Lecture 14: Introduction to parameter estimation

Complex Systems 530

3/26/20

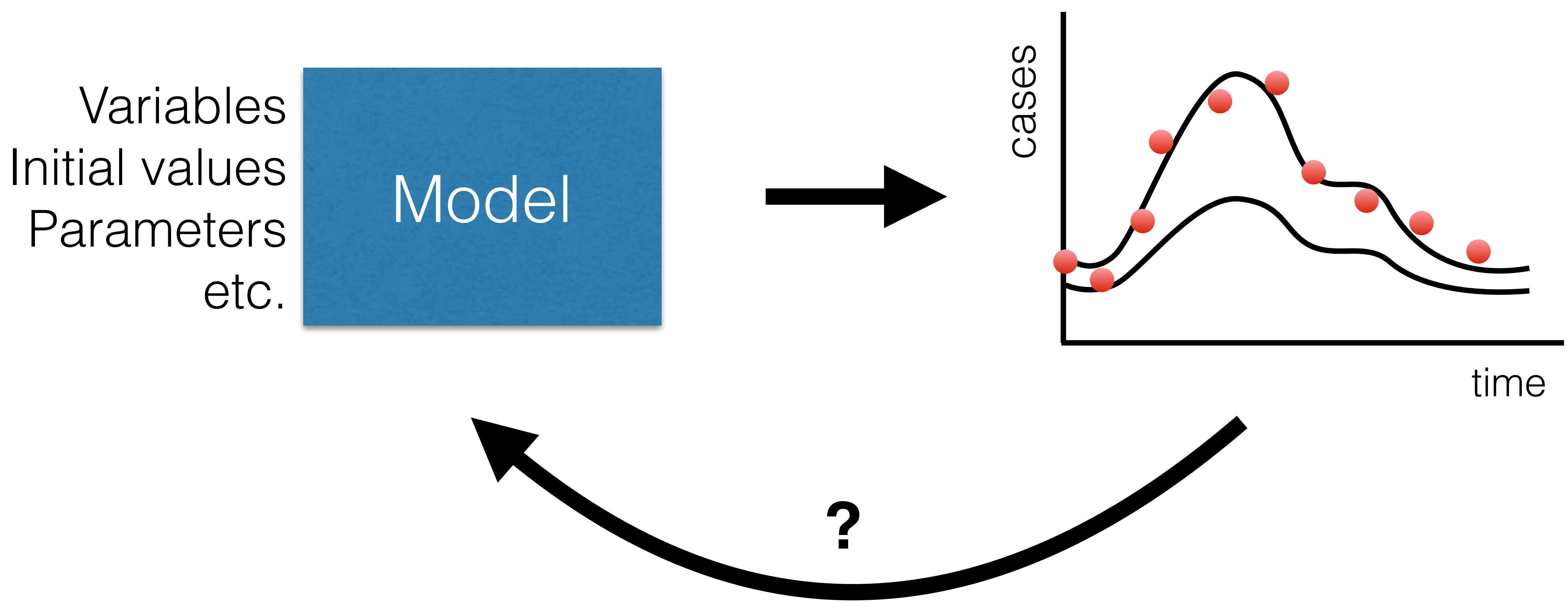
Outline

- Quick update about lab
- Today
  - Intro to parameter estimation (for models in general, not just ABMs)
  - Some of the challenges involved in using these tools for ABMs
- Next time: Bayesian & sampling based approaches, intro to MCMC

Connecting Models with Data

- Depending on parameters, models can give very different results
- How to figure out parameters for model?
    - Direct measurements of parameters often difficult
- If we have data on what is observed in the real world, this may be able to tell us something about what parts of parameter space are more realistic?
- How to connect models with data?

General idea

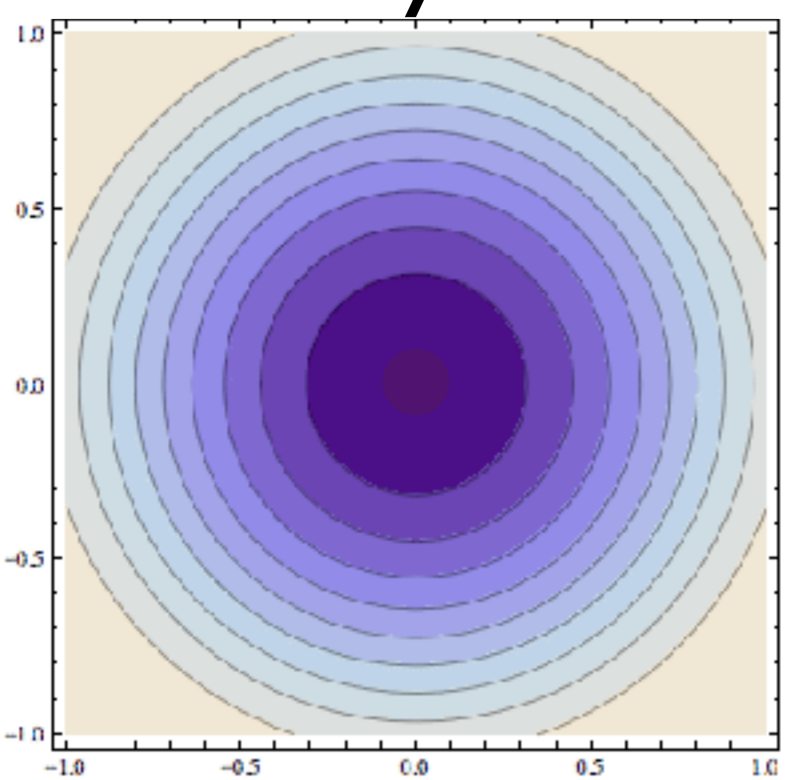


Parameter estimation

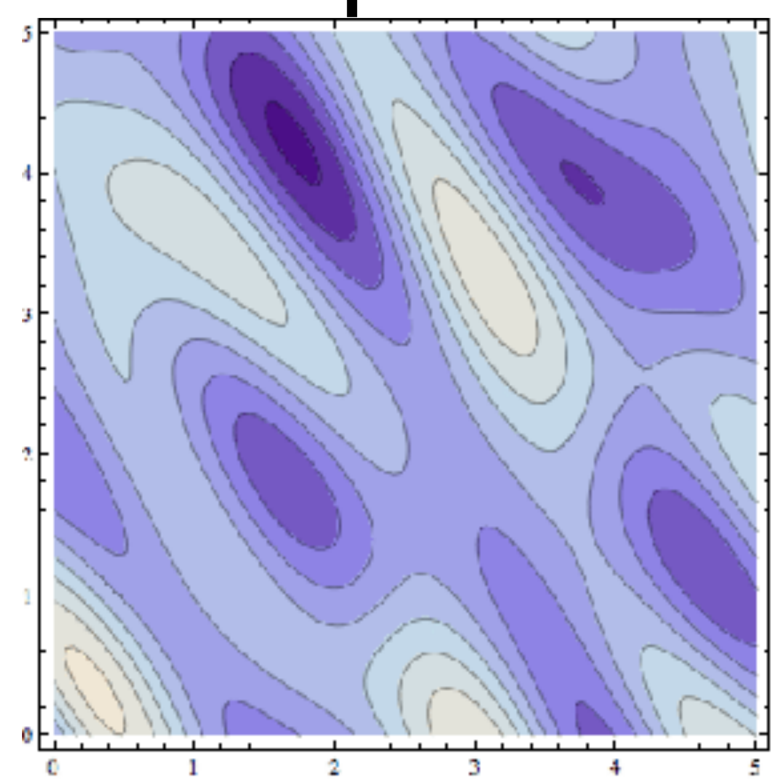
Parameter estimation goals

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

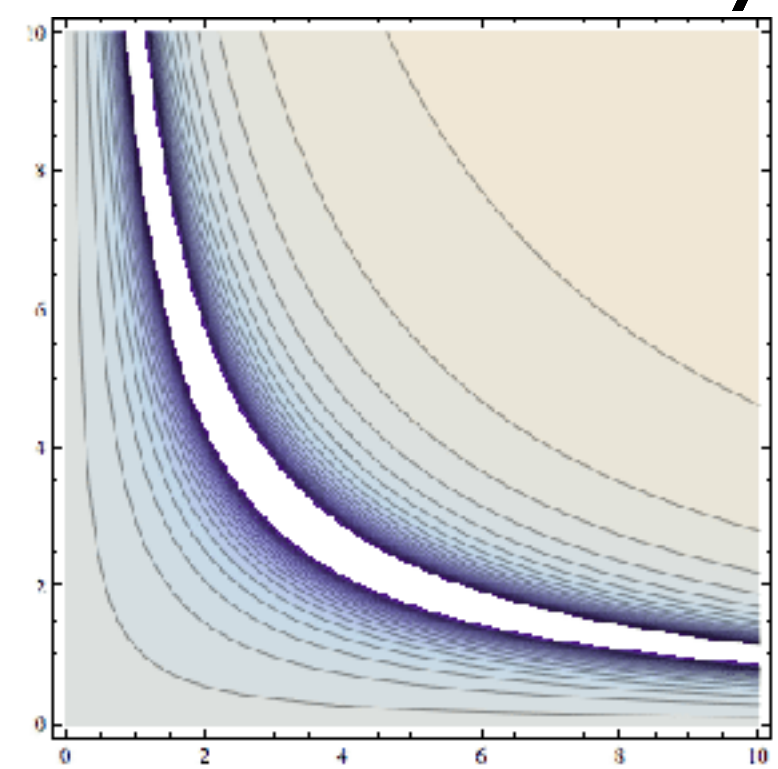
Yay!



Multiple Mins



Unidentifiability

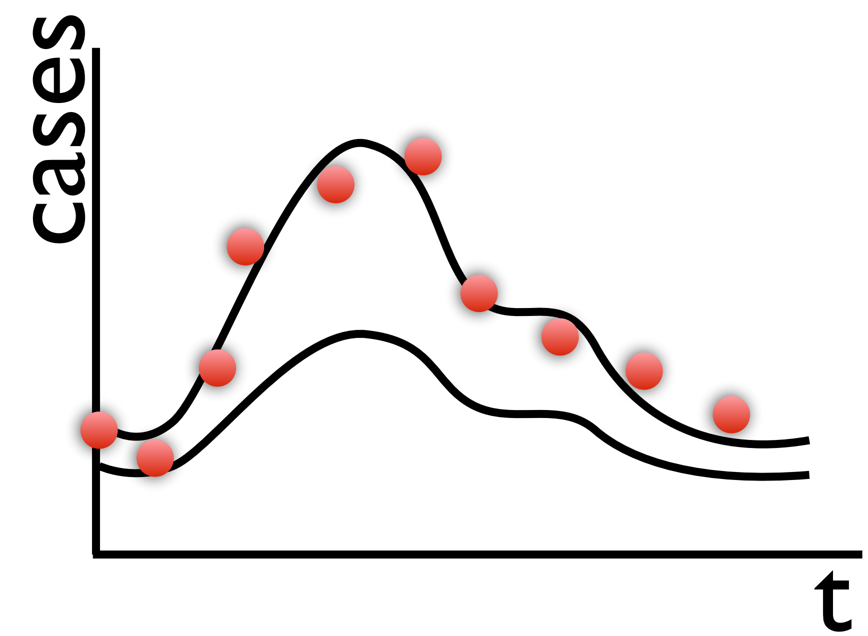


However

- Parameter estimation is one way to connect models with data—not the only one!
- **Just because a model does not precisely fit the data quantitatively, does not mean it cannot bring useful insight!**
- Usefulness of models is not just about prediction or data fitting
- Qualitative patterns are important
- Often may not have every detail of the mechanism, or may not have enough data to characterize fully, but models can be useful to reason and get intuition about the system—often moreso than a model that 'fits' the data better (e.g. think about a mechanistic model vs a spline)

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.

Many things can go wrong!

- Data issues - bias, noise, missing data, not enough data
- Model issues
  - Model misspecification
  - Unidentifiability—particularly for complex models like ABMs, we can expect that many different parameter sets will fit the data equally well

How to frame this statistically?

- **Maximum Likelihood Approach**
- Idea: treat our model as a statistical model, where we suppose we know the general form of the density function (based on the model output) but not the parameter values (**discuss**)
- 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**

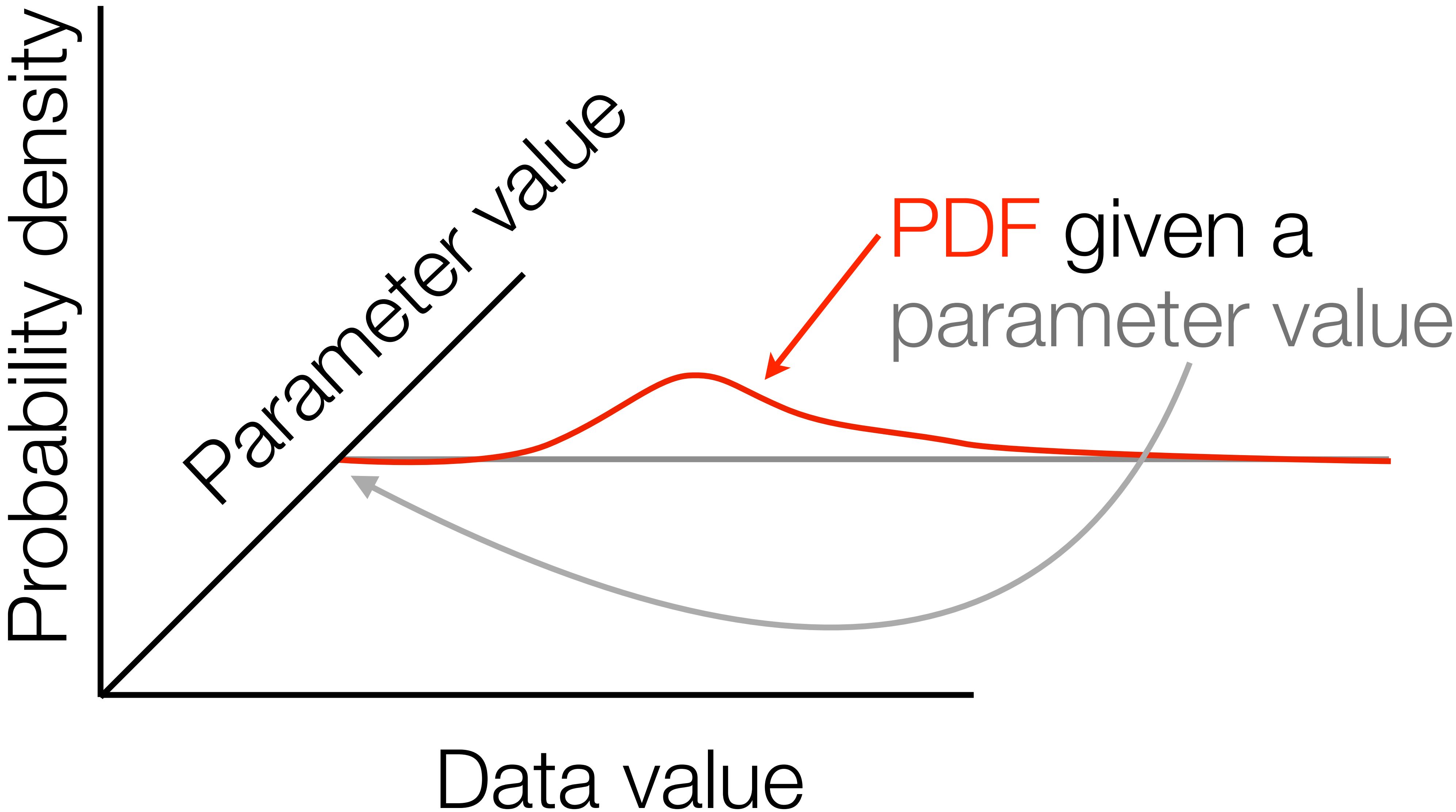
$$P(z \mid p) = f(z, p) = L(p \mid z)$$

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

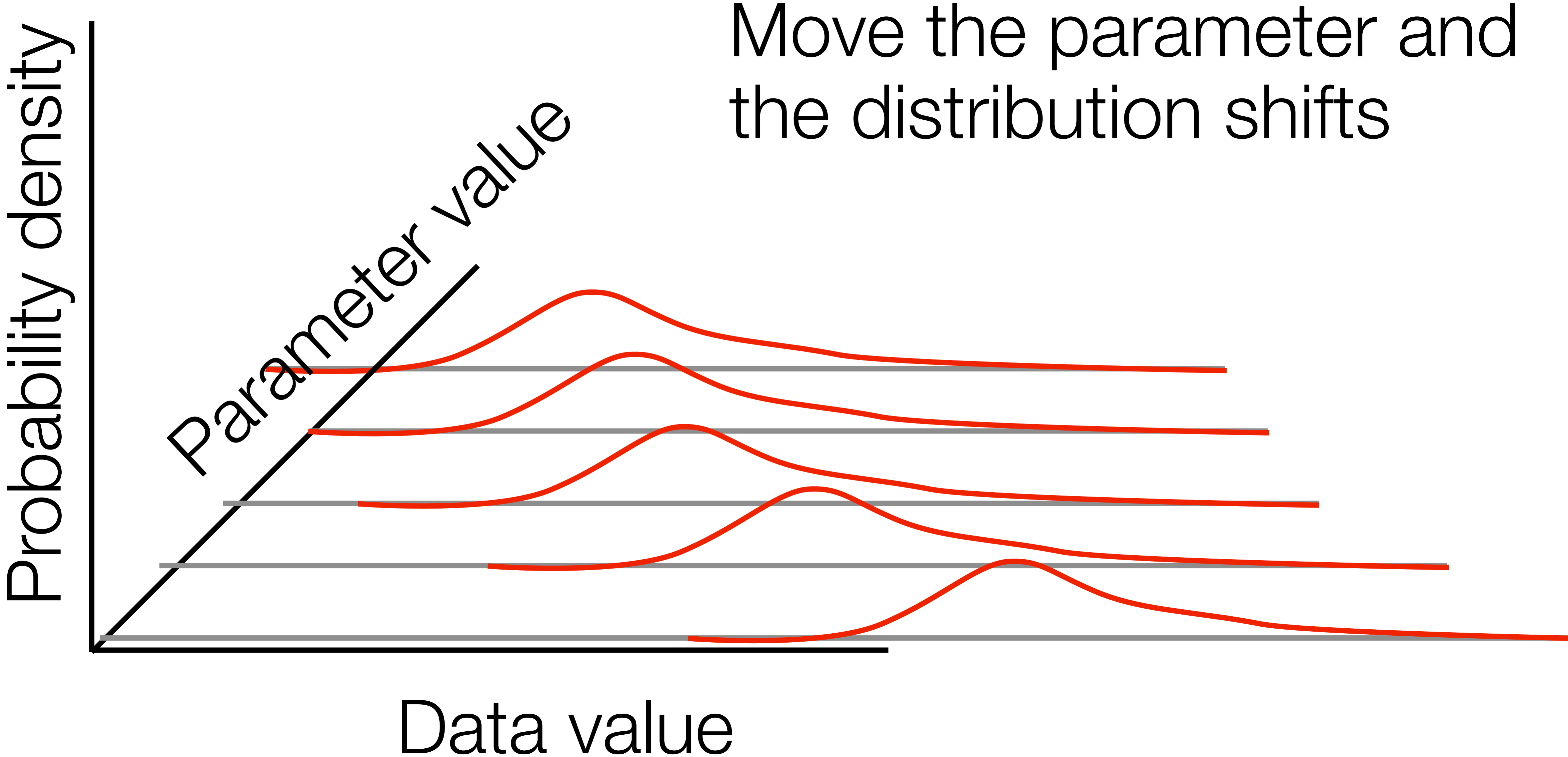
- E.g. $f(z \mid \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma}} \exp\left(-\frac{(z-\mu)^2}{2\sigma^2} \right) = L(\mu, \sigma^2 \mid z)$

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

Likelihood Function

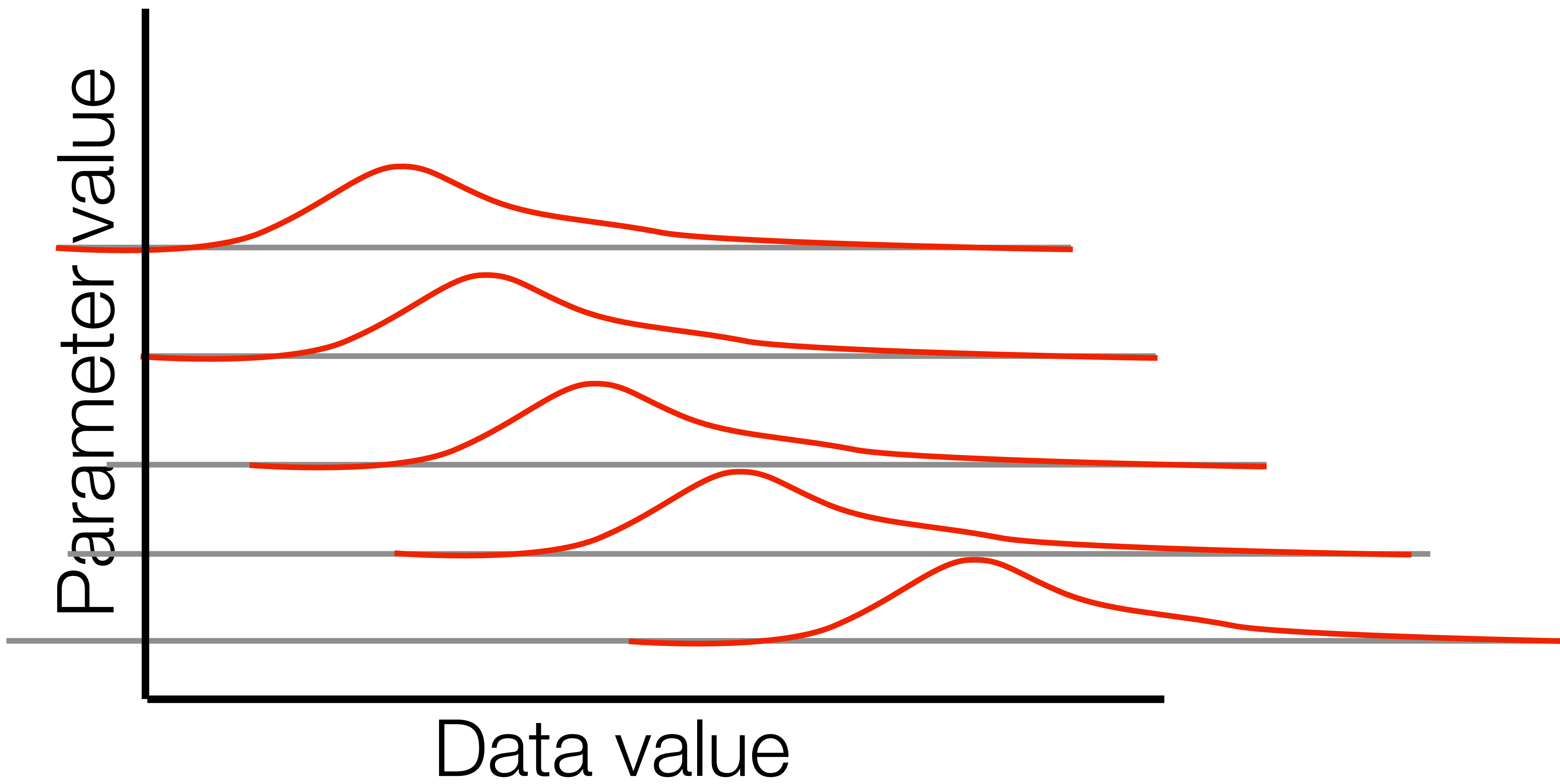


Likelihood Function

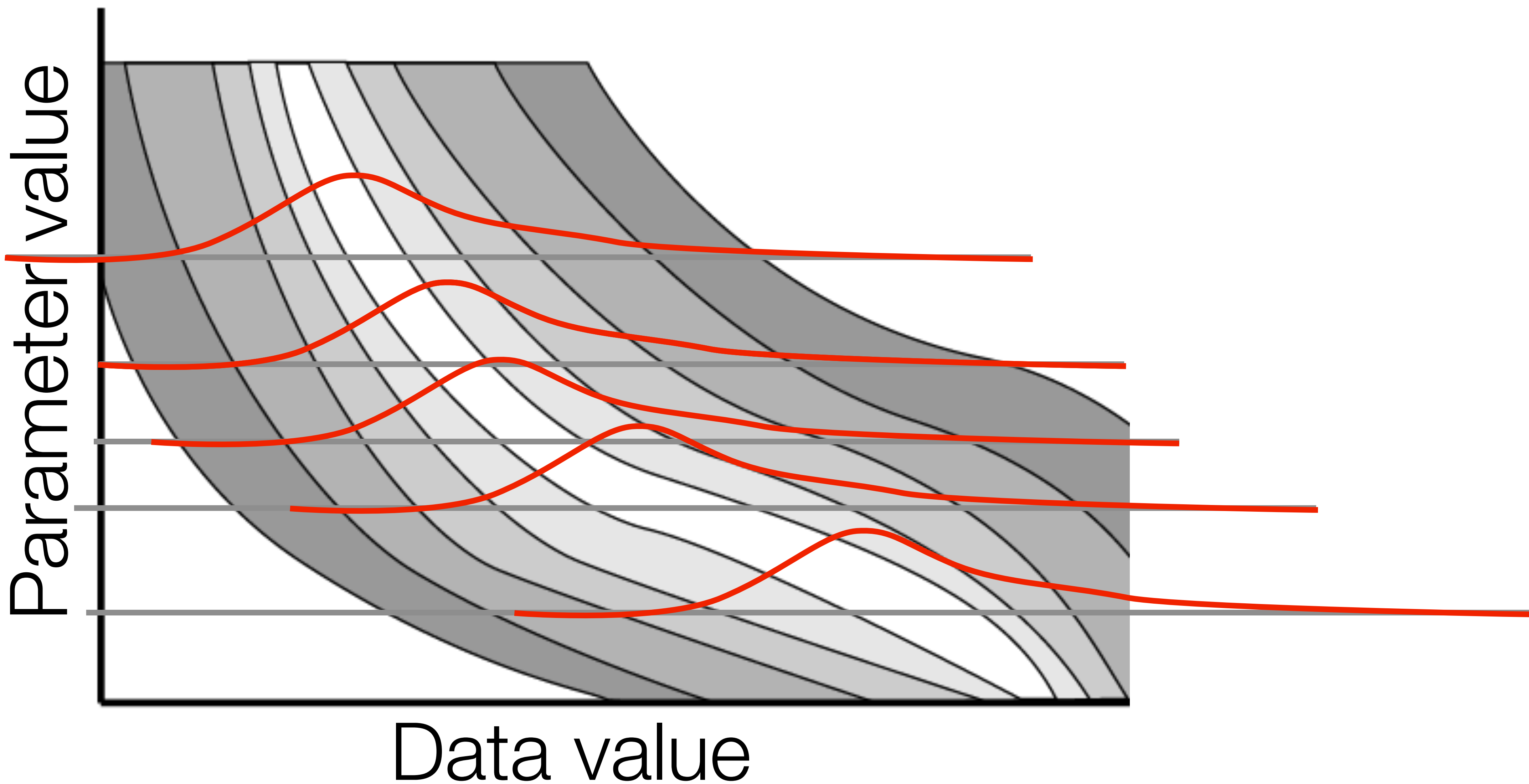


Move the parameter and the distribution shifts

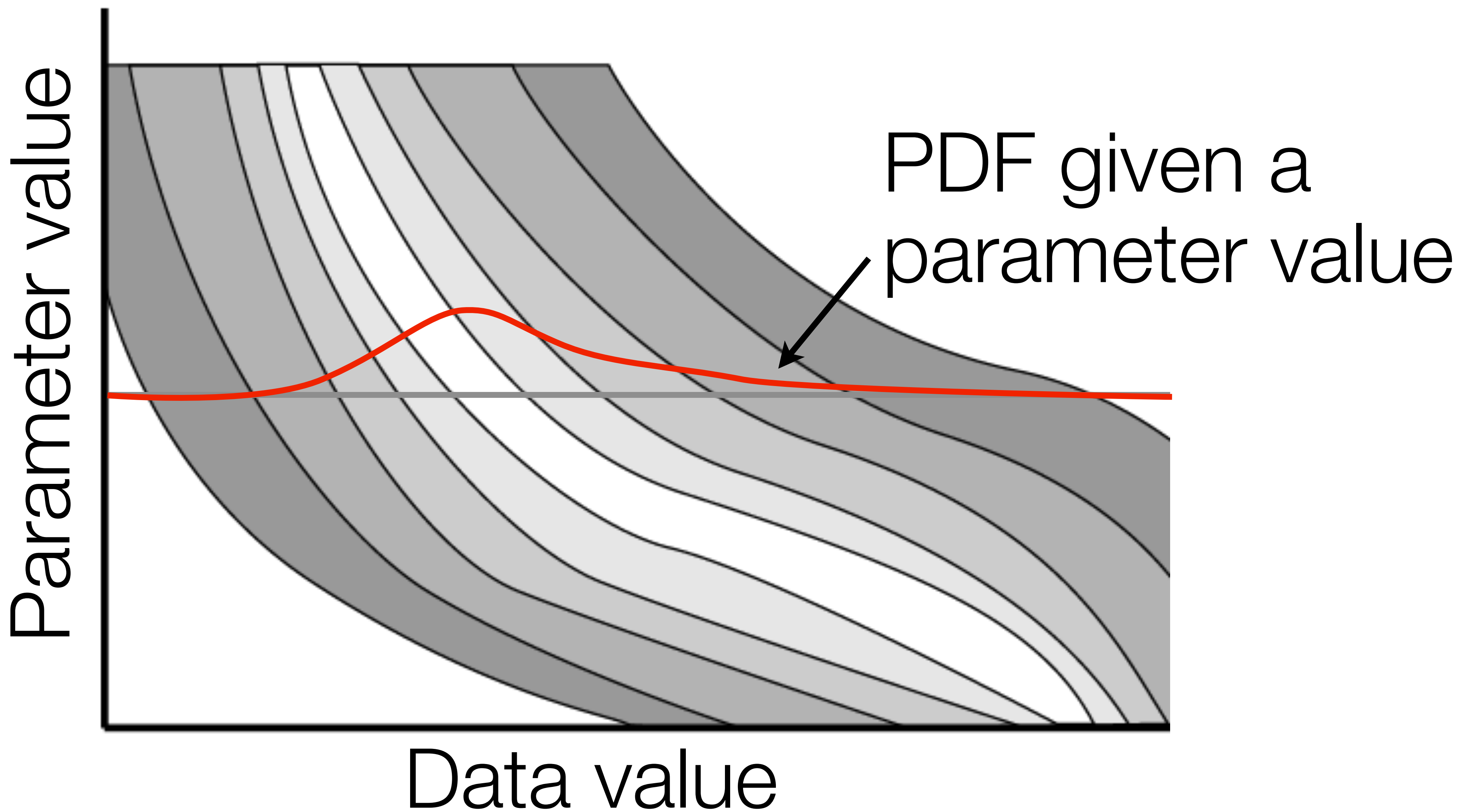
Likelihood Function



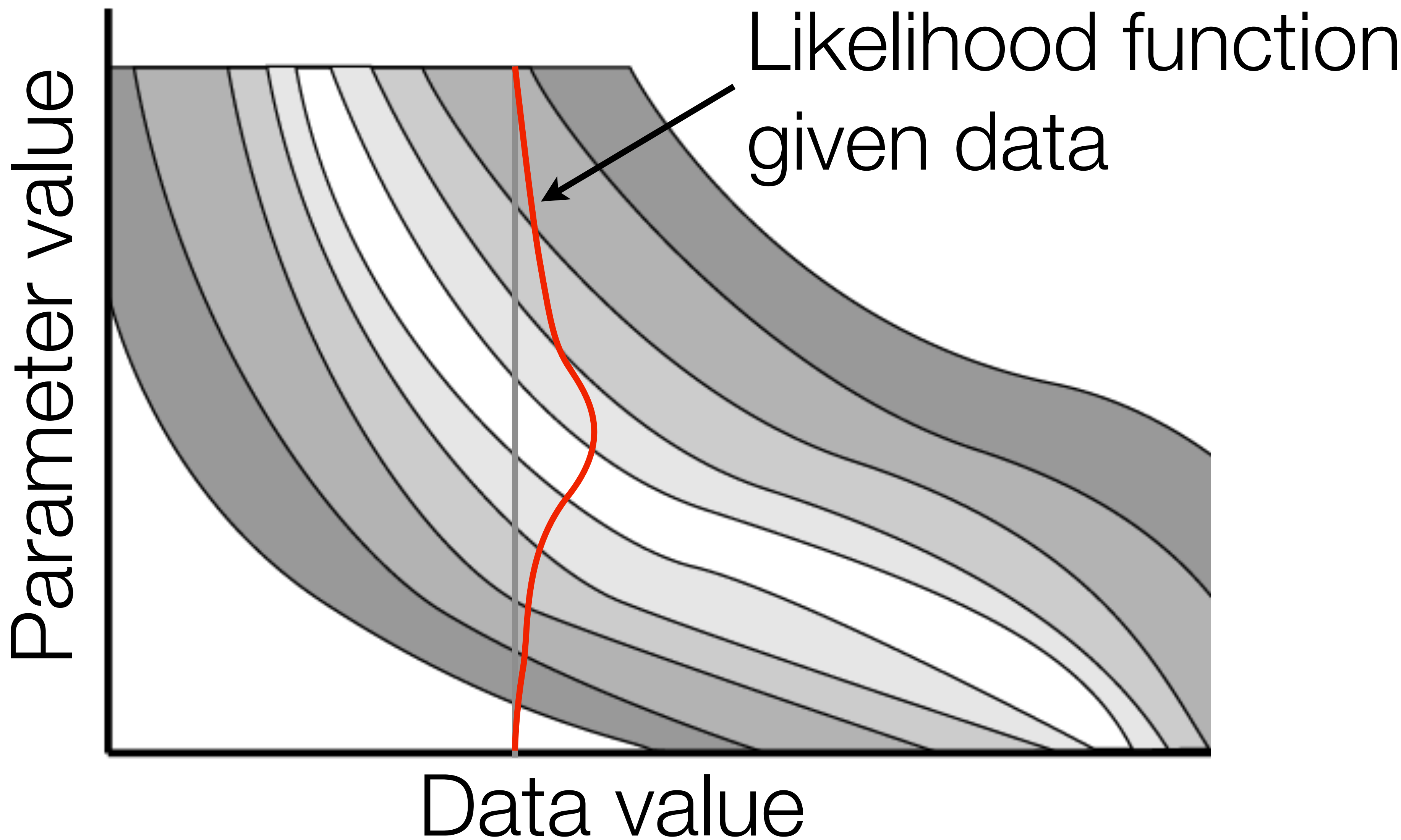
Likelihood Function



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)

Example: deterministic mean field model

- E.g. something like the Erdos-Renyi network SIR mean field model we derived before:

$$s_{t+1} = s_t - b\, s_t i_t$$

$$i_{t+1} = i_t + b\, s_t i_t - pr\, i_t$$

$$r_{t+1} = (1 - s_t - i_t) + pr\, i_t$$

- Or you can think of any other simple deterministic model if you'd prefer (e.g. some simple CA models would also work as the example here)

Example: difference equation model with Gaussian measurement error

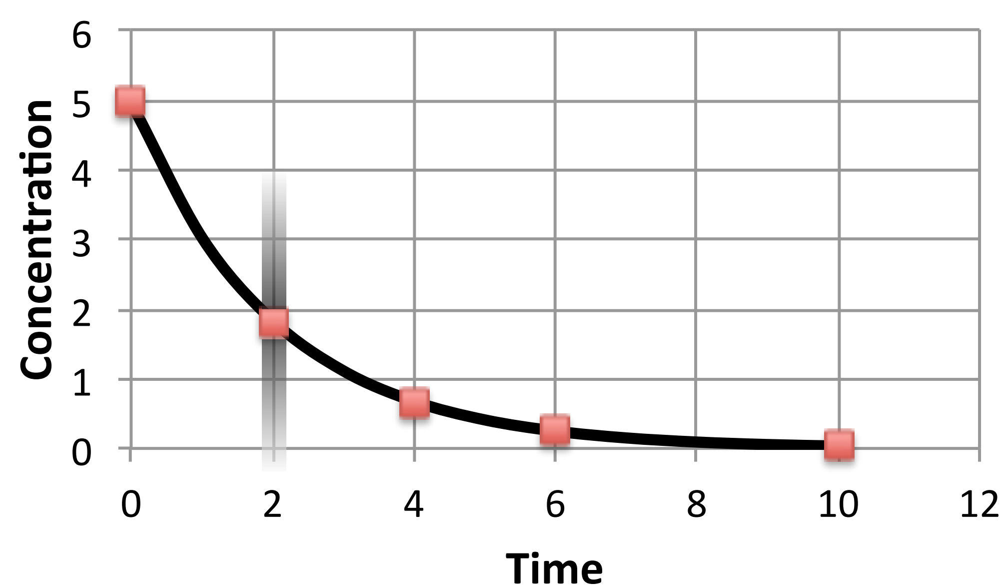
- Model:

$$x(t+1) = f(x,t,p)$$
$$y(t) = g(x,t,p)$$

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

Example: difference equation model with Gaussian measurement 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: difference equation model with Gaussian measurement 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: difference equation model with Gaussian measurement 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: difference equation model with Gaussian measurement 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: difference equation model with Gaussian measurement error

$$
\begin{aligned}
L\left(y\left(t_i, p\right), \sigma^2 \mid z_1, \ldots, z_n\right) &= f\left(z_1, \ldots, z_n \mid y\left(t_i, p\right), \sigma^2\right) \\
&= \prod_{i=1}^{n} f\left(z_i \mid y\left(t_i, p\right), \sigma^2\right) \\
&= \left(\frac{1}{2\pi\sigma^2}\right)^{n/2} \exp\left(- \frac{\sum_{i=1}^{n}\left(z_i - y\left(t_i, p\right)\right)^2}{2\sigma^2} \right)
\end{aligned}
$$

Example: difference equation model with Gaussian measurement 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: difference equation model with Gaussian measurement 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: difference equation model with Gaussian measurement error

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

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

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

Example: difference equation model with Gaussian measurement error

- Similarly for denominator:

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

- This is just least squares!

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

Maximum Likelihood

- 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: $$x(t+1) = f(x,t,p)$$ $$y(t) = g(x,t,p)$$

- Data z_i is assumed to be Poisson with mean $y(t_i)$

- Assume all data points are independent

- Poisson PMF:

$$f\left(z_i \mid y(t_i)\right) = \frac{y(t_i)^{z_i}\, e^{-y(t_i)}}{z_i\,!}$$

Example - Poisson ML

- Likelihood function:

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

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

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

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:

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

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

Maximum likelihood for deterministic models

- Basic approach for deterministic models - suppose only measurement error (otherwise distribution is determined by the model stochasticity and 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 ABMs & other kinds of models

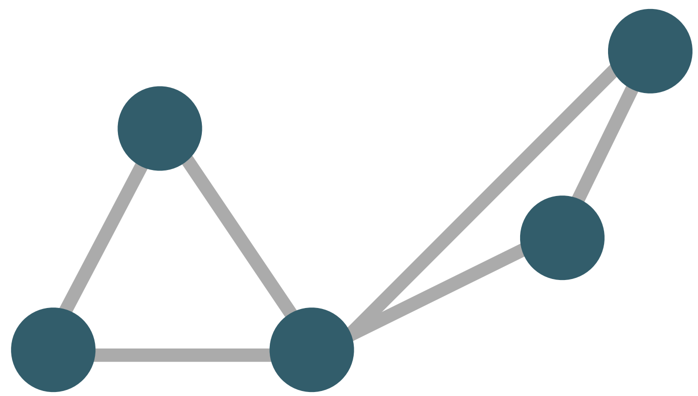
- Can be quite different!
- May require more computation to evaluate (e.g. stochastic models)
- May also be structured quite differently! (e.g. network or individual-based models)

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?



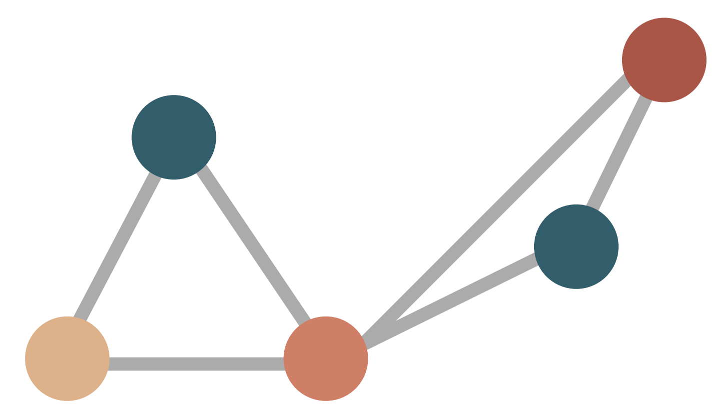
Tiny Network Example

- 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)

Very (very!) brief intro to Bayesian 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 estimation

Prior distribution

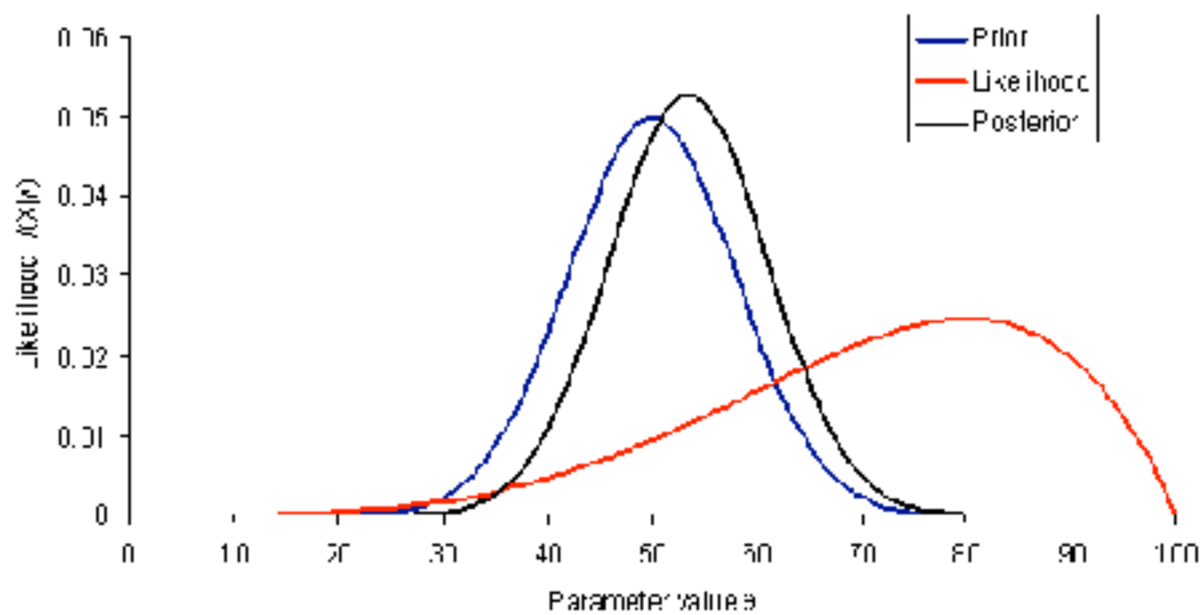
Likelihood

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

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

Sampling-based approaches to parameter estimation

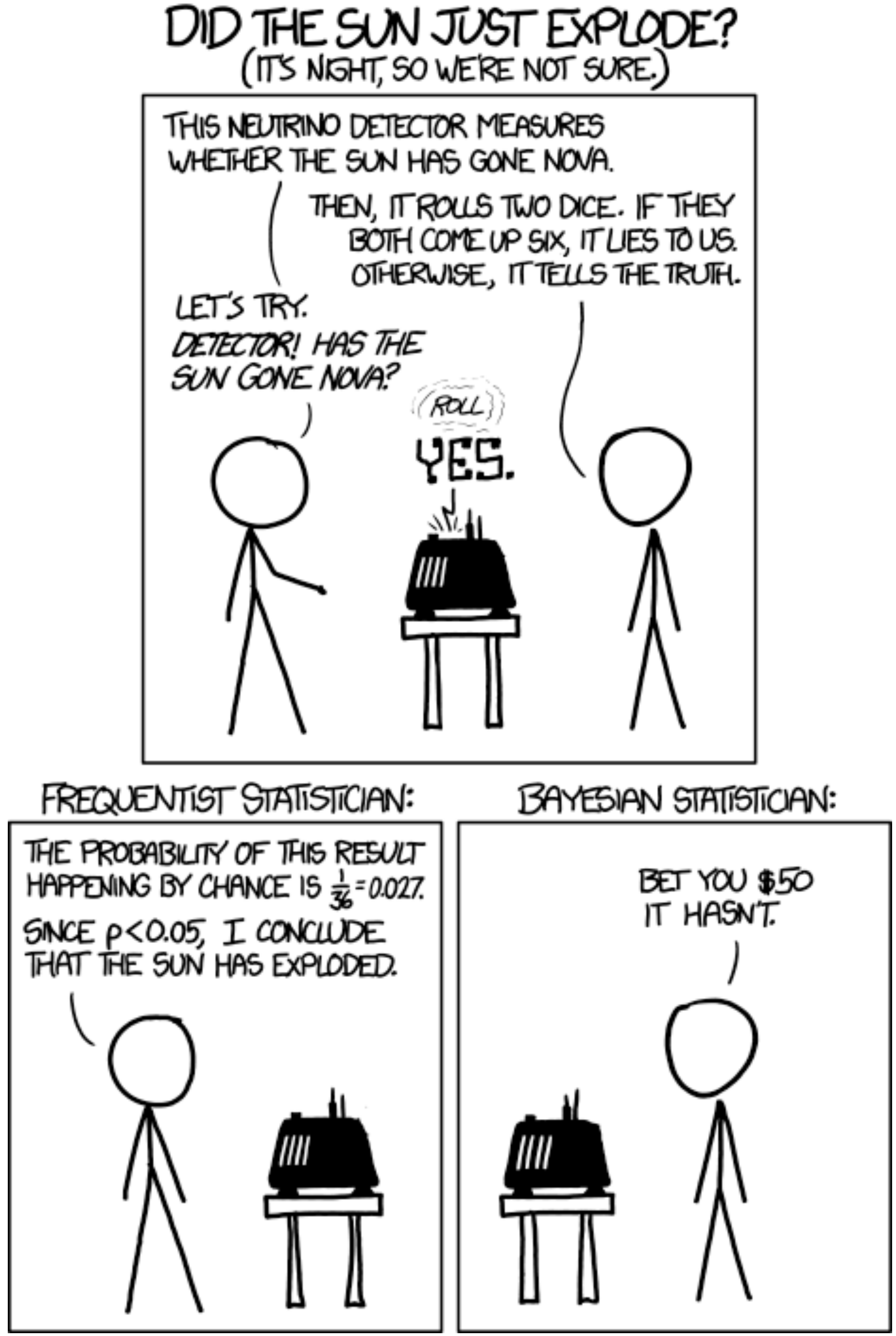
- In our maximum likelihood example, we were able to write down our likelihood explicitly, in terms of equations (e.g. using a normal distribution and the model equations)

- However, for more complex models, or for Bayesian estimation, it's often difficult or impossible to write down an equation for the posterior/likelihood/etc.

Sampling-based approaches to parameter estimation

- Instead—we can use sampling based approaches to sample from the posterior/likelihood—this is often more tractable for ABMs and other complex models

- More on this next time!



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