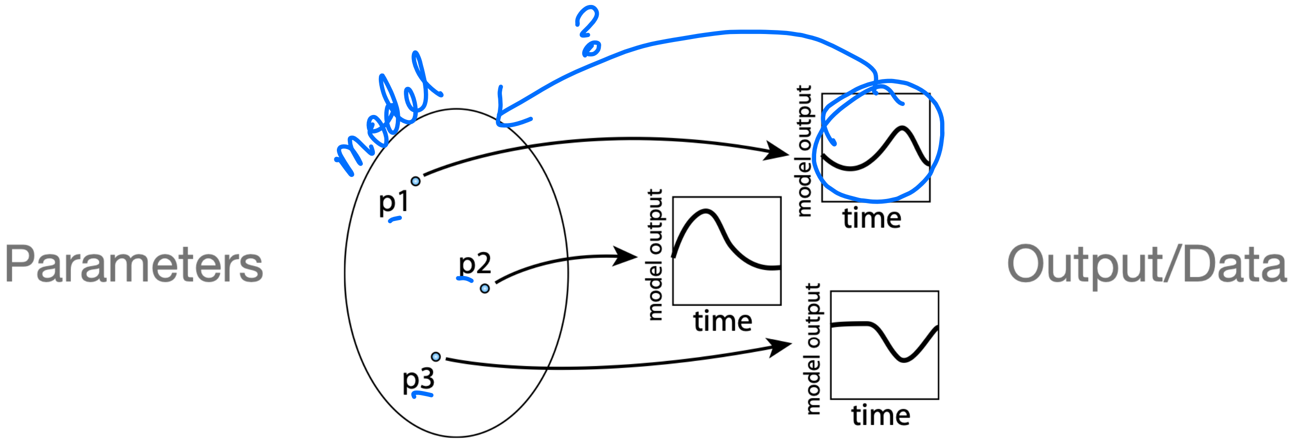
Introduction to Numerical Methods for Identifiability

MSRI Algebraic Geometry Summer School 2022

Jessica Conrad, MSPH



How can we apply numerical methods to identifiability?

Consider the following approach:

- Assume your model is structurally identifiable and choose a set of "true" parameters
  - Generate simulated output data from these "true" parameters
- Attempt to fit your simulated data using a range of parameter values and solve for the "best" parameter set to reproduce the simulated data
- IF your original parameter set is returned, your model may be identifiable
  - What does it mean if the simulated data is without noise?
  - With noise?

How can I check my model using statistics?

- Begin by reframing the question:
  - Given output data **z**, what parameter set **p** generates this?
  - Consider "Parameter estimation" instead of "identifiability"
    - How are these concepts related?
  - What parameter or distribution set *most likely* generated the data?

Why would I want to explore numerical methods?

- Most numerical methods can explore both structural and practical identifiability
- Wide range of applicability to different models
- Relatively fast implementation
- Typically restricted to local identifiability, but global methods exist too

Maximum Likelihood Method

- **Basic Idea:** Reframe compartmental model as a statistical model where we assume the general form of the density function, but not parameter values
- Then if we knew the parameters, we could frame a probability: $P(z \mid p)$

data parameters

Likelihood Function:

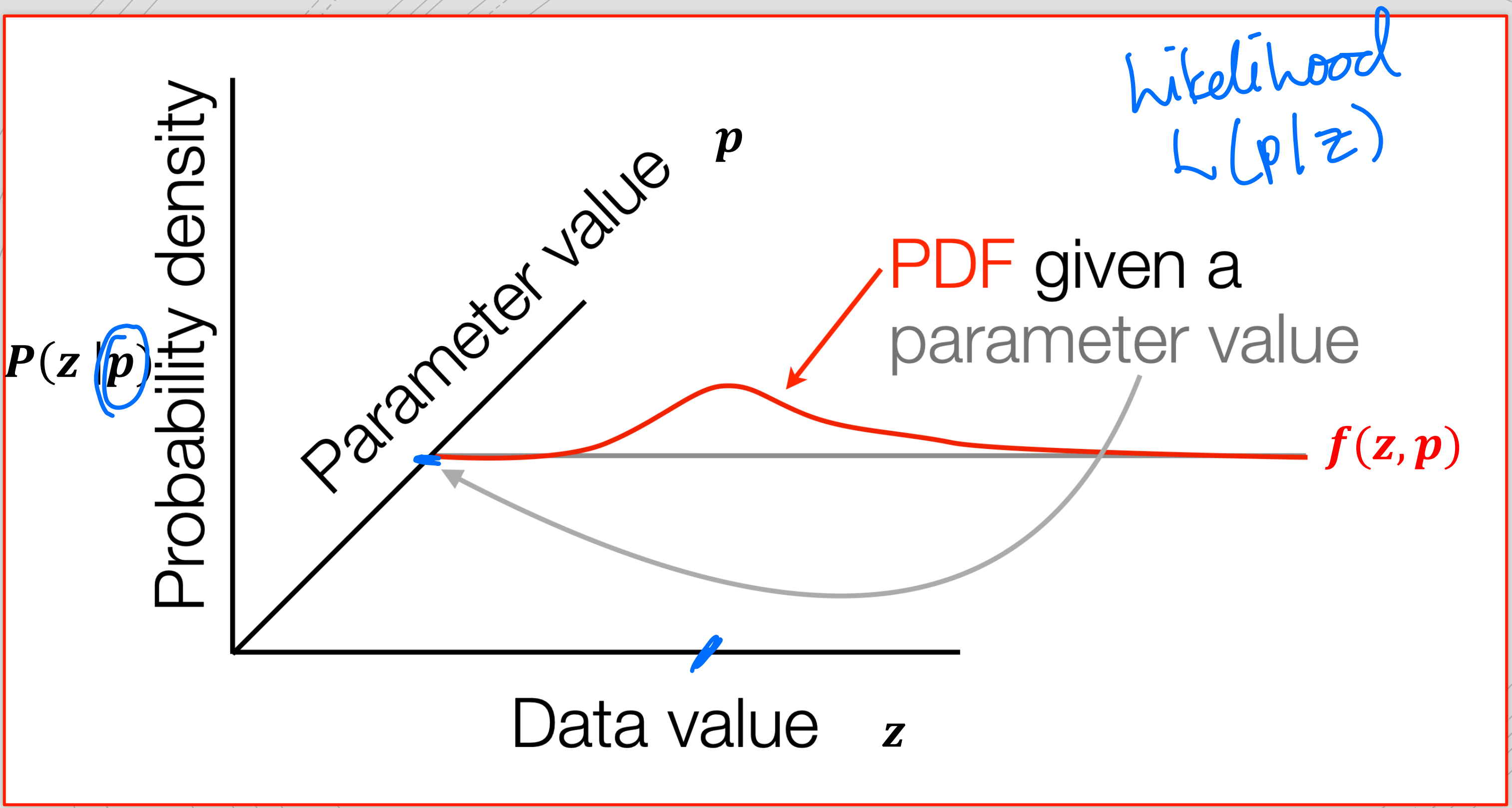
$$P(z \mid p) = f(z, p) = L(p \mid z)$$

- Now we rethink the probability distribution P, as a function of the data f
- We want to find the parameter set p that maximizes the likelihood L given data z

Example:

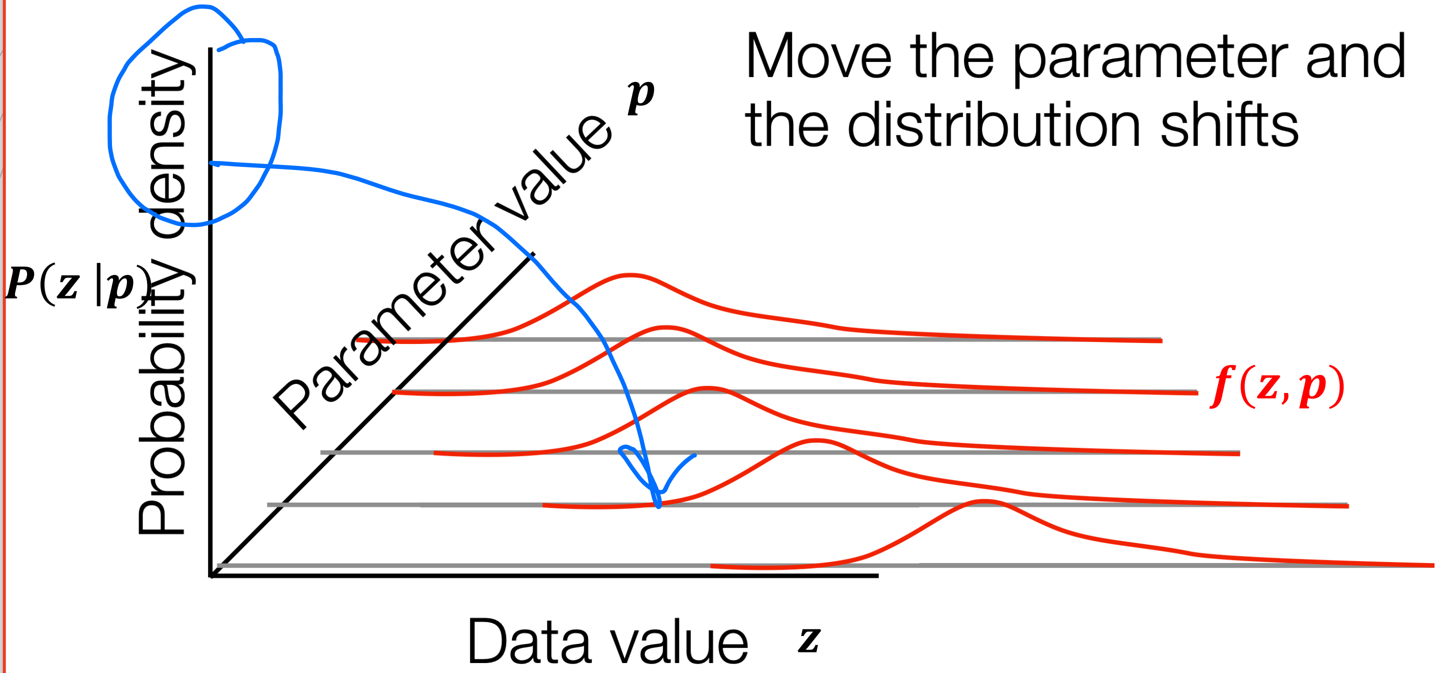
$$f(z \mid \mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left(-\frac{(z-\mu)^2}{2\sigma^2}\right) = L(\mu, \sigma^2 \mid z)$$

$P(z|p)$



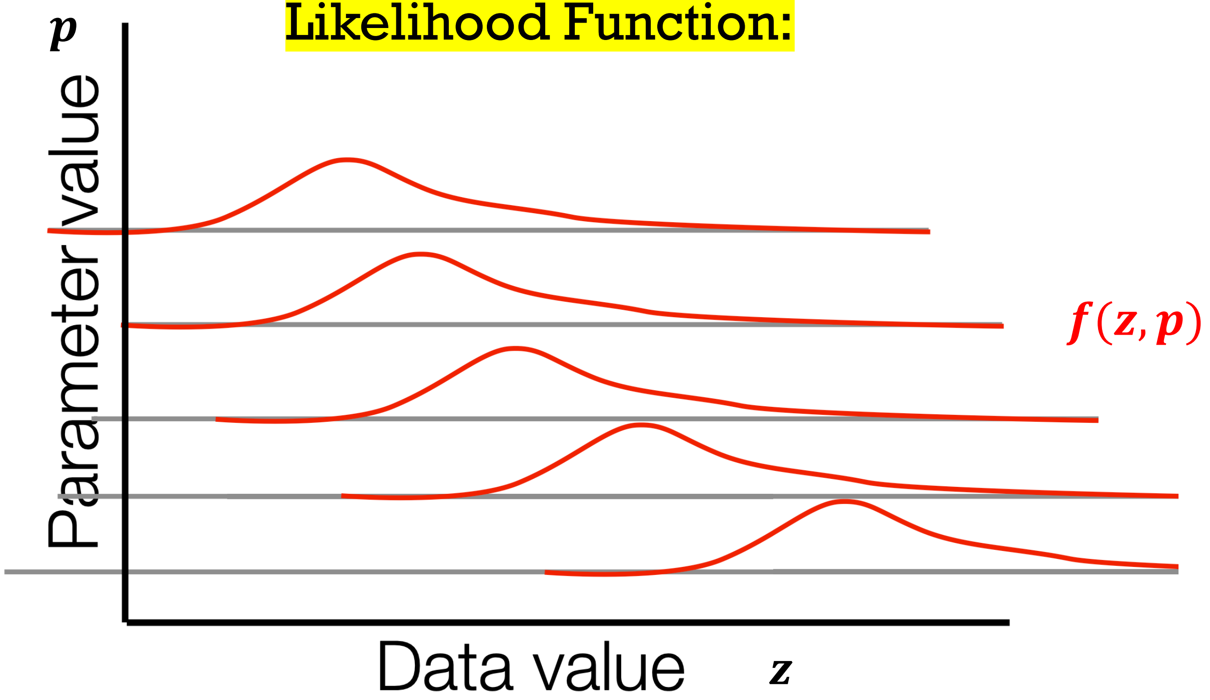
Original Slide from Dr. Marisa Eisenberg

Move the parameter and the distribution shifts



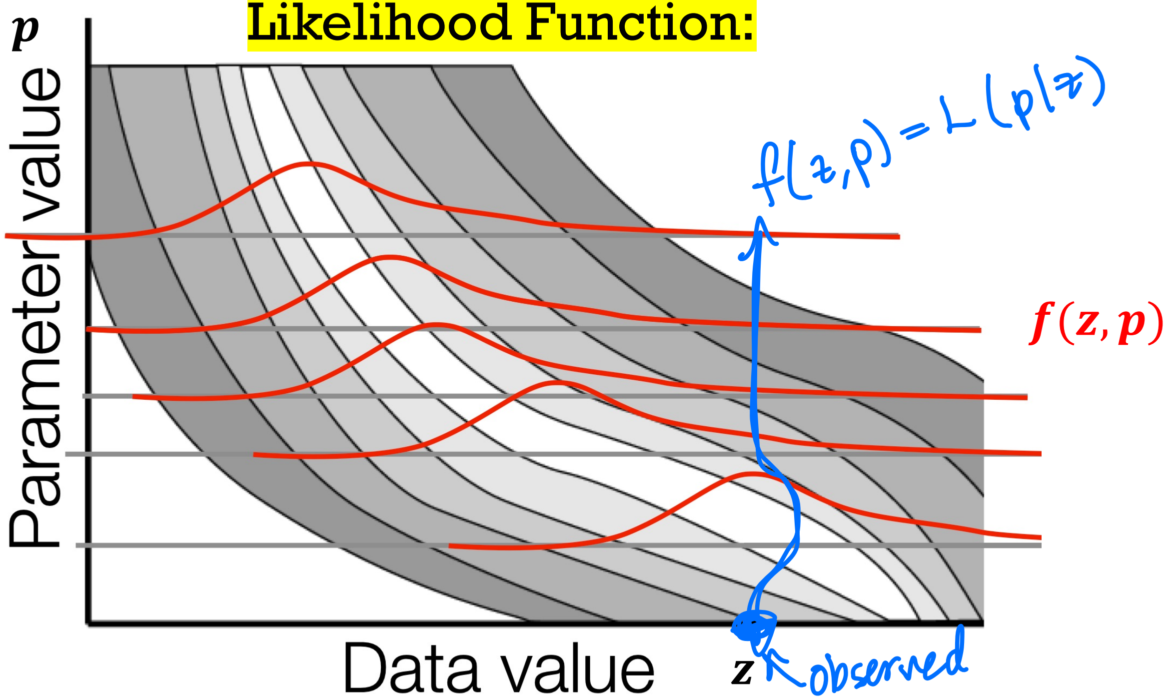
Original Slide from Dr. Marisa Eisenberg

Likelihood Function:



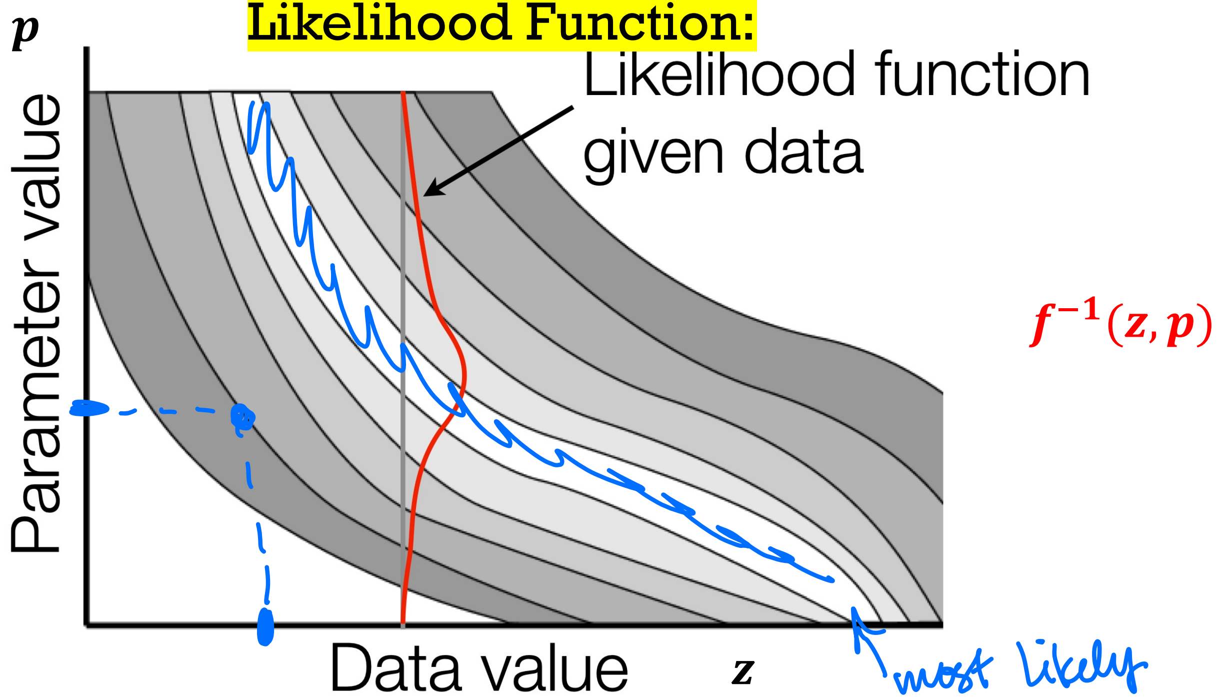
Original Slide from Dr. Marisa Eisenberg

Likelihood Function:



Original Slide from Dr. Marisa Eisenberg

Likelihood Function:



Original Slide from Dr. Marisa Eisenberg

Maximum Likelihood Method:

- **Consistency:** With sufficiently large number of observations, n, it is possible to find the value of p with arbitrary precision
- **Normality:** As n increases, the MLE tends to a Gaussian distribution with mean and covariance equal to the inverse of the Fisher information matrix
- **Efficiency**: Achieves Cramer-Rao bound as $n \to \infty$

Let's derive the likelihood function for the Gaussian example!

- We have an ODE model defined as:

$$\dot{x} = f(x, t; p)$$
$$y = g(x, t; p)$$

$\dot{x}$ - state var.
output

- Now sample data at times $t_1, t_2, t_3, \ldots, t_n$
  - Data at t_i is defined as $z_i = y(t_i) + e_i$ error
  - Assume error e_i is Gaussian and unbiased, with known variance σ^2 known constant
- View data z_i as a sample from **Gaussian distribution** with mean $y(x, t_i; p)$ and variance σ^2
  - We assume all measures are independent (Is this realistic?)

So let's calculate the likelihood function:

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

$z_i = y(t_i) + e_i$, $e_i \sim N(y, \sigma^2)$

Then our likelihood assuming independent observations:

$$L(y(t_i, p), \sigma^2 \mid z_1, z_2, \dots, z_n) = f(z_1, z_2, \dots z_n \mid y(t_i, p), \sigma^2)$$

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

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

μ, σ^2 Parameters | output data

maximizing likelihood

What does this look like in practice?

- Rather than maximizing the likelihood, in practice we **minimize the negative log likelihood**
  - Log is well behaved and minimization algorithms are common

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

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

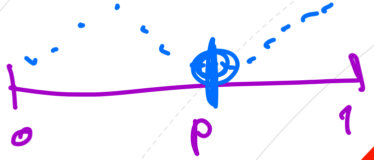
$$= \frac{n}{2}\ln(2\pi) + n\ln(\sigma) + \frac{\sum_{i=1}^{n}\left(z_i - y(t_i;p)\right)^2}{2\sigma^2}$$

constant constant

… Then our problem reduces to:

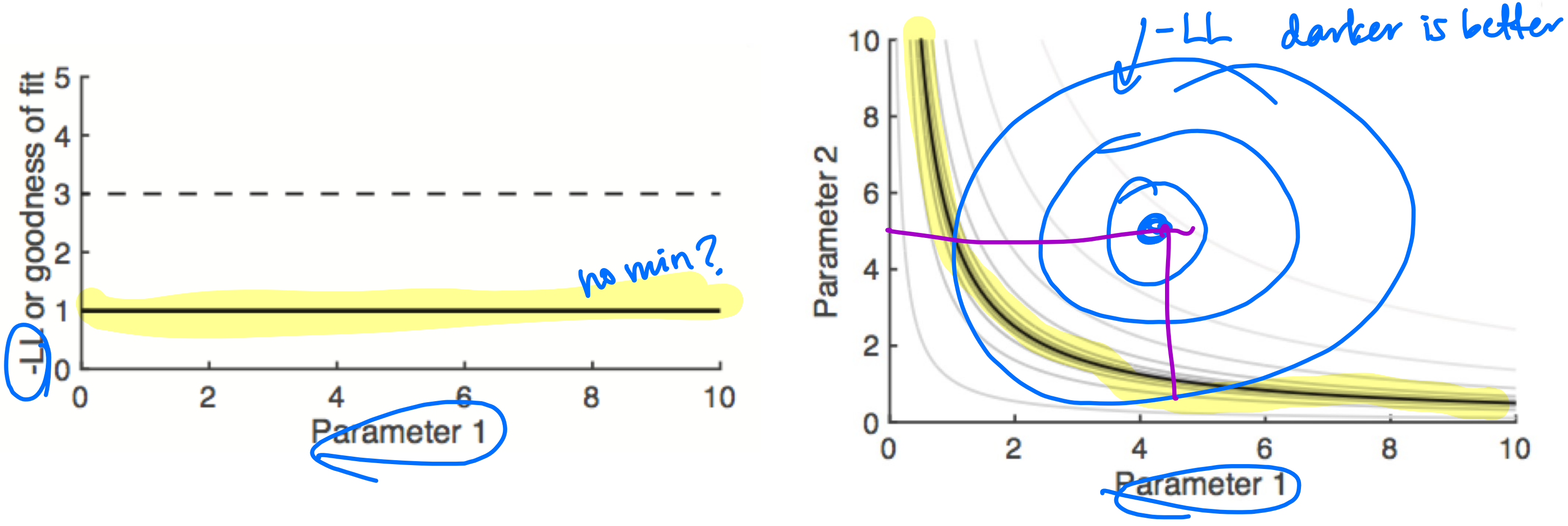
$$\min_p(-LL) = \min_p \sum_{i=1}^{n} \left(z_i - y(t_i;p)\right)^2$$

So what's the big idea?



- We will "profile" one parameter at a time using the likelihood function
  - Fix the parameter to a range of values and fitting all other parameters in the model
  - Calculate the likelihood value for every combination
- This will give the best fit at each point
- Plot the best likelihood values for each value of p_i

… This is known as the Profile Likelihood method!



What can we expect to see?

5 Minute Break!

☺

Time to Code!...