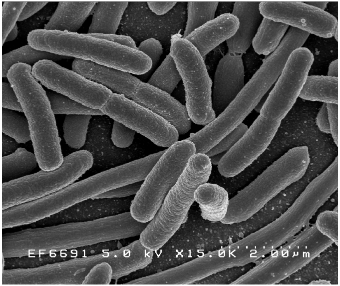
Modelling and statistics of branching processes

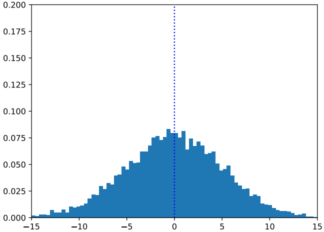
Aline Marguet, *Inria Grenoble – Rhône-Alpes*

Modelling the dynamics of a cellular population

Recent technological development: microfluidics, videomicroscopy, etc.

➥ Single-cell measurements: heterogeneity of cellular populations
➥ Individual-based models to capture this heterogeneity





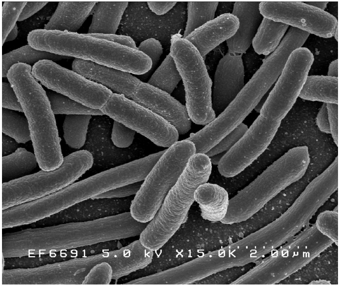
Development of structured individual-based models

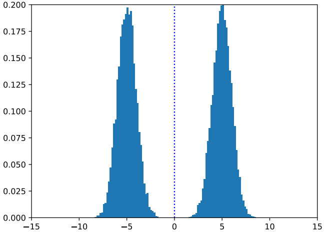
- Study of complex cellular mechanisms (division, ageing, inheritance)
- Inference of division rate, gene expression parameters

Modelling the dynamics of a cellular population

Recent technological development: microfluidics, videomicroscopy, etc.

➥ Single-cell measurements: heterogeneity of cellular populations
➥ Individual-based models to capture this heterogeneity



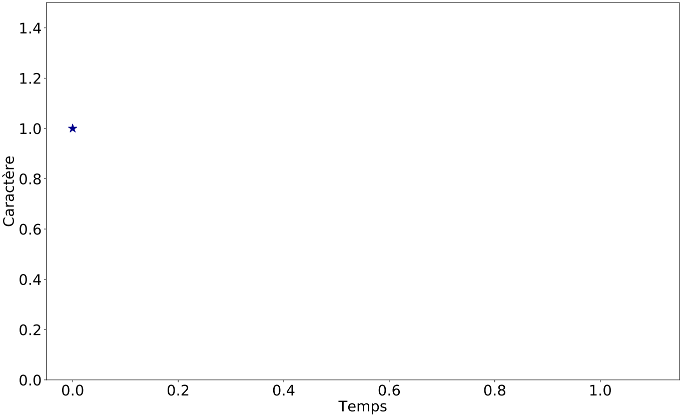


Development of structured individual-based models

- ♦ Study of complex cellular mechanisms (division, ageing, inheritance)
- ♦ Inference of division rate, gene expression parameters

Structured branching processes

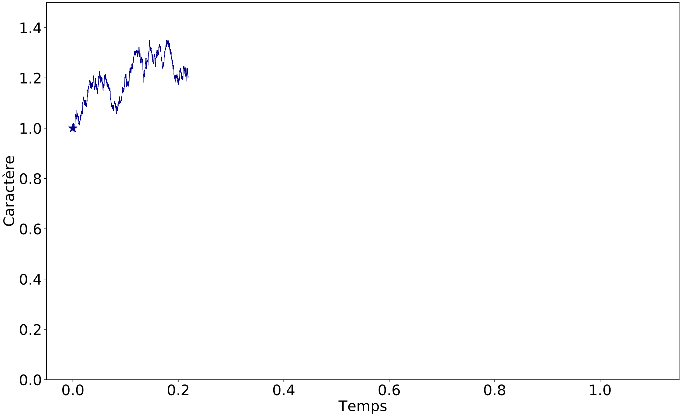
Model description



1. The trait $(X_t^u)_{t\geq 0}$ of each individual u follows a Markov process.

$$\text{Ex: } dX_t = r(X_t)dt + \sigma(X_t)dW_t, \quad X_0 = x.$$

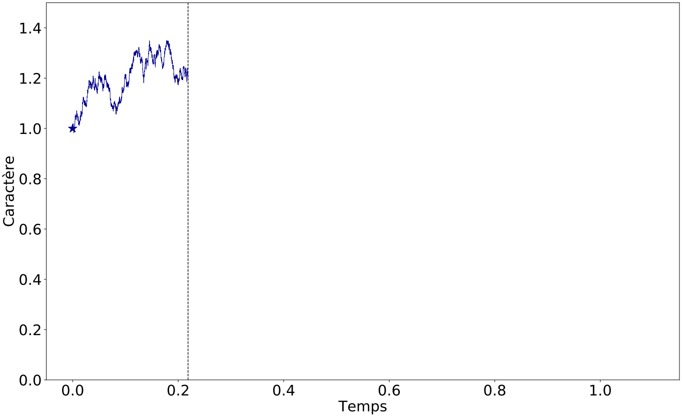
Model description



1. The trait $(X_t^u)_{t\geq 0}$ of each individual u follows a Markov process.

$$\text{Ex: } dX_t = r(X_t)dt + \sigma(X_t)dW_t, \quad X_0 = x.$$

Model description

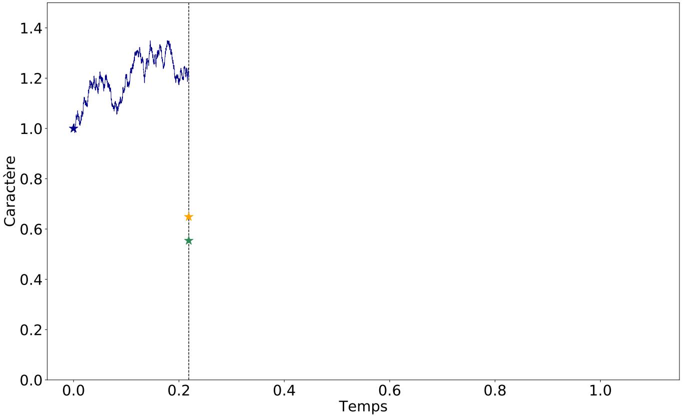


2. A cell u divides at time t at rate $B(X_t^u)$, i.e.

$$\mathbb{P}\left(\beta(u) > t | \alpha(u),\ (X_s^u, \alpha(u) \leq s \leq t)\right) = \exp\left(-\int_{\alpha(u)}^{t} B(X_s^u) ds\right)$$

where $\beta(u)$ denotes the lifetime of u and $\alpha(u)$ its time of birth.

Model description

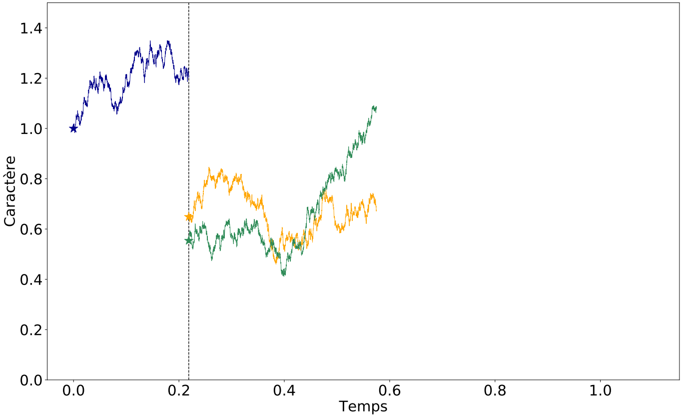


3.- At division, a cell with trait x is replaced by k daughter cells with probability $p_k(x)$.

- The traits at birth are given for the ith descendant by $F_i(x, \theta)$.

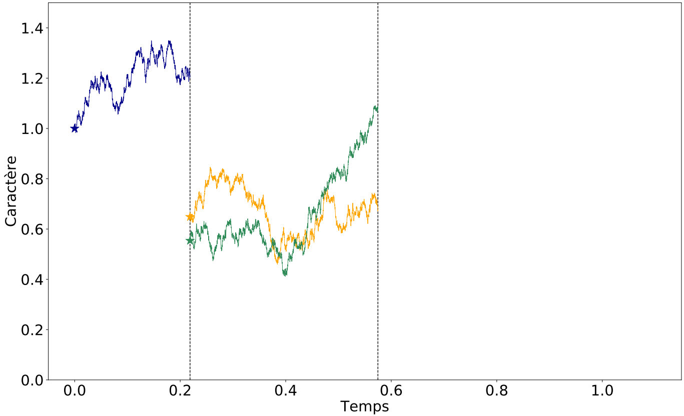
Ex: $F_1(x, \theta) = \theta x, F_2(x, \theta) = (1 - \theta)x,$ where $\theta \sim \kappa(d\theta)$.

Model description



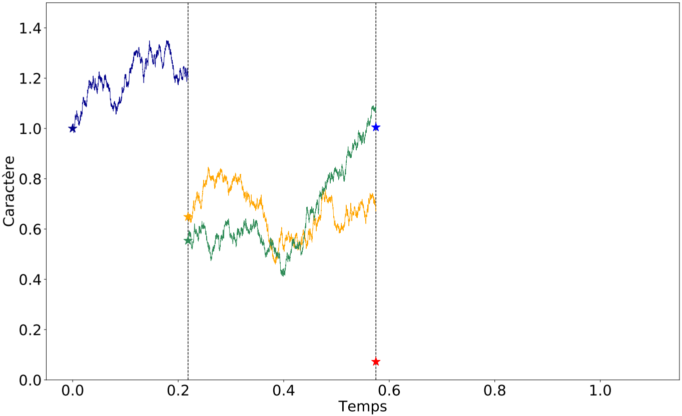
4. Conditionally to the trait of their ancestor, the daughter cells evolves independently following the same dynamics.

Model description



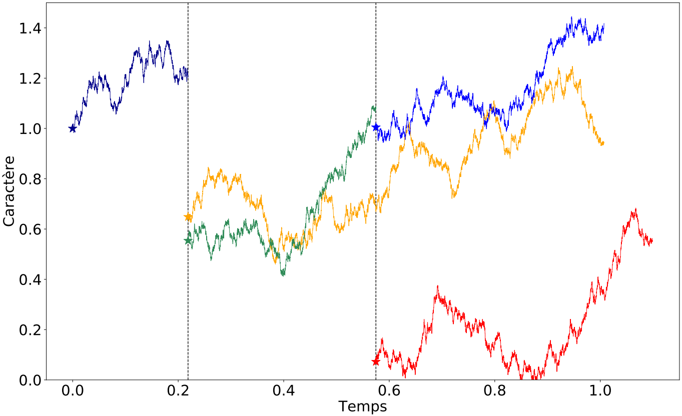
4. Conditionally to the trait of their ancestor, the daughter cells evolves independently following the same dynamics.

Model description



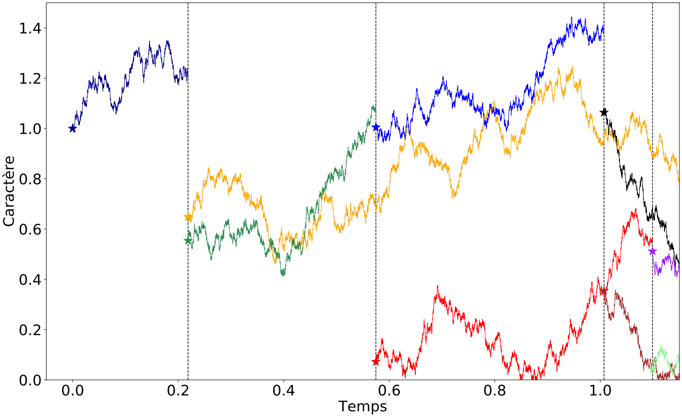
4. Conditionally to the trait of their ancestor, the daughter cells evolves independently following the same dynamics.

Model description



4. Conditionally to the trait of their ancestor, the daughter cells evolves independently following the same dynamics.

Model description



4. Conditionally to the trait of their ancestor, the daughter cells evolves independently following the same dynamics.

Examples

- Study of ageing: age-structured populations
- Study of the division mechanism: asymmetry, size-dependency, heritability
- Switching dynamics: different division rate for different subpopulations (dormant cells, etc.)

Example: study of the proliferation of an infection in a cellular population

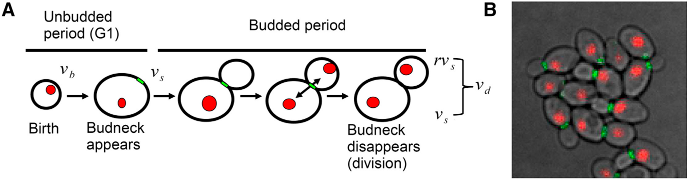


Figure: Image : Soifer, Robert & Amir, 2016.

- Asymmetric division : asymmetric sharing of the parasites in the daughter cells
- Strategy to eliminate the infection?

A. Marguet, C. Smadi, Long time behaviour of a general class of branching Markov processes, Preprint, 2019.

Inference in branching processes

Inference of the division rate

Inference of the division rate

- Piecewise Deterministic branching processes

- Stochastic flow

Inference of the division rate

- Piecewise Deterministic branching processes
  - Genealogical time, Size-structured population:

    M. Doumic, M. Hoffmann, N. Krell, L. Roberts, Statistical estimation of a growth-fragmentation model observed on a genealogical tree, 2015 (generalization to any flow along a branch, N. Krell, 2018).
  - Continuous time, Age-structured population:

    M. Hoffmann, A. Olivier Non-parametric estimation of the division rate of an age-dependent branching process, 2016.
- Stochastic flow

Inference of the division rate

- Piecewise Deterministic branching processes
  - Genealogical time, Size-structured population:

    M. Doumic, M. Hoffmann, N. Krell, L. Roberts, Statistical estimation of a growth-fragmentation model observed on a genealogical tree, 2015 (generalization to any flow along a branch, N. Krell, 2018).
  - Continuous time, Age-structured population:

    M. Hoffmann, A. Olivier Non-parametric estimation of the division rate of an age-dependent branching process, 2016.
- Stochastic flow
  - Genealogical time

    M. Hoffmann, A. Marguet. Statistical estimation in a randomly structured population, 2018.

Inference of the division rate

- Piecewise Deterministic branching processes
  - Genealogical time, Size-structured population:

    M. Doumic, M. Hoffmann, N. Krell, L. Roberts, Statistical estimation of a growth-fragmentation model observed on a genealogical tree, 2015 (generalization to any flow along a branch, N. Krell, 2018).
  - Continuous time, Age-structured population:

    M. Hoffmann, A. Olivier Non-parametric estimation of the division rate of an age-dependent branching process, 2016.
- Stochastic flow
  - Genealogical time

    M. Hoffmann, A. Marguet. Statistical estimation in a randomly structured population, 2018.
  - Continuous time ?

Genealogical time

Bifurcating Markov chains

We consider Ulam-Harris-Neveu notations: for $n, m \geq 0$, let

$$\mathbb{T} = \bigcup_{m \in \mathbb{N}} \{0,1\}^m, \quad \mathbb{T}_n = \bigcup_{m=0}^{n} \{0,1\}^m, \quad \mathbb{T}_n^\star = \mathbb{T}_n \setminus \{\emptyset\}.$$

Let X_u be the trait at birth of cell $u \in \mathbb{T}$. Assume that we are given observations

$$\mathbb{X}^n = (X_u)_{u \in \mathbb{T}_n}.$$

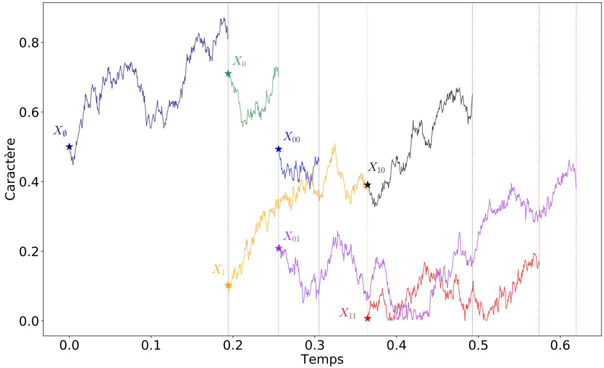
The process $(X_u, u \in \mathbb{T})$ is a bifurcating Markov chain with transition kernel $\mathcal{P}$ from $\mathbb{R}$ into $\mathbb{R} \times \mathbb{R}$ such that:

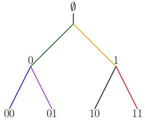
$$\mathbb{E}\Big[\prod_{u \in \mathbb{G}_m} \psi_u(X_u, X_{u0}, X_{u1}) \,\Big|\, \mathcal{F}_m \Big] = \prod_{u \in \mathbb{G}_m} \mathcal{P}\psi_u(X_u),$$

for all $m \geq 0$.

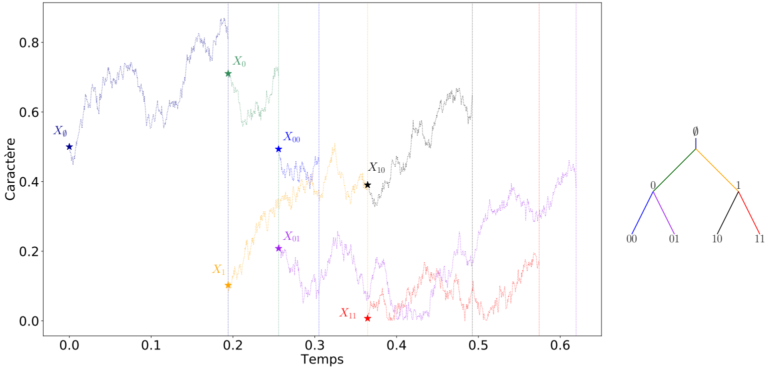
Guyon 2007.

Example of trajectory

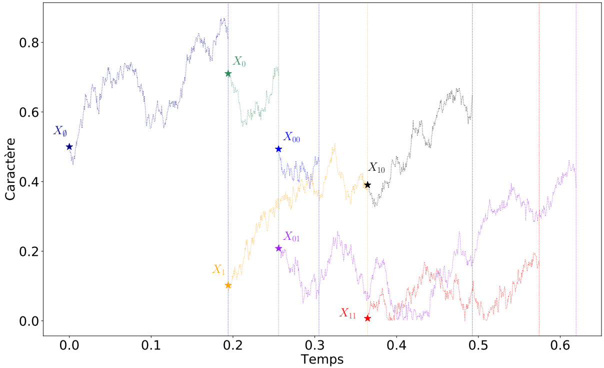


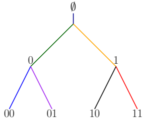


Example of trajectory



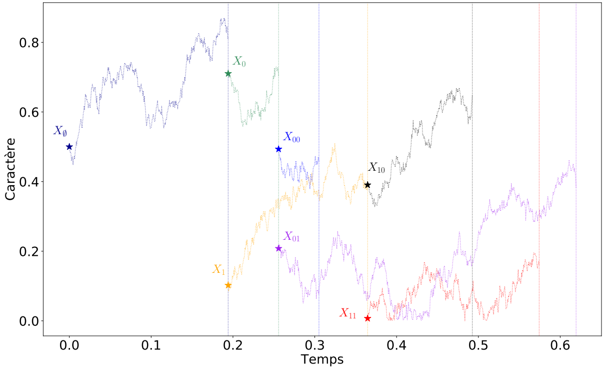
Example of trajectory

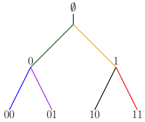




♦ Deterministic dynamic: as if we observed trajectories.

Example of trajectory





- Deterministic dynamic: as if we observed trajectories.
- Stochastic dynamic: unknown trajectories.

Kernel estimation

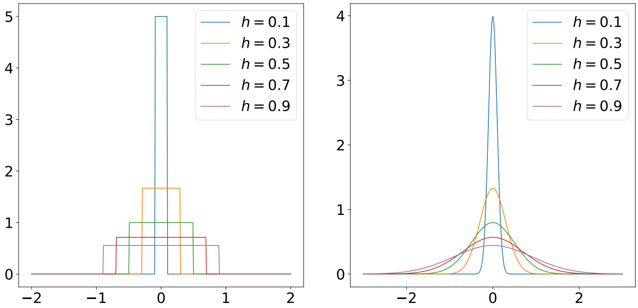
Kernel estimation

A function $G : \mathbb{R} \to \mathbb{R}$ is a kernel of order k if for all $i = 0, \dots, k$.

$$\int_{\mathbb{R}} x^i G(x) dx = \mathbf{1}_{\{i=0\}}$$

For $h > 0$, we define

$$G_h(y) = h^{-1} G(h^{-1} y), \quad \forall y \in \mathbb{R}.$$



Kernel estimator of the invariant measure

Consider the empirical measure

$$\mathcal{M}_n(\varphi) = \frac{1}{|\mathbb{T}_n^\star|} \sum_{u \in \mathbb{T}_n^\star} \varphi(X_u).$$

If $\mathcal{M}_n(\varphi)$ converges towards $\nu(\varphi)$, we have

$$\mathcal{M}_n(G_h(\cdot - x_0)) \xrightarrow[n \to \infty]{} \int_{\mathbb{R}} G_h(x - x_0)\nu(x)dx.$$

Kernel estimator of the invariant measure

Consider the empirical measure

$$\mathcal{M}_n(\varphi) = \frac{1}{|\mathbb{T}_n^\star|} \sum_{u \in \mathbb{T}_n^\star} \varphi(X_u).$$

If $\mathcal{M}_n(\varphi)$ converges towards $\nu(\varphi)$, we have

$$\mathcal{M}_n(G_h(\cdot - x_0)) \xrightarrow[n\to\infty]{} \int_{\mathbb{R}} G_h(x - x_0)\nu(x)dx.$$

An estimator of $\nu(x_0)$, for $x_0 \in \mathbb{R}$ is then given by

$$\widehat{\nu}_n(x_0) = \mathcal{M}_n(G_h(\cdot - x_0)).$$

Kernel estimator of the invariant measure

Consider the empirical measure

$$\mathcal{M}_n(\varphi) = \frac{1}{|\mathbb{T}_n^\star|} \sum_{u \in \mathbb{T}_n^\star} \varphi(X_u).$$

If $\mathcal{M}_n(\varphi)$ converges towards $\nu(\varphi)$, we have

$$\mathcal{M}_n(G_h(\cdot - x_0)) \xrightarrow[n\to\infty]{} \int_{\mathbb{R}} G_h(x - x_0)\nu(x)dx.$$

An estimator of $\nu(x_0)$, for $x_0 \in \mathbb{R}$ is then given by

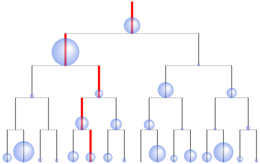
$$\widehat{\nu}_n(x_0) = \mathcal{M}_n(G_h(\cdot - x_0)).$$

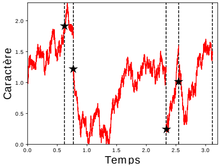
➥How *good* is this estimator?

Convergence of the empirical measure

Process of the tagged cell

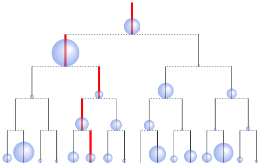
Let Y be the Markov chain corresponding to the trait of the tagged cell.

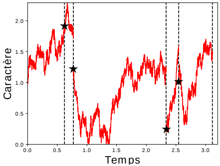




Process of the tagged cell

Let Y be the Markov chain corresponding to the trait of the tagged cell.





The transition operator of Y is given by $\mathcal{Q} = (\mathcal{P}_1 + \mathcal{P}_2)/2$, where

$$\mathcal{P}_1(x, dy) = \int_{\mathbb{R}} \mathcal{P}(x, dy dy_2) \quad \mathcal{P}_2(x, dy) = \int_{\mathbb{R}} \mathcal{P}(x, dy_1 dy).$$

Convergence of the tagged chain

Theorem

Under assumptions, $\mathcal{Q}$ admits an invariant measure ν. Moreover, there exist $C > 0$ and $\rho \in (0,1)$ such that for all $m \in \mathbb{N}$:

$$\left|\mathcal{Q}^m \varphi - \nu(\varphi)\right|_\infty \leq C\rho^m \left|\varphi - \nu(\varphi)\right|_\infty$$

Convergence of the tagged chain

Let

$$|\varphi|_V = \sup_{x\in\mathbb{R}} \frac{|\varphi(x)|}{1+V(x)}.$$

Theorem

Under assumptions, $\mathcal{Q}$ admits an invariant measure ν. Moreover, there exist $C>0$ and $\rho\in(0,1)$ such that for all $m\in\mathbb{N}$:

$$\left|\mathcal{Q}^m\varphi-\nu(\varphi)\right|_V \le C\rho^m\left|\varphi-\nu(\varphi)\right|_V$$

for all measurable function $\varphi:\mathbb{R}\to\mathbb{R}$ such that $|\varphi|_V<\infty$, where $V(x)=x^2$.

Convergence of estimator

Let $\beta > 0$ and let G be a kernel of order $k > \beta$. Let us set

$$h = |\mathbb{T}_n|^{\frac{-1}{2\beta+1}}$$

Theorem

Under assumptions ensuring the ergodicity of the tagged chain Y, we have

$$\Big(\mathbb{E}\big[\big(\widehat{\nu}_n(x_0) - \nu(x_0)\big)^2\big]\Big)^{1/2} \lesssim |\mathbb{T}_n|^{-\beta/(2\beta+1)},$$

uniformly in $\mathcal{Q}$ for $\mathcal{Q}$ in a given Hölder regularity class depending on β.

Estimation of the division rate

Size-structured population

Assume that the cells grow exponentially at rate a, and divide in two equal parts. Then, the transition kernel of the chain of successive size at birth along a branch is given by

$$\mathcal{Q}_B(x, dy) = \mathbf{1}_{2y \geq x} \frac{B(2y)}{ay} e^{-\int_{\frac{x}{2}}^{y} \frac{B(z)}{az} dz} dy.$$

Moreover, if ν_B is the invariant measure of the chain, we have

$$\nu_B(x) = \frac{B(2x)}{ax} \mathbb{E}_{\nu_B} \left[\mathbf{1}_{\{X_u \geq x,\ X_{u^-} \leq 2x\}} \right],$$

where X_u denotes the size at birth of cell u and u^- is the ancestor of u. Then, an estimator of B is given by

$$\widehat{B}(x) = a \frac{x}{2} \frac{n^{-1} \sum_{u \in \mathcal{U}_n} G_h(X_u - x/2)}{n^{-1} \sum_{u \in \mathcal{U}_n} \mathbf{1}_{\{X_u \geq x,\ X_{u^-} \leq 2x\}}}$$

M. Doumic, M. Hoffmann, N. Krell, L. Roberts, Statistical estimation of a growth-fragmentation model observed on a genealogical tree, 2015.

Randomly structured population

Assume that the dynamic of the trait follows

$$d\phi_x(t) = r(\phi_x(t))dt + \sigma(\phi_x(t))dW_t, \quad \phi_x(0) = x.$$

The dependency in B of the transition function is complex but explicit:

$$q(x,y) = \int_0^1 \frac{\kappa(z)}{z} B(y/z)\sigma(y/z)^{-2}\mathbb{E}\Big[\int_0^\infty e^{-\int_0^t B(\phi_x(s))ds} dL_t^{y/z}(\phi_x)\Big]dz,$$

where $L_t^y(\phi_x)$ is the local time at time t at position y of ϕ_x and

$$Q(x,dy) = q(x,y)dy$$

.

M. Hoffmann, A. Marguet. Statistical estimation in a randomly structured population, 2018.

Randomly structured population

Assume that the dynamic of the trait follows

$$d\phi_x(t) = r(\phi_x(t))dt + \sigma(\phi_x(t))dW_t, \quad \phi_x(0) = x.$$

The dependency in B of the transition function is complex but explicit:

$$q(x,y) = \int_0^1 \frac{\kappa(z)}{z} B(y/z)\sigma(y/z)^{-2}\mathbb{E}\Big[\int_0^\infty e^{-\int_0^t B(\phi_x(s))ds} dL_t^{y/z}(\phi_x)\Big] dz,$$

where $L_t^y(\phi_x)$ is the local time at time t at position y of ϕ_x and

$$Q(x,dy) = q(x,y)dy$$

.

➥ Likelihood contrast in a parametric framework with (r,σ,κ) known.

M. Hoffmann, A. Marguet. Statistical estimation in a randomly structured population, 2018.

Parametric model

Assume that the division rate B belongs to a class

$$\mathcal{B} = \big\{ B : [0, L] \to \mathbb{R}, B(x) = B_0(\vartheta, x), x \in [0, L], \vartheta \in \Theta \big\},$$

where $x \mapsto B_0(x, \vartheta)$ is known up to a parameter $\vartheta \in \Theta$, and $\Theta \subset \mathbb{R}^d$ is a compact set.

Aim

Estimate ϑ from $(X_u, u \in \mathbb{T}_n)$.

Likelihood contrast

A likelihood contrast is given by:

$$\mathcal{L}_n\big(\vartheta, (X_u, u \in \mathbb{T}_n)\big) = \prod_{u \in \mathbb{T}_n^\star} q_\vartheta(X_{u^-}, X_u),$$

where X_{u^-} is the trait of the ancestor of u. We consider the estimator of ϑ given by:

$$\widehat{\vartheta}_n \in \operatorname*{argmax}_{\vartheta \in \Theta} \left\{ \frac{1}{\mathbb{T}_n^\star} \sum_{u \in \mathbb{T}_n^\star} \log\left(q_\vartheta\left(X_{u^-}, X_u\right)\right) \right\}.$$

M. Hoffmann, A. Marguet. Statistical estimation in a randomly structured population, 2018.

Likelihood contrast

A likelihood contrast is given by:

$$\mathcal{L}_n\big(\vartheta, (X_u, u \in \mathbb{T}_n)\big) = \prod_{u \in \mathbb{T}_n^\star} q_\vartheta(X_{u^-}, X_u),$$

where X_{u^-} is the trait of the ancestor of u. We consider the estimator of ϑ given by:

$$\widehat{\vartheta}_n \in \operatorname*{argmax}_{\vartheta \in \Theta} \left\{ \frac{1}{\mathbb{T}_n^\star} \sum_{u \in \mathbb{T}_n^\star} \log\left(q_\vartheta\left(X_{u^-}, X_u\right)\right) \right\}.$$

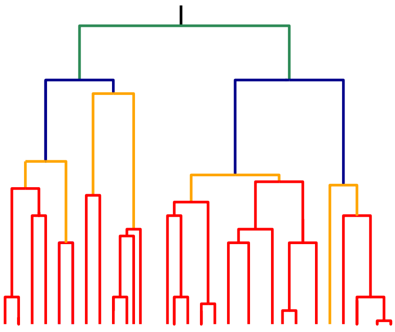
Under regularity assumptions on B, we prove that

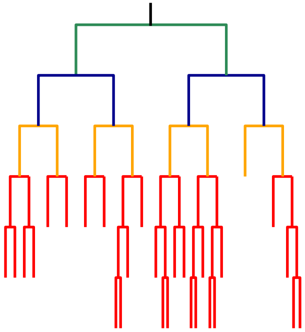
- $\widehat{\vartheta}_n$ converges in probability to ϑ when n tends to infinity.
- $\widehat{\vartheta}_n$ is asymptotically efficient.

M. Hoffmann, A. Marguet. Statistical estimation in a randomly structured population, 2018.

Continuous time

Data : continuous time vs genealogical time





- Random size of the population
- Sampling bias: more cells coming from fertile lineages
- Censored data: the cells at the bottom of the tree have not divided yet

Behavior of a typical individual

For all $t > 0$, $x_0 \in \mathcal{X}$, for all measurable positive function $F : \mathbb{D}([0,t], \mathcal{X}) \to \mathbb{R}$, we have:

$$\mathbb{E}_{\delta_{x_0}}\left[\sum_{u \in V_t} F\left(X_s^u, s \leq t\right)\right] = \mathbb{E}_{\delta_{x_0}}(N_t)\mathbb{E}_{x_0}\left[F\left(Y_s^{(t)}, s \leq t\right)\right],$$

with $\left(Y_s^{(t)}\right)_{s \leq t}$ a time-inhomogeneous Markov process.

- Tool: First moment semigroup transformation (Del Moral 2004, Bansaye 2013, Champagnat et Villemonais 2016, Cloez 2017)
- Interest : A unique process to summarize the dynamic of the whole population

A. Marguet, Uniform sampling in a structured branching population, 2019.

Estimation of the division rate: age-structured population

- Many-to-One formula

$$\mathbb{E}\left[\sum_{u\in V_t} g(X_t^u)\right] = \frac{e^{\lambda_B t}}{m}\mathbb{E}\left[g(\chi_t)B(\chi_t)^{-1}H_B(\chi_t)\right],$$

where χ is an auxiliary process.

- Convergence of the empirical measure: using the geometric ergodicity of χ.
- Representation of B: combining

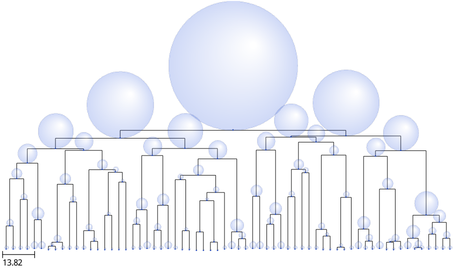
$$B(x) = \frac{B(x)e^{-\int_0^x B(z)dz}}{1-\int_0^x B(y)e^{-\int_0^y B(z)dz}dy}$$

and the invariant measure of χ.

M. Hoffmann, A. Olivier Non-parametric estimation of the division rate of an age-dependent branching process, 2016.

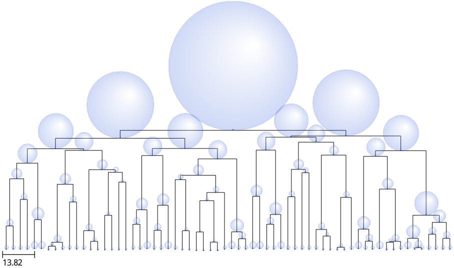
Perspectives

- estimation of B continuous time & stochastic flow
- study of asymmetry in division
- dependency of the division mechanism in some trait



Perspectives

- estimation of B continuous time & stochastic flow
- study of asymmetry in division
- dependency of the division mechanism in some trait



Thank you for your attention!

Asymptotic normality

Let us define $\Psi(\vartheta)$ the Fisher information matrix which coefficients are given for all $1 \leq i, j \leq d$ by:

$$\Psi(\vartheta)_{i,j} = \nu_\vartheta \mathcal{Q}(\vartheta) \left(\frac{\partial_{\vartheta_i} q_\vartheta \partial_{\vartheta_j} q_\vartheta}{q_\vartheta^2} \right).$$

Theorem

Under regularity assumptions and if $\Psi(\vartheta)$ is invertible, for all ϑ in the interior of Θ, we have:

$$|\mathbb{T}_n|^{1/2} \left(\widehat{\vartheta}_n - \vartheta \right) \to \mathcal{N}\left(0, \Psi(\vartheta)^{-1}\right),$$

in law when n tends to infinity, where $\mathcal{N}\left(0, \Psi(\vartheta)^{-1}\right)$ is the d-dimensional normal law with mean 0 and the inverse of the Fisher information matrix $\psi(\vartheta)$ as covariance matrix.