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

[1]Milliman, [2]Université Paris-Diderot, [3]INRIA, [4]Sorbonne Université, [5]Université Paris-Dauphine, [6]Université de Rennes 1, [7]Université Paris 1, [8]Ecole Polytechnique, [9]Université Paris-Sud, [10]Université de Nice, [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 $\mathcal{Z}^N$ 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 $\mathcal{H}_N$ and $\mathcal{H}$)
  2. What is N?
  3. What do we mean by a proxy $\mathcal{Z}^N$ 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).

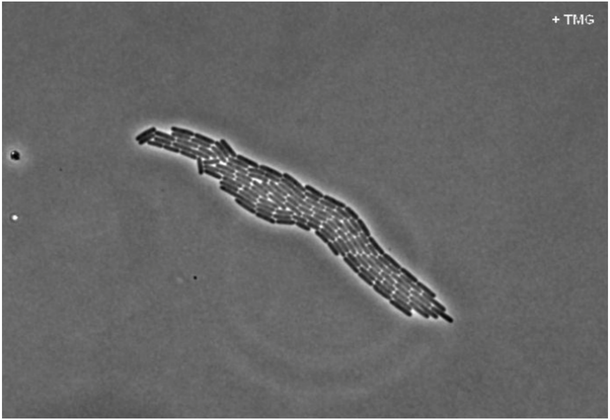

Figure: Evolution of a *E. Coli* population.

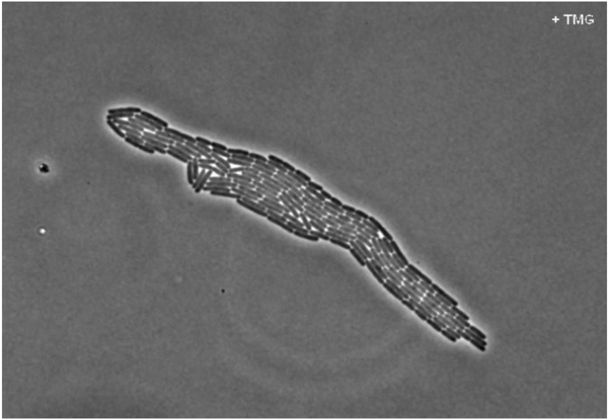

Figure: Evolution of a *E. Coli* population.

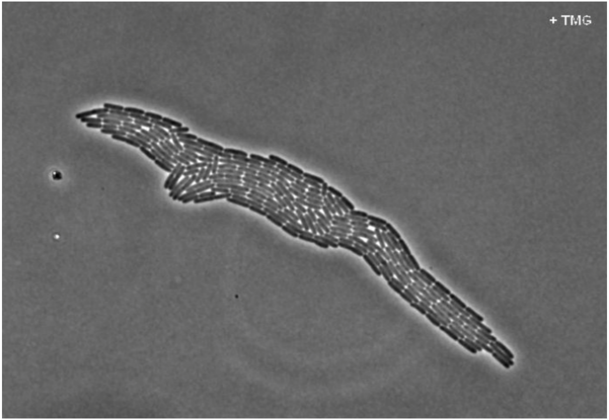

Figure: Evolution of a *E. Coli* population.

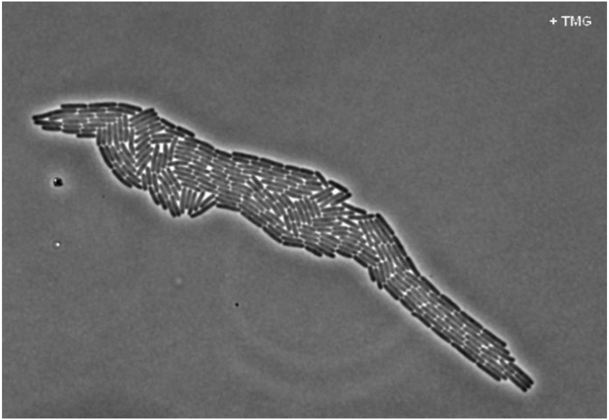

Figure: Evolution of a *E. Coli* population.

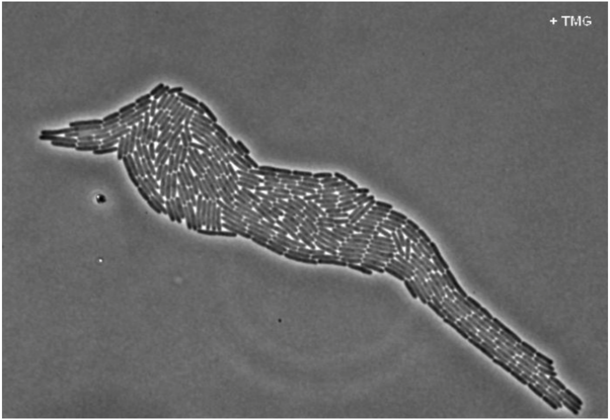

Figure: Evolution of a *E. Coli* population.

Growth-fragmentation models, informal description

- The growth of a cell size follows the deterministic evolution

$$\frac{dX(t)}{dt} = \kappa\big(X(t)\big)dt.$$

- 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 $\varphi \geq 0$:

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

  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 $\langle M, \mathbf{1} \rangle$ = size of the population M.

Example 1.1: An age-structured model

- $\big(A_i(t)\big)_{1\leq i\leq N_t}$ = all the (ordered) ages of the cell population at time t.
- $Z_t = \sum_{i=1}^{N_t} \delta_{A_i(t)}$ with $N_t = \langle Z_t, \mathbf{1}\rangle$.
- The division rate $a \mapsto B(a)$ is a function of age only!

Associated SDE:

$$Z_t = \tau_t Z_0 + \int_0^t \sum_{i\leq \langle Z_{s-},\mathbf{1}\rangle} \int_{0\leq\theta\leq B(a_i(Z_{s-}))} \big(2\delta_{t-s} - \delta_{a_i(Z_{s-})+t-s}\big) Q(ds, di, d\theta)$$

- Q : Poisson random measure, intensity $ds\big(\sum_{k\geq 1}\delta_k(di)\big)d\theta$.
- $\tau_t \sum_i \delta_{a_i} = \sum_i \delta_{a_i+t}$.

Example 1.1: An age-structured model

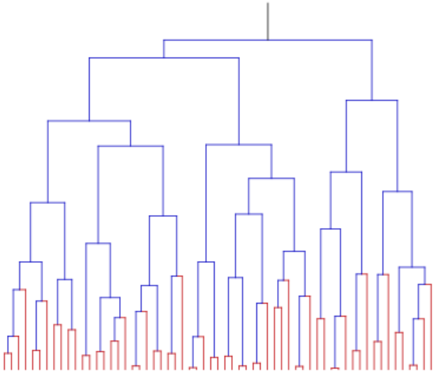

Figure: *A sample path of* $Z_t(da)_{0\leq t\leq T}$ *with* $B(a) = a^2$ *and* $T = 7$.

Example 1.1: limiting object

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

$$\begin{cases} \partial_t g(t,a) + \partial_a g(t,a) + B(a)g(t,a) = 0 \\ g(0,a) = g_0(a),\ g(t,0) = 2\int_0^\infty B(a)g(t,a)da. \end{cases}$$

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 $\mathcal{H}_N$ and $\mathcal{H}$)
  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 $\langle Z_T, \mathbf{1} \rangle$.
  - Z^N is $(Z_t)_{0 \leq t \leq T}$.
  - f is $(t, a) \mapsto g(t, a)$ or $a \mapsto B(a)$.
  - $\mathcal{H}^N$ and $\mathcal{H}$ are the SDE and the renewal equation.

Example 1.2: A size-structured model

- $\big(X_i(t)\big)_{1\leq i\leq N_t}$ = all the sizes of the cell population at time t.
- $Z_t = \sum_{i=1}^{N_t} \delta_{X_i(t)}$ with $N_t = \langle Z_t, \mathbf{1}\rangle$.
- The division rate $x \mapsto B(x)$ is a function of size only!

Associated SDE:

$$Z_t = \phi_{Z_0}(t) + \int_0^t \sum_{i\leq\langle Z_{s-},\mathbf{1}\rangle} \int_{0\leq\theta\leq B(a_i(Z_{s-}))} \left(2\delta_{\phi_{\frac{x_i(Z_{s-})}{2}}(t-s)} - \delta_{\phi_{x_i(Z_{s-})}(t-s)}\right) Q(ds, di, d\theta)$$

- $\frac{d}{dt}\phi_x(t) = \kappa\big(\phi_x(t)\big)$ with $\phi_x(0) = x$.
- $\phi_{\sum_i \delta_{x_i}}(t) = \sum_i \delta_{\phi_{x_i}(t)}$.

Example 1.2: limiting object

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

$$\begin{cases} \partial_t g(t,x) + \partial_x\big(\kappa(x)g(t,x)\big) + B(x)g(t,x) = 4B(2x)g(t,2x) \\ g(0,x) = g_0(x) \text{ and } g(t,0) = 0, t > 0. \end{cases}$$

Example 1.2: identification of the objects of interest

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

Observation schemes, proxy of Z^N

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

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

(b) We observe Z at a terminal time $t = T$. Two situations
  - $\mathcal{Z}^N = Z_T$.
  - $\mathcal{Z}^n$ 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) $\mathcal{Z}^N$ 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

$$\mathbb{T} = \bigcup_{m \in \mathbb{N}} \mathbb{G}_m, \quad \mathbb{G}_m = \{0,1\}^m, \quad (\mathbb{G}_0 = \emptyset).$$

- To each cell or node $u \in \mathbb{T}$, we associate a node with size at birth ξ_u and lifetime ζ_u.
- To each $u \in \mathbb{T}$, 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)

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

- Example 1.2 (size model)

$$Z_t = \sum_{u \in \mathbb{T}} \delta_{\phi_{\xi_u}(t-b_u)} \mathbf{1}_{\{b_u \le t < b_u + \zeta_u\}}.$$

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

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

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

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

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

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

Temporal data

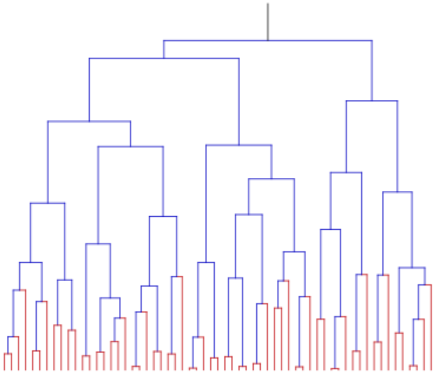

Figure: *Genealogical tree observed up to $T = 7$ for a time-dependent division rate $B(a) = a^2$ (60 cells). In blue: $\mathring{\mathcal{T}}_T$. In red: $\partial\mathcal{T}_T$.*

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

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

$$\begin{cases} (Z_t)_{0\le t\le T} = \left\{\zeta_u^T, \xi_u^T, u \in \mathcal{T}_T\right\}, \\ Z_T = \left\{\xi_u^T, u \in \partial\mathcal{T}_T\right\}, \end{cases}$$

with $\xi_u^T = \xi_u$ if $d_u \le T$ and $\phi_{\xi_u}(T - b_u)$ otherwise.

Observation scheme (c): genealogical data

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

$$\mathbb{T}_n = \bigcup_{m=0}^{n} \mathbb{G}_m.$$

- Observation scheme (c): informally, for some $n \geq 1$, we observe (ζ_u, ξ_u) along a subset of $\mathbb{T}_n$, called a ϱ-regular tree.

- Definition
  $\mathbb{U}_n \subseteq \mathbb{T}_n$ is a ϱ-regular tree if
  1. $u \in \mathbb{U}_n \implies u^- \in \mathbb{U}_n$ (u^- parent of u)
  2. $\left|\mathbb{U}_n \cap \mathbb{G}_n\right| \approx 2^{n\varrho}$ $(0 \leq \varrho \leq 1)$.

- Two extreme cases
  - Dense case: $\mathbb{U}_n = \mathbb{T}_n$, with $|\mathbb{U}_n| = 2^{n+1} - 1$ (and $\varrho = 1$).
  - Sparse case: $\mathbb{U}_n$ a single line along $\mathbb{T}_n$, with $|\mathbb{U}_n| = n$ (and $\varrho = 0$).

Observation scheme (c): genealogical data

- Take a ϱ-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, $T = T\big(N, (Z_t)_{t \geq 0}\big) \to \infty$ as $n \to \infty$.

Temporal versus geneaological data: selection bias!

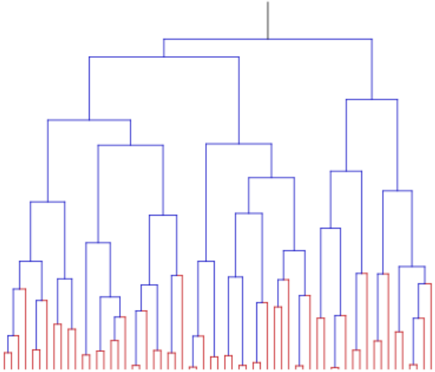

Figure: *Genealogical tree observed up to $T = 7$ for a time-dependent division rate $B(a) = a^2$ (60 cells). In blue: $\mathring{\mathcal{U}}_T$. In red: $\partial\mathcal{U}_T$.*

Genealogical data

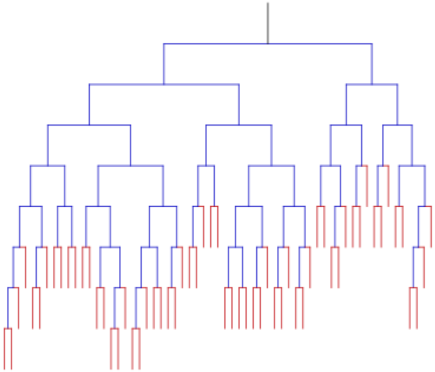

Figure: *The same outcome organised at a genealogical level.*

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 $\phi_x(t)$ = 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)

$$d\phi_x(t) = \kappa\big(\phi_x(t)\big)\,dt, \;\; \phi_x(0) = x \in (0, \infty).$$

- We may think of more general flows in between jumps.

Example 2: more general flows

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

$$\boxed{d\phi_x(t) = \kappa(\phi_x(t))dt + \sigma(\phi_x(t))dW_t, \quad \phi_x(0) = x}$$

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

$$\boxed{B\big(\phi_x(t)\big)dt \ \text{ during } \ [t, t+dt]}$$

  $B : \mathcal{X} \to [0, \infty)$ division rate.
- At division, a particle with trait y is replaced by two particles with traits $y\vartheta$ and $y(1-\vartheta)$, where $\mathrm{Law}(\vartheta) = r(dy)$.
- Parameters of the model : (κ, σ, B, r).

Bifurcating Markov chains

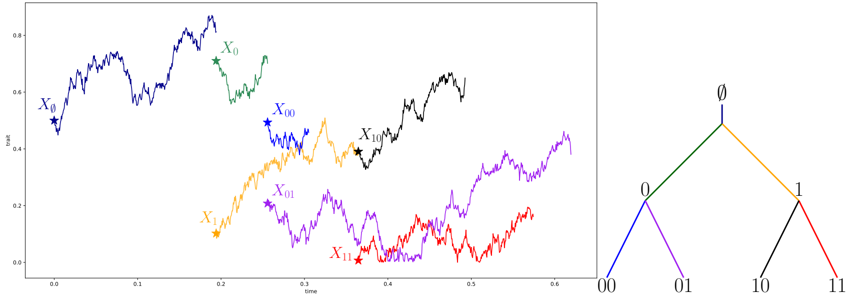

Figure: Example of a trajectory of a BMC and its associated genealogy.

Bifurcating Markov chains

- $\mathcal{X}$ a state space and $\mathcal{P}$: $\mathcal{X} \to \mathcal{X} \times \mathcal{X}$ a Markov kernel.
- $\big(\Omega, \mathcal{F}, (\mathcal{F}_m)_{m\geq 0}, \mathbb{P}\big)$ a filtered probability space.
- Notation: $|u| = m$ for $u \in \mathbb{G}_m$.
  $u0 = (u,0) \in \mathbb{G}_{m+1}, u1 = (u,1) \in \mathbb{G}_{m+1}$.
  $\mathcal{P}\psi(x) = \int_{\mathcal{X}\times\mathcal{X}} \psi(x, y_1, y_2)\mathcal{P}(x, dy_1 dy_2)$.

Definition

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

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

for every $m \geq 0$ and any family of (bounded) functions $(\psi_u)_{u\in\mathbb{G}_m}$.

Representation of $\mathcal{P}$ for Examples 1.1 and 1.2

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

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

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

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

with

$$\boxed{\mathcal{Q}(x, y) = \frac{B(2y)}{\frac{1}{2}\kappa\big(\phi_x^{-1}(2y)\big)} \exp\Big(-\int_{2x}^{y} \frac{B(2z)}{\frac{1}{2}\kappa\big(\phi_x^{-1}(2y)\big)} dz\Big) \mathbf{1}_{\{y \geq x/2\}}}$$

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

$$\mathbb{P}(\zeta_u \geq t + dt, |\zeta_u \geq t, X_{u^-}) = B\big(\phi_{X_{u^-}}(t)\big) dt.$$

Representation of $\mathcal{P}$ for Example 2

- Stochastic flow $\rightsquigarrow$ formulas become more intricate.
- Under apporiate 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

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

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

$$\int_0^t \psi(s, \phi_x(s)) ds = \int_0^t \int_{\mathcal{X}} \psi(s, y) dL_s^y(\phi_x)$$