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 $u \in \mathcal{U}$, ζ_u = life length.
- Particles split with a rate $B(a)$,

$$\mathbb{P}\big(\zeta_u \in [a, a+da) | \zeta_u \geq a\big) = B(a)da.$$

- 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.

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

Easy case

- Observation of n particles up to a fixed generation,

$$\left(\zeta_u, u \in \mathcal{U}_n\right)$$

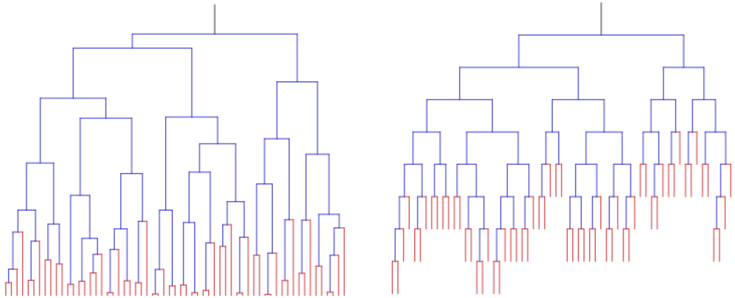
where $\mathcal{U}_n :=$ particles up to the generation $\lfloor \log_2(n) \rfloor$.

$\rightarrow$ i.i.d.

- Other possibility: observation 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.)

$\rightarrow$ Intricate dependence and sampling bias.

Our observation scheme, cont.

- Observation between time 0 and time T.
- More precisely, we observe
  - $\left(\zeta_u, u \in \mathcal{U}_T\right)$ where $\mathcal{U}_T :=$ particles that divide before T,
  - $\left(\zeta_u^*, u \in \mathcal{U}_T^*\right)$ where $\mathcal{U}_T^* :=$ living particles at time T.

$\rightarrow$ Simple model ("renewal on a tree").
$\rightarrow$ Complexity comes from the observation scheme.

Behaviour of the empirical mean

- For nonnegative measurable bounded test functions ϕ, when $T \to \infty$,

$$\frac{1}{N_T} \sum_{u \in \mathcal{U}_T} \phi(\zeta_u) \xrightarrow{\mathbb{P}} \int_0^\infty \phi(a) f_{\text{biased}}(a) da.$$

(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 λ_B ?

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

$$\begin{cases} \partial_t n(t,a) + \underbrace{\partial_a\big(n(t,a)\big)}_{\text{Aging}} + \underbrace{B(a)n(t,a)}_{\text{Division}} = 0, \\ \\ n(t,a=0) = 2\int_0^\infty B(a)n(t,a)da, \\ \\ n(t=0,a) = n^{(0)}(a), \end{cases}$$

- λ_B is the first eigenvalue in this problem.

A weighted estimator of the division rate

- Estimator of B at point a

$$\widehat{B}_T(a) := \frac{\widehat{f_B}(a)}{\left(1 - \widehat{F_B}(a)\right) \vee \varpi}$$

where

$$\widehat{f_B}(a) := \frac{1}{N_T} \sum_{u \in \mathcal{U}_T} K_h\left(\zeta_u - a\right)\left(e^{\widehat{\lambda_B} a}/2\right),$$

$$\widehat{F_B}(a) := \frac{1}{N_T} \sum_{u \in \mathcal{U}_T} \left(e^{\widehat{\lambda_B} \zeta_u}/2\right) \mathbb{1}_{\{\zeta_u \leq a\}}$$

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

- $\widehat{\lambda_B}$ is an estimator of λ_B (built with the ζ_u and ζ_u^*).

Rate for the estimator $\widehat{B}_T$?

- 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: $\left(e^{-\lambda_B T/2}\right)^{2\beta/(2\beta+1)}$ because

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

  with $\beta =$ Hölder-regularity.
- The magnitude of the speed depends on B.

Rate of convergence of $\widehat{B}_T$

Theorem

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

$$h := c_0 \big(\exp(-\widehat{\lambda_B} T) \big)^{1/(1+2\beta)},$$

then, if β is large enough,

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

is tight, uniformly over the class of B.

$\rightarrow$ 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 ($N_T/\mathbb{E}[N_T] \xrightarrow{\mathbb{L}^2} W$ with $\mathrm{Var}(W) > 0$).
- Size-structured model
  - Genealogical obs. scheme: ?

Outline

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 τ *(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 $u \in \mathcal{U}$,
  - ζ_u = life length,
  - ξ_u = size at birth.
- Particles split at a division rate B,

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

- Markovian dependence between sizes at birth.
- $\big(\xi_u, u \in \mathcal{U}\big)$ 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 ν its invariant measure.
- $\mathcal{U}_n :=$ subtree up to the generation $\lfloor \log_2(n) \rfloor$.
- We proved

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

for all $\kappa \geq 1$, for test functions such that

$$||\phi||_\infty \lesssim \sqrt{\frac{n}{\ln(n)}}.$$

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 ($N_T/\mathbb{E}[N_T] \xrightarrow{\mathbb{L}^2} W$ with $\mathrm{Var}(W) > 0$).
- Size-structured model
  - Genealogical obs. scheme: adaptive optimal estimator.
  - (Continuous obs. scheme)

Merci de votre attention

Bansaye, Delmas, Marsalle and Tran. Limit theorems for Markov processes indexed by continuous time Galton-Watson trees. *The Annals of Applied Probability*, 2011.

Cloez. Limit theorems for some branching measure-valued processes. arXiv:1106.0660, 2011.

Doumic, Hoffmann, Krell and Robert. *Statistical estimation of a growth-fragmentation model observed on a genealogical tree. Bernoulli*, to appear.

Down, Meyn and Tweedie. Exponential and uniform ergodicity of Markov processes. *The Annals of Probability*, 1995.

Guyon. Limit theorems for bifurcating Markov chains. Application to the detection of cellular aging. *The Annals of Applied Probability*, 2007.

Hairer, Mattingly. Yet another look at Harris' ergodic theorem for Markov chains. *Seminar on Stochastic Analysis, Random Fields and Applications VI Progress in Probability*, 2011

Harris. *The theory of branching processes.* Springer, 1963.

Perthame. *Transport equations in biology.* Birckhauser Frontiers in mathematics edition, 2007.