Mathematical tools for pharmacometrics

S. Benzekry

sebastien.benzekry@inria.fr

Team composition

MATHEMATICAL MODELING

Pr D. Barbolosi
Dr S. Benzekry
Dr F. Gattacceca
Pr. Iliadis (emeritus)

PHARMACOLOGY

Pr J. Ciccolini
Dr R. Fanciullino
Dr A. Rodallec

MEDICINE

Dr D. Boulate (thoracic surgery)
Pr L. Greillier (thoracic oncology)
Dr X. Muracciole (radiotherapy)
Pr S. Salas (medical oncology)

Postdoctoral fellows

- Paul Dufossé (Inria, sup. S. Benzekry)
- Patricia Piris (CNRS/Inserm, sup. A. Rodallec)

PhD students

- Célestin Bigarré (Inria – Inserm, sup. S. Benzekry and X. Muracciole)
- Mathilde Dacos (APHM, sup. R. Fanciullino)
- Anthéa Deschamps (APHM, sup. F. Gattacceca)
- Mourad Hamimed, PharmD (AMU, sup. J. Ciccolini)
- Clémence Marin (AMU, sup. J. Ciccolini)
- Linh Nguyen Phuong (ICI – Laënnec, AMU, sup. S. Benzekry and S. Salas)
- Loic Osanno (APHM, sup. R. Fanciullino)
- Jessica Ou (AMU, sup. F. Gattacceca)

Research engineers

- Mohamed Boussena (Inria, sup. S. Benzekry)
- Sophie Marolleau (AMU, sup. J. Ciccolini)

Administrative assistant

- Sandrine Boute

COMPO: COMPutational pharmacology and clinical Oncology

MATHEMATICAL MODELING

Pr D. Barbolosi
Dr S. Benzekry
Dr F. Gattacceca
Pr. Iliadis (emeritus)

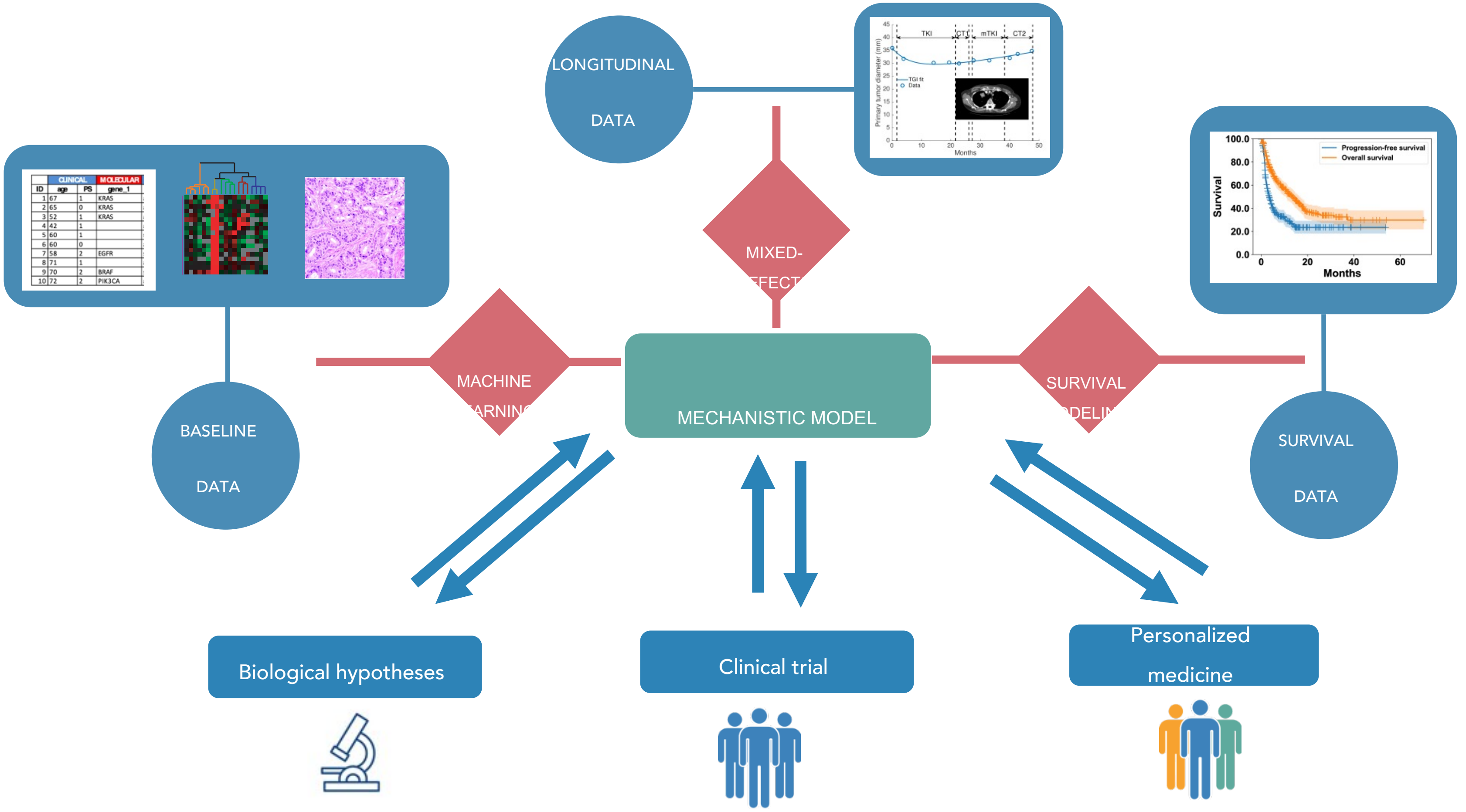
PHARMACOLOGY

Pr J. Ciccolini
Dr R. Fanciullino
Dr A. Rodallec

MEDICINE

Dr D. Boulate (thoracic surgery)
Pr L. Greillier (thoracic oncology)
Dr X. Muracciole (radiotherapy)
Pr S. Salas (medical oncology)

Mechanistic learning



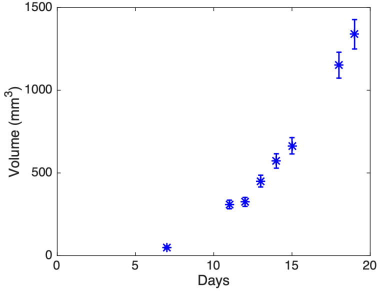
1. Introduction

Data

Theory

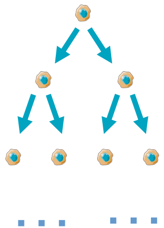
Mathematical model

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

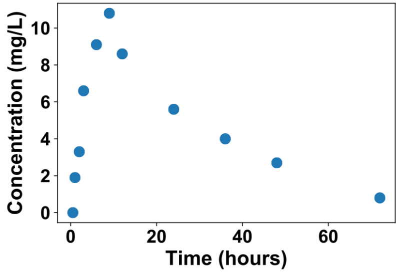


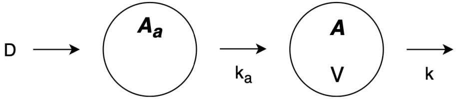
?

Statistical model



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

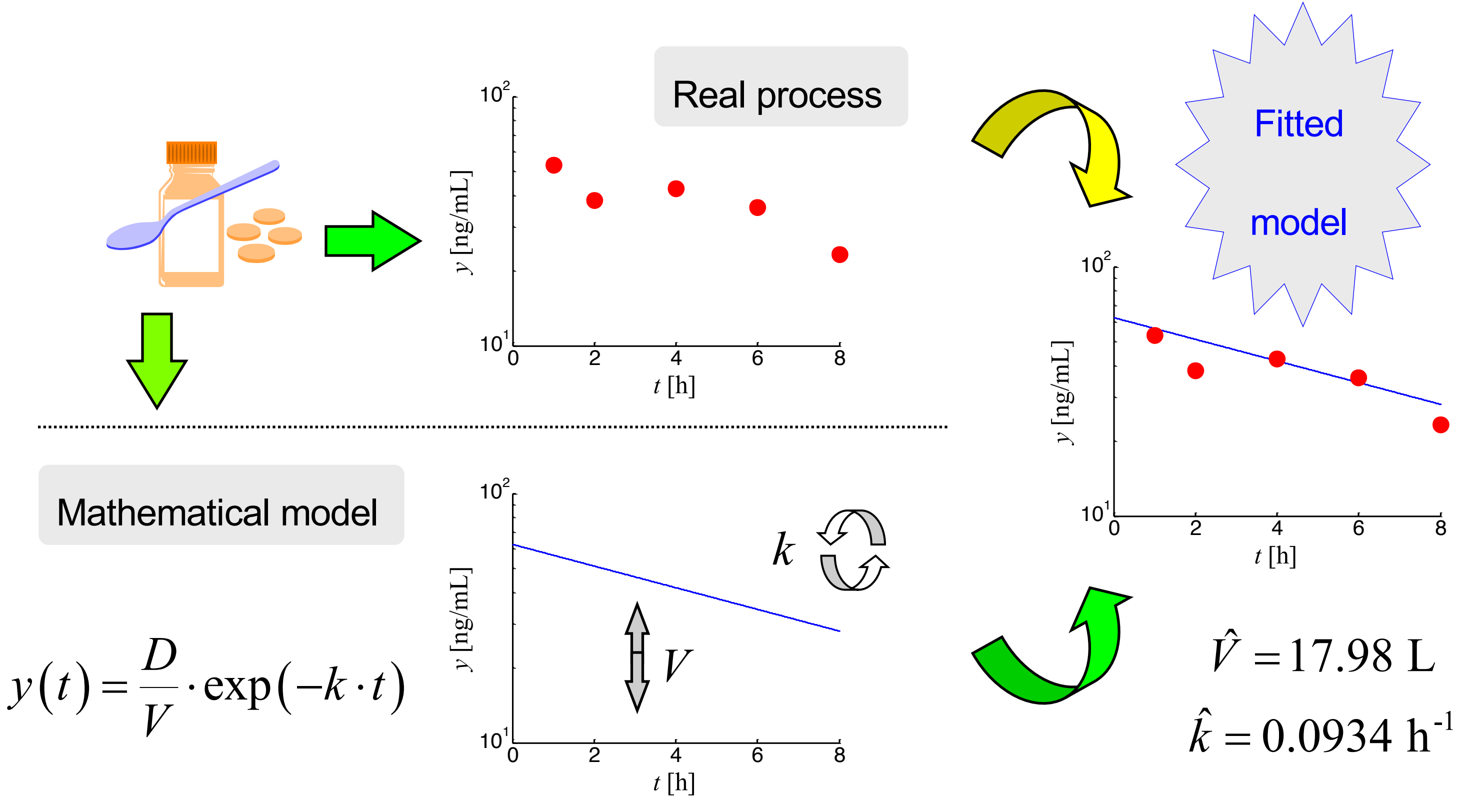




$$\begin{cases} \frac{dA_a}{dt} = -k_a A_a \\ \frac{dA}{dt} = k_a A_a - kA \\ A_a(t=0) = D, \quad (t=0) = 0 \end{cases}$$

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

Real process and mathematical model



Choose the best model

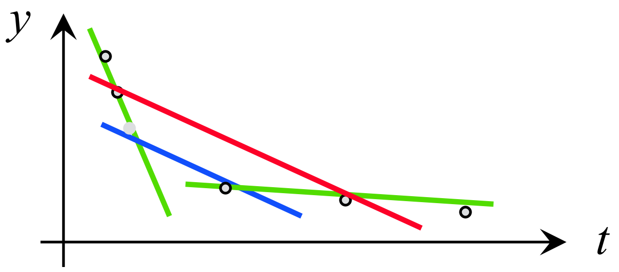
Rule 1 : The model should be a necessary and sufficient description of the real process

- necessary : good fitting on observed data.
- sufficient : without redundancy of parameters in the structure.

Rule 2 : The model should be parsimonious :

- adequately represent the real process.
- involve the smallest number of parameters.

- Ex :
  - 1-cpt model is misspecified.
  - 3-cpt model is redundant.
  - 2-cpt is the best model.



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
- $(y_1, \cdots, y_n)$ are realizations 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 $\hat{\theta}$:

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

Math interlude: proba/stats

Random variable

- **Definition**
  - A random variable is a numerical quantity whose value is determined by the outcome of a **random** experiment or process

- **Type**
  - *Continuous* : temperature in one month, Cmax
  - *Discrete:* toss of a coin, number of students passing 1st year PACES, toxicity grade

Mathematical tools : PDF, CDF

- Continuous mathematic functions

❶ Probability Density Function (PDF) $f(x)$

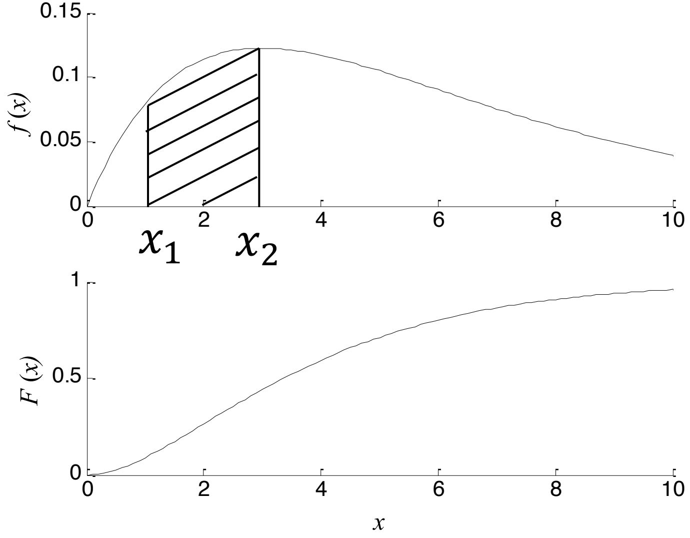
to compute an interval probability

$$\Pr\{\, x_1 \le X < x_2 \,\} = \int_{x_1}^{x_2} f(x) \cdot dx$$

❷ Cumulative Distribution Function (CDF) $F(x)$

to compute an interval probability

$$\Pr\{\, x_1 \le X < x_2 \,\} = F(x_2) - F(x_1)$$



Note : $\dfrac{dF(x)}{dx} = f(x)$ and area under the PDF curve expresses probability.

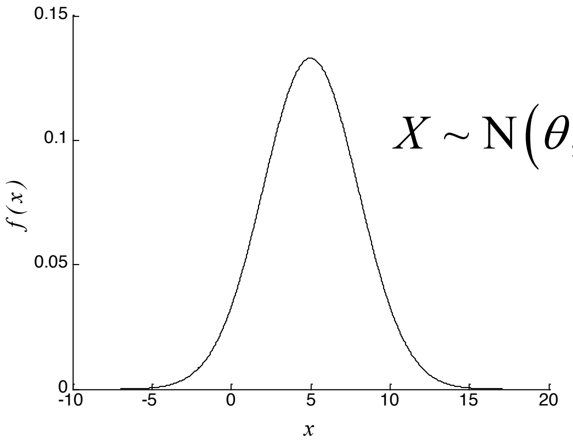
Sample vs. model characteristics

	Realization x	Random variable X
	Sample	**Model**
	Randomly drawn from a population $x_j \quad j=1:n$	Probability density function $f(x)$
Central tendency	Average $\bar{x} \equiv \text{ave}(x_j) = \frac{1}{n}\sum_{j=1}^{n} x_j$	Expectation $\text{E}[X] = \int_x x \cdot f(x) \cdot dx$
Variability	Empirical variance $s^2 \equiv \text{var}(x_j) = \frac{1}{n}\sum_{j=1}^{n}(x_j - \bar{x})^2$	Variance $\text{V}[X] = \int_x (x - \text{E}[X])^2 \cdot f(x) \cdot dx$

The Gaussian density

❶ Distribution model

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \cdot \exp\left[-\frac{1}{2} \cdot \left(\frac{x-\theta}{\sigma}\right)^2\right]$$



$$X \sim \mathrm{N}\left(\theta, \sigma^2\right)$$

- Parameters θ σ

❷ Distribution characteristics

$$\mathrm{E}[X] = \theta \quad \mathrm{V}[X] = \sigma^2$$

❸ Sample characteristics

$\bar{x}$ s^2

Parameter estimates

❶ + ❷ + ❸ ⇨

$$\theta = \mathrm{E}[X] = \bar{x}$$

$$\sigma^2 = \mathrm{V}[X] = s^2$$

Uncertainty vs. information

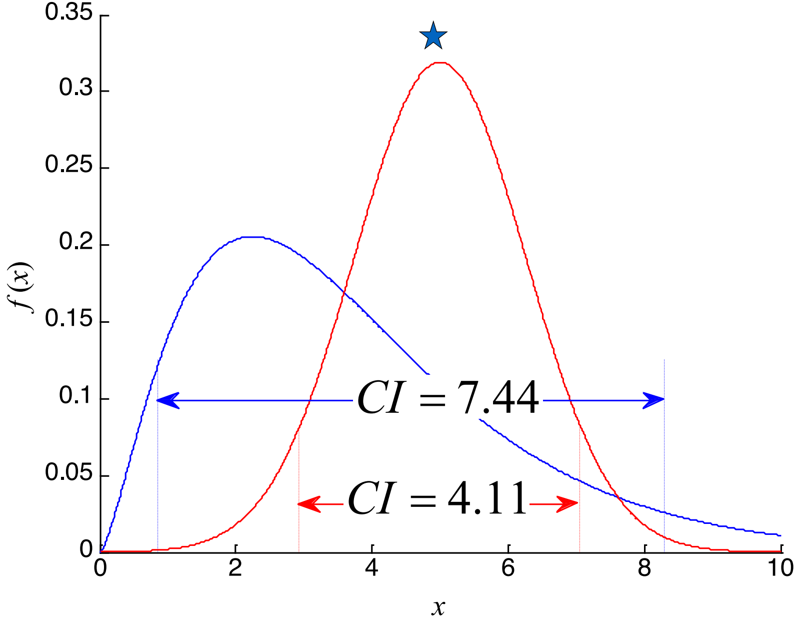
- The confidence interval quantifies the uncertainty for a random variable.

  It may be associated with “precision” or “dispersion”.

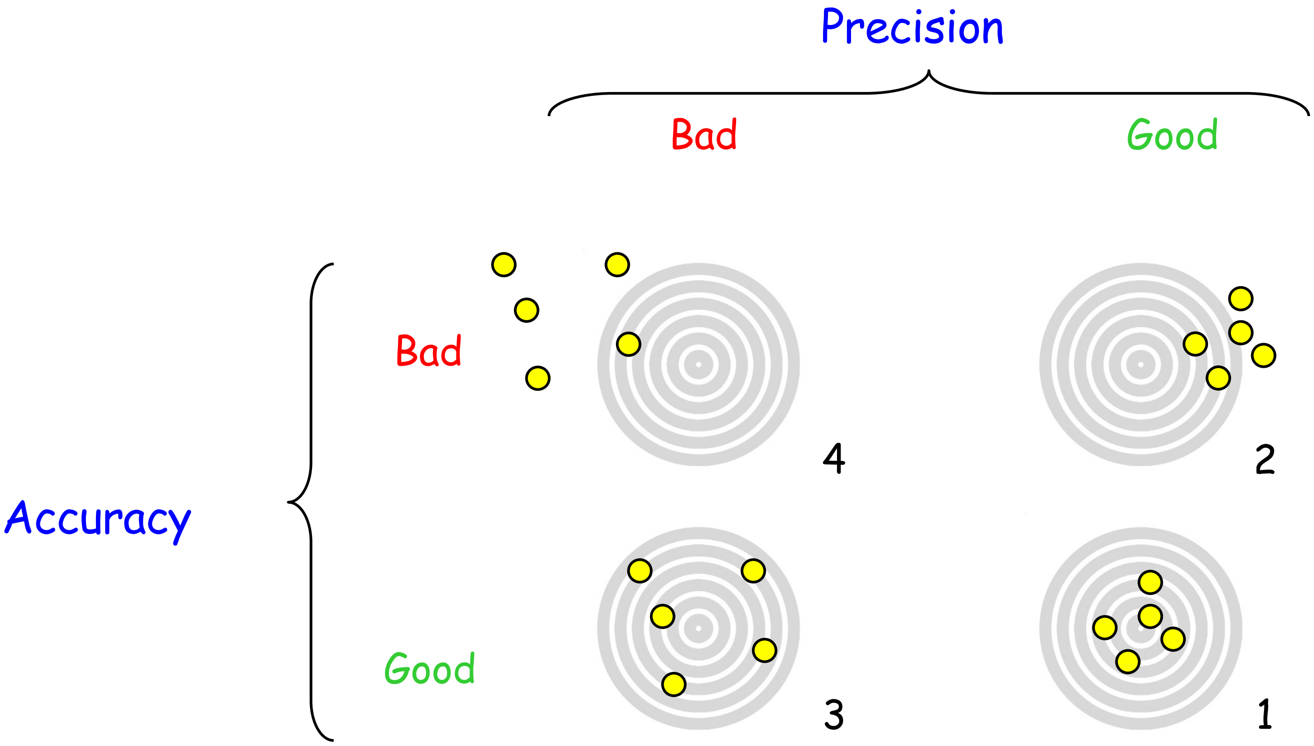
The narrower the confidence interval,

the higher the information

	information	
	low	high
precision	imprecise	accurate
dispersion	wide	narrow



Accuracy (bias) and precision (variance)

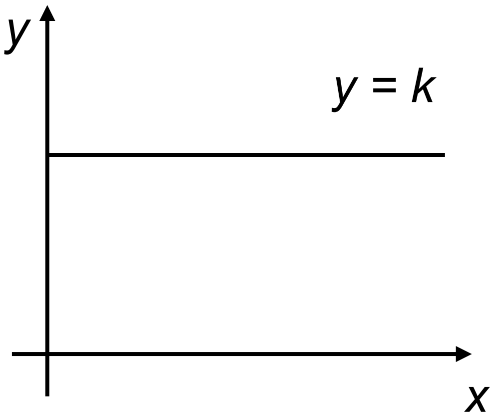


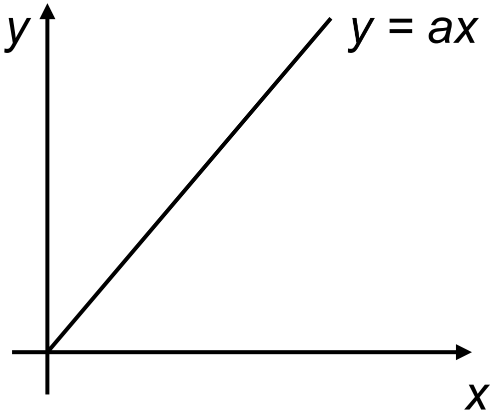
2. Linear regression

Math interlude: linear algebra

Linearity: the simplest mathematical relationship

- Variable y (ex: Cmax) that depends on variable x (ex: dose)





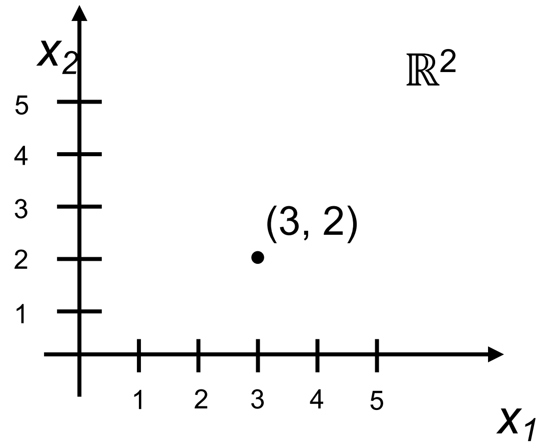
How to extend linearity to several variables $(x_1, \ldots, x_n)$?

Multiple variables = vectors

- a single (real) number is called a scalar : 1, -2, 3.5, 5/7, π, etc…
- an ordered set of numbers is called a vector: (x_1, x_2, x_3), (2, -1, 5) ≠ (-1, 2, 5)
- mathematically, it's an element of $\mathbb{R}^n$, n is called the dimension
- a vector can be written as row or column

(2, -1, 5)

$$\begin{pmatrix} 2 \\ -1 \\ 5 \end{pmatrix}$$



Matrices

- A matrix is a rectangular array of numbers with a given number of rows (*m*) and columns (*n*)

3 columns

$$\begin{pmatrix} 3 & -1 & 5 \\ -2 & 12 & -7 \end{pmatrix}$$ 2 rows

$$\begin{array}{c|cccc} & 1 & 2 & \dots & n \\ 1 & a_{11} & a_{12} & \dots & a_{1n} \\ 2 & a_{21} & a_{22} & \dots & a_{2n} \\ 3 & a_{31} & a_{32} & \dots & a_{3n} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ m & a_{m1} & a_{m2} & \dots & a_{mn} \end{array} = (a_{i,j})$$

- A *m x n* matrix can be applied to a vector in $\mathbb{R}^n$ and gives a vector in $\mathbb{R}^m$

$$M : \begin{array}{ccc} \mathbb{R}^n & \to & \mathbb{R}^m \\ x & \mapsto & M \cdot x \end{array}$$

$$\begin{pmatrix} 1 \\ 0 \\ 2 \end{pmatrix}$$ 3 rows

2 rows $$\begin{pmatrix} 3 & -1 & 5 \\ -2 & 12 & -7 \end{pmatrix}$$

3 columns

$$= \begin{pmatrix} 1 \times 3 + 0 \times (-1) + 2 \times 5 \\ \dots \end{pmatrix}$$

$$= \begin{pmatrix} 13 \\ -16 \end{pmatrix}$$ 2 rows

Matrix?

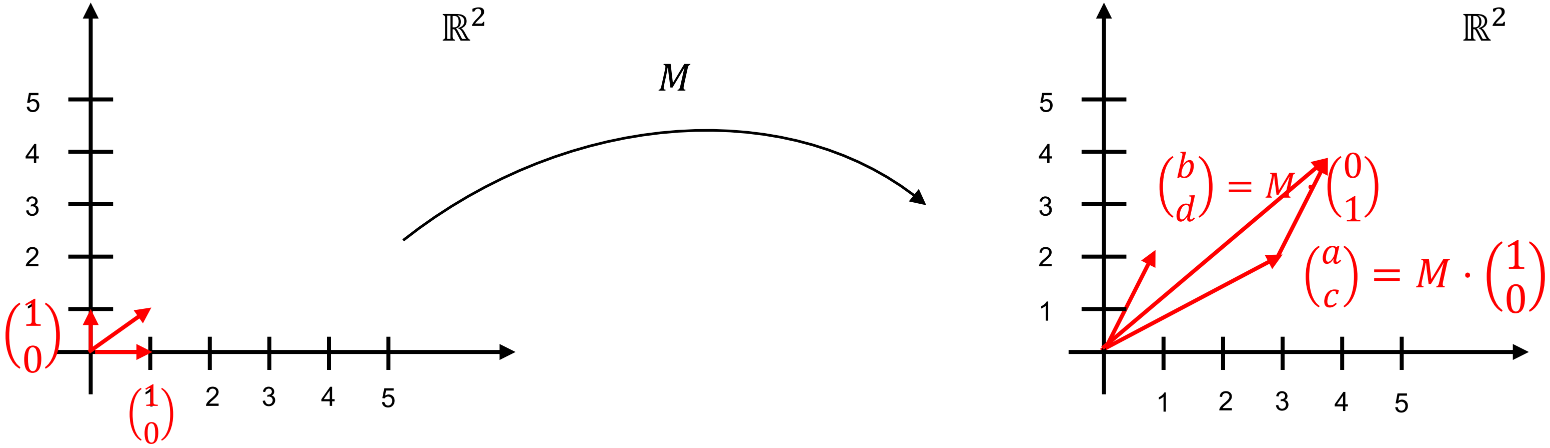
- $M = \begin{pmatrix} a & b \\ c & d \end{pmatrix} : \mathbb{R}^2 \to \mathbb{R}^2$ $\qquad M \cdot \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = \begin{pmatrix} ax_1 + bx_2 \\ cx_1 + dx_2 \end{pmatrix} = \begin{pmatrix} a \\ c \end{pmatrix} x_1 + \begin{pmatrix} b \\ d \end{pmatrix} x_2$

- Respects addition

$$M \cdot (x + y) = M \cdot x + M \cdot y$$

- And scalar multiplication

$$M \cdot (\lambda\, x) = \lambda M \cdot x$$



Matrix definitions

- Definition of terms
  - An array of numbers written as below is called a matrix of order $m \times n$.
  - Column vector if $n = 1$
  - Row vector if $m = 1$
  - Scalar if $m = n = 1$
  - Square matrix if $m = n$
  - A square matrix may be :
    - Symmetric if $a_{ij} = a_{ji} \quad i, \neq j$
    - Diagonal if $a_{ij} = 0 \quad i \neq j$
  - A diagonal matrix with $a_{ii} = 1 \quad i = 1 : n$ is the identity matrix I.

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix}$$

- Notation : matrix A (cap), matrix element a (lower), vector $\underline{a}$ (lower, underlined).

Special matrices

- Square matrix $\begin{pmatrix} a_{1,1} & \cdots & a_{1,n} \\ \vdots & \ddots & \vdots \\ a_{n,1} & \cdots & a_{n,n} \end{pmatrix}$

- Symmetric matrix: $M^T = M$ $\begin{pmatrix} 1 & 5 & -3 \\ 5 & 2 & -1 \\ -3 & -1 & -4 \end{pmatrix}$

- Diagonal matrix $\begin{pmatrix} d_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & d_n \end{pmatrix}$

- Identity matrix $I = \begin{pmatrix} 1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1 \end{pmatrix}$

$$\forall M \in M_{n,n}, \qquad M \cdot I = I \cdot M = M$$

Commonly used vectors in PKs

Sampling times	Observations at sampling times	Model parameters	Predictions at the sampling times and using model parameters
$\underline{t} = \begin{Bmatrix} t_1 \\ t_2 \\ \dots \\ t_m \end{Bmatrix}$	$\underline{y} = \begin{Bmatrix} y_1 \\ y_2 \\ \dots \\ y_m \end{Bmatrix}$	$\underline{x} = \begin{Bmatrix} V \\ k \\ \dots \\ x_p \end{Bmatrix}$	$\underline{y}(\underline{t}, \underline{x}) = \begin{Bmatrix} y(t_1, \underline{x}) \\ y(t_2, \underline{x}) \\ \dots \\ y(t_m, \underline{x}) \end{Bmatrix}$

- Followed by subscript, they denote vectors associated with a given individual or time point
  - $\underline{y}_j$ are the observations for the $j-$th time point
  - $\underline{x}_i$ are parameters for the $i-$th individual.

Matrix example 1: data array

n columns (covariates)

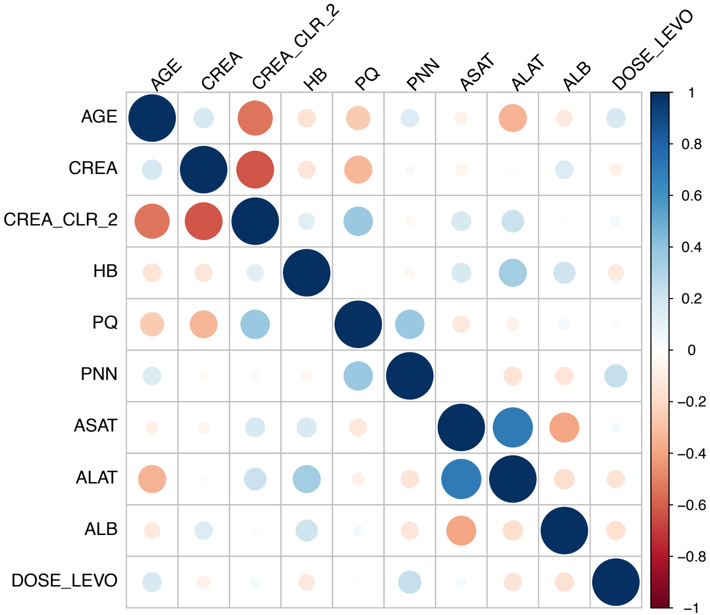
m patients

AGE	BSA	CREA	CREA_CLR_2	HB	PQ	PNN	ASAT	ALAT	ALB	DOSE_LEVO
64	2	53	140	106	225	10.4	93	101	25.6	450
37	1.5	37	195	152	185	5.75	25	70	44	300
61	1.75	30	200	112	428	24.71	34	32	34.8	5600
49	2	71.6	120	86	138	14.06	15	23	38.1	1250
49	2	56	136	78	103	4.03	30	67	36.1	0
33	1.79	54	211	79	441	16.95	22	34	36.6	0
33	1.79	71	152	119	300	5.42	30	54	34.3	0
33	1.79	52	177	105	146	4.86	21	34	34.3	0
36	1.87	67	128	129	279	4.76	45.5	118	39.1	0
47	2	56	149	85	400	8.76	15	15	35.4	3450
47	2	77	130	101	501	8.76	12	19	35.2	0
47	2	73	154	99	202	6.92	82	16	35.4	3140
59	1.83	71	125	86	22	0.2	35	85	37.4	2700

Matrix example 2: correlation matrix

$(x_1, \cdots, x_n)$ n vectors, $C_{i,j} = corr(x_i, x_j)$

	AGE	CREA	CREA_CLR_2	HB	PQ	PNN	ASAT	ALAT	ALB	DOSE_LEVO
AGE	1.00	0.18	-0.54	-0.15	-0.25	0.14	-0.07	-0.35	-0.11	0.16
CREA	0.18	1.00	-0.63	-0.13	-0.34	-0.03	-0.06	-0.02	0.14	-0.07
CREA_CLR_2	-0.54	-0.63	1.00	0.12	0.38	-0.03	0.17	0.22	-0.02	0.04
HB	-0.15	-0.13	0.12	1.00	0.01	-0.04	0.17	0.34	0.21	-0.11
PQ	-0.25	-0.34	0.38	0.01	1.00	0.38	-0.13	-0.07	0.05	-0.02
PNN	0.14	-0.03	-0.03	-0.04	0.38	1.00	0.00	-0.15	-0.14	0.23
ASAT	-0.07	-0.06	0.17	0.17	-0.13	0.00	1.00	0.71	-0.39	0.04
ALAT	-0.35	-0.02	0.22	0.34	-0.07	-0.15	0.71	1.00	-0.18	-0.14
ALB	-0.11	0.14	-0.02	0.21	0.05	-0.14	-0.39	-0.18	1.00	-0.16
DOSE_LEVO	0.16	-0.07	0.04	-0.11	-0.02	0.23	0.04	-0.14	-0.16	1.00

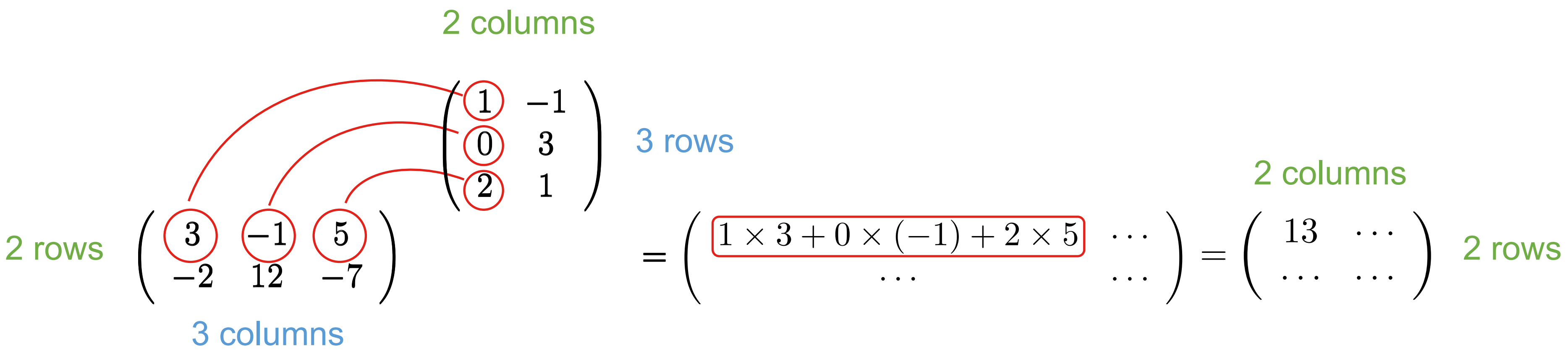


Matrix algebra

- Transposition $\quad A \quad m\times n \qquad C = A^T \quad n\times m \qquad c_{ij} = a_{ji}$
- Addition $\quad A, B \quad m\times n \qquad C = A + B \quad m\times n \qquad c_{ij} = a_{ij} + b_{ij}$
- Multiplication $\quad A \quad m\times n, \quad B \quad n\times q \qquad C = A\cdot B \quad m\times q \qquad c_{ij} = \sum_{k=1}^{n} a_{ik}\cdot b_{kj}$
- Square matrix
  - Determinant (characteristic scalar) $|A|$ Ex. $|A| = a_{11}\cdot a_{22} - a_{12}\cdot a_{21}$ for $n = 2$
  - Inversion $\quad C = A^{-1} \qquad C\cdot A = A\cdot C = I \qquad c_{ij} = (-1)^{i+j}\dfrac{|A^T|_{i,j}}{|A|}$
- Rules

$$(A\cdot B)^T = B^T\cdot A^T \qquad (A\cdot B)^{-1} = B^{-1}\cdot A^{-1} \qquad |A^{-1}| = (|A|)^{-1}$$

Matrix multiplication



- The number of rows of the second matrix must be the same as the number of columns of the first matrix
- It is only possible to multiply a matrix in $M_{m,n}$ by a matrix in $M_{n,p}$ and it gives a $M_{m,p}$ matrix

Linearity

- Let :

$$\text{Input}_1(t) \Rightarrow \boxed{T} \Rightarrow \text{Output}_1(t)$$

$$\text{Input}_2(t) \Rightarrow \boxed{T} \Rightarrow \text{Output}_2(t)$$

- For the input combination :

$$w_1 \cdot \text{Input}_1(t) + w_2 \cdot \text{Input}_2(t) \Rightarrow \boxed{T} \Rightarrow \ ?$$

- The operator T is linear, if and only if ? $= w_1 \cdot \text{Output}_1(t) + w_2 \cdot \text{Output}_2(t)$

PK example - 1

- For the 1-cpt model $$y(t) = \frac{D}{V} \cdot \exp(-k \cdot t)$$

associate : - $y(t)$ to $\mathrm{Output}(t)$

- D to $\mathrm{Input}(t)$

- The model is a linear because

when $\frac{D_1}{V} \cdot \exp(-k \cdot t) \rightarrow y_1(t)$ and $\frac{D_2}{V} \cdot \exp(-k \cdot t) \rightarrow y_2(t)$

then $$\Rightarrow \frac{w_1 \cdot D_1 + w_2 \cdot D_2}{V} \cdot \exp(-k \cdot t) = w_1 \cdot y_1(t) + w_2 \cdot y_2(t)$$

PK linearity (proportionality)

PK example - 2

- For the 1-cpt model

$$y(t)=\frac{D}{V}\cdot\exp(-k\cdot t)$$

associate : - k to $\text{Input}(t)$

- $y(t)$ to $\text{Output}(t)$

- The model is a nonlinear operator because

when $\frac{D}{V}\cdot\exp(-k_1\cdot t)\rightarrow y_1(t)$ and $\frac{D}{V}\cdot\exp(-k_2\cdot t)\rightarrow y_2(t)$

then ⇒

$$\frac{D}{V}\cdot\exp\left[-(w_1\cdot k_1+w_2\cdot k_2)\cdot t\right]\neq w_1\cdot y_1(t)+w_2\cdot y_2(t)$$

Parametric nonlinearity

Linear regression

Linear system: Equation of a line

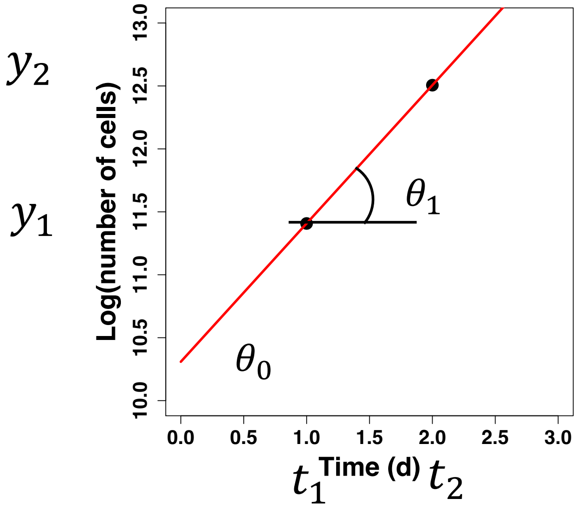
$$y = \theta_0 + \theta_1 t$$

$$\begin{cases} y_1 = 1\times\theta_0 + t_1\times\theta_1 \\ y_2 = 1\times\theta_0 + t_2\times\theta_1 \end{cases} \Leftrightarrow \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} = \begin{pmatrix} 1 & t_1 \\ 1 & t_2 \end{pmatrix} \cdot \begin{pmatrix} \theta_0 \\ \theta_1 \end{pmatrix}$$

$$y = M \cdot \theta \Rightarrow \theta = M^{-1} \cdot y$$

M^{-1} ?? $\quad M^{-1} := \text{“}\frac{1}{M}\text{”}, \quad M \cdot M^{-1} = \text{“}1\text{”} = I$

is $M \neq \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$ sufficient?

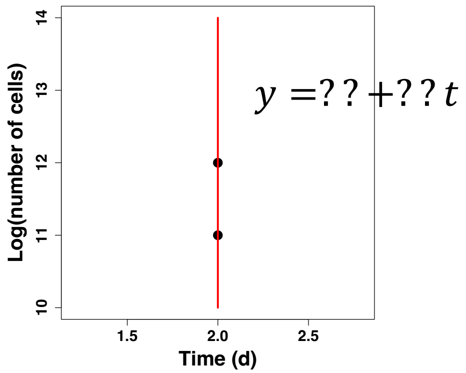


$$\begin{cases} 11.4 = 1\times\theta_0 + 1\times\theta_1 \\ 12.5 = 1\times\theta_0 + 2\times\theta_1 \end{cases} \Leftrightarrow \begin{pmatrix} 11.4 \\ 12.5 \end{pmatrix} = \begin{pmatrix} 1 & 1 \\ 1 & 2 \end{pmatrix} \cdot \begin{pmatrix} \theta_0 \\ \theta_1 \end{pmatrix}$$

$$\theta_0 = 10.3, \theta_1 = 1.1$$

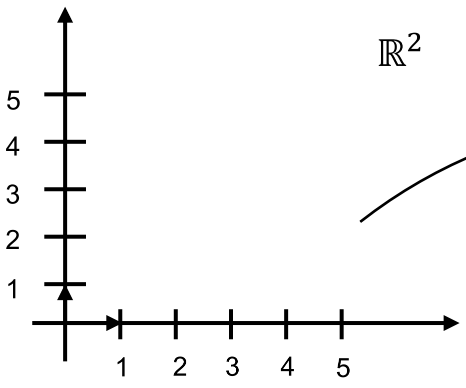
Doubling time = $\frac{\ln 2}{\theta_1} \times 24 = 15.1$ hours

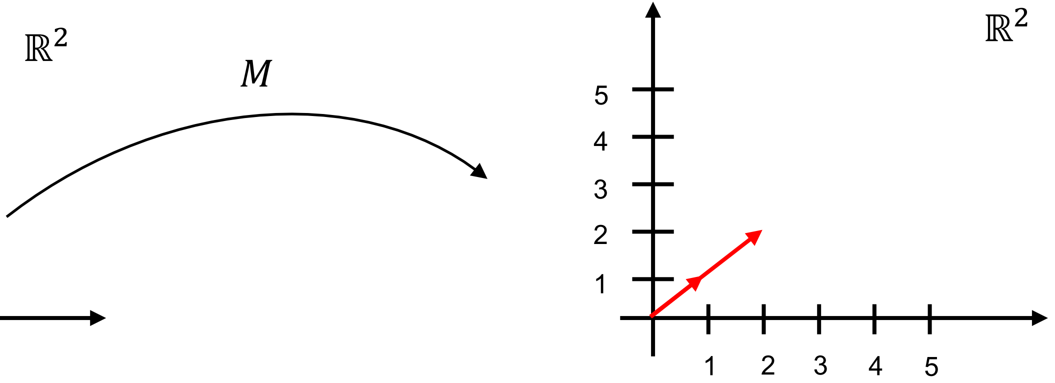
Invertible matrix



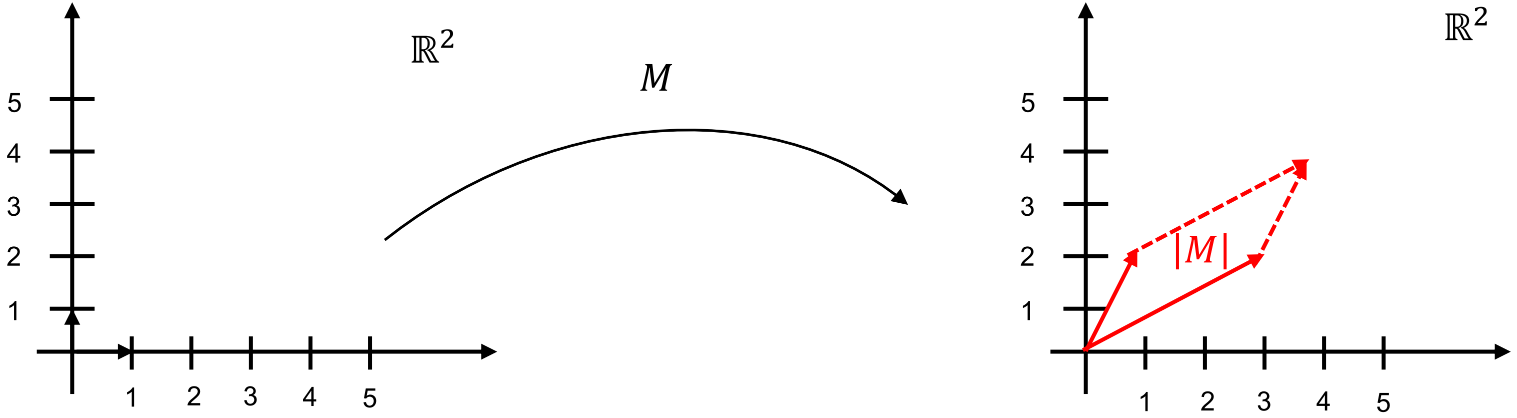
$$\begin{cases} 11 = \theta_0 + 2\times\theta_1 \\ 12 = \theta_0 + 2\times\theta_1 \end{cases} \Leftrightarrow \begin{pmatrix} 11 \\ 12 \end{pmatrix} = \begin{pmatrix} 1 & 2 \\ 1 & 2 \end{pmatrix} \cdot \begin{pmatrix} \theta_0 \\ \theta_1 \end{pmatrix}$$

$M = \begin{pmatrix} 1 & 2 \\ 1 & 2 \end{pmatrix}$ is not invertible because its column (and row) vectors are colinear





Determinant

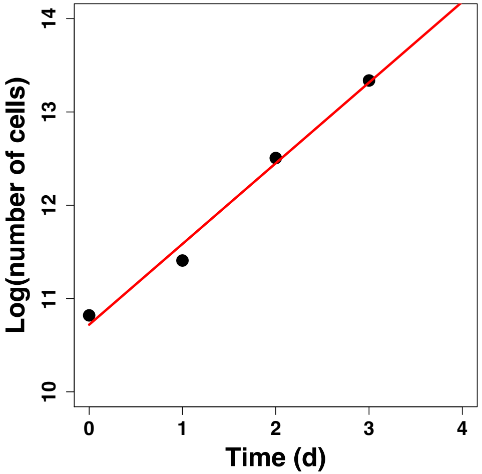


- The determinant of M, denoted $|M|$, is the area of the parallelogramm spanned by the column vectors of M
- For $M = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$ it is given by $ad - bc$.
- It can be generalized in any dimension and is a measure of the colinearity (and correlation) of the vectors
- $|M| \neq 0 \Leftrightarrow M$ is invertible $\Leftrightarrow$ the column (and row) vectors of M are independent

Linear regression

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

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



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

M rectangular: no solution (more equations than variables)

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

$\times M^T (\in M_{2,n})$

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

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

one unique solution

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

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

Linear least-squares: statistical properties

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

$$\hat{\theta}_{LS} = (M^T M)^{-1} M^T \cdot Y \quad \Leftrightarrow \hat{\theta}_{LS} = \underset{\theta \in \mathbb{R}^p}{\operatorname{argmin}} \parallel Y - M \cdot \theta \parallel^2 = \underset{\theta \in \mathbb{R}^p}{\operatorname{argmin}} \sum_{j=1}^{J} \left(y_j - (M \cdot \theta)_j\right)$$

Theorem:

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

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

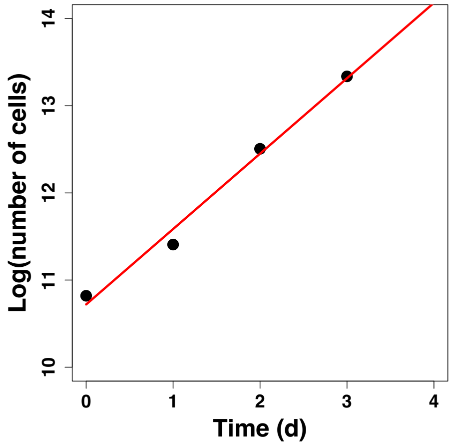
From this, standard errors and confidence intervals can be computed on the parameter estimates

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

Example: tumor growth

$$nb_j \simeq N_0 e^{\lambda t_j} = N_0 e^{\frac{\log(2)}{DT} t_j}$$

$$y = \theta_1 + \theta_2 t + \varepsilon$$



$$\ln(nb_j) \simeq \log(N_0) + \lambda t_j$$

$\hat{\lambda} = \widehat{\theta_2} = 0.865 \Longrightarrow \widehat{DT} = \frac{\log(2)}{\hat{\lambda}} = 19.2$ hours

$se(\widehat{\theta_2}) = 0.004, \quad \text{rse}(\widehat{\theta_2}) = \frac{se(\widehat{\theta_2})}{\widehat{\theta_2}} = 0.005 = 0.5\%$

$CI(DT) = (18.8, 19.7)$ hours

$s = 0.15$

Variance of the least-squares estimator

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

$$M = \begin{pmatrix} 1 & t_1 \\ \vdots & \vdots \\ 1 & t_n \end{pmatrix} \Longrightarrow M^T M = \begin{pmatrix} 1 & \cdots & 1 \\ t_1 & \cdots & t_n \end{pmatrix} \begin{pmatrix} 1 & t_1 \\ \vdots & \vdots \\ 1 & t_n \end{pmatrix} = \begin{pmatrix} n & \sum t_j \\ \sum t_j & \sum t_j^2 \end{pmatrix}$$

$$(M^T M)^{-1} = \frac{1}{n^2 Var(t)} \begin{pmatrix} \sum t_j^2 & -\sum t_j \\ -\sum t_j & n \end{pmatrix}$$

$$Var(\widehat{\theta_0}) = \sigma^2 \left(\frac{1}{n} + \frac{\bar{t}^2}{Var(t)^2} \right)$$

$$Var(\widehat{\theta_1}) = \frac{\sigma^2}{n \cdot Var(t)}$$

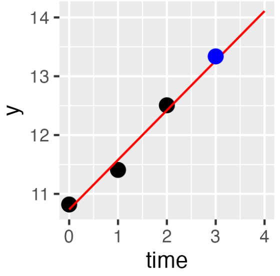
Impact of the number of time points

$$Var(\widehat{\theta_1}) = \frac{\sigma^2}{\mathbf{n} \cdot Var(t)}$$

3 time points

$\widehat{\theta_2} = 0.865, \quad \mathrm{rse}(\widehat{\theta_2}) = 2.58\%, \quad DT = 19.7\ h\ (14.9\ - 29.3)$

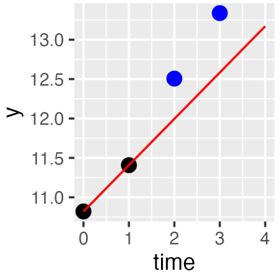
$s = 0.21$



2 time points

$\widehat{\theta_2} = 0.58, \quad \mathrm{rse}(\widehat{\theta_2}) = NaN, \quad DT = 28.3\ h\ (NaN\ - NaN)$

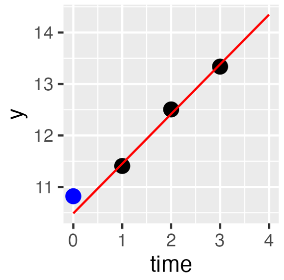
$s = NaN$



Impact of the sampling times

$$Var(\widehat{\theta_1}) = \frac{\sigma^2}{n \cdot \boldsymbol{Var(t)}}$$

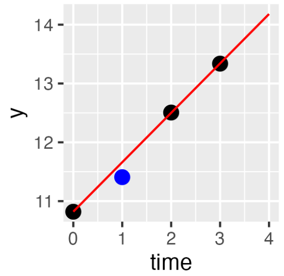
Close points (*Var(t)* low)



$$\widehat{\theta_2} = 0.965, \quad \text{rse}(\widehat{\theta_2}) = 0.62\%, \quad DT = 17.2\ h\ (16.0\ - 18.7)$$

$$s = 0.11$$

Dispersed times (*Var(t)* high)



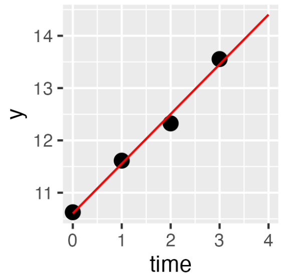
$$\widehat{\theta_2} = 0.840, \quad \text{rse}(\widehat{\theta_2}) = 0.01\%, \quad DT = 19.809\ h\ (19.806\ - 19.812)$$

$$s = 0.006$$

Impact of the residual error

$$Var(\widehat{\theta_1}) = \frac{\boldsymbol{\sigma^2}}{n \cdot Var(t)}$$

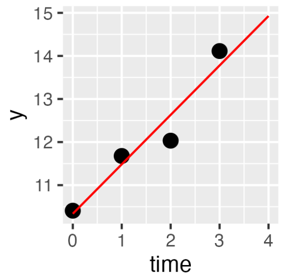
Low error ($\sigma = 0.15$)



$\widehat{\theta_2} = 0.950$, $\text{rse}(\widehat{\theta_2}) = 0.49\%$, $DT = 17.5\, h\ (17.2 - 17.9)$

$s = 0.152$

High error ($\sigma = 0.5$)



$\widehat{\theta_2} = 1.15$, $\text{rse}(\widehat{\theta_2}) = 4.47\%$, $DT = 12.2\, h\ (14.5 - 17.9)$

$s = 0.506$

Confidence interval and prediction interval

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

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

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

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

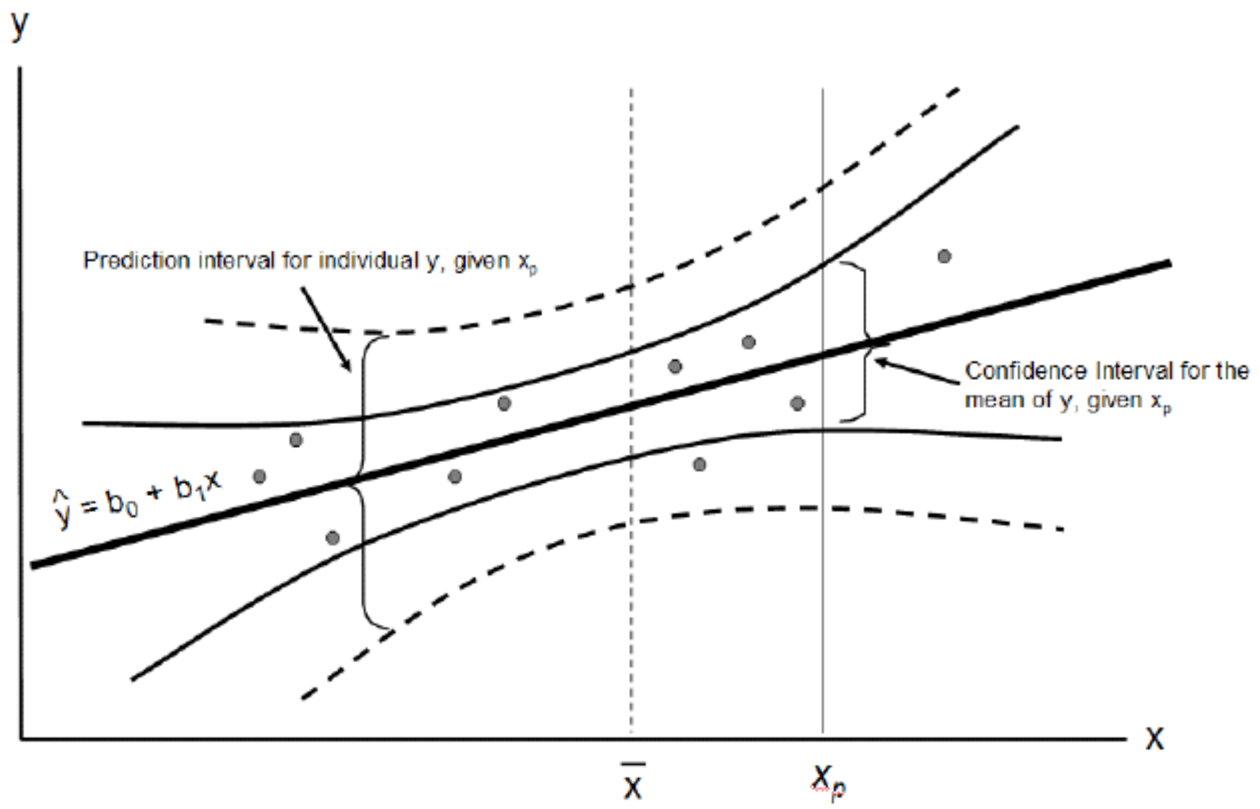
- Uncertainty on parameter estimate $\hat{\theta}$

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

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

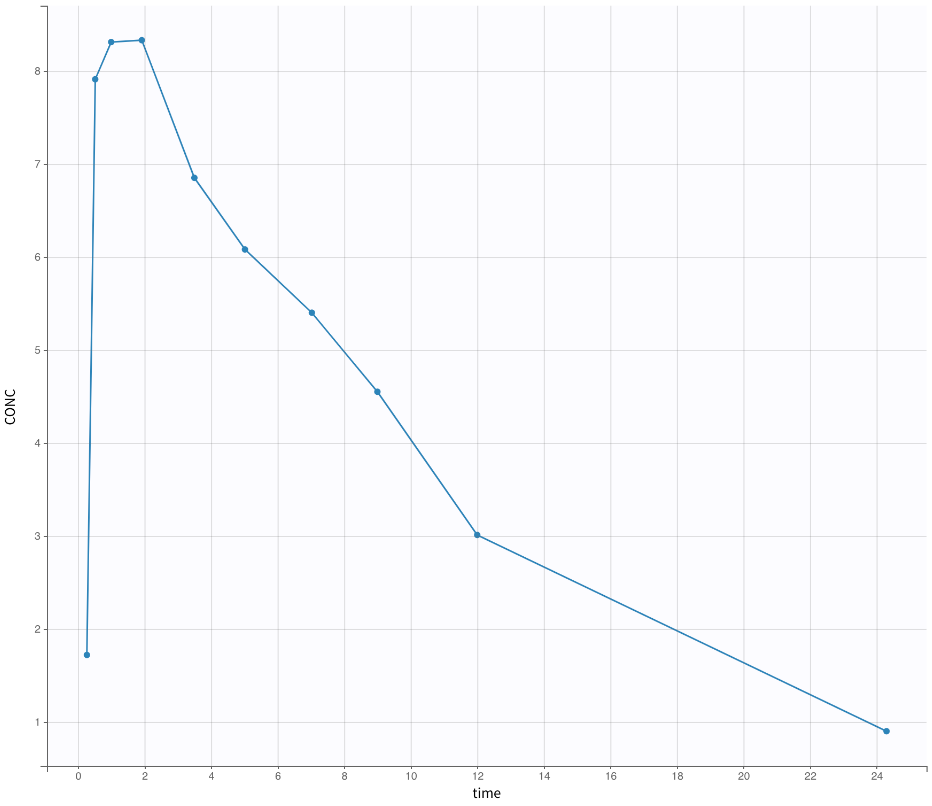
$$y_{new} \sim \mathcal{N}\left(\widehat{M_{new}}, Var\left(\widehat{M_{new}}\right) + \sigma^2 I\right)$$

Confidence interval vs prediction interval



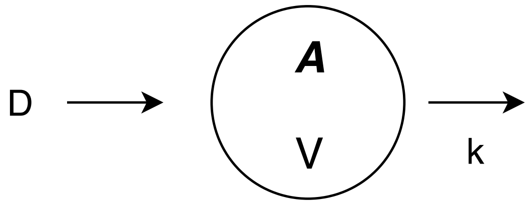
2. Nonlinear regression

Data



Structural model M : intravenous one-compartment model

$$\begin{cases} \frac{dA}{dt} = -kA \\ A(t=0) = D \end{cases} \qquad C(t) = \frac{A(t)}{V};$$



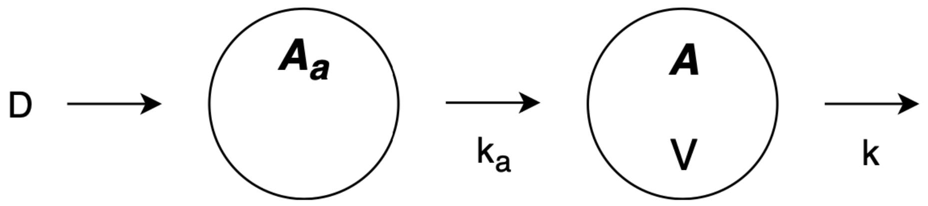
$$A(t) = De^{-k \cdot t}$$

$$M(t;\theta) = C(t) = \frac{D}{V}e^{-kt}, \qquad \theta = (V,k)$$

Structural model M : One-compartment model with absorption

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

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



$$\begin{pmatrix} A_a{}' \\ A' \end{pmatrix} = \begin{pmatrix} -k_a & 0 \\ k_a & -k \end{pmatrix} \cdot \begin{pmatrix} A_a \\ A \end{pmatrix}$$

(Eigenvalues $-k_a$ and $-k$)

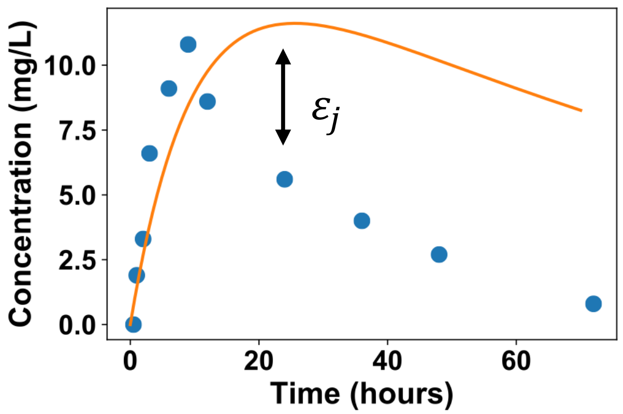
$$A_a(t) = De^{-k_a t}$$

$$A(t) = D\frac{k_a}{k_a - k}\left(e^{-kt} - e^{-k_a t}\right)$$

$$M(t;\theta) = C(t) = \frac{A(t)}{V} = \frac{D}{V}\frac{k_a}{k_a - k}\left(e^{-kt} - e^{-k_a t}\right), \theta = (V, k_a, k)$$

Nonlinear regression

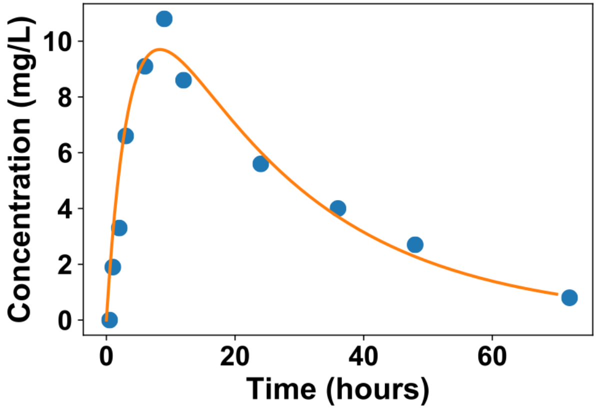
$$Y = M(t;\theta^*) + \varepsilon$$



How??

Nonlinear regression

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



$$\hat{\theta}_{LS} = \underset{\theta \in \mathbb{R}^p}{\operatorname{argmin}} \sum \left(y_j - M(t_j;\theta)\right)^2$$

Why??

Likelihood maximization

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

The likelihood is defined by

$$L(\theta) = p(y_1, \cdots, y_n | \theta)$$

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

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

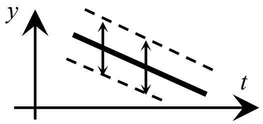
Error model: ε

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

- ε is what drives the **randomness** in the data
- It corresponds to the **intra-individual** variability
- It is also associated to the **measurement error**
- Gaussian (normally-distributed) error: $\varepsilon_j = \sigma_j \cdot \epsilon, \quad \epsilon \sim \mathcal{N}(0,1)$

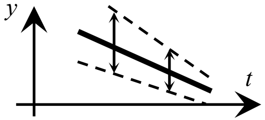
Constant (homoscedastic)

$$\sigma_j = a$$



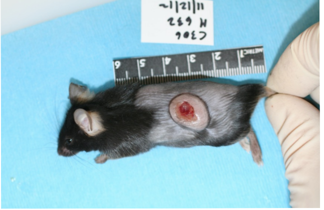
Proportional (heteroscedastic)

$$\sigma_j = bM(t_j, \hat{\theta})$$

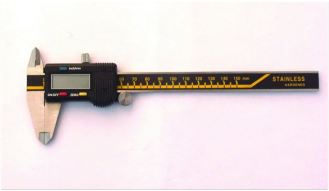


Combined (heteroscedastic)

$$\sigma_j = a + bM(t_j, \hat{\theta})$$



Error models for tumor volume

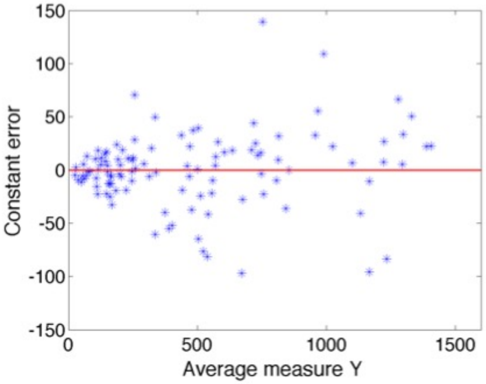


$$\varepsilon_j \ i.i.d \ \mathcal{N}(0, \sigma_j)$$

Constant

$$\sigma_j = a, \qquad \forall j$$

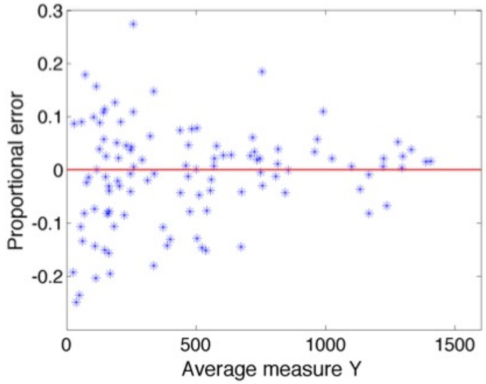
$p = 0.004$



Proportional

$$\sigma_j = b\, M(t_j, \hat{\theta}\,)$$

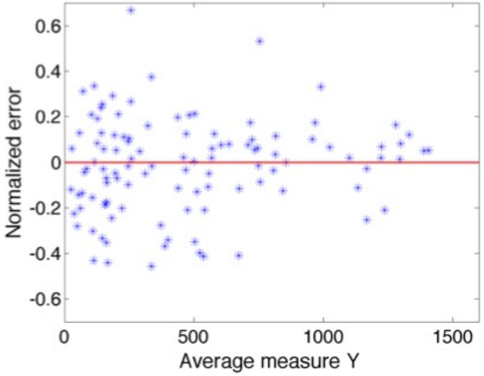
$p = 0.083$



Specific

$$\sigma_i = \begin{cases} \sigma M\left(t_j, \hat{\theta}\right)^{\alpha}, & M\left(t_j, \hat{\theta}\right) \geq V_m \\ \sigma V_m^{\alpha}, & M\left(t_j, \hat{\theta}\right) < V_m \end{cases}$$

$p = 0.2$



Maximum likelihood ⇔ Least-squares (constant error)

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

Independence assumption

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

$$p(y_j | \theta, a) = \frac{1}{a\sqrt{2\pi}} e^{-\frac{\left(y_j - M\left(t_j, \theta\right)\right)^2}{2a^2}}$$

$$L(\theta, a) = \frac{1}{\left(a\sqrt{2\pi}\right)^n} e^{-\frac{\sum_j \left(y_j - M\left(t_j, \theta\right)\right)^2}{2a^2}}$$

Maximize $L(\theta, a)$ ⇔ minimize $F(\theta, a) = -\log(L(\theta, a))$

$$F(\theta, a) = n\log\left(a\sqrt{2\pi}\right) + \frac{\sum_j \left(y_j - M\left(t_j, \theta\right)\right)^2}{2a^2}$$

$$\frac{\partial F}{\partial a}(\hat{\theta}, \hat{a}) = 0 \Rightarrow \hat{a} = \frac{1}{n} \sum_j \left(y_j - M\left(t_j, \hat{\theta}\right)\right)^2$$

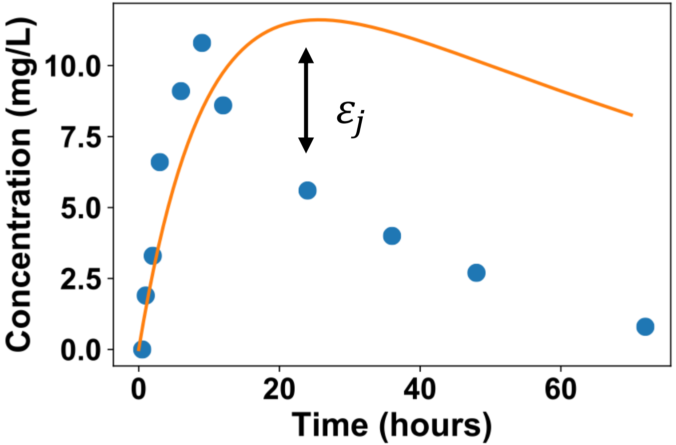
$$\Rightarrow \hat{\theta} = \underset{\theta}{\operatorname{argmin}} \sum_j \left(y_j - M\left(t_j, \theta\right)\right)^2$$

3. Uncertainty quantification: standard errors

Nonlinear regression: least-squares

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

$$\hat{\theta}_{LS} = \underset{\theta \in \mathbb{R}^p}{\operatorname{argmin}} \sum \left(y_j - M(t_j;\theta)\right)^2$$

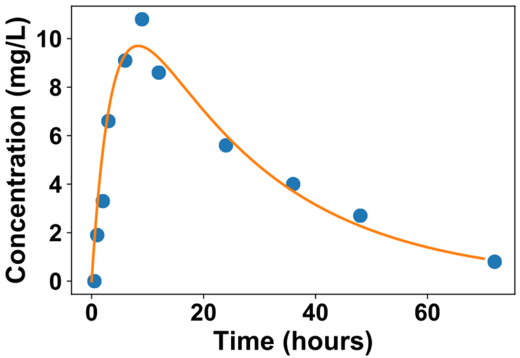


Linearization: $M(t,\theta) = M(t,\theta^*) + J \cdot (\theta - \theta^*) + o(\theta - \theta^*), \quad J = D_\theta M(t,\theta^*)$

Nonlinear regression: least-squares

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

$$\hat{\theta}_{LS} = \underset{\theta\in\mathbb{R}^p}{\operatorname{argmin}} \sum \left(y_j - M(t_j;\theta)\right)^2$$



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

Proposition:

Assume $\varepsilon \sim \mathcal{N}(0, a^2 I)$. Then, for large n, approximately

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

⇒ standard errors, confidence intervals

Sensitivity matrix

$$y_j = M\left(t_j, \theta^*\right) + \varepsilon_j \Leftrightarrow \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} = \begin{pmatrix} M(t_1, \theta^*) \\ \vdots \\ M(t_n, \theta^*) \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix} \Leftrightarrow y = M(t, \theta^*) + \varepsilon$$

- Suppose we have a first guess θ_0 close to θ^*.

$$M(t, \theta^*) \simeq M(t, \theta_0\,) + D_\theta M(t, \theta_0) \cdot (\theta^* - \theta_0)$$

- If there is an exact solution ($\varepsilon = 0$)

$$y \simeq M(t, \theta_0\,) + J \cdot (\theta^* - \theta_0)$$

$$\theta^* \simeq \theta_0 + J^{-1} \cdot \left(y - f(t, \theta_0)\right)$$

- In general, $\varepsilon \neq 0$ and $n > p \Rightarrow$ least-squares

$$\hat{\theta} = (J^T \cdot J)^{-1} \cdot J^T \cdot \left(y - M(t, \theta_0)\right)$$

p

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

n

= Sensitivity matrix

Components of uncertainty: data + model sensitivity

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

- a = measurement error = uncertainty on the data
- $J^T J \simeq J^2, \; J = D_\theta M(t, \theta)$
- $(J^T J\,)^{-1}$ small $\Leftrightarrow$ $J^T J$ high $\Leftrightarrow$ $D_\theta M(t,\theta)$ high $\Leftrightarrow$ M is highly **sensitive** to θ

Asymptotic properties of the maximum likelihood estimator

$Y = M(t;\theta^*) + \varepsilon$ ← **no assumption on ε**

$L(\theta) = p(y_1, \cdots, y_n|\theta)$

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

Proposition

Under regularity assumptions on L, *when* $\mathrm{n} \to +\infty$

1. $\hat{\theta}_{\mathrm{MV}} \longrightarrow \theta^*$ *(consistency)*

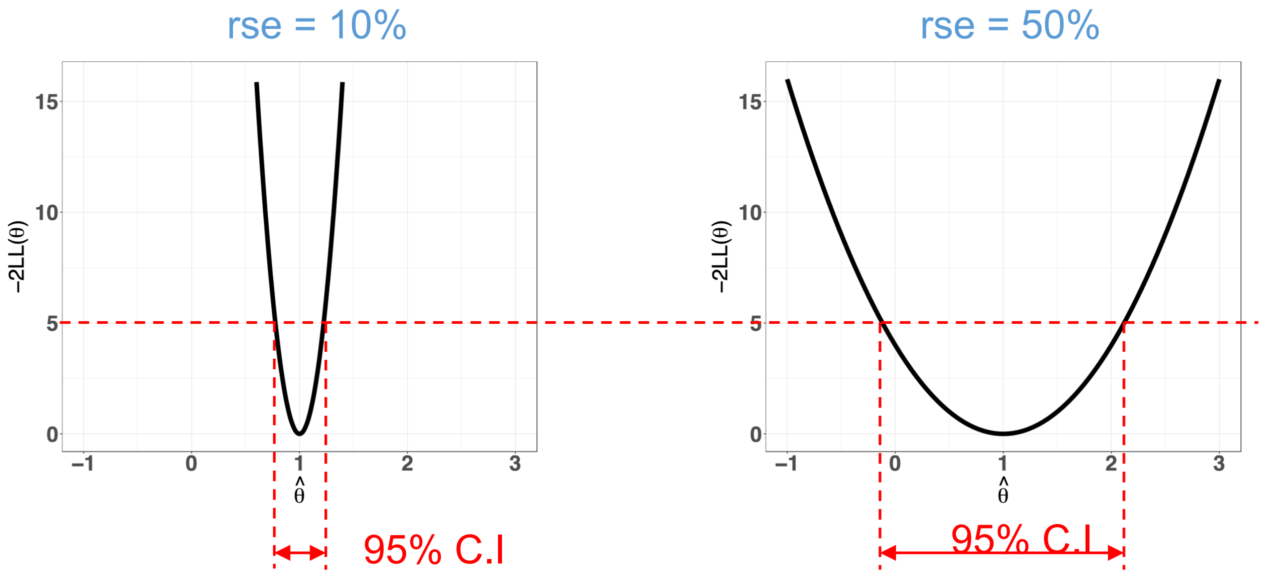
2. $\hat{\theta}_{\mathrm{MV}}$ *is asymptotically of minimal variance (it reaches the Cramér-Rao bound):*

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

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

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

Precision of the estimates



Math interlude: two-dimensional distribution

1D gaussian distribution

- One complexity step beyond linearity: quadratic relationship

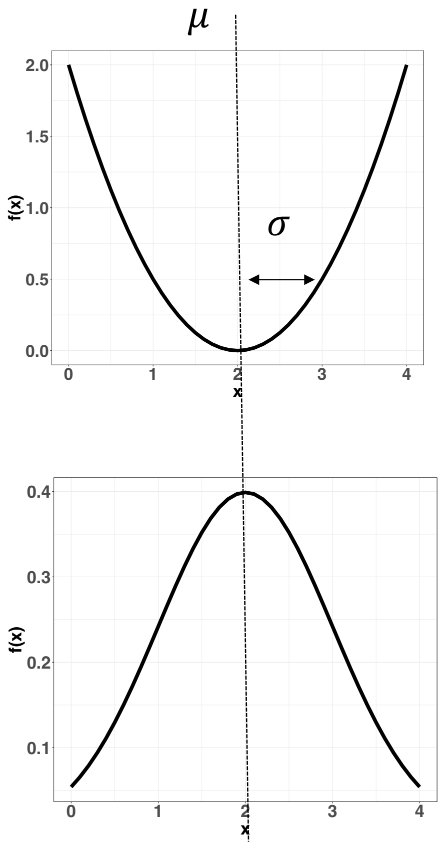
$$f: \begin{array}{ccc} \mathbb{R} & \rightarrow & \mathbb{R} \\ x & \mapsto & ax^2 \end{array}$$

$$f(x) = \frac{(x-\mu)^2}{2\sigma^2}$$

- Normal distribution

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

in 2D??

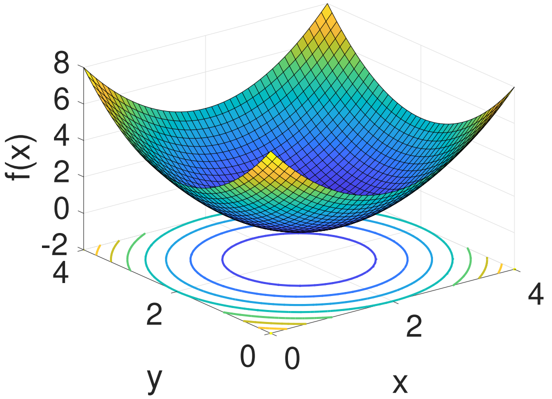


Quadratic form 2D: matrix form

- Quadratic form in $\mathbb{R}^2$

$$f: \begin{array}{ccc} \mathbb{R}^2 & \rightarrow & \mathbb{R} \\ x = (x_1, x_2) & \mapsto & ax_1^2 + 2bx_1x_2 + cx_2^2 \end{array}$$

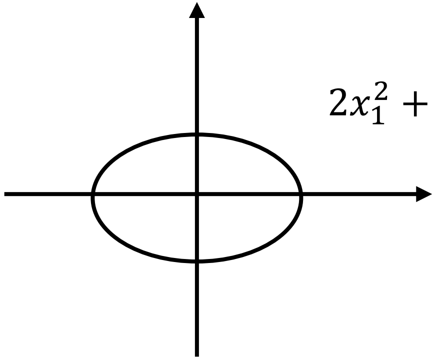
$$f(x) = (x_1, x_2) \cdot \begin{pmatrix} a & b \\ b & c \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = x^T \cdot M \cdot x$$



- M is a symmetric matrix

- If M is diagonal

$$(x_1, x_2) \cdot \begin{pmatrix} 2 & 0 \\ 0 & 3 \end{pmatrix} \cdot \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} = 2x_1^2 + 3x_2^2$$



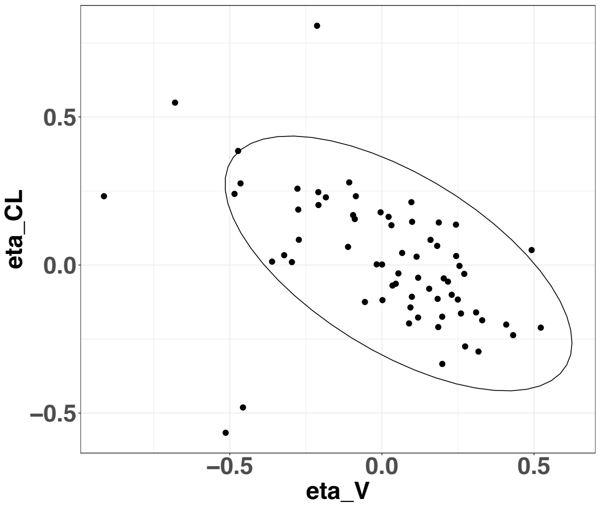
$$2x_1^2 + 3x_2^2 = C$$

Covariance/correlation matrix

- Two (or more) variables x and y (ex: V and CL)

$$\Sigma = \begin{pmatrix} Var(x) & Cov(x,y) \\ Cov(x,y) & Var(y) \end{pmatrix}$$

$$Cov(x,y) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$$



- Correlation matrix

$$R = \begin{pmatrix} 1 & r(x,y) \\ r(x,y) & 1 \end{pmatrix}$$

$$r(x,y) = \frac{Cov(x,y)}{\sigma_x \sigma_y}$$

note: $\widehat{\theta_1} = r(x,y)\frac{\sigma_y}{\sigma_x}$

Multivariate normal distribution

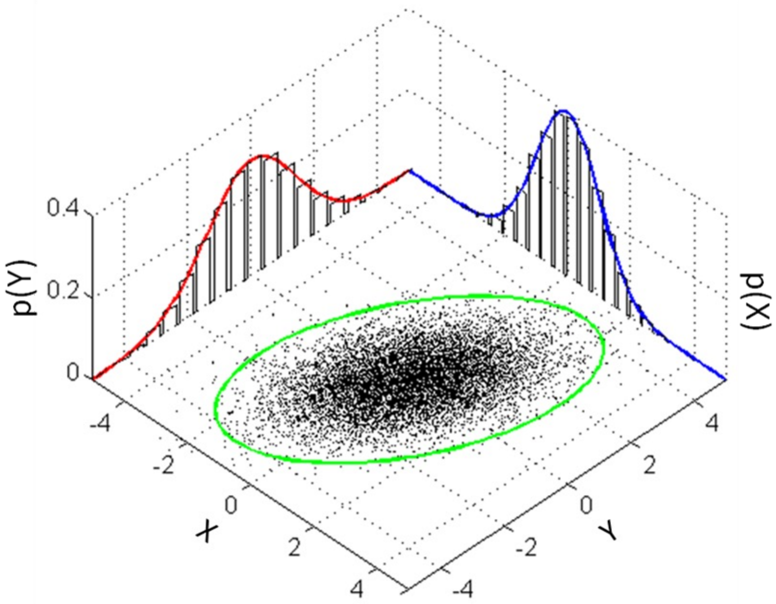
- Generalization of the normal distribution in dimension n

$$x \to \boldsymbol{x} = (x, y), \mu \to \boldsymbol{\mu} = (\mu_1, \mu_2)$$

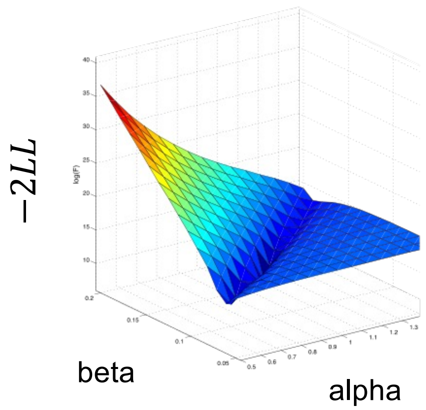
$$\sigma^2 \to \Sigma = \begin{pmatrix} Var(x) & Cov(x,y) \\ Cov(x,y) & Var(y) \end{pmatrix}$$

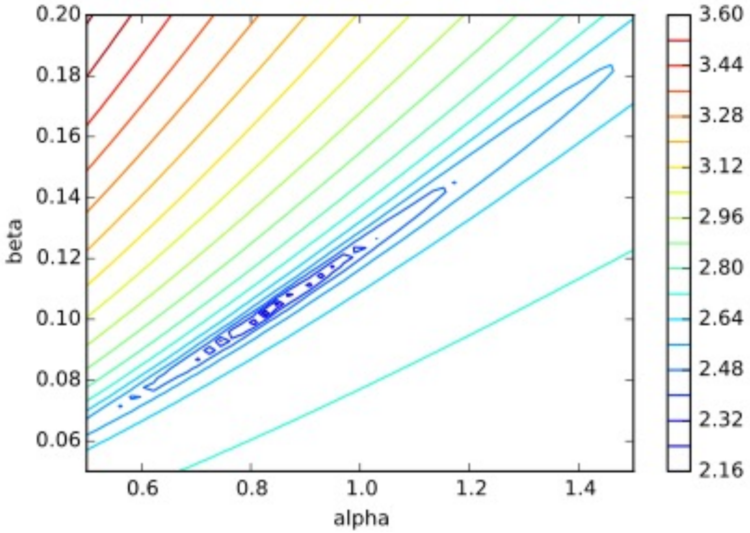
$$\frac{(x-\mu)^2}{2\sigma^2} \to \frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T \Sigma^{-1} (\boldsymbol{x}-\boldsymbol{\mu})$$

$$f(x) = \frac{1}{2\pi\sqrt{|\Sigma|}} e^{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T \Sigma^{-1} (\boldsymbol{x}-\boldsymbol{\mu})}$$



Correlation between estimates





Correlation matrix of the estimates

	R.S.E.(%)			
alpha_pop	3.09	1		
beta_pop	5.68	**0.98574**	1	
b	23.8	-0.00055974	0.022018	1

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

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

Precision vs. dispersion

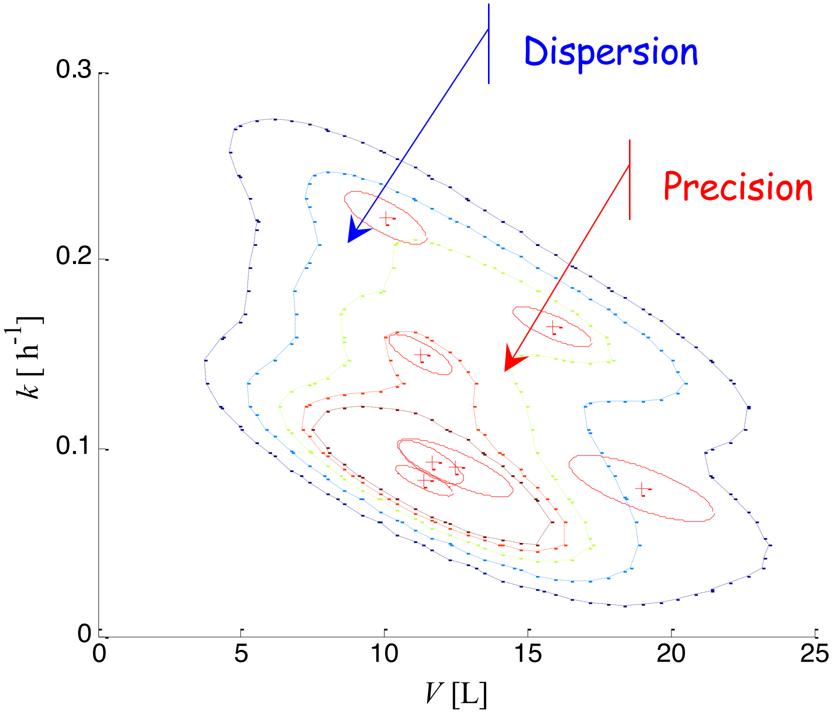
- 7 individuals – 2 overlap
- Intra- vs. inter-individual variability

Dispersion (standard deviation, std)

is conditioned by

precision (standard error, sem)

- Precision and dispersion are described by the variance – covariance matrix



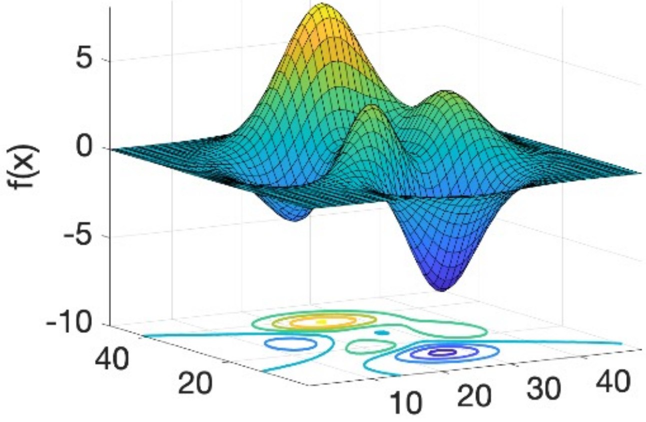
4. Minimization: theory

Minimization problem

$$\hat{\theta} = \underset{\theta}{\operatorname{argmin}} \sum_j \left(y_j - M(t_j, \theta\ )\right)^2$$

$$\hat{\theta} = \underset{\theta=(V,k_a,k)}{\operatorname{argmin}} \sum_j \left(y_j - \frac{D}{V}\frac{k_a}{k_a - k}\left(e^{-kt_j} - e^{-k_a t_j}\right)\right)^2$$

$$\hat{\theta} = \underset{\theta=(\theta_1,\theta_2,\theta_3)}{\operatorname{argmin}} J(\theta_1, \theta_2, \theta_3)$$



θ_1 θ_2

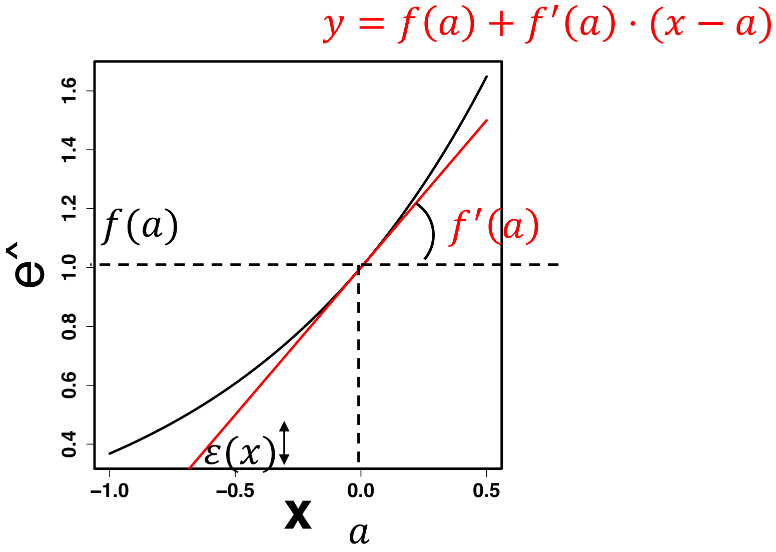
Derivative

- Linear approximation of a (nonlinear) function $\mathbb{R} \to \mathbb{R}$, in the neighborhood of a

$$f(x) = f(a) + f'(a) \cdot (x - a) + \varepsilon(x)$$

- Formally

$$f'(a) = \lim_{x \to a} \frac{f(x+a) - f(a)}{x - a}$$



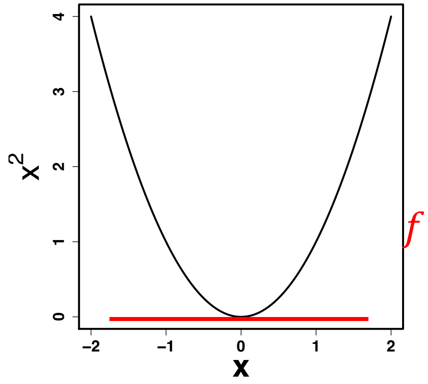
Extremum. Necessary condition

- If f has a minimum or maximum in a

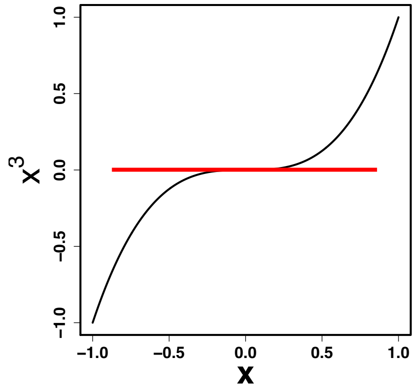
$$f(x) \simeq f(a) + f'(a) \cdot (x - a)$$

$\Rightarrow f'(a) = 0$

- But this is not a sufficient condition



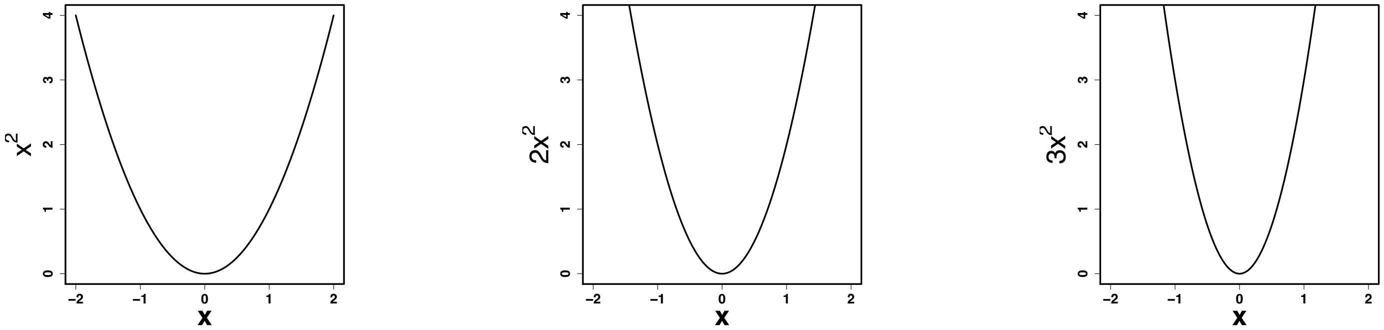
$f'(0) = 0$



$f'(0) = 0$

Quadratic approximation. Second derivative

- Linear function $f(x) = \theta_0 + \theta_1 x \Leftrightarrow$ line characterized by slope
- Quadratic function $f(x) = \theta_0 + \theta_1 x + \theta_2 x^2 \Leftrightarrow$ characterized by curvature



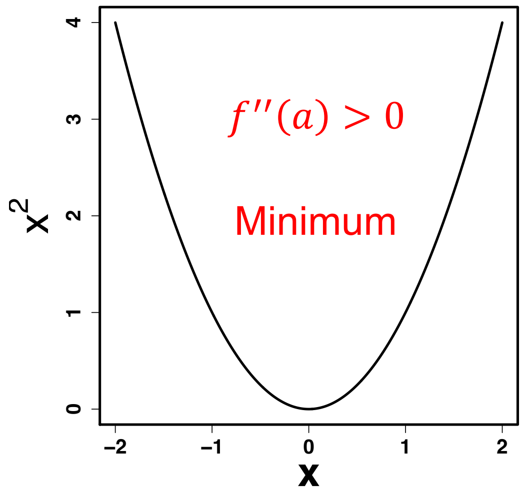
- Same slope as f and same curvature. Taylor's formula

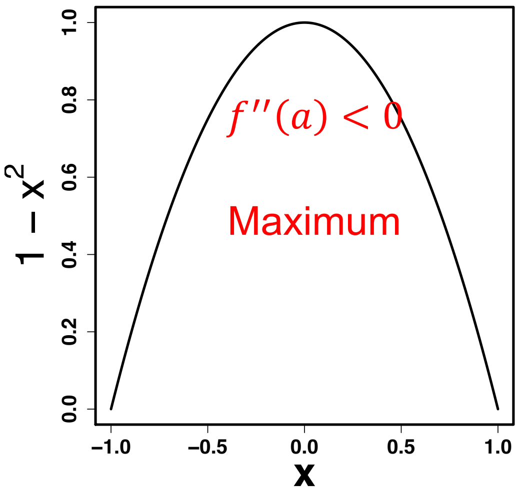
$$f(x) = f(a) + f'(a) \cdot (x - a) + \frac{f''(a)}{2} \cdot (x - a)^2 + \varepsilon(x)$$

Extremum. Sufficient condition

$$f(x) \simeq f(a) + f'(a) \cdot (x-a) + \frac{f''(a)}{2} \cdot (x-a)^2$$

- a extremum





$f: \mathbb{R}^n \to \mathbb{R}$. Gradient

- Linear function $\mathbb{R}^n \to \mathbb{R} \Rightarrow \langle a, x \rangle = a_1 x_1 + \cdots + a_n x_n$

$$\boldsymbol{f(x) \simeq f(a) + \langle \nabla f(a), x - a \rangle = f(a) + \frac{\partial f}{\partial_{x_1}}(a) \cdot (x_1 - a_1) + \cdots + \frac{\partial f}{\partial_{x_n}}(a) \cdot (x_n - a_1)}$$

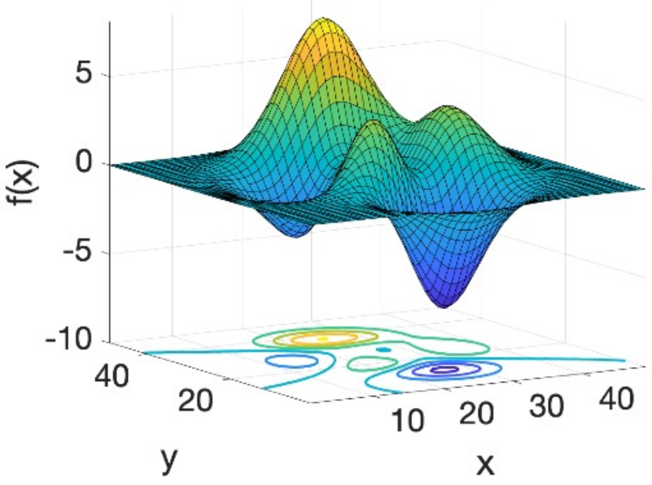
- $\nabla f(a) = \left(\frac{\partial f}{\partial x_1}(a), \cdots, \frac{\partial f}{\partial x_n}(a) \right) \in \mathbb{R}^n$

- $\frac{\partial f}{\partial_{x_i}}(a)$ is called the partial derivative of f in a in the direction x_i

Gradient descent algorithms

- f has a minimum (or maximum) in $a \Rightarrow \boldsymbol{\nabla f(a) = 0}$, i.e. $\frac{\partial f}{\partial x_i}(a) = 0 \quad \forall i$

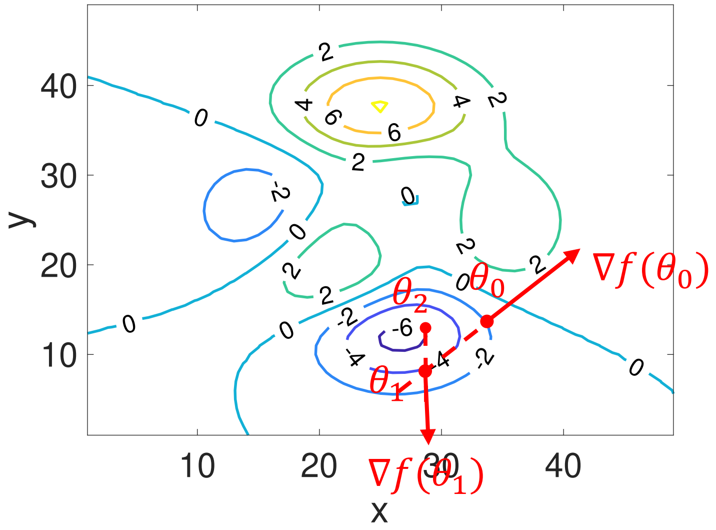
$\Rightarrow$ To minimize f $\qquad \theta_{n+1} = \theta_n - \lambda \nabla f(\theta_n)$



$\nabla f(\theta_n) =$

steepest descent

direction



$f: \mathbb{R}^n \to \mathbb{R}$. Hessian matrix

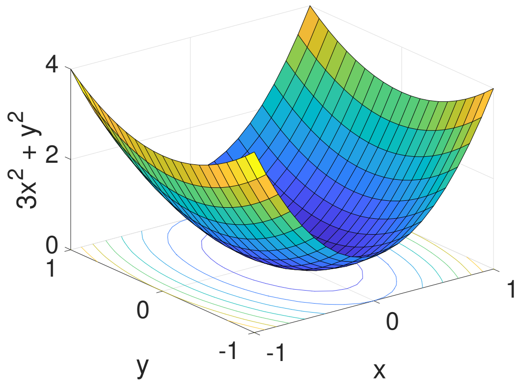
- Second derivative?
- Quadratic form: $\begin{array}{ccc} \mathbb{R}^n & \to & \mathbb{R} \\ x & \mapsto & x^T \cdot M \cdot x \end{array}$, M symmetric matrix
- Matrix of second partial derivatives = Hessian matrix

$$H = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2} & \cdots & \frac{\partial^2 f}{\partial x_1 \partial x_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_n \partial x_1} & \cdots & \frac{\partial^2 f}{\partial x_n^2} \end{pmatrix}$$

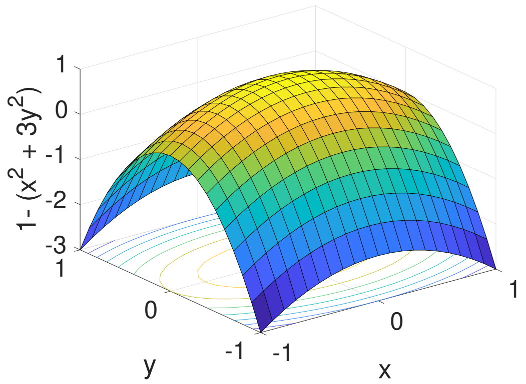
- $\frac{\partial f}{\partial x_i \partial x_j} = \frac{\partial f}{\partial x_j \partial x_i} \Rightarrow H$ is symmetric

Curvature

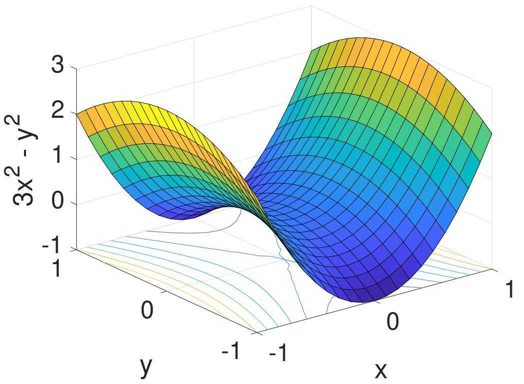
- The Hessian matrix extends the notion of (local) curvature to $f: \mathbb{R}^n \to \mathbb{R}$
- The eigenvectors give the principal directions and associated eigenvalues the curvatures



$$H = \begin{pmatrix} 6 & 0 \\ 0 & 1 \end{pmatrix}$$



$$H = \begin{pmatrix} -1 & 0 \\ 0 & -6 \end{pmatrix}$$

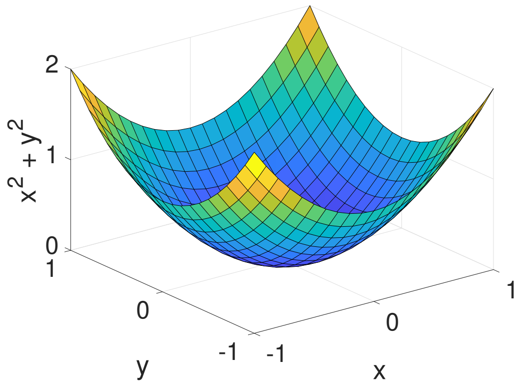


$$H = \begin{pmatrix} 6 & 0 \\ 0 & -1 \end{pmatrix}$$

Extremum in dimension n

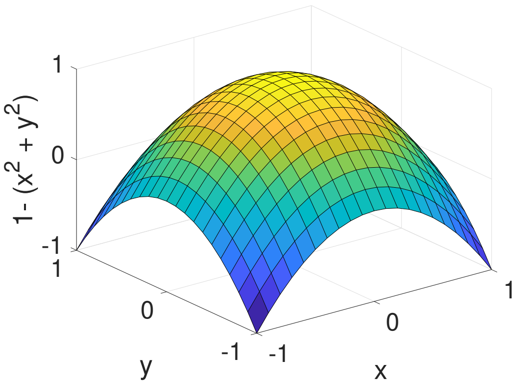
$$f(x) = f(a) + \langle \nabla f(a), x - a \rangle + \frac{1}{2}(x-a)^T \cdot H(a) \cdot (x-a) + \varepsilon(x)$$

Minimum



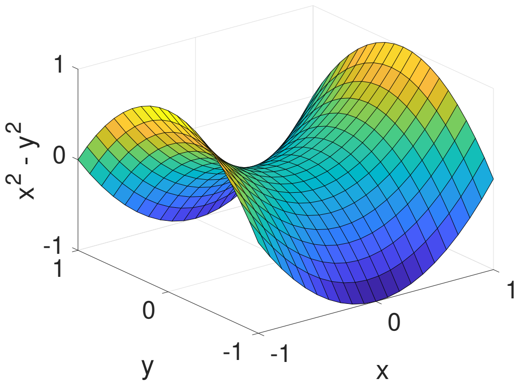
H positive definite

Maximum



H negative definite

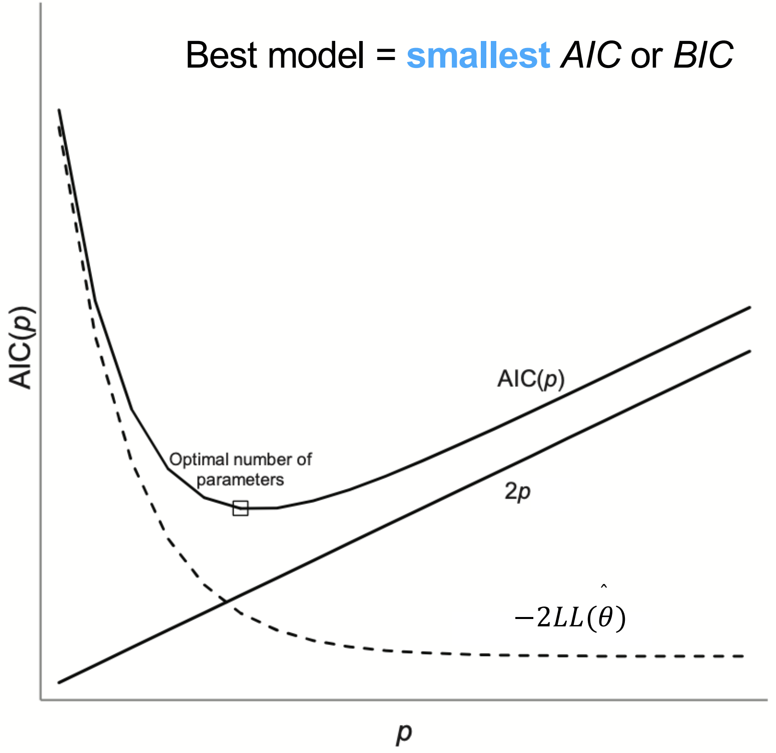
Saddle



- Note: Hessian of the objective function (-2LL) = R matrix in NONMEM

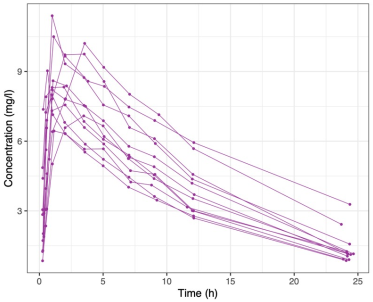
Information criteria

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

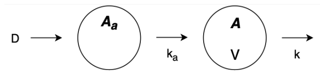


Population modeling: the two-steps approach

Population data



Individual structural model



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

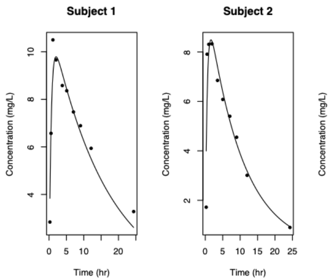
Individual fits

$$Y^1 = M(t;\theta^1) + \varepsilon$$

$$Y^2 = M(t;\theta^2) + \varepsilon$$

$$\vdots$$

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



$$\rightarrow \widehat{\theta^1}, \dots, \widehat{\theta^N}$$

$$\hat{\theta}_{pop} = \frac{1}{N}\sum_{i=1}^{N} \hat{\theta}_i$$

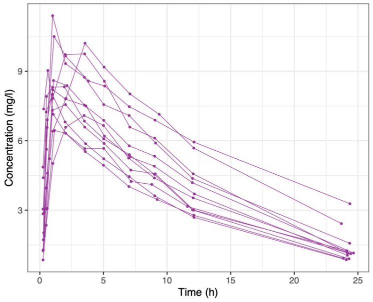
$$\widehat{\Omega} = Var - Cov(\hat{\theta}_i)$$

$$\mathcal{N}(\widehat{\boldsymbol{\theta}}_{pop}, \widehat{\boldsymbol{\Omega}})$$

Population model

Population modeling: mixed-effects approach

Population data



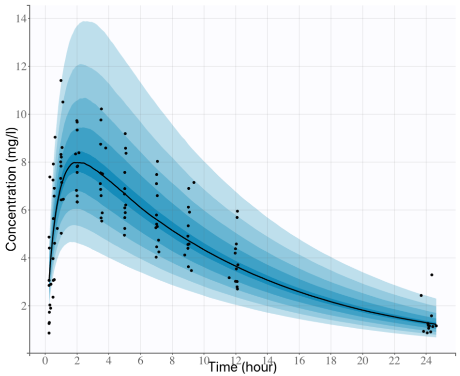
Population model

$$\theta^i = \theta_{pop} + \eta^i, \eta^i \sim \mathcal{N}(0, \Omega)$$

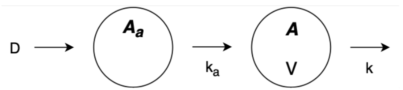
fixed effects

random effects

Population fit (MLE)

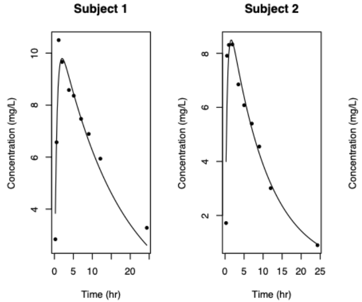


Individual structural model



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

Individual fit



Optimization algorithms for NLME

- FO = First Order. Linearizes for small values of random effects and residual error
  Sheiner, Comput Biomed Res, 1972, Sheiner and Beal, J Pharmacokinet Biopharm, 1980

⇒ Fast, but inaccurate for large ω or large residual error

- FOCE (NONMEM) = First-Order Conditional Estimation. Improvement over FO in terms of estimation of ω.
  Lindstrom and Bates, Biometrics, 1990
- FOCE-I = FOCE with interaction between intra- (ε) and inter- (η) variability.

⇒ to be used when proportional (or combined) error model but more computationally demanding

- LAPLACE = Same as FOCE except second-derivative (Hessian) approximation. Used for nonnormal (or lognormal) densities.
  Wolfinger, Biometrika, 1993
- SAEM (NONMEM, Monolix) = Stochastic Approximation of Expectation-Maximization
  Delyon, Lavielle, Moulines, Annals of Statistics, 1999

⇒ Slower, but converges better to global minimum

Plan et al., AAPS J, 2012

Bauer, CPT: PSP, 2019

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.

- Didactic and beautiful illustration of mathematical concepts: 3Blue1Brown on youtube

  https://www.youtube.com/@3blue1brown