Markov Chain Monte Carlo

Quan Zhang, Ph.D.

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

Maximum likelihood estimation (MLE) finds 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).

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

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

$$P(x) = (n!/(n-x)!)p^x(1-p)^{n-x}$$

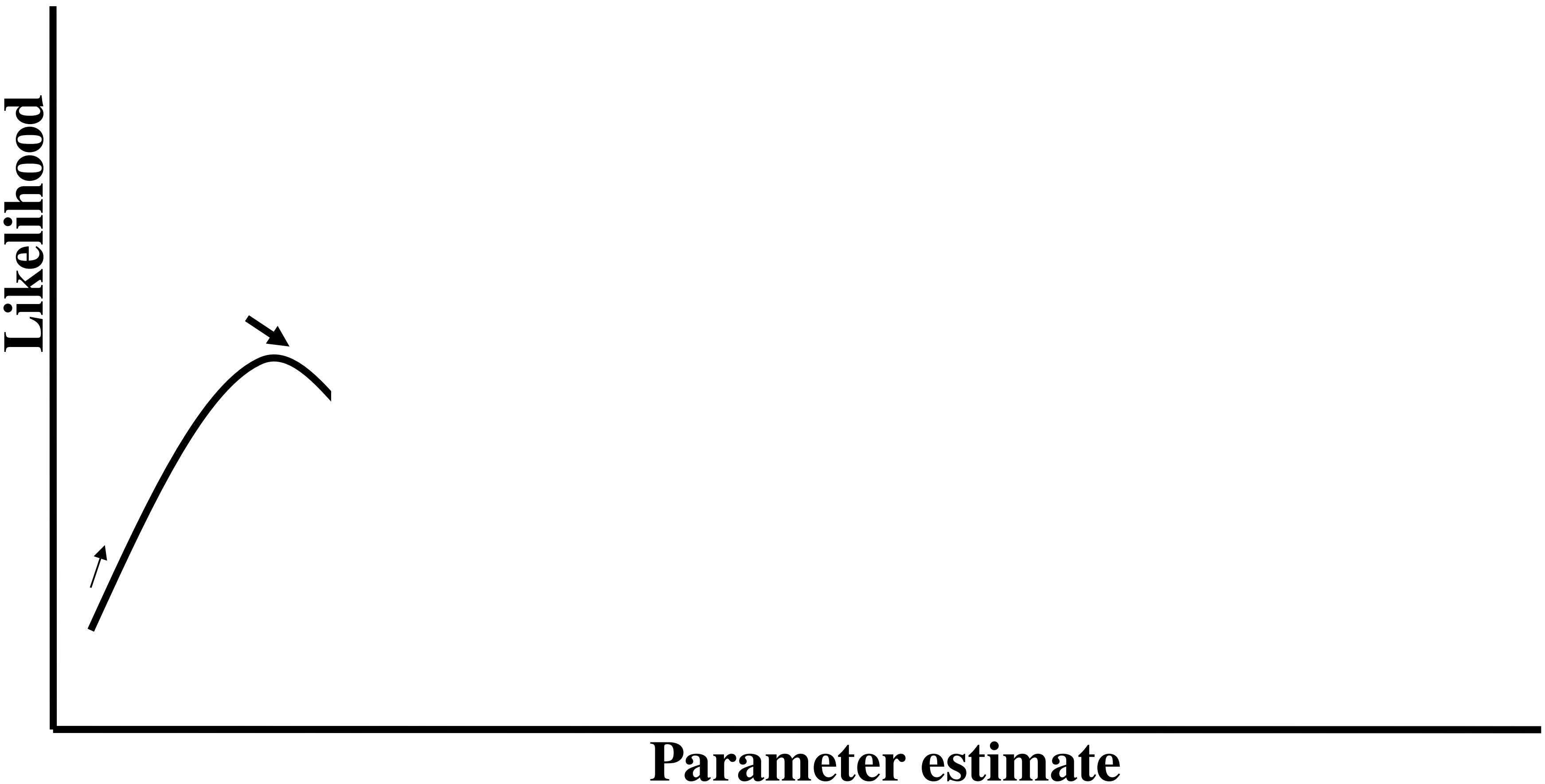
How do we calculate likelihoods and estimate parameters from a Bayesian perspective?

From a Bayesian perspective, we want to find the posterior distribution of the parameters whose samples are around the “peak” of the likelihood.

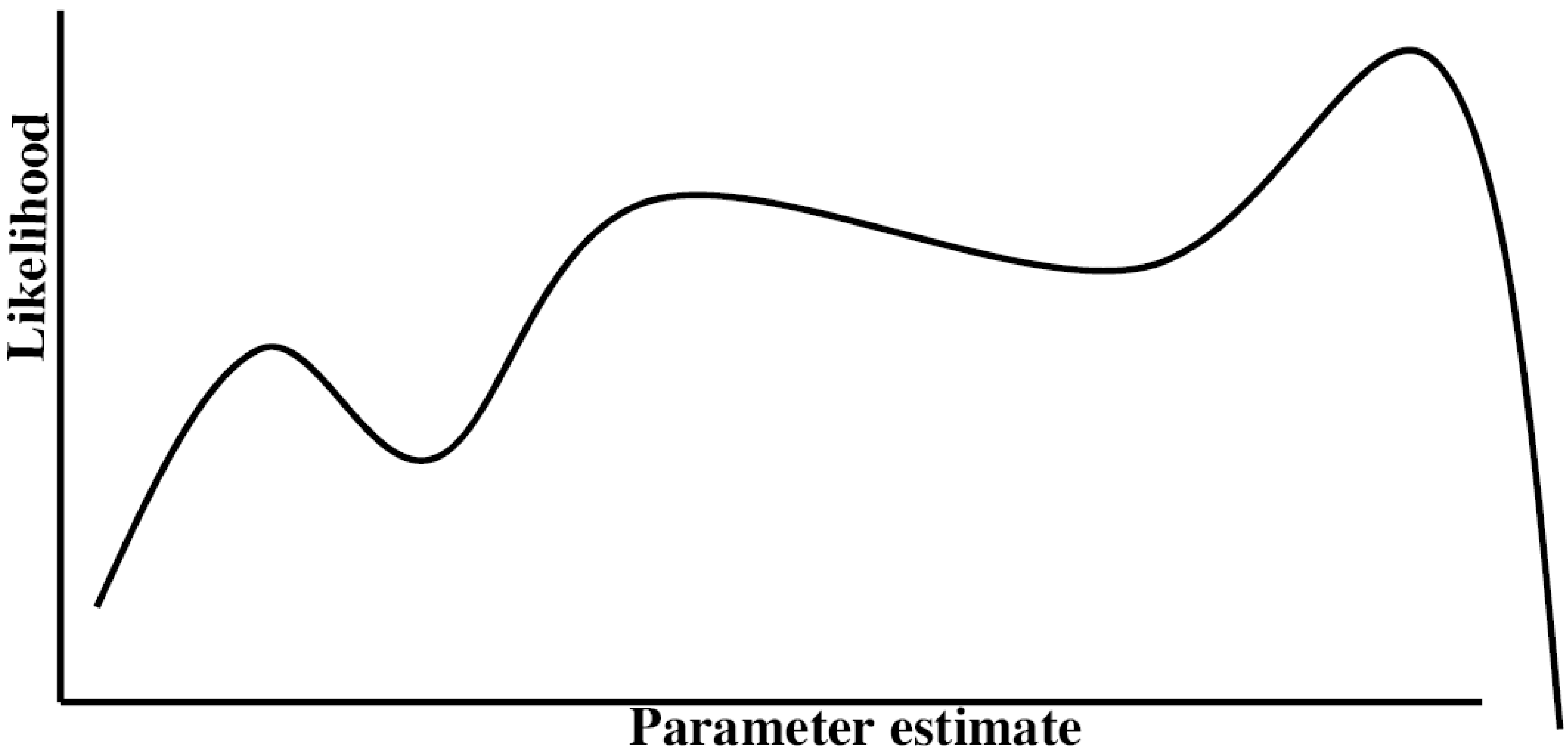
posterior $\propto$ likelihood * prior

-analogy to walking up hill.

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



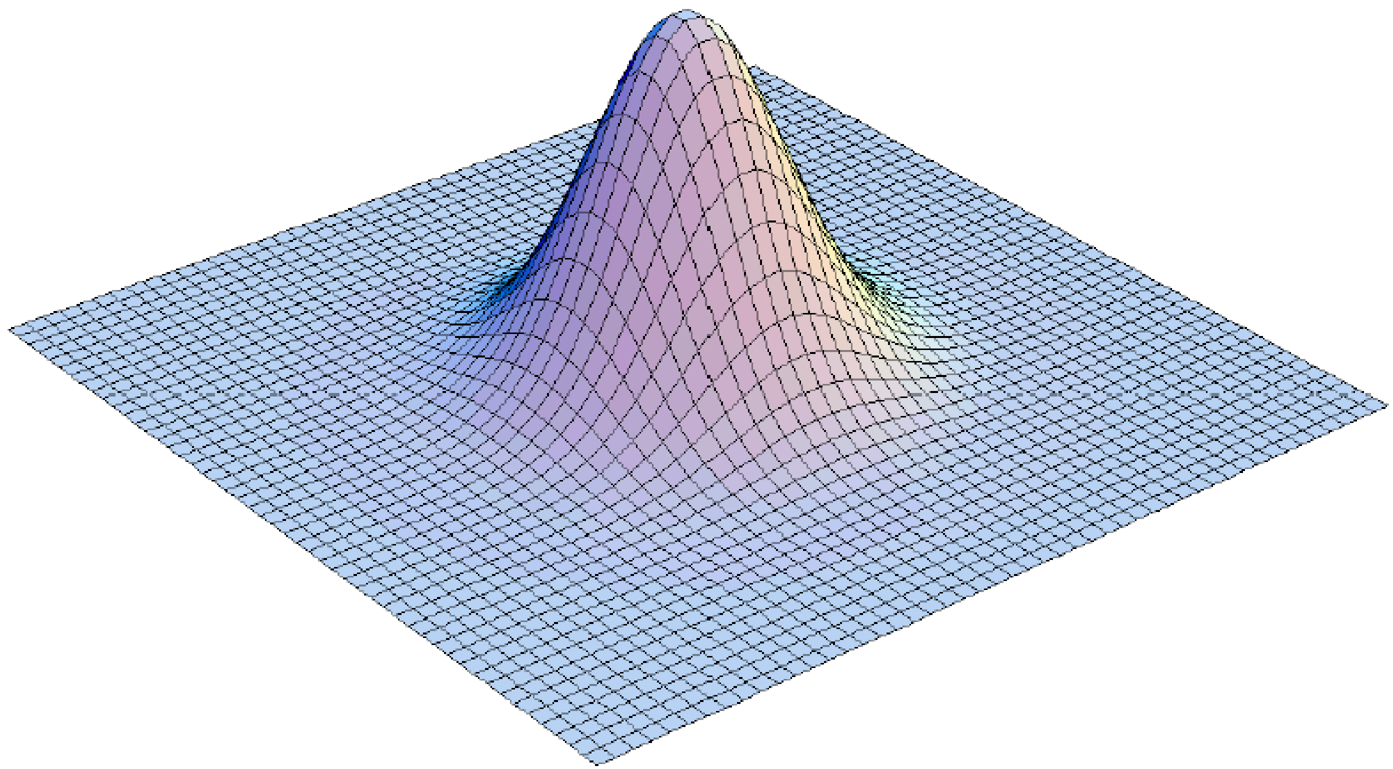
Problem of multiple peaks and valleys



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

- A Markov chain is a stochastic model describing a sequence of possible events in which the probability of each event depends only on the state attained in the previous event.
- We are focused on discrete-time Markov chains
- Example

  $b_t = 0$ means a plate is not broken on the tth day

  $b_t = 1$ means the plate is broken on the tth day

  $p(b_{t+1} \mid b_1, b_2, \ldots, b_t) = p(b_{t+1} \mid b_t)$

Markov Chain Monte Carlo (MCMC)

- MCMC constructs a Markov chain of parameters of interest, and obtains a sample from their posterior distribution by recording the states of the chain
- Example

  $y_i = \boldsymbol{x}_i' \boldsymbol{\beta} + \epsilon_i$

  MCMC:

  - Set initial value $\boldsymbol{\beta}_0$
  - Sample $\boldsymbol{\beta}_1$ from $p(\boldsymbol{\beta}_1 \mid \boldsymbol{\beta}_0)$ that is defined by the MCMC
  - ...
  - Sample $\boldsymbol{\beta}_t$ from $p(\boldsymbol{\beta}_t \mid \boldsymbol{\beta}_{t-1})$ that is defined by the MCMC
  - ...
  - Sample $\boldsymbol{\beta}_T$ from $p(\boldsymbol{\beta}_T \mid \boldsymbol{\beta}_{T-1})$ that is defined by the MCMC

  Use $\{\boldsymbol{\beta}_1, \boldsymbol{\beta}_2, \ldots, \boldsymbol{\beta}_T\}$ as samples from the posterior of β.

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

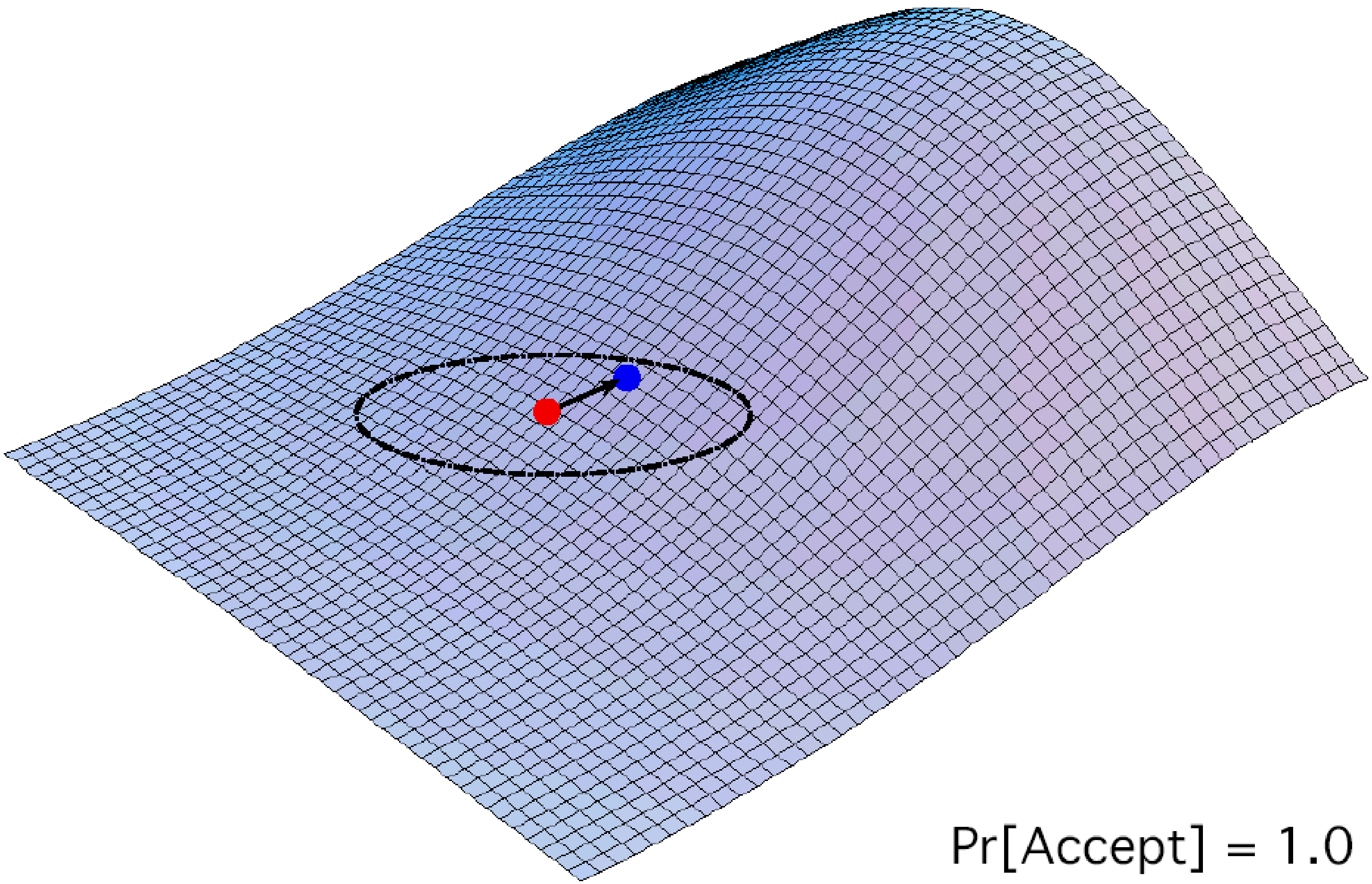
- Start with proposed state
- Perturb old state and calculate probability of new state
- Test if new state is better than old state, accept if ratio of new to old is greater than a randomly drawn number between 0 and 1.
- Move to new state if accepted, if not stay at old state

Markov Chain Monte Carlo

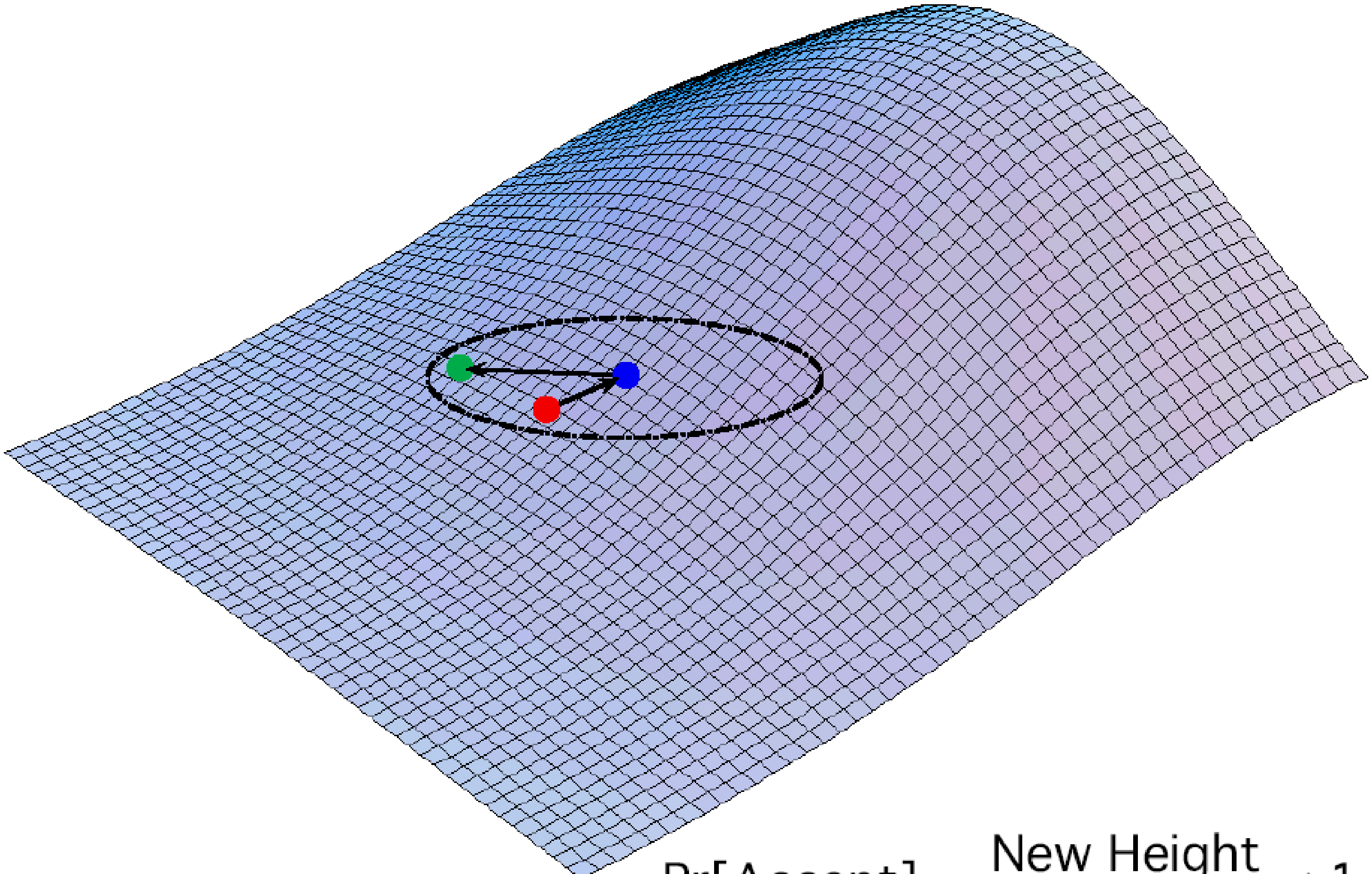
- Start with proposed state
- Perturb old state and calculate probability of new state
- Test if new state is better than old state, accept if ratio of new to old is greater than a randomly drawn number between 0 and 1.
- Move to new state if accepted, if not stay at old state
- Start over

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.

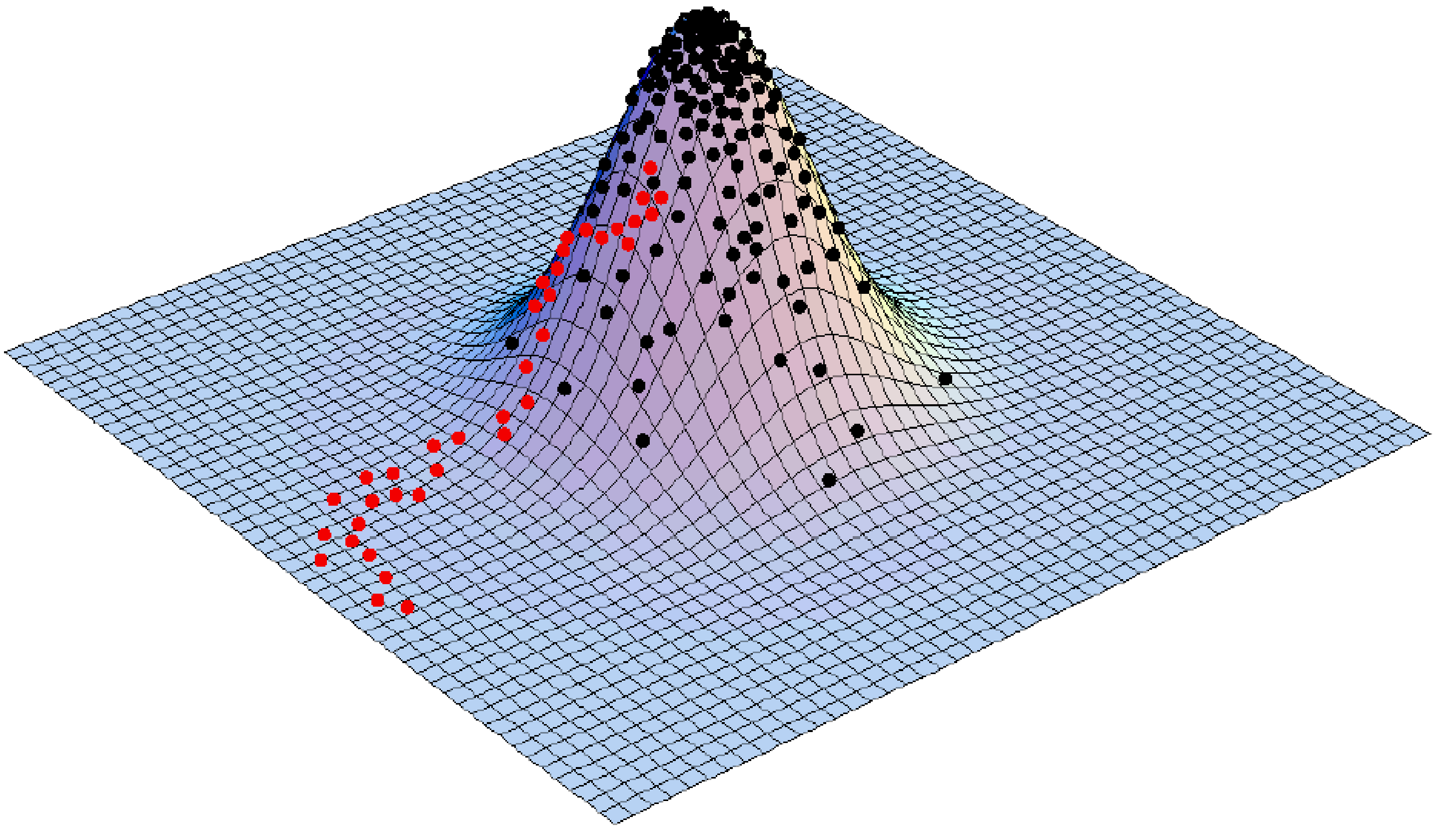


Pr[Accept] = 1.0



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

Repeat steps until you find the peak.



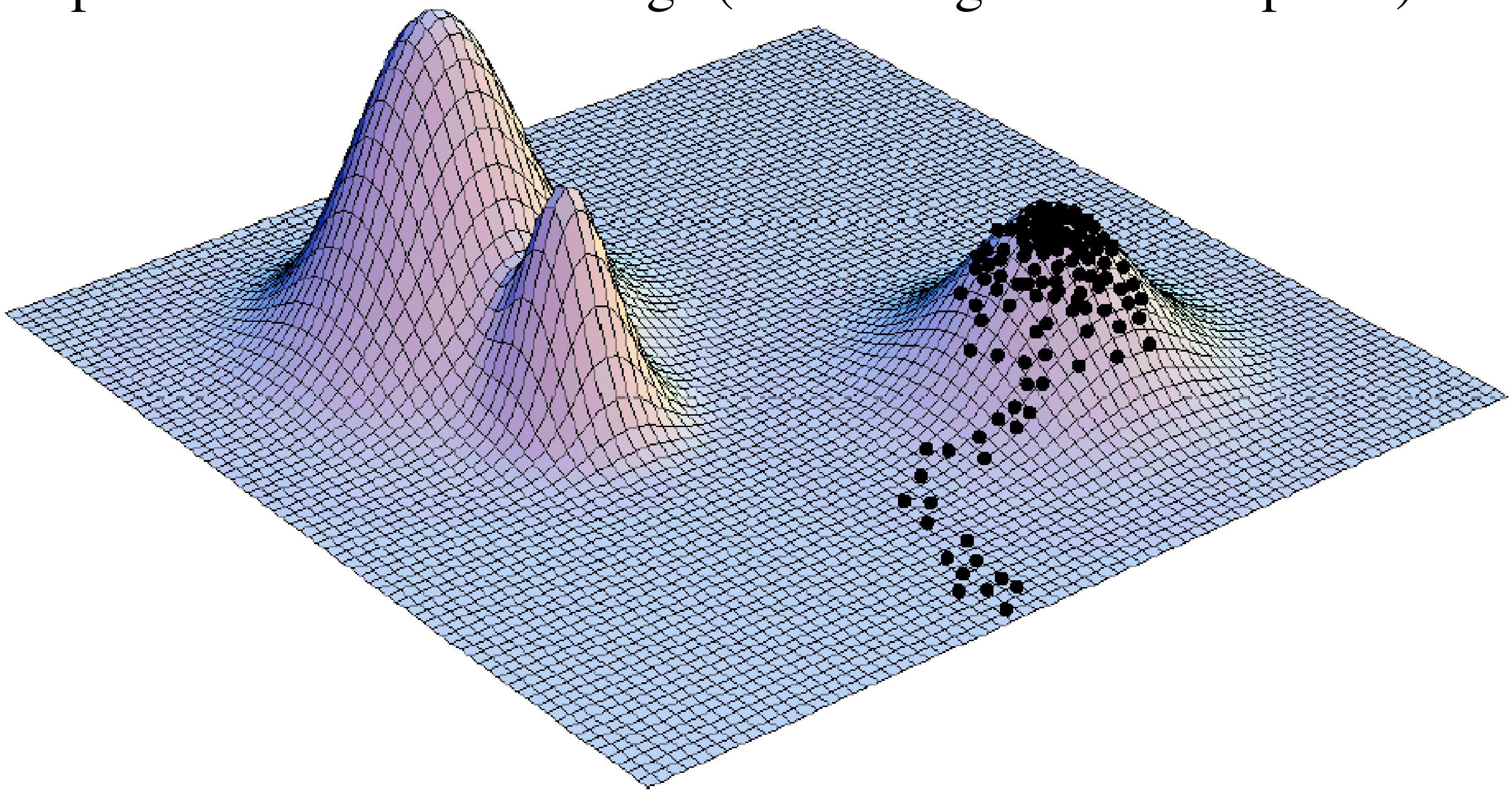
What is the “answer”

- 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)



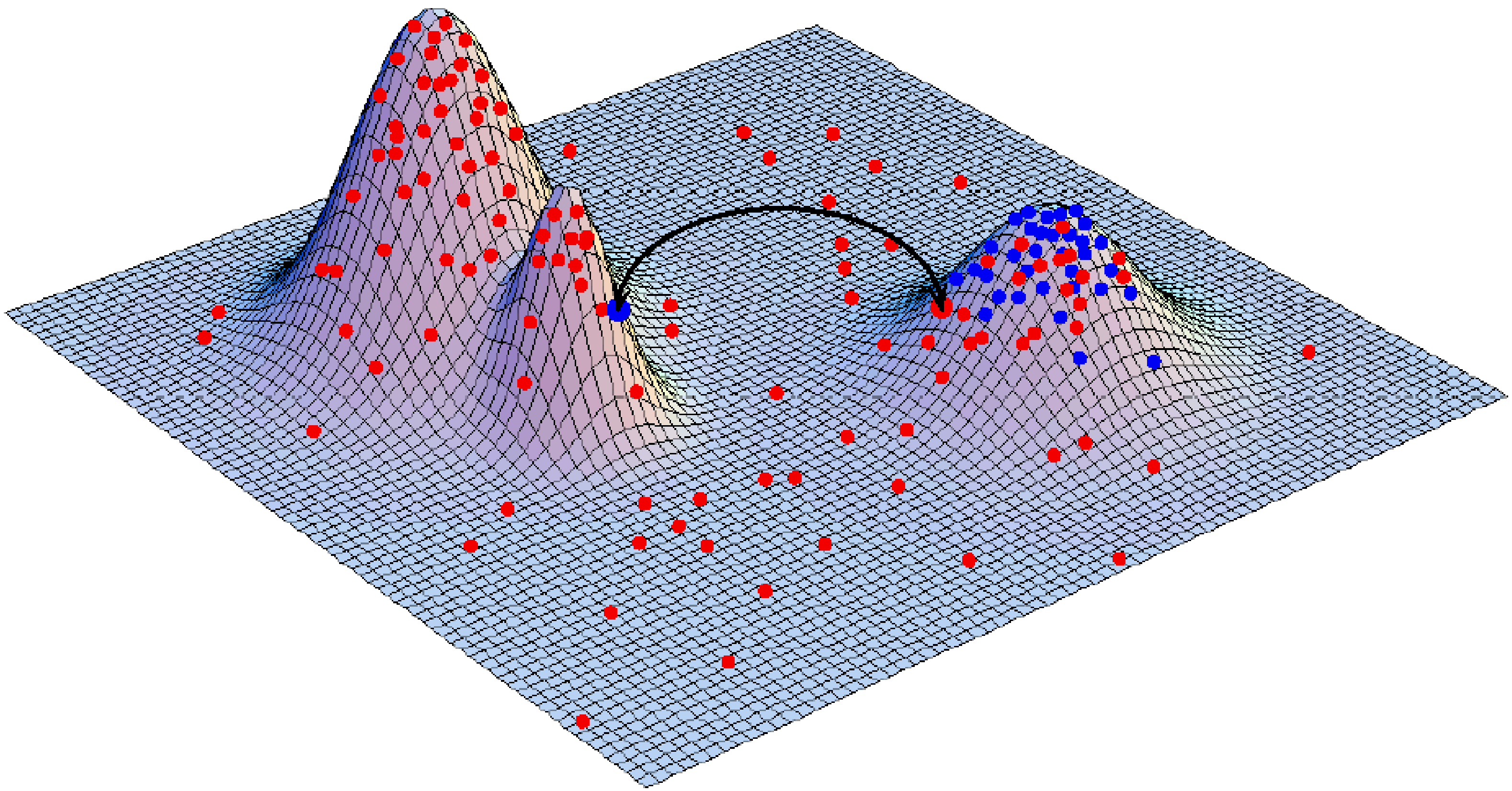
Convergence = tested all of likelihood surface and found maximum
But this can be hard.

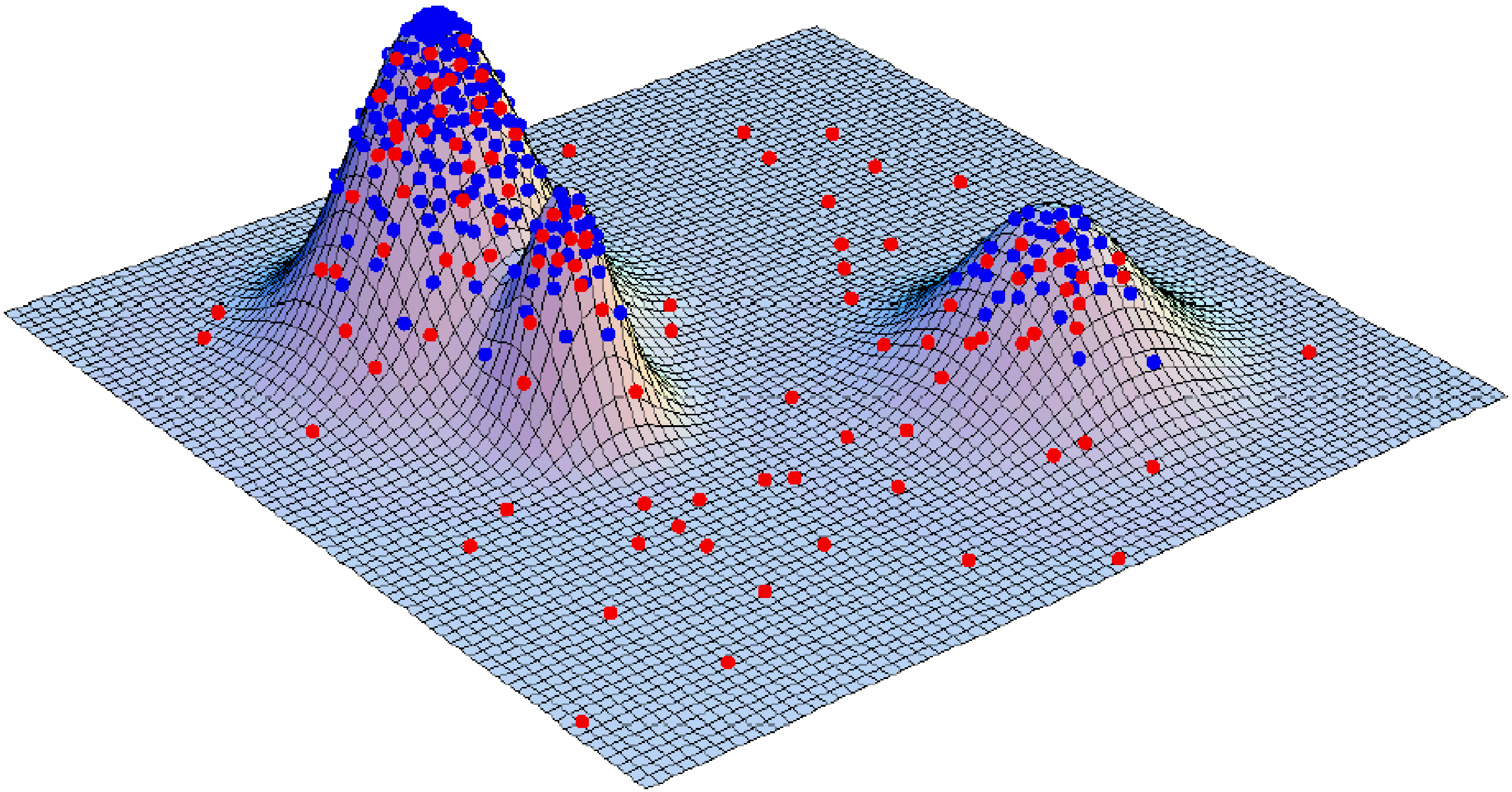
For better mixing/convergence:

1. **Multiple chains starting different initial values**
2. **Run for very long time**
3. **Increase the step size at the beginning**
4. **Use good proposals**
5. **Swap between chain**

...

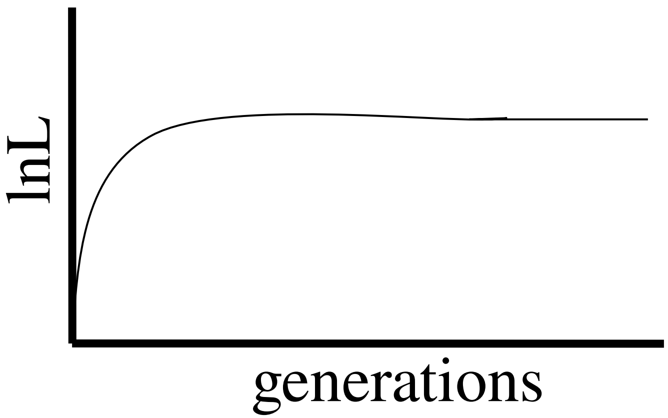
Swap between chains



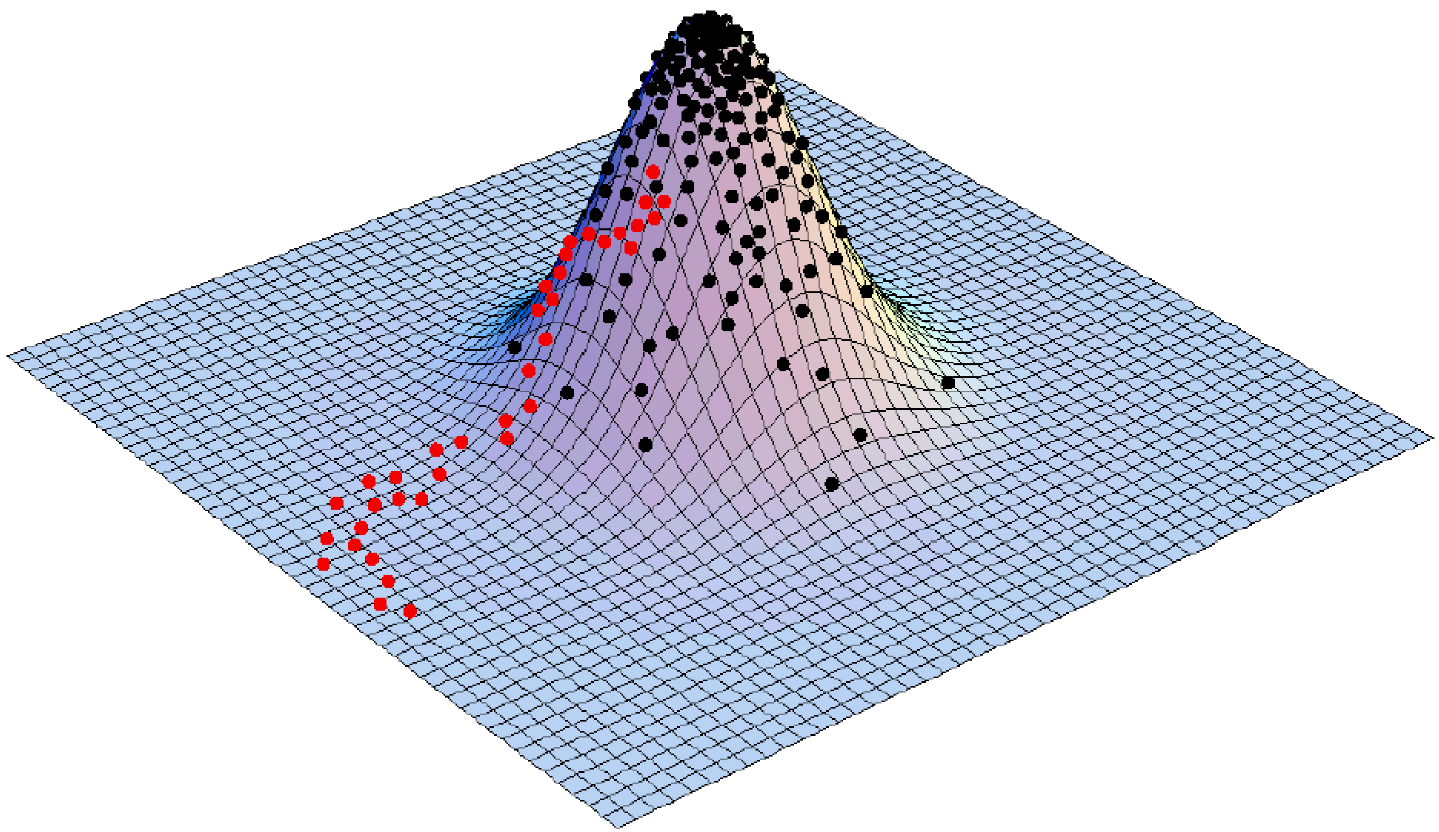


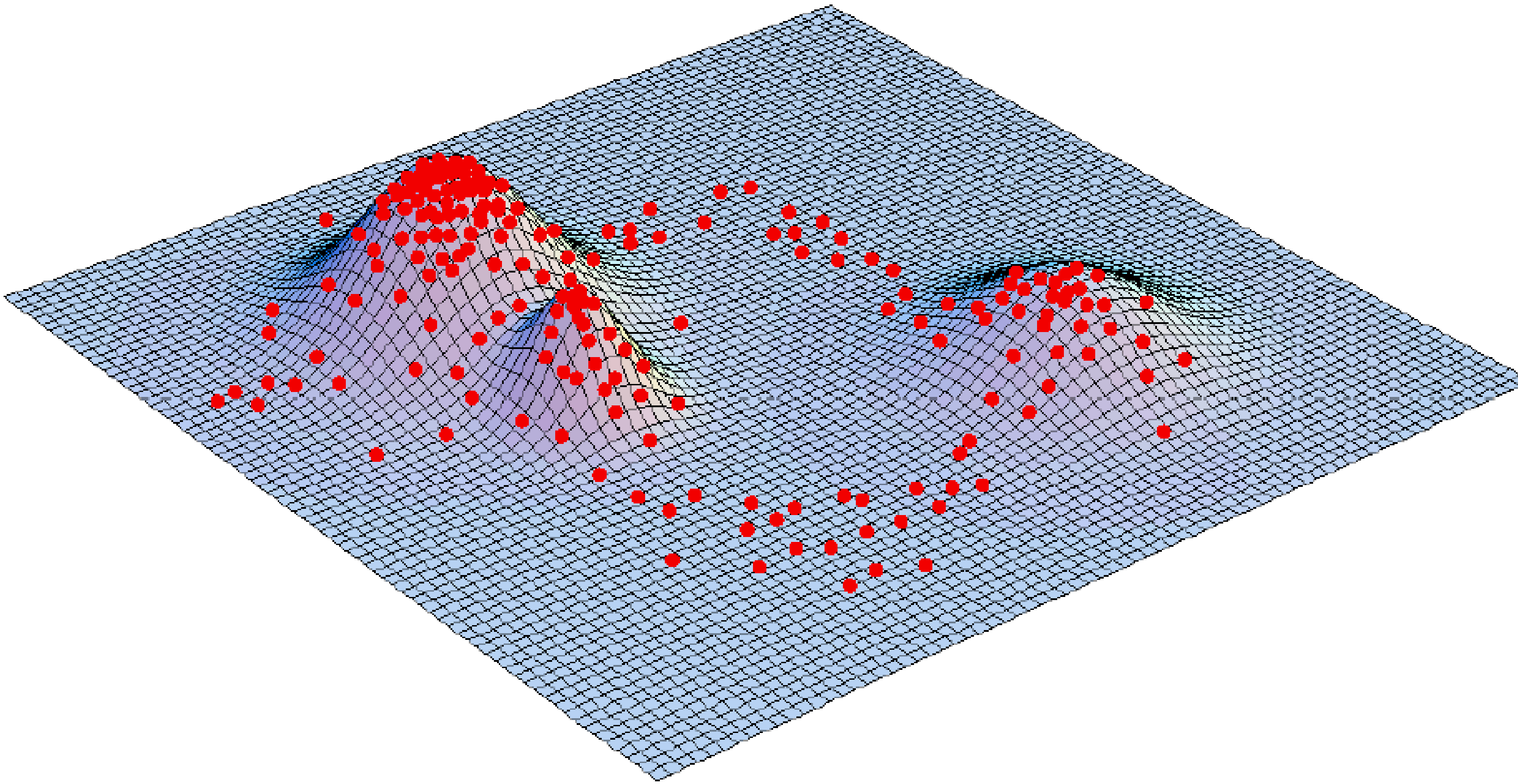
What other information can you get from MCMC methods?

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

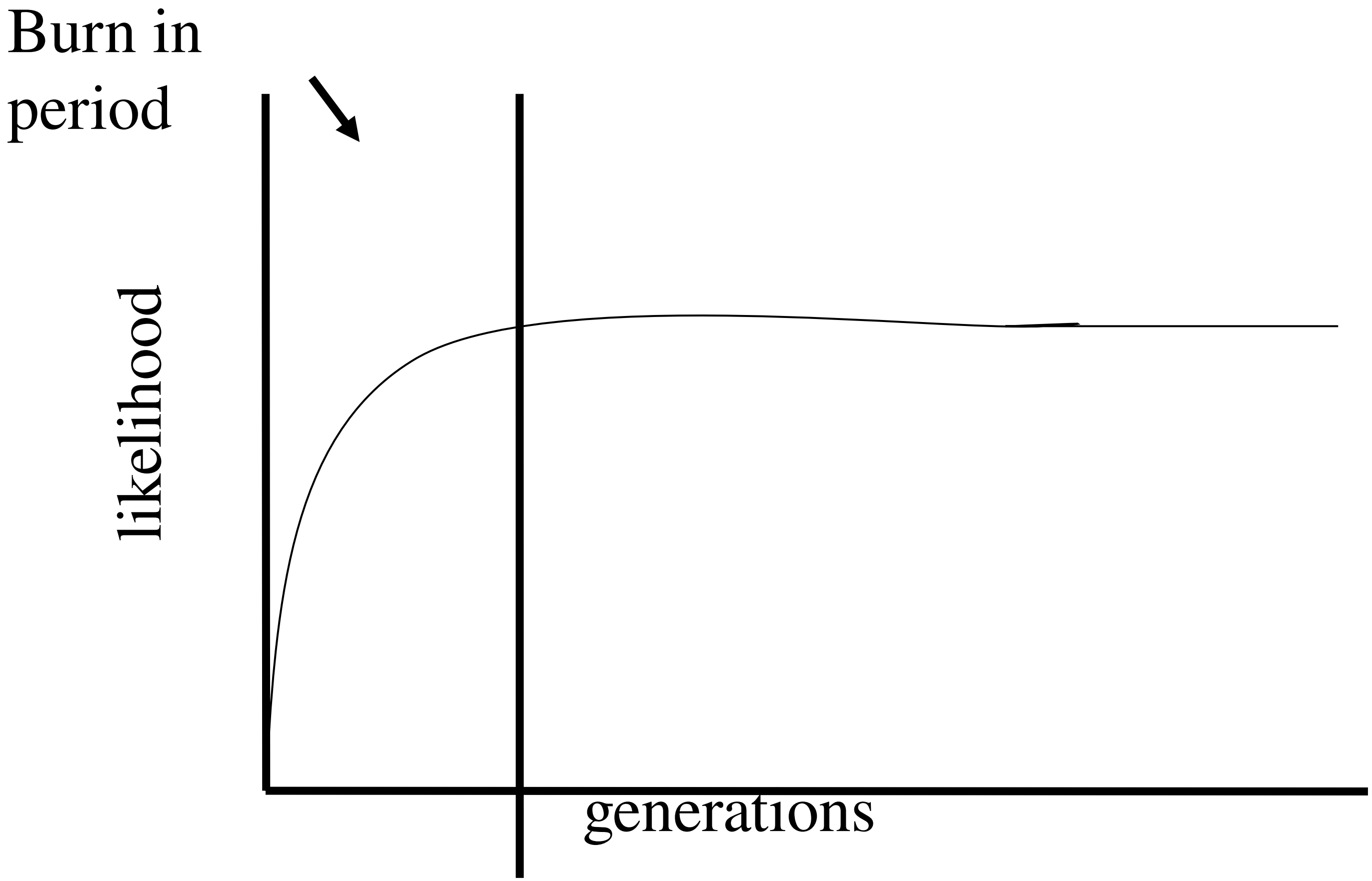


Red dots = “burn in” period.





Red dots = “burn in” period.



Note: burning-in is a common practice, but not necessary at all.

Some things to consider when running MCMC analyses

- Length of a chain
- Number of chains
- Burn-in period
- Convergence