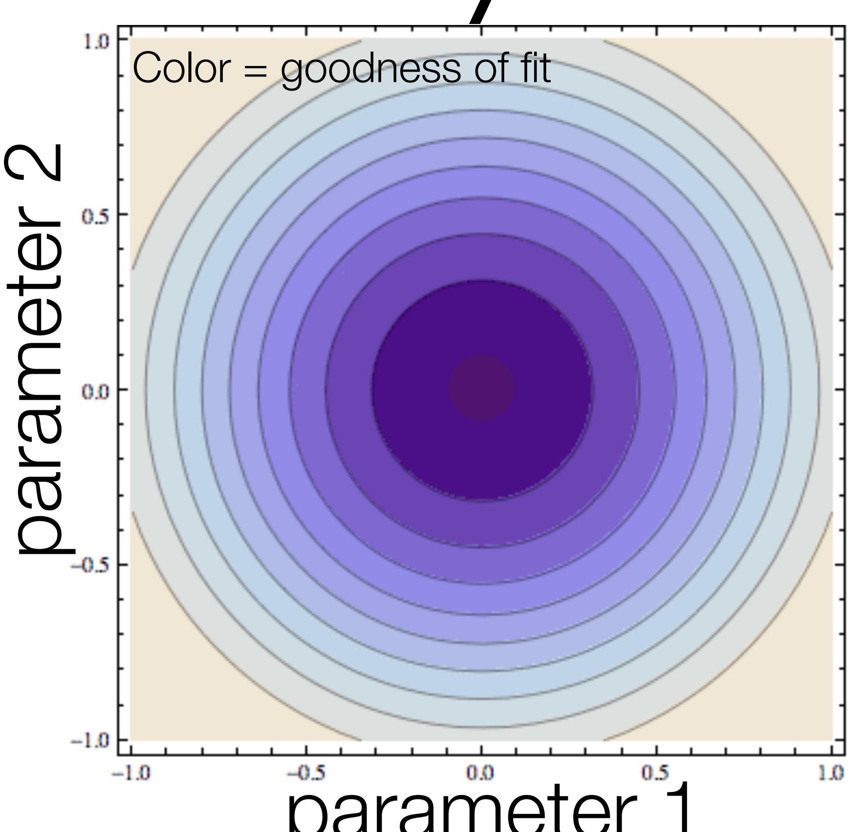
Parameter Estimation & Maximum Likelihood

Marisa Eisenberg
Epid 814

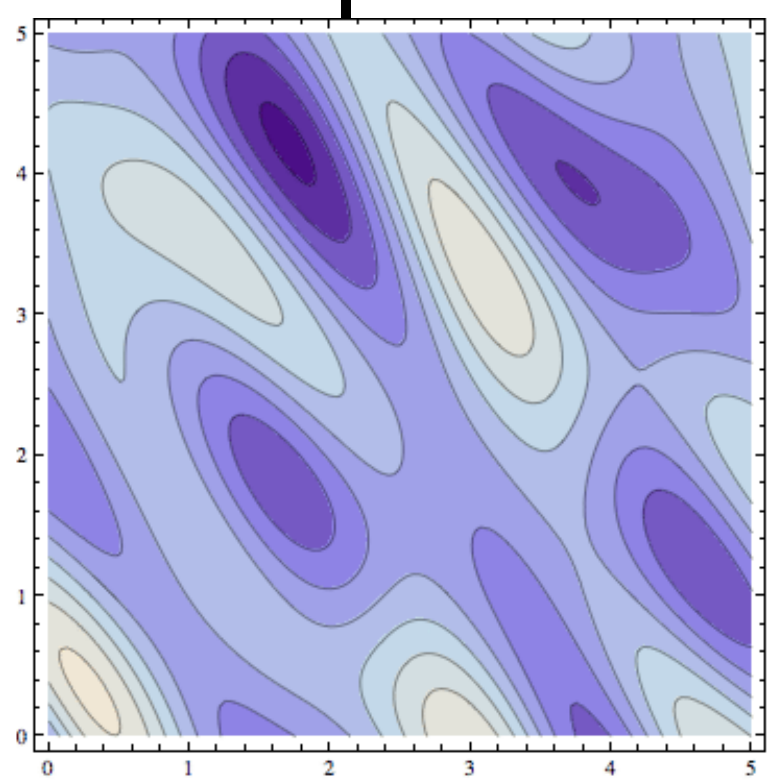
Parameter Estimation

- In general—search parameter space to find optimal fit to data
- Or to characterize distribution of parameters that matches data

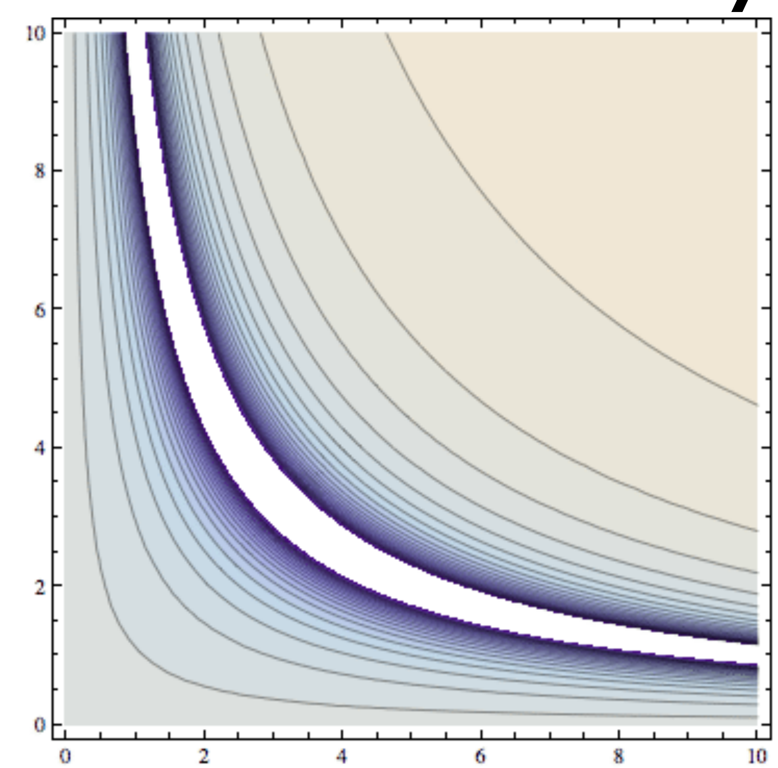
Yay!



Multiple Mins

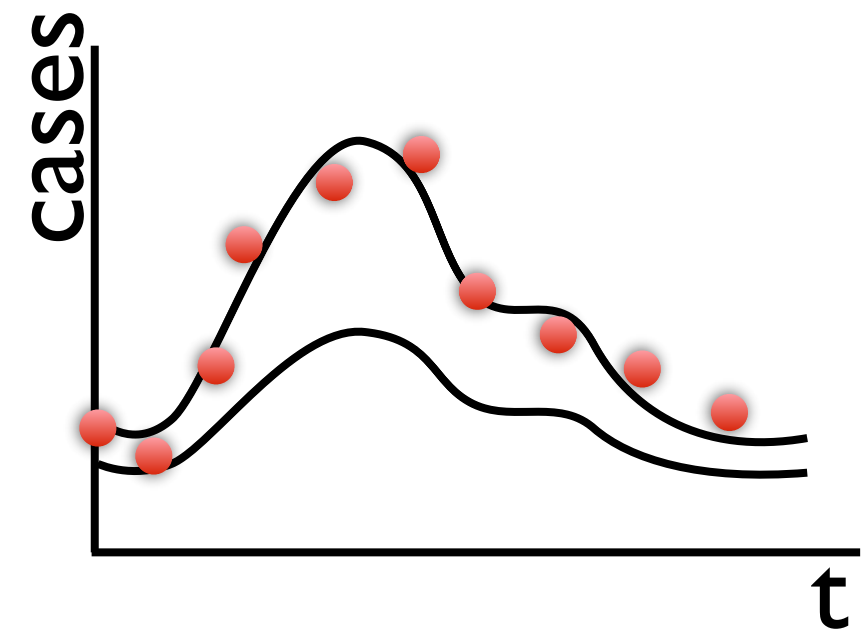


Unidentifiability



Parameter Estimation

- Basic idea: parameters that give model behavior that more closely matches data are 'best' or 'most likely'



- Frame this from a statistical perspective (inference, regression)
    - Can determine 'most likely' parameters or distribution, confidence intervals, etc.

How to frame this statistically?

- **Maximum Likelihood Approach**
- Idea: rewrite the ODE model as a statistical model, where we suppose we know the general form of the density function but not the parameter values
- Then if we knew the parameters we could calculate probability of a particular observation/data:

$$P(\underset{\text{data}}{z} \mid \underset{\text{parameters}}{p})$$

Maximum Likelihood

- **Likelihood Function**

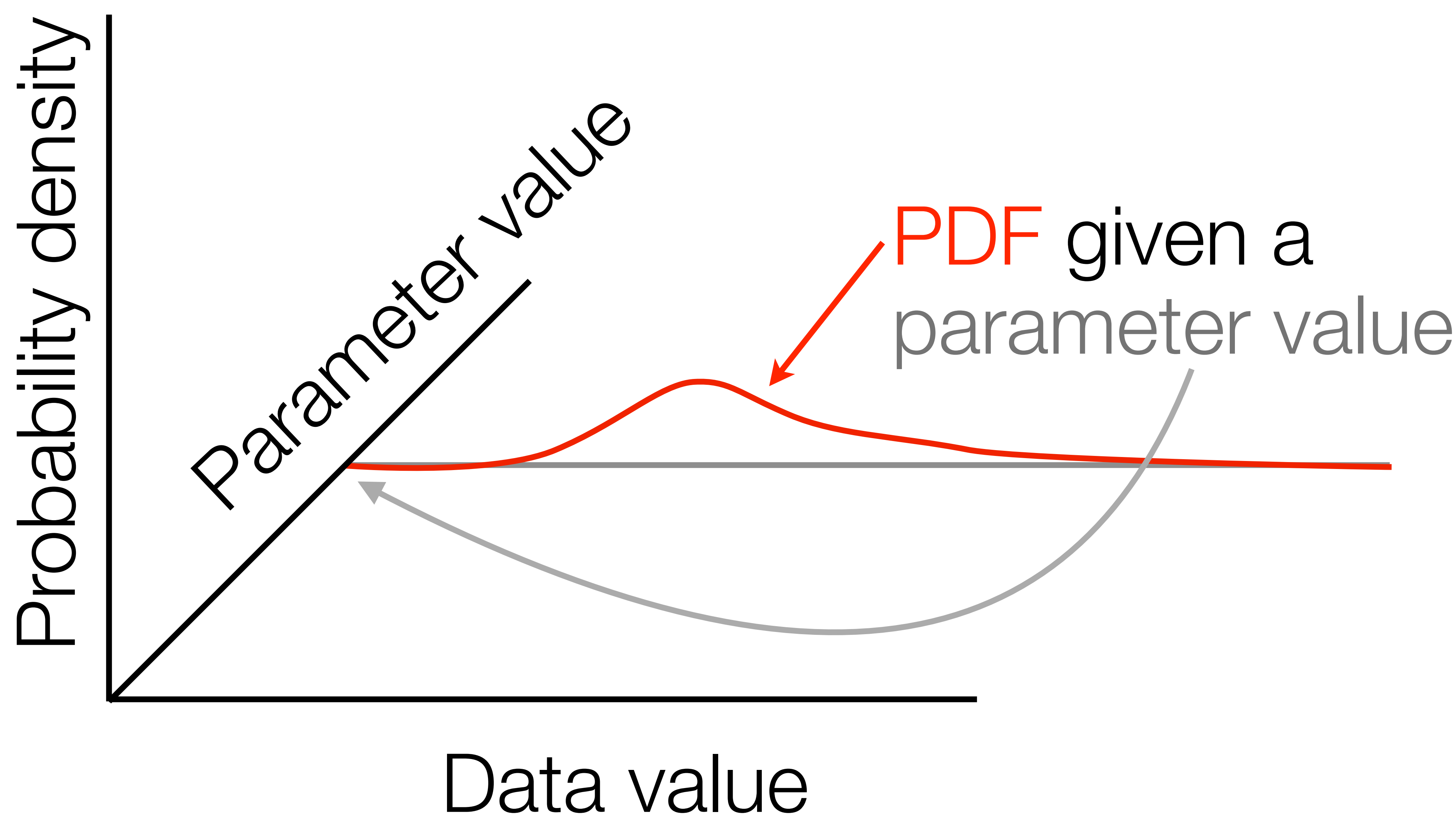
$$P(z \mid p) = f(z, p) = L(p \mid z)$$

- Re-think the distribution as a function of the data instead of the parameters

- E.g. $f(z \mid \mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(z-\mu)^2}{2\sigma^2}\right) = L(\mu, \sigma^2 \mid z)$

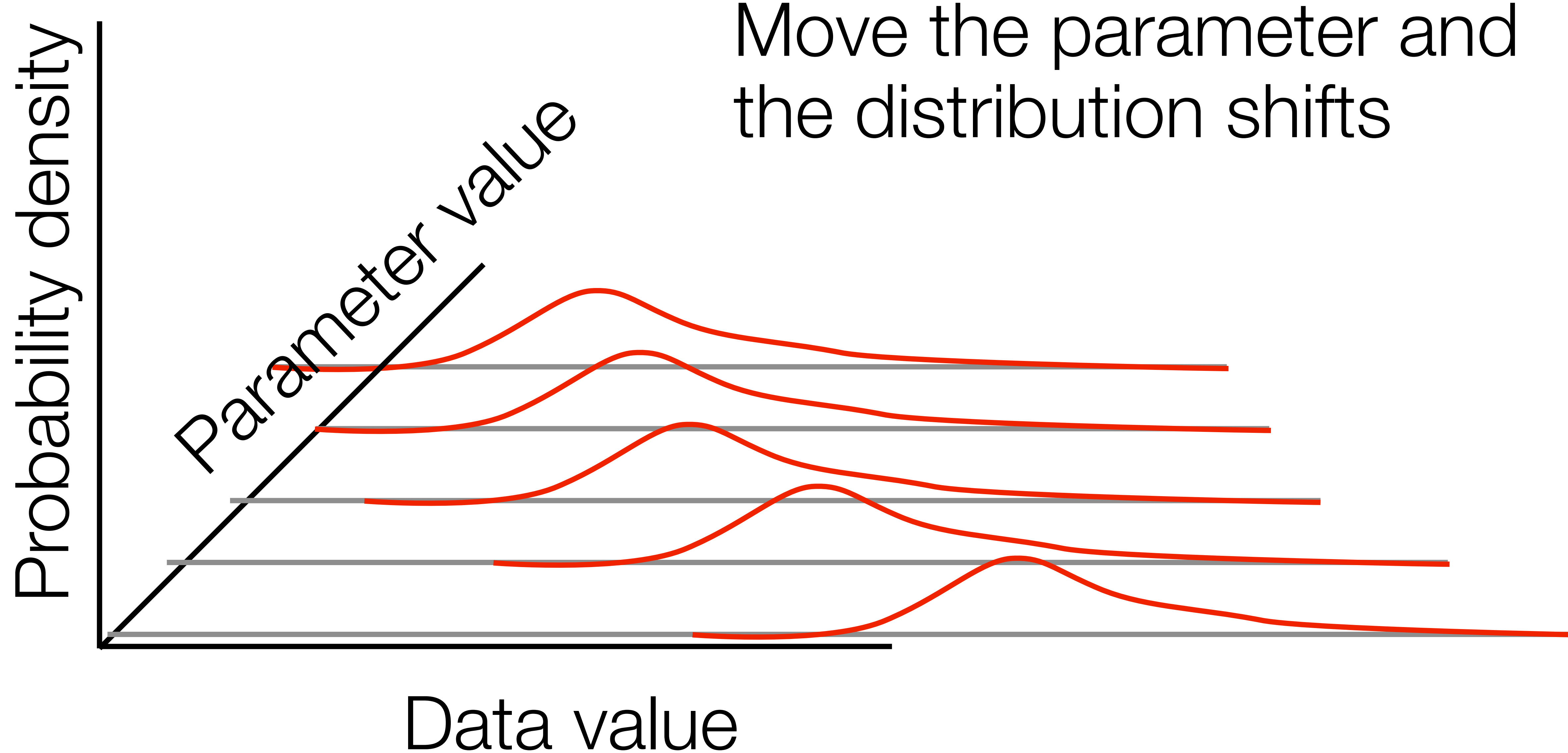
- Find the value of p that maximizes L(p|z) - this is the maximum likelihood estimate (**MLE**) (most likely given the data)

Likelihood Function

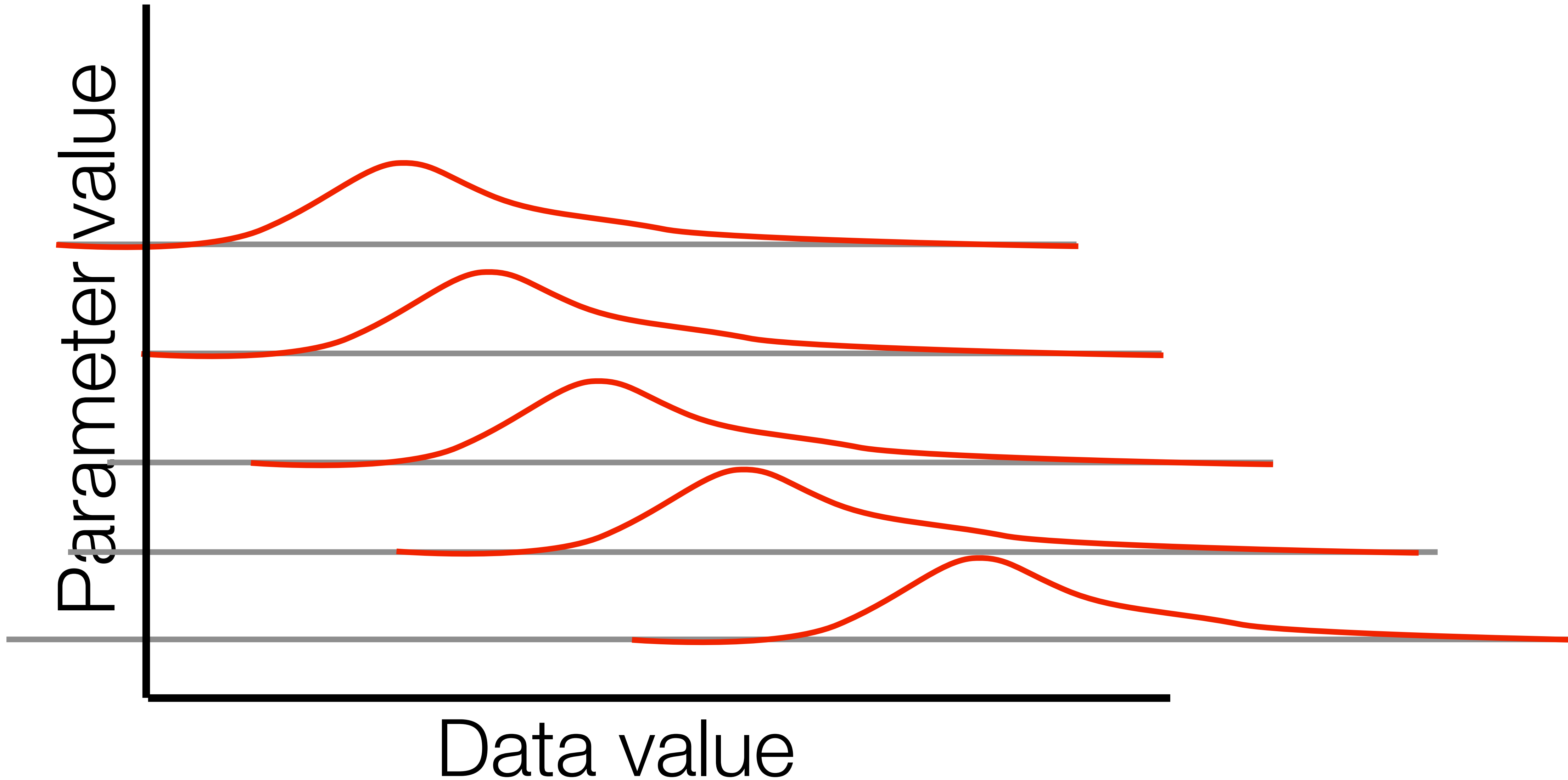


Likelihood Function

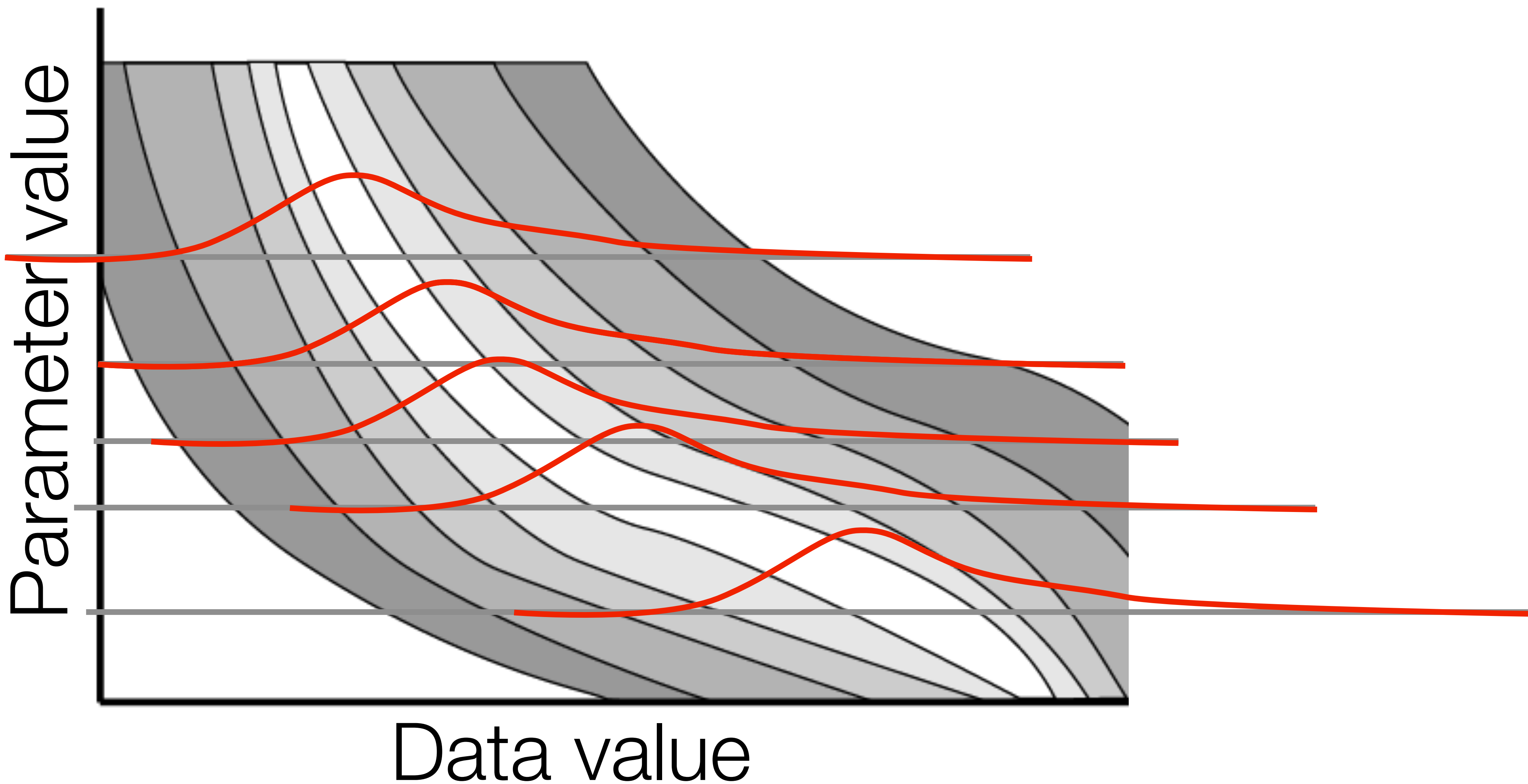
Move the parameter and the distribution shifts



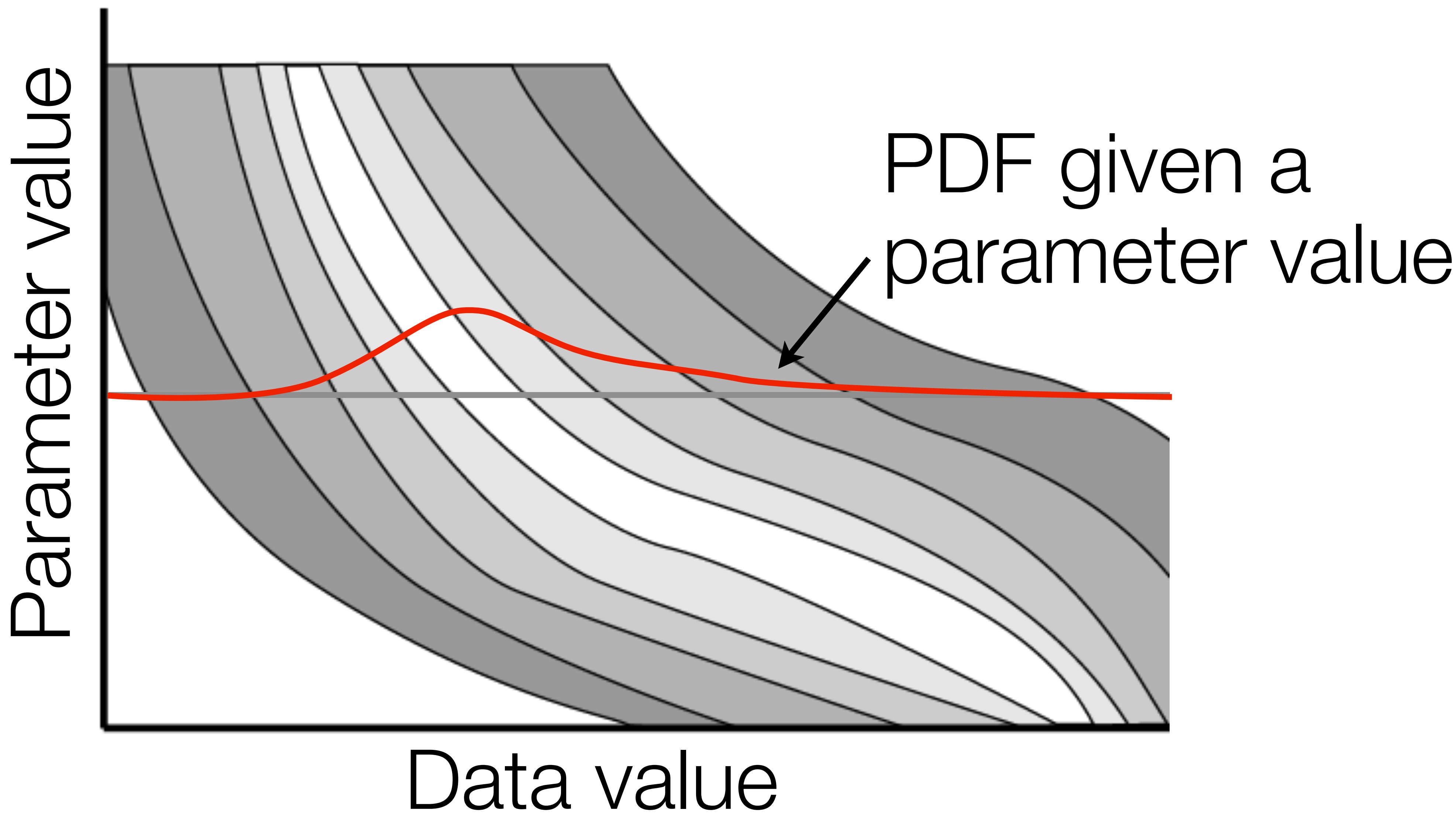
Likelihood Function



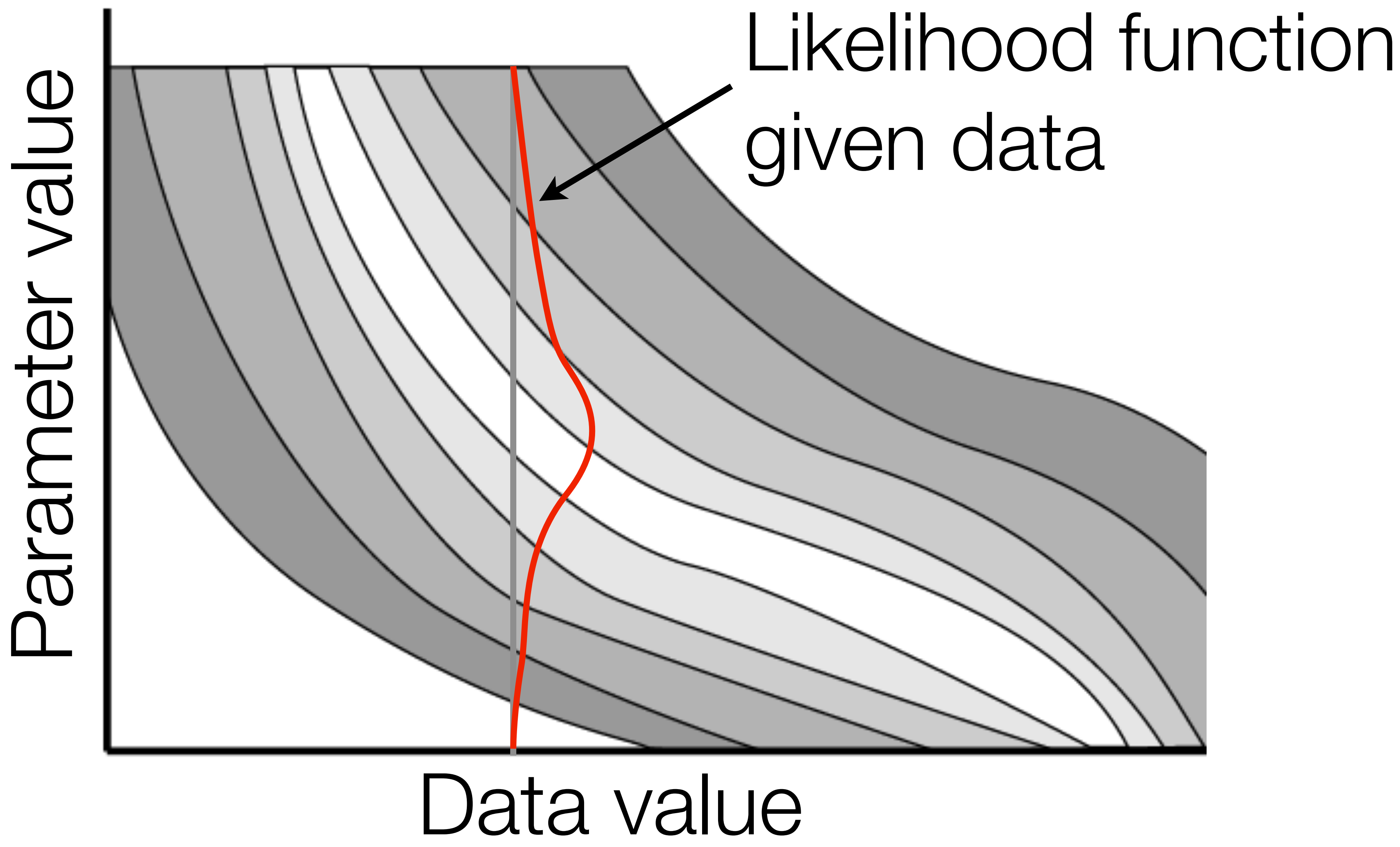
Likelihood Function



Likelihood Function



Likelihood Function



Maximum Likelihood

- **Consistency** - with sufficiently large number of observations n, it is possible to find the value of p with arbitrary precision (i.e. converges in probability to p)

- **Normality** - as the sample size increases, the distribution of the MLE tends to a Gaussian distribution with mean and covariance matrix equal to the inverse of the Fisher information matrix

- **Efficiency** - achieves CR bound as sample size$\longrightarrow\infty$ (no consistent estimator has lower asymptotic mean squared error than MLE)

Likelihood functions

- In general, your likelihood is just the probability distribution of your data, written in terms of your model
- Then, we ‘re-think’ of that distribution as a function of the parameters with the data fixed

Likelihood functions

- For example—what might a model and likelihood function be for the following situations:
  - Measure: 3 coin tosses,
    Parameter to estimate: coin bias (i.e. % heads)
  - Measure: incidence of bicycle accidents each year
    Parameter to estimate: rate of bicycle accidents
  - Measure: age information (maybe other covariates) and current happiness levels in a sample of people
    Parameters to estimate: effect of age & other covariateson happiness level
  - Measure: incidence of bicycle accidents each year
    Parameter to estimate: daily probability of a bicycle accident per square meter

Example - ODE Model with Gaussian Error

- Model:

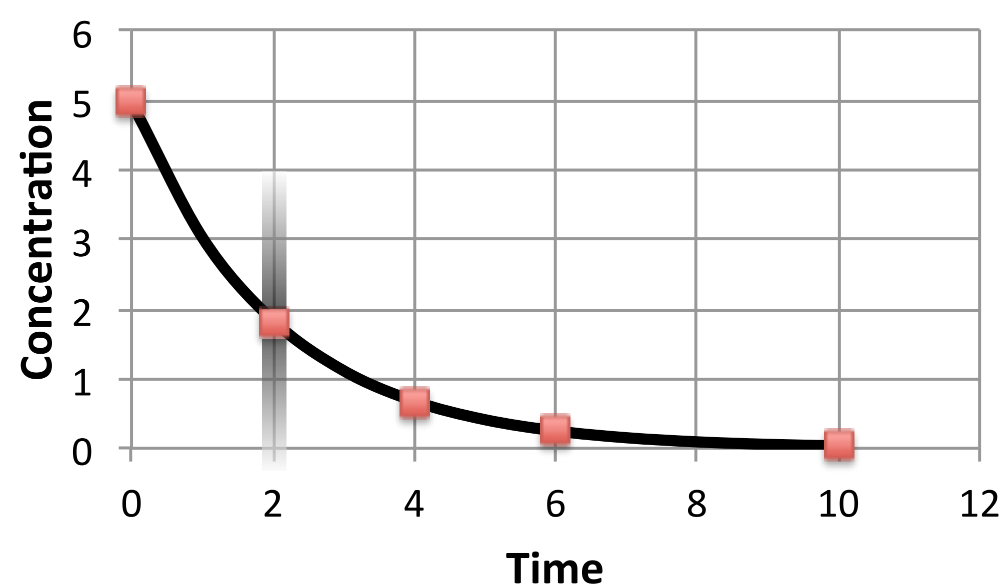
$$\dot{x} = f\left(x, t, p\right)$$

$$y = g(x, t, p)$$

- Suppose data is taken at times $t_1, t_2, \ldots, t_n$
- Data at t_i = $z_i = y\left(t_i\right) + e_i$
- Suppose error is gaussian and unbiased, with known variance σ^2 (can also be considered an unknown parameter)

Example - ODE Model with Gaussian Error

- The measured data z_i at time i can be viewed as a sample from a Gaussian distribution with mean y(x, t_i,p) and variance σ^2



- Suppose all measurements are independent (is this realistic?)

Example - ODE Model with Gaussian Error

- Then the likelihood function can be calculated as:

Gaussian PDF: $$f\left(z_i \mid \mu,\sigma^2\right) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{\left(z_i-\mu\right)^2}{2\sigma^2}\right)$$

Example - ODE Model with Gaussian Error

- Then the likelihood function can be calculated as:

Gaussian PDF: $$f\left(z_i \mid \mu,\sigma^2\right) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{\left(z_i-\mu\right)^2}{2\sigma^2}\right)$$

Formatted for model: $$f\left(z_i \mid y\left(x,t_i,p\right),\sigma^2\right) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{\left(z_i-y\left(t_i,p\right)\right)^2}{2\sigma^2}\right)$$

Example - ODE Model with Gaussian Error

- Then the likelihood function can be calculated as:

Gaussian PDF: $$f\left(z_i \mid \mu, \sigma^2\right) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{\left(z_i - \mu\right)^2}{2\sigma^2}\right)$$

Formatted for model: $$f\left(z_i \mid y\left(x, t_i, p\right), \sigma^2\right) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{\left(z_i - y\left(t_i, p\right)\right)^2}{2\sigma^2}\right)$$

Likelihood function assuming independent observations:

$$L\left(y\left(t_i, p\right), \sigma^2 \mid z_1, \ldots, z_n\right) = f\left(z_1, \ldots, z_n \mid y\left(t_i, p\right), \sigma^2\right)$$

$$= \prod_{i=1}^{n} f\left(z_i \mid y\left(t_i, p\right), \sigma^2\right)$$

Example - ODE Model with Gaussian Error

$$L\left(y\left(t_i,p\right),\sigma^2 \mid z_1,\ldots,z_n\right) = f\left(z_1,\ldots,z_n \mid y\left(t_i,p\right),\sigma^2\right)$$

$$= \prod_{i=1}^{n} f\left(z_i \mid y\left(t_i,p\right),\sigma^2\right)$$

$$= \left(\frac{1}{2\pi\sigma^2}\right)^{n/2} \exp\left(-\frac{\sum_{i=1}^{n}\left(z_i - y\left(t_i,p\right)\right)^2}{2\sigma^2}\right)$$

Example - ODE Model with Gaussian Error

- It is often more convenient to minimize the Negative Log Likelihood (-LL) instead of maximizing the Likelihood
    - Log is well behaved, minimization algorithms common

$$-LL = -\ln\left(\left(\frac{1}{2\pi\sigma^2}\right)^{n/2}\exp\left(-\frac{\sum_{i=1}^{n}\left(z_i - y\left(t_i, p\right)\right)^2}{2\sigma^2}\right)\right)$$

Example - ODE Model with Gaussian Error

$$-LL = -\ln\left(\left(\frac{1}{2\pi\sigma^2}\right)^{n/2}\exp\left(-\frac{\sum_{i=1}^{n}\left(z_i - y\left(t_i, p\right)\right)^2}{2\sigma^2}\right)\right)$$

$$-LL = -\left(-\frac{n}{2}\ln\left(2\pi\right) - n\ln\left(\sigma\right) - \frac{\sum_{i=1}^{n}\left(z_i - y\left(t_i, p\right)\right)^2}{2\sigma^2}\right)$$

Example - ODE Model with Gaussian Error

$$-LL = \frac{n}{2}\ln(2\pi) + n\ln(\sigma) + \frac{\sum_{i=1}^{n}\left(z_i - y(t_i, p)\right)^2}{2\sigma^2}$$

If σ is known, then first two terms are constants & will not be changed as p is varied—so we can minimize only the 3rd term and get the same answer

$$\min_p(-LL) = \min_p\left(\frac{\sum_{i=1}^{n}\left(z_i - y(t_i, p)\right)^2}{2\sigma^2}\right)$$

Example - ODE Model with Gaussian Error

- Similarly for denominator:

$$\min_p\left(-LL\right) = \min_p\left(\frac{\sum_{i=1}^{n}\left(z_i - y\left(t_i, p\right)\right)^2}{2\sigma^2}\right) = \min_p\left(\sum_{i=1}^{n}\left(z_i - y\left(t_i, p\right)\right)^2\right)$$

- This is just least squares!

- So, least squares is equivalent to the ML estimator when we assume a constant known variance

Let’s code this likelihood function for an SIR model!

- Switch to R and code together

Maximum Likelihood Summary for ODEs

- Can calculate other ML estimators for different distributions
- Not always least squares-ish! (mostly not)
- Although surprisingly, least squares does fairly decently a lot of the time

Example - Poisson ML

- For count data (e.g. incidence data), the Poisson distribution is often more realistic than Gaussian
- Likelihood function?

Example - Poisson ML

- Model: $$\dot{x} = f\left(x, t, p\right)$$ $$y = g(x, t, p)$$
- Data z_i is assumed to be Poisson with mean $y\left(t_i\right)$
- Assume all data points are independent
- Poisson PMF: $$f\left(z_i \mid y\left(t_i\right)\right) = \frac{y\left(t_i\right)^{z_i} e^{-y\left(t_i\right)}}{z_i!}$$

Example - Poisson ML

- Likelihood function:

$$L\left(y(t,p) \mid z_1,\ldots,z_n\right) = f\left(z_1,\ldots,z_n \mid y(t,p)\right)$$

$$= \prod_{i=1}^{n} f\left(z_i \mid y(t,p)\right)$$

$$= \prod_{i=1}^{n} \frac{y(t_i)^{z_i}\, e^{-y(t_i)}}{z_i\,!}$$

Poisson ML

- Negative log likelihood:

$$-LL = -\ln\left(\prod_{i=1}^{n} \frac{y(t_i)^{z_i}\, e^{-y(t_i)}}{z_i\,!}\right)$$

$$= -\sum_{i=1}^{n} \ln\left(\frac{y(t_i)^{z_i}\, e^{-y(t_i)}}{z_i\,!}\right)$$

$$= -\sum_{i=1}^{n} z_i \ln\left(y(t_i)\right) + \sum_{i=1}^{n} y(t_i) + \sum_{i=1}^{n} \ln(z_i)$$

- Last term is constant

Example - Poisson ML

- Poisson ML Estimator:

$$\min_p(-LL) = \min_p\left(-\sum_{i=1}^{n} z_i \ln\left(y\left(t_i\right)\right) + \sum_{i=1}^{n} y\left(t_i\right)\right)$$

- Other common distributions - negative binomial (overdispersion), zero-inflated poisson or negative binomial, etc.

Maximum Likelihood Summary for ODEs

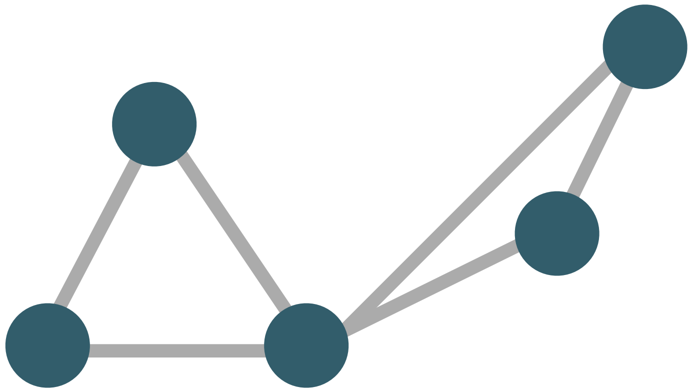
- Basic approach - suppose only measurement error
- Data is given by distribution where model output is the mean
- Suppose each time point of data is independent
- Use PDF/PMF to calculate the likelihood
- Take the negative log likelihood, minimize this over the parameter space

Maximum Likelihood for other kinds of models

- Can be quite different!
- May require more computation to evaluate (e.g. stochastic models)
- May also be structured quite differently! (e.g. network or individual-based models)

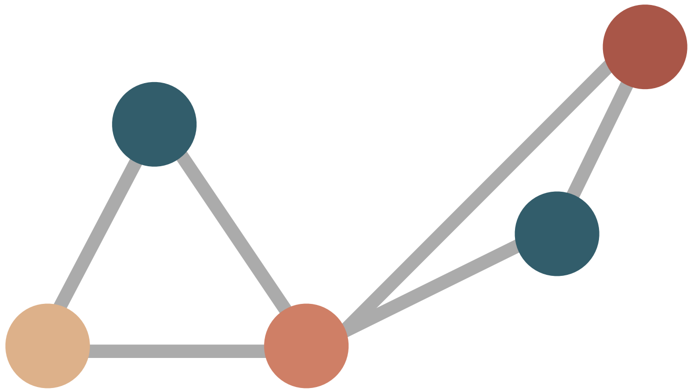
Tiny Network Example

- Data: infection pattern on the network
- Model: suppose constant probability *p* of infecting along an edge from someone who got sick before you
- What's the likelihood?



Tiny Network Example

- Data: infection pattern on the network
- Model: suppose constant probability *p* of infecting along an edge, assuming we start with first case
- What's the likelihood?
- Let's see how we would calculate it for a specific data set



- L(p,data) = P(susc nodes did not get sick)
  x P(infected nodes did get sick)

(note not actually independent though!)

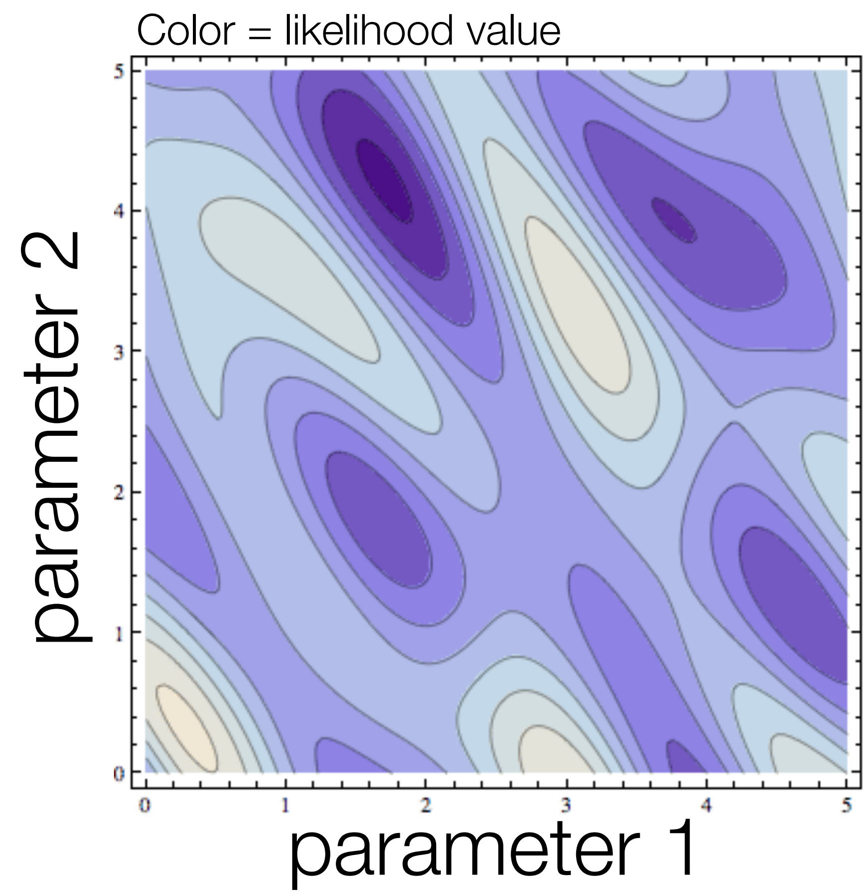
Now that we can write down a likelihood function, how do we find the maximum likelihood estimate?

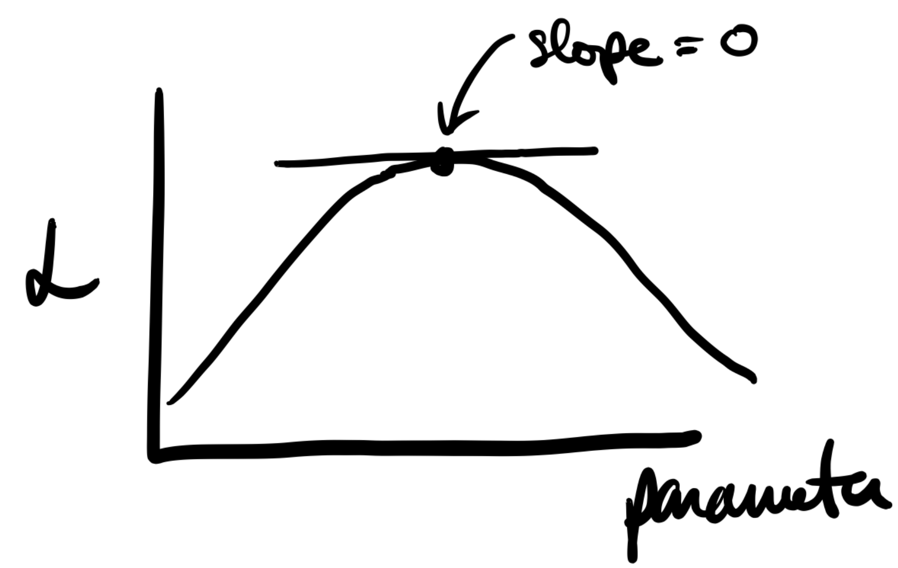
- For $\mathcal{L}(\theta, z)$ how to find

$$\hat{\theta} = \arg\max_{\theta \in \Theta} \mathcal{L}(\theta, z)$$

- For simple examples (e.g. coin toss, linear regression model, simple Poisson model), we can calculate what values of the parameters will maximize L explicitly! (Take derivatives of L and set = 0)

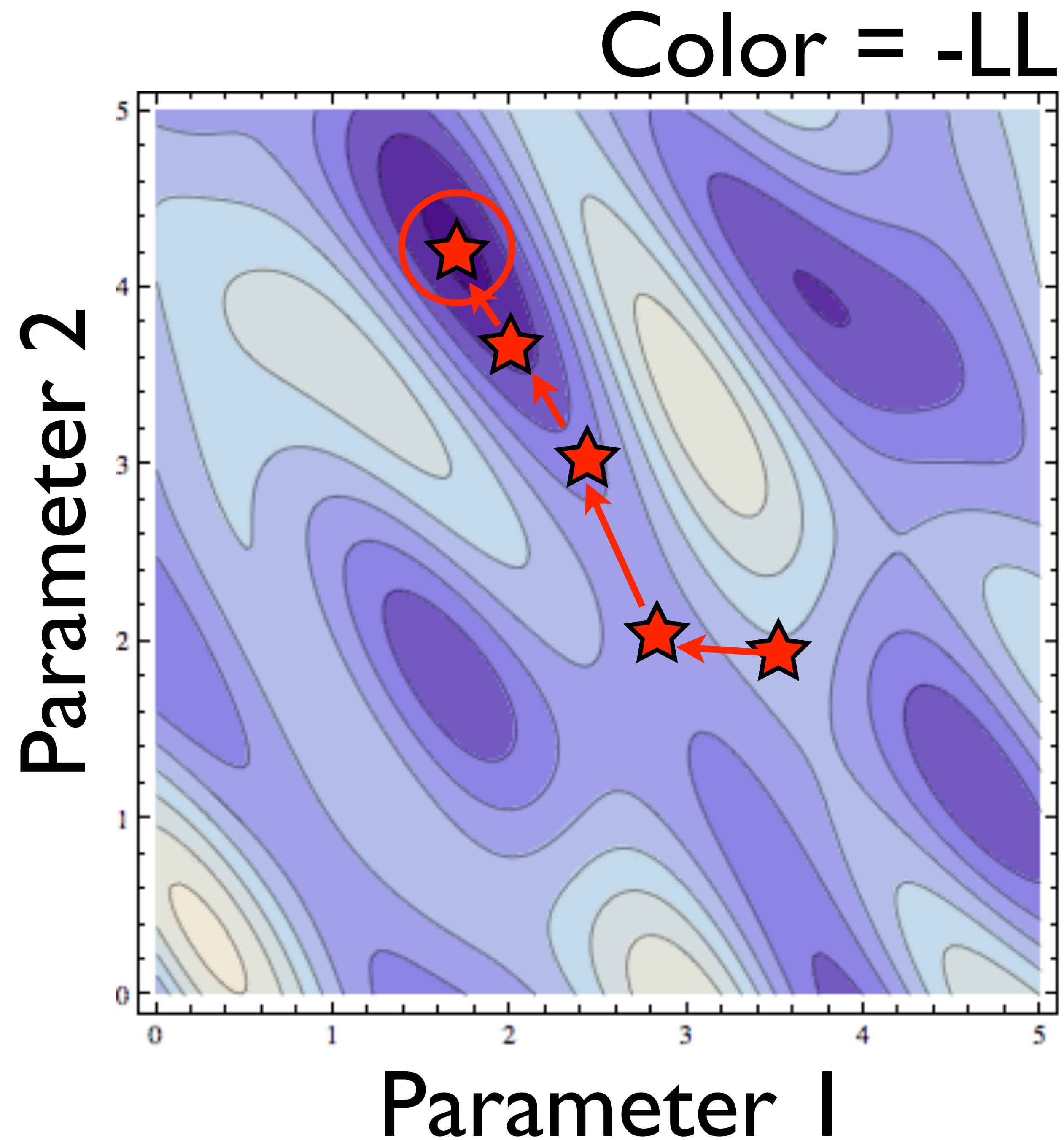
- But what if more complicated? This may not be possible—need to use numerical optimization. Most complex systems models fall into this category





Parameter Estimation Algorithms (Optimizers)

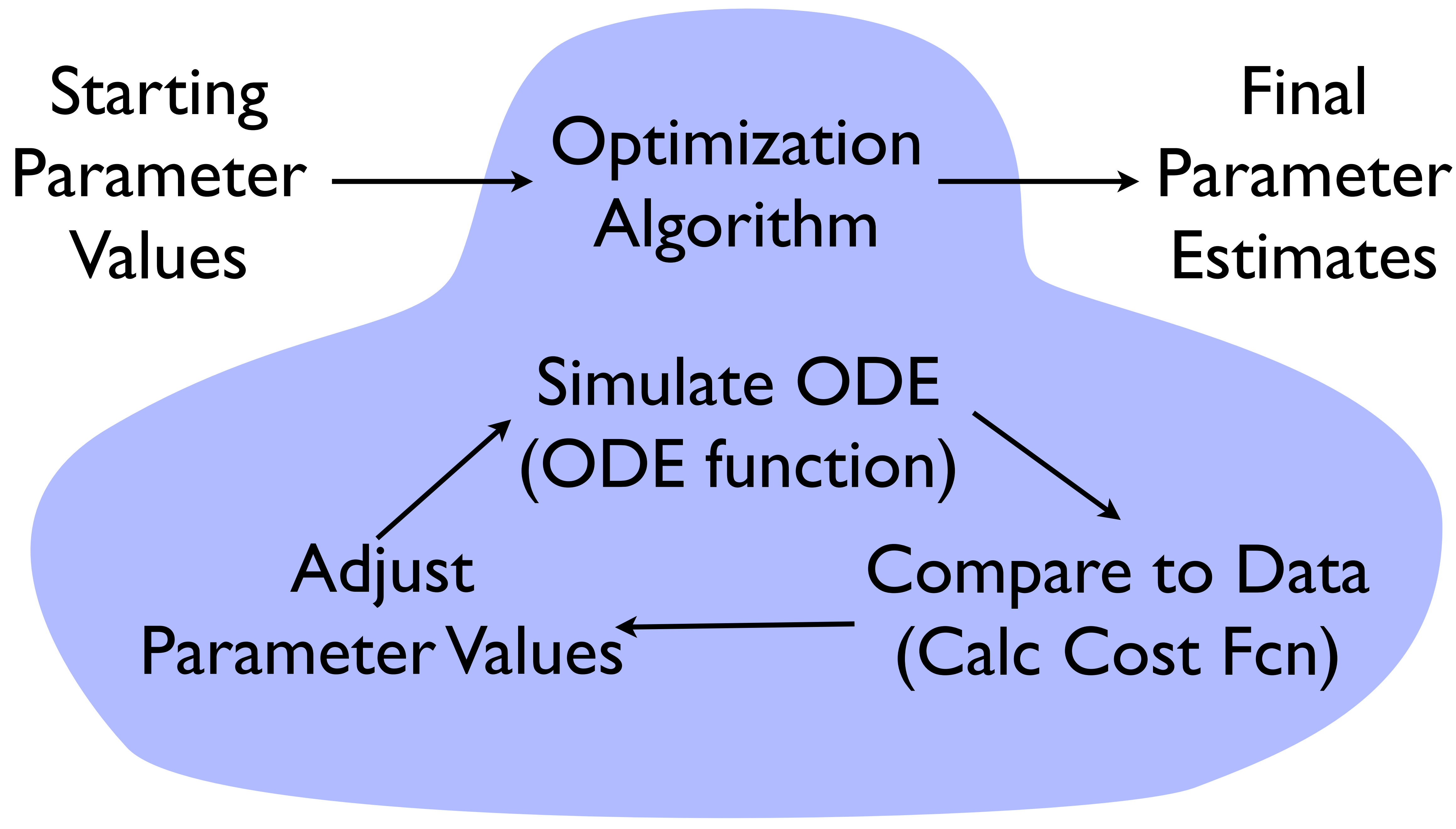
- Starting values for parameters
- Optimization algorithm searches parameter space to minimize RSS or -LL
- Converges once it finds a minimum



Parameter Estimation in R

- Need several pieces
  - ODE function that allows you to pass parameters
  - Cost function - something to calculate the RSS or -LL
  - Optimization function
    (e.g. optim)

Basic Idea


Starting Parameter Values
Optimization Algorithm
Final Parameter Estimates
Simulate ODE (ODE function)
Compare to Data (Calc Cost Fcn)
Adjust Parameter Values

Let’s code this up in R using the SIR model from before!

- Switch to R & code together

Very (very!) brief intro to Bayesian Approaches to Parameter Estimation

- Allows one to account for prior information about the parameters
  - E.g. previous studies in a similar population
- Update parameter information based on new data
- Recall Bayes' Theorem:

$$P(p \mid z) = P(params \mid data) = \frac{P(z \mid p) \cdot P(p)}{P(z)}$$

Very (very!) brief intro to Bayesian Approaches to Parameter Estimation

- Allows one to account for prior information about the parameters
  - E.g. previous studies in a similar population
- Update parameter information based on new data
- Recall Bayes' Theorem:

Prior distribution

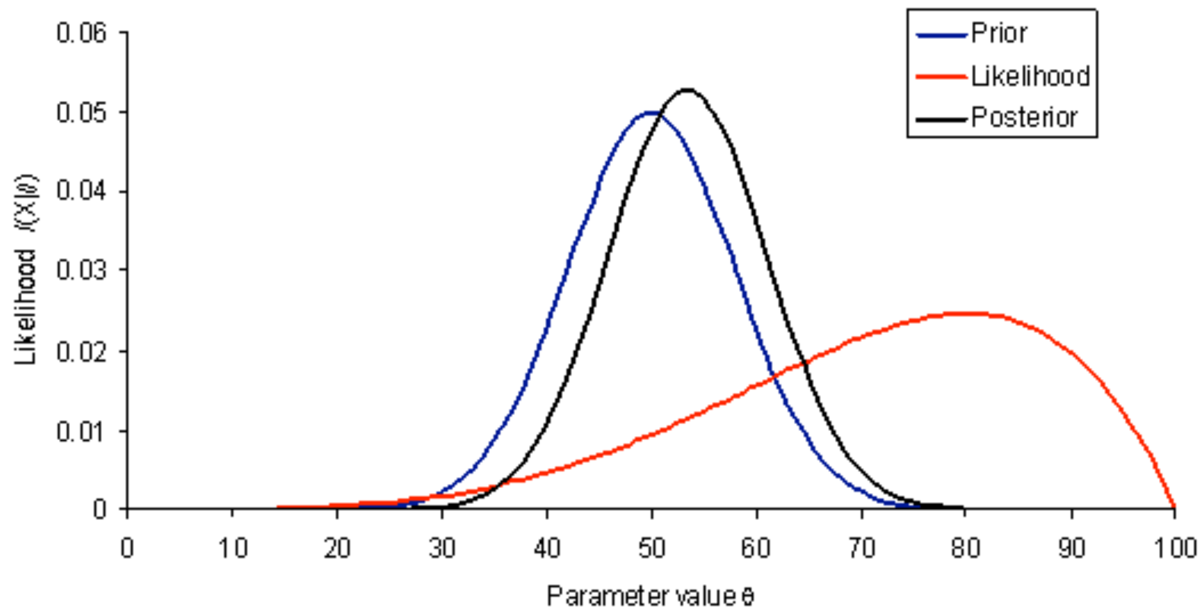
Likelihood

$$P(p \mid z) = P(params \mid data) = \frac{P(z \mid p) \cdot P(p)}{P(z)}$$

Normalizing constant
(can be difficult to calculate!)

Bayesian Parameter Estimation

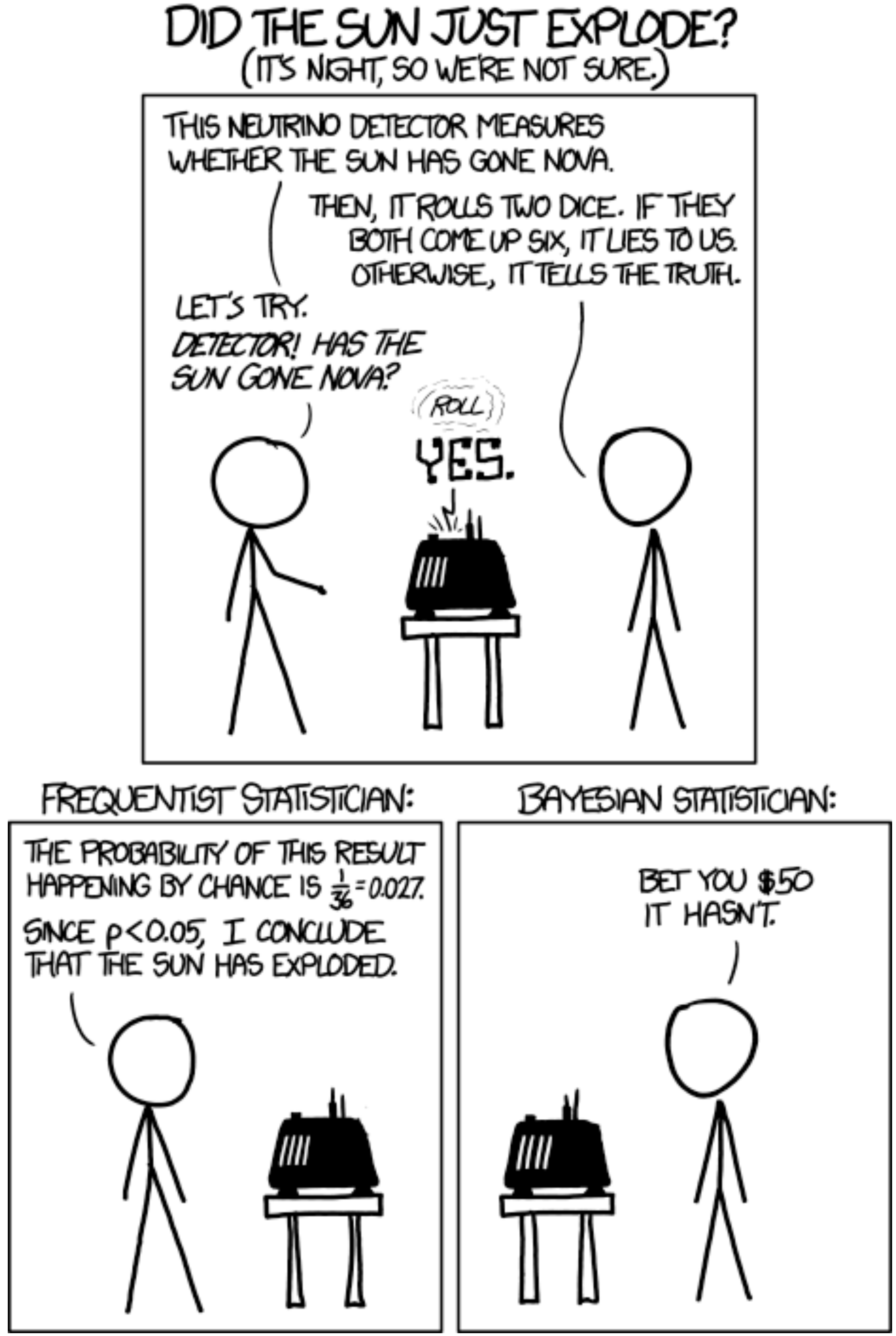
- From prior distribution & likelihood distribution, determine the posterior distribution of the parameter



- Can repeat this process as new data is available

Bayesian Parameter Estimation

- Treats the parameters inherently as distributions (belief)
- Philosophical battle between Bayesian & frequentist perspectives
- Word of caution on choosing your priors
- Denominator issues - MAP Approach



from XKCD:
http://xkcd.com/1132/