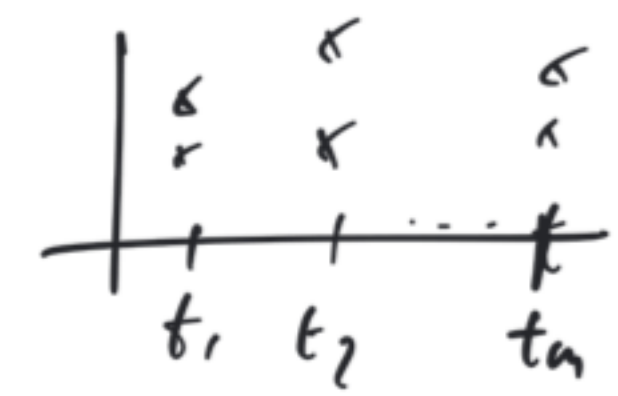
PARAMETER ESTIMATION

- PARAMETRIC MODEL FAMILY, M_θ, $\theta = \theta_1, \dots, \theta_k$ PARAMETERS $\theta_j \in \mathbb{R}_{\geq 0}$

  $\forall M_\theta$, LET $x(t|\theta)$ BE THE STATE OF M_θ AT TIME t.

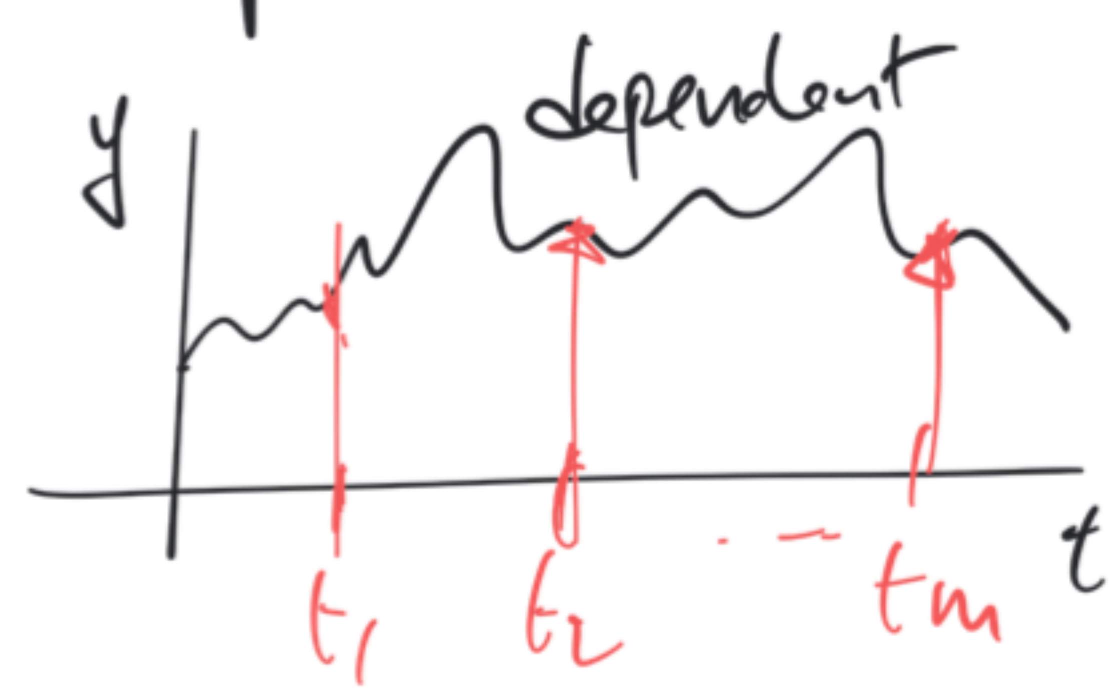
  (M_θ can be ODE, or CTMC, a STOCHASTIC HYBRID MODEL, ...)

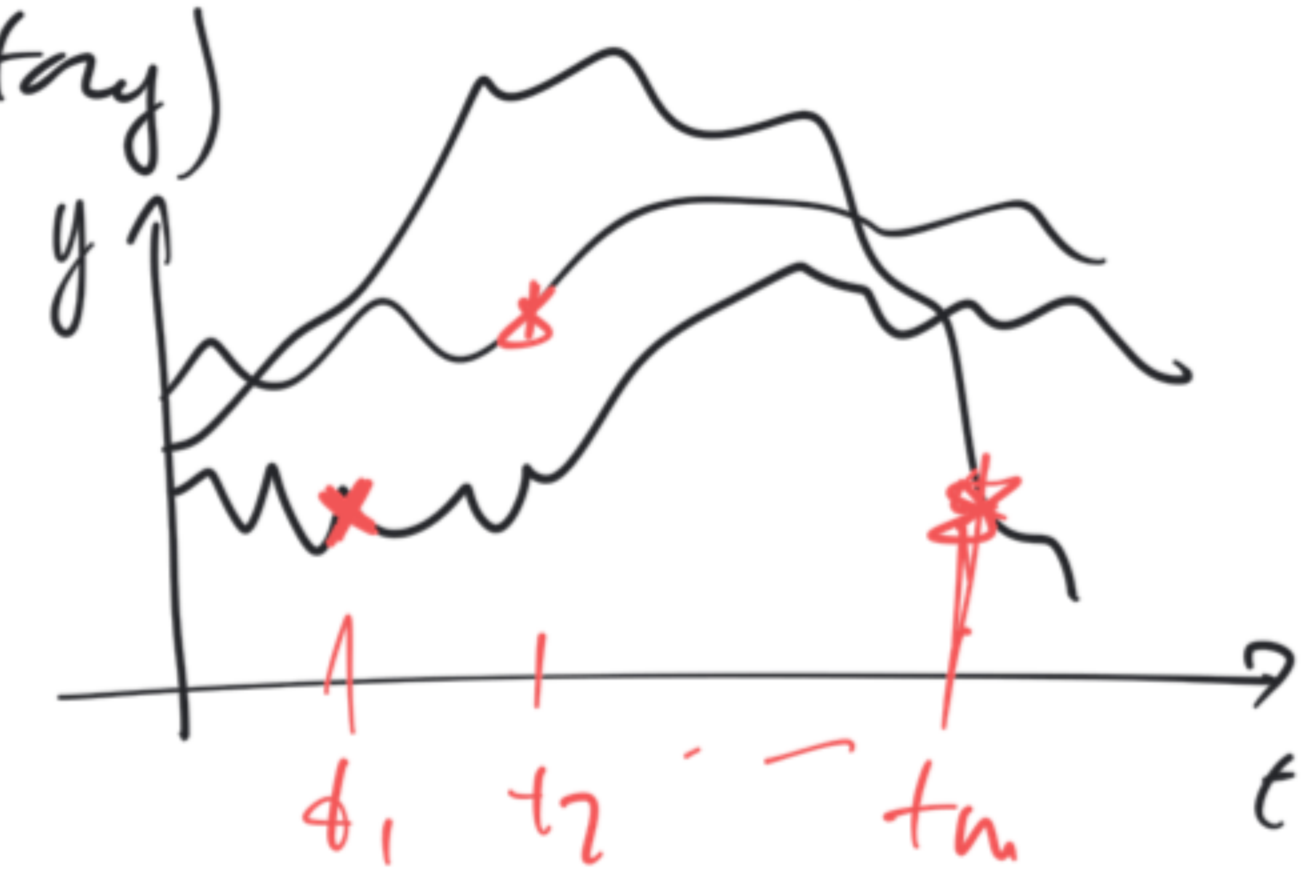
- OBSERVATIONS $y_j(t_i)$, $\underbrace{j = 1, \dots, N \text{ indip}}$, $\underline{i = 1, \dots m \text{ times}}$



  FOR FIXED θ, $y_j(t_1), \dots, y_j(t_m)$ can be either independent or dependent (from the same trajectory)

-





- MODEL OF EXPERIMENTAL NOISE (measurement errors)

  ↳ GAUSSIAN NOISE $\mathcal{N}(0, \sigma)$ $\Rightarrow$ $\underline{y_j(t_i) \sim x(t_i, \theta_{true}) + \varepsilon}$, $\varepsilon \sim \mathcal{N}(0, \sigma)$

GOAL: FIND $\theta^* \in \underbrace{\Theta}_{\text{parameter space}}$ s.t. $x(t, \theta^*)$ is "as close as possible" to $y_j(t_i)$

OBSERVATIONS

- AVERAGE CONCENTRATION OF PROTEIN P AT A FIXED t_i ACROSS A CELL POPULATION
  $\Rightarrow$ $y_i(t_1)$ ARE TIME INDEPENDENT, and they are averages. $\leftarrow$ stoch. or determ. models

- SINGLE CELL DATA OF P, AT A FIXED TIME
  $\Rightarrow$ $y_1(t_1)$ ARE TIME INDEP., but of a single stochastic run

- SINGLE CELL OBSERVED AT DIFFERENT TIME POINTS $t_1, \dots, t_m$
  $\Rightarrow$ $y_1(t_i)$ are time DEPENDENT

stochastic models

---

2 WAYS: MAXIMUM LIKELIHOOD OR BAYESIAN INFERENCE.

LIKELIHOOD: $f(Y \mid \theta) : \Theta \longrightarrow \mathbb{R}$

function of θ

$$f(Y \mid \theta) = P(Y \mid M_\theta) = \prod_{j=1}^{N} P(y_j(t_1), \dots, y_j(t_m) \mid M_\theta)$$

MAXIMUM LIKELIHOOD

$$\boxed{\theta^* = \arg\max_\theta f(Y \mid \theta)}$$

LIKELIHOOD FOR ODE (RRE)

1) AVERAGE DATA (indip. wrt time)

$$P(y_j(t_1), \to y_j(t_m) \mid \vartheta) = \underbrace{\prod_i P(y_j(t_i) \mid \vartheta)}$$

Problem: compute $P(y_j(t_i) \mid \vartheta)$

$x(t_i) = x(t_i, \vartheta)$

$$P(\underbrace{y_j(t_i)}, \underbrace{x(t_i)} \mid \vartheta) = P(y_j(t_i) \mid x(t_i), \vartheta) \cdot \underbrace{P(x(t_1) \mid \vartheta)}_{\delta_{\underbrace{x(t_i,\vartheta)}}}$$

SOLUTION OF ODE WITH PARAM. ϑ AT TIME t_i.

$= P(y_j(t_1) \mid x(t_i, \vartheta))$ ~ this is the obs. noise

$= \underbrace{\mathcal{N}(x(t_i,\vartheta), \sigma^2)}_{1\text{-dim}}$ $\left[\text{or } \underbrace{\mathcal{N}(x(t_1,\vartheta), \Sigma)}_{n\text{-dim}}, \ \Sigma = \sigma^2 I_n \right]$

$\propto \exp\left(-(y_j(t_i) - x(t_i,\vartheta))^2 / 2\sigma^2\right)$

let's take the log!

$$L(\theta) = \log f(D \mid \theta)$$

$$= \sum_{j=1}^{N} \sum_{i=1}^{m} -\left(y_j(t_i) - x(t_i, \theta)\right)^2 / \sigma^2$$

(FIXING σ^2)

I need to optimize (minimize)

$$f(\theta) = \underbrace{\sum_{i,j} \left(y_j(t_i) - x(t_i, \theta)\right)^2}_{\text{sum of squares.}} \qquad \theta^* = \operatorname{argmax} f(\theta)$$

$$\left(\sigma^2 = \frac{f(\theta^*)}{N \cdot m} \right)$$

IDENTIFIABILITY

$\emptyset \longrightarrow X$, k_1 mass action $\rightsquigarrow$

$X \longrightarrow \emptyset$, k_2

$$\boxed{\dot{X} = k_1 - k_2 X}$$

$k_1 - k_2 X = 0$ (equilibrium)

$$x(t) \xrightarrow[t \to \infty]{} \frac{k_1}{k_2}$$

$k_1 = \alpha_1 \cdot \gamma \qquad k_2 = \alpha_2 \cdot \gamma$

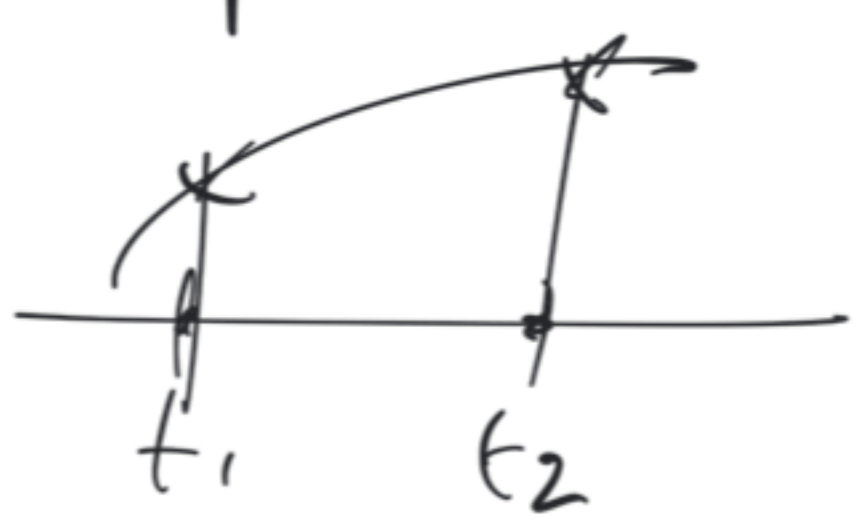
$\frac{k_1}{k_2} = \frac{\alpha_1}{\alpha_2}$ indip. from γ.

Se osservo

$y_j(t)$, for t large so that the system is at steady state.

$$y_j(t) \simeq y_j(\infty) \simeq \frac{k_1}{k_2} = q \qquad \underline{y_j \sim N(q, \sigma^2)}$$

I can find q but not k_1 and k_2.



$$e^{-k_2 t}$$

With observations at time t_1 NOT at steady state, and $t_2 > t_1$, I can identify my system.

$\emptyset \to X$

$\emptyset \to Y$

$X \to \emptyset$

$Y \to \emptyset$

but we observe $Z = X + Y$

We cannot distinguish between rates of X and of Y.