Statistical inference for structured models

Some connections between nonparametric estimation and PDEs. Lecture I: the stochastic models

Marc Hoffmann, Université Paris-Dauphine PSL

Spring school SFB 1294, March 2019
Acknowledgement

Based on a series of works jointly written with

Alexandre Boumezoued\textsuperscript{1}, Sylvain Delattre\textsuperscript{2}, Marie Doumic\textsuperscript{3,4}, Nicolas Fournier\textsuperscript{4}, Paulien Jeunesse\textsuperscript{5}, Nathalie Krell\textsuperscript{6}, Eva Löcherbach\textsuperscript{7}, Aline Marguet\textsuperscript{3,8}, Adélaïde Olivier\textsuperscript{9}, Vincent Rivoirard\textsuperscript{5}, Patricia Reynaud-Bouret\textsuperscript{10}, Lydia Robert\textsuperscript{11}.

\textsuperscript{1}Milliman, \textsuperscript{2}Université Paris-Diderot, \textsuperscript{3}INRIA, \textsuperscript{4}Sorbonne Université, \textsuperscript{5}Université Paris-Dauphine, \textsuperscript{6}Université de Rennes 1, \textsuperscript{7}Université Paris 1, \textsuperscript{8}Ecole Polytechnique, \textsuperscript{9}Université Paris-Sud, \textsuperscript{10}Université de Nice, \textsuperscript{11}INRA.
Disclaimer

- Lecture notes in progress!
- Will be updated according to the wishes of the audience (hopefully!)
- Bibliography still to be processed!
Overview and informal structure of the study

- Work in progress!
- Falls into the scope of statistical inference with PDE.
- **Statistical setting:** We have (i) data $Z^N$ and (ii) a parameter of interest $f$. Asymptotics are taken as $N \to \infty$.
- **Structure of the problem:**

\[
\mathcal{H}_N(Z^N) = 0 \text{ for some SDE } \mathcal{H}_N,
\]
\[
Z^N \to \xi \text{ limiting object,}
\]
\[
\mathcal{H}(\xi, f) = 0 \text{ for some PDE } \mathcal{H}.
\]

- **Objective:** recover $f$ from the observation of $Z^N$, or a proxy $Z^N$ of $Z^N$. 
Overview and informal structure of the study

We need to specify several objects and a methodology
1. What are $Z^N$ and $f$? (and therefore the meaning of $H_N$ and $H$)
2. What is $N$?
3. What do we mean by a proxy $Z^N$ of $Z^N$?
4. What do we mean by recovering $f$ (as $N \to \infty$)?

We do not have a “nice theory” at this stage. We will rather explore these questions via several examples.

We have a relatively complete picture in some cases.

For other examples, we have more questions than answers!
Paradigmatic examples

1. Cell division: growth-fragmentation models
   ▶ Age-structured models and the renewal equation
   ▶ Size-structured models

2. General bifurcating models

3. Human population models for demography \(\rightsquigarrow\) Lecture II
   ▶ Cohort effects in human mortality
   ▶ Towards nonlinearity

4. Models of interacting neurons \(\rightsquigarrow\) Lecture IV
   ▶ Spikes models
   ▶ Hawkes models

5. More nonlinear models in a mean-field limit \(\rightsquigarrow\) Lecture IV
Paradigmatic examples

Cell division: growth-fragmentation models

General bifurcating models
Example 1: Growth-fragmentation models

- We consider (simple) branching processes with deterministic evolution between jump times.
- Such models appear as toy models for population growth in cellular biology.
- We wish to statistically estimate the parameters of the model, in order to ultimately discriminate between different hypotheses related to the mechanisms that trigger cell division.
Example 1: Growth-fragmentation models

- We structure the model by state variables for each individual like size, age, growth rate, DNA content and so on.
- The evolution of the particle system is described by a common mechanism:
  1. Each particle grows by ingesting a common nutrient = deterministic evolution.
  2. After some time, depending on a structure variable, each particle gives rise to $k = 2$ offsprings by cell division = branching event.
- Goal: estimate the branching rate as a function of age or size (or both).
Figure: Evolution of a *E. Coli* population.
Figure: Evolution of a *Escherichia coli* population.
Figure: Evolution of a *E. Coli* population.
Figure: Evolution of a *E. Coli* population.
Figure: Evolution of a *E. Coli* population.
Growth-fragmentation models, informal description

- The growth of a cell size follows the deterministic evolution
  \[ \frac{dX(t)}{dt} = \kappa(X(t)) \, dt. \]

- A mother cell of size \( x \) and age \( a \) splits into two daughter cells with a division rate \( B \) depending on its age \( a \) or size \( x \).

- 1 cell of size \( x \) gives birth to 2 cells of size \( x/2 \).
Example 1: mathematical description

- $\mathcal{M}_F$ the set of finite point measures on $\mathbb{R}_+ = [0, \infty)$.
- For $M \in \mathcal{M}_F$ and test function $\varphi \geq 0$:

$$\langle M, g \rangle = \int_{[0,\infty)} \varphi(s) M(ds) = \sum_{i=1}^{\langle M, 1 \rangle} \varphi(x_i)$$

for a finite (ordered) family $(x_i) \leftrightarrow M$ of nonnegative $x_i$.
- **Evaluation maps**: $x_i : \mathcal{M}_F \to [0, \infty) : M \mapsto x_i(M) = x_i$.
- In particular $\langle M, 1 \rangle$ = size of the population $M$. 
Example 1.1: An age-structured model

\[
(A_i(t))_{1 \leq i \leq N_t} = \text{all the (ordered) ages of the cell population at time } t.
\]

\[
Z_t = \sum_{i=1}^{N_t} \delta A_i(t) \text{ with } N_t = \langle Z_t, 1 \rangle.
\]

- The division rate \( a \mapsto B(a) \) is a function of age only!

Associated SDE:

\[
Z_t = \tau_t Z_0 + \int_0^t \sum_{i \leq \langle Z_s, 1 \rangle} \int_{0 \leq \theta \leq B(a_i(Z_s -) + t - s)} \left(2\delta_{t-s} - \delta_{a_i(Z_s -) + t - s}\right) Q(ds, di, d\theta)
\]

- \( Q \) : Poisson random measure, intensity \( ds \left( \sum_{k \geq 1} \delta_k(di) \right) d\theta. 
- \( \tau_t \sum_i \delta a_i = \sum_i \delta a_i + t. \)
Example 1.1: An age-structured model

Figure: A sample path of $Z_t(da)_{0 \leq t \leq T}$ with $B(a) = a^2$ and $T = 7$. 
Example 1.1: limiting object

- \( N = N_T = \langle Z_T, 1 \rangle \to \infty \text{ as } T \to \infty. \)
- Here \( N \) is random and the asymptotics are transferred to \( T.\)
- Heuristically \( Z_T \approx E[Z_T(da)] = \xi_T(da) = g(T, a)da. \)
- \( g(t, x) \) is a weak solution to the renewal equation:

\[
\begin{align*}
\partial_t g(t, a) + \partial_a g(t, a) + B(a)g(t, a) &= 0 \\
g(0, a) &= g_0(a), \quad g(t, 0) = 2 \int_0^\infty B(a)g(t, a)da.
\end{align*}
\]
Example 1.1: identification of the objects of interest

We need to specify several objects and a methodology

1. What are $Z^N$ and $f$? (and therefore the meaning of $H_N$ and $H$)
2. What is $N$?
3. What do we mean by a proxy of $Z^N$?
4. What do we mean by recovering $f$ (as $N \to \infty$)?

We have identified the following objects

- $N$ is $\langle Z_T, 1 \rangle$.
- $Z^N$ is $(Z_t)_{0 \leq t \leq T}$.
- $f$ is $(t, a) \mapsto g(t, a)$ or $a \mapsto B(a)$.
- $H^N$ and $H$ are the SDE and the renewal equation.
Example 1.2: A size-structured model

- \((X_i(t))_{1 \leq i \leq N_t}\) = all the sizes of the cell population at time \(t\).

- \(Z_t = \sum_{i=1}^{N_t} \delta X_i(t)\) with \(N_t = \langle Z_t, 1 \rangle\).

- The division rate \(x \mapsto B(x)\) is a function of size only!

Associated SDE:

\[
Z_t = \phi Z_0(t) + \int_0^t \sum_{\delta \leq \langle Z_{s-}, 1 \rangle} \int_{0 \leq \theta \leq B(a_i(Z_{s-}))} \left(2\delta \phi_{x_i(Z_{s-})}(t-s) - \delta \phi_{x_i(Z_{s-})}(t-s) \right) Q(ds, di, d\theta)
\]

- \(\frac{d}{dt} \phi_x(t) = \kappa(\phi_x(t))\) with \(\phi_x(0) = x\).

- \(\phi \sum_i \delta x_i(t) = \sum_i \delta \phi x_i(t)\).
Example 1.2: limiting object

- $N = N_T = \langle Z_T, 1 \rangle \to \infty$ as $T \to \infty$.
- $N$ is random and the asymptotics are transferred to $T$.
- Heuristically $Z_T \approx \mathbb{E}[Z_T(dx)] = \xi_T(dx) = g(T, x)dx$
- $g(t, x)$ is a weak solution to the transport-fragmentation equation:

\[
\begin{align*}
\partial_t g(t, x) + \partial_x (\kappa(x)g(t, x)) + B(x)g(t, x) &= 4B(2x)g(t, 2x) \\
g(0, x) &= g_0(x) \text{ and } g(t, 0) = 0, t > 0.
\end{align*}
\]
Example 1.2: identification of the objects of interest

- Similarly to Example 1.1, we can identify the following objects
  - $N$ is $\langle Z_T, 1 \rangle$.
  - $Z^N$ is $(Z_t)_{0 \leq t \leq T}$.
  - $f$ is $(t, x) \mapsto g(t, x)$ or $x \mapsto B(x)$.
  - $H^N$ and $H$ are the SDE and the transport-fragmentation equation.
Observation schemes, proxy of $Z^N$

We will distinguish several observation schemes $Z^N$ depending on the observation devices at hand:

(a) We observe the whole path $Z^N = (Z_t)_{0 \leq t \leq T}$

(b) We observe $Z$ at a terminal time $t = T$. Two situations
   - $Z^N = Z_T$.
   - $Z^n$ is a $n$-sample of $Z_T$, with $n \ll N = \langle Z_T, 1 \rangle$. Our data is thus a proxy of $Z_T$ with $n \to \infty$ (as $N \to \infty$).

(c) $Z^N$ is realised as a subsample of size $N$ of $Z_T$ at certain stopping times that correspond to branching events. Again, we need $N$ large as $T \to \infty$. 
Observing $Z$ via a genealogical representation

- We elaborate on the observation schemes (a), (b) and (c) by means of a genealogical representation.
- The population evolution is associated with an infinite marked binary tree

$$
T = \bigcup_{m \in \mathbb{N}} G_m, \quad G_m = \{0, 1\}^m, \quad (G_0 = \emptyset).
$$

- To each cell or node $u \in T$, we associate a node with size at birth $\xi_u$ and lifetime $\zeta_u$.
- To each $u \in T$, we associate a birth time $b_u$ and a time of death $d_u$ so that $\zeta_u = d_u - b_u$. 
The process $Z$ via a genealogical representation

We have the following identity between point measures

- **Example 1.1** (age model)

$$Z_t = \sum_{u \in \mathbb{T}} \delta_{t-b_u} 1_{\{b_u \leq t < b_u + \zeta_u\}}$$

- **Example 1.2** (size model)

$$Z_t = \sum_{u \in \mathbb{T}} \delta_{\phi_{\xi_u}(t-b_u)} 1_{\{b_u \leq t < b_u + \zeta_u\}}.$$
Observation scheme (a) and (b): temporal data

We introduce random subsets of $\mathbb{T}$

$\mathcal{T}_T = \{ u \in \mathbb{T}, b_u \leq T \} = \mathcal{\hat{T}}_T \cup \partial \mathcal{T}_T$, with

$\mathcal{\hat{T}}_T = \{ u \in \mathbb{T}, d_u \leq T \}$ and $\partial \mathcal{T}_T = \{ u \in \mathbb{T}, b_u \leq T < d_u \}$.

In Example 1.1 (age model) for observation schemes (a) and (b), we have

$$(Z_t)_{0 \leq t \leq T} = \{ \zeta_u^T = \min(d_u, T) - b_u, u \in \mathcal{T}_T \},$$

$Z_T = \{ \zeta_u^T, u \in \partial \mathcal{T}_T \}.$$
Figure: Genealogical tree observed up to $T = 7$ for a time-dependent division rate $B(a) = a^2$ (60 cells). In blue: $\hat{T}_T$. In red: $\partial T_T$. 

Temporal data
Observation scheme (a) and (b): temporal data

- $|\mathcal{T}_T|$ and $|\partial \mathcal{T}_T|$ are of the same order of magnitude.
- In Example 1.2 (size model), for observation schemes (a) and (b), we have

\[
(Z_t)_{0 \leq t \leq T} = \{ \zeta^T_u, \xi^T_u, u \in \mathcal{T}_T \},
\]

\[
Z_T = \{ \xi^T_u, u \in \partial \mathcal{T}_T \},
\]

with $\xi^T_u = \xi_u$ if $d_u \leq T$ and $\phi_{\xi_u}(T - b_u)$ otherwise.
Observation scheme (c): genealogical data

- Introduce the **binary tree** up to the first $n$-generations

$$
T_n = \bigcup_{m=0}^{n} G_m.
$$

- Observation scheme (c): informally, for some $n \geq 1$, we observe $(\zeta_u, \xi_u)$ along a subset of $T_n$, called a $\varrho$-regular tree.

**Definition**

$U_n \subseteq T_n$ is a $\varrho$-regular tree if

1. $u \in U_n \implies u^- \in U_n$ ($u^-$ parent of $u$)
2. $|U_n \cap G_n| \approx 2^{n\varrho}$ ($0 \leq \varrho \leq 1$).

**Two extreme cases**

- **Dense case:** $U_n = T_n$, with $|U_n| = 2^{n+1} - 1$ (and $\varrho = 1$).
- **Sparse case:** $U_n$ a single line along $T_n$, with $|U_n| = n$ (and $\varrho = 0$).
Observation scheme (c): genealogical data

- Take a $\varrho$-regular tree $\mathbb{U}_n$ with size $N = \kappa 2^{\varrho n}$.
- Define the (now random) time
  \[
  T = \inf\{ t \geq 0, \ Z_t \text{ has visited all the nodes of } \mathbb{U}_n \}\]
- The observation scheme is then $\mathcal{Z}^N = (\xi_u, \zeta_u)_{u \in \mathbb{U}_n}$, extracted from $(Z_t)_{0 \leq t \leq T}$.
- $N = \kappa 2^{\varrho n} \to \infty$ as $n \to \infty$.
- Now, $T = T(N, (Z_t)_{t \geq 0}) \to \infty$ as $n \to \infty$. 
Temporal versus genealogical data: selection bias!

Figure: Genealogical tree observed up to $T = 7$ for a time-dependent division rate $B(a) = a^2$ (60 cells). In blue: $\mathcal{U}_T$. In red: $\partial \mathcal{U}_T$. 
Genealogical data

Figure: The same outcome organised at a genealogical level.

The handling of genealogical data (via discrete Markov chain techniques) will prove significantly easier than temporal data.
Paradigmatic examples

Cell division: growth-fragmentation models

General bifurcating models
General bifurcating models

- Observation scheme (c) \(\rightsquigarrow\) digression on bifurcating Markov chains models.
- Bifurcating Markov chains = Markov chains on binary trees.
- Extension to non-deterministic evolution between jumps.
- So far, evolution given by \(\phi_x(t) = \text{value of the trait at time } t \) with initial value \(x\) at \(t = 0\):
  - Example 1.1 (age model)
    \[ \phi_0(t) = t. \]
  - Example 1.2 (size model)
    \[ d\phi_x(t) = \kappa(\phi_x(t)) \, dt, \quad \phi_x(0) = x \in (0, \infty). \]
- We may think of more general flows in between jumps.
Example 2: more general flows

- Binary division triggered by a \textit{trait} \( x \in X \subseteq \mathbb{R} \).
- The trait \textit{stochastically evolves} according to

\[
d\phi_x(t) = \kappa(\phi_x(t))dt + \sigma(\phi_x(t))dW_t, \quad \phi_x(0) = x
\]

\( r, \sigma : X \rightarrow X \) regular functions, \((W_t)_{t \geq 0}\) standard BM.
- A \textbf{branching event} occurs with probability

\[
B(\phi_x(t))dt \quad \text{during} \quad [t, t + dt]
\]

\( B : X \rightarrow [0, \infty) \) division rate.
- At division, a particle with trait \( y \) is replaced by \textbf{two particles with traits} \( y\vartheta \) and \( y(1 - \vartheta) \), where \( \text{Law}(\vartheta) = r(dy) \).
- Parameters of the model: \((\kappa, \sigma, B, r)\).

Bifurcating Markov chains

Figure: Example of a trajectory of a BMC and its associated genealogy.
Bifurcating Markov chains

- $\mathcal{X}$ a state space and $\mathcal{P} : \mathcal{X} \to \mathcal{X} \times \mathcal{X}$ a Markov kernel.
- $(\Omega, \mathcal{F}, (\mathcal{F}_m)_{m \geq 0}, \mathbb{P})$ a filtered probability space.
- **Notation:** $|u| = m$ for $u \in \mathcal{G}_m$.
  
  $u_0 = (u, 0) \in \mathcal{G}_{m+1}$, $u_1 = (u, 1) \in \mathcal{G}_{m+1}$.
  
  $\mathcal{P} \psi(x) = \int_{\mathcal{X} \times \mathcal{X}} \psi(x, y_1, y_2) \mathcal{P}(x, dy_1dy_2)$.

**Definition**

A family of $\mathcal{X}$-valued r.v. $(\mathcal{X}_u)_{u \in \mathbb{T}}$ is a BMC with transition $\mathcal{P}$ if $\mathcal{X}_u$ is $\mathcal{F}_{|u|}$-measurable and

$$
\mathbb{E}\left[ \prod_{u \in \mathcal{G}_m} \psi_u(\mathcal{X}_u, \mathcal{X}_{u_0}, \mathcal{X}_{u_1}) \mid \mathcal{F}_m \right] = \prod_{u \in \mathcal{G}_m} \mathcal{P} \psi_u(\mathcal{X}_u)
$$

for every $m \geq 0$ and any family of (bounded) functions $(\psi_u)_{u \in \mathcal{G}_m}$. 
Representation of $\mathcal{P}$ for Examples 1.1 and 1.2

- **Example 1.1** $X_u = \zeta_u$ for the age model:

  $$\mathcal{P}(x, dy_1 \, dy_2) = \delta_0(dy_1) \otimes \delta_0(dy_2)$$

- **Example 1.2** $X_u = \xi_u$ for the size model:

  $$\mathcal{P}(x, dy_1 \, dy_2) = Q(x, y_1) dy_1 \otimes \delta_{y_1}(dy_2)$$

with

$$Q(x, y) = \frac{B(2y)}{\frac{1}{2} \kappa(\phi^{-1}_x(2y))} \exp \left( - \int_{2x}^{y} \frac{B(2z)}{\frac{1}{2} \kappa(\phi^{-1}_x(2y))} dz \right) 1\{y \geq x/2\}$$

obtained under appropriate regularity properties on the flow $\phi$ via $X_u = 2\phi X_{u-}(\zeta_u)$ and the fact that

$$\mathbb{P}(\zeta_u \geq t + dt, |\zeta_u \geq t, X_{u-}) = B(\phi X_{u-}(t)) dt.$$
Representation of $\mathcal{P}$ for Example 2

- Stochastic flow $\sim$ formulas become more intricate.
- Under appropriate regularity conditions, we have

$$\mathcal{P}(x, dy_1 dy_2) = p(x, y_1, y_2) dy_1 dy_2$$

with $p(x, y_1, y_2)$ given by

$$r\left(\frac{y_1}{y_1 + y_2}\right) B(y_1 + y_2) \mathbb{E}\left[\int_0^\infty e^{-\int_0^t B(\phi_x(s)) ds} \frac{dL_{t}^{y_1+y_2}(\phi_x)}{\sigma(y_1 + y_2)^2}\right]$$

- $L_t^y(\phi_x)_{t \geq 0}$ is the local time of $\phi_x(t)_{t \geq 0}$.
- Occupation times formula

$$\int_0^t \psi(s, \phi_x(s)) ds = \int_0^t \int_x \psi(s, y) dL_s^y(\phi_x)$$