NONPARAMETRIC IDENTIFICATION OF PHARMACOKINETIC MODELS VIA GAUSSIAN PROCESSES

Giuseppe DE NICOLAO

University of Pavia
Italy

Outline

- Bayesian Learning
- Population Data Analysis
- Signal Estimation
- Hyperparameter Estimation: Empirical Bayes and MCMC
- Application to PK Data

Bayesian Learning

Bayesian Learning

Model: $y_i = f(x_i) + v_i \qquad v \sim N(0, \sigma^2 I)$

Problem: estimate $f(\cdot): \mathrm{R}^d \mapsto \mathrm{R}$ from the

training set $D_n = \{(x_i, y_i), i = 1, \ldots, n\}$

Solution: Bayesian approach

Bayes Theorem:

$$\underbrace{p(f \mid D_n)}_{\textit{posterior}} = \frac{\underbrace{p(D_n \mid f)}_{\textit{likelihood}} \underbrace{p(f)}_{\textit{prior}}}{p(D_n)} \; ?$$

Bayesian Learning

$f(\cdot)$ ~ Gaussian Process (GP)

GP properties:

- easy to compute
- flexibility
- determined by their mean and autocovariance function $K(x_1, x_2)$

$$K(x_1, x_2) = K(x_1, x_2; \underbrace{\lambda_1^{\,2}, \dots, \lambda_r^{\,2}}_{\textit{hyperparameters}}) \Rightarrow \boxed{\text{GCV, OCV, } C_p\text{, ML, MCMC, …}}$$

Bayesian Learning

$f(\cdot) \sim \text{GP} \Rightarrow p(f \mid D_n)$ is Gaussian

if E[f] = 0

$$\boxed{\hat{f}(x) = E[f \mid D_n] = \overline{K}(x)^T Var[y]^{-1} y}$$

where $\overline{K}(x) = [K(x, x_1) \cdots K(x, x_n)]^T$

Remark:

Let $\theta = Var[y]^{-1} y \Rightarrow \boxed{\hat{f}(x) = \sum_{i=1}^{n} K(x, x_i)\theta_i}$

Radial *K*

RBF

Bayesian Learning

Tychonov Regularization $\Longleftrightarrow$ Bayesian Estimation

$$\Downarrow$$

$$\boxed{f_\gamma(x) = \arg\min_f \sum_{j=1}^{n} (y_j - f(x_j))^2 + \underbrace{\gamma \|f\|_K^2}_{\text{"smoothness penalty}}}$$

Solution:

$$\hat{f}_\gamma(x) = \sum_{j=1}^{n} K(x, x_j)\tilde{\theta}_j$$

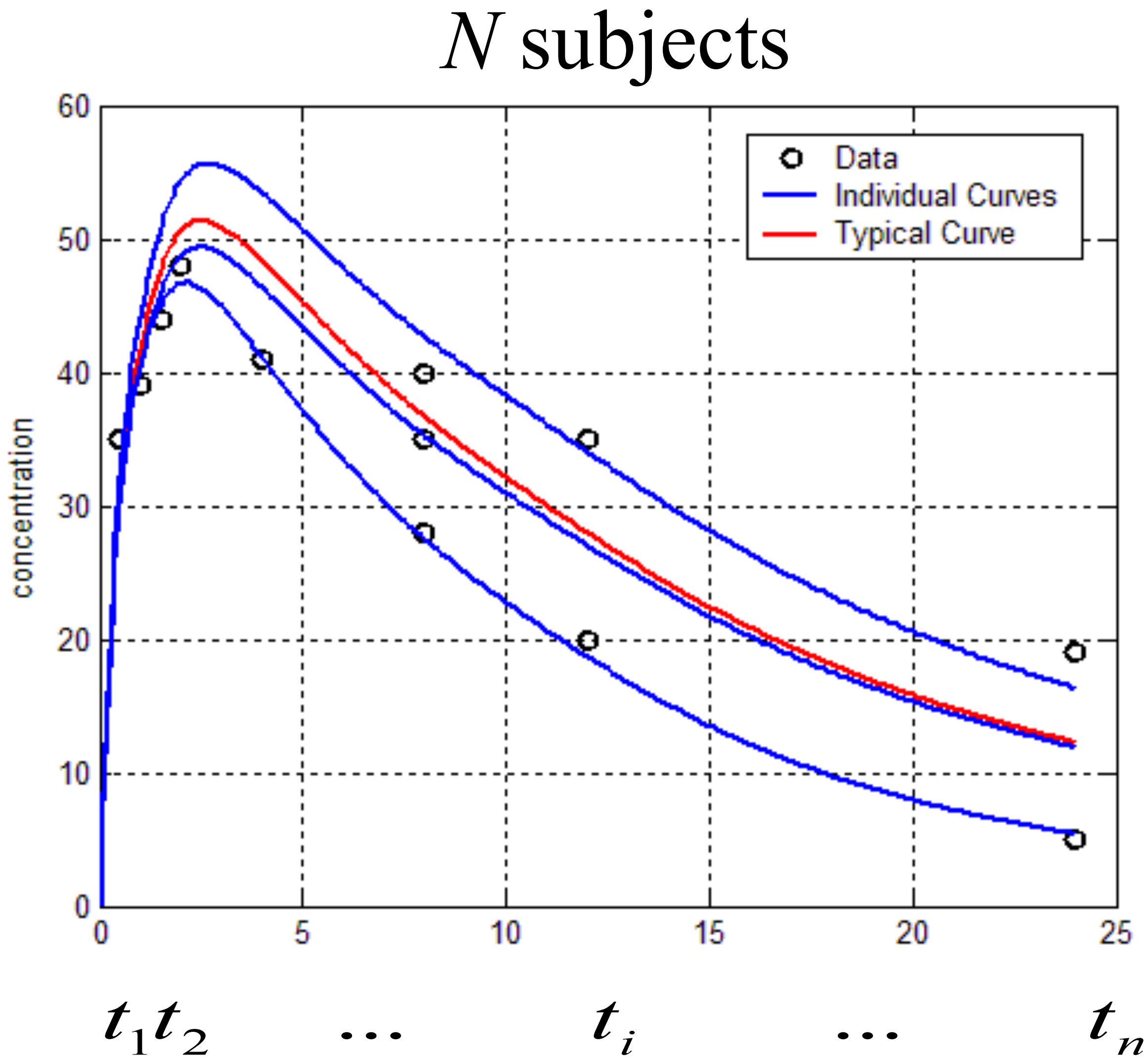
Remark:

if $\quad \hat{\gamma} = Var[v] \quad \Rightarrow \quad \tilde{\theta} = \theta \quad \Rightarrow \quad \boxed{\hat{f}_\gamma(x) = \hat{f}(x)}$

Bayes estimator

Population Data Analysis

Pharmacokinetic data



Noisy data

$$y_k = x(t_k) + v_k$$

Main goals:

- Typical curve
- Individual curves

State of the art

- ***Parametric models:***

  Example: compartmental models

  Drawback: a structural model is needed

- ***Semiparametric models:***

  Example: regression splines [*Verotta et al.,1997*]

  Drawback : number and location of spline knots

- ***Nonparametric models:***

  Example: Random Walk [*Magni et al., 2002*]

  Drawback : Discrete-time model

Open problems

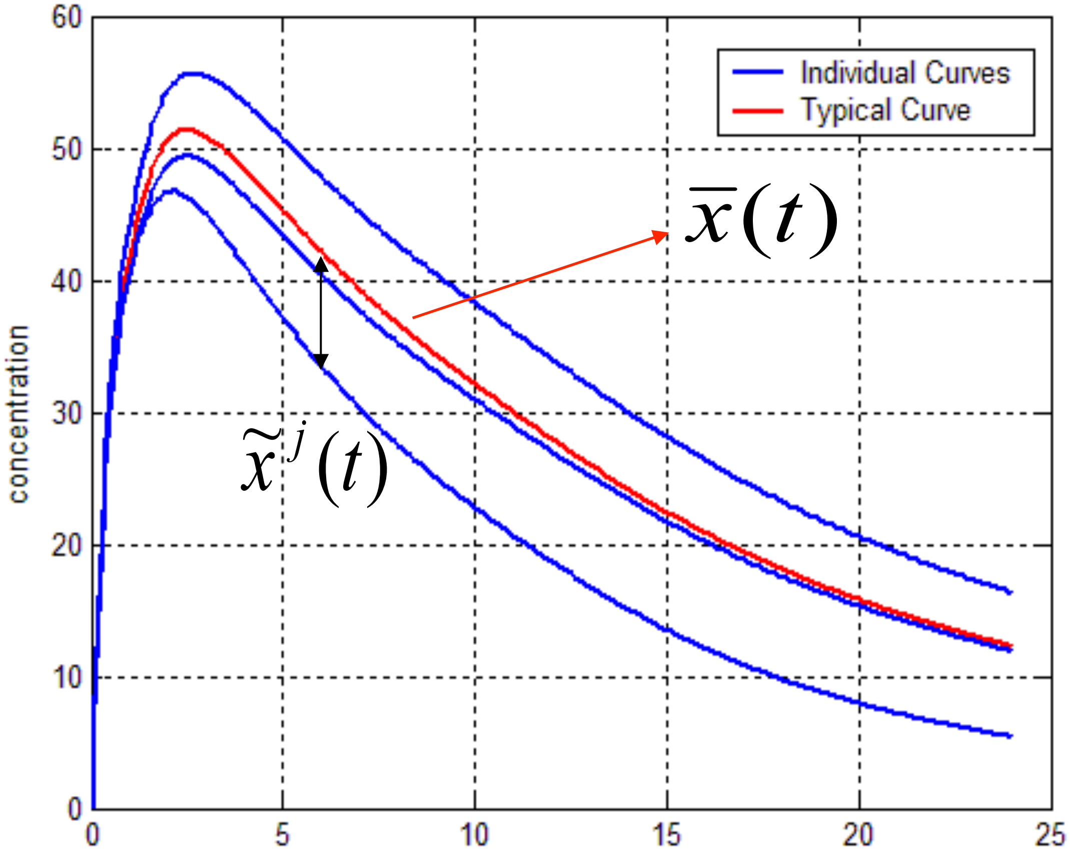
Population analysis able to estimate

continuous time signals

without resorting to parametric models

Approach: Signals modelled as continuous-time stochastic processes

Population curves



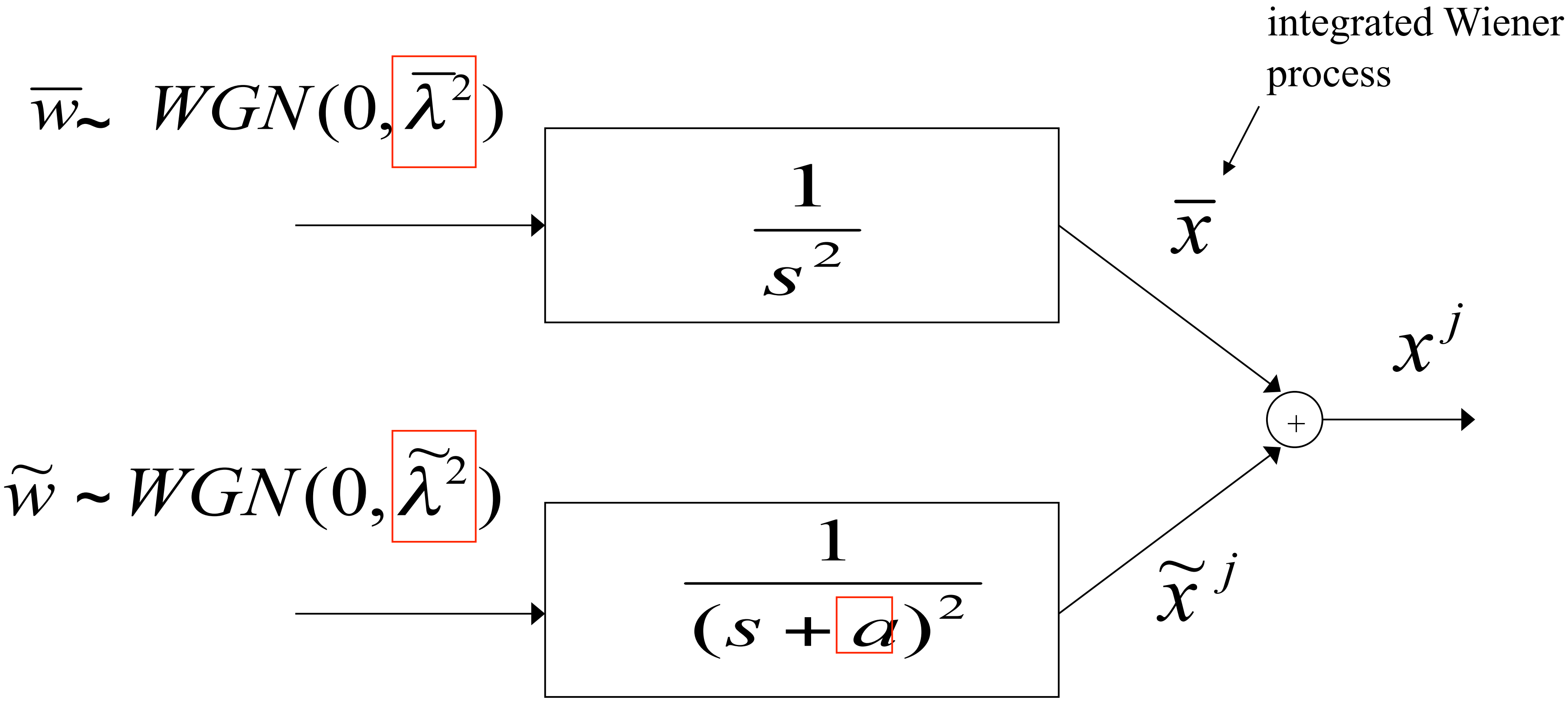
$$x^j(t) = \bar{x}(t) + \tilde{x}^j(t)$$

The *j*-th individual curve $x^j(t)$ is given by the sum of:

- the typical curve: $\bar{x}(t)$
- the individual shift: $\tilde{x}^j(t)$

Models and hyperparameters

Unknown hyperparameters:

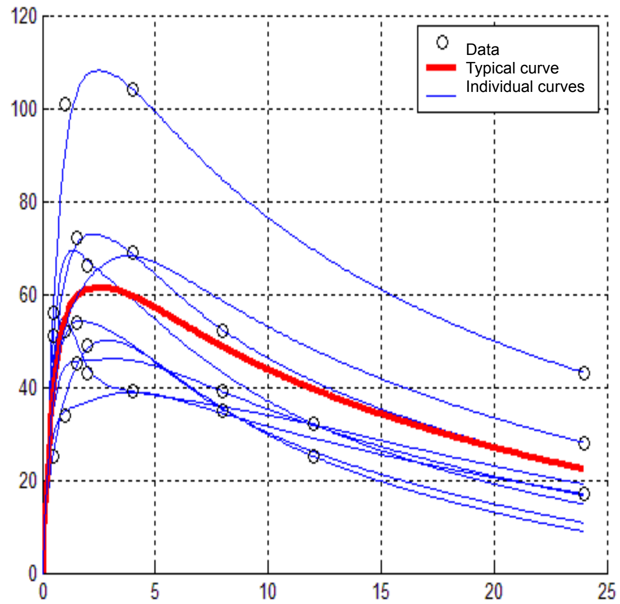


Assumption: zero initial state (can be relaxed)

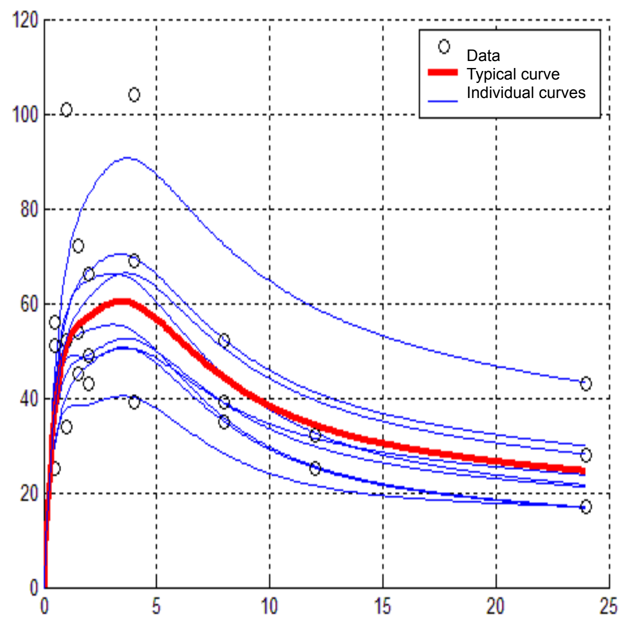
Nonparametric model

$\widetilde{\lambda}^2$ *effect*

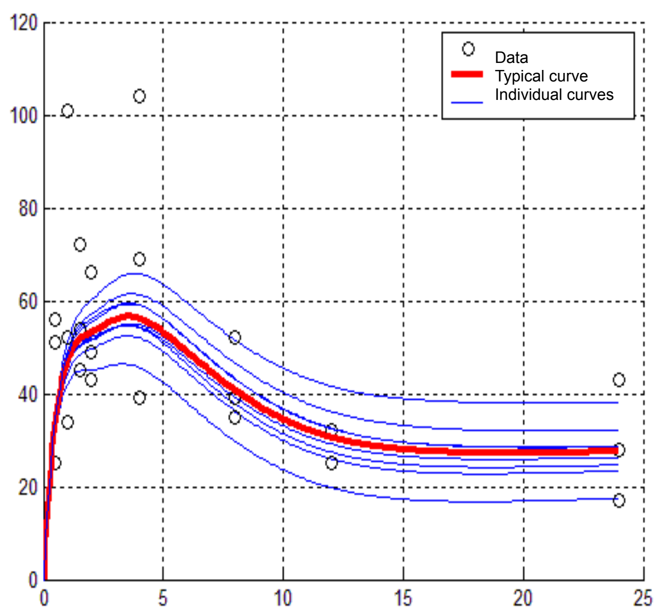
$\widetilde{\lambda}^2 = 10000$



$\widetilde{\lambda}^2 = 1000$



$\widetilde{\lambda}^2 = 100$



Nonparametric model

Signal Estimation

Bayesian estimation

Remarks :

- $\overline{x}(t)$ is a gaussian random variable with zero mean
- $Y = \overline{X} + \widetilde{X} + V$ array of all the available data
- $\left(\overline{x}(t), Y\right)$ are jointly gaussian

$$\hat{\overline{x}}(t) = E[\overline{x}(t) \mid Y] = Cov[\overline{x}(t), Y]Var[Y]^{-1}Y$$

estimate of $\overline{x}(t)$

Bayesian estimation

$$\bar{R}(t,\tau) = \bar{\lambda}^2 \bar{\gamma}(t,\tau): \text{ } \textit{autocovariance of } \bar{x}(t)$$

$$\tilde{R}(t,\tau) = \tilde{\lambda}^2 \tilde{\gamma}(t,\tau): \text{ } \textit{autocovariance of } \tilde{x}(t)$$

Remark 1:

$$Cov[\bar{x}(t), Y] = [\bar{R}(t,t_1) \ldots \bar{R}(t,t_n)]$$

t_i (knots), $i = 1,\ldots,n$ $\quad$ $n = n°$ *observations*

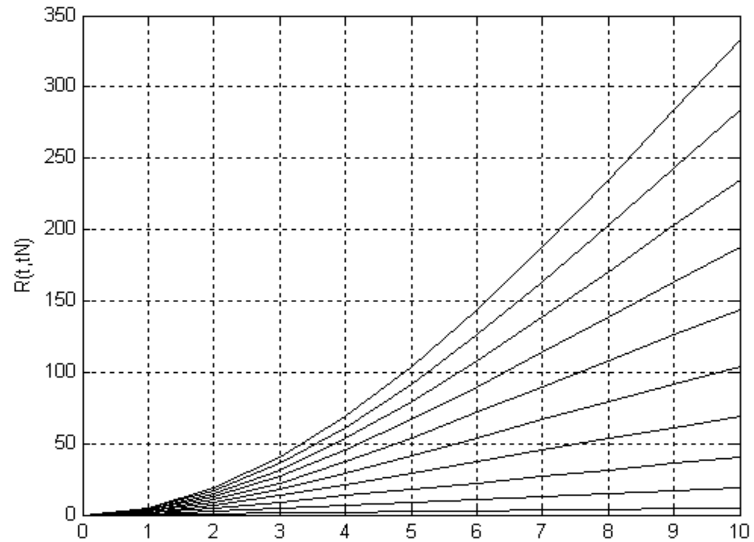
Remark 2:

$Var[Y]$ depends on both $\bar{R}(t,t_i)$ and $\tilde{R}(t,t_i)$

Autocovariance functions

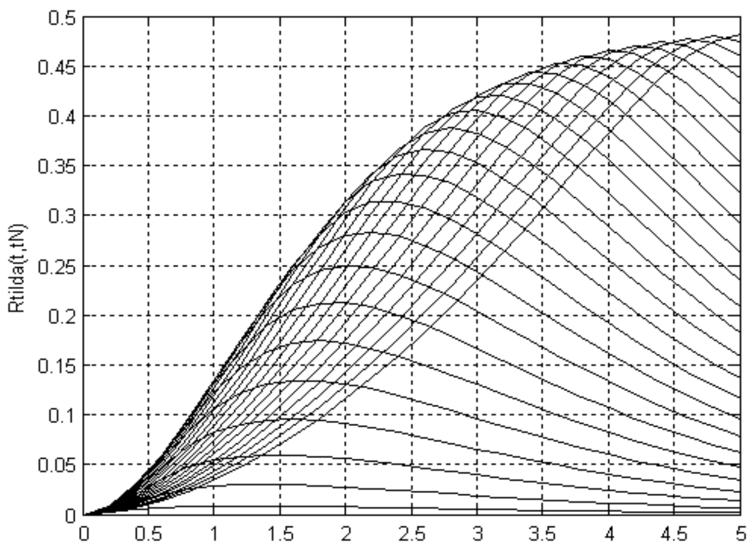
Typical curve:

$$\bar{\gamma}(t,\tau)=\begin{cases}\dfrac{t^2}{2}\left(\tau-\dfrac{t}{3}\right) & t\le\tau\\[2ex] \dfrac{\tau^2}{2}\left(t-\dfrac{\tau}{3}\right) & t>\tau\end{cases}$$



Individual shifts:

$$\tilde{\gamma}(t,\tau)=\begin{cases}e^{a(\tau-t)}\left(\dfrac{t^2e^{2at}}{2a}-\dfrac{t\cdot e^{2at}}{2a^2}-\dfrac{e^{2at}-1}{4a^3}\right)+ \\ \quad +(\tau-t)e^{a(\tau-t)}\dfrac{2at\cdot e^{2at}-e^{2at}+1}{4a^2} & t\le\tau\\[2ex] e^{a(t-\tau)}\left(\dfrac{\tau^2e^{2a\tau}}{2a}-\dfrac{\tau\cdot e^{2a\tau}}{2a^2}-\dfrac{e^{2a\tau}-1}{4a^3}\right)+ \\ \quad +(t-\tau)e^{a(t-\tau)}\dfrac{2a\tau\cdot e^{2a\tau}-e^{2a\tau}+1}{4a^2} & t>\tau\end{cases}$$



Estimate expressions

Hypothesis: known hyperparameters

Estimate of the typical curve:

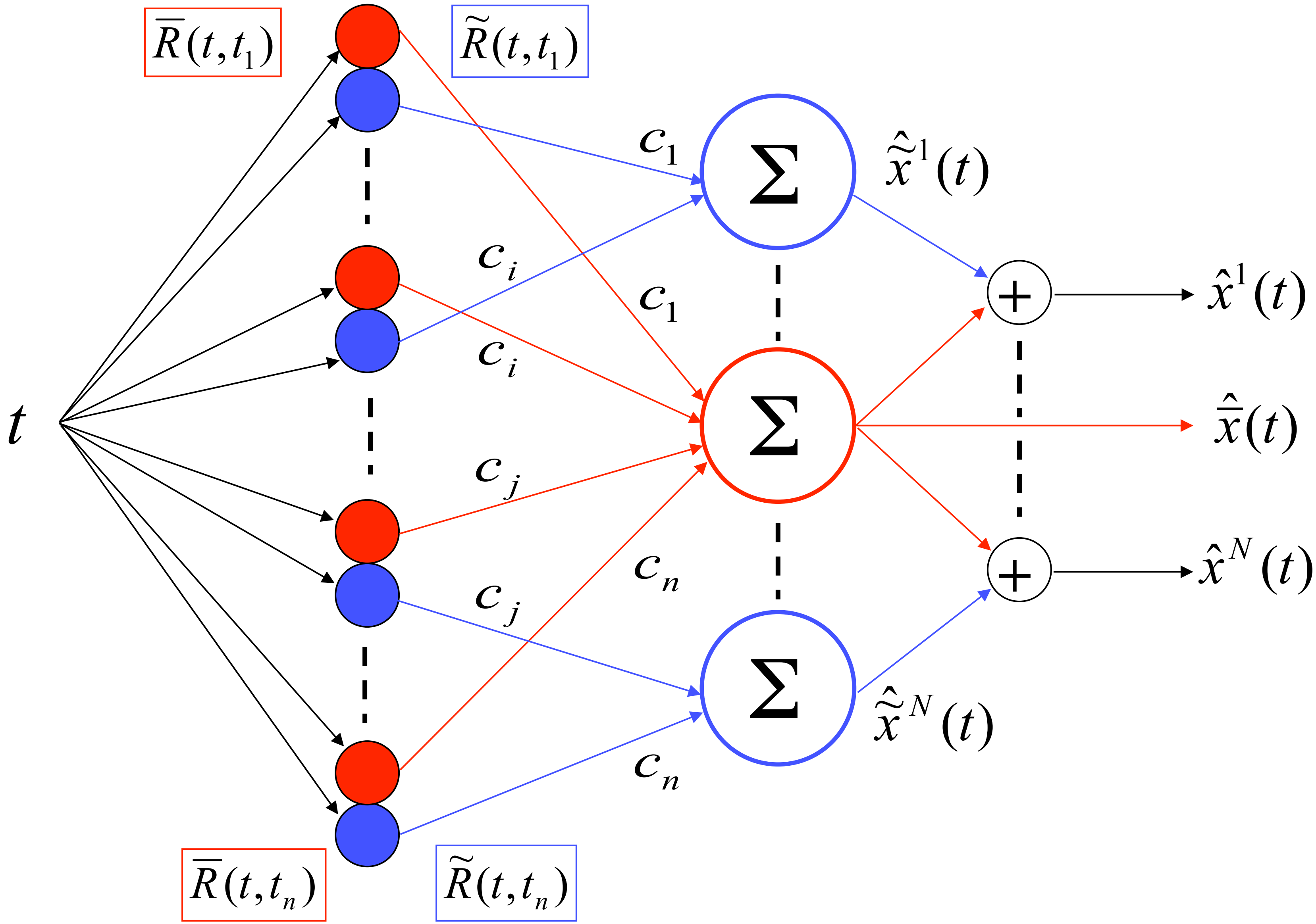
$$\boxed{\hat{\bar{x}}(t) = \sum_{i=1}^{n} \bar{R}(t,t_i) \cdot c_i} \longrightarrow \text{Cubic spline}$$

Estimate of the individual curve:

$$\boxed{\hat{x}^j(t) = \sum_{i=1}^{n} \bar{R}(t,t_i) \cdot c_i + \sum_{i=1}^{n} \tilde{R}(t,t_i) \cdot c_i}$$

Array of weights: $\boxed{c = Var[Y]^{-1} Y}$

Regularization Network



Signal estimation

Hyperparameter Estimation: Empirical Bayes and MCMC

Empirical Bayes

Combination between

Classic approach:

Bayesian approach:

Hyperparameter estimation: likelihood maximization

$$\theta = [\bar{\lambda}^2, \tilde{\lambda}^2, a\ ]$$

$$\theta_{ML} = \arg\max(f(Y \mid \theta))$$

Estimation of the typical curve:

$$\hat{\bar{x}}(t) = E[\bar{x}(t) \mid Y, \theta = \theta_{ML}]$$

MCMC algorithms

Bayesian Inference:

$$\boxed{E[f(\theta)\,|\,Y] = \int_{\Theta} f(\theta) P(\theta\,|\,Y)\, d\theta}$$

Problem: the integral does not admit an analytic solution

⬇

MONTE CARLO INTEGRATION

MCMC algorithms

MONTE CARLO INTEGRATION

Given a sufficient amount of samples $[\theta_1 ... \theta_K]$ drawn from the posterior:

$$E[f(\theta) \mid Y] \approx \frac{1}{K} \sum_{k=1}^{K} f(\theta_k)$$

Problem: how to draw the samples?

MCMC algorithms

MARKOV CHAINS

Construct a Markov chain which converges to the posterior distribution that has to be sampled

→

METROPOLIS-HASTINGS
(Gibbs Sampler)
algorithm

MCMC

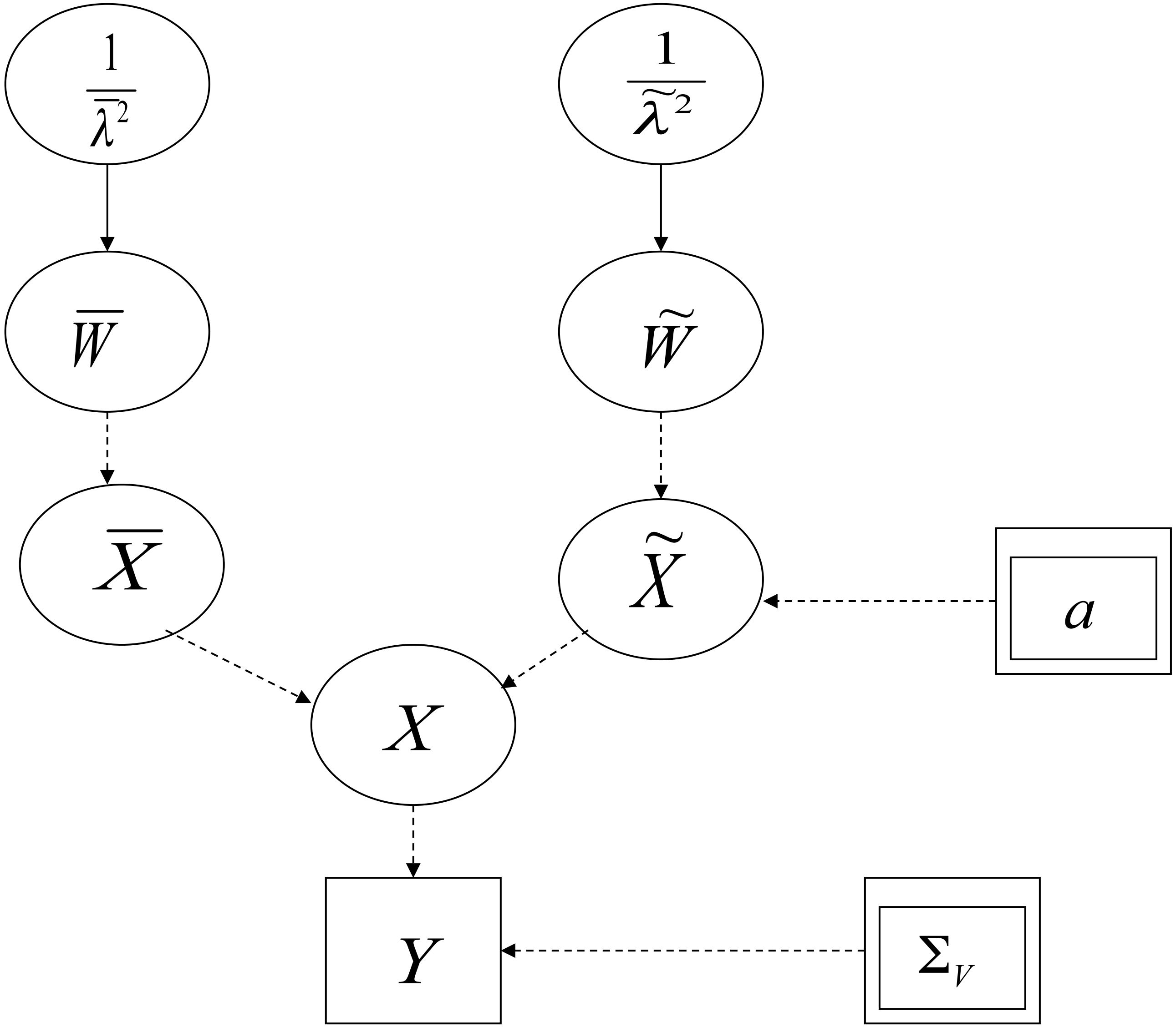
Problem: MCMC simulation of the curves in each time instant t

Computationally burdensome

Solution: MCMC simulation of the curves in correspondence with the nodes

only n random variables are sampled

Bayesian network



Hyperparameter estimation

MCMC

Prior distributions:

$$p\left(\frac{1}{\bar{\lambda}^2}\right) = \Gamma(\gamma_1, \gamma_2)$$

$$p\left(\frac{1}{\tilde{\lambda}^2}\right) = \Gamma(\gamma_3, \gamma_4)$$

$$p\left(\bar{W}\right) = N\left(0, \bar{\lambda}^2 I_n\right)$$

$$p\left(\tilde{W}\right) = N\left(0, \tilde{\lambda}^2 I_n\right)$$

Full conditional:

$$p\left(\frac{1}{\bar{\lambda}^2} \mid \cdot\right) \propto \Gamma\left(\gamma_1 + \frac{n}{2}, \gamma_2 + \frac{\bar{W}^T \bar{W}}{2}\right)$$

$$p\left(\frac{1}{\tilde{\lambda}^2} \mid \cdot\right) \propto \Gamma\left(\gamma_3 + \frac{n}{2}, \gamma_4 + \frac{\tilde{W}^T \tilde{W}}{2}\right)$$

$$p\left(\bar{W}\right) \propto e^{-\frac{1}{2}\left(\frac{\bar{W}^T \bar{W}}{\bar{\lambda}^2}\right)} e^{\left(-\frac{1}{2}(Y-\bar{X})^T (Var[\tilde{X}]+\Sigma_V)^{-1} (Y-\bar{X})\right)}$$

$$p\left(\tilde{W}\right) \propto e^{-\frac{1}{2}\left(\frac{\tilde{W}^T \tilde{W}}{\tilde{\lambda}^2}\right)} e^{\left(-\frac{1}{2}(Y-\tilde{X})^T (Var[\bar{X}]+\Sigma_V)^{-1} (Y-\tilde{X})\right)}$$

MCMC: the typical curve

Given the posterior distribution: $p(\bar{\lambda}^2, \tilde{\lambda}^2 \mid Y)$

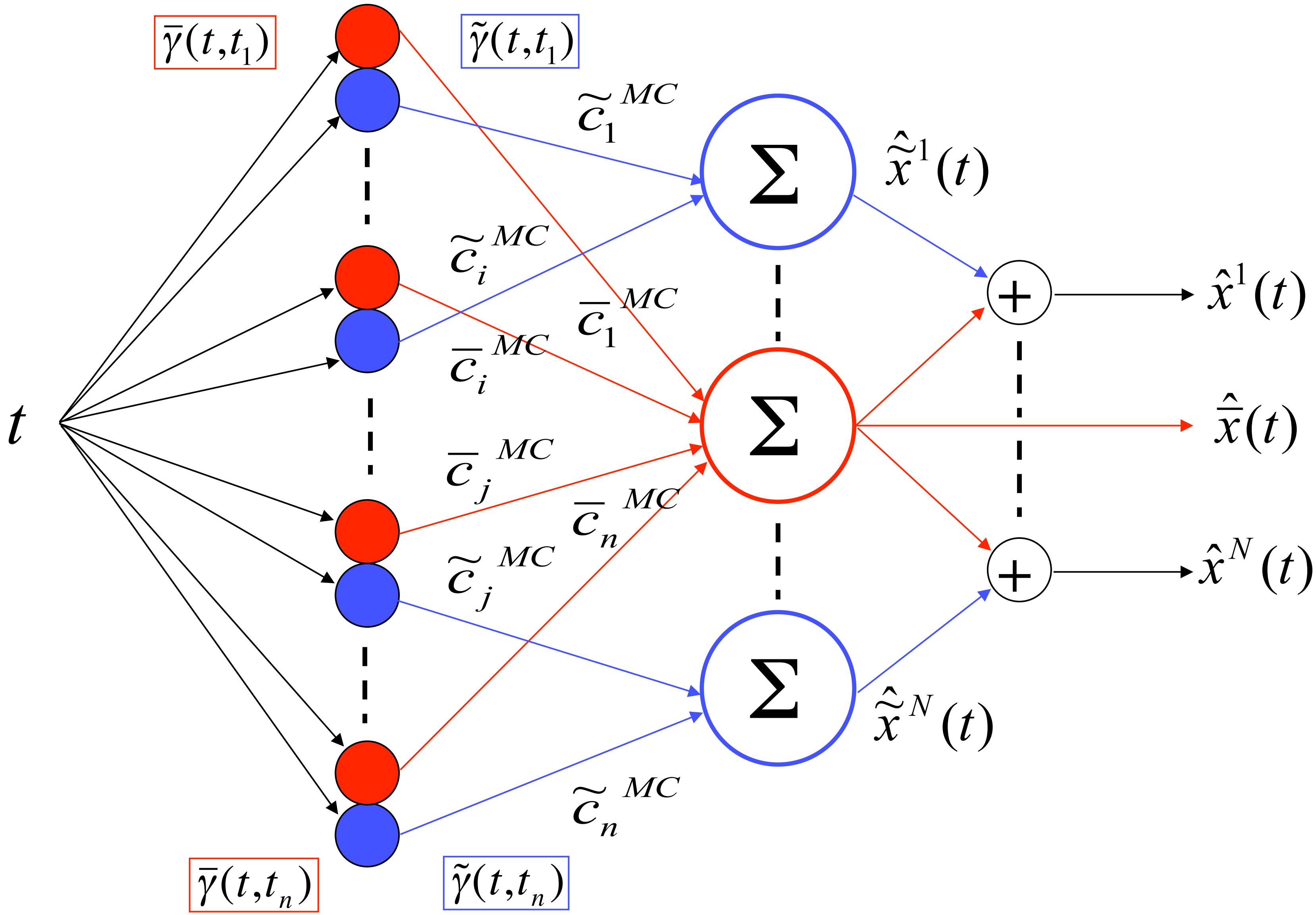
(hp: α known)

$$\hat{\bar{x}}(t) = E[\bar{x}(t) \mid Y] = \int E[\bar{x}(t) \mid Y, \theta] p(\theta \mid Y) d\theta$$

$$\cong \sum_{k=1}^{K} \frac{E[\bar{x}(t) \mid Y, \bar{\lambda}_k^{\,2}, \tilde{\lambda}_k^{\,2}]}{K} = \sum_{i=1}^{n} \bar{\gamma}(t, t_i) \cdot \bar{c}_i^{MC}$$

with $\bar{c}_i^{MC} = \sum_{k=1}^{K} \frac{\bar{\lambda}_k^{\,2} c_i(\bar{\lambda}_k^{\,2}, \tilde{\lambda}_k^{\,2})}{K}$ (K samples)

Modified Regularization Network

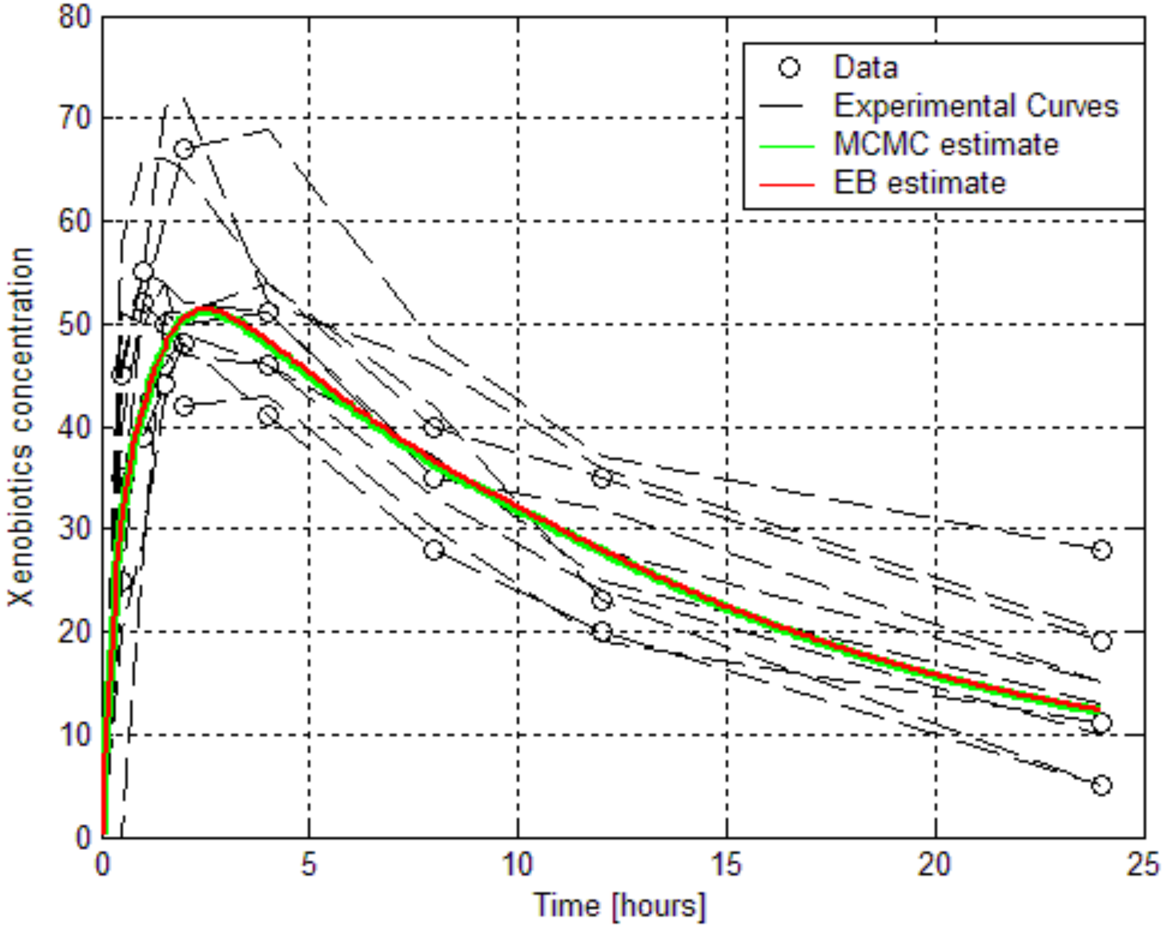


Hyperparameter estimation

Application to PK Data

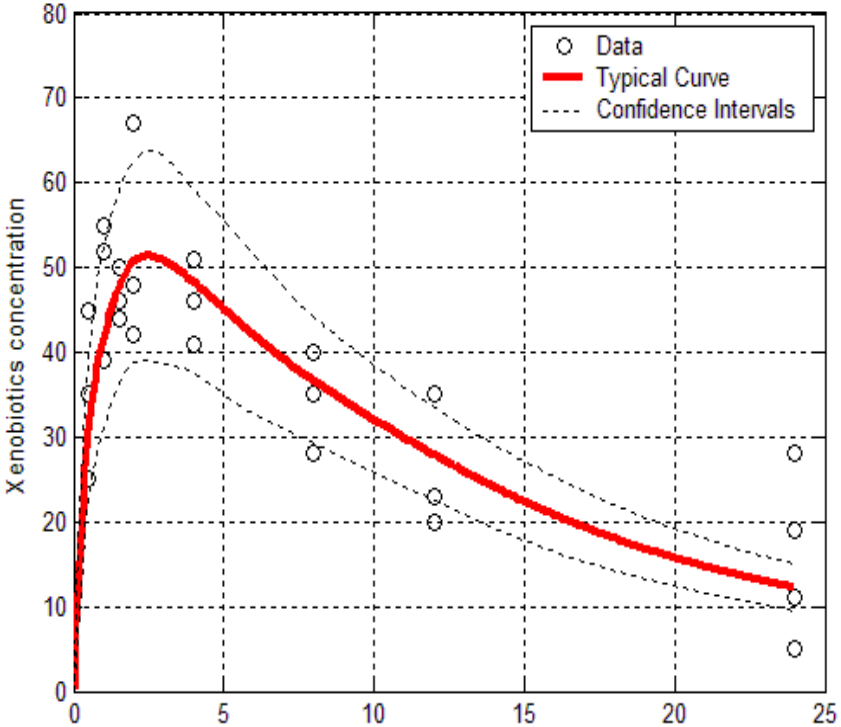
Pharmacokinetic data

N=10 subjects

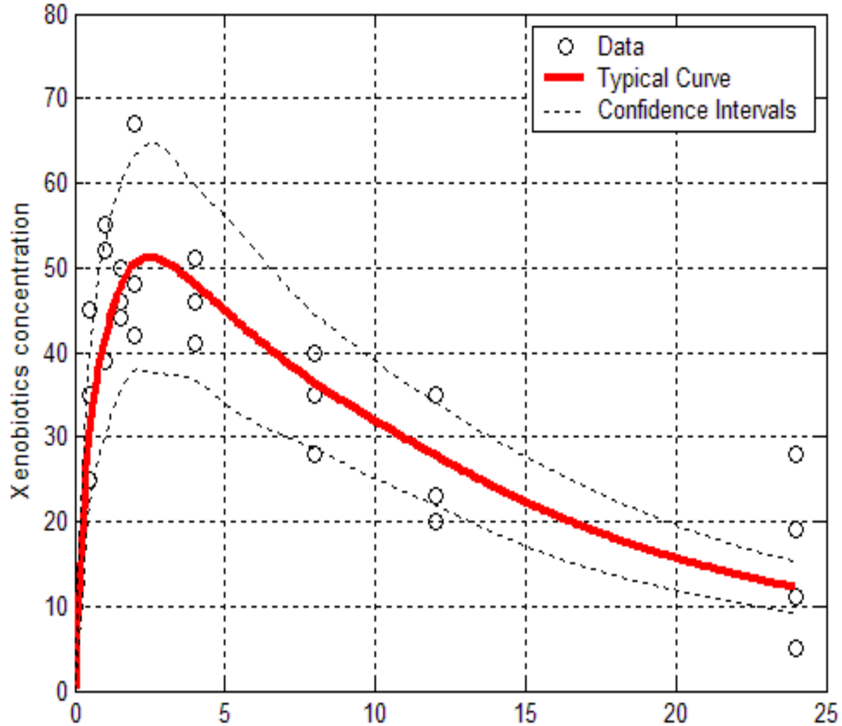


Typical curve estimates

EB estimate

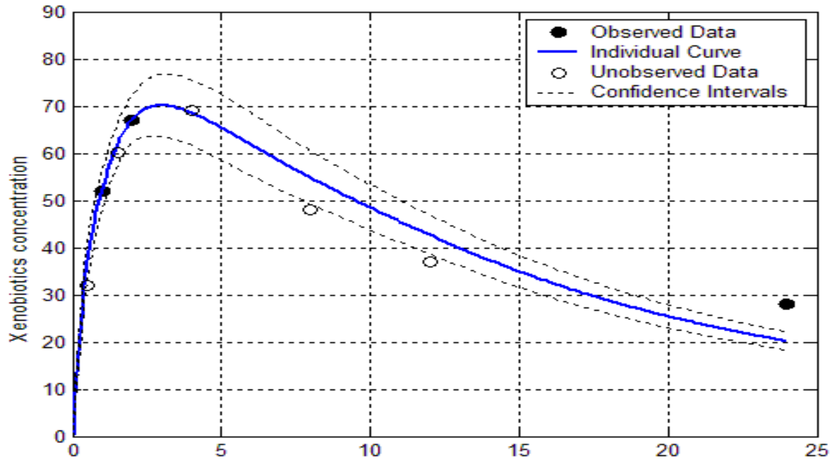


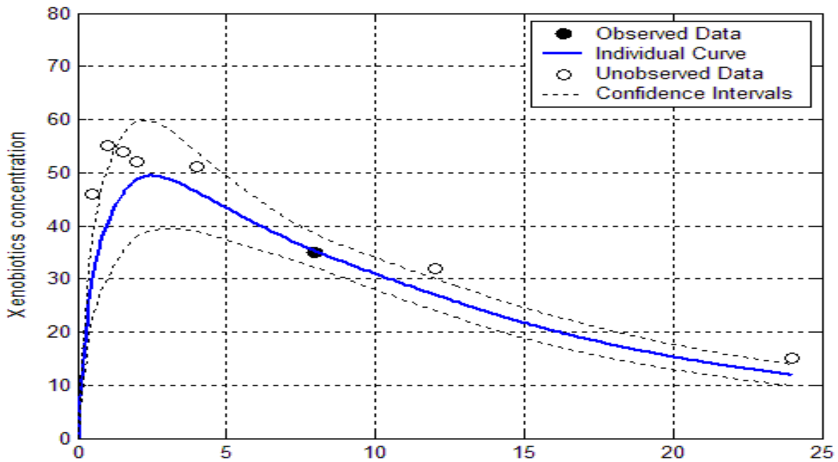
MCMC estimate



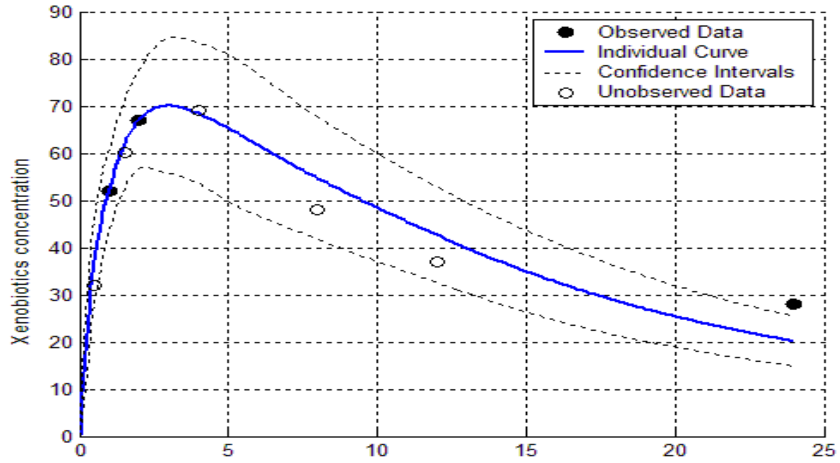
Individual curve estimates

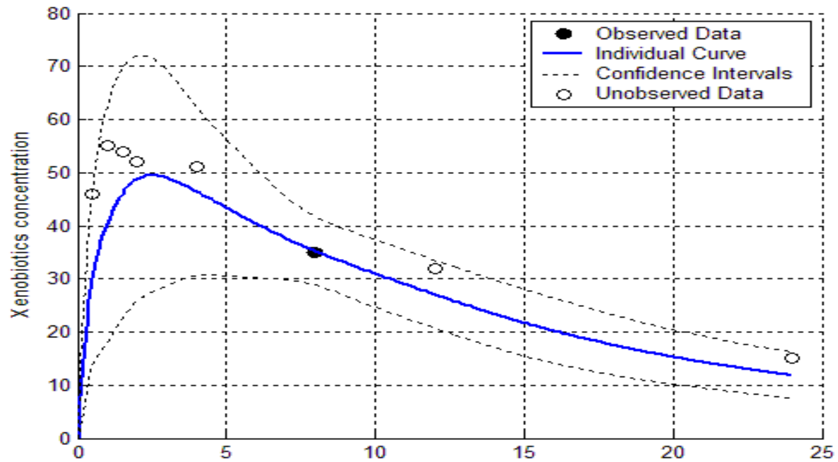
EB estimate





MCMC estimate

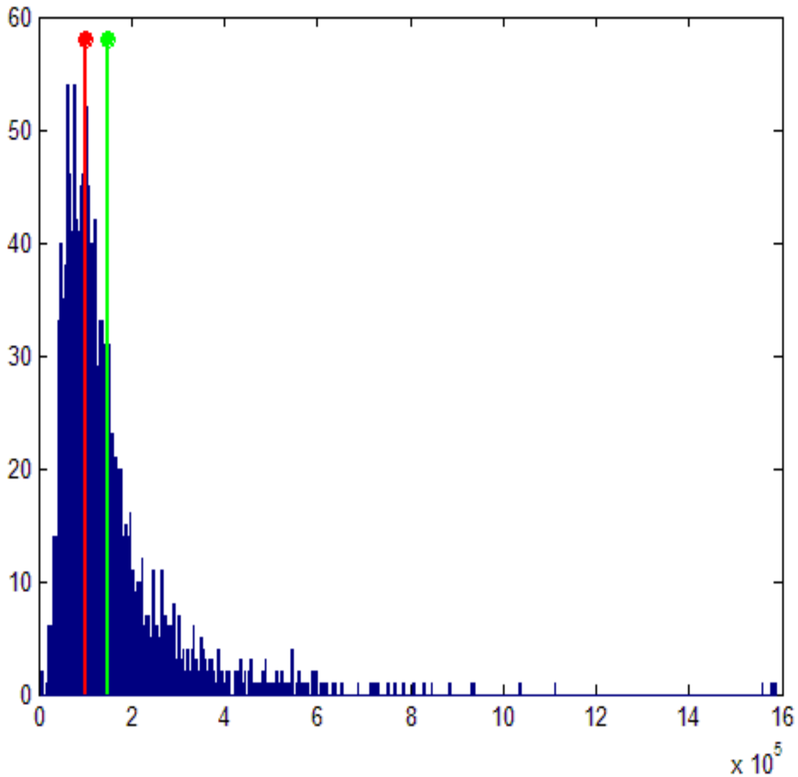


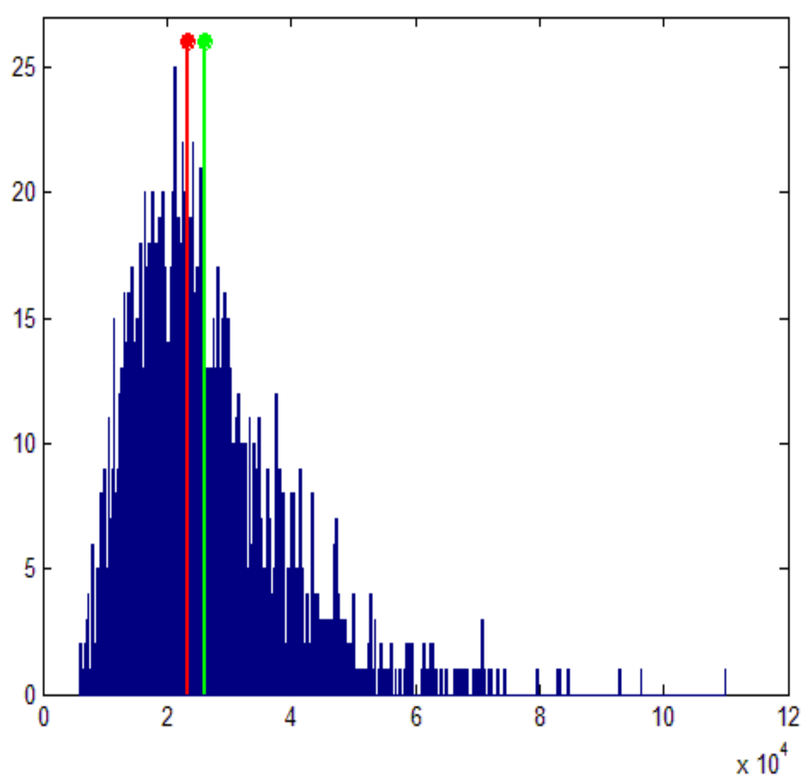


Hyperparameter posterior

Histogram of $\overline{\lambda}^2$

Histogram of $\widetilde{\lambda}^2$

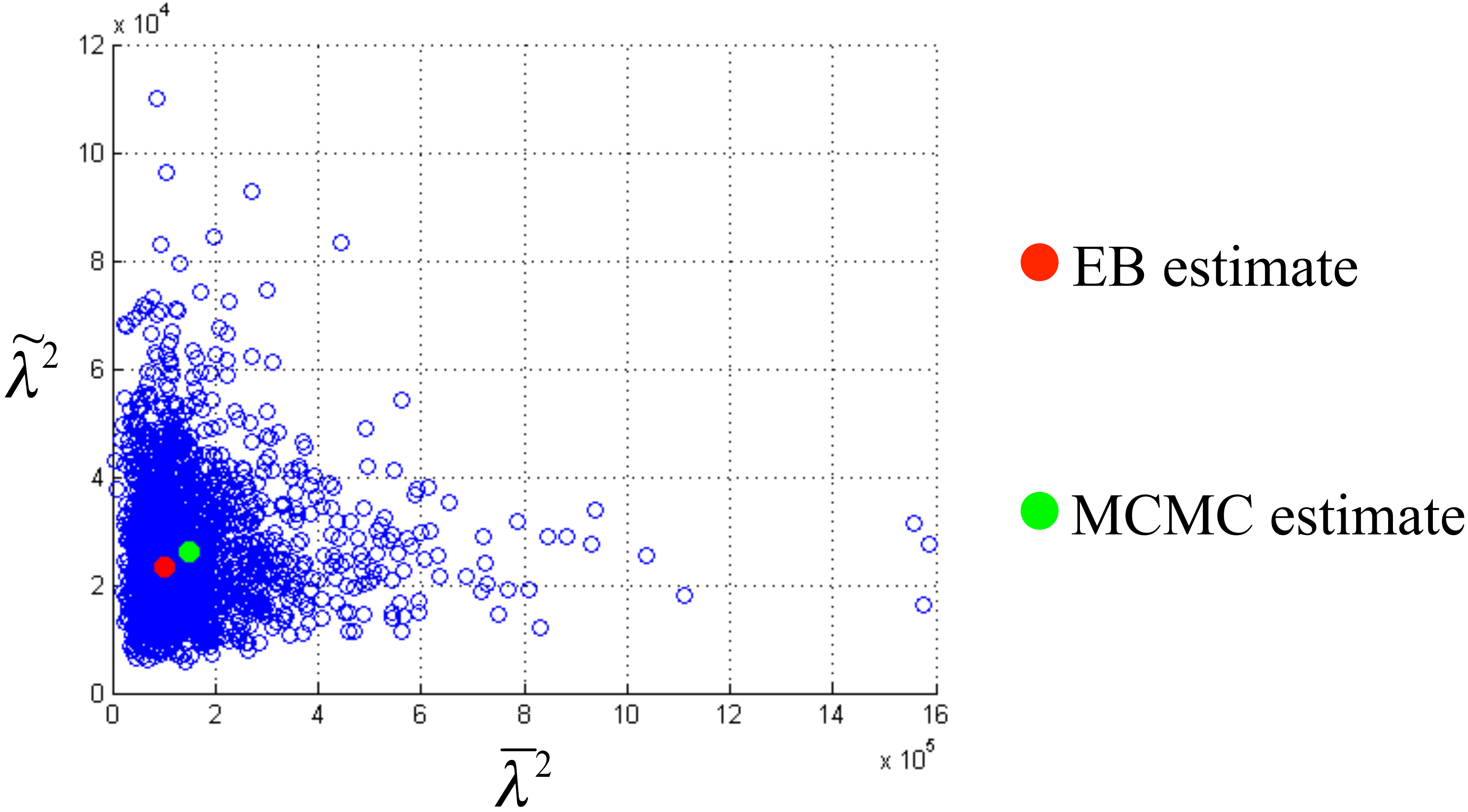




● EB estimate ● MCMC estimate

Application to PK data

Scatter plot of the couples $(\overline{\lambda}^2, \widetilde{\lambda}^2)$



Conclusions

- Continuous- time estimates solving a finite dimensional problem
- Estimate of the typical curve: cubic spline
- Regularization Network structure
- Hyperparameter estimation: Empirical Bayes and MCMC (the latter does not underestimate confidence intervals)
- MCMC estimates for the intersample without drawing additional intermediate samples