Nonparametric estimation of the division rate of a Piecewise Deterministic Markov Process: models on trees

Adélaïde Olivier (CREST & Paris-Dauphine) supervised by Marc Hoffmann (Paris-Dauphine)

Colloque des Jeunes Probabilistes et Statisticiens, Forges-les-Eaux

2014, April 7th
Outline

1. **Age-structured model**
   - Model
   - Estimation of the division rate
   - Rate of convergence

2. **Size-structured model (joint work with Valère Bitseki)**
Outline

1. **Age-structured model**
   - Model
   - Estimation of the division rate
   - Rate of convergence

2. **Size-structured model (joint work with Valère Bitseki)**
Description of the system

- We consider a system of particles. We focus on the age of the particles.
- Two phenomenons:
  - aging (deterministic),
  - and splitting (random).
- A particle of age $a$ splits at a rate $B(a)$ to give birth to two new particles of age 0.
- Applications: cellular division, fragmentation, other models on trees.
A marked binary tree

- Genealogical tree $\mathcal{U} := \bigcup_{k=0}^{\infty} \{0, 1\}^k$.
- For $u \in \mathcal{U}$, $\zeta_u =$ life length.
- Particles split with a rate $B(a)$,
  \[ \mathbb{P}(\zeta_u \in [a, a + da]|\zeta_u \geq a) = B(a)da. \]
- $B$ is a hazard rate,
  \[ B(a) = \frac{f_B(a)}{1 - F_B(a)}, \quad a \geq 0, \]
  with $f_B$ a density, $F_B$ a c.d.f.

→ Aim: estimate the division rate $B$.
→ This can be very easy or very difficult depending on the observation scheme we choose.
Easy case

- Observation of \( n \) particles \textit{up to a fixed generation},

\[ (\zeta_u, u \in U_n) \]

where \( U_n := \) particles up to the generation \( \lfloor \log_2(n) \rfloor \).

\( \rightarrow \) \text{i.i.d.}

- Other possibility: observation \textit{up to a fixed time} \( T \).
Our observation scheme

Observation between time 0 and time $T$,

Continuous representation
Generational representation

(Blue: non-censored data, red: censored data.)

→ Intricate dependence and sampling bias.
Our observation scheme, cont.

- Observation between time 0 and time $T$.
- More precisely, we observe
  
  - $(ζ_u, u ∈ U_T)$ where $U_T :=$ particles that divide before $T$,
  - $(ζ^*_u, u ∈ U^*_T)$ where $U^*_T :=$ living particles at time $T$.

→ Simple model ("renewal on a tree").
→ Complexity comes from the observation scheme.
For nonnegative measurable bounded test functions $\phi$, when $T \to \infty$,

$$\frac{1}{N_T} \sum_{u \in U_T} \phi(\zeta_u) \xrightarrow{\Pr} \int_0^\infty \phi(a) f_{\text{biased}}(a) da.$$ 

(Bansaye et al., 2011 and Cloez, 2011)

In our setting,

$$f_{\text{biased}}(a) = 2e^{-\lambda_B a} f_B(a).$$

For a binary tree, when $T$ is large,

$$\mathbb{E}[N_T] \sim c_B e^{\lambda_B T}.$$ 

(Harris, 1963)

The magnitude of $\mathbb{E}[N_T]$ depends on $B$. 
What is $\lambda_B$?

- **Macroscopic description** of the system: $n(t, a) = $ density of particles of age $a$ at time $t$,

\[
\begin{align*}
\partial_t n(t, a) + \partial_a (n(t, a)) + B(a)n(t, a) = 0, \\
n(t, a = 0) &= 2 \int_0^\infty B(a)n(t, a)da, \\
n(t = 0, a) &= n^{(0)}(a),
\end{align*}
\]

- $\lambda_B$ is the first eigenvalue in this problem.
A weighted estimator of the division rate

- Estimator of $B$ at point $a$

\[ \hat{B}_T(a) := \frac{\hat{f}_B(a)}{(1 - \hat{F}_B(a)) \lor \varpi} \]

where

\[ \hat{f}_B(a) := \frac{1}{N_T} \sum_{u \in U_T} K_h(\zeta_u - a)(e^{\hat{\lambda}_B a}/2), \]

\[ \hat{F}_B(a) := \frac{1}{N_T} \sum_{u \in U_T} (e^{\hat{\lambda}_B \zeta_u}/2)1\{\zeta_u \leq a\} \]

with $\varpi > 0$ a threshold, $h > 0$ a bandwidth, $K : \mathbb{R} \to [0, \infty)$ a kernel function and $K_h(x) := h^{-1}K(h^{-1}x)$ for $x \in \mathbb{R}$.

- $\hat{\lambda}_B$ is an estimator of $\lambda_B$ (built with the $\zeta_u$ and $\zeta_u^*$).
Rate for the estimator $\hat{B}_T$?

- Rate of convergence in the law of large numbers?
- Main tools:
  - Many-to-one formulas,
  - Control of the ergodicity rate of nonreversible Markov processes.

- Target speed for the estimator: $(e^{-\lambda_B T/2})^{2\beta/(2\beta+1)}$ because
  \[ \mathbb{E}[N_T] \sim c_B e^{\lambda_B T} \]
  with $\beta = \text{Hölder-regularity}$.

- The magnitude of the speed depends on $B$. 
Rate of convergence of $\hat{B}_T$

**Theorem**

Let $\mathcal{D}$ be compact set. Assume that $B \in \mathcal{B} \cap \mathcal{H}^\beta(\mathcal{D})$. Specify

$$h := c_0 \left( \exp(-\hat{\lambda}_B T) \right)^{1/(1+2\beta)},$$

then, if $\beta$ is large enough,

$$\left( \exp(\lambda_B T/2) \right)^{2\beta/(2\beta+1)} \| \hat{B}_T - B \|_{L^2(\mathcal{D})}$$

is tight, uniformly over the class of $B$.

→ **Nonadaptive** choice of the bandwidth $h$. 
Partial conclusion

- **Age-structured model**
  - Genealogical obs. scheme: i.i.d situation.
  - Continuous obs. scheme: nonadaptive estimator converging at our target rate. Future:
    - Lower bound, adaptivity (concentration inequality).
    - Main difficulties: intricate dependence, random number of observations ($N_T / \mathbb{E}[N_T] \xrightarrow{L^2} W$ with $\text{Var}(W) > 0$).

- **Size-structured model**
  - Genealogical obs. scheme: ?
Outline

1. **Age-structured model**
   - Model
   - Estimation of the division rate
   - Rate of convergence

2. **Size-structured model (joint work with Valère Bitseki)**
Description of the system

- We consider a system of particles. We focus on the size of the particles.

- Two phenomenons:
  - exponential growth at a rate $\tau$ (deterministic),
  - and splitting (random).

- A particle of size $y$ splits at a rate $B(y)$ to give birth to two particles of size $y/2$.

  $\rightarrow$ The size at birth of a particle depends on the size of its ancestor.
Structure of dependence

- For $u \in U$,
  - $\zeta_u =$ life length,
  - $\xi_u =$ size at birth.

- Particles split at a division rate $B$,

$$
\mathbb{P}(\zeta_u \in [t, t + dt) \mid \zeta_u \geq t, \xi_u) = B(\xi_u e^{\tau t})dt.
$$

- Markovian dependence between sizes at birth.

- $(\xi_u, u \in U)$ is a bifurcating Markov chain. BMC introduced in Guyon, 2007.

- Estimation for $B$ by Doumic, Hoffmann, Krell and Robert, 2012, observing the system up to a fixed generation, but nonadaptive.
Concentration inequality

Let \((\xi_u, u \in \mathcal{U})\) be a general bifurcating Markov chain. Denote by \(\nu\) its invariant measure.

\(\mathcal{U}_n := \text{subtree up to the generation} \left\lfloor \log_2(n) \right\rfloor\).

We proved

\[
P_\nu \left( \frac{1}{n} \sum_{u \in \mathcal{U}_n} \phi(\xi_u) - \langle \phi, \nu \rangle > C_\kappa \left( \frac{\ln(n)}{n} \right)^{1/2} \right) \leq n^{-\kappa/2},
\]

for all \(\kappa \geq 1\), for test functions such that

\[
\|\phi\|_\infty \lesssim \sqrt{n \ln(n)}.
\]
Conclusion

- Age-structured model
  - Genealogical obs. scheme: i.i.d situation.
  - Continuous obs. scheme: nonadaptive estimator converging at our target rate. Future:
    - Lower bound, adaptivity (concentration inequality).
    - Main difficulties: intricate dependence, random number of observations ($N_T/\mathbb{E}[N_T] \xrightarrow{L^2} W$ with $\text{Var}(W) > 0$).

- Size-structured model
  - Genealogical obs. scheme: adaptive optimal estimator.
  - (Continuous obs. scheme)
Merci de votre attention


