Nonlinear regression

S. Benzekry

Master 2 Pharmacocinétique

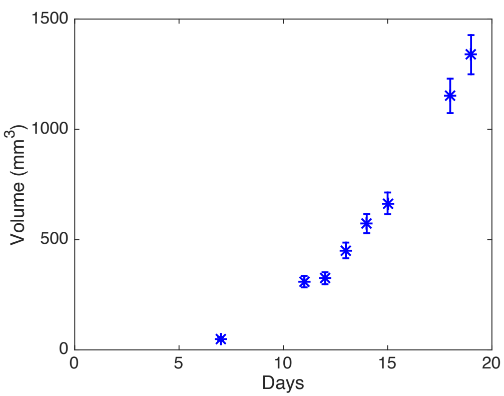
1. Fitting a model

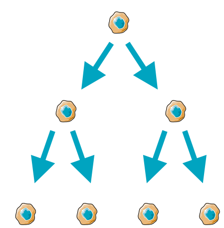
Data

Theory

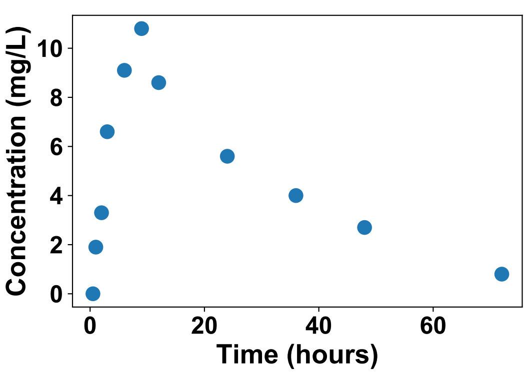
Mathematical model

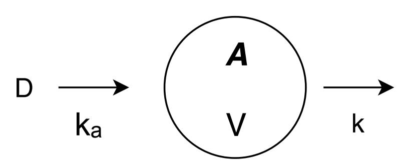
$$M : \begin{array}{ccc} \mathbb{R} \times \mathbb{R}^p & \to & \mathbb{R} \\ (t, \theta) & \mapsto & M(t, \theta) \end{array}$$





$$M(t, \theta) = e^{\theta t}$$





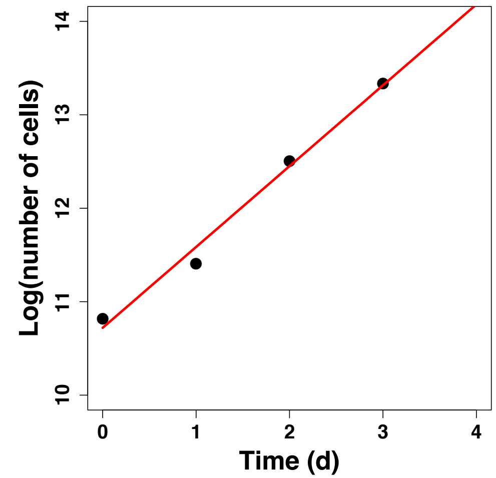
$$\begin{cases} \frac{dA_a}{dt} = -k_a A_a \\ \frac{dA}{dt} = k_a A_a - kA \\ A_a(t=0) = D, \quad (t=0) = 0 \end{cases}$$

$$C(t) = \frac{A(t)}{V}.$$

1.1 Fitting a *linear* model

Linear regression

$$y = \theta_0 + \theta_1 t + \varepsilon$$



Question: what is the « best » linear approximation of y ?

$$\begin{array}{cc} & \begin{array}{c} \\ \\ \end{array} \\ n & \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \overset{2}{\begin{pmatrix} 1 & t_1 \\ \vdots & \vdots \\ 1 & t_n \end{pmatrix}} \cdot \begin{pmatrix} \theta_0 \\ \theta_1 \end{pmatrix} \end{array} \longrightarrow$$

M rectangular

no solution

$$\Leftrightarrow y = M \cdot \theta$$

$$\times\ M^T (\in M_{2,n}) \qquad \Rightarrow \underbrace{M^T y}_{M_{2,n} \cdot M_{n,1}} = \underbrace{M^T M}_{M_{2,n} \cdot M_{n,2}} \cdot\ M_{2,1} \theta \longrightarrow$$

one unique solution

(if the square matrix $M^T M$ is invertible)

$$M_{2,1} \qquad M_{2,2} \cdot M_{2,1}$$

$$\hat{\theta} = \left(M^T M\right)^{-1} M^T y$$

1.2 General theory

Formalism

- **Observations**: *n* couples of points (t_j, y_j), with $y_j \in \mathbb{R}$ (or $\mathbb{R}^m$).

  We will denote $y = (y_1, \cdots, y_n) \in \mathbb{R}^n$ and $t = (t_1, \cdots, t_n)$.

- **Structural model:** a function

$$M : \begin{array}{ccc} \mathbb{R} \times \mathbb{R}^p & \to & \mathbb{R} \\ (t, \theta) & \mapsto & M(t, \theta) \end{array}$$

- The (unknown) vector of **parameters** $\theta \in \mathbb{R}^p$

Goal = find θ

Statistical model

$$y_j = M(t_j; \theta^*) + e_j$$

- « True » parameter θ^*
- e_j = **error** = measurement error + structural error
- See $(y_1, \cdots, y_n)$ as realisations of **random variables**

$$Y_j = M(t_j; \theta^*) + \varepsilon_j$$

Y_j, ε_j = r.v.

y_j, e_j = realizations

- $(y_1, \cdots, y_n)$ = **sample** with probability density function $p(y \,|\, \theta^*)$
- An **estimator** of θ^* is a random variable function of Y, denoted $\hat{\theta}$:

$$\hat{\theta} = h(Y_1, \cdots, Y_n)$$

Linear least-squares: statistical properties

$$Y = M\theta^* + \varepsilon$$

$$\hat{\theta}_{LS} = \underset{\theta \in \mathbb{R}^p}{\text{argmin}} \; \| Y - M\theta \|^2 \Leftrightarrow \hat{\theta}_{LS} = \left(M^T M\right)^{-1} M^T Y$$

Proposition:

Assume that $\varepsilon \sim \mathcal{N}\left(0, \sigma^2 I\right)$*, then*

$$\hat{\theta}_{LS} \sim \mathcal{N}\left(\hat{\theta}^*, \sigma^2 \left(M^T M\right)^{-1}\right)$$

From this, standard errors and confidence intervals can be computed on the parameter estimates

$$se\left(\hat{\theta}_{LS,p}\right) = s\sqrt{\left(M^T M\right)^{-1}_{p,p}} \qquad IC_\alpha\left(\theta_{LS,p}\right) = \theta^* \pm t^{\alpha/2}_{n-p} s\sqrt{\left(M^T M\right)^{-1}_{p,p}} \qquad s^2 = \frac{1}{n-p} \left\| y - M\hat{\theta}_{LS} \right\|^2$$

Statistical test for the model parameters

$$\hat{\theta} \sim \mathcal{N}\left(\hat{\theta}^{*}, \sigma^2 \left(M^T M\right)^{-1}\right)$$

For $k = 1,2\ldots,p$

$$t_k = \frac{\theta_k - \theta_k^*}{se_k}$$

t-distribution with $n - p$ degrees of freedom

$\Longrightarrow$ t-test (Wald test)

H_0 : « $\beta_k = 0$ » versus H_1 : « $\beta_k \neq 0$ »

Under the null hypothesis, $t_{stat} = \dfrac{\hat{\beta}_k}{se_k}$ follows a t-distribution with $n - d$ degrees of freedom

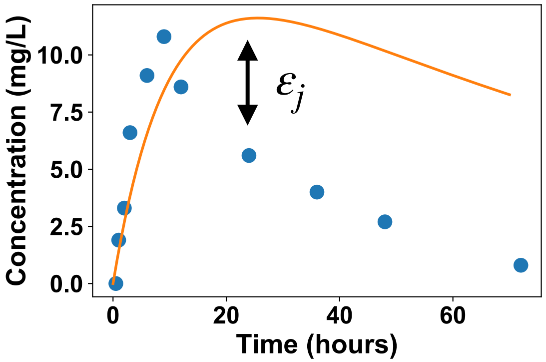
p-value:

$$\mathbb{P}\left(|t_{n-p}| \geq t_{stat}\right) = 2\left(1 - \mathbb{P}\left(t_{n-p} \leq t_{stat}\right)\right)$$

Nonlinear regression: least-squares

$$Y = M(t;\theta^*) + \varepsilon$$

$$\hat{\theta}_{LS} = \underset{\theta \in \mathbb{R}^p}{\text{argmin}} \left\| Y - M(t;\theta) \right\|^2$$



Linearization: $M(t,\theta) = M(t,\theta^*) + J \cdot (\theta - \theta^*) + o(\theta - \theta^*), \quad J = D_\theta M(t,\theta^*)$

Proposition:
Assume $\varepsilon \sim \mathcal{N}\left(0, \sigma^2 I\right)$. Then, for large n, approximately

$$\hat{\theta}_{LS} \sim \mathcal{N}\left(\hat{\theta}^*, \sigma^2 \left(J^T J\right)^{-1}\right)$$

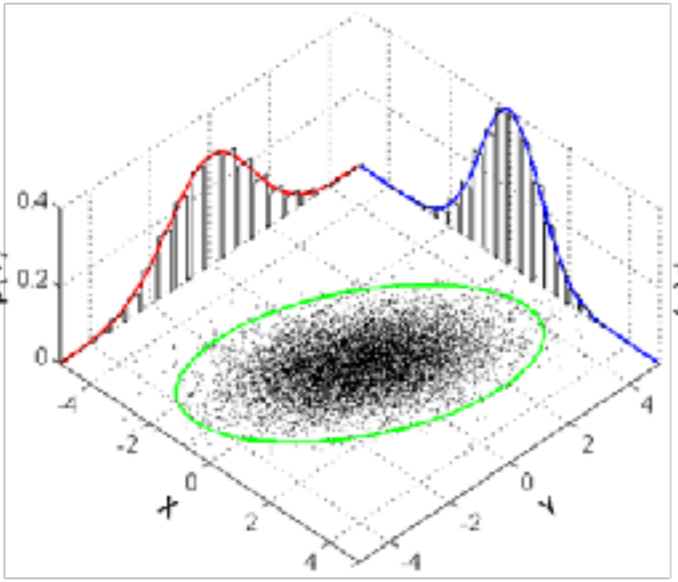
⇒ standard errors, confidence intervals

Sensitivity matrix

$$J = D_\theta M(t,\hat{\theta}) = \begin{pmatrix} \frac{\partial M}{\partial \theta_1}\left(t_1,\hat{\theta}\right) & \cdots & \frac{\partial M}{\partial \theta_p}\left(t_1,\hat{\theta}\right) \\ \vdots & \ddots & \vdots \\ \frac{\partial M}{\partial \theta_1}\left(t_n,\hat{\theta}\right) & \cdots & \frac{\partial M}{\partial \theta_p}\left(t_n,\hat{\theta}\right) \end{pmatrix}$$

$$var\left(\hat{\theta}_{LS}\right) = \sigma^2 \left(J^T J\right)^{-1}$$

- $J^T J$ is a $p \times p$ symmetric matrix
- It is invertible if and only if $rank(J) = p$
- Column k of $J = 0 \Leftrightarrow M(t,\hat{\theta})$ does not depend on θ_k
- Line i of $J = 0 \Leftrightarrow M(t_i,\hat{\theta})$ does not depend on θ



Nonlinear regression: Likelihood maximization

$$Y = M(t; \theta^*) + \varepsilon$$

The likelihood is defined by

$$L(\theta) = p(y_1, \cdots, y_n | \theta) = \prod_{j=1}^{n} p(y_j | \theta)$$

It is the probability to observe y if the parameter is θ.

The maximum likelihood estimator (MLE) is the value of θ that maximizes the likelihood

$$\hat{\theta}_{MV} = \underset{\theta}{\mathrm{argmax}} \, L(\theta)$$

Asymptotic properties of the MLE

Proposition:

Under regularity assumptions on L, when $n \to +\infty$

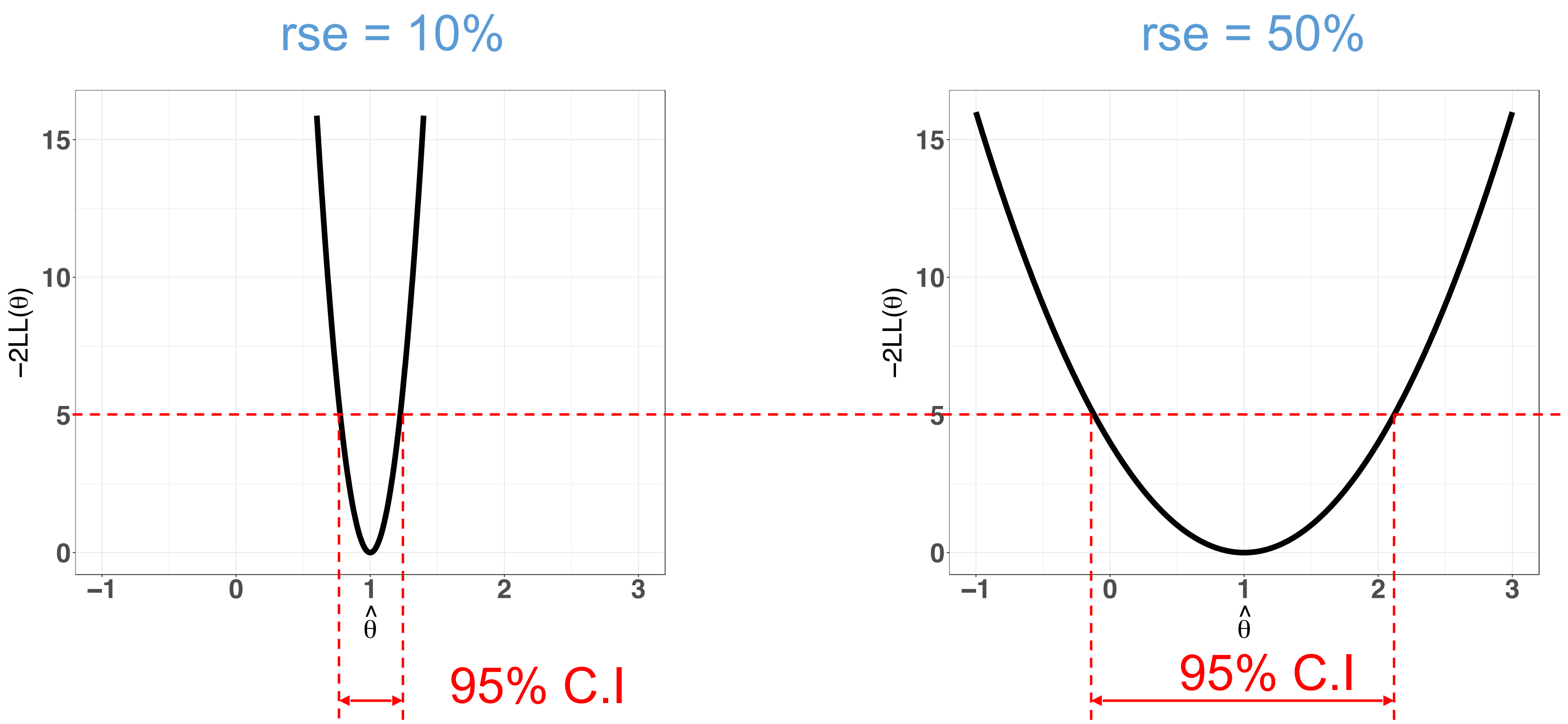
1. $\hat{\theta}_{MV} \longrightarrow \theta^*$ *(consistency)*
2. $\hat{\theta}_{MV}$ *is asymptotically of minimal variance (it reaches the Cramér-Rao bound):*

$$\sqrt{n}\left(\hat{\theta}_{MV} - \theta^*\right) \rightharpoonup \mathcal{N}\left(0, I_{\theta^*}^{-1}\right)$$

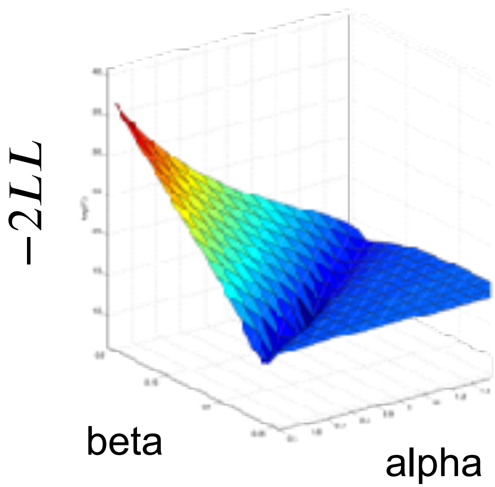
where I_{θ^*} is the Fisher information matrix

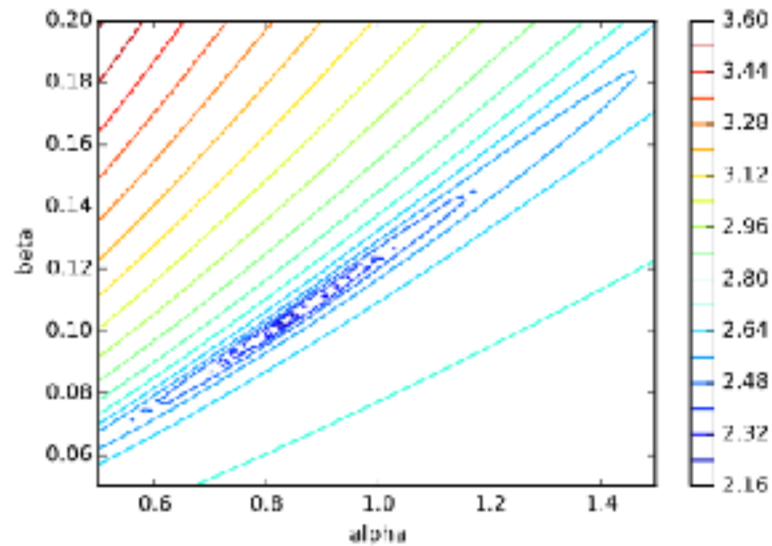
$$\left(I_{\theta^*}\right)_{j,k} = \mathbb{E}\left[\left\{\frac{\partial \log(p(Y|\theta^*))}{\partial \theta_j}\right\}\left\{\frac{\partial \log(p(Y|\theta^*))}{\partial \theta_k}\right\}\right] = \mathbb{E}\left[-\left(\frac{\partial^2 \log\left(p\left(Y|\theta^*\right)\right)}{\partial \theta_j \partial \theta_k}\right)\right].$$

Precision of the estimates



Correlation between estimates





Correlation matrix of the estimates

	R.S.E.(%)			
alpha_pop	3.09	1		
beta_pop	5.65	0.96974	1	
b	23.3	-0.0055971	0.0032018	1

	MIN	MAX	MAX/MIN
Eigen values	0.014	2	1.4e+2

small r.s.e on alpha and beta, but large correlation

MLE: normal errors

$$Y_j = M\left(t_j; \theta^*\right) + \varepsilon_j, \quad \varepsilon_j \sim \mathcal{N}(0, \sigma)$$

$$p(y_j \mid \theta, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{\left(y_j - M(t_j, \theta)\right)^2}{2\sigma^2}}, \quad L(\theta, \sigma) = \frac{1}{\left(\sigma\sqrt{2\pi}\right)^n} e^{-\frac{\| y - M(t, \theta) \|^2}{2\sigma^2}}$$

Maximize $L(\theta, \sigma) \Leftrightarrow$ minimize $F(\theta, \sigma) = -\log\left(L(\theta, \sigma)\right)$

$$F(\theta, \sigma) = n \log\left(\sigma\sqrt{2\pi}\right) + \frac{\| y - M(t, \theta) \|^2}{2\sigma^2}$$

$$\frac{\partial F}{\partial \sigma}\left(\hat{\theta}, \hat{\sigma}\right) = 0 \Rightarrow \hat{\sigma} = \frac{1}{n} \| y - M(t, \hat{\theta}) \|^2$$

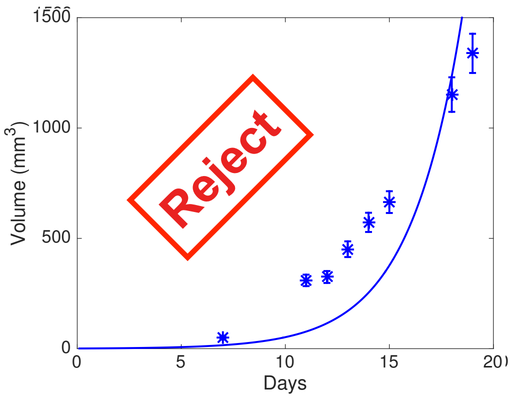
$$\Rightarrow \hat{\theta} = \underset{\theta}{\text{argmin}} \| y - M(t, \theta) \|^2$$

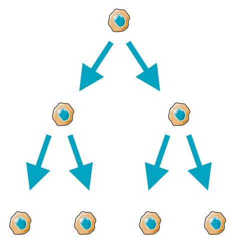
Maximum likelihood $\Leftrightarrow$ Least-squares

Application: tumor growth

What are **minimal** biological processes able to recover the **kinetics** of (experimental) tumor growth?

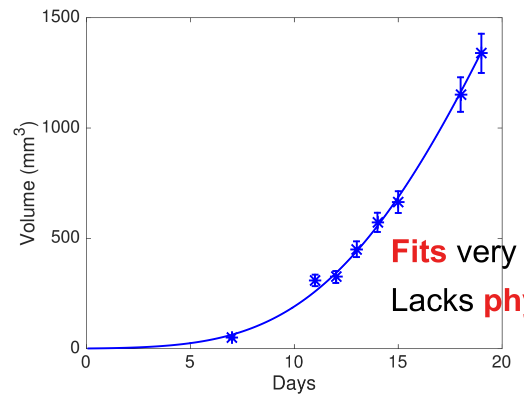
Exponential





$$\frac{dV}{dt} = aV$$

Gompertz

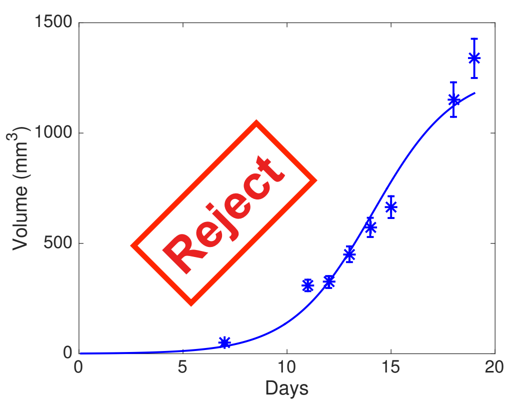


$$\frac{dV}{dt} = \alpha e^{-\beta t} V$$

Fits very well

Lacks **physiological** interpretation

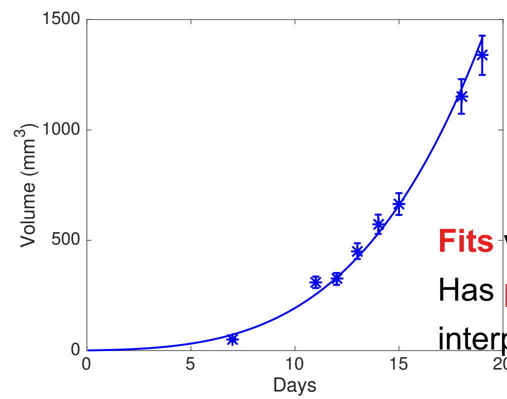
Logistic



Competition

$$\frac{dV}{dt} = aV\left(1 - \frac{V}{K}\right)$$

Power law



$$\frac{dV}{dt} = \alpha V^{\gamma}$$

Fits very well

Has **physiological** interpretation

Benzekry et al., PloS Comp Biol, 2014

Goodness of fit metrics

Sum of Squared Errors

$$SSE^i = \sum_{j=1}^{n^i} \left(\frac{V_j^i - V(t_j^i, \hat{\theta}^i)}{\hat{\sigma}_j^i} \right)^2$$

Akaike Information Criterion

$$AIC^i = -2l(\hat{\theta}^i) + 2p$$

number of parameters

Model	SSE	AIC	RMSE	R2	p > 0.05	#
Power law	**0.164(0.0158 - 0.646)[1]**	**-18.4(-43.2 - 1.63)[1]**	**0.415(0.145 - 0.899)[1]**	**0.97(0.801 - 0.998)[1]**	100	2
Gompertz	0.176(0.019 - 0.613)[2]	-16.9(-48.2 - 1.1)[2]	0.433(0.156 - 0.875)[2]	0.971(0.828 - 0.997)[2]	100	2
Logistic	0.404(0.0869 - 0.85)[3]	-5.41(-18.4 - 3.88)[3]	0.665(0.331 - 1)[3]	0.908(0.712 - 0.989)[3]	100	2
Exponential	1.9(0.31 - 3.56)[4]	10.7(-5.38 - 23.1)[4]	1.4(0.595 - 1.95)[4]	0.69(0.454 - 0.944)[4]	15	1

Root Mean Squared Errors

$$RMSE^i = \sqrt{\frac{1}{n-p} SSE^i}$$

R^2

$$R^{2,j} = 1 - \frac{\sum_j \left(V_j^i - V(t_j^i; \hat{\theta}^i) \right)^2}{\sum_j \left(V_j^i - \overline{V^i} \right)^2}$$

Parameter values and identifiability

Model	Par.	Unit	Median value (CV)	NSE (%) (CV)
Power law	α	$mm^{3(1-\gamma)} \cdot day^{-1}$	0.886 (30.8)	8.17 (52.5)
	γ	-	0.788 (7.56)	2.28 (58.6)
Gompertz	α_0	day^{-1}	1.68 (23.5)	6.11 (82.9)
	β	day^{-1}	0.0703 (28)	8.35 (92.9)
Logistic	a	day^{-1}	0.474 (13.3)	2.93 (23.3)
	K	mm^3	1.92e+03 (36.7)	15.8 (28.7)
Exponential	a	day^{-1}	0.356 (12.9)	2.53 (19.4)
Generalized logistic	a	$[day^{-1}]$	2555 (148)	2.36e+05 (137)
	K	$[mm^3]$	4378 (307)	165 (220)
	α	-	0.0001413 (199)	2.36e+05 (137)

NSE = Normalized Standard Error ⟷ practical identifiability

$$\hat{\theta} \sim \mathcal{N}\left(\theta^*, \hat{\sigma}^2 \left(J \cdot J^T\right)^{-1}\right)$$

$$se\left(\hat{\theta}^k\right) = \sqrt{\hat{\sigma}^2 \left(J \cdot J^T\right)_{k,k}}$$

Likelihood ratio test

- Comparing two models

$$Y = M_1\left(t;\theta_1\right) + \varepsilon_1 \Longrightarrow L_1(\theta_1, y) \qquad Y = M_2\left(t;\theta_2\right) + \varepsilon_2 \Longrightarrow L_2(\theta_2, y)$$

- Models have to be nested: $\theta_1 \subset \theta_2$ (ex: M_1 = one compartment, M_2 = two compartments)

$$LRT_{stat} = 2\left(LL_2\left(\hat{\theta}_2\right) - LL_1\left(\hat{\theta}_1\right)\right)$$
$$= n\log\left(\frac{\sum_{j=1}^{n}\left(y_j - M_1(t_j, \theta_1)\right)^2}{\sum_{j=1}^{n}\left(y_j - M_2(t_j, \theta_2)\right)^2}\right)$$

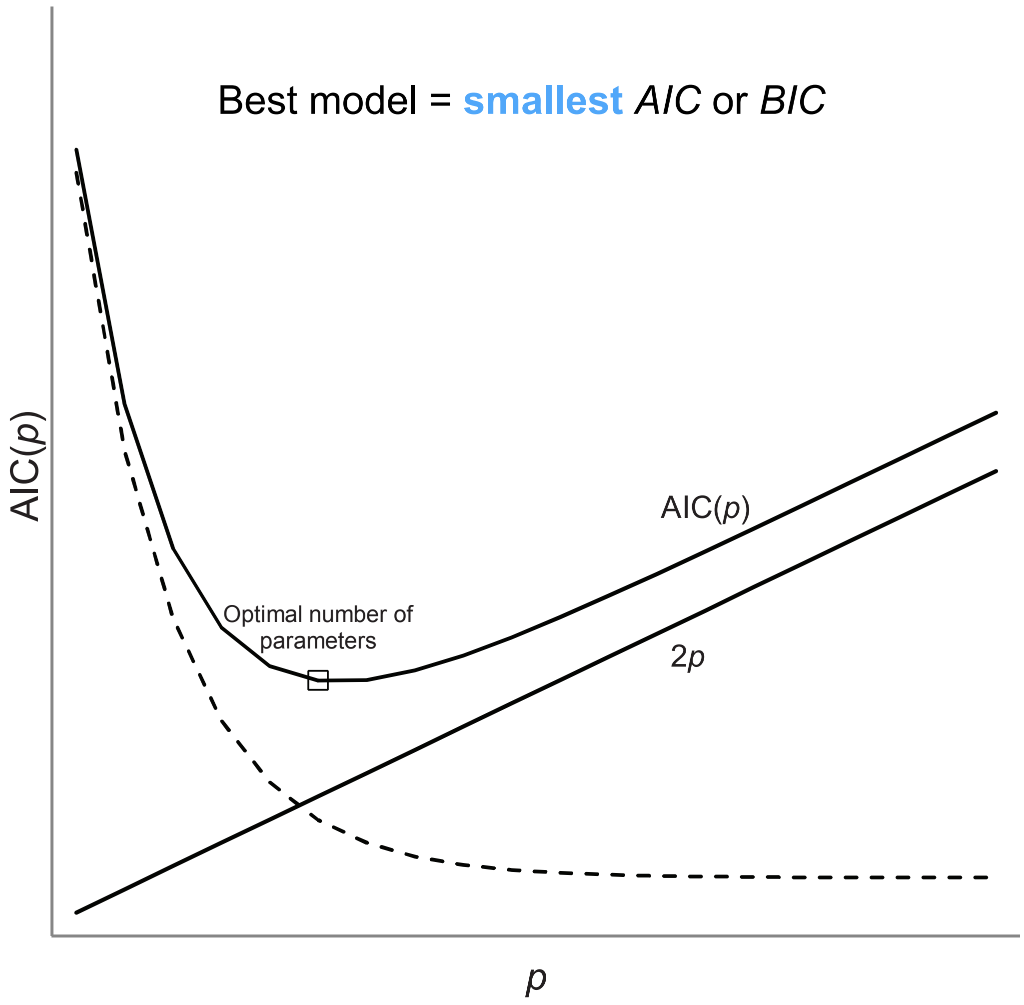
is approximated by a χ^2 distribution with $p_2 - p_1$ degrees of freedom

Null hypothesis *H₀* = « Model M_1 (small model) is better »

Information criteria

$$AIC = -2LL(\hat{\theta}) + 2p \qquad BIC = -2LL(\hat{\theta}) + \log(n)p$$

Best model = **smallest** *AIC* or *BIC*



Confidence interval and prediction interval

$$Y = M(t;\theta) + \varepsilon$$

- Prediction at new time t_{new} $\qquad \widehat{M_{new}} = M\left(t_{new}, \hat{\theta}\right)$

- Uncertainty on parameter estimate $\hat{\theta} \Longrightarrow$ confidence interval on $\widehat{M_{new}}$

$$\widehat{M_{new}} \sim \mathcal{N}\left(M_{new}, Var\left(\widehat{M_{new}}\right)\right)$$

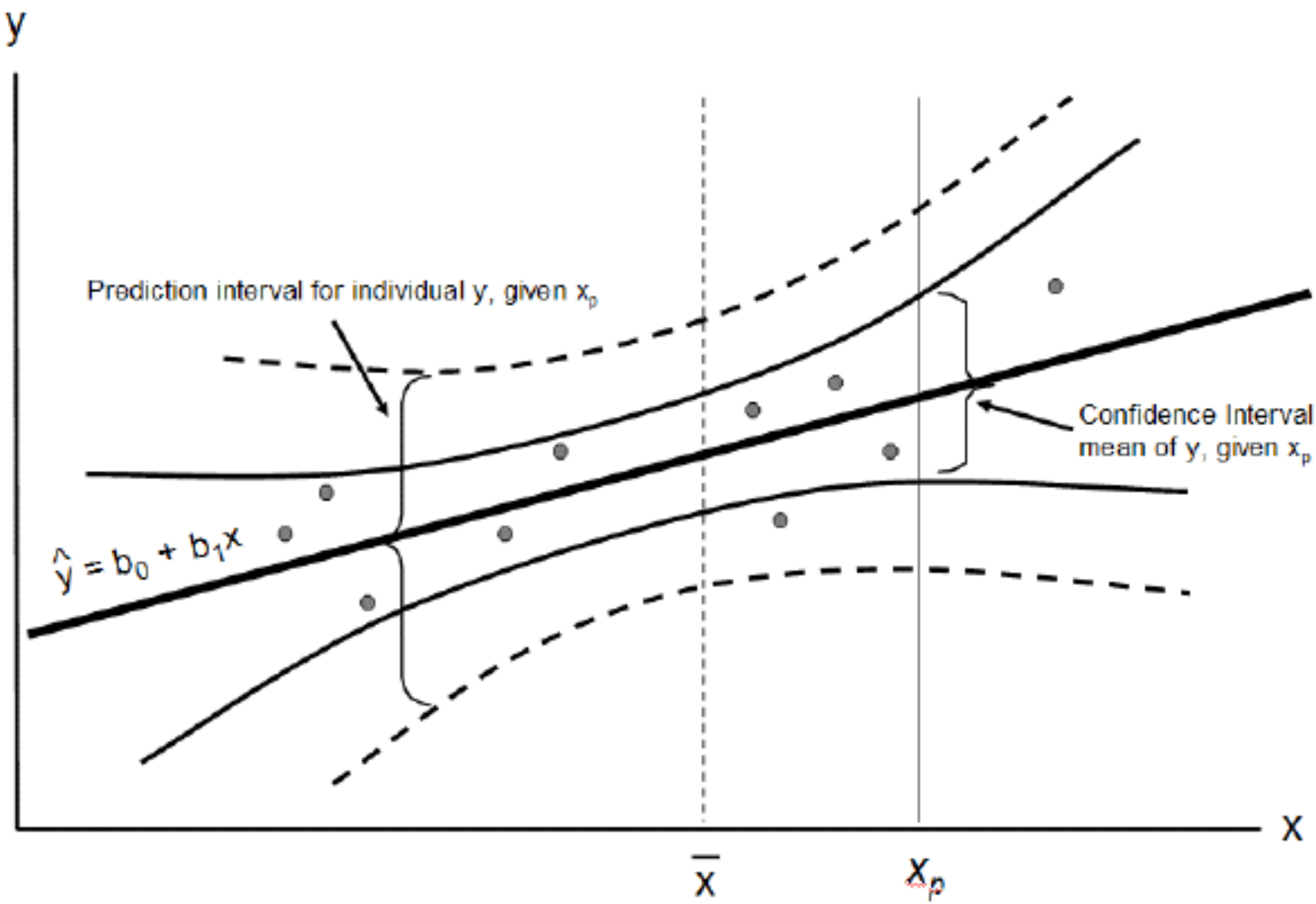
- Uncertainty on parameter estimate $\hat{\theta}$

+ uncertainty on observation ε (e.g. measurement error) $\Longrightarrow$ prediction interval on $\widehat{M_{new}}$

$$y_{new} = M_{new} + \varepsilon$$

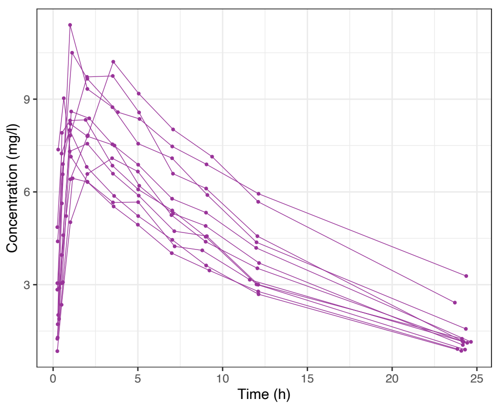
$$y_{new} \sim \mathcal{N}\left(\widehat{M_{new}}, Var\left(\widehat{M_{new}}\right) + \sigma^2 I\right)$$

Confidence interval vs prediction interval



Mixed-effects modeling

Population data

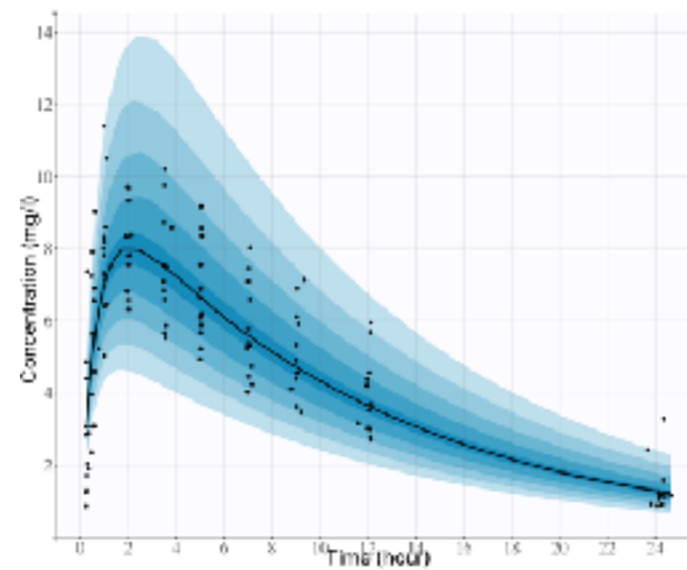


$$\psi^i = \psi_{pop} + \eta^i, \quad \eta^i \sim \mathcal{N}(0, \Omega)$$

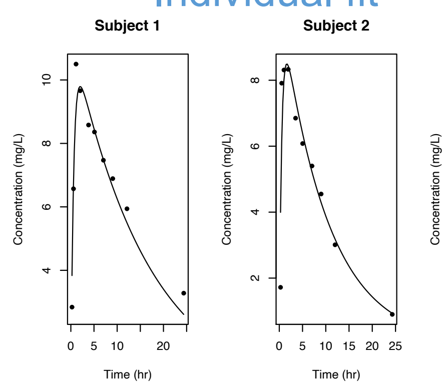
fixed effects

random effects

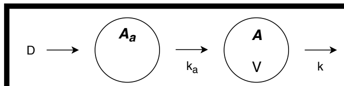
Population fit (MLE)



Individual fit



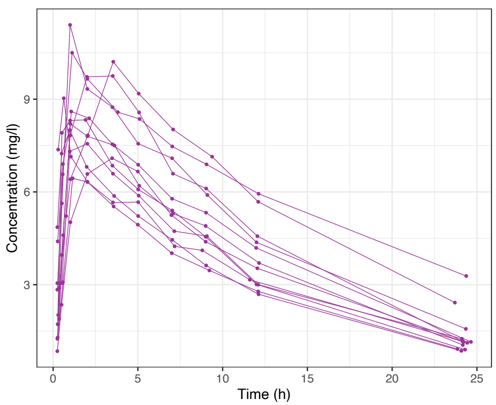
Individual structural model



$$\begin{cases} \frac{dA_a}{dt} = -k_a A_a \\ \frac{dA}{dt} = k_a A_a - kA \\ A_a(t=0) = D, \quad (t=0) = 0 \end{cases} \qquad C(t) = \frac{A(t)}{V}.$$

Population modeling: the two-steps approach

Population data



$$Y^{1} = M\left(t;\theta^{1}\right) + \varepsilon$$

$$Y^{2} = M\left(t;\theta^{2}\right) + \varepsilon$$

$$\vdots$$

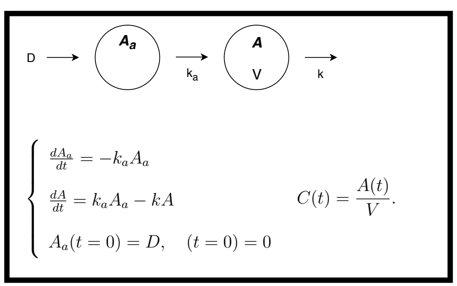
$$Y^{N} = M\left(t;\theta^{N}\right) + \varepsilon$$

$$\widehat{\theta_{pop}} = \frac{1}{N}\sum_{i=1}^{N}\theta^{i}$$

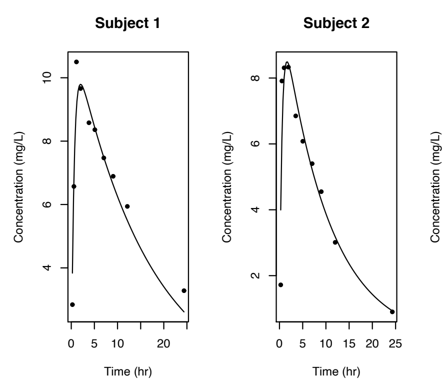
$$\widehat{\Omega} = VCov\left(\hat{\theta}^{i}\right)$$

$$\mathcal{N}\left(\widehat{\theta_{pop}}, \widehat{\Omega}\right)$$

Individual structural model



Individual fit



References

- Course « Statistics in Action with R » by Marc Lavielle

  http://sia.webpopix.org/index.html

- Seber, G. A., & Wild, C. J. (2003). Nonlinear regression. Hoboken (NJ): Wiley-Interscience.