Estimating the Division Kernel of a Size-structured Population

Van Ha Hoang
(Start 2^{nd}-year PhD under supervision of Viet Chi Tran (Lille 1),
Vincent Rivoirard (Paris Dauphine)
and Thanh Mai Pham Ngoc (Paris Sud))

Laboratoire Paul Painlevé, Université Lille 1

Outline

Motivations

- We want to study the evolution of the quantity of toxicities in a cell population under binary division.
- We introduce an individual-based model that describes discrete population in continuous time. The dynamics of the population are specified at the level of the individual SDEs and branching events.
- In this model, a cell which contains a toxicity $x \in \mathbb{R}_+$ divides in continuous time and the toxicity grows inside the cell. When the cell divides, the toxicity is shared randomly in the two daughter cells.

The continuous time model

- Along branches: The toxicity $(X_t, t \geq 0)$ satisfies

$$dX_t = \kappa g(X_t)dt \tag{1}$$

with $X_0 = 0$.

- Genealogical tree:
  - The cell divides at a rate $B(x)$ and the toxicity increases with rate $\kappa g(x)$ where g is a continuous positive function and $\kappa > 0$.
  - When a cell divides, a random fraction Γ of the toxicity goes in the first daughter cell and a fraction $(1 - \Gamma)$ in the second one. We assume that Γ is a random variable in $[0, 1]$ with distribution $H(d\gamma)$.

The continuous time model

Assumption

- The division rate $B(x)$ is continuous and bounded by a positive constant $\bar{B}$.
- There exists a constant $C > 0$ and $x_0 > 0$ such that $\forall x > x_0$, $|g(x)| \leq Cx$.

- $H(d\gamma)$ is a symmetric distribution on $(0, 1)$.

Examples:

- $H(d\gamma) = \delta_{\frac{1}{2}}(d\gamma)$,
- $H(d\gamma) = \mathcal{U}[0, 1]$,

The continuous time model

- We aim to establish a partial differential equation that expresses the evolution when the population is large and estimate $H(d\gamma)$.
- For the similar models, we refer to Bansaye and Tran [2] and a lot of literature for discrete time (Guyon, Delmas and Marsalle, etc)
- For the statistic: we refer to Doumic et al [3].

Empirical measure

- Let V_t be the set of living cells at time t, we define

$$Z_t^n(dx) = \frac{1}{n} \sum_{i \in V_t} \delta_{X_t^i}(dx), \qquad n \in \mathbb{N}^* \tag{2}$$

  is the random point measure on $\mathcal{M}_F(\mathbb{R}_+)$, the space of finite measures, is embedded with the topology of weak convergence.
- For a measure $\mu \in \mathcal{M}_F(\mathbb{R}_+)$ and a positive function f, we use the notation $\langle \mu, f \rangle = \int_{\mathbb{R}+} f d\mu$.
- The parameter n is related to the large population limit which corresponds to $n \to +\infty$.

Moment and martingale properties

- Z_t^n is described by a SDE driven by a Poisson Point measure and under the moment conditions, the SDE has a solution $(Z_t^n)_{t\geq 0} \in \mathbb{D}([0,T], \mathcal{M}_F(\mathbb{R}_+))$.
- If Z_0^n is such that $\mathbb{E}(\langle Z_0^n, 1\rangle^2) < +\infty$, then for all test function $f(x,t) \in \mathcal{C}_b^1(\mathbb{R}_+ \times \mathbb{R}_+, \mathbb{R})$

$$\langle Z_t^n, f_t\rangle = \langle Z_0^n, f_0\rangle + V_t^{n,f} + M_t^{n,f}$$

where

$$\begin{aligned} V_t^{n,f} &= \int_0^t \int_{\mathbb{R}_+} \big(\partial_s f_s(x) + \kappa g(x)\partial_x f_s(x)\big) Z_s^n(dx)ds \\ &+ \int_0^t \int_{\mathbb{R}_+} \int_0^1 [f_s(\gamma x) + f_s((1-\gamma)x) - f_s(x)]\, B(x)H(d\gamma)Z_s^n(dx)ds \end{aligned} \tag{3}$$

where M_t^n is a continous square integrable martingale with quadratic variation

$$\langle M^{n,f}\rangle_t = \frac{1}{n}\int_0^t \int \int_0^1 \Big[f_s(\gamma x) + f_s((1-\gamma)x) - f_s(x)\Big]^2 B(x)H(d\gamma)Z_s^n(dx)ds \tag{4}$$

Convergence in the large population limit

Theorem

Consider the sequence $(Z^n)_{n\in\mathbb{N}^*}$, if Z_0^n converges in distribution to $\mu_0 \in \mathcal{M}_F(\mathbb{R}_+)$ as $n \to +\infty$ then $(Z^n)_{n\in\mathbb{N}^*}$ converges in distribution in $\mathbb{D}\left([0, T], \mathcal{M}_F(\mathbb{R}_+)\right)$ as $n \to +\infty$ to $\mu \in \mathcal{C}\left([0, T], \mathcal{M}_F(\mathbb{R}_+)\right)$, where μ is the unique solution of

$$\begin{aligned}\langle \mu_t, f_t\rangle =&\langle \mu_0, f_0\rangle + \int_0^t \int_{\mathbb{R}_+} \left(\partial_s f_s(x) + \kappa g(x)\partial_x f_s(x)\right)\mu_s(dx)ds \\ &+ \int_0^t \int_{\mathbb{R}_+} \int_0^1 \left[f_s(\gamma x) + f_s((1-\gamma)x) - f_s(x)\right] B(x)H(d\gamma)\mu_s(dx)ds \quad (5)\end{aligned}$$

with $f_t(x) \in \mathcal{C}_b^{1,1}(\mathbb{R}_+ \times \mathbb{R}_+, \mathbb{R})$ is a test function.

Sketch of the proof

Tightness

- We will prove that $(Z^n)_{n\in\mathbb{N}^*}$ is tight in $\mathbb{D}([0,T],\mathcal{M}_F(\mathbb{R}_+))$, where $\mathcal{M}_F(\mathbb{R}_+)$ is embedded the topology of weak convergence.
- Aldous's criterion [1] has to be checked for $V_t^{n,f}$ and $\langle M^{n,f}\rangle_t$ (cf. Arzelá-Ascoli theorem for the similar proof) and additional work from vague topology $(\mathcal{M}_F, v)$ to weak topology $(\mathcal{M}_F, w)$.

Uniqueness: we use the martingale and moment arguments.

Sketch of the proof

Aldous's criterion:

1. $\forall t \in \mathcal{T}$ dense in $\mathbb{R}_+$, $\left(\langle M^{n,f}\rangle_t\right)_{n\in\mathbb{N}^*}$ and $\left(V_t^{n,f}\right)_{n\in\mathbb{N}^*}$ are tight in $\mathbb{R}_+$.
2. $\forall T \geq 0$, $\forall \epsilon > 0$, $\forall \eta > 0$, $\exists \delta > 0$, $n_0 \in \mathbb{N}$ such that

$$\sup_{n\geq n_0} \mathbb{P}\left(\left|\langle M^{n,f}\rangle_{T_n} - \langle M^{n,f}\rangle_{S_n}\right| \geq \eta\right) \leq \epsilon$$

and

$$\sup_{n\geq n_0} \mathbb{P}\left(\left|V_{T_n}^{n,f} - V_{S_n}^{n,f}\right| \geq \eta\right) \leq \epsilon$$

for every couples of stopping-times $(S_n, T_n)_{n\in\mathbb{N}^*}$ such that $S_n \leq T_n \leq T$ and $T_n \leq S_n + \delta$.

Sketch of the proof

- Let μ be a limiting value of (Z^n) and

$$\Psi_t(\mu) = \langle \mu_t, f\rangle - \langle \mu_0, f\rangle - \int_0^t \int_{\mathbb{R}_+} \left(\partial_s f_s(x) + \kappa g(x)\partial_x f_s(x)\right)\mu_s(dx)ds$$
$$+ \int_0^t \int_{\mathbb{R}_+} \int_0^1 \left[f_s(\gamma x) + f_s((1-\gamma)x) - f_s(x)\right] B(x)H(d\gamma)\mu_s(dx)ds$$

- To prove that μ satisfies (5) almost surely, we will show that $\Psi_t(\mu) = 0$ for every t.

We know that, $\forall n \in \mathbb{N}^*$ the process $M_t^{f,\phi(n)} = \Psi_t\left(Z_t^{\phi(n)}\right)$ is a square integrable martingale. Thus, for $t \in \mathbb{R}_+$ we have

$$\mathbb{E}\left(\left|M_t^{f,\phi(n)}\right|\right)^2 \leq \mathbb{E}\Big(\langle M^{f,\phi(n)}\rangle_t\Big) \leq \frac{tC}{\phi(n)}\mathbb{E}\Big(\sup_{s\in[0,t]} \langle Z_s^{\phi(n)}, 1\rangle\Big)$$

Then, $\forall t \in \mathbb{R}_+, \lim_{n\to+\infty} \mathbb{E}\left(\left|M_t^{f,\phi(n)}\right|\right) = 0$

Sketch of the proof

To show that, $\forall t \in \mathbb{R}_+$, $\Psi_t(\mu) = 0$ almost surely, we need to prove that

$$\lim_{n\to+\infty} \mathbb{E}\left(\left|M_t^{f,\phi(n)}\right|\right) = \mathbb{E}\left((\Psi_t(\mu)|\right)$$

We know that $(Z^{\phi(n)})_{n\in\mathbb{N}^*}$ converges in distribution to μ. Since μ is continous almost surely, $f \in \mathcal{C}^1$ class with bounded derivatives and from moment assumption then Ψ_t is continous and

$$\lim_{n\to+\infty} \Psi_t(Z^{\phi(n)}) \xrightarrow{d} \Psi_t(\mu)$$

This ends the proof of the Theorem.

Size-structured population equation

Proposition:

We assume that $g(x) = 1$ for all $x \in \mathbb{R}_+$. This implies that $X_t = x_0 + \kappa(t - t_0)$ if toxicity is x_0 at times t_0. Then, we have the following results:

i. If $\mu_0(dx) = n_0(x)dx$ then $\forall t \in \mathbb{R}_+$, $\mu_t(dx) = n_t(x)dx$.

ii. $n(t,x) \in \mathcal{C}^{1,1}(\mathbb{R}_+, \mathbb{R}_+)$ and satisfies the PDE:

$$\frac{\partial n(t,x)}{\partial t} + \kappa \frac{\partial n(t,x)}{\partial x} + B(x)n(t,x) = 2\int_0^1 \frac{1}{\gamma} B\left(\frac{x}{\gamma}\right) n\left(t, \frac{x}{\gamma}\right) H(d\gamma) \tag{6}$$

We now focus to estimate the the division kernel in two cases: 1) case of complete data of divisions. 2) Else, use the stationary distribution approximation.

Case of complete data of divisions

- We observed the evolution of the cell population in a given time interval $[0, T]$.
- At the i^{th} division time T_i, denote j_i the individual who splits into two daughters $X_{t_i}^{j_i 1}$ and $X_{t_i}^{j_i 2}$. We define

$$\Gamma_i^1 = \frac{X_{t_i}^{j_i 1}}{X_{t_i -}^{j_i}} \quad \text{and} \quad \Gamma_i^2 = \frac{X_{t_i}^{j_i 2}}{X_{t_i -}^{j_i}}$$

  the random fractions that go into the daughter cells, with the convention $\frac{0}{0} = 0$.
- The couples $(\Gamma_i^1, \Gamma_i^2)_{i \in \mathbb{N}^*}$ are independent with distribution (Γ^1, Γ^2) where $\Gamma^1 \sim H(d\gamma)$ and $\Gamma^2 = 1 - \Gamma^1$.
- Assume that $H(d\gamma) = h(\gamma) d\gamma$ where $h(\gamma)$ is a density function. We now construct an estimator of h based on $(\Gamma_i^1, \Gamma_i^2)_{i \in \mathbb{N}^*}$.

Case of complete data of divisions

Definition

Let N_T be the random number of divisions in the time interval $[0, T]$. For $\gamma \in (0,1)$, define

$$\hat{h}_T(\gamma) = \frac{1}{N_T} \sum_{i=1}^{N_T} K_\ell(\gamma - \Gamma_i^1), \tag{7}$$

where K is a kernel function, $\ell > 0$ is the bandwidth to be chosen and $K_\ell(\cdot) = \frac{1}{\ell} K(\frac{\cdot}{\ell})$. We write $\hat{h}(\gamma)$ instead of $\hat{h}_T(\gamma)$ for convenience.

Example:

- $K(x) = \frac{1}{2}\mathbb{I}(|x| \leq 1)$ (the rectangular kernel),
- $K(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-x^2/2\right)$ (the Gaussian kernel).

Case of complete data of divisions

Proposition:

If $B(x) = R > 0$ and $h \in \mathcal{H}(\beta)$ ($h \in \mathcal{C}^{\lfloor\beta\rfloor}$ and $h^{(\beta)}$ is $\beta - \lfloor\beta\rfloor$ Hölder continuous). Then the kernel estimator $\hat{h}_T(\gamma)$ satisfies

$$\sup_{h\in\mathcal{H}(\beta)} \mathbb{E}\|h - \hat{h}_T\|_2^2 \leq Ce^{-\varrho T}. \tag{8}$$

where

$$\varrho = \frac{N_0}{N_0+1}\frac{2\beta R}{2\beta+1}. \tag{9}$$

Remark: compare with $n^{-\frac{2\beta}{2\beta+1}}$ for n observations.

Case of the stationary distribution approximation

- If we don't have complete data, $\hat{h}_T$ can not be computed anymore. Then, we consider the case when we have a stationary distribution of x at a given time.
- To estimate h, we assume that we have n data, each data being obtained from the measurement of an individual cell picked at random, after the system has evolved for a long time so that the approximation $n(t,x) \approx N(x)e^{\lambda t}$ is valid (The growth rate λ will be explained later).
- Each data is viewed as the outcome of random variable X_i, each X_i having density function $N(x)$.
- We observed a random sample $(X_1, X_2, \ldots, X_n)$.

Case of the stationary distribution approximation

- The PDE that describes the evolution of cell population

$$\partial_t n(t,x)+\alpha\partial_x n(t,x)+B(x)n(t,x) = 2\int_0^1 \frac{1}{\gamma}B\left(\frac{x}{\gamma}\right) n\left(t,\frac{1}{\gamma}\right) H(d\gamma), \tag{10}$$

- Assume that $H(d\gamma)$ has a density: $H(d\gamma) = h(\gamma)d\gamma$ and set $y = x/\gamma$ then equation (10) becomes:

$$\partial_t n(t,x) + \alpha\partial_x n(t,x) = 2\int_0^\infty B(y)n(t,y)h\left(\frac{x}{y}\right)\frac{dy}{y} - B(x)n(t,x), \tag{11}$$

where $h(x/y) = 0$ if $y < x$.

Case of the stationary distribution approximation

By general relative entropy principle (see [6]), it is proved that under suitable assumption on h and B, one has

$$\int_0^\infty |n(t,x)e^{-\lambda t} - \rho N(x)|\phi(x)dx \to 0 \quad \text{as } t \to \infty, \tag{12}$$

where $\rho = \int_0^\infty n(t=0,u)\phi(u)du$ and (λ, N) is the unique solution of the following eigenvalue problem

$$\begin{cases} \alpha\partial_x N(x) + \lambda N(x) = 2\int_0^\infty B(y)N(y)h\left(\frac{x}{y}\right)\frac{dy}{y} - B(x)N(x), \quad x \geq 0 \\ B(0)N(0) = 0, \quad \int N(x)dx = 1, \quad N(x) \geq 0, \quad \lambda > 0 \end{cases} \tag{13}$$

Case of the stationary distribution approximation

- We denote $G = BN$ and equation (13) can be expressed in terms of G:

$$\alpha\partial_x N(x) + \lambda N(x) = 2\int_0^\infty G(y)h\left(\frac{x}{y}\right)\frac{dy}{y} - G(x). \tag{14}$$

- We define the operators

$$L(N)(x) := \alpha\partial_x N(x) + \lambda N(x), \tag{15}$$

$$\mathcal{L}(G,h)(x) := 2\int_0^\infty G(y)h\left(\frac{x}{y}\right)\frac{dy}{y} - G(x), \tag{16}$$

then

$$L(N)(x) = \mathcal{L}(G,h)(x). \tag{17}$$

- We introduce the functions

$$\tilde{G}(u) = G(e^u), \quad \tilde{h}(u) = h(e^u) \quad \text{and } \tilde{N}(u) = N(e^u),$$

- Applying Fourier transform into both sides of (17), we obtain

$$\mathcal{F}[\tilde{h}](\xi) = \frac{1}{2}\frac{\mathcal{F}[L(\tilde{N})](\xi)}{\mathcal{F}[\tilde{G}](\xi)} + \frac{1}{2}, \qquad \xi \in \mathbb{R} \tag{18}$$

Case of the stationary distribution approximation

Proposition:

Assume that N satisfies $\int_{\mathbb{R}_+} x^{-k} N(x)dx < \infty$ for $k =, 1, 2$. For $\gamma \in (0, 1)$, the function h is explicitly given for a.e by

$$h(\gamma) = \frac{1}{2}\mathcal{F}^{-1}\left[A(\xi)\right]\big(\ln(\gamma)\big), \tag{19}$$

where

$$A(\xi) = \frac{\alpha(1+i\xi)\mathbb{E}\left[X_1^{-2-i\xi}\right] + \lambda\mathbb{E}\left[X_1^{-1-i\xi}\right]}{\mathbb{E}\left[B(X_1)X_1^{-1-i\xi}\right]}, \qquad \xi \in \mathbb{R}, \tag{20}$$

with $X_1 \sim N(x)dx$ and i is the unit imaginary number.

For the estimator: we replace the expectation by empirical means. Convergence is a work in progress.

References

- D. Aldous. Stopping times and tightness. *Annal of Probability*, 6, 1978.
- Vincent Bansaye and Viet Chi Tran. Branching feller diffusion for cell division with parasite infection. *ALEA, Lat. Am. J. Probab. Math. Stat.*, 8, 2011.
- M. Doumic, M. Hoffmann, P. Reynaud-Bouret, and V. Rivoirard. Nonparametric estimation of the division rate of a size-structured population. *SIAM Journal on Numerical Analysis*, 50, 2012.
- N. Fournier and S. Méléard. A microscopic probabilistic description of a locally regulated population and macroscopic approximations. *The Annals of Applied Probability*, 14, 2004.
- Sylvie Méléard and Viet Chi Tran. Slow and fast scales for superprocess limits of age-structured populations. *Stochastic Processes and their Applications*, 122, 2012.
- B. Perthame. *Transport equations arising in biology*. Frontiers in Mathematics. Birchkhauser, 2007.