Q: What is a nonidentifiable model?

When we say 'nonidentifiable' we are referring to whether or not parameters in a given mathematical model are identifiable.

*A model is called sloppy if it has many unknown parameters.

Two classes of Parameter identifiability

① Structural: A model property. Refers to whether or not parameters can be uniquely determined in case where we have infinite noise-free data available.

Example

② Practical: A property of the model & a specific dataset. Refers to how well parameters can be estimated when we have a finite, noisy dataset.

Example.

Examples (We will go back to these once we develop some tools to solve the identification problem).

① Structurally non-identifiable model:

Production decay model

$$\frac{dx}{dt} = p - (k_1+k_2)x$$

↳ Modeling context:

$x(t)$: density of biological cells.

Cells are produced at a constant rate $p>0$ and cells undergo two types of linear decay at rates k_1 and k_2, respectively.

$\left\{ \begin{array}{l} k_1 \text{ could represent cell death due to apoptosis \&} \\ k_2 \text{ could represent cell death due to necrosis} \end{array} \right\}$

Exact solution can be computing using an integrating factor:

$$(*) \quad x(t) = \frac{p}{k_1+k_2} + \left(x_0 - \frac{p}{k_1+k_2}\right) e^{-(k_1+k_2)t}$$

Goal is to estimate $\theta = (k_1, k_2, p)^T$ from a particular set of finite noisy data $y^{obs}(t)$.

This production-decay model is structurally non-identifiable, since it's dynamics is controlled $\underline{k_1+k_2}$ & so there are infinitely-many choices of θ that give same values of (k_1+k_2, p).

↳ Side Note: There's software for structural identification for models ... DAISY is one.

② Practical non-identifiable model.

Spatio temporal invasion of cell population w/ reaction diffusion model in context of a scratch assay.

Modeling Context: Cells migrate + proliferate leading to closure of scratched region.

Two cell types we consider:

1) Cells in G1 phase that fluoresce red - r
2) Cells in S/G2/M phase that fluoresce green - g

Model to describe Scratch assay experiment:

$$\frac{\partial r}{\partial t} = D_r \frac{\partial^2 r}{\partial x^2} - k_r r + \underbrace{2k_g g(1-s)}_{\text{second transition produces 2 daughter cells in G1 phase}}$$

$$\frac{\partial g}{\partial t} = D_g \frac{\partial^2 g}{\partial x^2} + k_r r - \underbrace{k_g g(1-s)}_{\text{green} \to \text{red transition subject to contact inhibition}}$$

$r(x,t) \geq 0$: G1 phase cells non-dimensional density

$g(x,t) \geq 0$: G2 phase cells non-dimensional density.

$s(x,t) = r(x,t) + g(x,t)$: total non-dim. density

Parameters : $\Theta = (D_r, D_g, k_r, k_g)^T$

diffusivity of cells in G1 phase " " in G2

$k_r > 0$: rate at which cells transition from G1 → G2

$k_g > 0$: rate at which cells transition from G2 → G1

* Solved Numerically through 'Method of Lines'

Turns out that D_r & D_g are not well identified by the data here, while others are

Makes sense because the solution of model with reduced (or increased) D_r can match data reasonably well with D_g increased (or reduced) to compensate.

Therefore, model is practically non-identifiable.

Tools for Identifiability Analysis

① Likelihood function → Finding optimal θ

$(X_1, X_2, \dots, X_n)$ collection of random samples from probability distribution with density function $f(x; \theta)$

θ : unknown vector of $\underline{K}$ model parameters.

Given a realization $x = (x_1, x_2, \dots, x_n)$, the likelihood function is defined as:

$$L(\theta | x) = \prod_{i=1}^{n} f(x_i ; \theta)$$

how likely is it to observe an outcome where θ is variable

↳ Note: x will be our observed data : y^{obs}.

Log-likelihood function : (used for convenience)

$$\ell(\theta | x) = \log(L(\theta | x)) = \sum_{i=1}^{n} \log(f(x_i ; \theta))$$

* Note: The likelihood function we consider depends on noise model used to relate individual measurements to output of the model.

↳ Example: Additive Gaussian, f will be the Gaussian distribution.

(1a) Maximum Likelihood Estimation (MLE)

↳ Results in the 'best-fit' set of parameters with the following optimization problem that is solved with the Nelder-Mead algorithm with simple bound constraints,

(MLE) $\hat{\theta} = \arg\max_{\theta} \left[\ell(\theta \mid y^{obs}) \right]$

(1b) Normalized log likelihood function $\bar{\ell}$.

(Convenient b/c it sort of centers ℓ about max)

$$\bar{\ell}(\theta \mid y^{obs}) = \ell(\theta \mid y^{obs}) - \ell(\hat{\theta} \mid y^{obs})$$

So that:

$$\bar{\ell}(\hat{\theta} \mid y^{obs}) = 0.$$

→ Asymptotic Confidence sets. ↗ contains θ s.t likelihoods not sig. lower than max. at confidence level, according to chi^2

(2) We define asymptotic confidence sets for the parameter θ with a likelihood based threshold: approx valid for large samples.

Threshold: $\bar{\ell}^* = -\frac{\Delta_{q,n}}{2}$ } qth quantile of χ^2 distribution w/ n dof.

$\Rightarrow$ q % capture true time

Asym. conf. sets implicitly defined by values of θ satisfying $\bar{\ell}(\theta \mid y^{obs}) \geq \bar{\ell}^*$.

defines a 'region' where the true θ would likely contain

③ Profile-likelihood for parameter identifiability assessment. Determines if parameter is identifiable or not.

↳ Partition $\theta = (\psi, \omega)^T$

ψ: interest — either one – univariate profile, or two – bivariate profile (contour plot)

ω: nuisance

Profile likelihood function is: maximize over ω given ψ as a fixed value

$$\bar{\ell}_p(\psi \mid y^{obs}) = \underset{\omega \mid \psi}{\operatorname{argmax}}\ \bar{\ell}(\psi, \omega \mid y^{obs})$$

Implicitly defines function $\omega^*(\psi)$ of optimal values of ω for each value of ψ.

Evaluated on uniform discretization of ψ.

How to reparameterize model if likelihood assessment says a parameter is not identifiable?

Two Approaches.

- 'Intuitive' Approach - Reparameterize by 'guessing' based on model form
- Data-Based Approach

Tools for Data Informed Reparameterization

(4) Observed Fisher information matrix (FIM) at the MLE, $I(\hat{\theta})$
eigendecomposition

$$I(\hat{\theta}) = -\nabla\nabla^T \ell(\theta \mid y^{obs}) \Big|_{\theta=\hat{\theta}}$$

$\downarrow$

$k \times k$ square, symmetric positive semi-definite matrix.

$\lambda_k \rightarrow$ eigenvalues, $v_k \rightarrow$ eigenvectors.

↳ For structurally non-identifiable models, I is rank deficient & eigenvectors associated w/ zero eigenvalues span non-identifiable parameter space. Eigenvectors associated with non-zero eigenvalues span identifiable param space.

↳ For practically non-identifiable models, I has full rank.

(5) Eigenparameterization

Regardless of non-identifiability type, the eigendecomposition of FIM gives way to combine parameters resulting in a reduced identifiable model.

→ important so that α_k corresponds to an indpdt mode of information in likelihood. Decouples effects of parameters.

v_k's are mutually orthogonal, eigenparameters α_k can be written as a linear combination of original model parameters. (Note eigenvectors are normalized $(v_k)_j \in [-1, 1]$)

change of basis

$$\alpha_k = \sum_{j=1}^{k} (v_k)_j \, \theta_j$$

defines indpt directions of curvature. Strength of curv given by λ_k. v_p

For some problems it is useful to work with the log parameterization:

$$\log(\alpha_k) = \sum_{j=1}^{k} (v_k)_j \log(\theta_j).$$

5a Reparameterization can either be direct or approximate

↳ Direct: Apply relation between α and θ as stated.

↳ Approximate: Elements of V_K are close to 0. Smaller terms are omitted from summation
↳ Leads to reduced model involving both original & eigenparameters.

(6) Likelihood-based Prediction intervals.

Variability in y^{obs} → Var in θ → Var in prediction
↳ 2 types: Exact & Approximate.

(6a) Exact Prediction Intervals

→ Rejection sampling for θ using full $\bar{\ell}$
Obtain M samples θ_m $m = 1, 2, \ldots, M$
within 95% confidence region where

$$\bar{\ell}(\theta_m | y^{obs}) \geq \bar{\ell}^* = -\frac{\Delta_{0.95, k}}{2}$$

→ For each θ_m, model solved for $x(t)$ & width of distribution of noise model involved by computing 5% and 95% quantiles of associated noise model, denote these $x_{0.05}(t)$ & $x_{0.95}(t)$.

↳ With this prediction interval is $x(t) \in [x_{0.05}(t), x_{0.95}(t)]$

- Repeat for each θ_m, take union of resulting M prediction intervals to get exact prediction interval. Accurate for large enough M.

* Having a large number of unknowns, potentially correlated parameters, can make problem computationally infeasible to sample full loglikelihood.

(6b) Approximate Profile-Wise Prediction Intervals.

Construction is essentially the same with a few key distinctions

① Sample from univariate profile

likelihood function

$\bar{l}_p(\psi_k \mid y^{obs})$ ψ_k is kth scalar interest param $k = 1, 2, \ldots, K$

With rejection sampling, they find M samples of $\psi_{k,m}$ & $\omega_{k,m}$ s.t.

$$\bar{l}((\psi_{k,m}, \omega_{k,m}) \mid y^{obs}) \geq \bar{l}^* = -\frac{\Delta_{0.95,k}}{2}$$

for $k = 1, 2, \ldots, K$

Construct prediction interval $x(t) \in [x_{0.05}(t), x_{0.95}(t)]$ for M param. combinations.

2. Process must be repeated for the $K-1$ remaining univariate profile LL function.

3. Take union over $M \times K$ approximate prediction intervals

4. Due to lack of correlation structure in log likelihood, we can use smaller M

Back to Example 1:

Model: $\frac{dx}{dt} = p - (k_1 + k_2)x$, $x(t) \approx 0$.

Recall: Exact solution is:

$$x(t) = \frac{p}{k_1+k_2} + \left(x(0) - \frac{p}{k_1+k_2}\right) e^{-(k_1+k_2)t}$$

Want to Estimate $\theta = (k_1, k_2, p)^T$ from a finite, noisy data set $y^{obs}(t)$.

Form of model + solution makes model structurally non-identifiable, as said before and is clear, there are many choices of θ that result in same values for (k_1+k_2, p).

Try generate noisy data w/ $x(0) = 100$ & true parameter values $\theta = (0.1, 0.1, 1.0)^T$.
y^{obs} constructed by taking 11 equally spaced values of $x(t)$ across $0 \le t \le 20$.

& each value is corrupted w/ additive Gaussian noise w/ $\sigma = 3$. We write this as

$$y^{obs}(t) \mid \theta \underset{\text{distributed}}{\sim} \underset{\text{normally}}{N}(x(t), \sigma^2)$$

$\uparrow$ Mean of noise model is solution of ODE.

Therefore log-likelihood function is

$$\ell(\theta \mid y^{obs}) = \sum_{i=1}^{I} \log\left[\phi\left(y^{obs}(t_i) ; x(t_i), \sigma^2\right)\right]$$

Where

$$\phi\left(y^{obs}(t_i) ; x(t_i), \sigma^2\right) = \underbrace{\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(y^{obs}(t_i) - x(t_i))^2}{2\sigma^2}}}_{\text{Gaussian pdf.}}$$

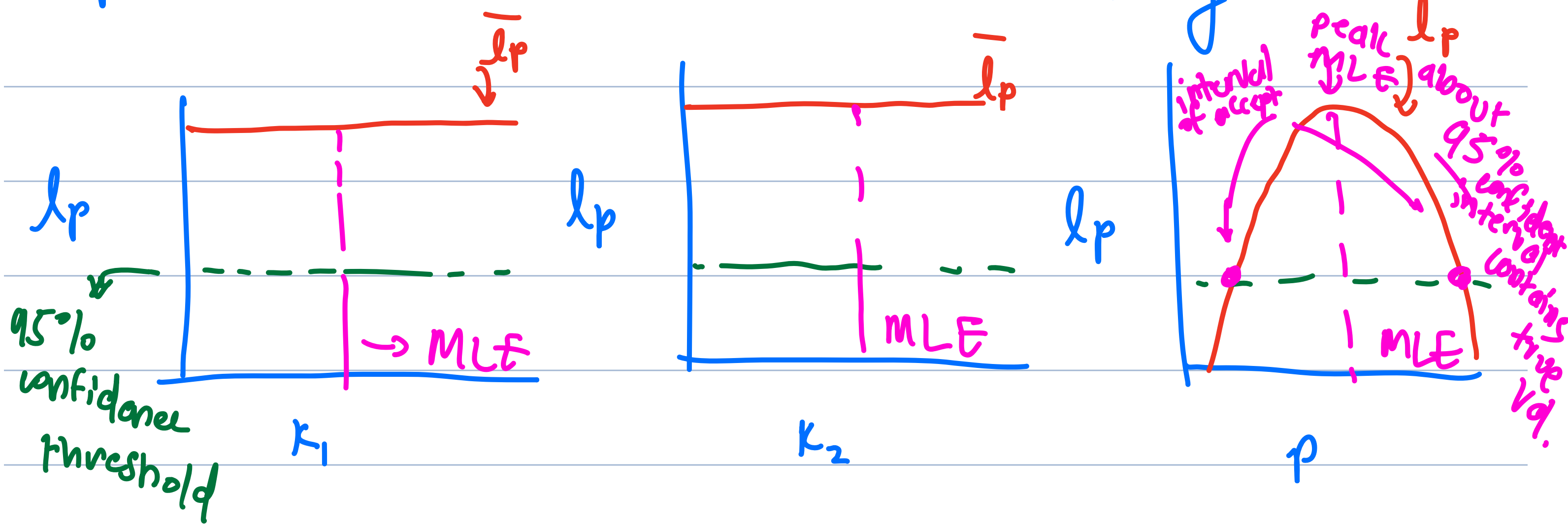
Downside of using Gaussian noise $\rightarrow$ can lead to non-physical negative densities if p is small enough or k_1/k_2 is large enough.

Finding the MLE through solving the optimization problem results in (θ that maximizes ℓ)

$$\hat{\theta} = (\hat{k}_1, \hat{k}_2, \hat{p}) = (0.105, 0.106, 0.966)^T$$

↳ Model w/ $\hat{\theta}$ compares well to noisy data.

Are the parameters identifiable? Compute univariate profile likelihood functions for each parameter to determine the following:



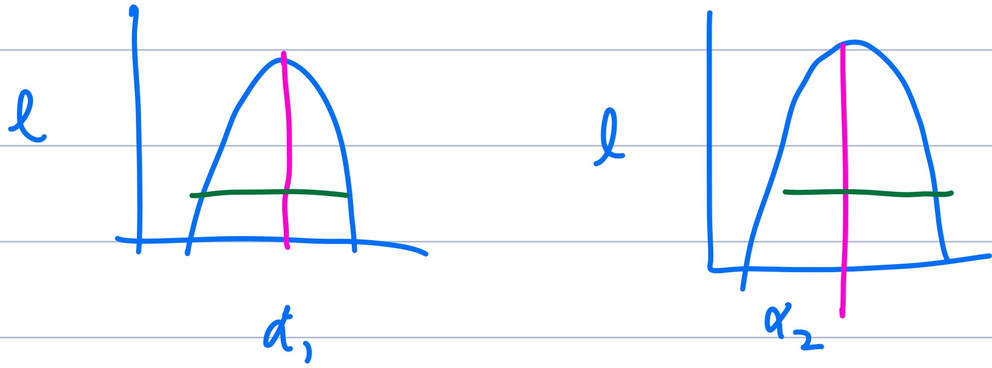
These plots indicate k_1, k_2 are not identifiable no well defined peak for any k_1, k_2.

↳ Reparameterization by Intuition

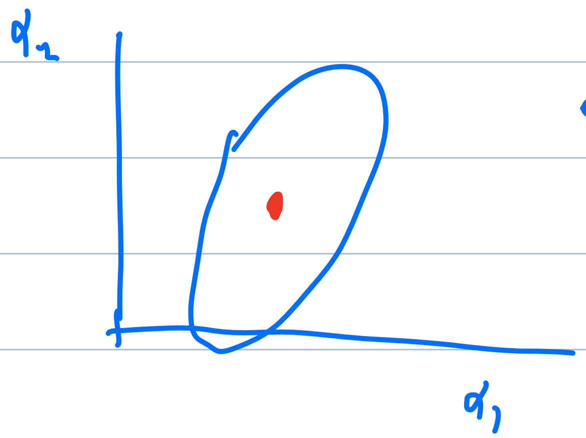
$(q_1, q_2)^T = (k_1 + k_2, p)^T$

↳ Model evaluation with the MLE defined by these parameters compares well with the data. Log-likelihood function has a single well-defined peak at MLE $\hat{\theta} = (\hat{q}_1, \hat{q}_2)^T = (0.210, 0.966)^T$

Both univariate profiles contain a well-defined peak about the MLE



Bivariate profile. - Normalized loglikelihood function with MLE as red dot where $\bar{\ell}=0$
95% threshold is curve.



plot indicates there is a positive correlation between α_1 & α_2.

The correlation makes intuitive sense. Since it would be possible to match the same set of data by increasing (or decreasing) the total decay rate α_1 while simultaneously increasing (or decreasing) the production rate α_2.

Now we want to compute the prediction interval, which quantifies how variability in measurements translates to variability in predictions.

We use rejection sampling to obtain $M = 1000$ samples of Θ in region where $\overline{\ell} \geq \overline{\ell^*}$ and construct our prediction interval for the model solution.

As it turns out, the correlation structure in reparameterization makes sampling within confidence set complicated. i.e. Sampling does not learn the shape of the set.

↳ Reparameterization by Data.

We consider a reparameterization of the full $\Theta = (k_1, k_2, p)^T$ by considering eigendecomp. of FIM evaluated at MLE.

↳ Computing FIM allows us to see w/ 3 parameters the 3×3 FIM has rank 2 with one zero eigenvalue.

To see why this is, based on the form of our loglikelihood function:

$$\ell(\theta \mid y^{obs}) = \sum_{i=1}^{I} \left[\log\left(\frac{1}{\sigma\sqrt{2\pi}}\right) - \frac{1}{2\sigma^2}\left(y(t_i) - \frac{p}{k_1+k_2} - \left(x(0) - \frac{p}{k_1+k_2}\right) \cdot \right.\right.$$

$$\left.\left. \cdots \exp\left(-t_i\,(k_1+k_2)\right)\right)^2 \right]$$

that $\frac{\partial \ell}{\partial k_1} = \frac{\partial \ell}{\partial k_2}$

Therefore

$$I = \begin{bmatrix} \frac{\partial^2 \ell}{\partial k_1^2} & \frac{\partial^2 \ell}{\partial k_1^2} & \frac{\partial^2 \ell}{\partial k_1 \partial p} \\ \frac{\partial^2 \ell}{\partial k_1^2} & \frac{\partial^2 \ell}{\partial k_1^2} & \frac{\partial^2 \ell}{\partial k_1 \partial p} \\ \frac{\partial^2 \ell}{\partial p\, \partial k_1} & \frac{\partial^2 \ell}{\partial p\, \partial k_1} & \frac{\partial^2 \ell}{\partial p^2} \end{bmatrix}$$

these are the same

Therefore, eigenvector corresponding to 0 eigenvalue spans non-identifiable param space.

The 2 eigenvectors associated with non-zero eigenvalues span the identifiable parameter space & define a linear relationship between eigenparameters &

model parameters for our data:

$$\begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} = \begin{bmatrix} 0.013409 & 0.013409 & 0.999820 \\ -0.706980 & -0.706980 & 0.018963 \end{bmatrix} \begin{bmatrix} k_1 \\ k_2 \\ p \end{bmatrix}$$

Working with $(\alpha_1, \alpha_2)^T$ leads to a reduced + identifiable model

Solution & model evaluated at the MLE, $\hat{\theta} = (\hat{\alpha}_1, \hat{\alpha}_2)^T$ fits data well.

- The loglikelihood function parameterized in terms of the eigenparameters is characterized by a single well-defined peak about MLE.

Both of corresponding univariate profiles have a single well-defined peak about MLE.

Note that estimates of $(\hat{\alpha}_1, \alpha_2)^T$ can be reinterpreted through linear equation above.

The benefit of working with the eigenparameters clear when 95% confidence threshold $\bar{\ell}^* = -\frac{\Delta_{0.95,2}}{2}$ is superimposed

on bivariate profile likelihood contour plot which confirms there is a lack of correlation between α_1 and α_2 in orthogonal reparameterization.

The benefit of eigen-reparameterization are:

① Lack of correlation between α_1 and α_2 means mapping how variability in θ impacts variability in predictions $x(t)$ is simpler through rejection sampling.

② Exact likelihood based prediction interval is accurately approximated by sampling from $k=2$ univariate profile likelihood functions & taking union of profile-wise prediction intervals to give approx. prediction interval.

Such accuracy w/ parameter-wise predictions not always possible when likelihood is correlated.