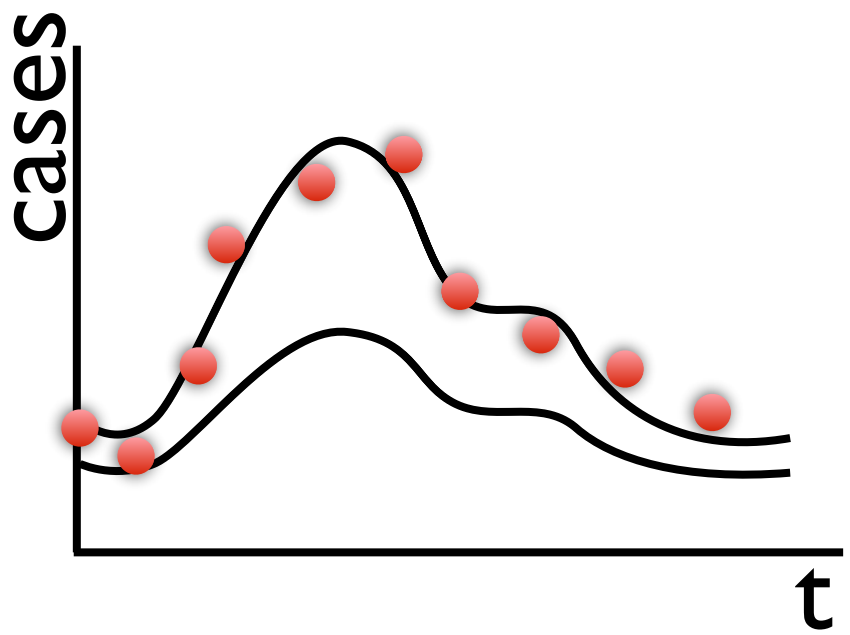
Introduction to Structural & Practical Identifiability

Marisa Eisenberg
University of Michigan, Ann Arbor

But first, a quick intro to parameter estimation

Parameter Estimation

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

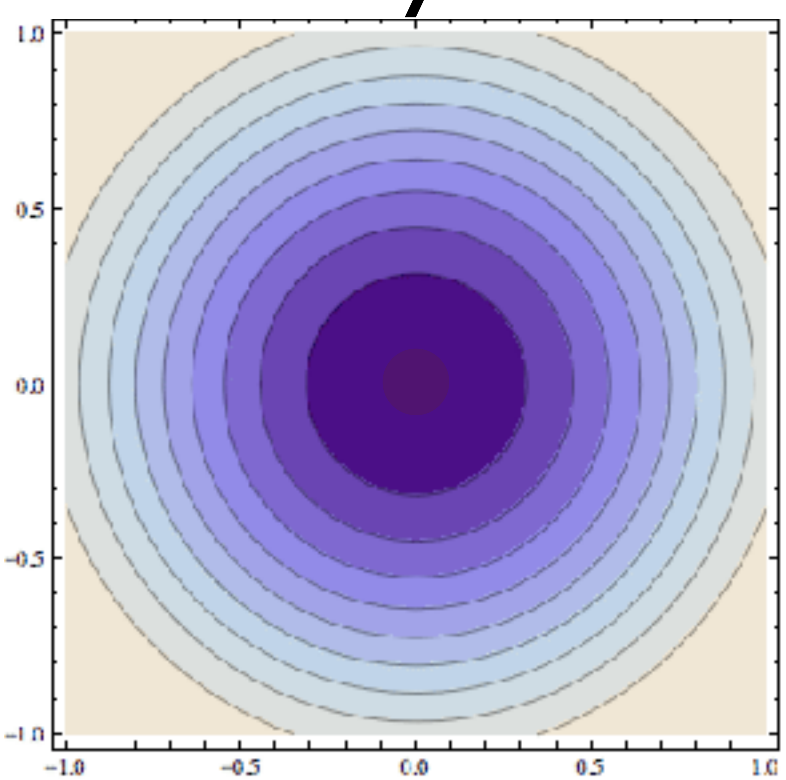


- Frame this from a statistical perspective (inference, regression)
  - Can determine 'most likely' parameters or distribution, confidence intervals, etc.

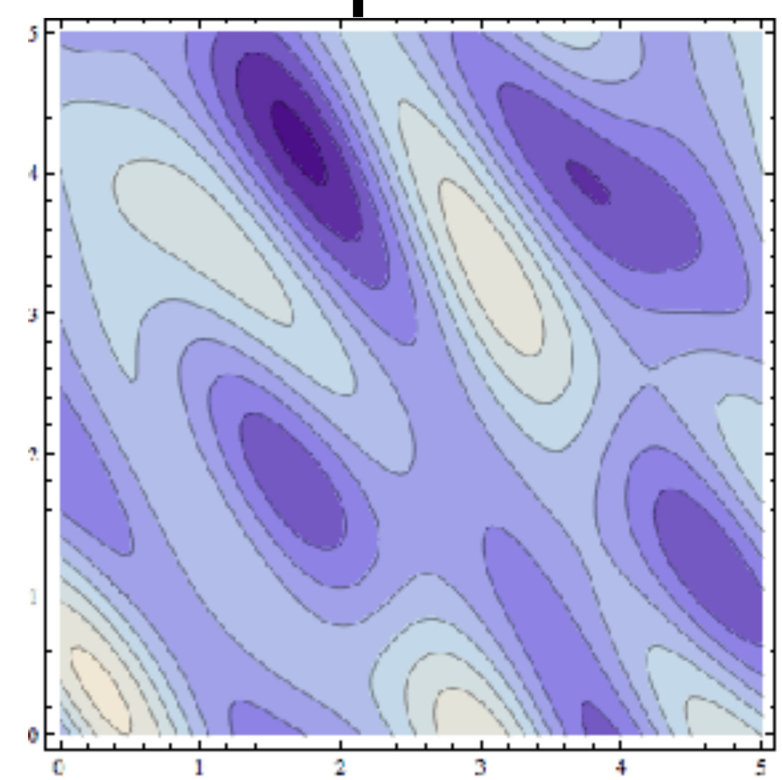
Parameter Estimation

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

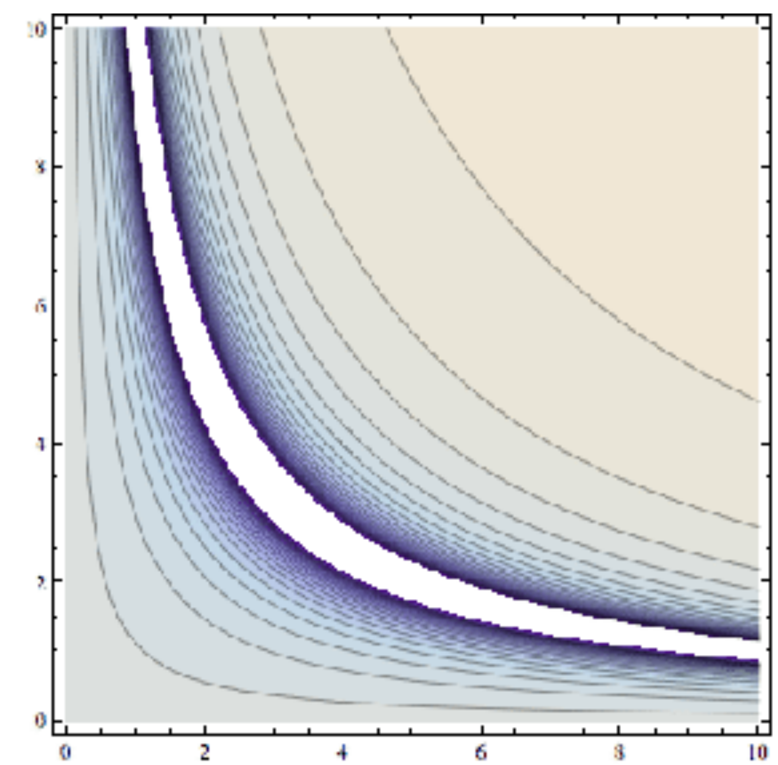
Yay!



Multiple Mins



Struct. UnID



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:

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

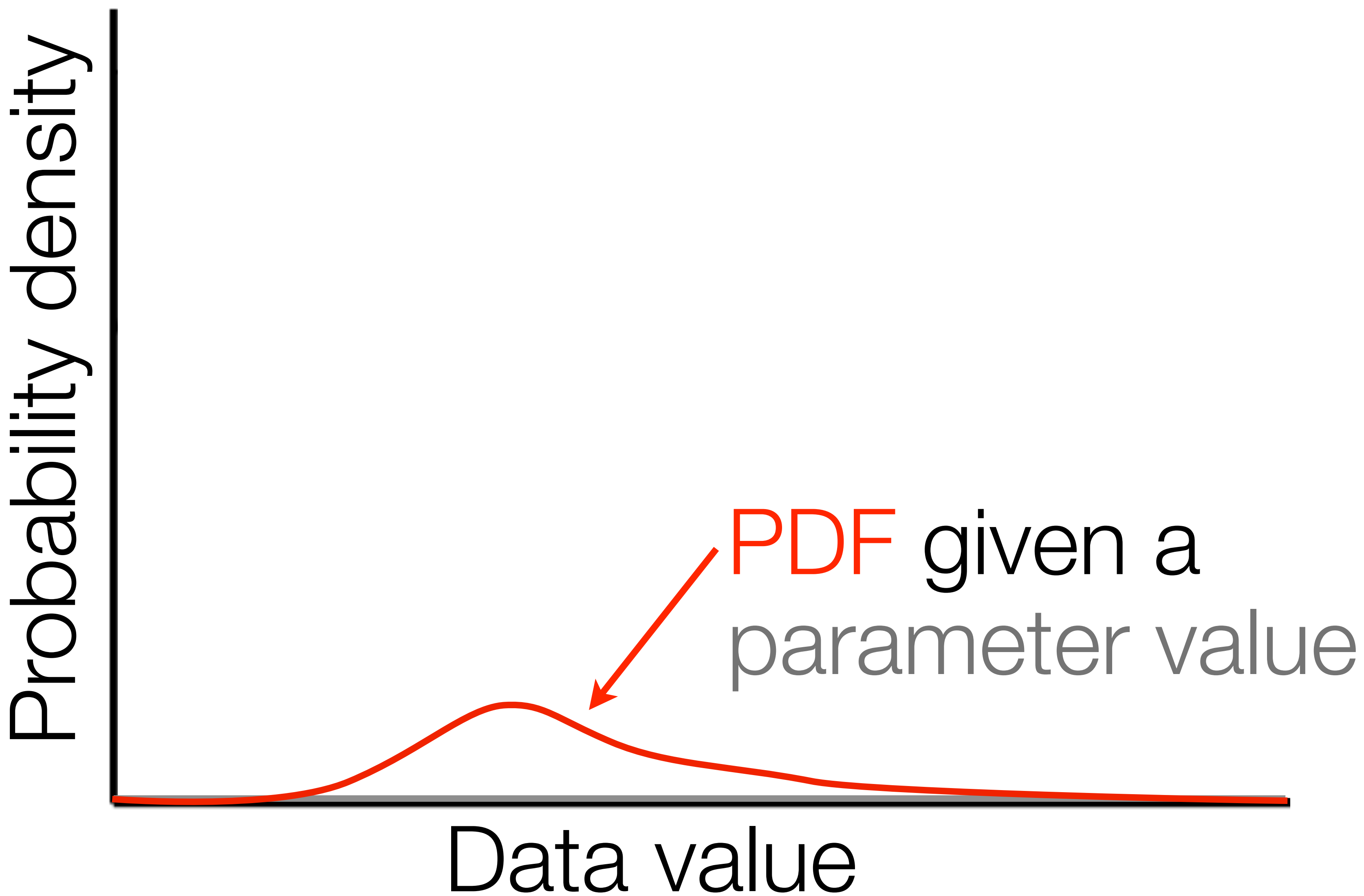
Maximum Likelihood

- **Likelihood Function**

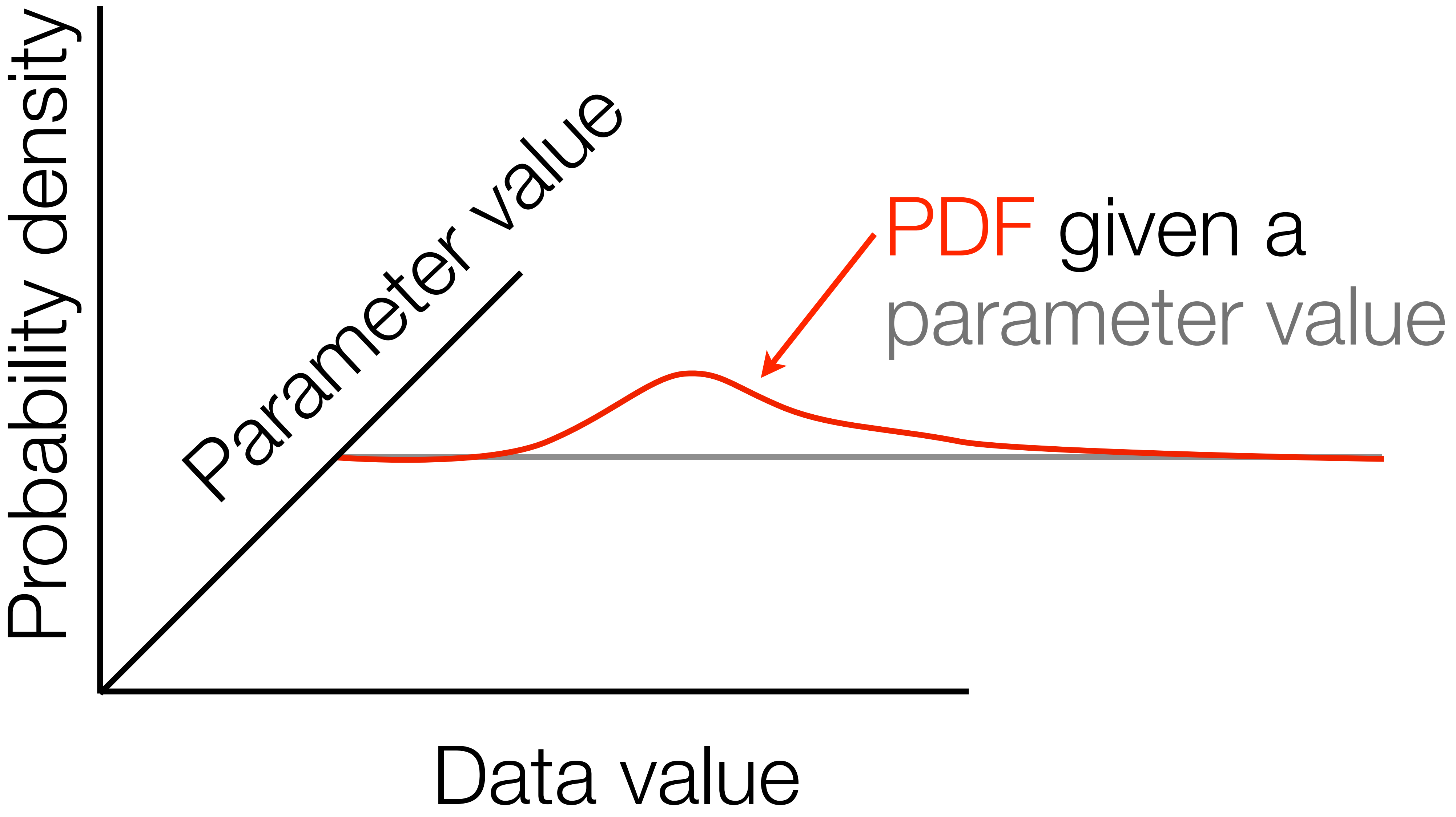
$$P(z \mid p) = f(z, p) = L(p \mid z)$$

- 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

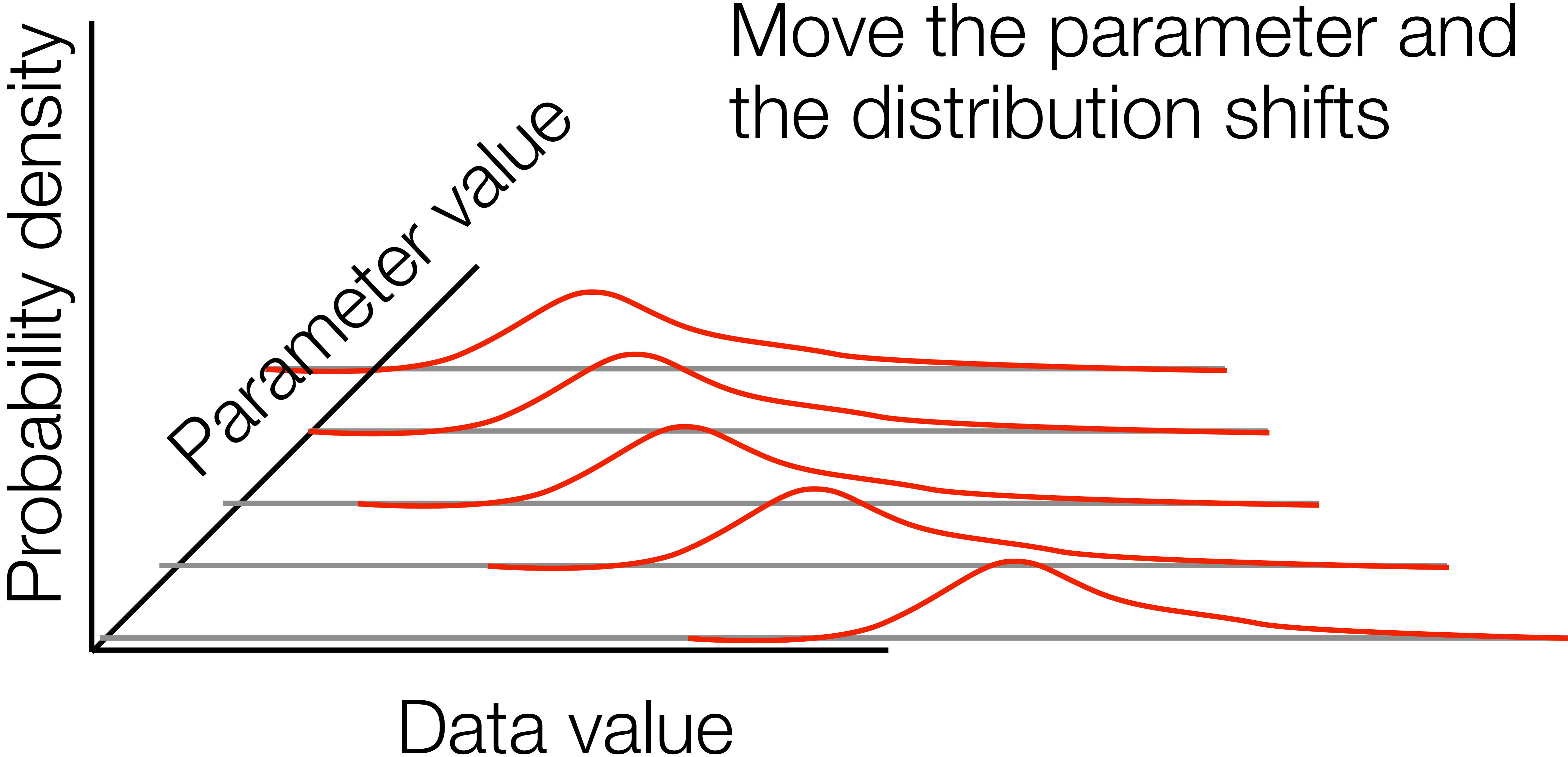


Likelihood Function

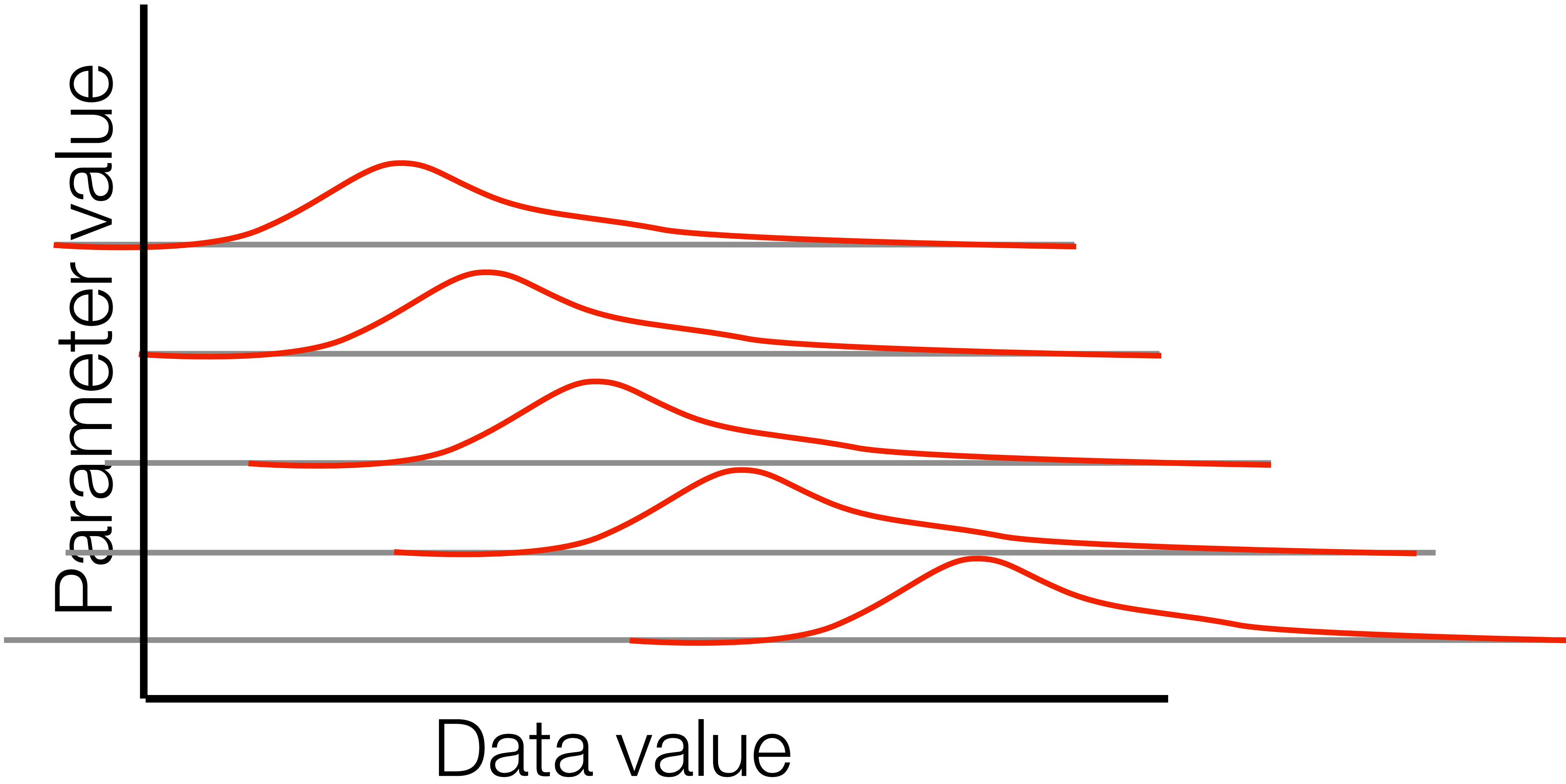


Likelihood Function

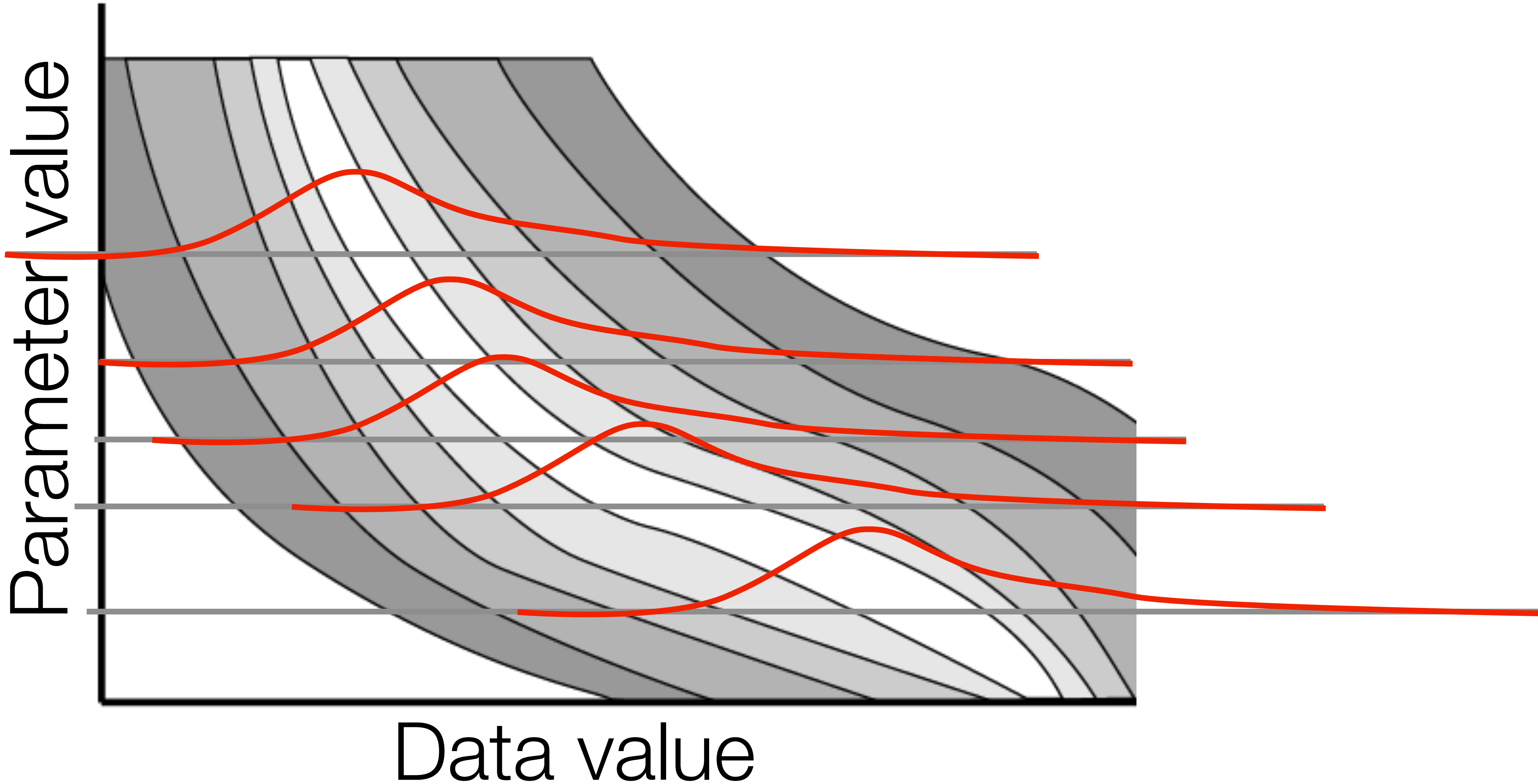
Move the parameter and the distribution shifts



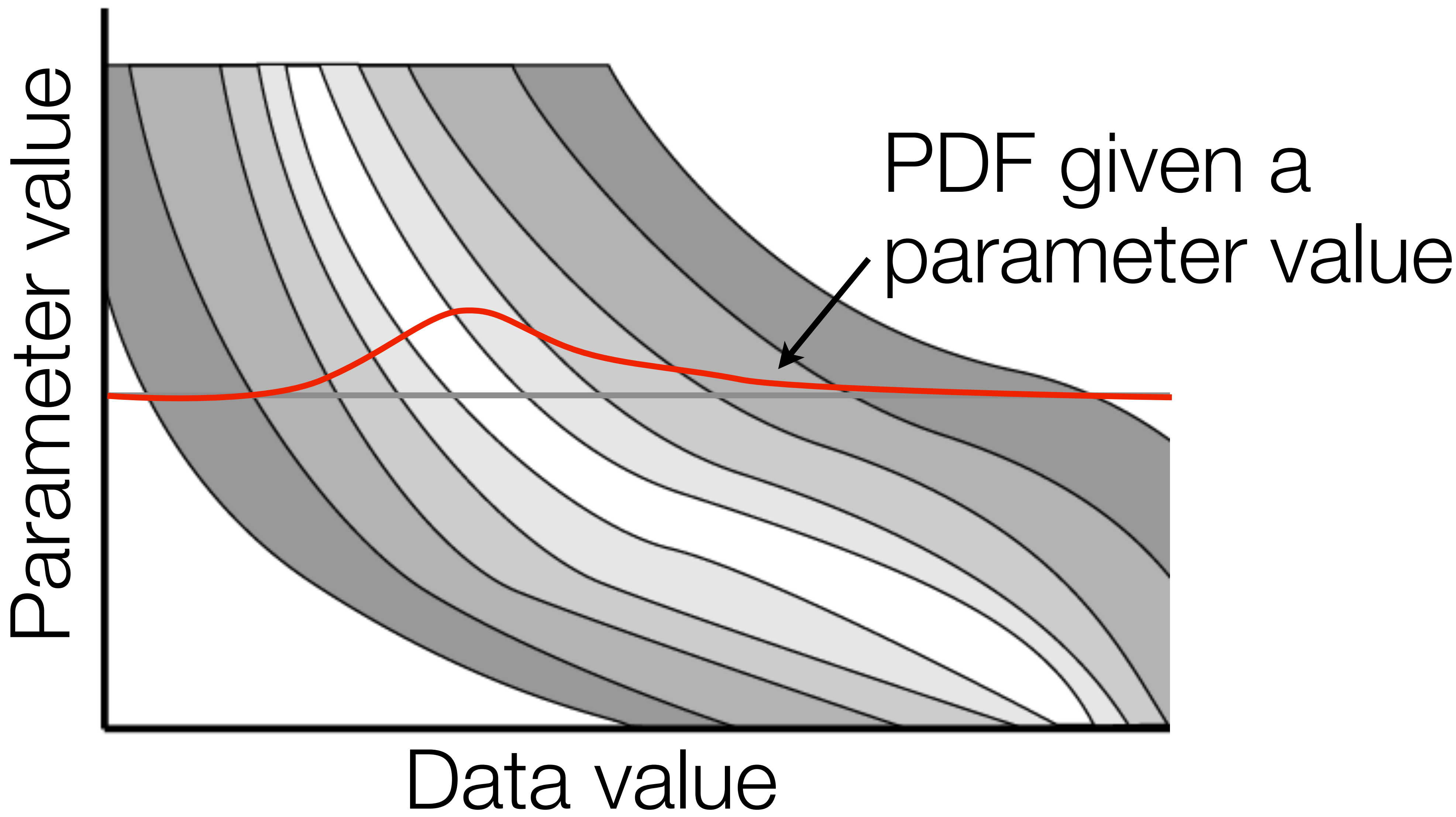
Likelihood Function



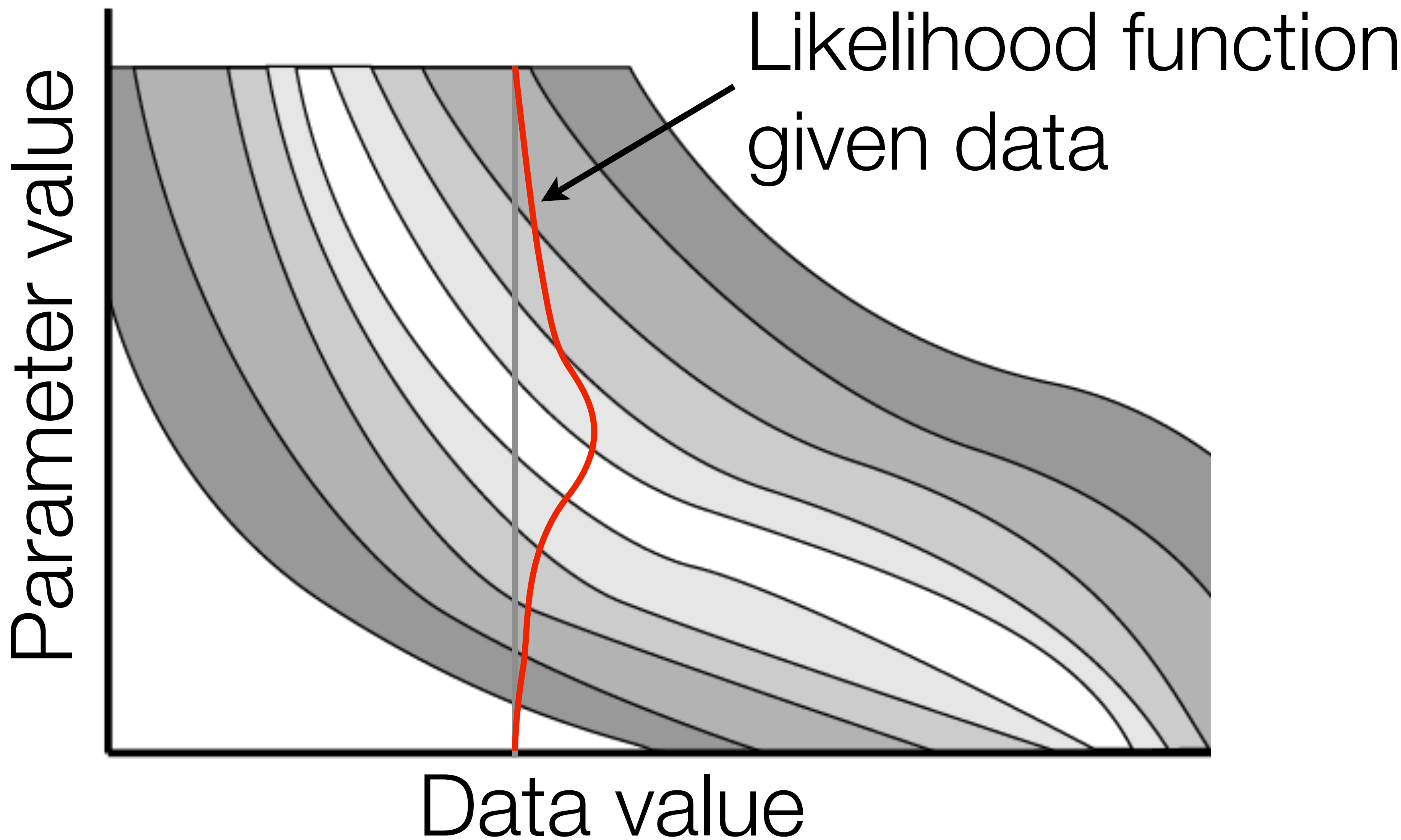
Likelihood Function



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 - ODE Model with Gaussian Error

- Model:

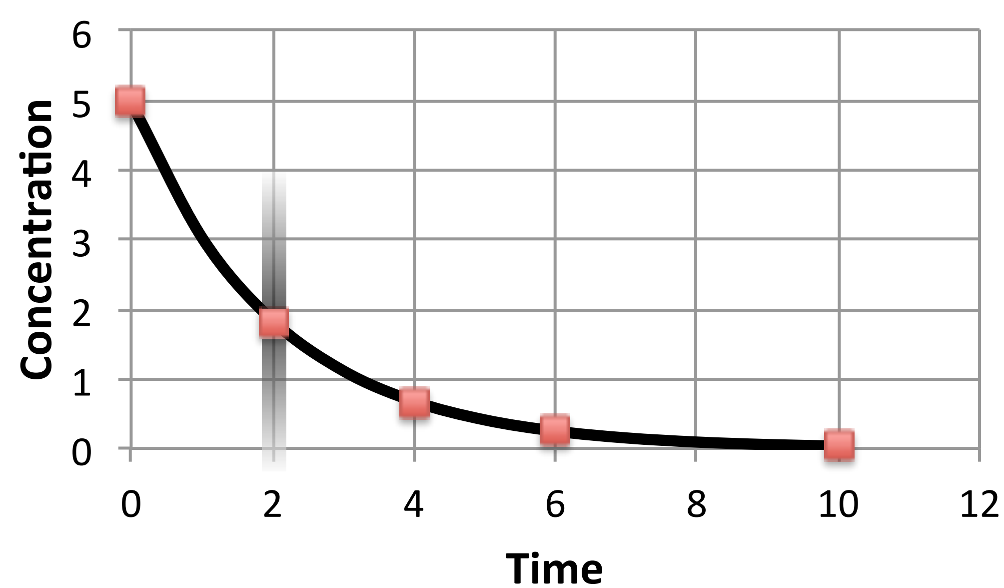
$$\dot{x} = f\left(x, t, p\right)$$

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

- 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

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

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

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

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 Summary for ODEs

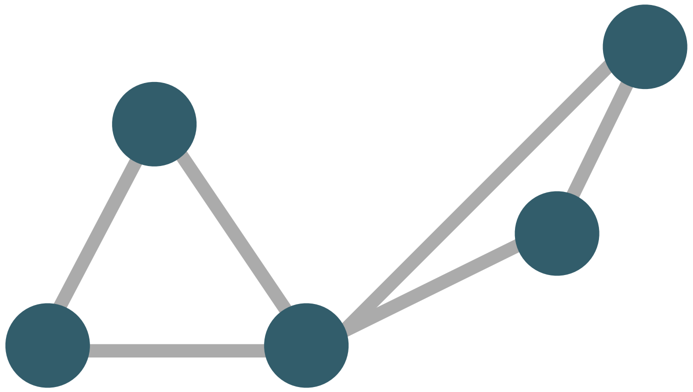
- Basic approach - suppose only measurement error
- Data is given by distribution where model output is the mean
- Suppose each time point of data is independent
- Use PDF/PMF to calculate the likelihood
- Take the negative log likelihood, minimize this over the parameter space

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)

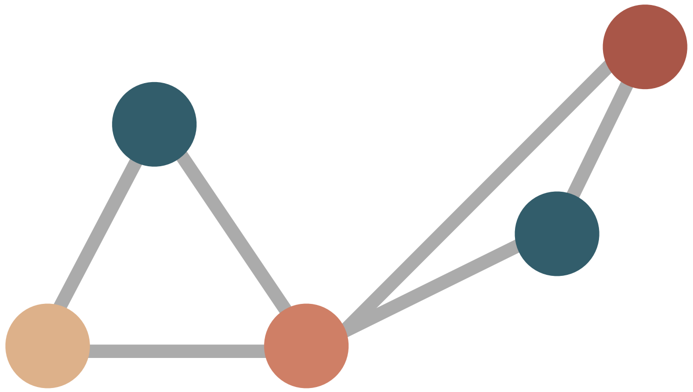
Tiny Network Example

- 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?
- Can calculate contributions from each node (for example)



- L(p,data) might be something like: P(susc nodes did not get sick) x P(infected nodes did get sick)

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:

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

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:

Likelihood

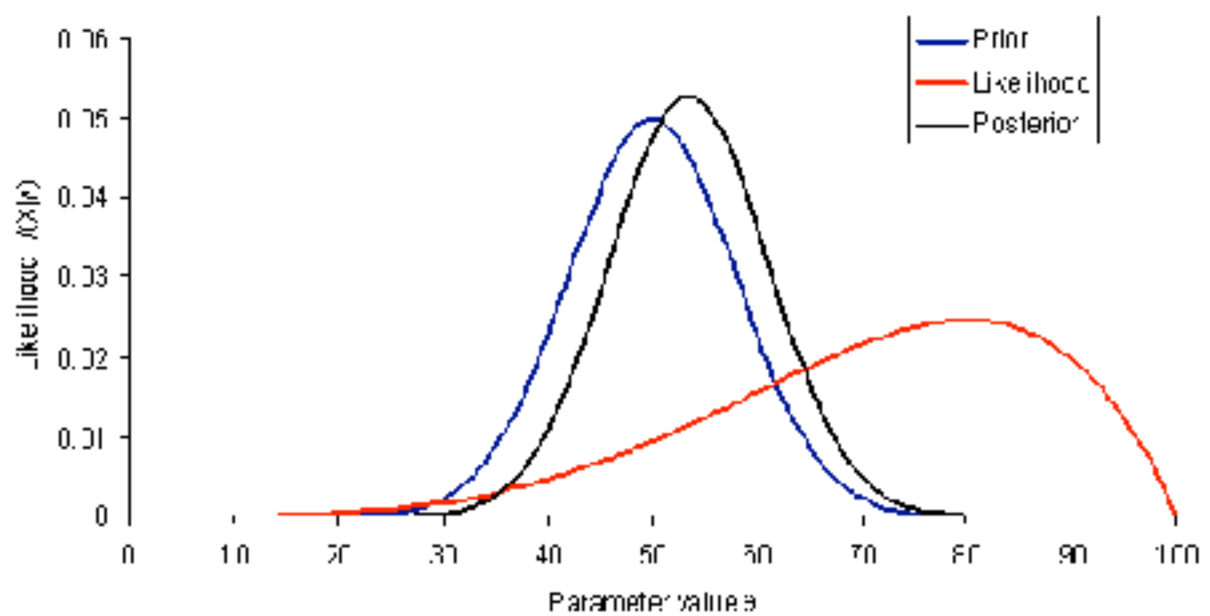
Prior distribution

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

Normalizing constant
(can be difficult to calculate!)

Bayesian Parameter Estimation

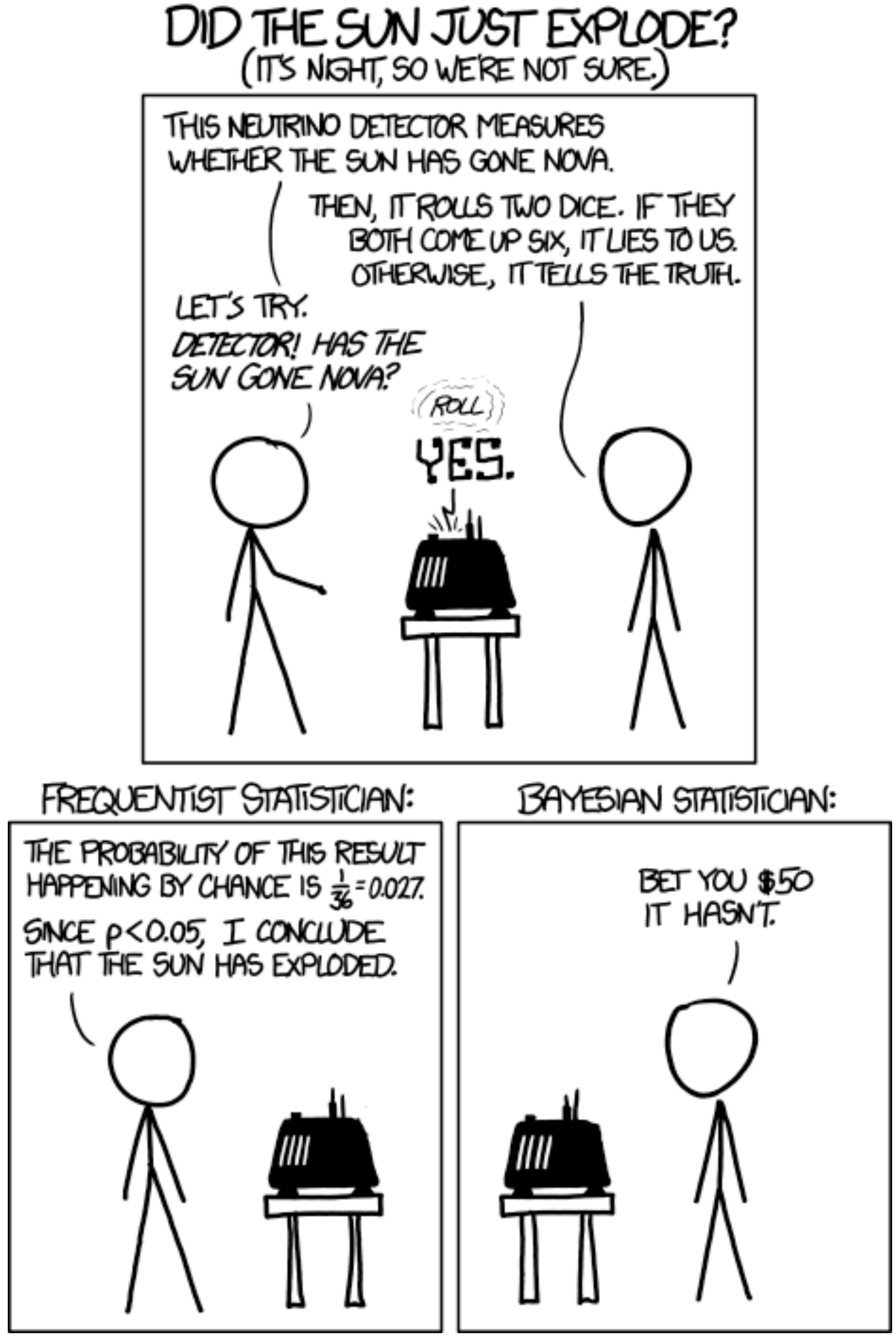
- From prior distribution & likelihood distribution, determine the posterior distribution of the parameter



- 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



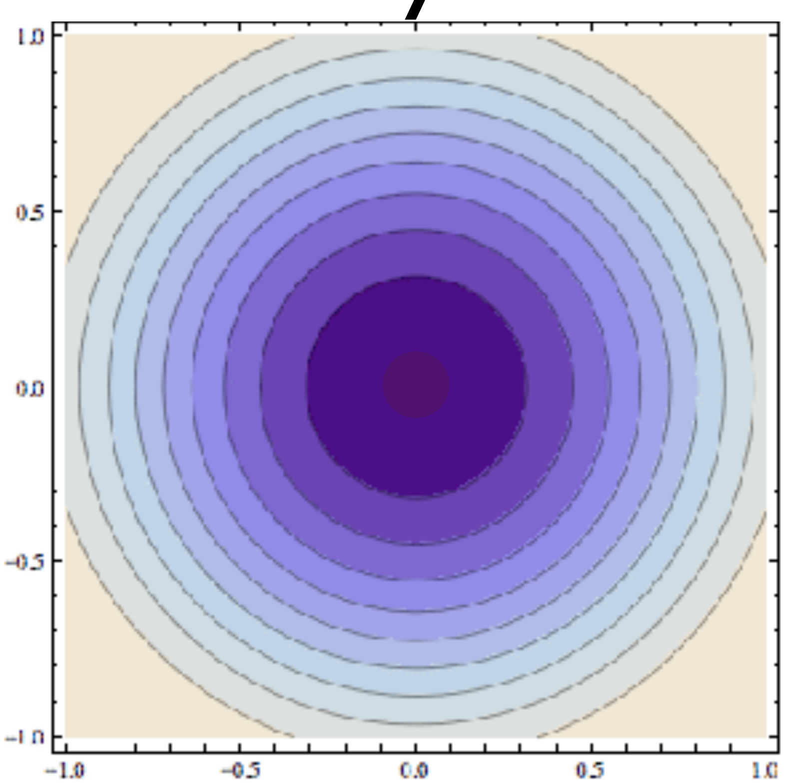
from XKCD:
http://xkcd.com/1132/

Identifiability

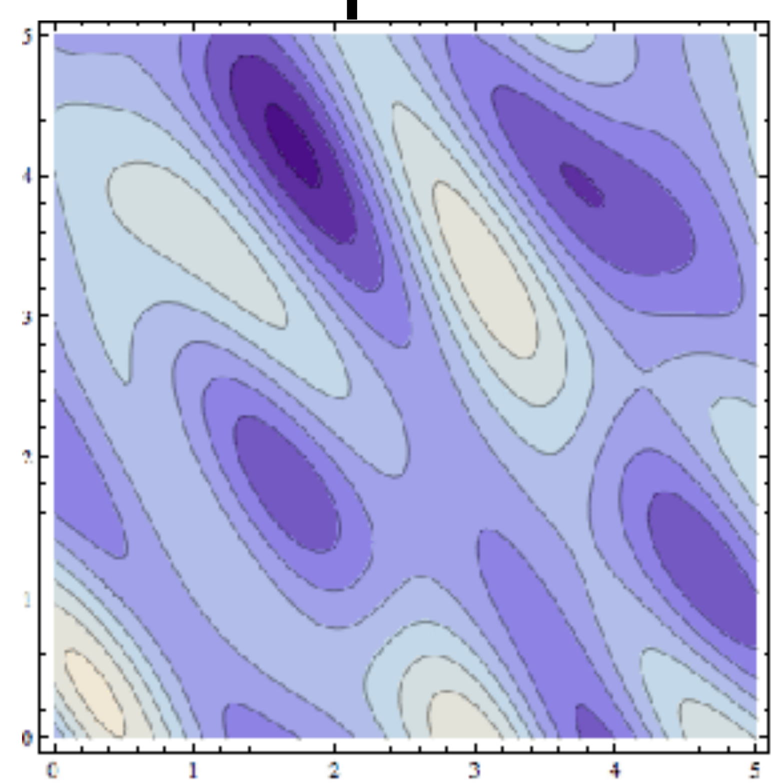
Parameter Estimation

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

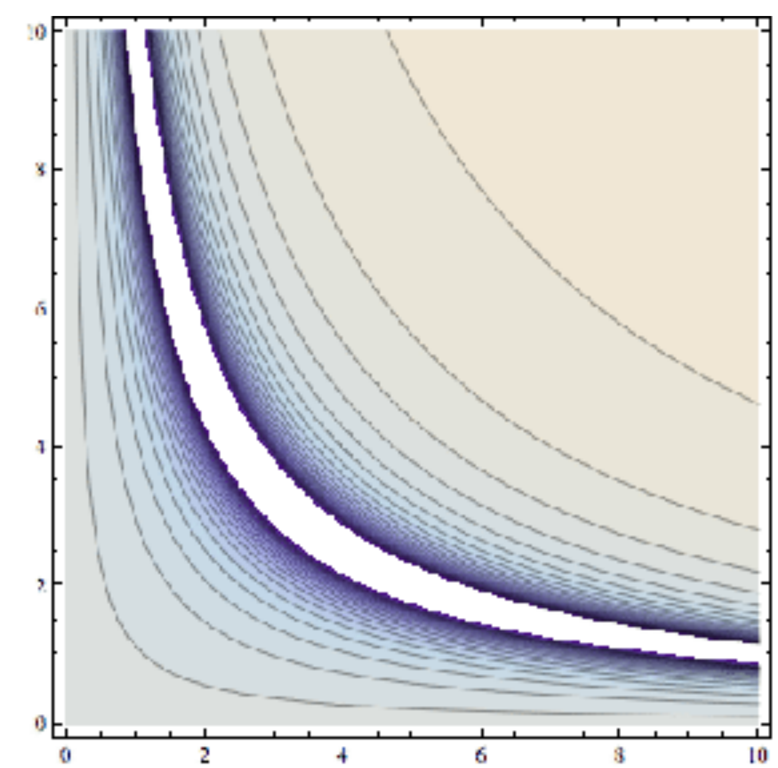
Yay!



Multiple Mins



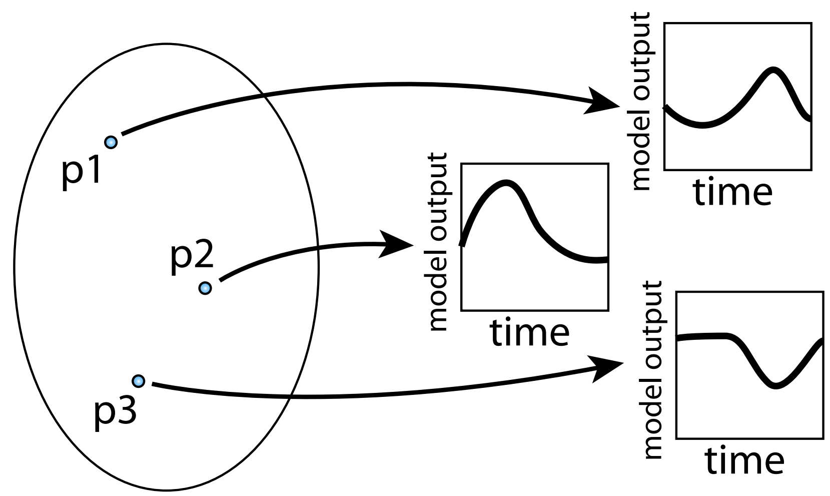
Struct. UnID



Identifiability

- Identifiability—Is it possible to uniquely determine the parameters from the data?

Parameters



Output/Data

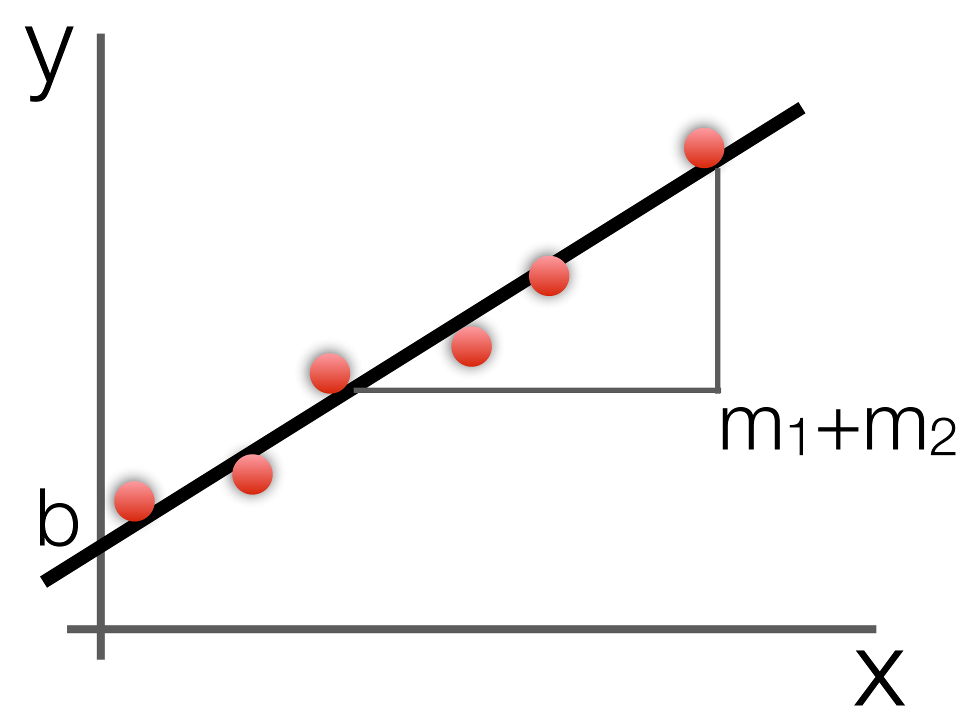
- Important problem in parameter estimation

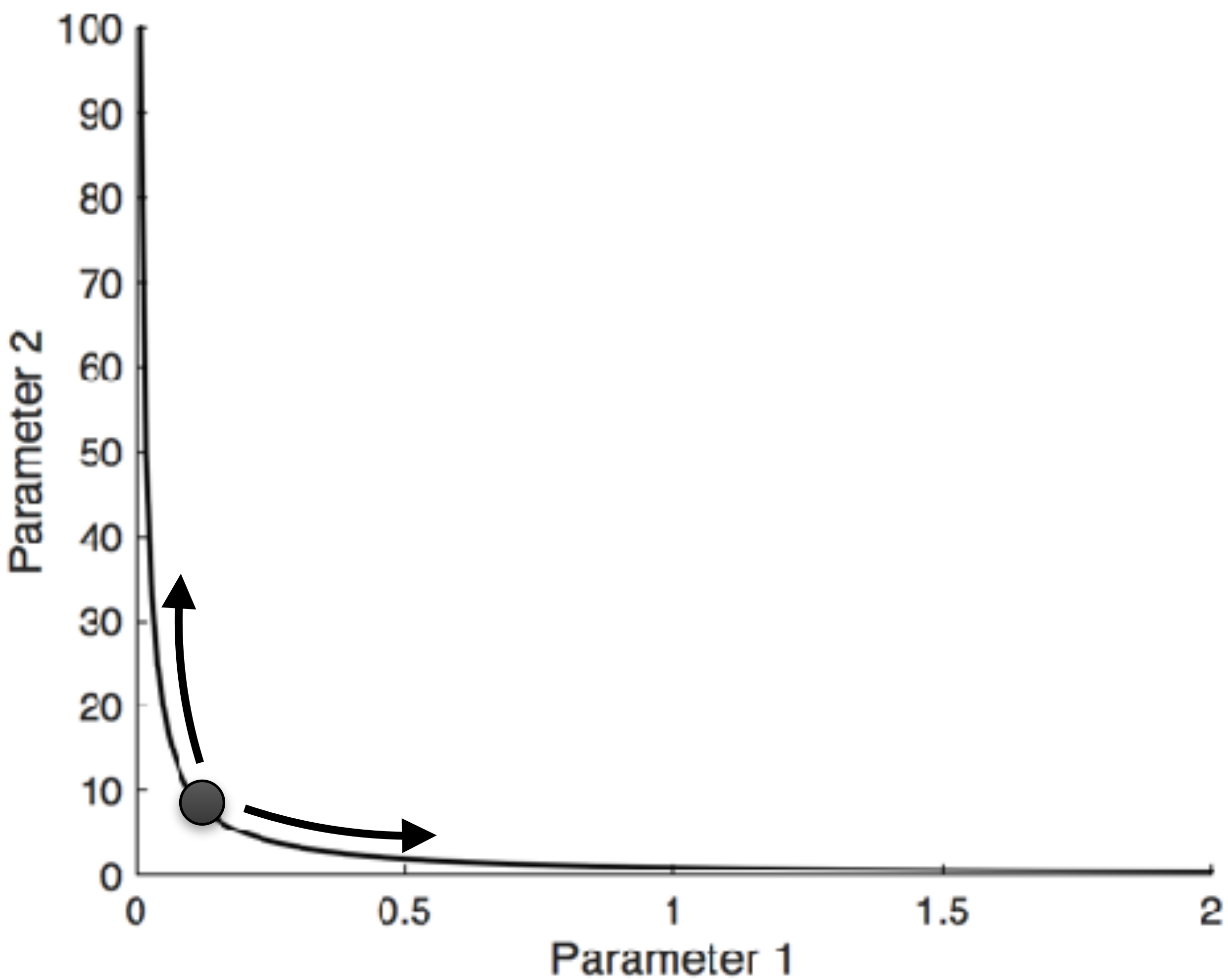
- Many different approaches - statistics, applied math, engineering/systems theory

Ollivier 1990, Ljung & Glad 1994, Evans & Chappell 2000, Audoly et al 2003, Hengl et al. 2007, Chis et al 2011

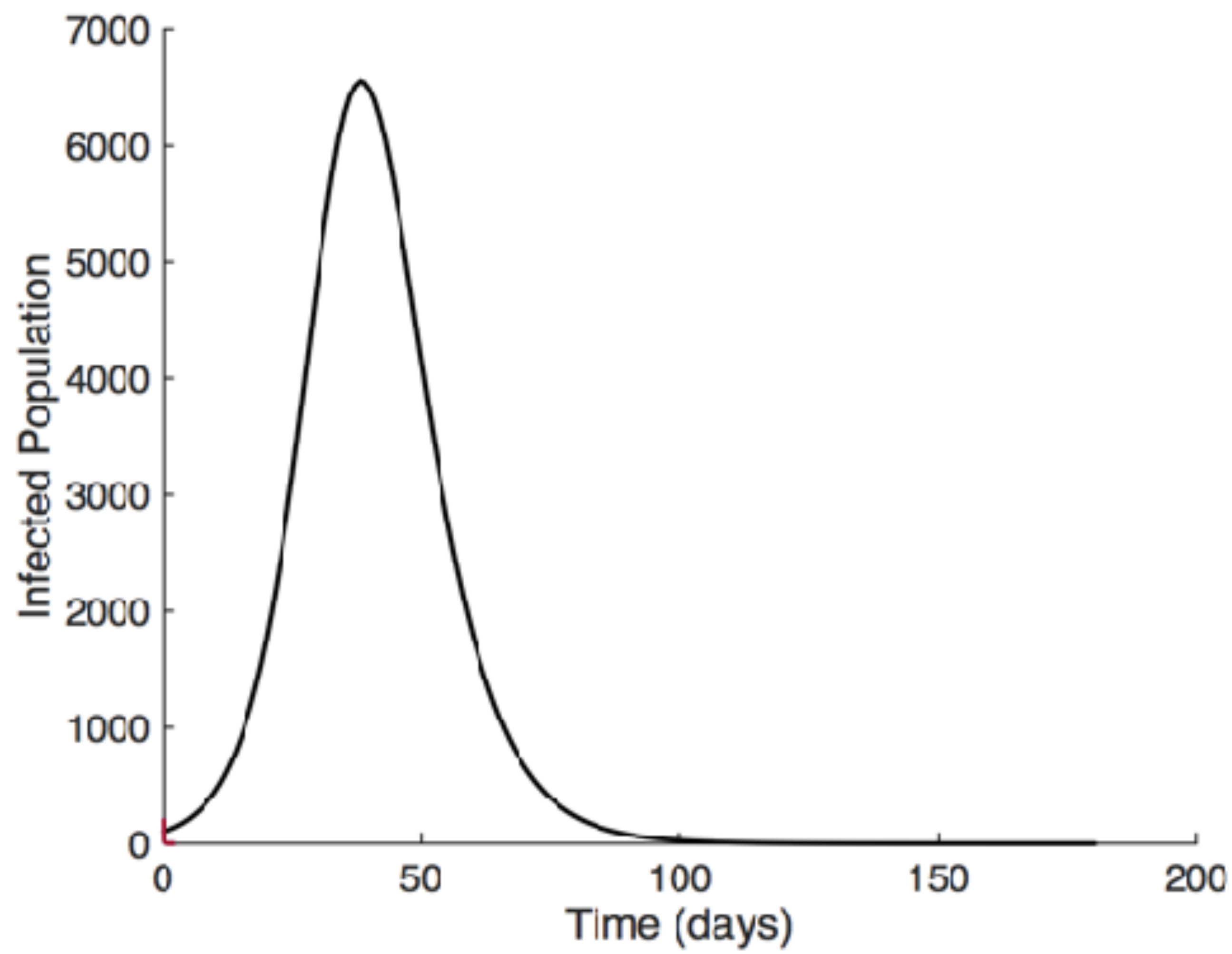
Identifiability

- Practical vs. Structural
    - Broad, sometimes overlapping categories
    - Noisy vs. perfect data
- Example: $y = (m_1 + m_2)x + b$
- Unidentifiability - can cause serious problems when estimating parameters
- Identifiable combinations

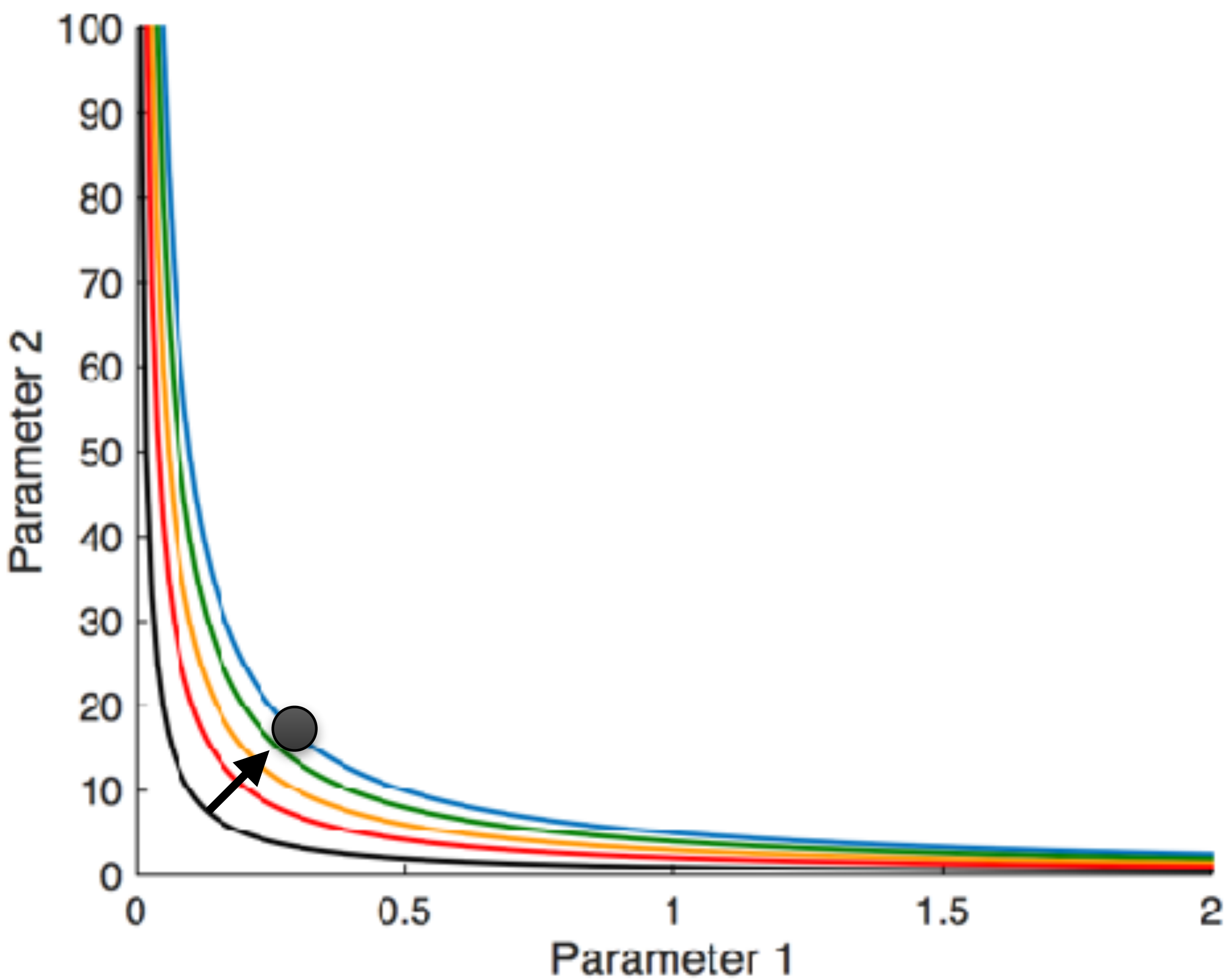




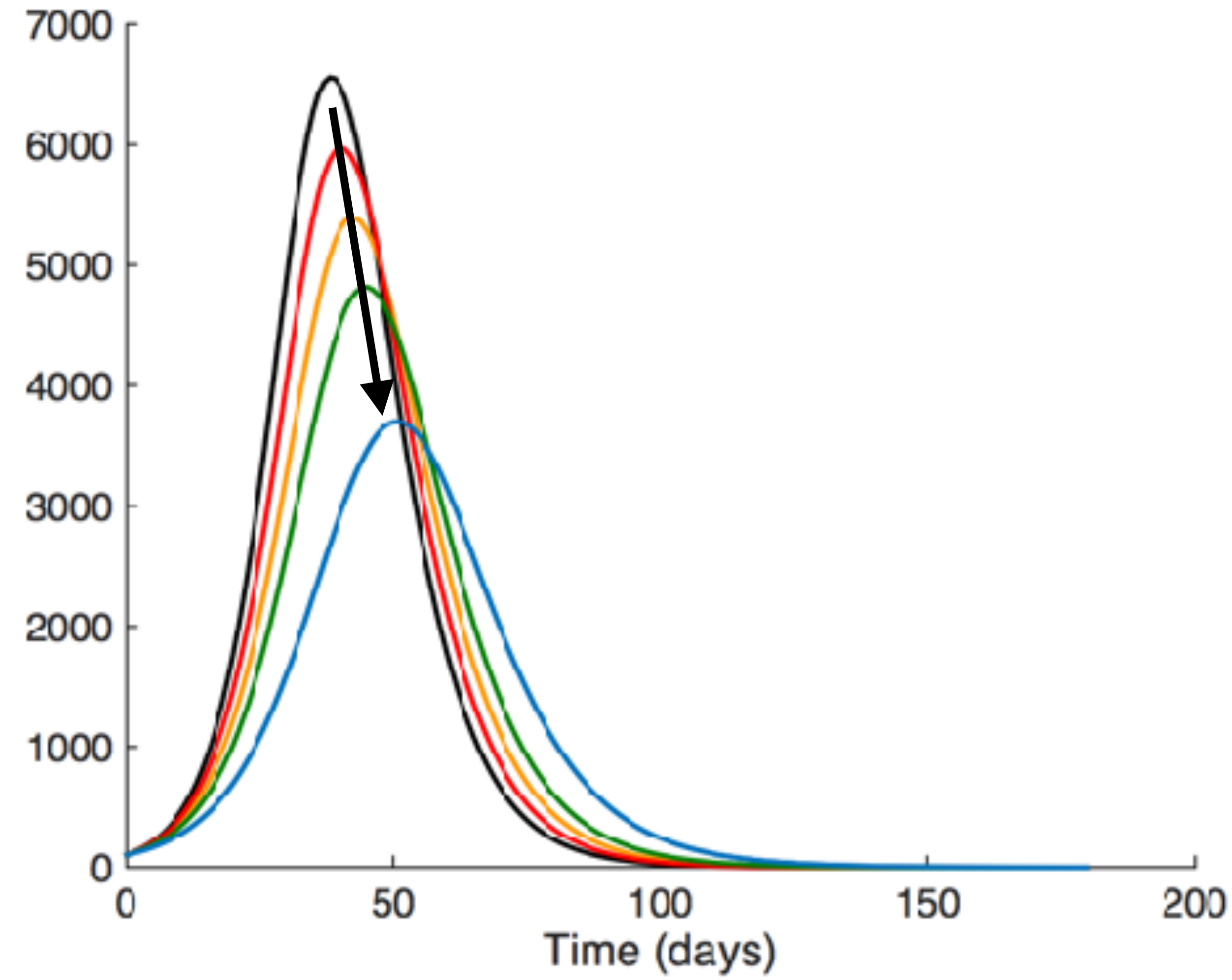
100
90
80
70
60
50
40
30
20
10
0
Parameter 2
0
0.5
1
1.5
2
Parameter 1



7000
6000
5000
4000
3000
2000
1000
0
Infected Population
0
50
100
150
200
Time (days)



100
90
80
70
60
50
40
30
20
10
0
0
0.5
1
1.5
2
Parameter 1
Parameter 2



7000
6000
5000
4000
3000
2000
1000
0
0
50
100
150
200
Time (days)

Structural Identifiability

- Assumes best case scenario - data is known perfectly at all times
- Unrealistic!
- But, necessary condition for practical identifiability with real, noisy data

Structural Identifiability

- Reveals identifiable combinations and how to restructure the model so that it is identifiable

- Can give a priori information, help direct experiment design

Categories to consider

- Structural vs. practical identifiability
- Analytical vs. numerical methods
- Global vs. local results (in parameter space)

Key Concepts

- Identifiability vs. unidentifiability
  - Practical vs. structural, local vs. global
  - When does unidentifiability matter?
- Identifiable Combinations
- Reparameterization
- Related questions: observability, distinguishability & model selection

Reparameterization

- Identifiable combinations - parameter combinations that can be estimated

- Once you know those, why reparameterize?

- Estimation issues - reparameterization provides a model that is input-output equivalent to the original but identifiable

- Often the reparameterized model has ‘sensible’ biological meaning (e.g. nondimensionalized, in terms of R_0, etc.)

Methods we'll talk about today

- Differential Algebra Approach - structural identifiability, global, analytical method
- Fisher information matrix - structural or practical, local, analytical or numerical method
- Profile likelihood - structural or practical, local, numerical method

Simple Methods

- If you have a small system, you can even plot the likelihood surface (typically can’t though—more on this with profile likelihoods)

- Simulated data approach

Analytical Methods for Structural Identifiability

Analytical Methods for Structural Identifiability

- **Laplace transform** - linear models only
- **Taylor series approach** - more broad application, but only local info & may not terminate
- **Similarity transform approach** - difficult to make algorithmic, can be difficult to assess conditions for applying theorem
- **Differential algebra approach** - rational function ODE models, global info

Bellman 1970, Cobelli & DiStefano 1980, Evans & Chappell 2000, Ollivier 1990, Ljung & Glad 1994, Audoly et al 2003

Analytical Methods for Structural Identifiability

- **Laplace transform** - linear models only
- **Taylor series approach** - more broad application, but only local info & may not terminate
- **Similarity transform approach** - difficult to make algorithmic, can be difficult to assess conditions for applying theorem
- **Differential algebra approach** - rational function ODE models, global info

Bellman 1970, Cobelli & DiStefano 1980, Evans & Chappell 2000, Ollivier 1990, Ljung & Glad 1994, Audoly et al 2003

Differential Algebra Approach

- Basic idea: use substitution & differentiation to eliminate all variables except for observed output (y)
- Clear (divide by) the coefficient for highest derivative term(s)
- This is called the **input-output equation(s)**
- Contains all structural identifiability info for the model

Differential Algebra Approach

- Use the coefficients to solve for identifiability of the model
- If unidentifiable, determine identifiable combinations
- Find identifiable reparameterization of the model?
- Easier to see with an example—

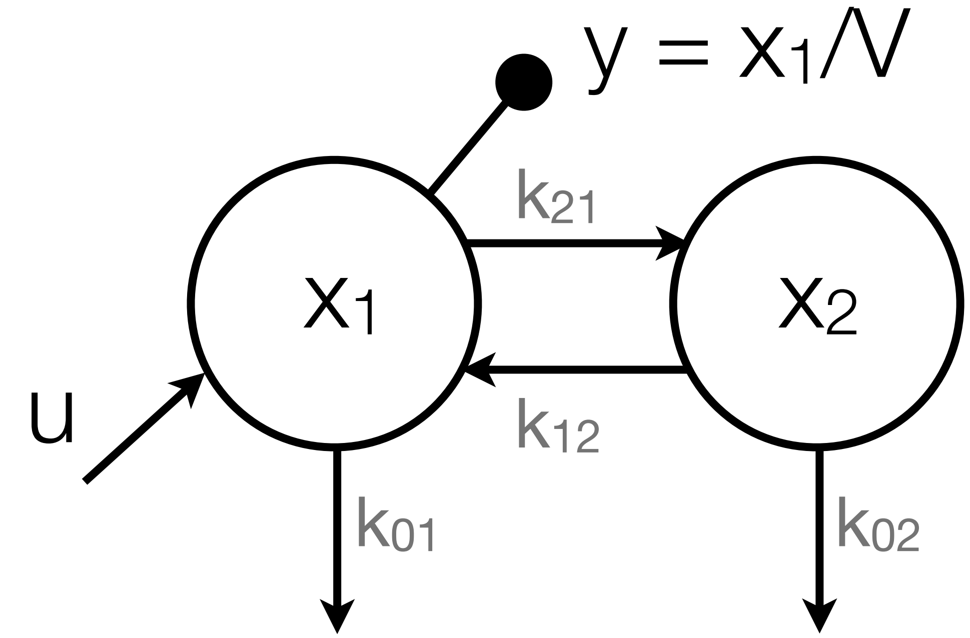
2-Compartment Example

- Linear 2-Comp Model

$$\dot{x}_1 = u + k_{12}x_2 - \left(k_{01} + k_{21}\right)x_1$$

$$\dot{x}_2 = k_{21}x_1 - \left(k_{02} + k_{12}\right)x_2$$

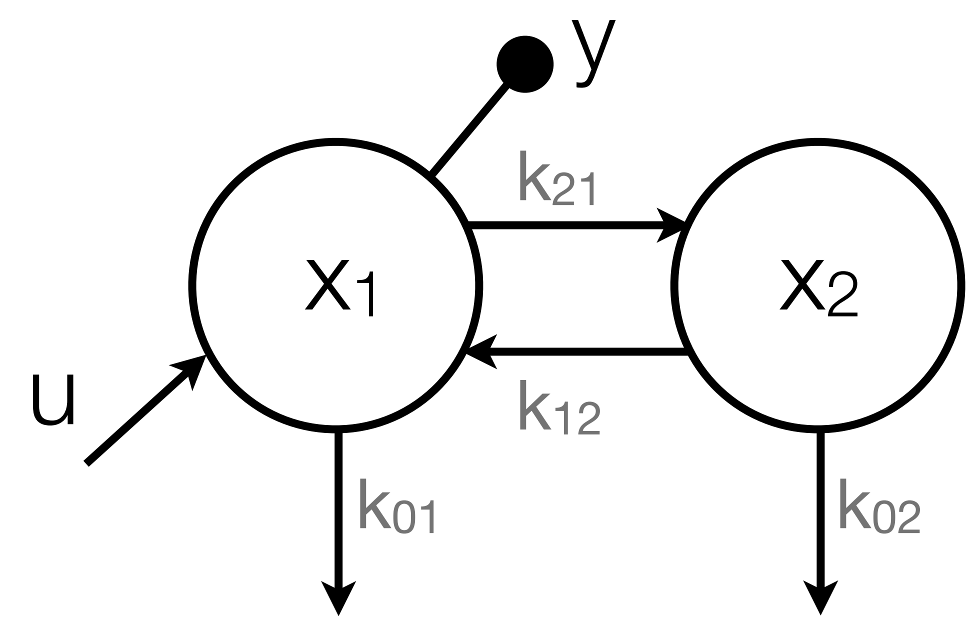
$$y = x_1 / V$$



- state variables (x)
- measurements (y)
- known input (u) (e.g. IV injection)

2-Compartment Example

$$\dot{x}_1 = u + k_{12}x_2 - \left(k_{01} + k_{21}\right)x_1$$
$$\dot{x}_2 = k_{21}x_1 - \left(k_{02} + k_{12}\right)x_2$$
$$y = x_1 / V$$

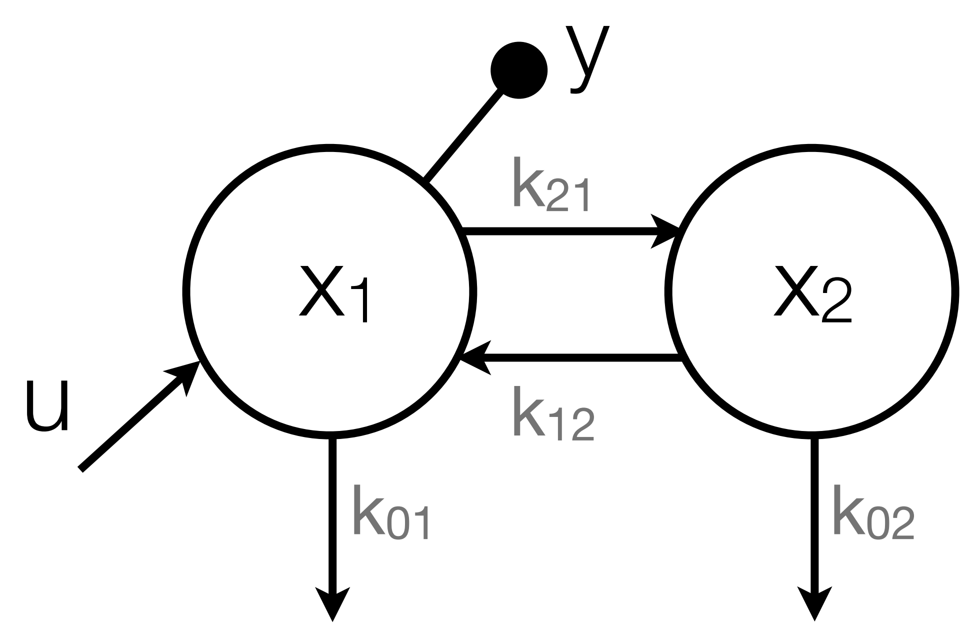


2-Compartment Example

$$y = x_1 / V$$

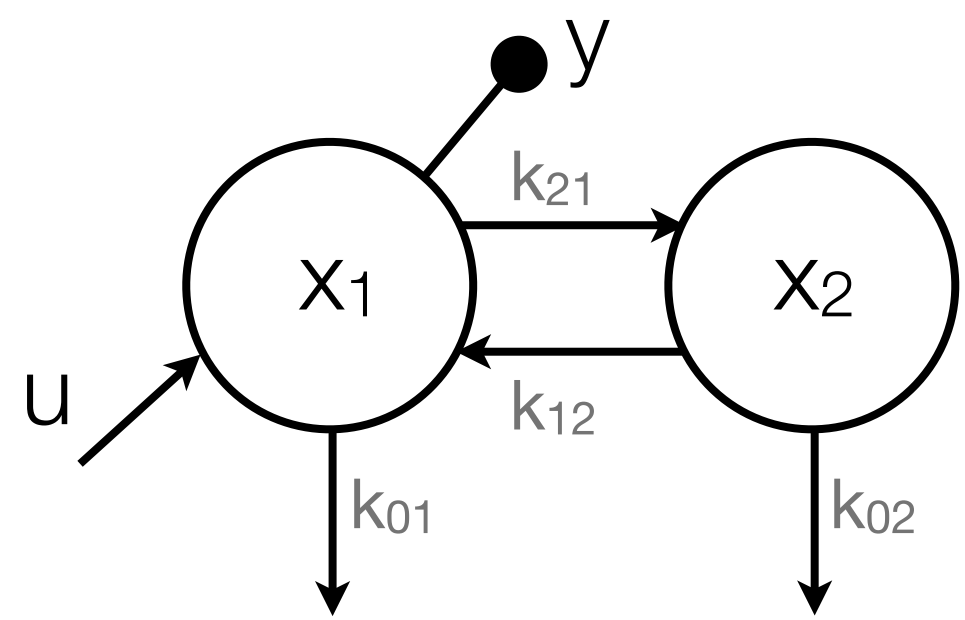
$$\dot{x}_1 = u + k_{12} x_2 - \left(k_{01} + k_{21}\right) x_1$$

$$\dot{x}_2 = k_{21} x_1 - \left(k_{02} + k_{12}\right) x_2$$



2-Compartment Example

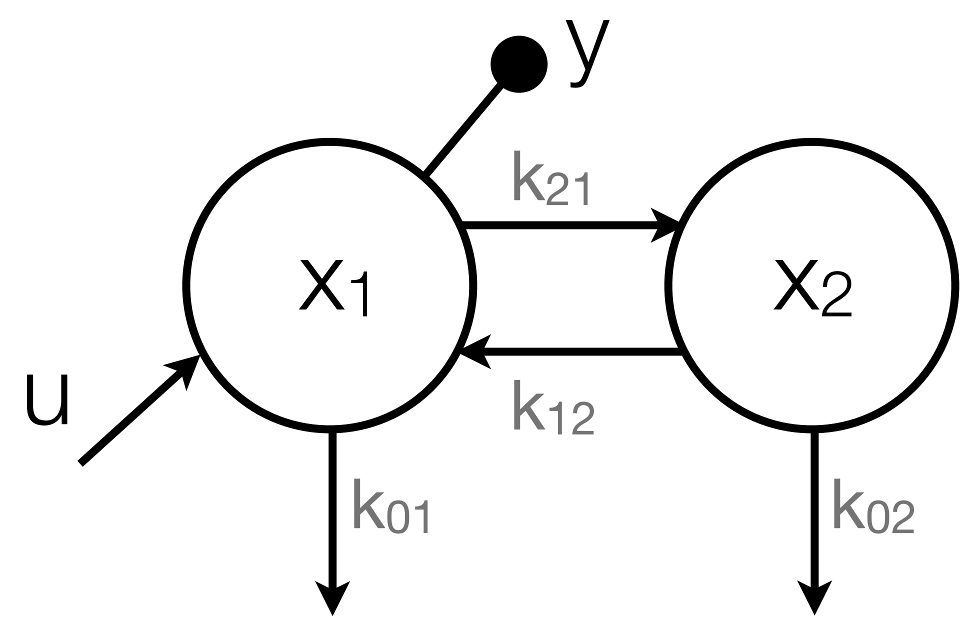
$$\dot{y}V = u + k_{12}x_2 - \left(k_{01} + k_{21}\right)yV$$
$$\dot{x}_2 = k_{21}x_1 - \left(k_{02} + k_{12}\right)x_2$$



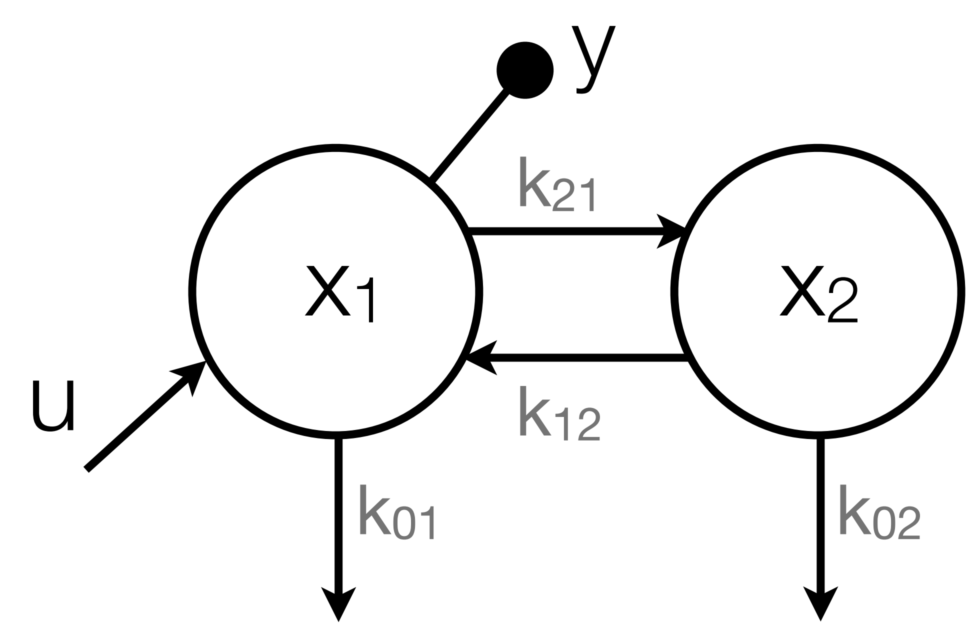
2-Compartment Example

$$\dot{x}_2 = k_{21}x_1 - (k_{02} + k_{12})x_2$$

$$\dot{x}_1 = u + k_{12}x_2 - (k_{01} + k_{21})x_1 \qquad yV$$

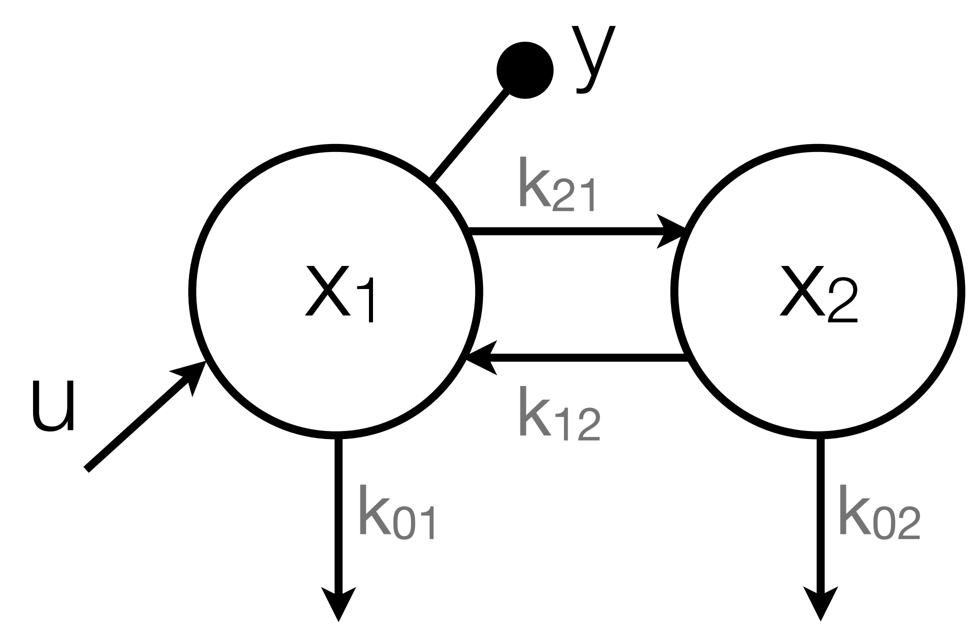


2-Compartment Example



$$\ddot{y} + \left(k_{01} + k_{21} + k_{12} + k_{02}\right)\dot{y} -$$
$$\left(k_{12}k_{21} - \left(k_{02} + k_{12}\right)\left(k_{01} + k_{21}\right)\right)y - u\left(k_{12} + k_{02}\right)/V - \dot{u}/V = 0$$

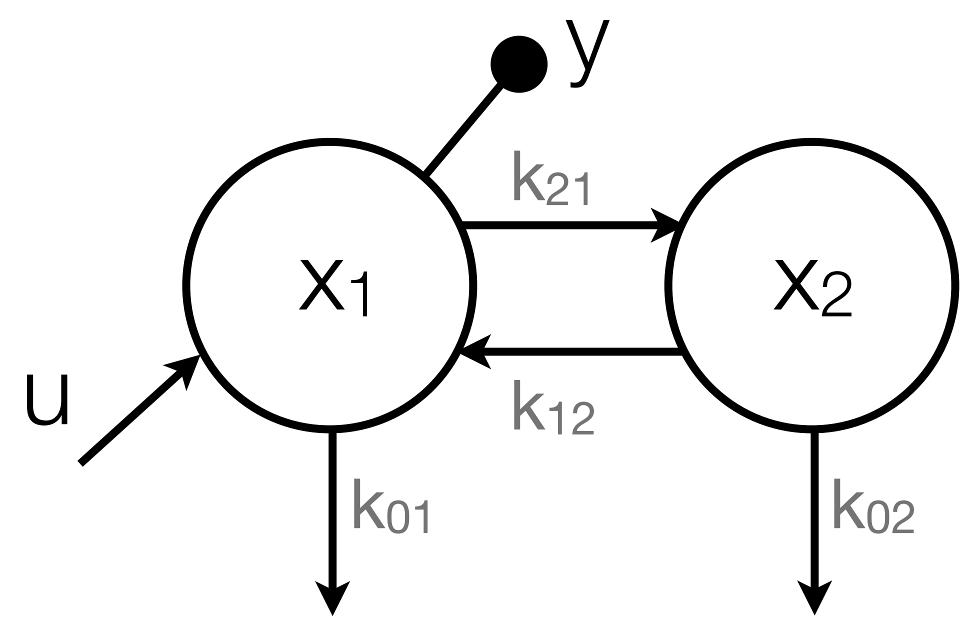
2-Compartment Example



$$\ddot{y} + \left(k_{01} + k_{21} + k_{12} + k_{02}\right)\dot{y} -$$

$$\left(k_{12}k_{21} - \left(k_{02} + k_{12}\right)\left(k_{01} + k_{21}\right)\right)y - u\left(k_{12} + k_{02}\right)/V - \dot{u}/V = 0$$

2-Compartment Example

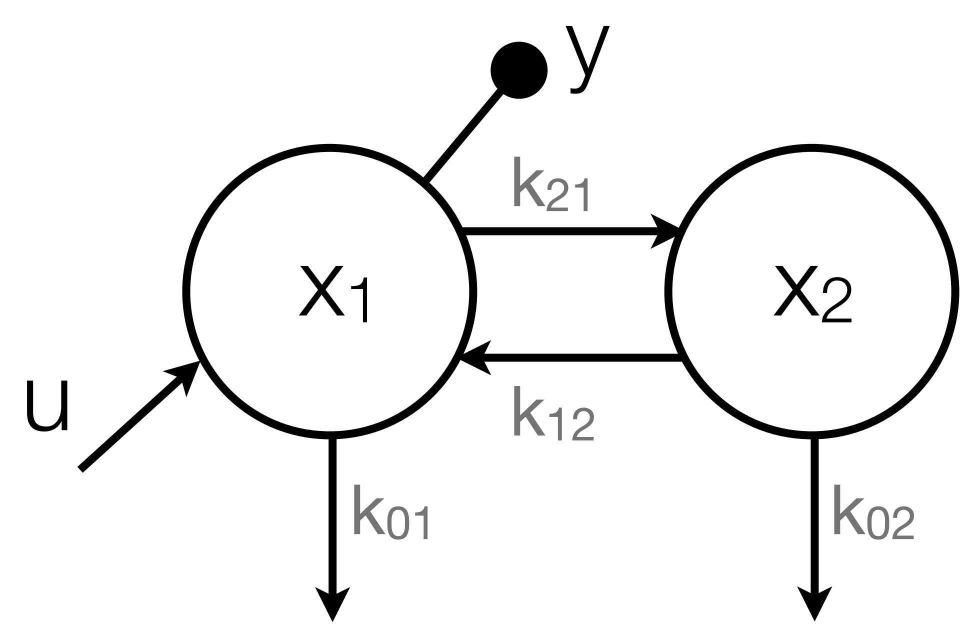


$$\ddot{y} + \left(k_{01} + k_{21} + k_{12} + k_{02}\right)\dot{y} -$$
$$\left(k_{12}k_{21} - \left(k_{02} + k_{12}\right)\left(k_{01} + k_{21}\right)\right)y - u\left(k_{12} + k_{02}\right)/V - \dot{u}/V = 0$$

$$\left(k_{01} + k_{21} + k_{12} + k_{02}\right) \qquad \left(k_{12} + k_{02}\right)/V$$

$$\left(k_{12}k_{21} - \left(k_{02} + k_{12}\right)\left(k_{01} + k_{21}\right)\right) \qquad 1/V$$

2-Compartment Example



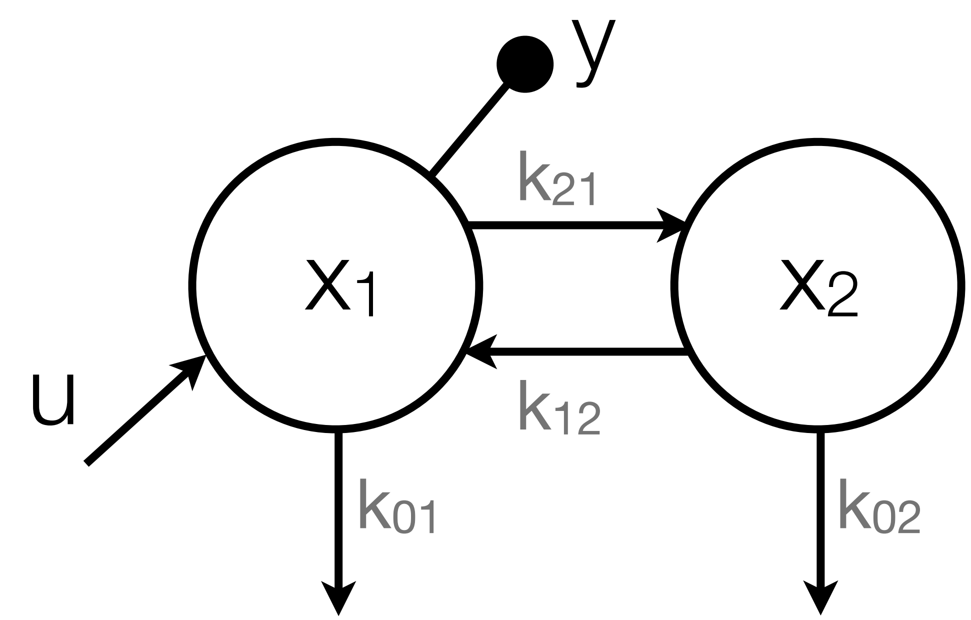
$$\left(k_{01} + k_{21} + k_{12} + k_{02}\right)$$

$$\left(k_{12} + k_{02}\right) / V$$

$$\left(k_{12}k_{21} - \left(k_{02} + k_{12}\right)\left(k_{01} + k_{21}\right)\right)$$

$$1 / V$$

2-Compartment Example



$$1 / V$$

$$\left(k_{12} + k_{02}\right) / V$$

$$\left(k_{01} + k_{21} + k_{12} + k_{02}\right)$$

$$\left(k_{12} k_{21} - \left(k_{02} + k_{12}\right)\left(k_{01} + k_{21}\right)\right)$$

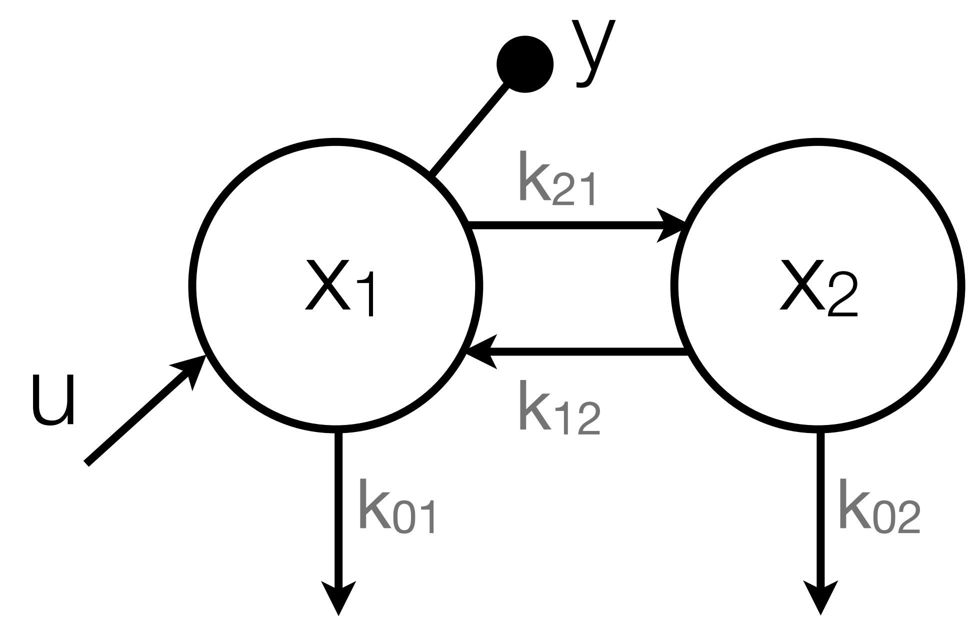
2-Compartment Example

$$1/V = a_1$$

$$\left(k_{12} + k_{02}\right)/V = a_2$$

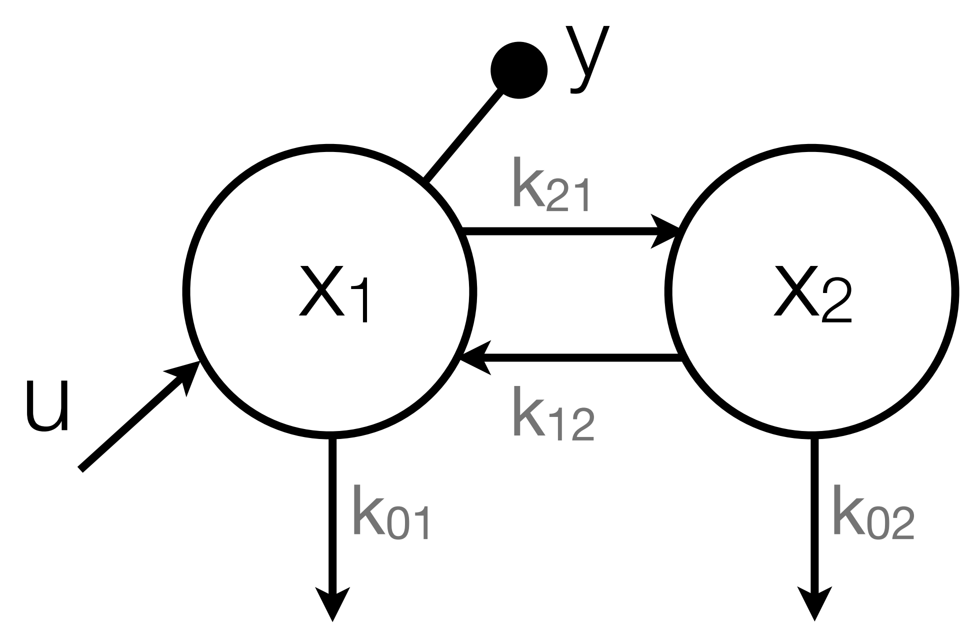
$$\left(k_{01} + k_{21} + k_{12} + k_{02}\right) = a_3$$

$$\left(k_{12}k_{21} - \left(k_{02} + k_{12}\right)\left(k_{01} + k_{21}\right)\right) = a_4$$



2-Compartment Example

$$\boxed{1/V = a_1 \Rightarrow V = 1/a_1}$$



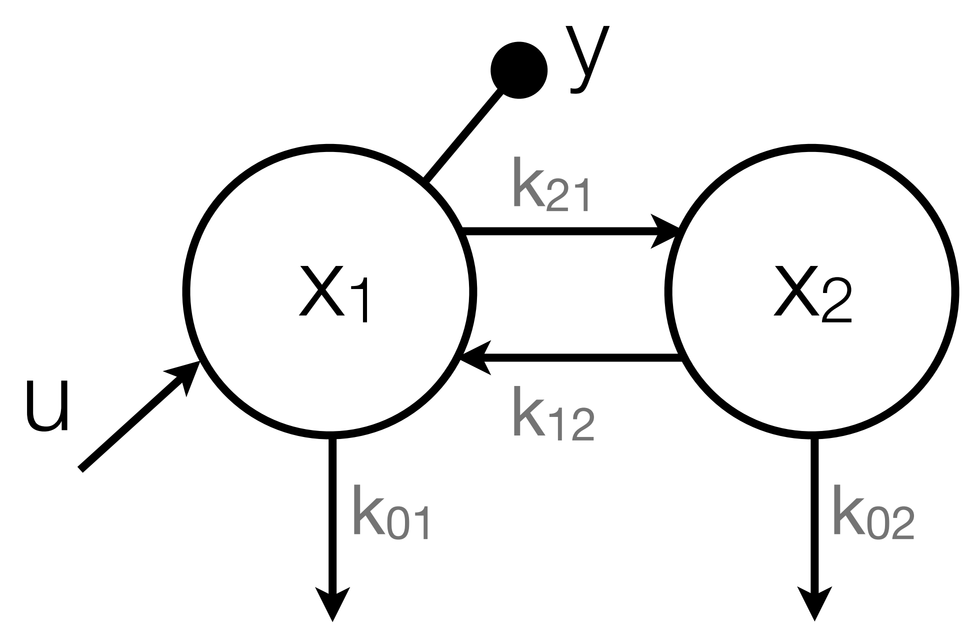
$$\left(k_{12} + k_{02}\right)/V = a_2$$

$$\left(k_{01} + k_{21} + k_{12} + k_{02}\right) = a_3$$

$$\left(k_{12}k_{21} - \left(k_{02} + k_{12}\right)\left(k_{01} + k_{21}\right)\right) = a_4$$

2-Compartment Example

$$\boxed{1/V = a_1 \Rightarrow V = 1/a_1}$$



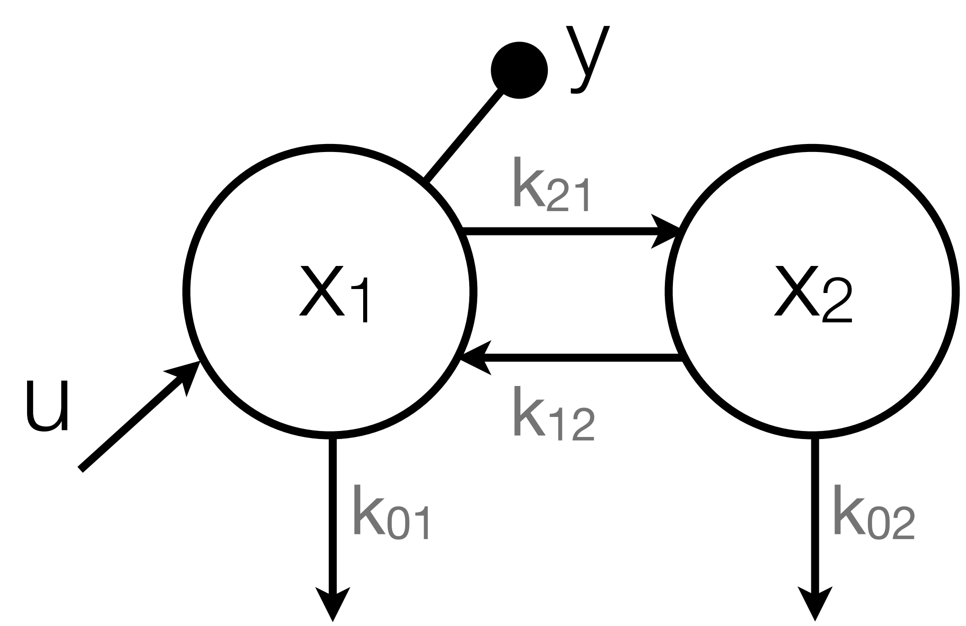
$$\left(k_{12} + k_{02}\right)/V = a_2$$

Unidentifiable

$$\left(k_{01} + k_{21} + k_{12} + k_{02}\right) = a_3$$

$$\left(k_{12}k_{21} - \left(k_{02} + k_{12}\right)\left(k_{01} + k_{21}\right)\right) = a_4$$

2-Compartment Example



Unidentifiable

$$\boxed{1/V = a_1 \Rightarrow V = 1/a_1}$$

$$\boxed{\left(k_{12} + k_{02}\right)}/V = a_2$$

$$\left(k_{01} + k_{21} + k_{12} + k_{02}\right) = a_3$$

$$\left(k_{12}k_{21} - \left(k_{02} + k_{12}\right)\left(k_{01} + k_{21}\right)\right) = a_4$$

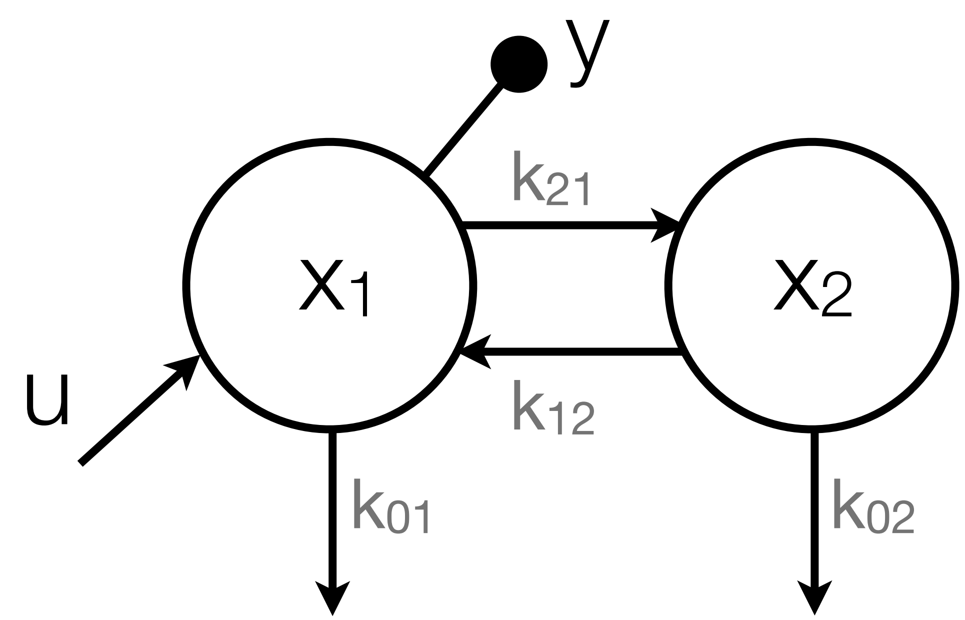
2-Compartment Example

$$\boxed{1/V = a_1 \Rightarrow V = 1/a_1}$$

$$\boxed{\left(k_{12} + k_{02}\right)}/V = a_2$$

$$\left(\boxed{k_{01} + k_{21}} + k_{12} + k_{02}\right) = a_3$$

$$\left(k_{12}k_{21} - \left(k_{02} + k_{12}\right)\left(k_{01} + k_{21}\right)\right) = a_4$$



Unidentifiable

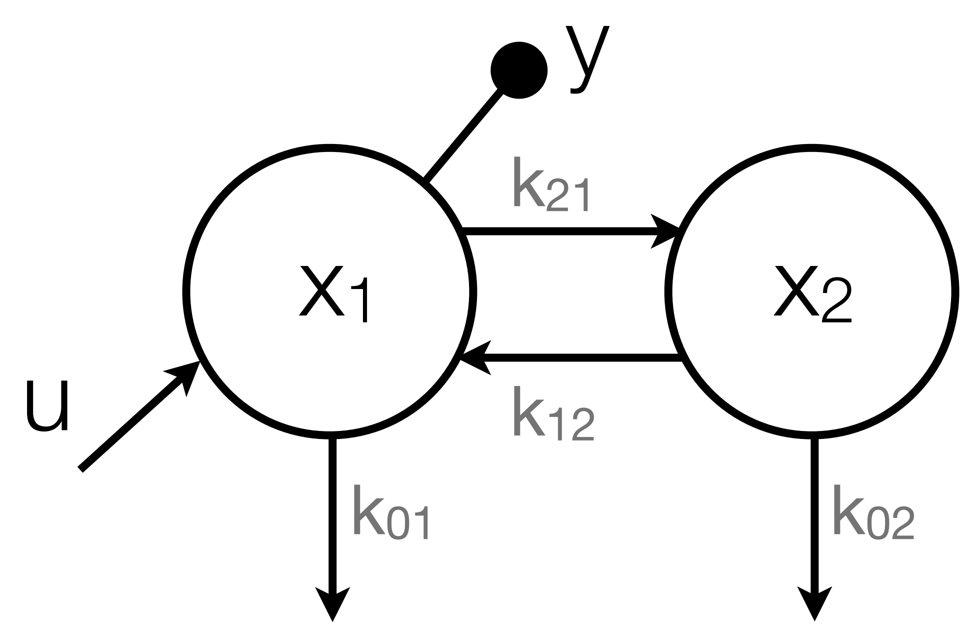
2-Compartment Example

$$\boxed{1/V = a_1 \Rightarrow V = 1/a_1}$$

$$\boxed{\left(k_{12} + k_{02}\right)}/V = a_2$$

$$\boxed{\left(k_{01} + k_{21}\right.} + k_{12} + k_{02}\left.\right) = a_3$$

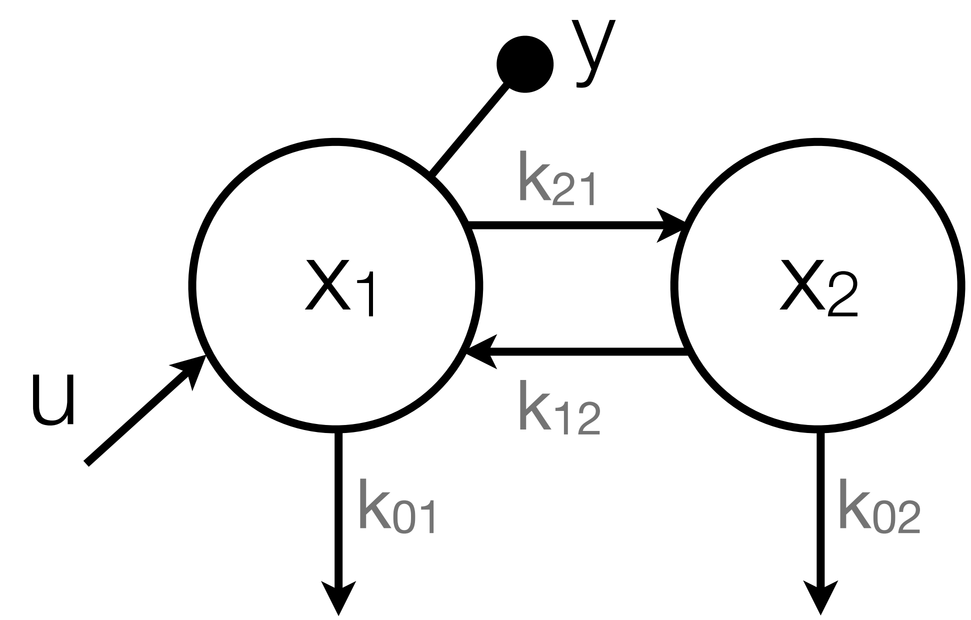
$$\boxed{\left(k_{12}k_{21}\right.} - \left(k_{02} + k_{12}\right)\left(k_{01} + k_{21}\right)\left.\right) = a_4$$



Unidentifiable

2-Compartment Example

$$\dot{x}_1 = u + k_{12}x_2 - \left(k_{01} + k_{21}\right)x_1$$
$$\dot{x}_2 = k_{21}x_1 - \left(k_{02} + k_{12}\right)x_2$$
$$y = x_1 / V$$

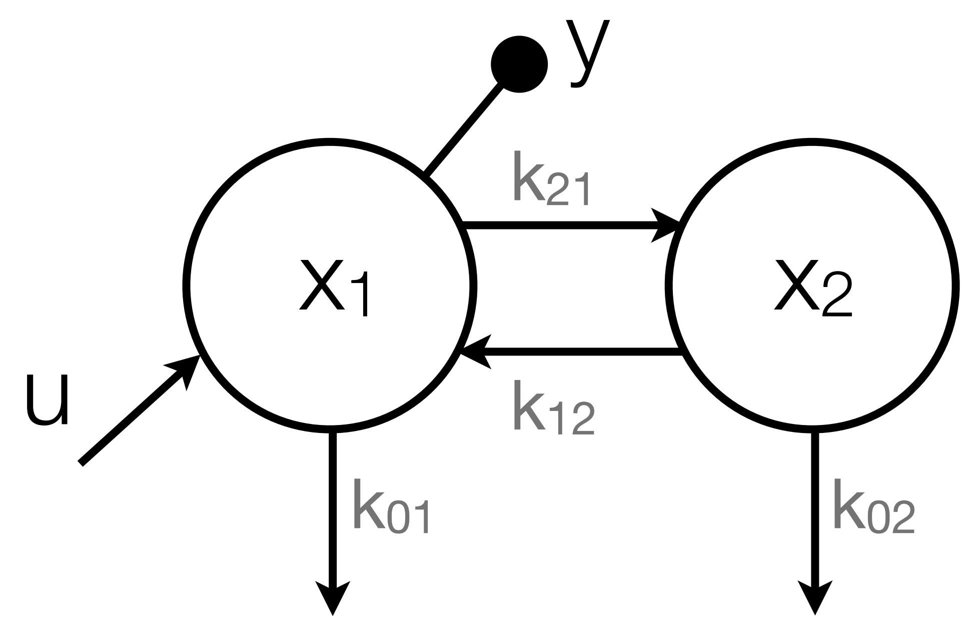


2-Compartment Example

$$\dot{x}_1 = u + \underline{k_{12}}x_2 - \left(\underline{k_{01} + k_{21}}\right)x_1$$

$$\dot{x}_2 = \underline{k_{21}}x_1 - \left(\underline{k_{02} + k_{12}}\right)x_2$$

$$y = x_1 / \underline{V}$$



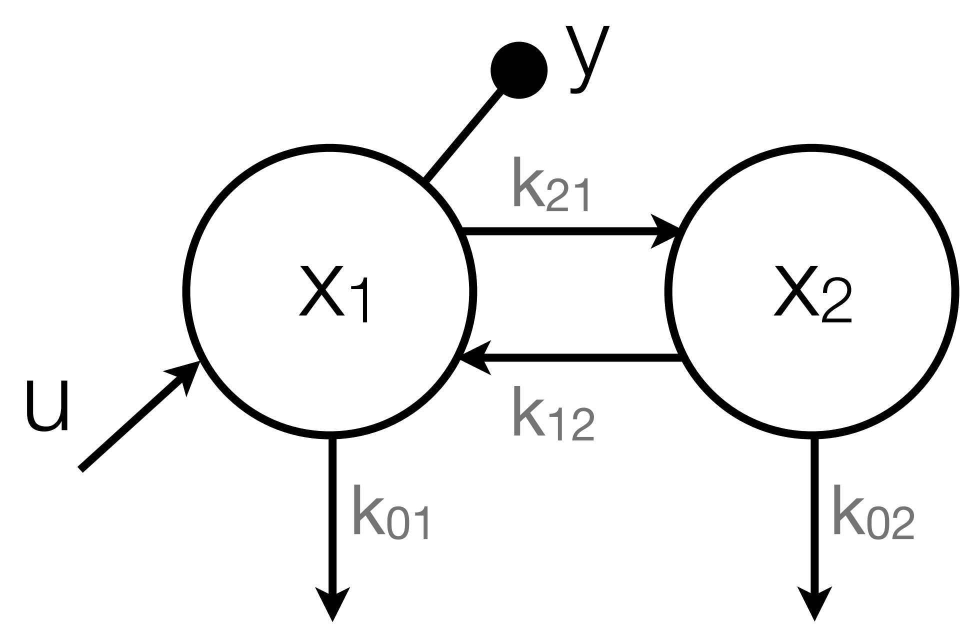
Let $\underline{x}_2 = k_{12}x_2$

2-Compartment Example

$$\dot{x}_1 = u + \underline{k_{12}}x_2 - \left(\underline{k_{01} + k_{21}}\right)x_1$$

$$\dot{x}_2 = \underline{k_{21}}x_1 - \left(\underline{k_{02} + k_{12}}\right)x_2$$

$$y = x_1 / \underline{V}$$



Let $\underline{x}_2 = k_{12}x_2$

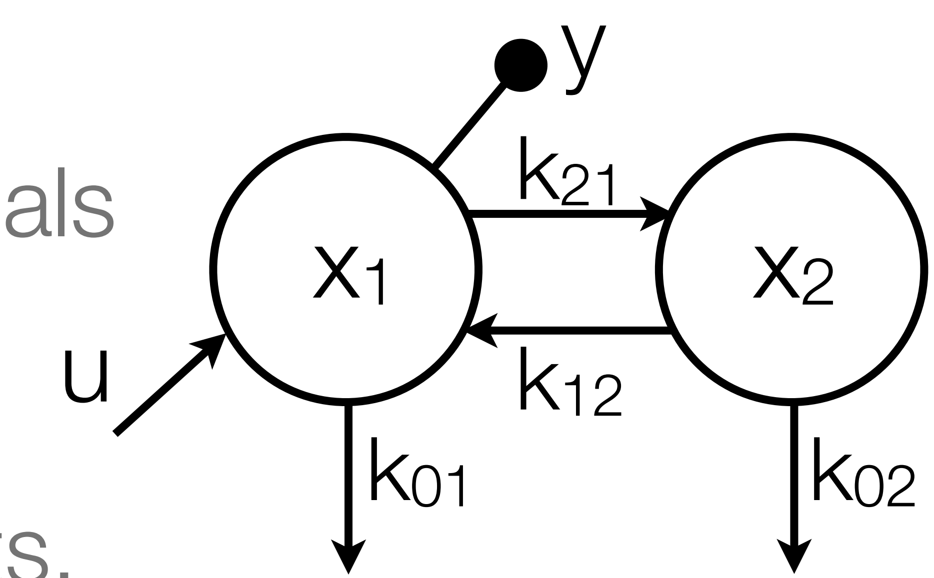
$$\dot{x}_1 = u + \underline{x}_2 - \left(\underline{k_{01} + k_{21}}\right)x_1$$

$$\underline{\dot{x}}_2 = \underline{k_{12}k_{21}}x_1 - \left(\underline{k_{02} + k_{12}}\right)\underline{x}_2$$

$$y = x_1 / \underline{V}$$

Or add information about one of the parameters

Differential Algebra Approach



- View model & measurement equations as differential polynomials
- Reduce the equations using Gröbner bases, characteristic sets, etc. to eliminate unmeasured variables (x)
- Yields **input-output equation(s)** only in terms of known variables (y, u)
- Use coefficients to test model identifiability

Ollivier 1990, Ljung & Glad 1994, Audoly et al 2003, etc.

Differential Algebra Approach

- From the coefficients, can often determine:
  - Simpler forms for identifiable combinations
  - Identifiable reparameterizations for model
- Not always easy by eye—use Gröbner bases & other methods to simplify
- Note about scaling as a useful first step (cf. nondimensionalization)

Differential Algebra Approach

- Convenient as a way to prove identifiability results for relatively broad classes of models

Numerical Methods for Identifiability Analysis

Numerical Approaches to Identifiability

- Analytical approaches can be slow, sometimes have limited applicability
- Wide range of numerical approaches
  - Sensitivities/Fisher Information Matrix
  - Profile Likelihood
  - Many others (e.g. Bayesian approaches, etc.)

Numerical Approaches to Identifiability

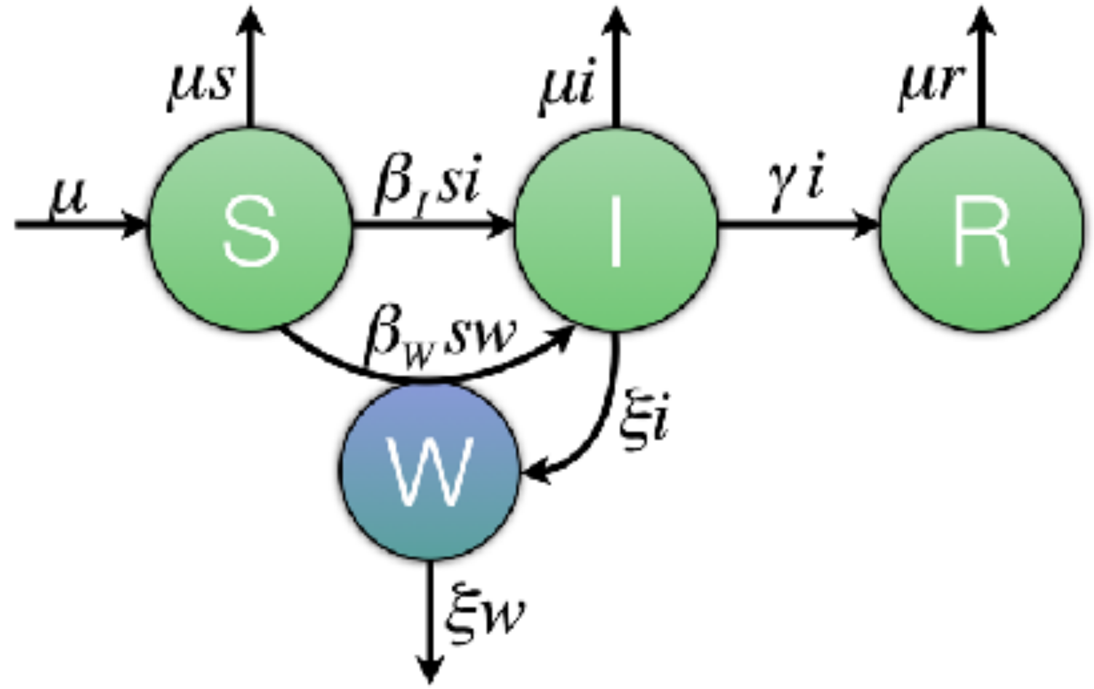
- Most can do both structural & practical identifiability
- Wide range of applicable models, often (relatively) fast
- Typically only local

Simple Simulation Approach

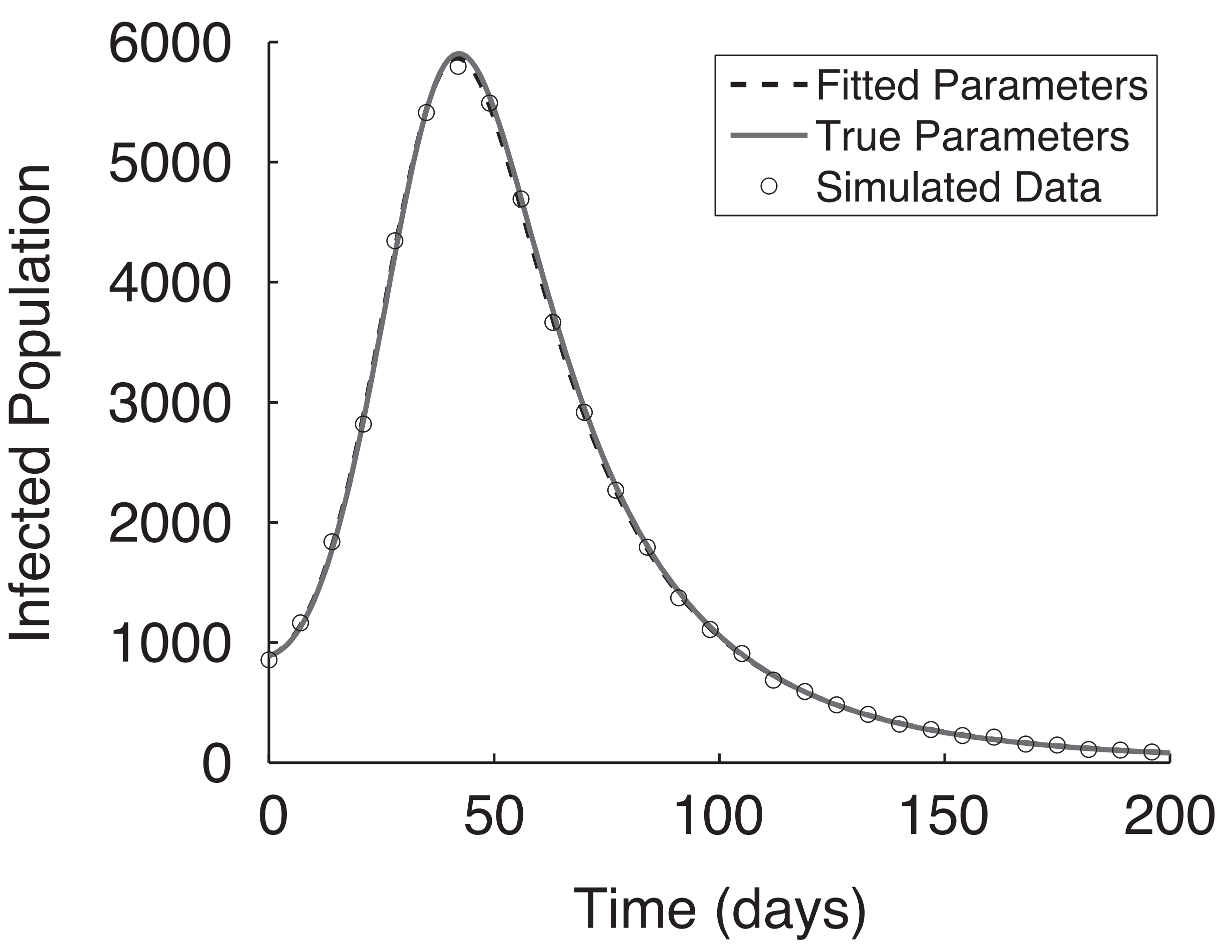
- Simulate data using a single set of ‘true’ parameter values
  - Without noise for structural identifiability
  - With noise for practical identifiability (in this case generate multiple realizations of the data)

Simple Simulation Approach

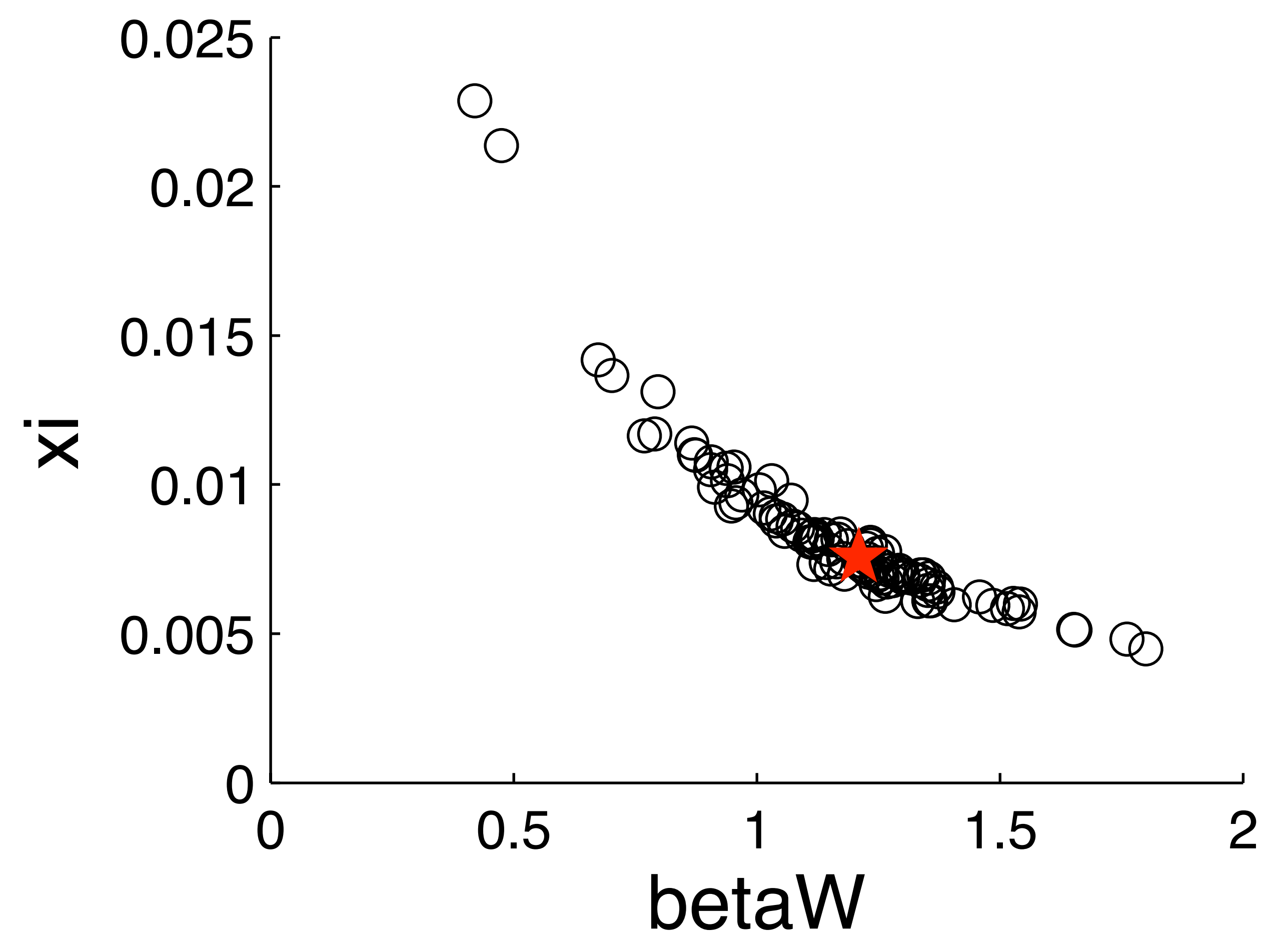
- Fit your simulated data from multiple starting points and see where your estimates land
- If they all return to the ‘true’ parameters, likely identifiable, if they do not—may be problems
- Note—unidentifiability when estimating with ‘perfect’, noise-free simulated data is most likely structural



μs
μi
μr
μ
S
$\beta_I si$
I
γi
R
$\beta_W sw$
ξi
W
ξw



Fitted Parameters
True Parameters
Simulated Data
Infected Population
Time (days)
0
1000
2000
3000
4000
5000
6000
0
50
100
150
200



xi
betaW
0
0.005
0.01
0.015
0.02
0.025
0
0.5
1
1.5
2

Parameter Sensitivities

- Output sensitivity matrix (design matrix)
- Closely related to identifiability
- Insensitive parameters
- Dependencies between columns

$$X = \begin{pmatrix} \frac{\partial y(t_1)}{\partial p_1} & \cdots & \frac{\partial y(t_1)}{\partial p_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial y(t_m)}{\partial p_1} & \cdots & \frac{\partial y(t_m)}{\partial p_n} \end{pmatrix}$$

Fisher Information Matrix

- FIM - N_P x N_P matrix

$$\left[\mathcal{I}(\theta)\right]_{i,j} = \mathrm{E}\left[\left(\frac{\partial}{\partial\theta_i}\log f(X;\theta)\right)\left(\frac{\partial}{\partial\theta_j}\log f(X;\theta)\right)\middle|\theta\right]$$

- Useful in testing practical & structural ID - represents amount of information that the output **y** contains about parameters **p**
- Cramer-Rao Bound: $\mathrm{FIM}^{-1} \leq \mathrm{Cov}(\mathbf{p})$
- Rank(FIM) = number of identifiable parameters/ combinations

Fisher Information Matrix

- For identifiability analysis, often more useful to consider (sometimes denoted the sensitivity FIM):

$$F = X^T X$$

$$X = \begin{pmatrix} \dfrac{\partial y(t_1)}{\partial p_1} & \cdots & \dfrac{\partial y(t_1)}{\partial p_n} \\ \vdots & \ddots & \vdots \\ \dfrac{\partial y(t_m)}{\partial p_1} & \cdots & \dfrac{\partial y(t_m)}{\partial p_n} \end{pmatrix}$$

- Can also derive as usual FIM with assumption of normally distributed measurement error with fixed variance (e.g. 1)

Identifiability & the FIM

- Covariance matrix/confidence interval estimates from Cramér-Rao bound: Cov $\geq$ FIM^{-1}
  - e.g. large confidence interval $\Rightarrow$ probably at least practically unID
  - Often can detect structural unID as 'near-infinite' (gigantic) variances in Cov ~ FIM^{-1}

Identifiability & the FIM

- **Rank of the FIM** is number of identifiable combinations/parameters - can do a lot by testing sub-FIMs and versions of the FIM
- Use FIM to find blocks of related parameters & how many to fix (not estimate)
- Identifiable combinations - can often see what parameters are related, but don’t know form
    - Interaction of combinations

Connections with sloppiness, active subspaces

- Use eigenvalues & eigenvectors to find sensitive/identifiable/stiff/active directions vs. insensitive/unidentifiable/sloppy/inactive
- E.g. in active subspaces, from Constantine (2015):

$$C = \int (\nabla f)(\nabla f)^T \rho(\theta)\ d\theta \qquad \nabla f = \begin{bmatrix} \frac{\partial f}{\partial \theta_1} \\ \vdots \\ \frac{\partial f}{\partial \theta_n} \end{bmatrix}$$

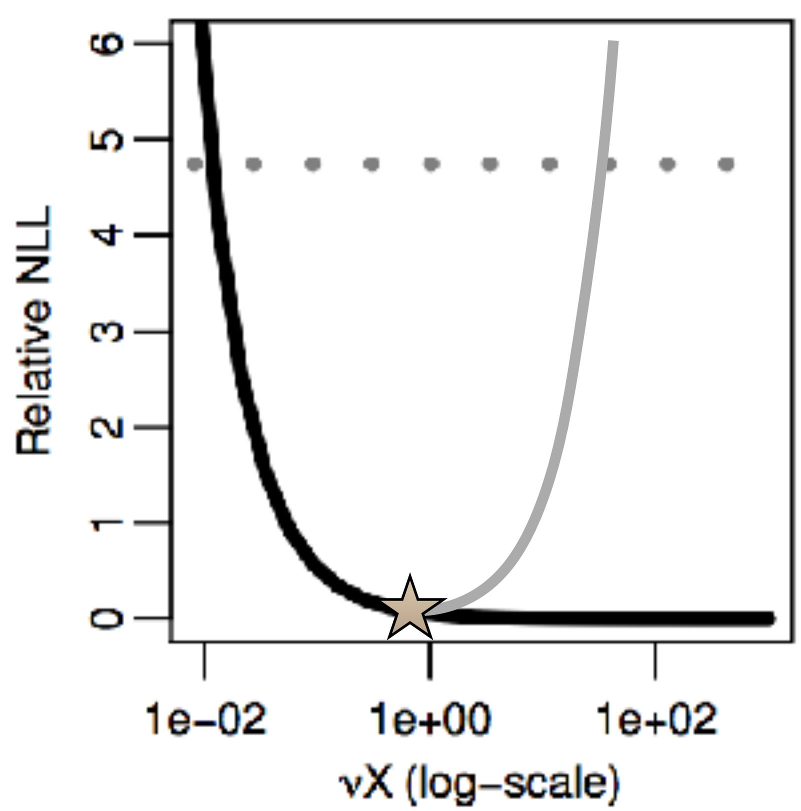
- Can write this as the weighted average sFIM:

$$C = \int F(f;\theta)\rho(\theta)\ d\theta.$$

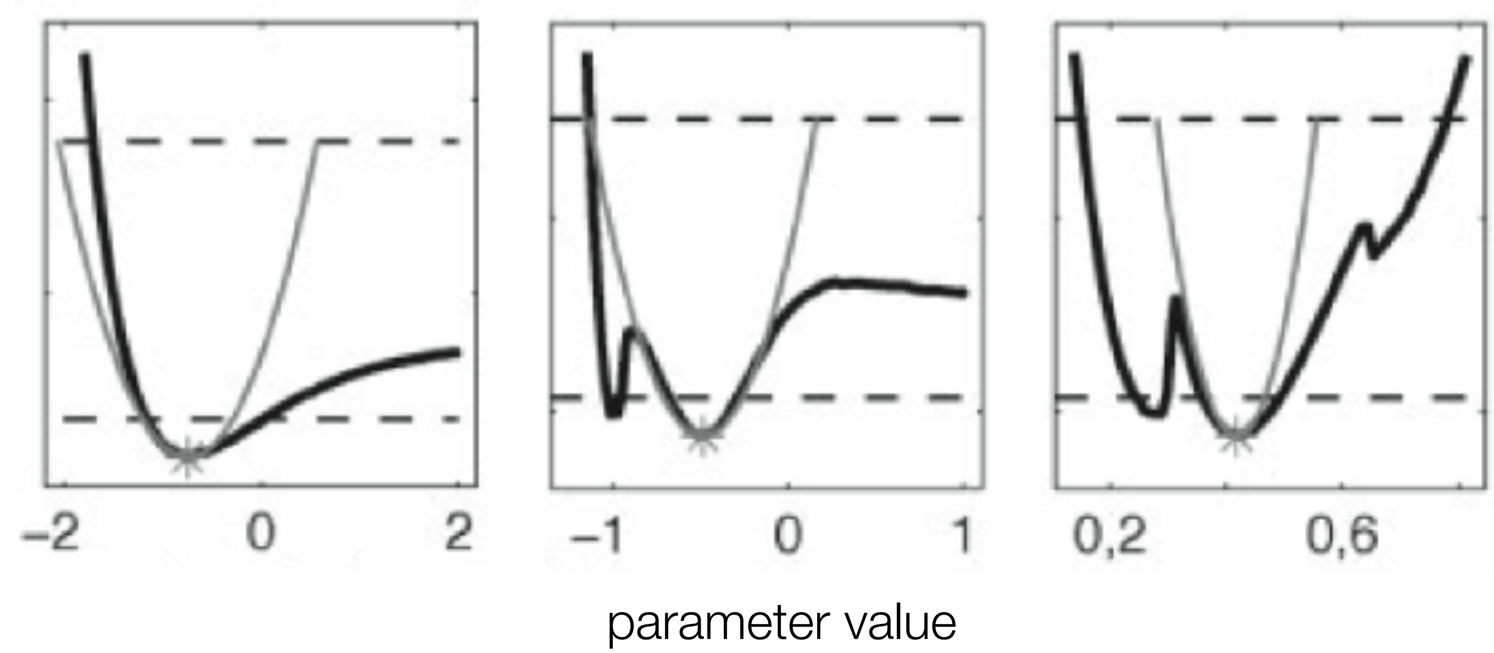
- In FIM form, QOI could be univariate or multivariate

Identifiability & the FIM

- But, be careful—FIM is local & asymptotic
- Local approximation of the curvature of the likelihood



Brouwer, Meza, Eisenberg 2017



Raue et al. 2010

Profile Likelihoods

Profile Likelihood

- Want to examine likelihood surface, but often high-dimensional

- Basic Idea: ‘profile’ one parameter at a time, by fixing it to a range of values & fitting the rest of the parameters

- Gives best fit at each point

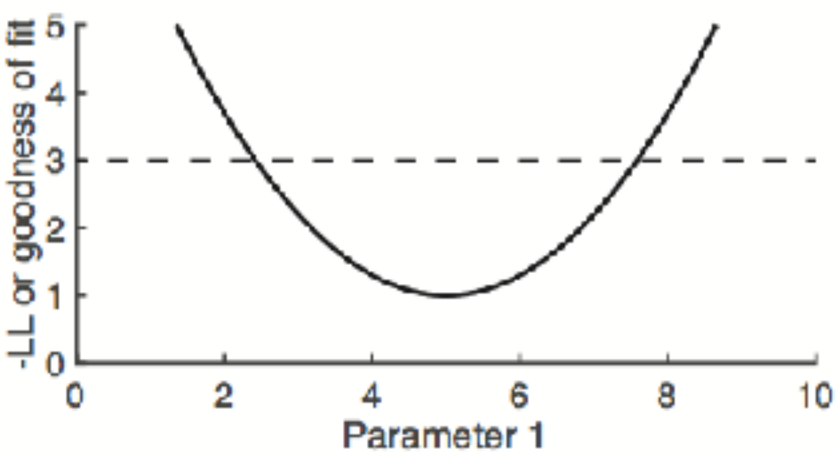
- Evaluate curvature of likelihood to determine confidence bounds on parameter (and to evaluate parameter uncertainty)

Profile Likelihood

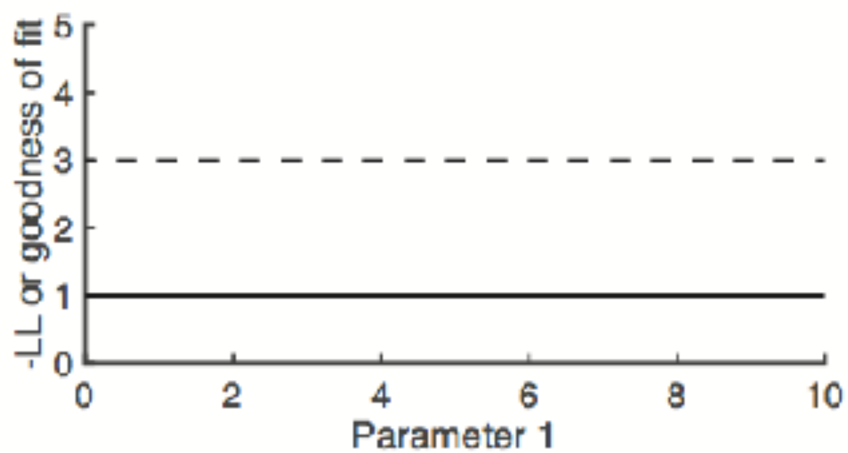
- Choose a range of values for parameter p_i
- For each value, fix p_i to that value, and fit the rest of the parameters
- Report the best likelihood/RSS/cost function value for that p_i value
- Plot the best likelihood values for each value of p_i— this is the profile likelihood

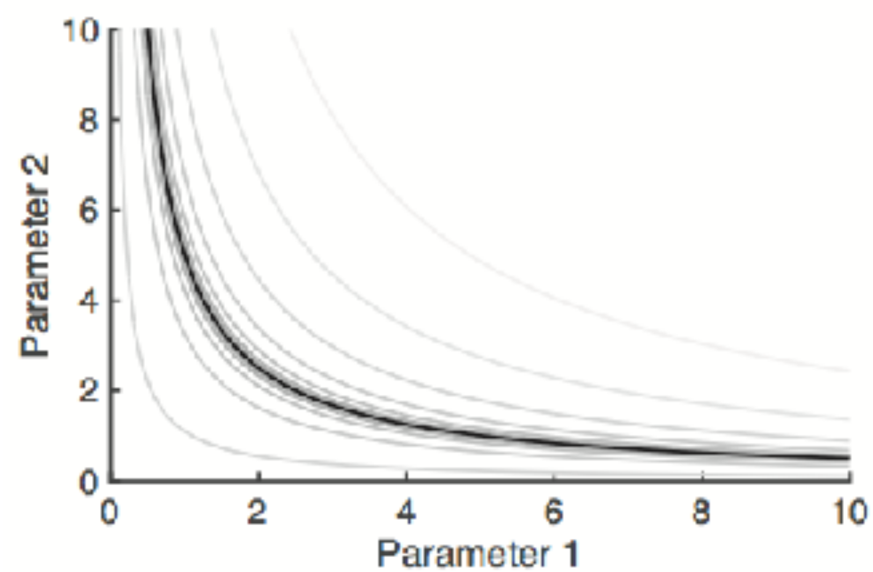
Profile Likelihoods

identifiable

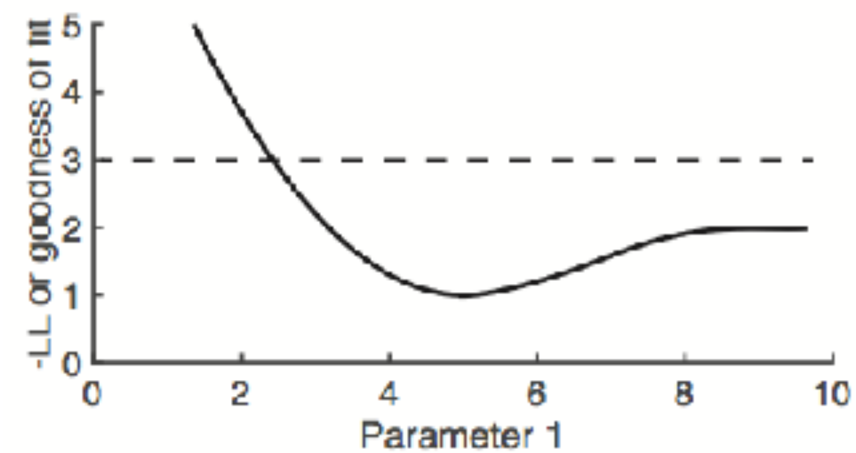


structurally unidentifiable





practically unidentifiable

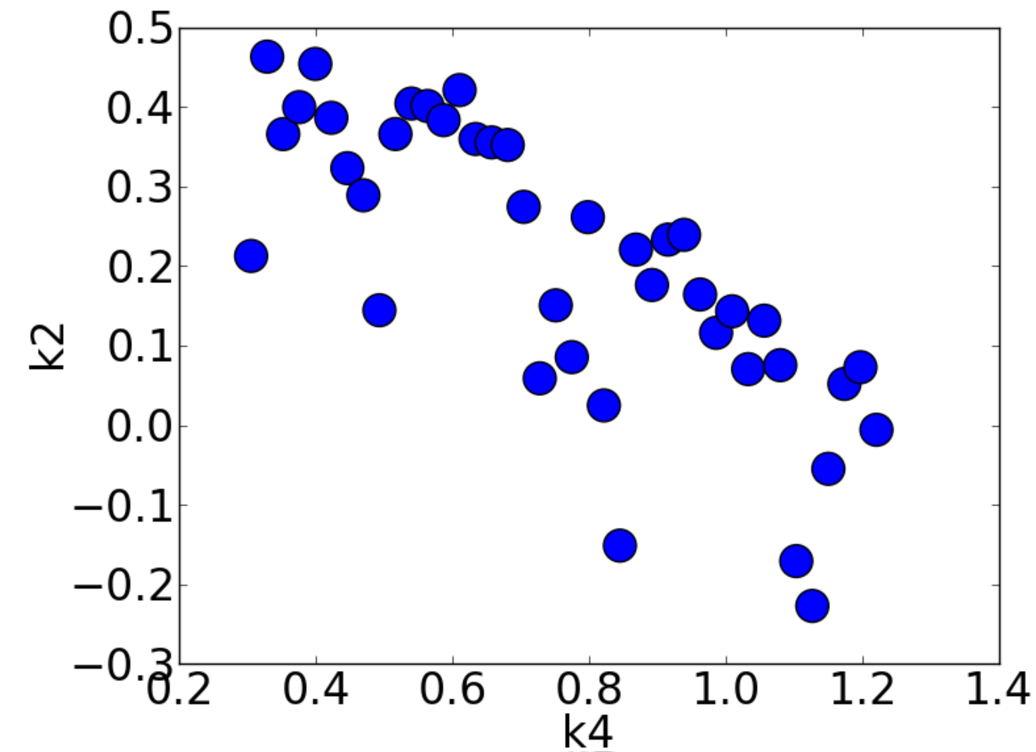


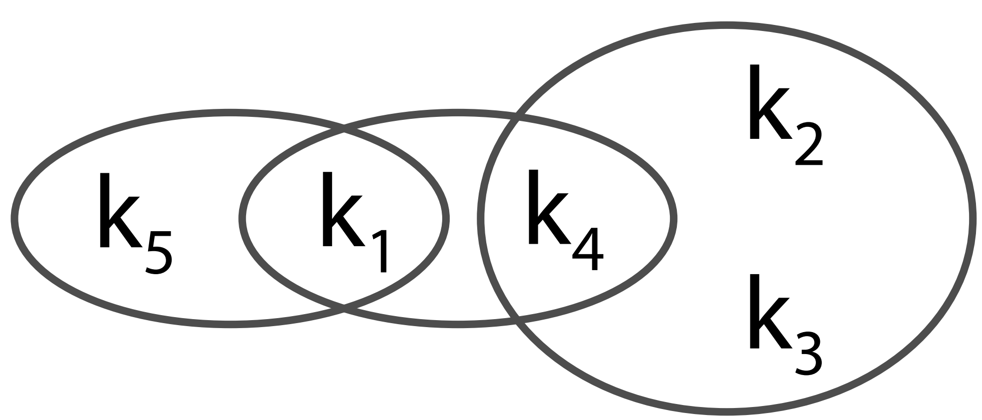
Potential issues with the profile likelihood

$$\dot{x}_1 = k_1 x_2 - (k_2 + k_3 + k_4) x_1$$

$$\dot{x}_2 = k_4 x_1 - (k_5 + k_1) x_2$$

$$y = x_1 / V$$

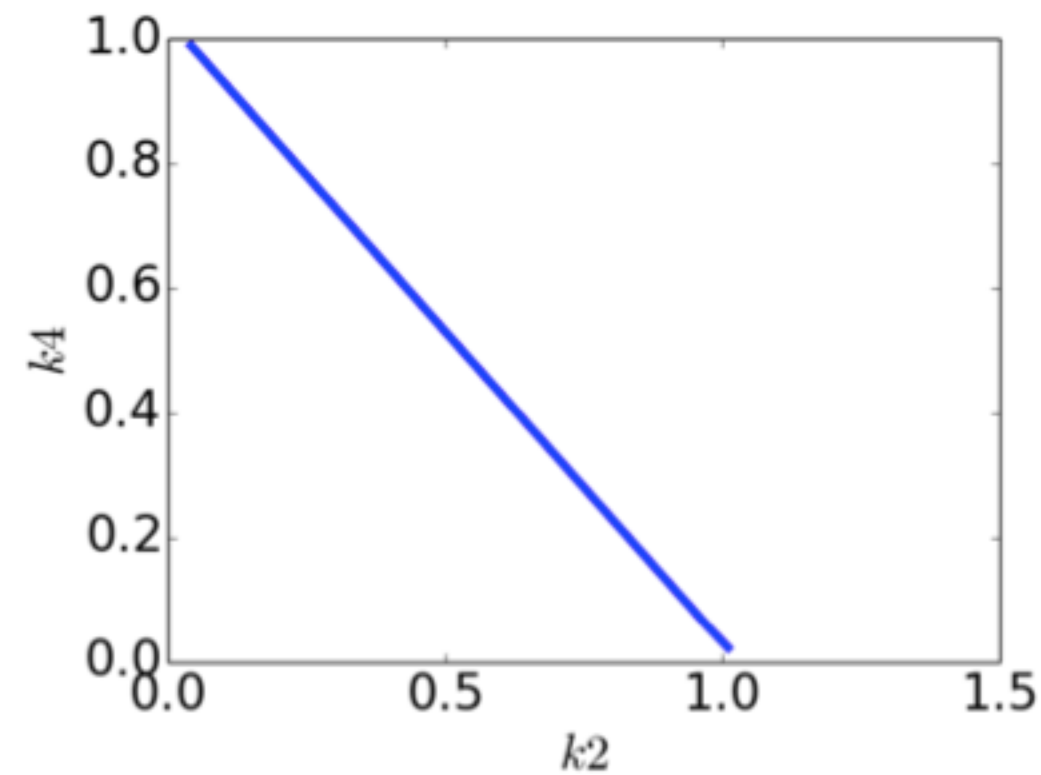


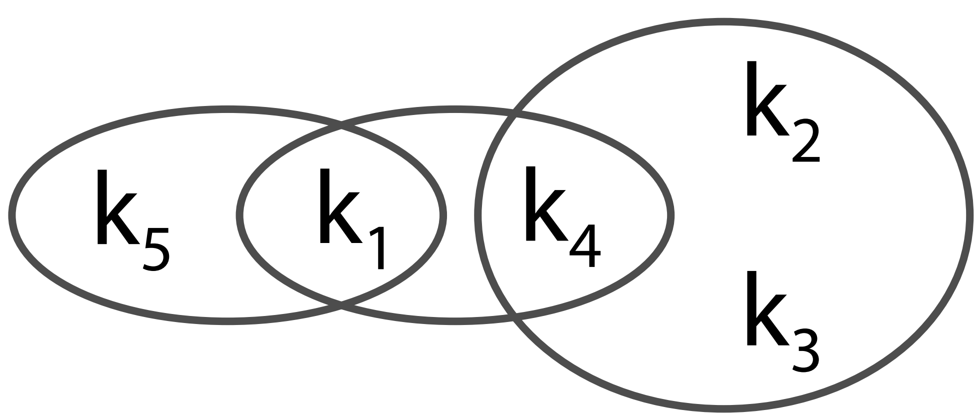


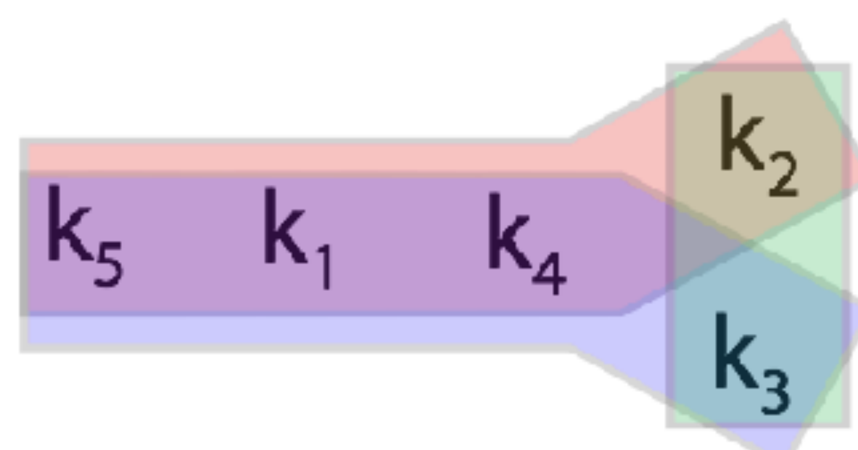
Eisenberg & Hayashi, Math Biosciences 2014

Potential issues with the profile likelihood

$$\dot{x}_1 = k_1 x_2 - (k_2 + k_3 + k_4) x_1$$
$$\dot{x}_2 = k_4 x_1 - (k_5 + k_1) x_2$$
$$y = x_1 / V$$







Eisenberg & Hayashi, Math Biosciences 2014

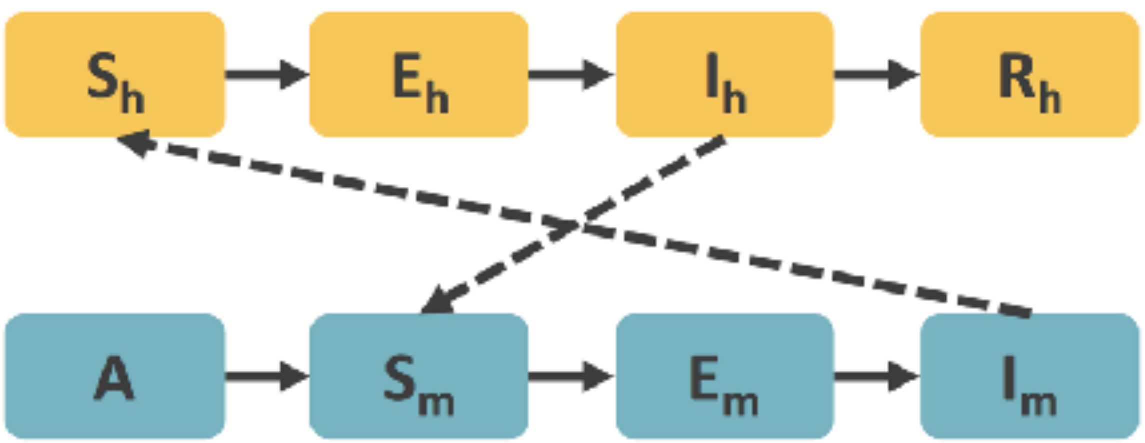
Profile Likelihood & ID

- Can generate confidence bounds based on the curvature of the profile likelihood
- Flat or nearly flat regions indicate identifiability issues
- Can generate simulated ‘perfect’ data to test structural identifiability

Profile Likelihood

- Can also help reveal the form of identifiable combinations
  - Look at relationships between parameters when profiling
  - However, can be problematic when too many degrees of freedom
- Can do the analogous thing in an MCMC or Bayesian context by looking at pairwise plots of parameter space samples

Dengue Model Example



$$\frac{dS_h}{dt} = \mu(1 - S_h) - \beta^*_{mh} S_h I_m$$

$$\frac{dE_h}{dt} = \beta^*_{mh} S_h I_m - \alpha E_h - \mu E_h$$

$$\frac{dI_h}{dt} = \alpha E_h - \eta I_h - \mu I_h$$

$$\frac{dR_h}{dt} = \eta I_h - \mu R_h$$

$$\frac{dA}{dt} = \xi^*(S_m + E_m + I_m)(1 - A) - \mu^*_a A$$

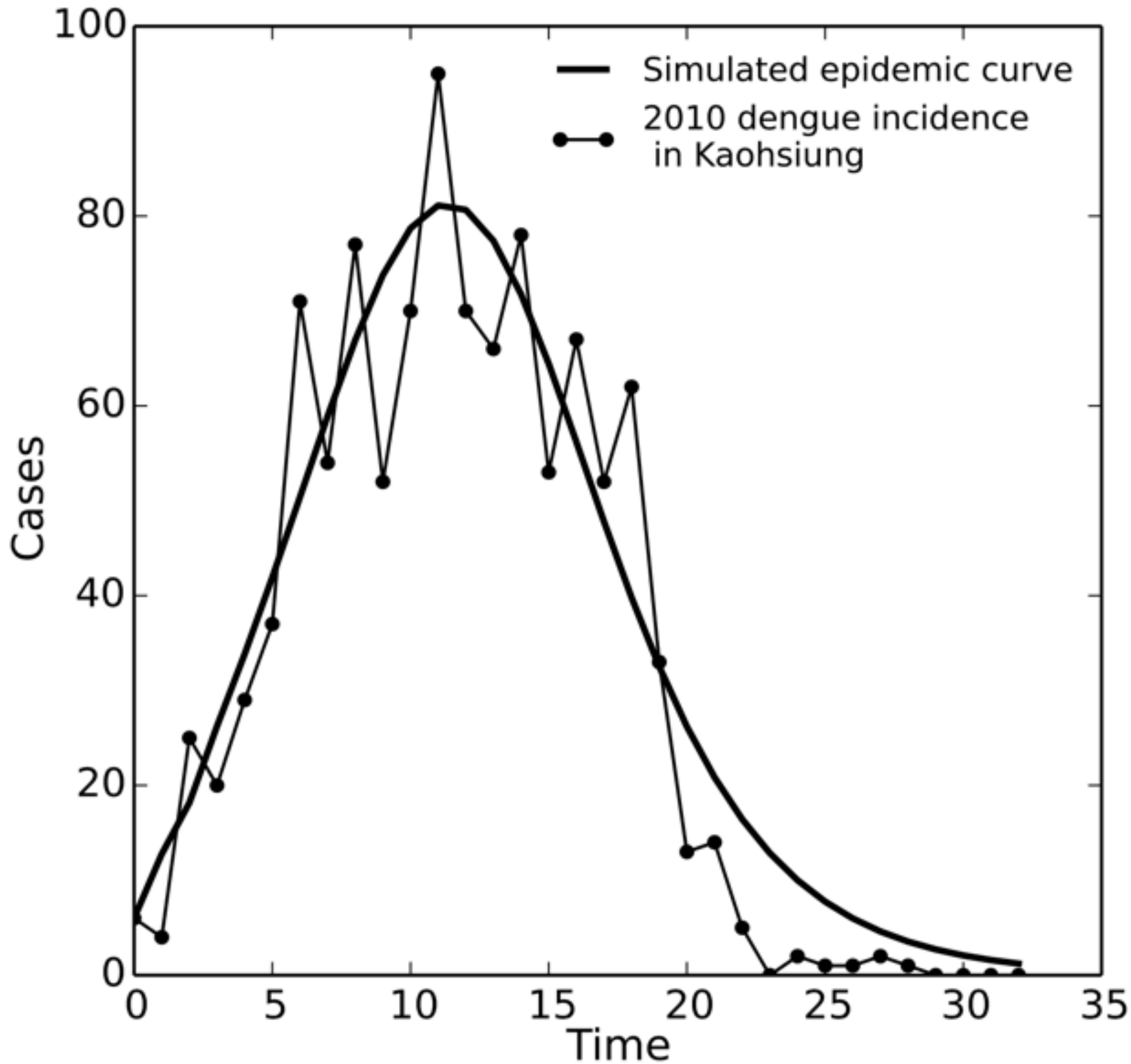
$$\frac{dS_m}{dt} = A - \beta_{hm} S_m I_h - \mu_m S_m$$

$$\frac{E_m}{dt} = \beta_{hm} S_m I_h - \gamma E_m - \mu_m E_m$$

$$\frac{I_m}{dt} = \gamma E_m - \mu_m I_m$$

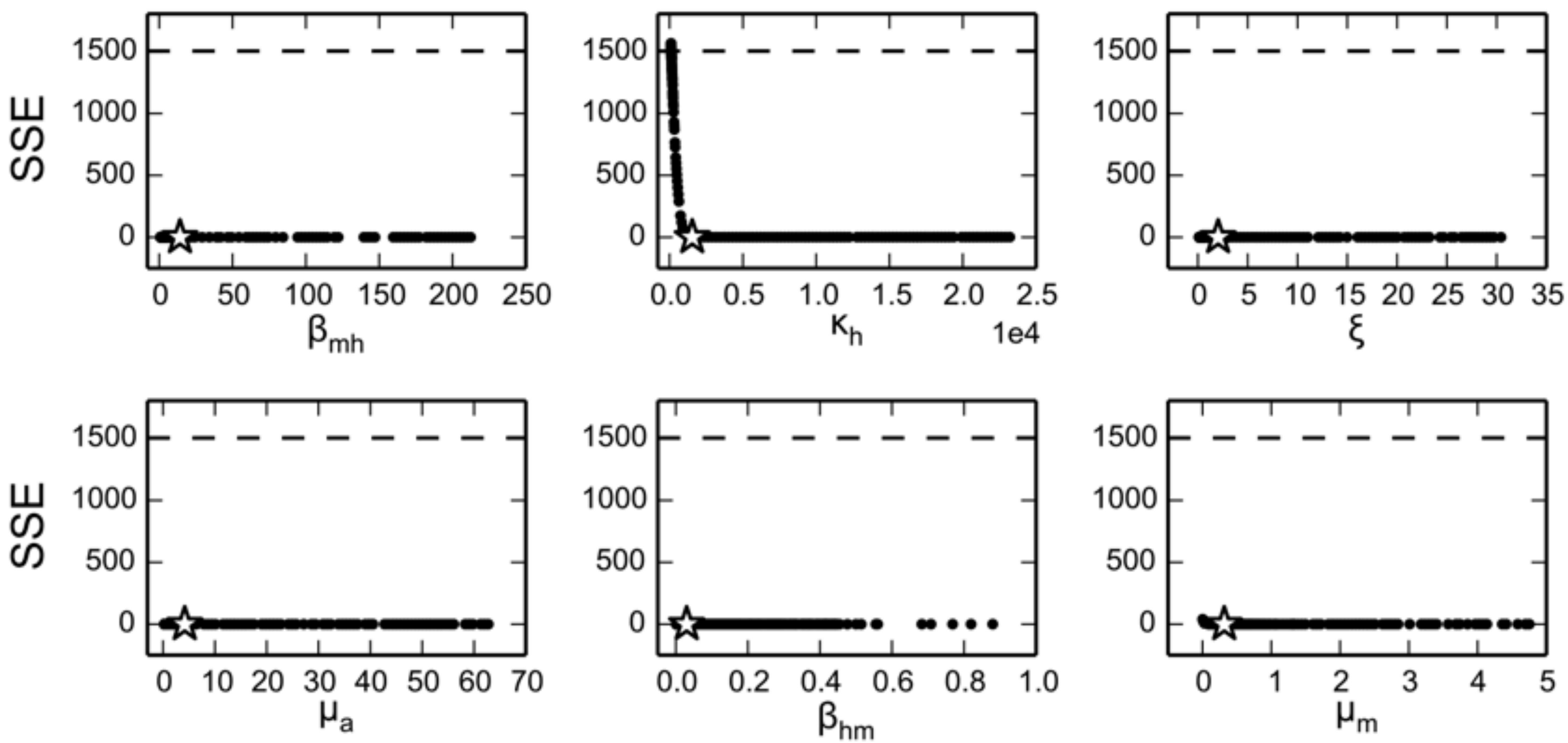
Measurement Model & Structural Identifiability

- Measure human incidence data, $y = \kappa_h \alpha E_h$, integrated to weekly incidence
- Differential algebra approach and FIM-based approaches show structural identifiability

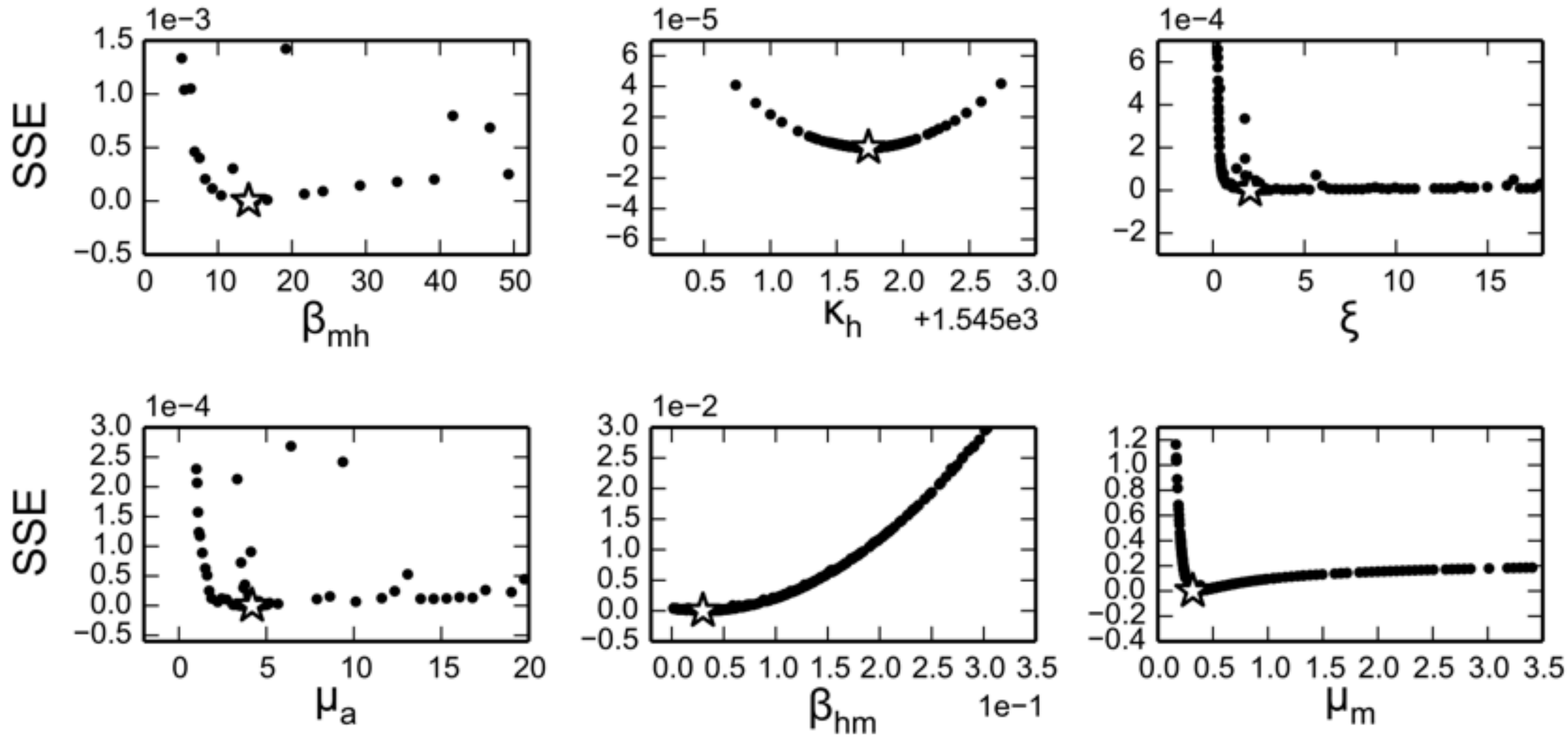


$\beta_{mh} = 14.15$
$\xi = 2.03$
$\beta_{hm} = 0.03$
$\mu_a = 4.18$
$\mu_m = 0.32$
$\kappa_h = 1546.74$

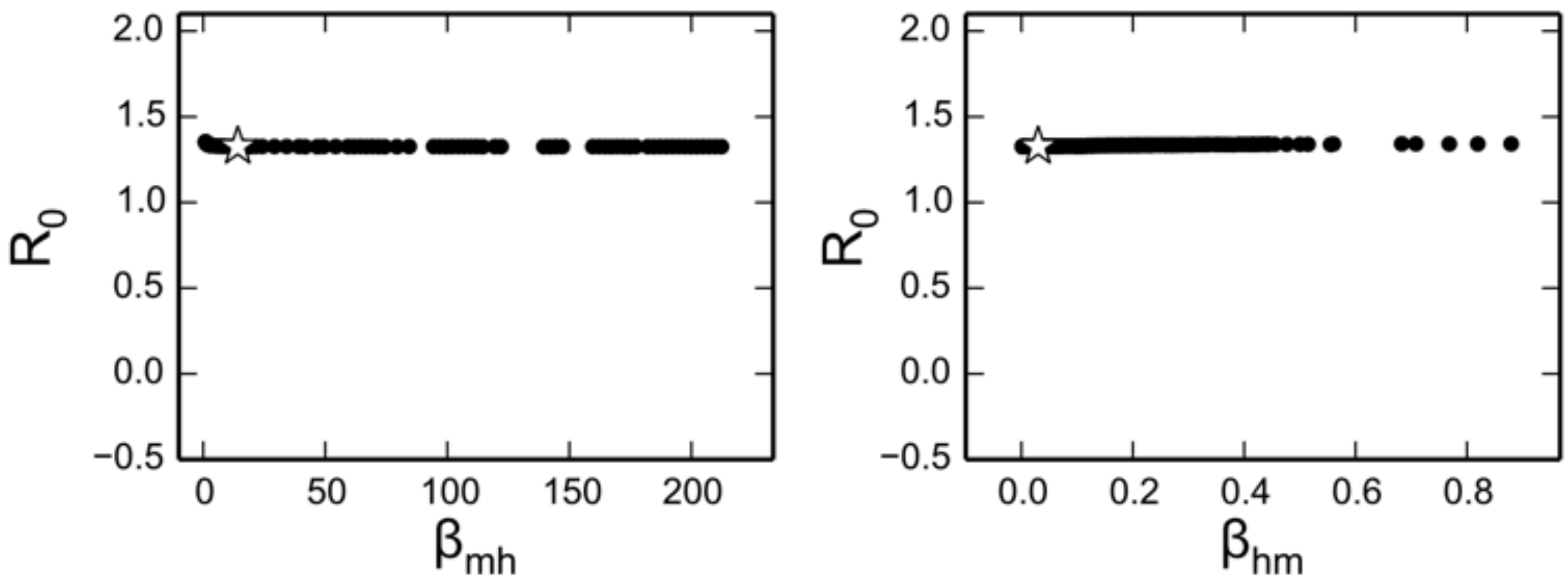
What about practical identifiability?



What about practical identifiability?

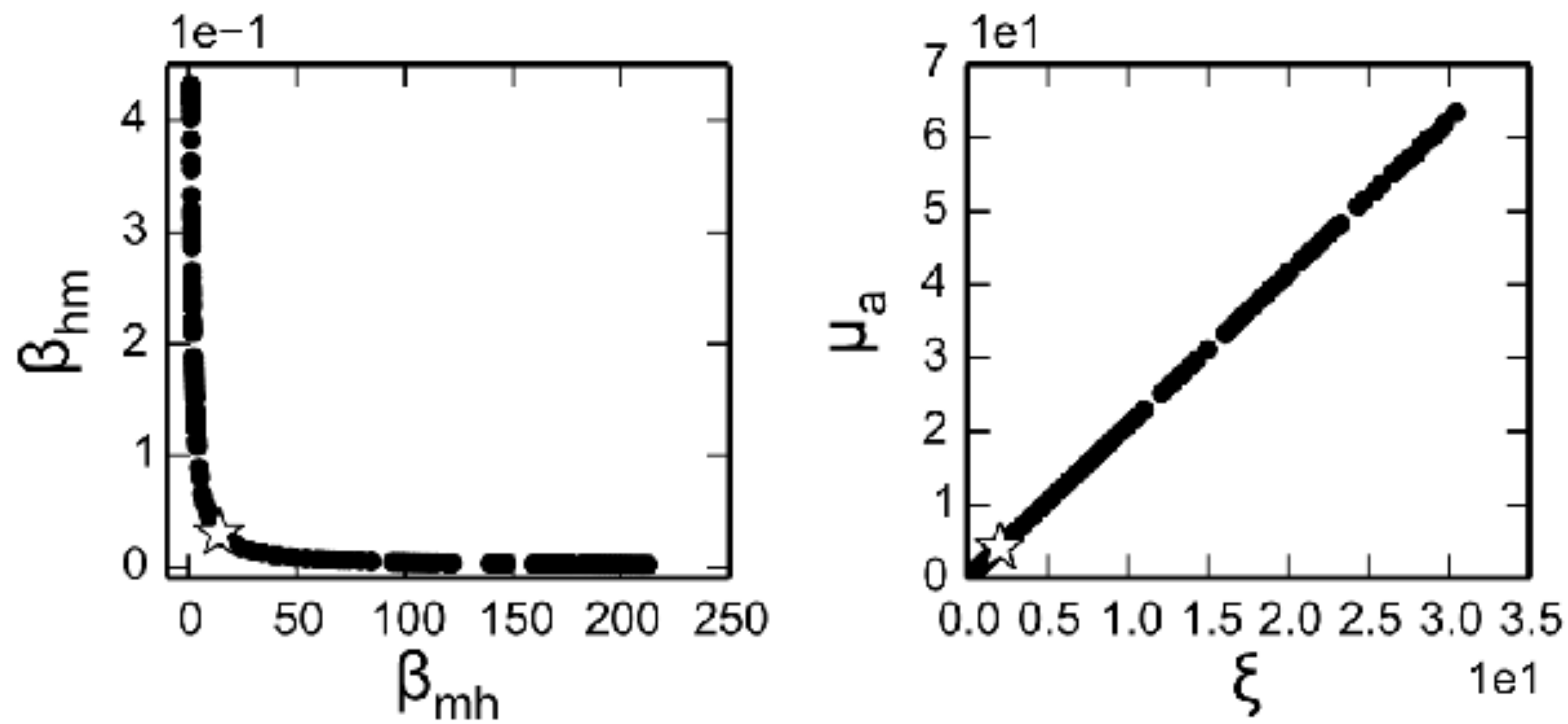


How does this affect R0?



$$\mathcal{R}_0 = \sqrt{\frac{S_m \alpha \beta_{hm} \beta_{mh} \gamma}{(\alpha + \mu)(\eta + \mu)(\gamma + \mu_m)\mu_m}}.$$

Practically Identifiable Combinations

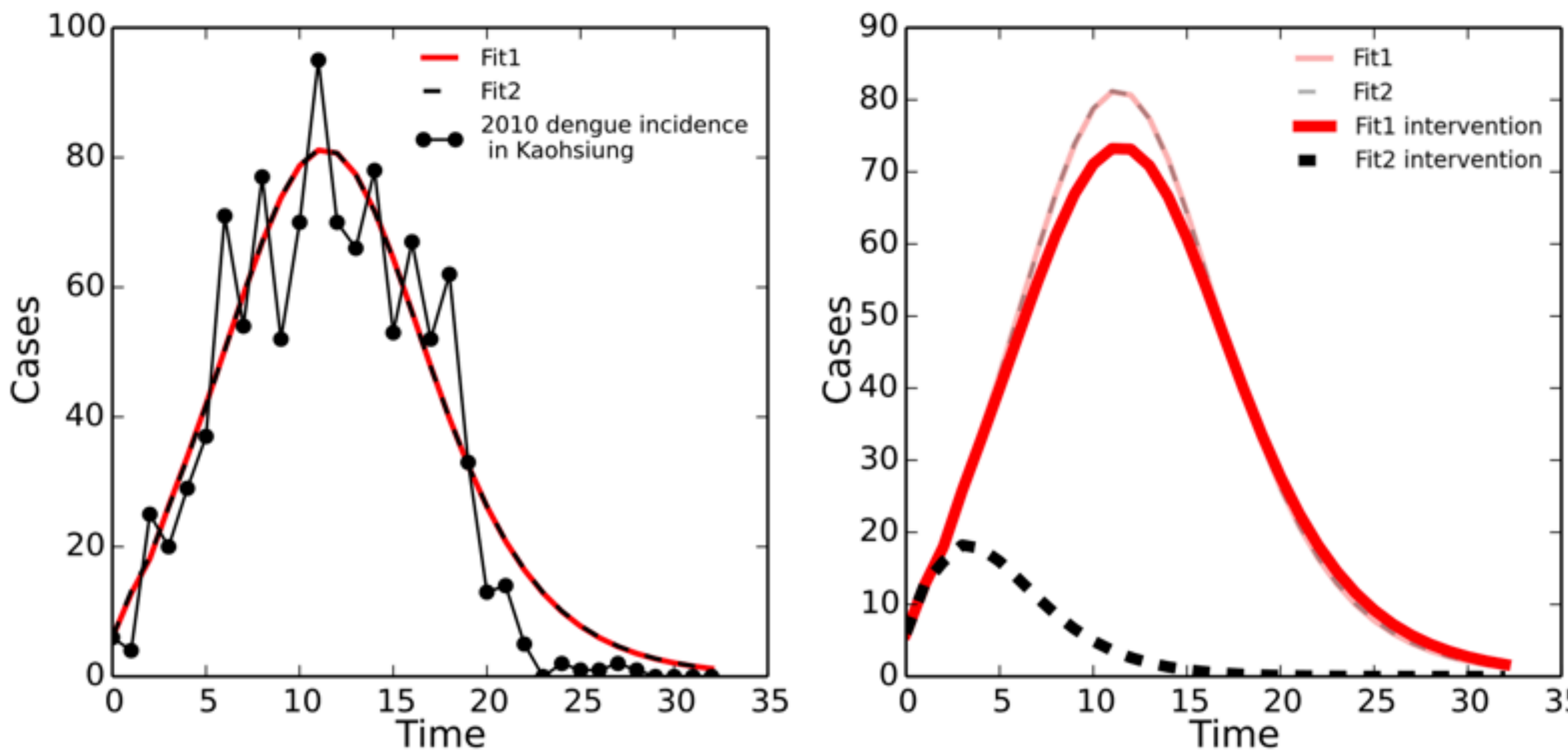


$$\mathcal{R}_0 = \sqrt{\frac{S_m \alpha \beta_{hm} \beta_{mh} \gamma}{(\alpha+\mu)(\eta+\mu)(\gamma+\mu_m)\mu_m}}.$$

Intervention predictions

Fit1:
$\beta_{mh} = 14.15$
$\xi = 2.03$
$\beta_{hm} = 0.03$
$\mu_a = 4.18$
$\mu_m = 0.32$
$\kappa_h =$
1546.74

Fit2:
$\beta_{mh} = 38.10$
$\xi = 0.13$
$\beta_{hm} = 0.02$
$\mu_a = 0.15$
$\mu_m = 0.42$
$\kappa_h =$
1625.42



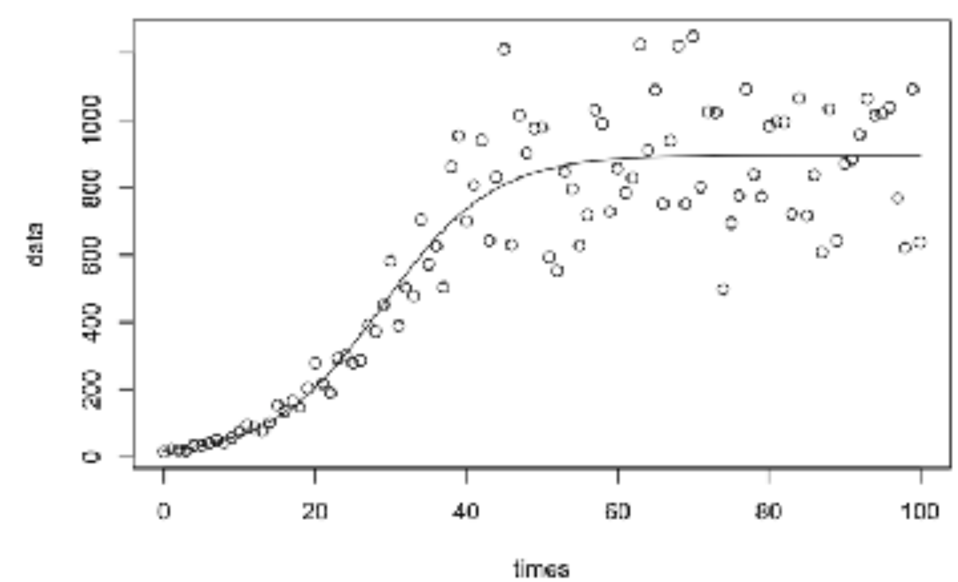
Kao & Eisenberg, *Epidemics*, 2018.

Sidenote: Identifiability in a Bayesian Context

- Unidentifiability can affect the performance of MCMC and other sampling methods, and can lead to broad, flat posteriors or heavy reliance on the prior

- Simple unidentifiable model example:

$$\frac{dS}{dt} = -\beta S I + \gamma I$$
$$\frac{dI}{dt} = \beta S I - \gamma I$$
$$y = kNI$$



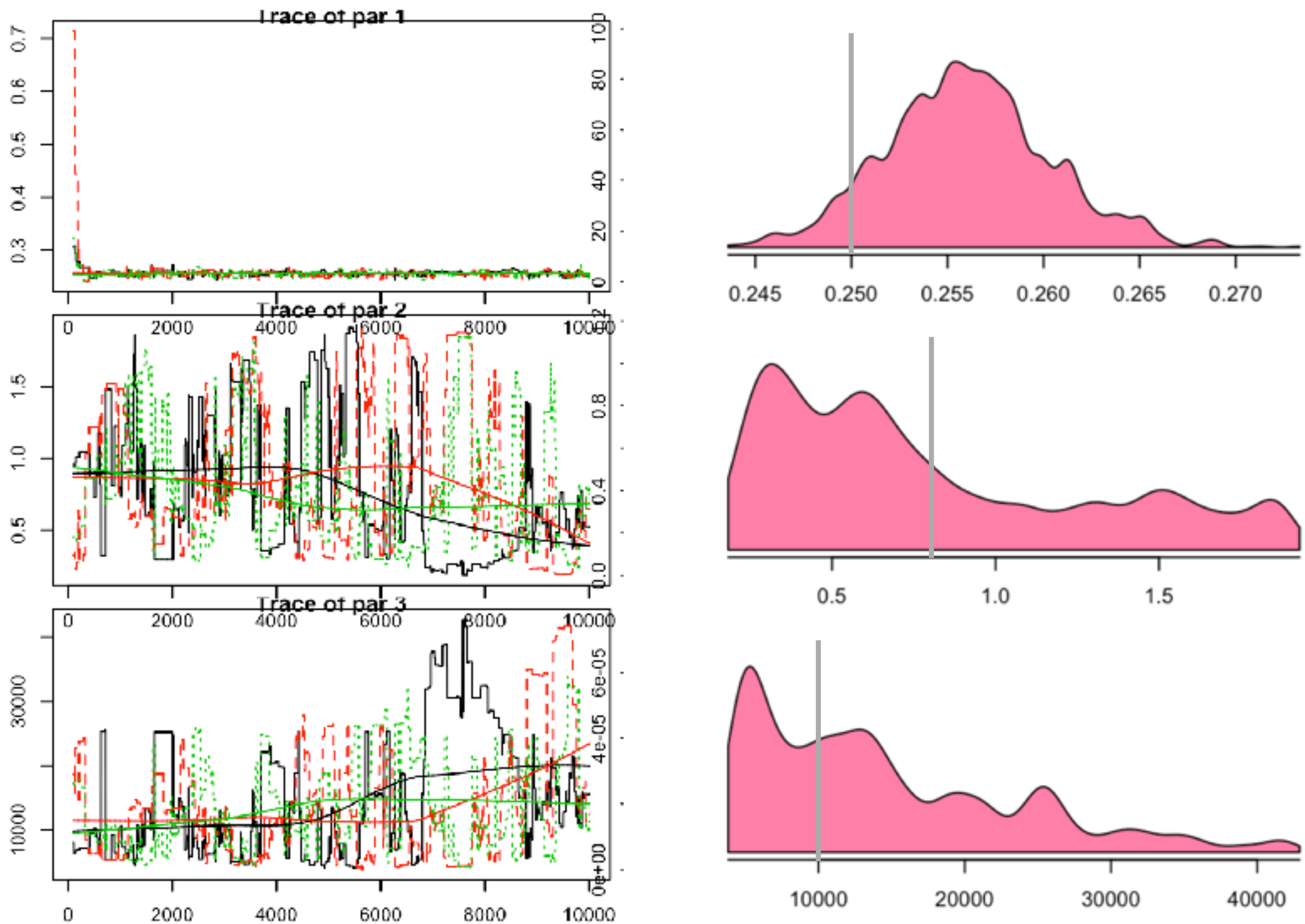
- Try MCMC (e.g. with Metropolis-Hastings or variants of)

Unidentifiable model

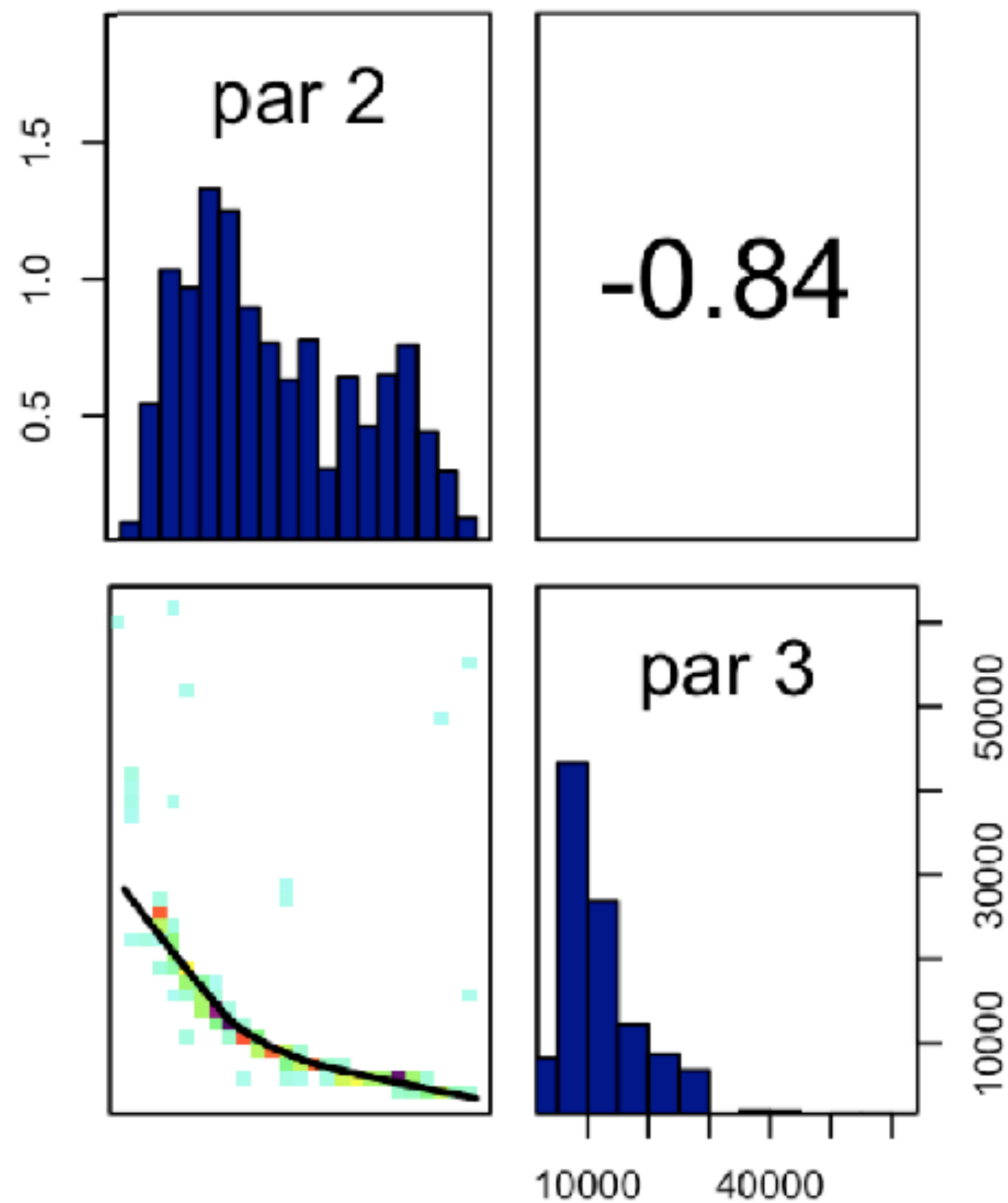
β

k

N



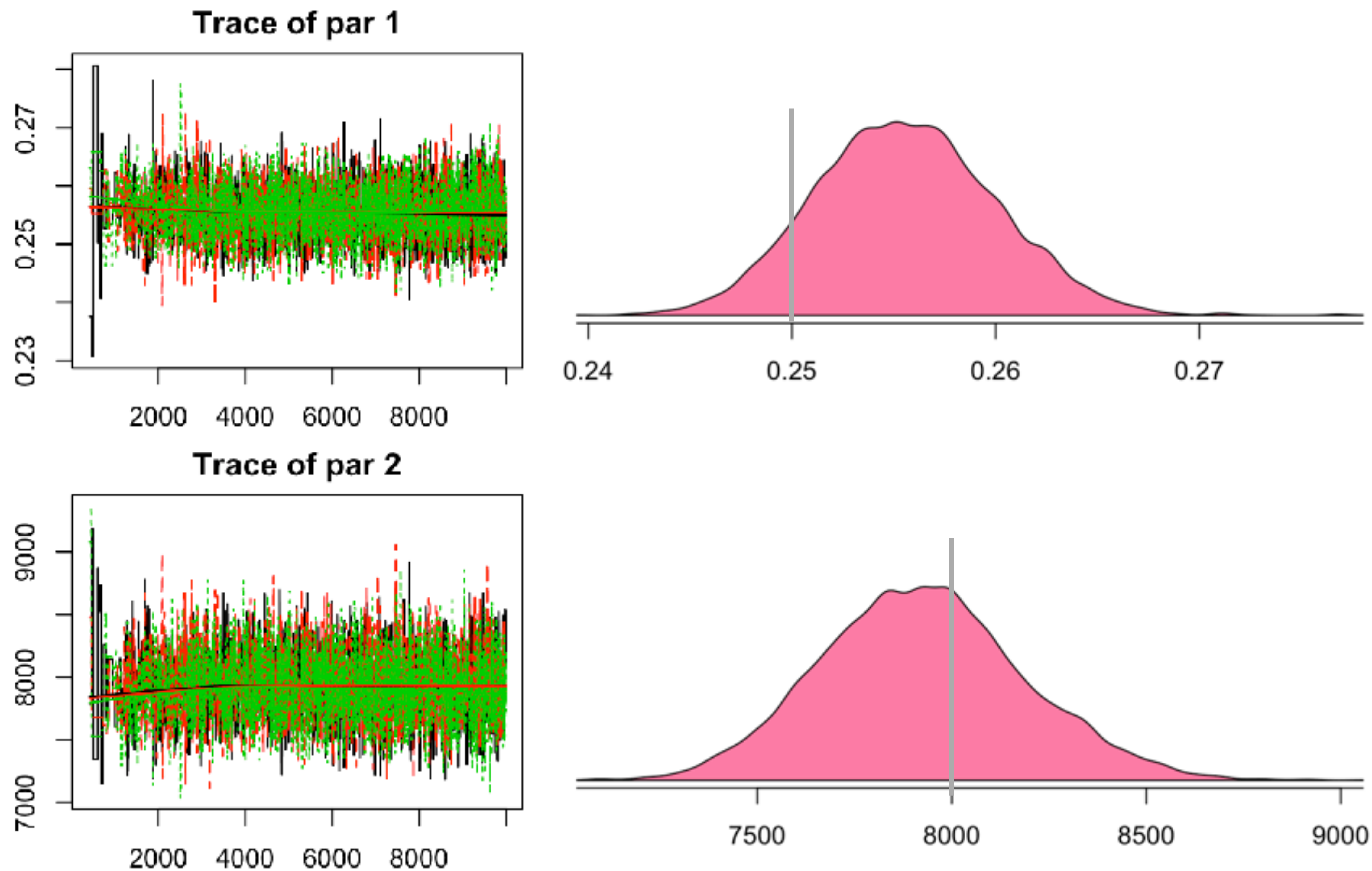
Correlation between k and N



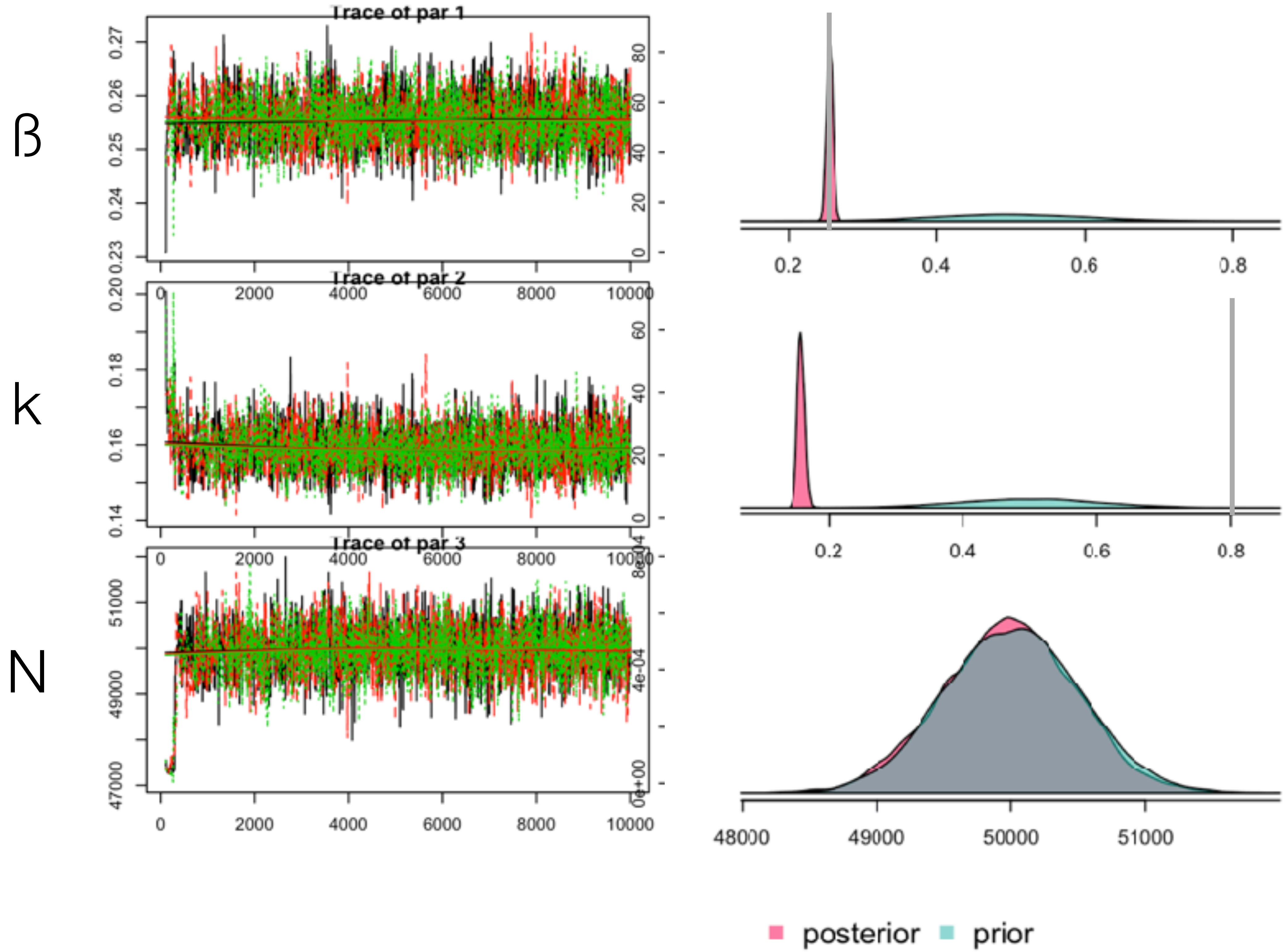
Reparameterize to make the model identifiable

β

kN



Adding a strong prior



Conclusions

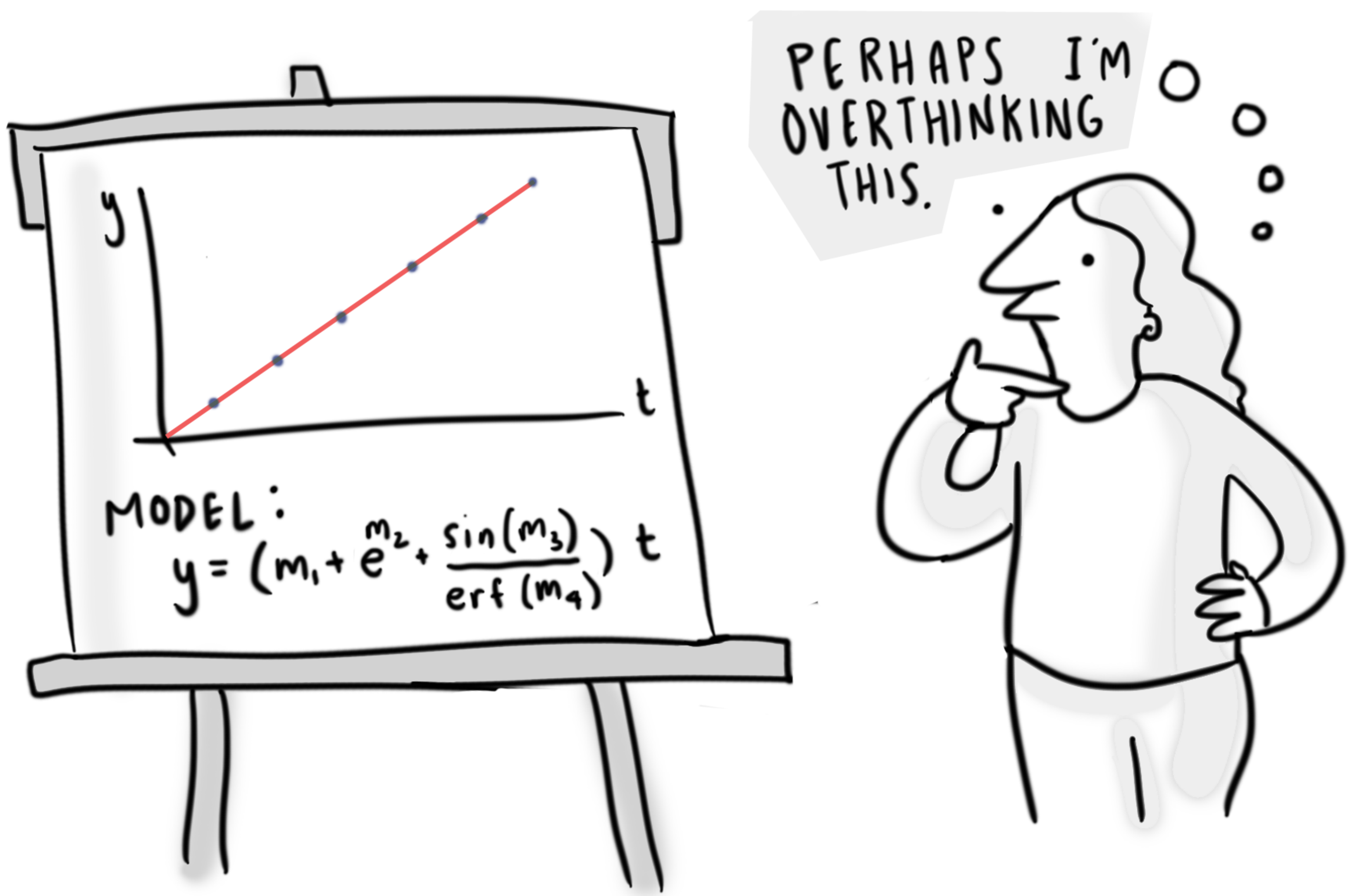
- Many related questions and potential issues when connecting models to data: observability, distinguishability & model selection, reparameterization & model/parameter reduction, and more
- Many other methods! (eigenvalues of FIM, sloppy models, active subspaces, Bayesian methods, & more)
- Depending on amount of data, model complexity, model type, and more, different approaches may work in different circumstances

Conclusions

- Identifiability—an important question to address when estimating model parameters
- Common problem in math bio (identifiability-robustness tradeoff)
- Many approaches, both numerical and analytical

End

Questions?



comic by Olivia Walch (UM):
http://imogenquest.net