (H_1)

With 2 additional parameters

Table 1
Numbers of Positively Selected Amino Acid Sites in HLA Identified by the SG and the Yang Methods^a

METHOD	INITIAL ω VALUE	ln L	NUMBER OF POSITIVELY SELECTED AMINO ACID SITES		ARS INDEX ^d
			ARS (57) ^b	Non-ARS (216) ^c	
SG (0.5, 0.95) ^e	N.A. ^f	N.A.	17	3 [0] ^g	30.2
Yang (M0) ^h	0.2, 0.4, 0.6, 0.8, 1, 2, 3, 3.14, 4, 5	-9114.23	N.A.	N.A.	N.A.
Yang (M1) ⁱ	N.A.	-7759.34	N.A.	N.A.	N.A.
Yang (M2) ^j	0.2, 0.4, 0.6, 0.8, 1 2, 3, 3.14, 4, 5	-7637.75 -7593.90	0 14	0 4 [0]	N.A. 17.3
Yang (M3) ^k	0.2	-8332.38	36	42 [14]	7.1
	0.4	-8668.06	0	0	N.A.
	0.6	-8549.94	0	0	N.A.
	<u>0.8</u>	<u>-8180.63</u>	<u>27</u>	<u>19 [4]</u>	<u>9.3</u>
	1	-8327.32	29	19 [4]	10.7
	2	-8464.39	0	0	N.A.
	3	-8479.37	0	0	N.A.
	3.14	-8611.37	0	0	N.A.
	4	-8333.05	35	37 [9]	7.7
	5	-8378.32	31	26 [6]	8.7
Yang (M7) ^l	N.A.	-7803.40	N.A.	N.A.	N.A.
Yang (M8) ^m	0.2	-8260.79	0	0	N.A.
	0.4	-8114.22	0	0	N.A.
	0.6	-7694.07	19	6 [0]	17.5
	<u>0.8</u>	<u>-7732.55</u>	<u>17</u>	<u>6 [0]</u>	<u>14.9</u>
	1	-7759.36	0	0	N.A.
	2	-7831.09	25	6 [0]	27.3
	3	-7770.29	22	4 [0]	33.3
	3.14	-7857.20	25	19 [4]	8.1
	4	-7772.56	24	5 [0]	30.7
	5	-7775.08	22	5 [0]	26.5

What other information can you
get from MCMC methods?

Bayes's theorem and posterior probability:

Example from Box 3, Lewis 2001

Urn A contains 40% black marbles

Urn B contains 80% black marbles

What is the likelihood that a black marble came from Urn A? Urn B?

What is the posterior probability that a black marble came from Urn A? Urn B?

$$\Pr[B|A] = \frac{\Pr[B] \times \Pr[A|B]}{\Pr[A]}$$

$\Pr[B|A]$ = Posterior probability of data
probability of hypothesis given data

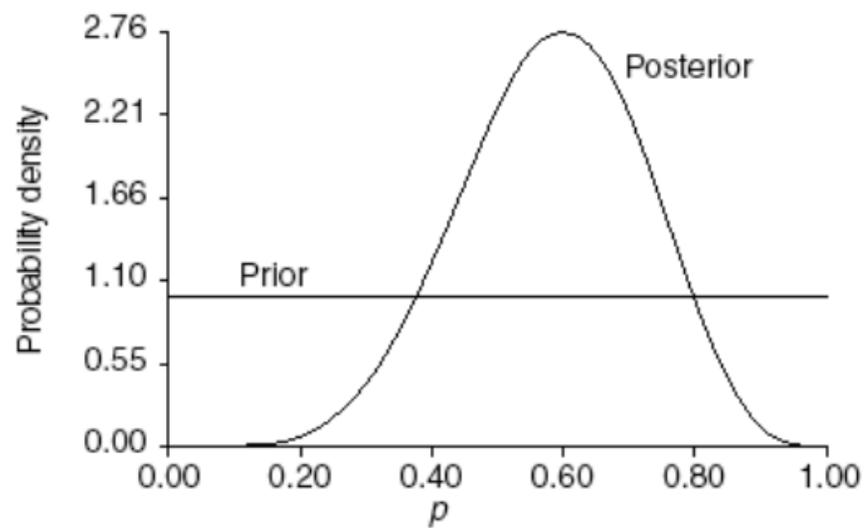
$\Pr[B]$ = Prior probability

$\Pr[A]$ = Unconditional probability of data
sum of $\Pr(B)\Pr(A|B)$ for all values B

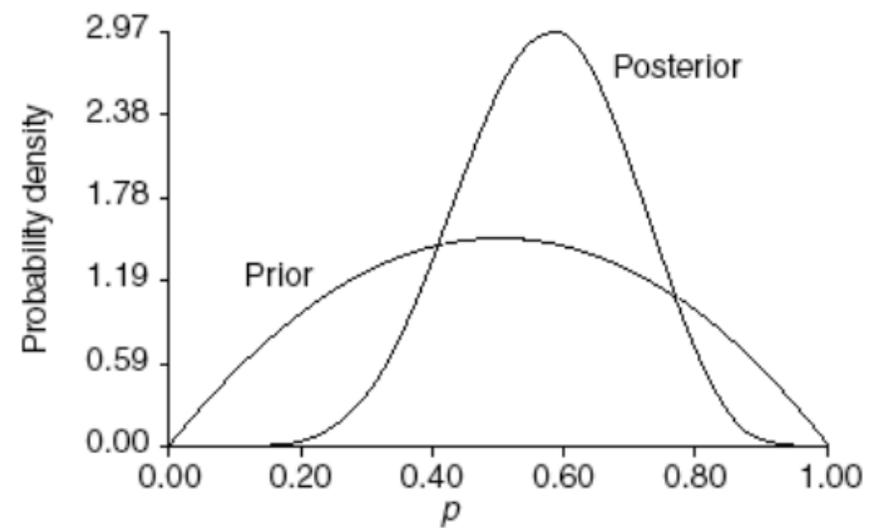
$\Pr[A|B]$ = Likelihood
probability of data given hypothesis

Prior and posterior probabilities

(a)



(b)

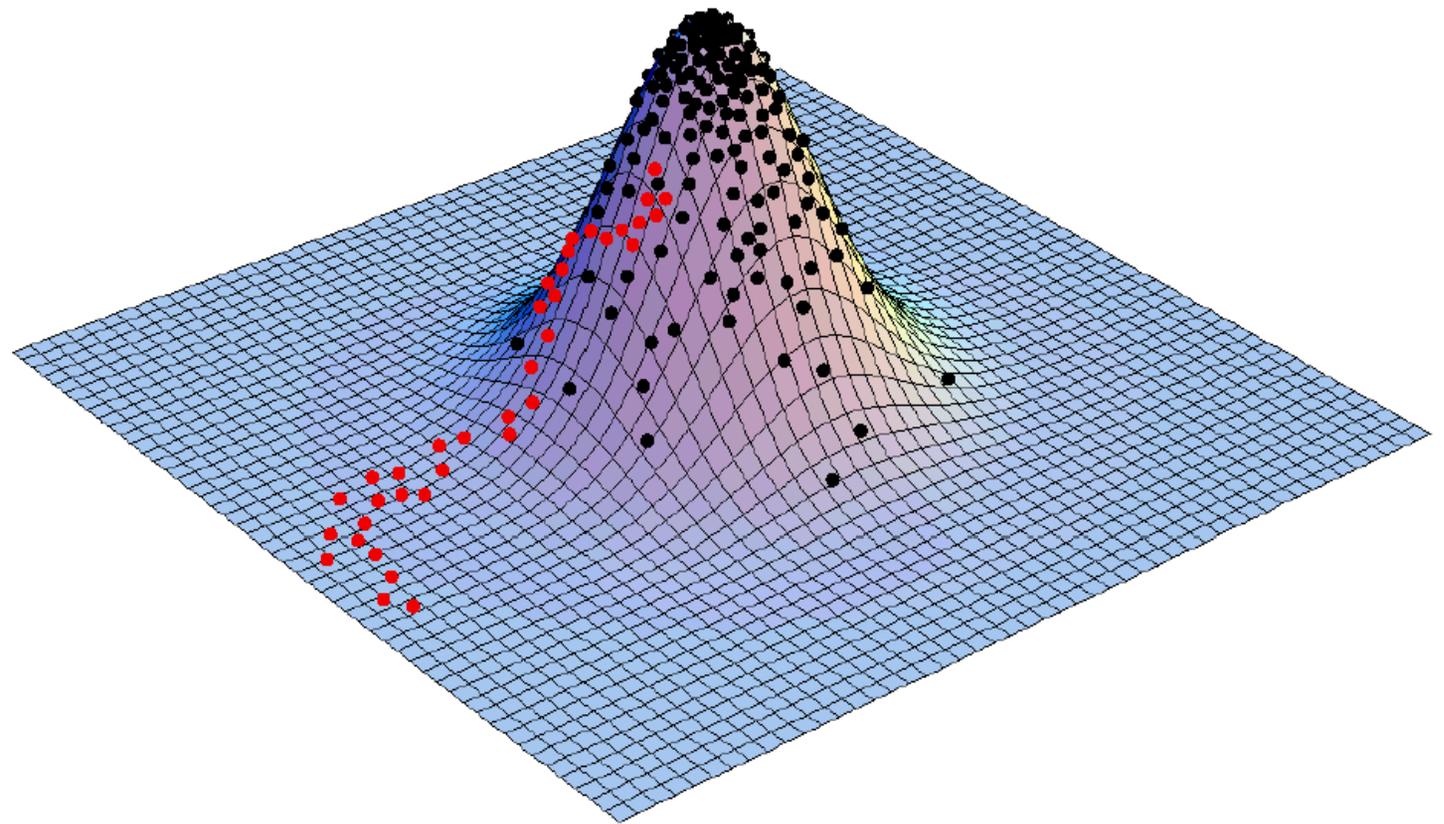
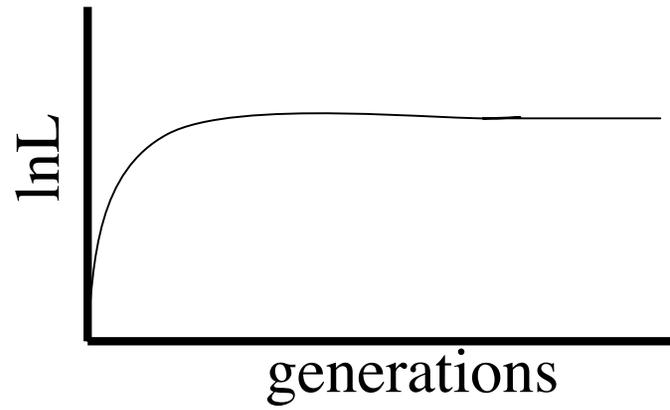


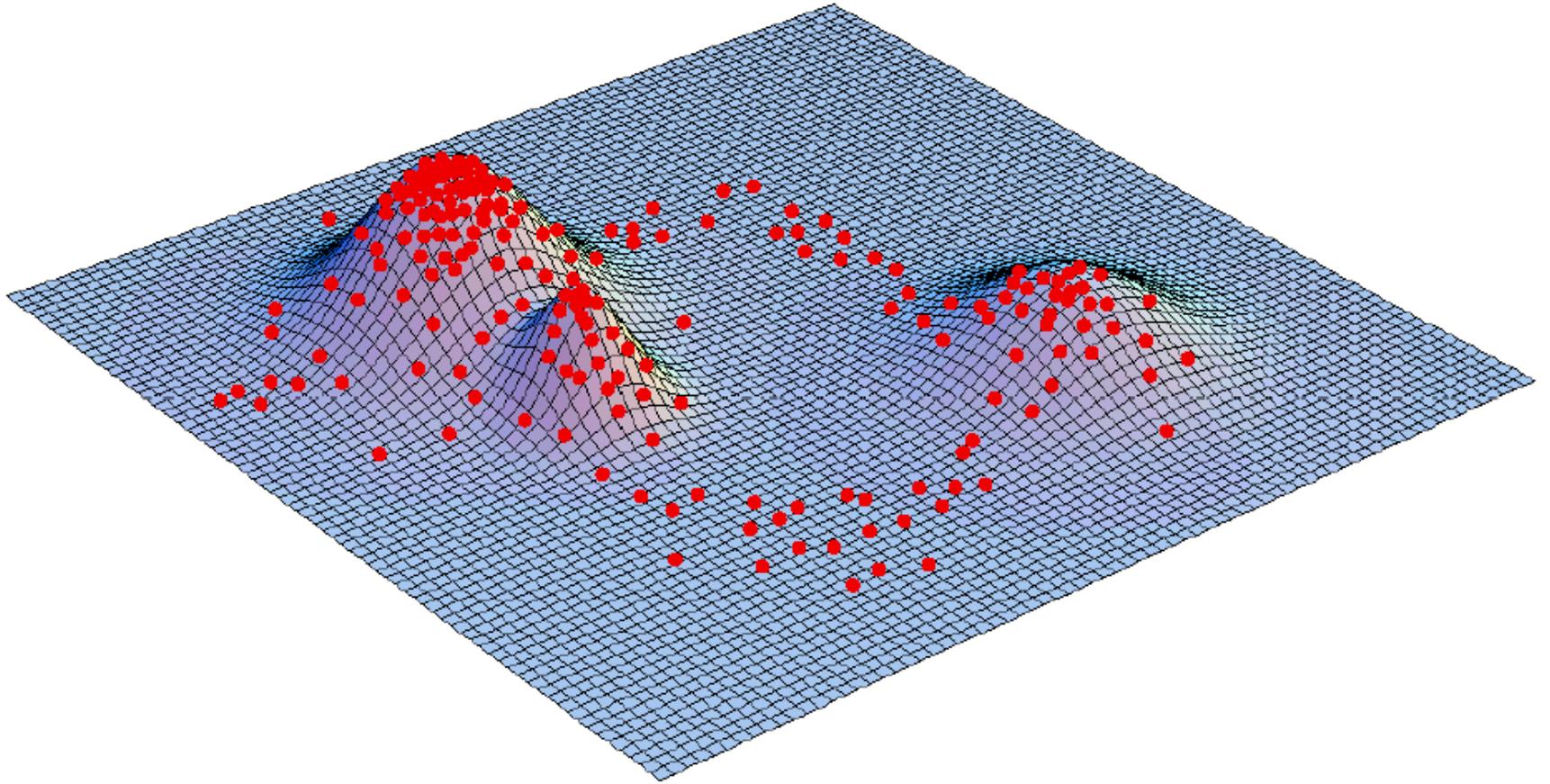
**Ass with likelihoods, this is difficult to calculate.
Can we use MCMC?**

$$\Pr[B|A] = \frac{\Pr[B] \times \Pr[A|B]}{\Pr[A]}$$

For an appropriately constructed and adequately run Markov chain, the proportion of the time any parameter value is visited is a valid approximation of the posterior probability of that parameter (= Bayes)

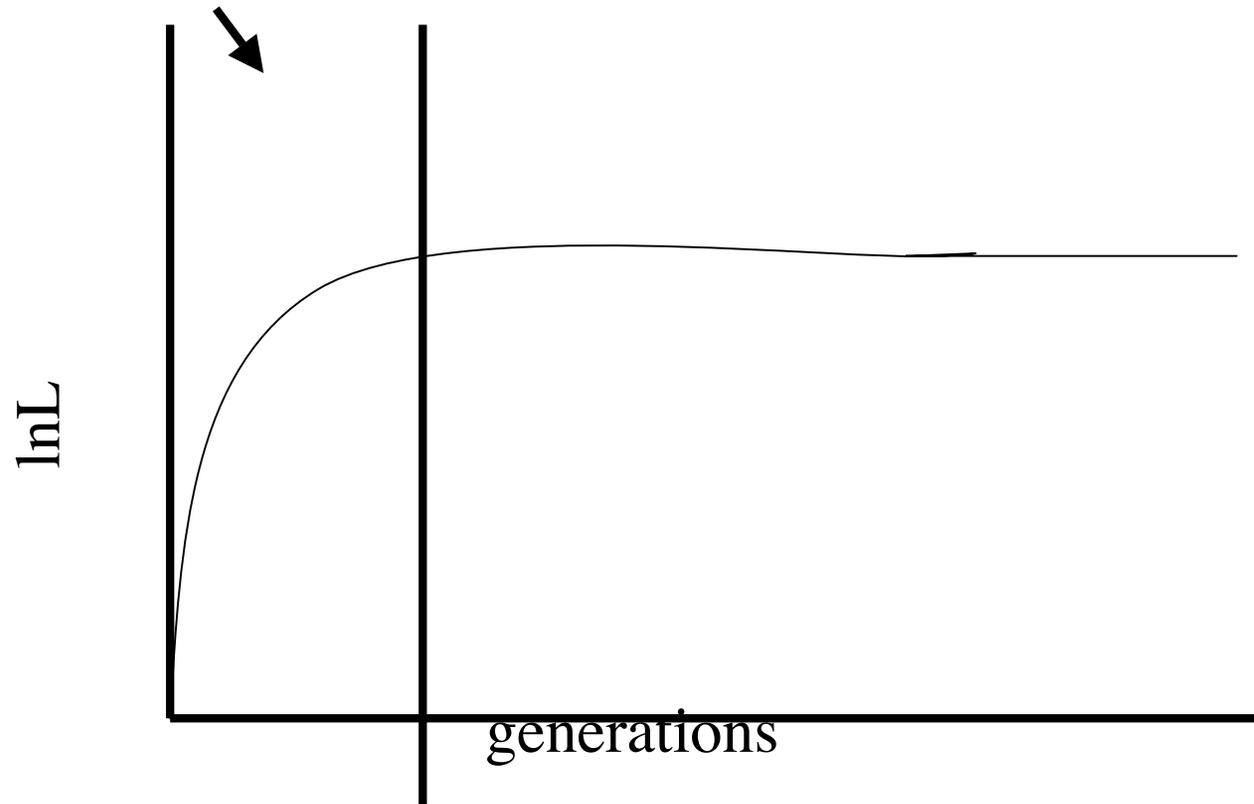
Red dots = “burn in” period.





Red dots = “burn in” period.

Burn in period



Advantages Bayesian perspective

- Gives probability of hypothesis of interest
 - Likelihood probability data given hypothesis
- Compare to bootstrap
- Disadvantage is subjectivity of prior

Two type of Bayes inference we will see

- Naïve Empirical Bayes (NEB)
 - Assumes parameter estimates from likelihood are exact
- Bayes Empirical Bayes (BEB)
 - Takes into account error in likelihood estimates of parameters

Some things to consider when running MCMC analyses

- Number of generations
- Number of chains
- Burn-in period
- Convergence

Some uses of MCMC

- Phylogeny
- Models of codon evolution (PAML-April 6)
- Conserved regions (Shadowing-April 13)
- Mutation rates
- Migration rates
- Population structure (Structure-April 20)