Nonlinear Modeling: Quality of parameter estimates

Introduction to Statistical Modelling

Prof. Joris Vankerschaver

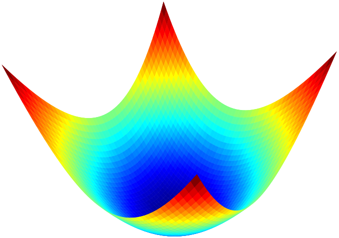
Learning outcomes

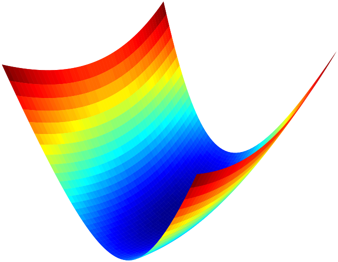
You should be able to

- Understand the interpretation of measurement noise
- Explain the role of the Fisher information matrix in quantifying parameter uncertainty
- Compute a confidence interval for a parameter
- Compute the correlation between two parameters

Quality of estimation

- Apart from obtaining parameter estimates, we want to know a measure of uncertainty for these values.
- Main idea: use objective function $J(\theta)$ to quantify uncertainty.
  - High curvature: low uncertainty (parameters well determined)
  - Low curvature: high uncertainty (not well determined)





Synthetic data

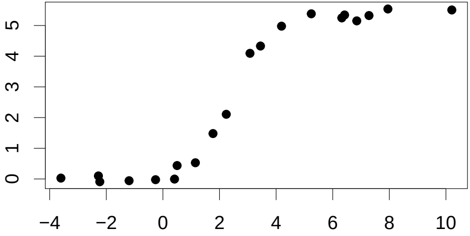
Model: logistic curve

$$y = \frac{A}{1 + \exp(k(x_{\text{mid}} - x))} + \epsilon,$$

where $\epsilon \sim N(0, \sigma^2)$.

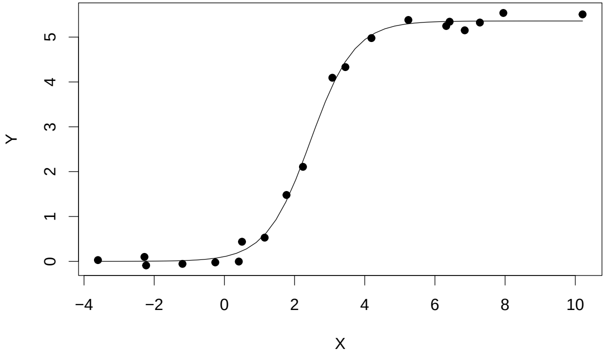
- Parameters: $A = 5.6$, $k = 1.4$, $x_{\text{mid}} = 2.5$.
- Measurement noise: $\sigma^2 = 0.2$ (the measure).

We take $n = 20$ data points from this model:



Model fit

- From now on, we "forget" the true parameters, and we will work with the data only.
- Nonlinear least squares: $A = 5.359$, $k = 1.597$, $x_{\text{mid}} = 2.500$.

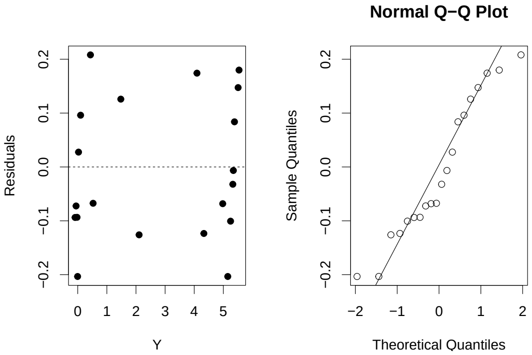


Measurement variance

- Typically measurement variance is not known.
- If model well-fitted: estimate from residuals:

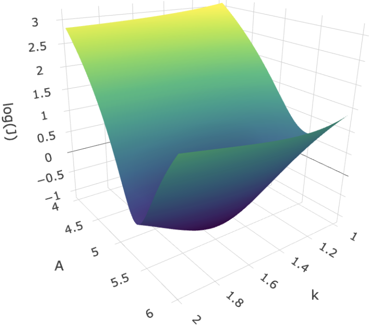
$$\sigma^2 \approx \frac{J(\theta_{\text{best}})}{N-p}.$$

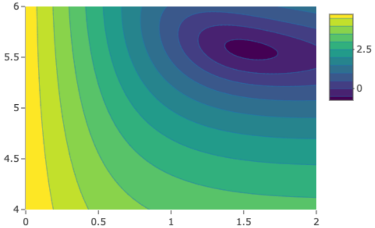
- Here $\sigma^2 \approx 0.523/17 = 0.031$ (true value $\sigma = 0.2^2 = 0.04$)



The loss surface

- Surface obtained by plotting $J(\theta)$ for all θ in some range.
- Optimal parameters are minima on this surface.
- When more than 2 parameters: focus on subset of parameters.
- **For visualization only.** (Higher dimensions: calculus)





Exact confidence region

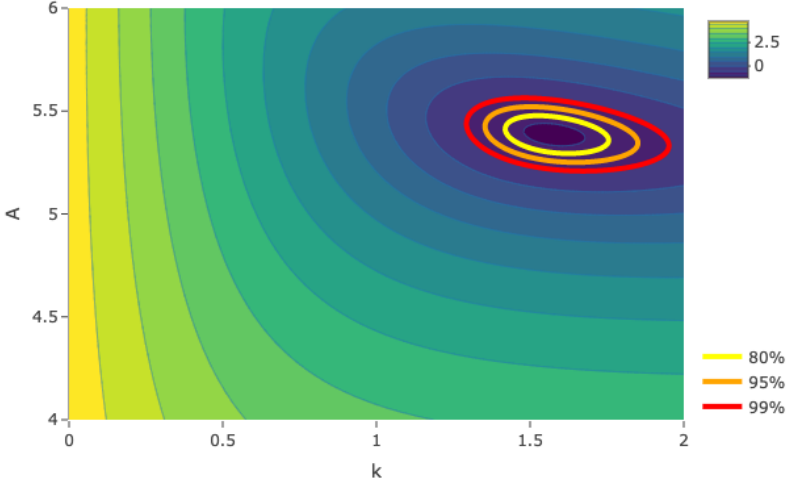
Confidence region: all θ such that

$$J(\theta) \leq \left(1 + \frac{p}{N-p} F_{p,N-p,1-\alpha}\right) \times J(\theta_{\text{best}}),$$

where $F_{p,N-p,1-\alpha}$ is quantile from F-distribution, α is significance level.

- Reasonably exact for models that are not too nonlinear.
- Easy to calculate numerically
- **Hard to describe or use explicitly**

Exact confidence region



Approximate confidence region

Taylor expansion to second order:

$$J(\theta) \approx J(\theta_{\mathsf{best}}) + \sum_{i=1}^{N} \underbrace{\frac{\partial J}{\partial \theta_i}(\theta_{\mathsf{best}})}_{=0}(\theta - \theta_{\mathsf{best}})_i + \frac{1}{2}\sum_{i,j=1}^{N} \frac{\partial^2 J}{\partial \theta_i \partial \theta_j}(\theta - \theta_{\mathsf{best}})_i(\theta - \theta_{\mathsf{best}})_j.$$

Confidence region becomes

$$(\theta - \theta_{\mathsf{best}})^T \mathcal{J} (\theta - \theta_{\mathsf{best}}) \leq p F_{p,N-p,1-\alpha}.$$

with $\mathcal{J}$ the **Fisher Information Matrix (FIM)**:

$$\mathcal{J}_{ij} = \frac{1}{\sigma^2}\sum_{k=1}^{N}\left(\frac{\partial y}{\partial \theta_i}(x_k, \theta_{\mathsf{best}})\frac{\partial y}{\partial \theta_j}(x_k, \theta_{\mathsf{best}})\right).$$

Interpretation of the FIM

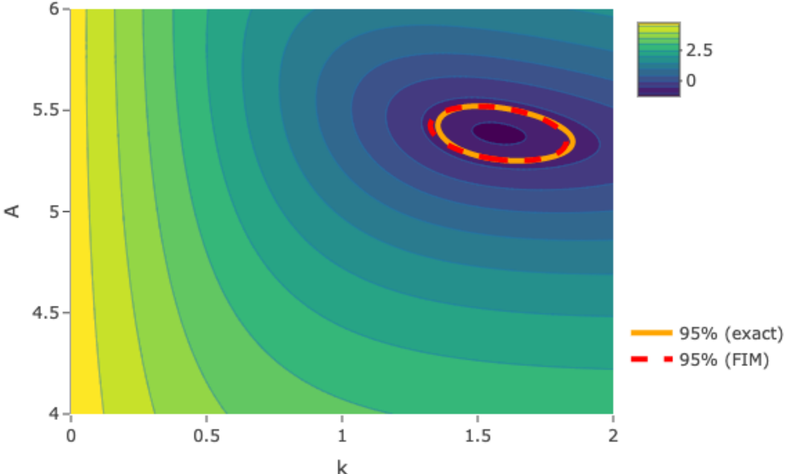
- The FIM tells us how much information the data give us about the model parameters.
- Alternatively, the FIM contains two ingredients:
  - The **sensitivity functions**, given by

    $$s_i(x, \theta) = \frac{\partial y}{\partial \theta_i}.$$

    Variables that are sensitive to perturbations in a parameter contain a lot of information about that parameter, and will contribute a lot to the FIM (and vice versa).
  - The **measurement noise** σ^2. Measurements with lots of noise contain less information about the parameters.

Approximate confidence region

- Level sets of quadratic approximation are ellipsoids.
- Good approximation to exact confidence region close to optimum.



Variance/covariance of parameters

- Often, we want to know variance of individual parameters and covariance between parameters.
- Encoded in the error covariance matrix:

$$C = \begin{bmatrix} \sigma^2_{\theta_1} & \mathrm{cov}(\theta_1, \theta_2) & \cdots & \mathrm{cov}(\theta_1, \theta_p) \\ \mathrm{cov}(\theta_2, \theta_1) & \sigma^2_{\theta_2} & \cdots & \mathrm{cov}(\theta_2, \theta_p) \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{cov}(\theta_p, \theta_1) & \mathrm{cov}(\theta_p, \theta_2) & \cdots & \sigma^2_{\theta_p} \end{bmatrix}.$$

- Diagonal: variances, off-diagonal: covariances
- Can be used to construct correlations between between parameters:

$$R_{ij} = \frac{\mathrm{cov}(\theta_i, \theta_j)}{\sigma_{\theta_i} \sigma_{\theta_j}}.$$

Computing the error covariance matrix

- The inverse of the FIM $\mathcal{J}$ is a lower bound for C:

$$C \geq \mathcal{J}^{-1}.$$

- **This is not an obvious result.**
- In practice, we just take $\mathcal{J}^{-1}$ as an estimate for C.
- Approximate confidence interval for parameter θ_i:

$$(\theta_{\mathsf{best}})_i \pm t_{N-p,1-\alpha/2}\sqrt{C_{ii}}.$$

Worked-out example: logistic model

To compute the FIM:

- Measurement noise: $\sigma^2 \approx 0.031$ (see earlier).
- Sensitivity functions:

$$\frac{\partial y}{\partial A} = \frac{1}{1 + \exp(k(x_{\mathsf{mid}} - x))}, \quad \frac{\partial y}{\partial k} = \ldots, \quad \frac{\partial y}{\partial x_{\mathsf{mid}}} = \ldots$$

  Often these functions have to be computed **numerically** (see next chapter).

Logistic model with synthetic data

- The inverse of the FIM is given by

$$\mathcal{J}^{-1} = \begin{bmatrix} 0.0057 & -0.0034 & 0.0020 \\ -0.0034 & 0.0169 & -0.0015 \\ 0.0020 & -0.0015 & 0.0040 \end{bmatrix}.$$

- Parameter estimates: $A = 5.359$, $k = 1.597$, $x_{\text{mid}} = 2.500$.
- 95% confidence intervals (low, high):

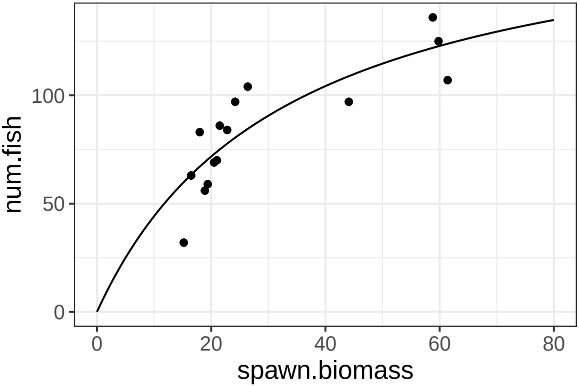
Parameter	Estimate	Low	High
A	5.36	5.20	5.52
k	1.60	1.32	1.87
x_{mid}	2.45	2.32	2.58

- Correlation between A and k:
$R = -0.0034/\sqrt{0.057 \times 0.0169} = -0.110.$

Worked-out example: stock-recruitment model

Optimal parameters:

- $\alpha = 5.75$
- $k = 33.16$



1. Measurement noise:

$$\sigma^2 = \frac{J(\theta_{\text{best}})}{N-p} = \frac{2809.01}{13} = 216.08$$

2. Sensitivity functions (for Beverton-Holt model):

$$\frac{\partial f}{\partial \alpha} = \frac{S}{1+S/k}, \quad \frac{\partial f}{\partial k} = -\frac{\alpha S^2}{(k+S)^2}.$$

Again, typically you would compute these derivatives numerically.

3. Fisher information matrix:

$$\mathcal{J} = \begin{bmatrix} 16.10 & -1.41 \\ -1.41 & 0.14 \end{bmatrix}$$

4. Error-covariance matrix:

$$C = \mathcal{J}^{-1} = \begin{bmatrix} 1.07 & 11.43 \\ 11.43 & 130.11 \end{bmatrix}$$

From previous slide:

$$\sigma_\alpha^2 = 1.07, \quad \sigma_k^2 = 130.11, \quad \mathsf{Cov}(\alpha, k) = 11.43.$$

5. 95% confidence intervals:

- For α:

$$\alpha_{\mathsf{best}} \pm 2.16 \times \sigma_\alpha = [3.53, 7.99]$$

- For k:

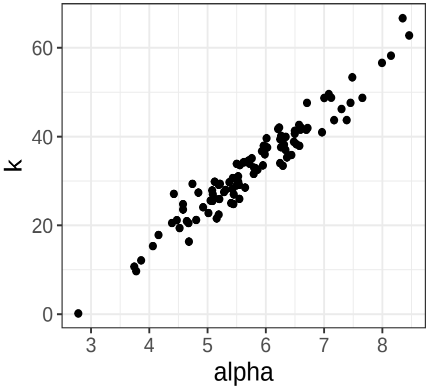
$$k_{\mathsf{best}} \pm 2.16 \times \sigma_k = [8.52, 57.80]$$

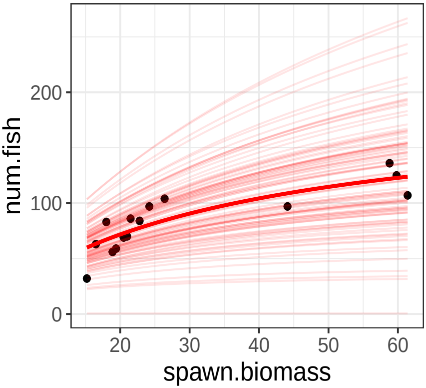
6. Parameter covariance:

$$R = \frac{\mathsf{Cov}(\alpha, k)}{\sigma_\alpha \times \sigma_k} = \frac{11.43}{1.04 \times 11.41} = 0.98.$$

Spaghetti plot

To get an idea of the variability in the confidence region, sample parameters from it, and plot resulting fitted curves.





Key takeaways

- Quality of parameter estimates depends on **model** and **data**, encoded by the FIM.
- The FIM provides a way of drawing elliptical confidence regions in parameter space.
- The FIM gives a lower bound for the error-covariance matrix.