Parametric inference for hypoelliptic stochastic diffusion

Anna Melnykova

Université de Cergy-Pontoise, Université de Grenoble Alpes

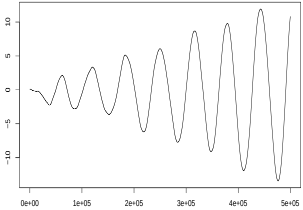
DynStoch 2018, June 2018
Porto, Portugal

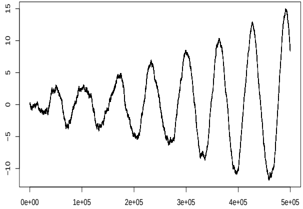
Example 1: Stochastic Damping Hamiltonian System

$$\begin{cases} dX_t = Y_t dt \\ dY_t = -(c(X_t, Y_t) Y_t + \bigtriangledown V(X_t)) dt + \sigma(X_t, Y_t) dW_t \end{cases}$$

- $V(X_t)$ — potential
- $c(X_t, Y_t)$ — damping coefficient
- $\sigma(X_t, Y_t)$ — diffusion coefficient

Examples: noisy Van der Pol oscillator, Kramer's oscillator, linear oscillator.



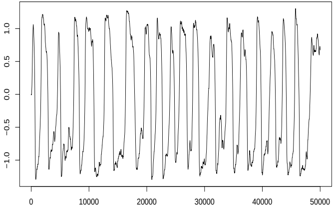


UNIVERSITÉ Grenoble Alpes

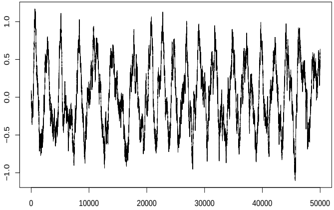
Example 2: Hypoelliptic FitzHugh-Nagumo model

The behaviour of the neuron is defined through

$$\begin{cases} dX_t = \frac{1}{\varepsilon}(X_t - X_t^3 - Y_t - s)dt \\ dY_t = (\gamma X_t - Y_t + \beta)dt + \sigma dW_t \end{cases}$$



- X_t — membrane potential
- Y_t — recovery variable
- s — magnitude of the stimulus current
- Parameters to be estimated are $\theta = (\gamma, \beta, \varepsilon, \sigma)$.



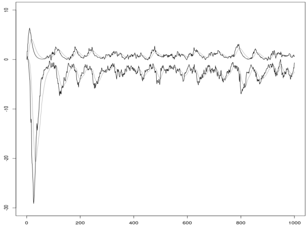
UNIVERSITÉ Grenoble Alpes

Example 3: Stochastic approximation of Hawkes process

Interacting system of neurons is described by:

$$dZ_t = A(Z_t)dt + B(Z_t)dW_t.$$

- κ populations ($= \kappa$ rough variables)
- $\sum_{i=1}^{\kappa} \eta_i$ memory variables (smooth variables)
- N neurons



$$A(z) = \begin{pmatrix} -\nu_1 z^1 + z^2 \\ -\nu_1 z^2 + z^3 \\ \vdots \\ -\nu_1 z^{\eta_1+1} + c_1 f_2(z^{\eta_1+2}) \\ -\nu_2 z^{\eta_1+2} + z^{\eta_1+3} \\ \vdots \\ -\nu_n z^{\kappa} + c_n f_1(z^1) \end{pmatrix}, \quad B(z) = \frac{1}{\sqrt{N}} \begin{pmatrix} 0 & 0 \\ \vdots & \vdots \\ 0 & \frac{c_1}{\sqrt{p_2}} \sqrt{f_2(z^{\eta_1+2})} \\ 0 & 0 \\ \vdots & \vdots \\ \frac{c_n}{\sqrt{p_1}} \sqrt{f_1(z^1)} & 0 \end{pmatrix},$$

Reference: Ditlevsen and Löcherbach (2017)

Model and assumptions

$$\begin{cases} dX_t = a_1(X_t, Y_t; \theta)dt \\ dY_t = a_2(X_t, Y_t; \theta)dt + b(X_t, Y_t; \sigma)dW_t, \end{cases} \tag{1}$$

- $Z_t := (X_t, Y_t)^T \in \mathbb{R} \times \mathbb{R}$,
- $A(Z_t; \theta) := (a_1(X_t, Y_t; \theta), a_2(X_t, Y_t; \theta))^T$ — drift term,
- $B(Z_t; \sigma) := \begin{pmatrix} 0 & 0 \\ 0 & b(X_t, Y_t; \sigma) \end{pmatrix}$ — degenerate diffusion coefficient,
- dW_t is a standard Brownian motion,
- $(\theta, \sigma) \in \Theta_1 \times \Theta_2$ — vector of the unknown parameters.

Goal:

Estimation of (θ, σ) from $(X_i, Y_i)^T$, $i \in 1, \dots N$ on time interval $[0, T]$, $T = N\Delta$.

$\mathbf{T} \to \infty, \Delta \to \mathbf{0}$

UNIVERSITÉ Grenoble Alpes

UNIVERSITÉ de Cergy-Pontoise

Model and assumptions

A1 $a_1(x, y; \theta)$ and $a_2(x, y; \theta)$ have bounded partial derivatives of every order, uniformly in θ. Furthermore $\partial_y a_1 \neq 0 \quad \forall (x, y) \in \mathbb{R}^2$ **(ensures hypoellipticity)**.

A2 Global Lipschitz and linear growth conditions **(ensures existence of a unique strong solution)**.

A3 Process Z_t is **ergodic** and there exists a unique invariant probability measure ν_0 with finite moments of any order.

A4 $a_1(Z_t; \theta)$ and $a_2(Z_t; \theta)$ are **identifiable**, that is $a_i(Z_t; \theta) = a_i(Z_t; \theta_0) \Leftrightarrow \theta = \theta_0$.

Difficulties:

- Degenerate diffusion coefficient $\longrightarrow$ non-invertible covariance matrix of the approximated transition density
- Each coordinate has a variance of different order $\longrightarrow$ numerical instabilities

Solution:

- Use a high-order scheme to ”catch” the propagated noise in all coordinates
- Build a quasi-maximum likelihood estimator based on the approximated density

Related works

- **Stochastic Damping Hamiltonian Systems:**
  - Ozaki (1989), consistency is later proven in León et al. (2018): *Local linearization scheme*
  - Samson and Thieullen (2012): *1-dimensional contrast, Euler Scheme*,
  - Pokern et al. (2007): *Bayesian approach*,
  - Cattiaux et al. (2014), Cattiaux et al. (2016): *non-parametric approach*
- **Linear homogeneous systems:** Le-Breton and Musiela (1985).
- **General systems:** Ditlevsen and Samson (2017): *1.5 strong order scheme*

Discretization: Local linearization scheme

For **Hamiltonian systems with constant diffusion coefficient**: see Ozaki (1989), and León et al. (2018) for consistency.

Generalization (Melnykova, 2018)

On each small time interval of size Δ we approximate (1) by

$$d\mathcal{Z}_s = J_\tau \mathcal{Z}_s ds + B(Z_\tau;\sigma)d\tilde{W}_s, \quad \mathcal{Z}_0 = Z_\tau, \quad s \in (\tau, \tau+\Delta]. \tag{2}$$

- $\mathcal{Z}_0$ – observation of the true process $\{Z_t\}$ at time τ
- J_τ is the Jacobian. **Assumption:** When Δ is small enough, $J_t = const$ and $J_t Z_t = A(Z_t;\theta)$.

Solution of (2) has an explicit form:

$$\mathcal{Z}_s = Z_\tau e^{J_\tau s} + \int_\tau^s e^{J_\tau (s-v)} B(Z_\tau;\sigma)d\tilde{W}_v, \quad \forall s \in (\tau, \tau+\Delta].$$

Discretization: Local linearization scheme

For **Hamiltonian systems with constant diffusion coefficient**: see Ozaki (1989), and León et al. (2018) for consistency.

Generalization (Melnykova, 2018)

On each small time interval of size Δ we approximate (1) by

$$d\mathcal{Z}_s = J_\tau \mathcal{Z}_s ds + B(Z_\tau;\sigma) d\tilde{W}_s, \quad \mathcal{Z}_0 = Z_\tau, \quad s \in (\tau, \tau + \Delta]. \tag{2}$$

- $\mathcal{Z}_0$ – observation of the true process $\{Z_t\}$ at time τ
- J_τ is the Jacobian. **Assumption:** When Δ is small enough, $J_t = \mathit{const}$ and $J_t Z_t = A(Z_t;\theta)$.

Solution of (2) has an explicit form:

$$\mathcal{Z}_s = Z_\tau e^{J_\tau s} + \int_\tau^s e^{J_\tau (s-v)} B(Z_\tau;\sigma) d\tilde{W}_v, \quad \forall s \in (\tau, \tau + \Delta].$$

Discretization: Local linearization scheme

First and second moment of $\mathcal{Z}_s$ on each Δ-interval:

$$\mathbb{E}[\mathcal{Z}_s] = Z_\tau e^{J_\tau s}$$

$$\Sigma(\mathcal{Z}_s; \theta, \sigma^2) = \mathbb{E}\left[\left(\int_\tau^s e^{J_\tau (s-v)} B(Z_\tau; \sigma) d\tilde{W}_v\right)\left(\int_\tau^s e^{J_\tau (s-v)} B(Z_\tau; \sigma) d\tilde{W}_v\right)^T\right].$$

The approximation of the solution of (1) at time $i\Delta$:

$$Z_{i+1} = \bar{A}(Z_i; \theta) + \bar{B}(Z_i; \theta, \sigma)\Xi_i,$$

- Ξ_i — standard Gaussian 2-dimensional random vector
- $\bar{A}$ is a discrete approximation of $\mathbb{E}[\mathcal{Z}_s]$.
- $\bar{B}$ — any matrix s. t. $\bar{B}\bar{B}^T = \Sigma(\mathcal{Z}_s; \theta, \sigma^2)$.

Discretization: Local linearization scheme

Proposition (Discretization of the covariance matrix)

The second-order Taylor approximation of matrix $\Sigma(\mathcal{Z}_\Delta;\theta,\sigma^2)$ *defined in* (10) *has the following form:*

$$b^2(Z_\tau;\sigma)\begin{pmatrix} (\partial_y a_1)^2\frac{\Delta^3}{3} & (\partial_y a_1)\frac{\Delta^2}{2}+(\partial_y a_1)(\partial_y a_2)\frac{\Delta^3}{3} \\ (\partial_y a_1)\frac{\Delta^2}{2}+(\partial_y a_1)(\partial_y a_2)\frac{\Delta^3}{3} & \Delta+(\partial_y a_2)\frac{\Delta^2}{2}+(\partial_y a_2)^2\frac{\Delta^3}{3}\end{pmatrix}+\mathcal{O}(\Delta^4),$$

where the derivatives are computed at time τ*.*

Contrast estimator

The contrast function is defined as follows:

$$\mathcal{L}(\theta,\sigma^2;Z_{0:N}) = \frac{1}{2}\sum_{i=0}^{N-1}(Z_{i+1}-\bar{A}(Z_i;\theta))^T\Sigma_\Delta^{-1}(Z_{i+1}-\bar{A}(Z_i;\theta))+\sum_{i=0}^{N-1}\log\det(\Sigma_\Delta).$$

The estimator is then:

$$(\hat{\theta},\hat{\sigma}^2) = \arg\min_{\theta,\sigma^2}\mathcal{L}(\theta,\sigma^2;Z_{0:N})$$

Theorem

Under assumptions (A1)-(A4) and $\Delta_N \to 0$ and $N\Delta_N \to \infty$ the following holds:

$$(\hat{\theta},\hat{\sigma}^2) \xrightarrow{\mathbb{P}_\theta} (\theta_0,\sigma_0^2)$$

Outline of the proof

- Ergodic theorem $\longrightarrow$ convergence in probability of the approximated density to true transition density
- Identifiability of the diffusion coefficient:

$$\arg\min_{\sigma^2} \lim_{\Delta\to 0, N\to\infty} \frac{1}{N}\mathcal{L}_{N,\Delta_N}(\theta,\sigma^2;Z_{0:N}) \to \sigma_0^2$$

- **Parameters of the drift term are identifiable under different scaling!**
  - Smooth variable:

$$\arg\min_{\theta} \lim_{\Delta\to 0, N\to\infty} \frac{\Delta_N}{N}\left[\mathcal{L}_{N,\Delta_N}(\theta,\sigma_0^2;Z_{0:N}) - \mathcal{L}_{N,\Delta_N}(\theta_0,\sigma_0^2;Z_{0:N})\right] \to \theta_0$$

  - Rough variable:

$$\arg\min_{\psi} \lim_{\Delta\to 0, N\to\infty} \frac{1}{N\Delta_N}\left[\mathcal{L}_{N,\Delta_N}(\varphi_0,\psi,\sigma_0^2;Z_{0:N}) - \mathcal{L}_{N,\Delta_N}(\varphi_0,\psi_0,\sigma_0^2;Z_{0:N})\right] \to \psi_0$$

Outline of the proof

- Ergodic theorem $\longrightarrow$ convergence in probability of the approximated density to true transition density
- Identifiability of the diffusion coefficient:

$$\arg\min_{\sigma^2} \lim_{\Delta\to 0, N\to\infty} \frac{1}{N}\mathcal{L}_{N,\Delta_N}(\theta, \sigma^2; Z_{0:N}) \to \sigma_0^2$$

- **Parameters of the drift term are identifiable under different scaling!**
  - Smooth variable:

$$\arg\min_{\theta} \lim_{\Delta\to 0, N\to\infty} \frac{\Delta_N}{N}\left[\mathcal{L}_{N,\Delta_N}(\theta, \sigma_0^2; Z_{0:N}) - \mathcal{L}_{N,\Delta_N}(\theta_0, \sigma_0^2; Z_{0:N})\right] \to \theta_0$$

  - Rough variable:

$$\arg\min_{\psi} \lim_{\Delta\to 0, N\to\infty} \frac{1}{N\Delta_N}\left[\mathcal{L}_{N,\Delta_N}(\varphi_0, \psi, \sigma_0^2; Z_{0:N}) - \mathcal{L}_{N,\Delta_N}(\varphi_0, \psi_0, \sigma_0^2; Z_{0:N})\right] \to \psi_0$$

Numerical performance: FitzHugh-Nagumo model

- Data (1000 trajectories) generated with $N = 500000, \Delta = 0.0001$
- Contrast is minimized with respect to subsampled data ($N = 50000, \Delta = 0.001$)
- Minimization: `optim` in R, method: Conjugate Gradient

	γ	β	ε	σ
Set 1:	**1.5**	**0.3**	**0.1**	**0.6**
Lin. contrast	1.477 (1.056)	0.289 (0.428)	0.100 (0.561)	0.672 (0.291)
1.5 scheme	1.497 (1.055)	0.299 (0.393)	0.099 (0.563)	0.597 (0.288)
Set 2:	**1.2**	**1.3**	**0.1**	**0.4**
Lin. contrast	1.199 (0.531)	1.315 (0.621)	0.102 (0.683)	0.472 (0.340)
1.5 scheme	1.221 (0.645)	1.324 (0.777)	0.088 (0.575)	0.398 (0.338)

Table: Comparison between presented scheme and the method presented in Ditlevsen and Samson (2017) (separate estimation of parameters presented in each equation)

UNIVERSITÉ Grenoble Alpes

UNIVERSITÉ de Cergy-Pontoise

Conclusions

Strong points:

- Straightforward implementation
- All parameters are estimated simultaneously

Weak points:

- Numerical instability due to the different order of variance
- Sensitivity to the initial value of the parameters
- Asymptotic normality?

Future work:

- Partial observation case
- Multidimensional system

Acknowledgements

Financial support: LabEx PERSYVAL-Lab, LABEX MME-DII, Laboratoire Jean Kuntzmann (University of Grenoble Alpes), AGM (University of Cergy-Pontoise).

Literature

Cattiaux, P., León, J. R., and Prieur, C. (2014). Estimation for stochastic damping hamiltonian systems under partial observation. ii drift term. *ALEA*, 11(1):p–359.

Cattiaux, P., León, J. R., Prieur, C., et al. (2016). Estimation for stochastic damping hamiltonian systems under partial observation. iii. diffusion term. *The Annals of Applied Probability*, 26(3):1581–1619.

Ditlevsen, S. and Löcherbach, E. (2017). Multi-class oscillating systems of interacting neurons. *SPA*, 127:1840–1869.

Ditlevsen, S. and Samson, A. (2017). Hypoelliptic diffusions: discretization, filtering and inference from complete and partial observations. *submitted*.

Le-Breton, A. and Musiela, M. (1985). Some parameter estimation problems for hypoelliptic homogeneous gaussian diffusions. *Banach Center Publications*, 16(1):337–356.

León, J., Rodriguez, L., and Ruggiero, R. (2018). Consistency of a likelihood estimator for stochastic damping hamiltonian systems. totally observed data.

Melnykova, A. (2018). Parametric inference for multidimensional hypoelliptic diffusion with full observations. *ArXiv e-prints*.

Ozaki, T. (1989). Statistical identification of nonlinear random vibration systems. *Journal of Applied Mechanics*, 56:186–191.

Pokern, Y., Stuart, A. M., and Wiberg, P. (2007). Parameter estimation for partially observed hypoelliptic diffusions. *J. Roy. Stat. Soc.*, 71(1):49–73.

Samson, A. and Thieullen, M. (2012). Contrast estimator for completely or partially observed hypoelliptic diffusion. *Stochastic Processes and their Applications*, 122:2521–2552.