Statistical analysis of stochastic processes and applications in life sciences

Maud Delattre

Université Paris-Saclay

Unité de recherche INRAE, MaIAGE (UR 1404), Jouy-en-Josas

Journées MAS 2022

30 août 2022

Contents

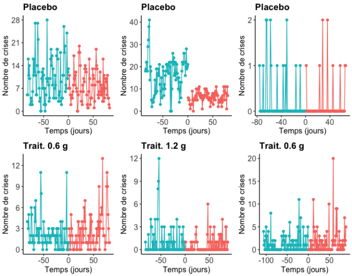
Contents

Introduction

- An exhaustive state of the art of statistics for stochastic processes is impossible.
- Many fields use observations collected over time (or according to a position variable).
- Due to their structure and properties, stochastic processes are relevant models to account for :
  - mechanistic aspects,
  - intrisic stochasticity in the dynamics,
  - dependence between consecutive observations (*e.g.* Markovian processes),
  - ...
- ... make it possible to distinguish between the stochastic modeling of the studied phenomenon and the statistical modeling of the available observations :
  - hidden states
  - aggregation of data
  - measurement errors
  - ...

Ex. 1 : Hidden Markov Models (HMM) in pharmacology

Evolution of the number of daily seizures over time in epileptic patients [D. et al., 2012]



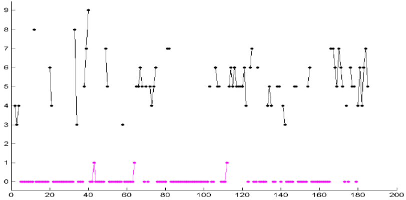
What clinicians need from a model :

- describe the mechanism behind the succession of seizures over time
- describe the effects of the drug on the progression of symptoms
- . . .

Ex. 1 : HMM in pharmacology

Aspects that the model needs to consider ([D. et al., 2012], [Altman R. M., 2007], . . .):

- Stochastic dynamic model
  - The existence of different health states with different levels of symptoms is a reasonable assumption.
  - Patients would tend to stay in the same state for several consecutive days.

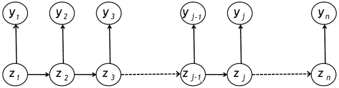


- Available observations :

  The states are not observed, only the daily numbers of seizures are.

Ex. 1 : HMM in pharmacology

- General model structure



- $(z_j)_{j=1,\dots,n}$: sequence of states, Markov chain

$$\mathcal{L}(z_j | z_{j-1}, \dots, z_1) = \mathcal{L}(z_j | z_{j-1}) \quad (1)$$

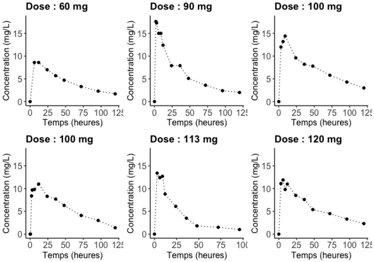
- $(y_j)_{j=1,\dots,n}$: sequence of observations, defined through their conditional distributions

$$\mathcal{L}(y_j | z_j = s) \ , \ s = 1, \dots, S \quad (2)$$

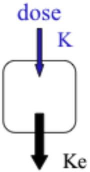
- Appropriate parameterization of (1) and (2) is likely to provide information about treatment effects
ex. :

$$\begin{aligned} \text{logit}(p_{2,1}) &= \theta_1 + \theta_2 \, \text{Dose} \\ \mathbb{E}(y_{ij} | z_{ij} = s) &= \theta_{3,s} + \theta_{4,s}(1 - 2^{-(t_{ij} - \theta_{5,s})/\theta_{6,s}}) \end{aligned}$$

Ex. 2 : pharmacokinetic (PK) studies



Evolution of the concentration of a drug in blood plasma over time



What clinicians need from a model :

- describe the mechanisms of drug diffusion in the body
- estimate the values of key pharmacokinetic parameters (and the effects of covariates on these values)
- ...

Noisy observations of ODEs (ordinary differential equations)

$$\begin{aligned} dZ(t) &= b(Z(t),\theta)dt \\ y_j &= g(Z(t_j),\theta) + \gamma^2(\theta)\epsilon_j \ , \ \epsilon_j \underset{i.i.d.}{\sim} \mathcal{N}(0,1) \end{aligned}$$

Ex. (intraveinous bolus injection)

$$dC(t) = -kC(t)dt$$
$$y_j = C(t_j) + \gamma^2\epsilon_j \ , \ \epsilon_j \underset{i.i.d.}{\sim} \mathcal{N}(0,1)$$

Kinetics are inherently irregular : ~~ODE~~ $\rightarrow$ SDE (stochastic differential equation) [Overgaard R. V. et al., 2005], [Tornoe C. W. et al., 2005]

$$\begin{aligned} dZ(t) &= b(Z(t),\theta)dt + \sigma(Z(t),\theta)dW(t) \\ y_j &= g(Z(t_j),\theta) + \gamma^2(\theta)\epsilon_j \ , \ \epsilon_j \underset{i.i.d.}{\sim} \mathcal{N}(0,1) \end{aligned}$$

Ex. (intraveinous bolus injection)

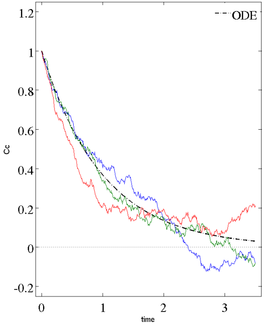
D. and Lavielle [2013]

$$\begin{aligned} dk(t) &= -\alpha(k(t) - k_0)dt + \sigma\sqrt{k(t)}dW(t) \\ dC(t) &= -k(t)C(t)dt \\ y_j &= C(t_j) + \gamma^2\epsilon_j \ , \ \epsilon_j \underset{i.i.d.}{\sim} \mathcal{N}(0,1) \end{aligned}$$

Effects of accounting for kinetics irregularities in the model through SDEs

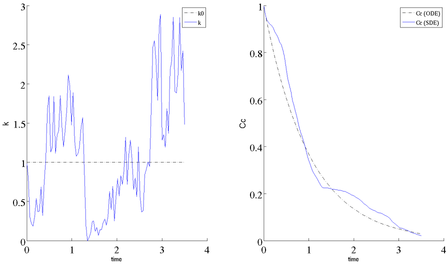
1)

$$dC(t) = -kC(t)dt + \sigma(C(t))dW(t)$$



2)

$$dk(t) = -\alpha(k(t) - k_0)dt + \sigma\sqrt{k(t)}dW(t)$$
$$dC(t) = -k(t)C(t)dt$$



Remarks

- Stochastic processes are used in many other areas : Hawkes processes (seismology, neuroscience, ...), branching processes (population evolution, spread of epidemics, ...), point processes, Levy processes, ...
- Advantage : large variety stochastic model
  - nature of the studied phenomena (discrete/continuous values, discrete/continuous time, ...)
  - specific assumptions on the temporal dependency between observations, the distribution of consecutive events, ...
- Drawback : sample paths are rarely completely observed (hidden states, data aggregation, measurement errors, ...)
  $\hookrightarrow$ require accurate statistical studies
- 2 examples above : several trajectories with extrinsic variability to be treated simultaneously but this is not the usual framework

Remarks

- Stochastic processes are used in many other areas : Hawkes processes (seismology, neuroscience, . . .), branching processes (population evolution, spread of epidemics, . . .), point processes, Levy processes, . . .
- Advantage : large variety stochastic model
  - nature of the studied phenomena (discrete/continuous values, discrete/continuous time, . . .)
  - specific assumptions on the temporal dependency between observations, the distribution of consecutive events, . . .
- Drawback : sample paths are rarely completely observed (hidden states, data aggregation, measurement errors, . . .)
  $\hookrightarrow$ require accurate statistical studies
- 2 examples above : several trajectories with extrinsic variability to be treated simultaneously but this is not the usual framework

Remarks

- Stochastic processes are used in many other areas : Hawkes processes (seismology, neuroscience, ...), branching processes (population evolution, spread of epidemics, ...), point processes, Levy processes, ...

- Advantage : large variety stochastic model
  - nature of the studied phenomena (discrete/continuous values, discrete/continuous time, ...)
  - specific assumptions on the temporal dependency between observations, the distribution of consecutive events, ...

- Drawback : sample paths are rarely completely observed (hidden states, data aggregation, measurement errors, ...)
  $\hookrightarrow$ require accurate statistical studies

- 2 examples above : several trajectories with extrinsic variability to be treated simultaneously but this is not the usual framework

Remarks

- Stochastic processes are used in many other areas : Hawkes processes (seismology, neuroscience, . . .), branching processes (population evolution, spread of epidemics, . . .), point processes, Levy processes, . . .
- Advantage : large variety stochastic model
  - nature of the studied phenomena (discrete/continuous values, discrete/continuous time, . . .)
  - specific assumptions on the temporal dependency between observations, the distribution of consecutive events, . . .
- Drawback : sample paths are rarely completely observed (hidden states, data aggregation, measurement errors, . . .)
  $\hookrightarrow$ require accurate statistical studies
- 2 examples above : several trajectories with extrinsic variability to be treated simultaneously but this is not the usual framework

Useful statistical issues for applications of stochastic processes

1) Estimation (either parametric or nonparametric)
   a. computational developments when exact computation of estimators is costly or not feasible
   b. theoretical study of estimators

2) Reconstruction of hidden states from observations (prediction)
   $\hookrightarrow$ characterization of the conditional distribution of the states given the observations

$$\mathcal{L}(Z_t | Y_{t-1}, \dots, Y_0) \quad \text{filtering}$$
$$\mathcal{L}(Z_t | Y_T, \dots, Y_0) \quad \text{smoothing}$$

   (e.g. Kalman methods and their derivations, particle methods, ...)

3) Model selection
   (e.g. selecting the number of states in a HMM)

1) Estimation (either parametric or nonparametric)
   a. computational developments when exact computation of estimators is costly or not feasible
   b. theoretical study of estimators

2) Reconstruction of hidden states from observations (prediction)
   $\hookrightarrow$ characterization of the conditional distribution of the states given the observations

$$\mathcal{L}(Z_t | Y_{t-1}, \dots, Y_0) \quad \text{filtering}$$
$$\mathcal{L}(Z_t | Y_T, \dots, Y_0) \quad \text{smoothing}$$

   (e.g. Kalman methods and their derivations, particle methods, . . .)

3) Model selection
   (e.g. selecting the number of states in a HMM)

Contents

Contents

General setting : mixed-effects models

Data :

- repeated measurements from several subjects
- similarly-shaped individual profiles
- inter-individual differences

Mixed effects models : hierarchical models s.t. [Pinheiro J. C. and Bates D. M., 2000], [Lavielle M., 2014]

1. 1st level : description of the intra-individual variability

$$Y_i \underset{ind.}{\sim} p(\cdot|\psi_i) \ , \ i = 1, \ldots, N$$

   - $y_i = (y_{i,1}, \ldots, y_{i,n_i})^\top$: observations for subject i
   - $\psi_i \in \mathbb{R}^d$: individual parameters

2. 2nd level : description of the inter-individual variability

$$\psi_i \underset{ind.}{\sim} p(\cdot; C_i, \theta)$$

   - $C_i \in \mathbb{R}^q$: covariates for individual i
   - θ : population parameter

General setting : SDEs with mixed-effects

SDEs with mixed-effects :

$$\begin{aligned} dZ_i(t) &= b(Z_i(t),\psi_i)dt + \sigma(Z_i(t),\psi_i)dW_i(t) \\ y_{ij} &= g(Z_i(t_{ij}),\psi_i) + \gamma^2(\psi_i)\epsilon_{ij} \ , \ \epsilon_{ij} \underset{i.i.d.}{\sim} \mathcal{N}(0,1) \\ \psi_i &\underset{i.i.d.}{\sim} g(\cdot,\theta) \end{aligned}$$

Maximum likelihood estimation : $\hat{\theta}_N = \arg\max_{\theta\in\Theta} L_N(\theta)$

→ difficult to compute

→ likelihood generally non explicit

$$L_N(\theta) = \prod_{i=1}^{N} \int p(Y_i|\psi_i)p(\psi_i;\theta)d\psi_i$$

where

$$p(Y_i|\psi_i) = \prod_{j=1}^{n_i} \int p(y_{ij}|Z_i(t_{ij}),\psi_i)p(Z_i(t_{ij})|y_{i,j-1},\ldots,y_{i,1},\psi_i)dZ_i(t_{ij})$$

→ Questions : How to compute $\hat{\theta}_N$ and what can we say about the properties of $\hat{\theta}_N$?

SDEs with mixed-effects :

$$\begin{aligned} dZ_i(t) &= b(Z_i(t), \psi_i)dt + \sigma(Z_i(t), \psi_i)dW_i(t) \\ y_{ij} &= g(Z_i(t_{ij}), \psi_i) + \gamma^2(\psi_i)\epsilon_{ij} \ , \ \epsilon_{ij} \underset{i.i.d.}{\sim} \mathcal{N}(0,1) \\ \psi_i &\underset{i.i.d.}{\sim} g(\cdot, \theta) \end{aligned}$$

Maximum likelihood estimation : $\hat{\theta}_N = \arg\max_{\theta \in \Theta} L_N(\theta)$

$\rightarrow$ difficult to compute

$\rightarrow$ likelihood generally non explicit

$$L_N(\theta) = \prod_{i=1}^{N} \int p(Y_i|\psi_i)p(\psi_i;\theta)d\psi_i$$

where

$$p(Y_i|\psi_i) = \prod_{j=1}^{n_i} \int p(y_{ij}|Z_i(t_{ij}), \psi_i)p(Z_i(t_{ij})|y_{i,j-1}, \ldots, y_{i,1}, \psi_i)dZ_i(t_{ij})$$

$\rightarrow$ Questions : How to compute $\hat{\theta}_N$ and what can we say about the properties of $\hat{\theta}_N$?

Contents

SAEM [Delyon et al, 1999] Iteration $m > 0$:

- S step *(Simulation)*: simulation of $\phi^{(m)}$ according to $p(\underbrace{\phi}_{latent} | \underbrace{Y}_{obs.}, \theta_{m-1})$
- SA step *(Stochastic approximation)*:

$$Q_m(\theta) = Q_{m-1}(\theta) + \gamma_m \left(\log p(Y, \phi^{(m)}; \theta) - Q_{m-1}(\theta) \right)$$

- M step *(Maximisation)*:

$$\hat{\theta}_m = \underset{\theta \in \Theta}{\mathrm{argmax}} \; Q_m(\theta)$$

where $(\gamma_m)_{m \geq 1}$ is a sequence of decreasing step sizes, s.t. $\gamma_1 = 1$ and $\lim_{m \to \infty} \gamma_m = 0$.

SAEM-MCMC [Kuhn et Lavielle, 2005] : coupling S step with MCMC procedures

Extension of the S step of SAEM-MCMC [D. and Lavielle, 2013]

SDEs with mixed effects

$\rightarrow$ two groups of latent variables $\phi = ((\psi_i, Z_i), i = 1, \ldots, N)$

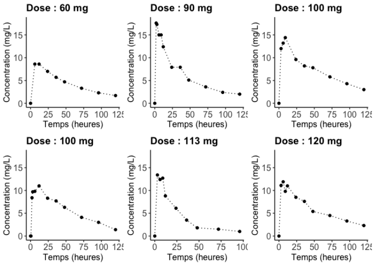
S step:

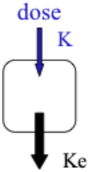
- simulation of $(\psi_i, Z_i)^{(c)} \sim p(\psi_i, Z_i | Y_i; \theta)$ is difficult and costly [Donnet S. and Samson A., 2014]
- marginalization w.r.t. Z_i
- simulation of $\psi_i \sim p(\psi_i | Y_i; \theta)$: Metropolis-Hastings
    - evaluation of the acceptance rate $\leftrightarrow p(\psi_i | Y_i; \theta) \propto p(Y_i | \psi_i) p(\psi_i; \theta)$
    - (extended) Kalman filter to evaluate

$$p(Y_i|\psi_i) = \prod_{j=1}^{n_i} \int p(y_{ij}|Z_i(t_{ij}), \psi_i) p(Z_i(t_{ij})|y_{i,j-1}, \ldots, y_{i,1}, \psi_i) dZ_i(t_{ij})$$

Advantage : reduced computational cost

Real data analysis [D. and Lavielle, 2013]





$$C_i(t) = \frac{D_i}{V_i} \frac{k_i}{k_i - k_{ei}(t)} [\exp(-k_{ei}(t)\, t) - \exp(-k_i\, t)]$$

$$y_{ij} = C_i(t) + \epsilon_{ij} \,,\ \epsilon_{ij} \underset{i.i.d.}{\sim} \mathcal{N}(0, \sigma^2)$$

Contents

Asymptotic studies of estimators

Need for simplified frameworks

1. $$d\,Z_i(t) = \underbrace{\phi_i^\top}_{\phi_i \sim \mathcal{N}(\mu,\Omega)} b(Z_i(t))dt + \gamma^{-1/2}\sigma(Z_i(t))\,dW_i(t), \quad i = 1,\ldots,N$$

2. $$d\,Z_i(t) = \varphi^\top b(Z_i(t))dt + \underbrace{\Gamma_i^{-1/2}}_{\Gamma_i \sim \Gamma(a,\lambda)} \sigma(Z_i(t))\,dW_i(t)$$

3. $$d\,Z_i(t) = \underbrace{\phi_i^\top}_{\phi_i|\Gamma_i = \gamma \sim \mathcal{N}(\mu,\gamma^{-1}\Omega)} b(Z_i(t))dt + \underbrace{\Gamma_i^{-1/2}}_{\Gamma_i \sim \Gamma(a,\lambda)} \sigma(Z_i(t))\,dW_i(t)$$

Observations :

- observations in discrete time and without noise : $y_{ij} = Z_i(t_j)$, $j = 0,\ldots,n$
- regularly spaced observations (Δ) on $[0,T]$ where $T < \infty$ (*i.e.* $n = T/\Delta$, $t_j = j\Delta$)
- $N \to \infty$, $\Delta \to 0$ ($n \to \infty$)

$\Rightarrow$ Contrasts derived from the Euler scheme of the N trajectories are explicit.

Summary of results [D. et al., 2017, 2018]

Results

Under suitable assumptions on the processes :

- *consistency and asymptotic normality when $N, n \to \infty$, $N/n \to 0$*
- *convergence rate $\sqrt{Nn}$: fixed effect in the diffusion coefficient (γ)*
- *convergence rate $\sqrt{N}$: all other parameters*

Remarks :

- If $N = 1$, the observation time interval $[0; T]$ is fixed, $\Delta \to 0$ ($n \to \infty$) only γ can be estimated consistently and the rate of convergence of the estimator is $\sqrt{n}$.
- Links with asymptotic results in [Nie L. and Yang M., 2005], [Nie L., 2006, 2007] in the standard nonlinear mixed-effects models framework.

Closing words

- Stochastic processes : useful models for many applications where data are collected over time
- Making statistics for such models is complex (non i.i.d. setting, . . .)
- Still an active research field

References

D. and Lavielle. Coupling the saem algorithm and the extended kalman filter for maximum likelihood estimation in mixed-effects diffusion models. *Stat. Interface*, 2013.

D., Genon-Catalot, and Larédo. Parametric inference for discrete observations of diffusion processes with mixed effects. *Stoch. Proc. Appl.*, 2017.

D., Genon-Catalot, and Larédo. Approximate maximum likelihood estimation for stochastic differential equations with random effects in the drift and the diffusion. *Metrika*, 2018.

D., Savic, Miller, Karlsson, and Lavielle. Analysis of exposure-response of ci-945 in patients with epilepsy: application of novel mixed hidden markov modelling methodology. *J. Pharmacokinet. Phar.*, 2012.

Altman R. M. Mixed Hidden Markov Models: An Extension of the Hidden Markov Model to the Longitudinal Data Setting. *Journal of the American Statistical Association*, 102(477):201–210, 2007.

Overgaard R. V., Jonsson N., Tornoe C. W., and Madsen H. Non-linear mixed-effects models with stochastic differential equations: Implementation of an estimation algorithm. *Journal of Pharmacokinetics and Pharmacodynamics*, 32(1):85–107, 2005.

Tornoe C. W., Overgaard R. V., Agerso H., Nielsen H. A., Madsen H., and Jonsson N. Stochastic differential equations in NONMEM : Implementation, application, and comparison with ordinary differential equations. *Pharmaceutical Research*, 22(8):1247–1258, 2005.

Pinheiro J. C. and Bates D. M. *Mixed-Effects Models in S and S-PLUS*. Statistics and Computing. Springer-Verlag New York Inc., 1st ed. edition, 2000.

Lavielle M. *Mixed Effects Models for the Population Approach. Models, Tasks, Methods and Tools.* Chapman and Hall/CRC Biostatistics Series. Chapman and Hall/CRC, 2014.

Donnet S. and Samson A. Using PMCMC in EM algorithm for stochastic mixed models: theoretical and practical issues. *Journal de la Société Francaise de Statistique*, 155:49–72, 2014.

Nie L. and Yang M. Strong consistency of the MLE in nonlinear mixed-effects models with large cluster size. *Sankhya: The Indian Journal of Statistics*, 67 (4):736–763, 2005.

Nie L. Strong consistency of the maximum likelihood estimator in generalized linear and nonlinear mixed-effects models. *Metrika*, 63(2):123–143, 2006.

Nie L. Convergence rate of the MLE in generalized linear and nonlinear mixed-effects models: Theory and applications. *Journal of Statistical Planning and Inference*, 137(6):1787–1804, 2007.