PARAMETER ESTIMATION - stochastic models.

ODE: $J(\theta) = \sum_{j,i} \left(y_j(t_i) - \underbrace{x(t_i, \theta)}_{\text{non-linear dependence on } \theta} \right)^2$ → OPTIMIZE (minimize) wrt θ.

non-linear dependence on θ

⇓

non-convex problem ⇒ many local minima

evaluate J requires solution of IVP (solving ODE).

In the stochastic case: which is the likelihood.

- POPULATION AVERAGES → (a)
- SINGLE CELL INDEPENDENT TIME MEASUREMENTS
- SINGLE CELL TRAJECTORY.

(a) POPULATION AVERAGES

$$Y(t) \sim \mathcal{N}\left(\mathbb{E}[X(t|\theta_{true})], \sigma^2 \right)$$

$y_j(t_i)$ observations

$$P(y_j(t_i)|\theta) = \int \boxed{P(y_j(t_i), \mathbb{E}[X(t_i)] = x \,|\, \theta)} \, dx$$

$$P(y_j(t_i) \,|\, \mathbb{E}[X(t_i)] = x) \cdot \underbrace{P(\mathbb{E}[X(t_i)] = x \,|\, \theta)}_{\delta_{E[X(t_i|\theta)]}}$$

$$p(y_j(t_i)|\vartheta) = P(y_j(t_i) \,|\, \mathbb{E}[x(t_i|\vartheta)])$$

$$\overset{!}{=} \mathcal{N}(\mathbb{E}[X(t_i|\vartheta)], \sigma^2) \propto \exp\left(\frac{(y_j(t_i) - \mathbb{E}[X(t_i|\vartheta)])^2}{\sigma^2}\right)$$

$$f(D, \vartheta) \propto \sum_{j,i} (y_j(t_i) - \underbrace{\mathbb{E}[X(t_i,|\vartheta)]})^2$$

Come calcolo $\mathbb{E}[X(t_i|\vartheta)]$

(Se tutti i rate sono lineari, allora è facile, basta risolvere le RRE)

Altrimenti devo approssimare

- STOCHASTIC APPROXIMATION (vedi next lectures)
- STATISTICAL APPROXIMATION: sample $X(t_i|\vartheta)$ n times, estimate $\mathbb{E}[X(t_i|\vartheta)] = \frac{1}{N}\sum_n x_k(t_i)$

(n ≥ 1000 to have decent estimates)

Computationally expensive

- SINGLE CELL / INDEP. TIME

$$\underbrace{P(y_j(t_i) \mid \theta)} = \int P(y_j(t_i) \mid X(t_i \mid \theta) = x) P(X(t_i \mid \theta) = x \mid \theta) \, dx$$

$$= \sum_{x \in S} \underbrace{P(y_j(t_i) \mid X(t_i \mid \theta) = x)} \underbrace{P(X(t_i) = x \mid \theta)}$$

for discrete S.

I "just" need to solve Kolmogorov equations at time t_i.

$$= \mathbb{E}_{P(X(t_i \mid \theta))} \left[\underbrace{P(y_j(t_i) \mid X(t_i \mid \theta) = x)}_{\propto \exp\left(-\frac{(y_j(t_i) - x)^2}{\sigma^2}\right)} \right]$$

$P(y_j(t_i) \mid x) = \mathcal{N}(x; \sigma^2)$

Warning $\log(\mathbb{E}[f(x)]) \geq \underbrace{\mathbb{E}[\log(f(x))]}$ So I can use $\mathbb{E}[\log(f(x))]$ as a lower bound on the likelihood

and maximize the lower bound.

• SINGLE CELL TRAJECTORY

$$P(y_j(t_1),\dots,y_j(t_k)\mid\vartheta) = \int P(y_j(t_1),\dots,y_j(t_k)\mid X(t_1)=x_1,\dots,X(t_k)=x_k)\cdot P(X(t_1)=x_1,\dots,X(t_k)=x_k\mid\vartheta)\,dx_1\dots dx_k$$

$$= \mathbb{E}\left[P(y_j(t_1),\dots,y_j(t_k)\mid X(t_1)=x_1\dots X(t_k)=x_k)\right]$$

$$P(x_1,\dots,x_k = x_1\dots x_k)$$

$$\mathbb{E}\left[\prod_i P(y_j(t_i)\mid x_i)\right] \geq \mathbb{E}\left[\sum_i \log P(y_j(t_i)\mid x_i)\right]$$

$$\log\mathbb{E} \geq \mathbb{E}\log$$

$$\sum_{x_1,\dots,x_k} \underbrace{P(y_j(t_1),\dots,y_j(t_k)\mid x_1,\dots,x_k)}_{} \cdot \underbrace{P(X(t_1)=x_1,\dots,X(t_k)=x_k \mid \theta)}_{}$$

$$P(y_j(t_1)\mid x_1)\cdots P(y_j(t_k)\mid x_k)$$

Independent (gaussian) noise of time points $t_1,\dots,t_k$

Memoryless

$$P(X(t_1)=x_1,\dots,X(t_k)=x_k) = \sum_{x_0} P(X(0)=x_0)\prod_{i=1}^{k} \boxed{P(X(t_i)=x_i \mid X(t_{i-1})=x_{i-1})}$$

$$= \sum_{x_0} P(X(0)=x_0, X(t_1)=x_1,\dots,X(t_k)=x_k)$$

Esempio: $t_1,\dots,t_k$ sia regolare, $t_i=i\cdot h$

$$P(\underbrace{X(t_i)}_{X(i\cdot h)=x_i}=x_i \mid \underbrace{X(t_{i-1})}_{X((i-1)\cdot h)=x_{i-1}}=x_{i-1}) = \boxed{P(X(h)=x_i \mid X(0)=x_{i-1})}$$

Can be computed by solving the matrix Kolmog. equations.

$$\mathcal{L}(D,\theta) = \prod_{j=1}^{N}\left[\sum_{x_0} P(X(0)=x_0|\theta)\prod_{i=1}^{K}\left(\sum_{x_i} p(y_j(t_i)|x_0)\, P(X(t_i)=x_i \,|\, X(t_{i-1})=x_{i-1})\right)\right]$$

likelihood

Ha complessità $K\cdot|S|$ invece di $|S|^K$!