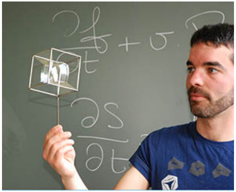
# Dynamics of phylogenies in a population with climate change

Chaire MMB

Viet Chi TRAN
Université Gustave Eiffel - France

September 22, 2021

# With...

A toy model with global warming

Genealogies and ancestral paths

# Model

★ Continuous time birth-death processes, stochastic evolution based on individual dynamics with spatial position ($x \in \mathbb{R}$), competition and environmental dependence.

# Model

★ Continuous time birth-death processes, stochastic evolution based on individual dynamics with spatial position ($x \in \mathbb{R}$), competition and environmental dependence.

★ **Large population:** the initial population size is proportional to $K$, with $K \to +\infty$,

# Model

★ Continuous time birth-death processes, stochastic evolution based on individual dynamics with spatial position ($x \in \mathbb{R}$), competition and environmental dependence.

★ **Large population:** the initial population size is proportional to $K$, with $K \to +\infty$,

★ **Asexual reproduction:** birth rate of 1. The individual at position $x \in \mathbb{R}$ gives birth to a new offspring at the same location.

★ **Motion/mutations:** during its life, the individual moves according to a Brownian motion with diffusion coefficient $\sigma > 0$ **or** according to a pure jump process with jump measure $\gamma m(x, y)dy$ (nonlocal mutation operator).

# Model

★ Continuous time birth-death processes, stochastic evolution based on individual dynamics with spatial position ($x \in \mathbb{R}$), competition and environmental dependence.

★ **Large population:** the initial population size is proportional to $K$, with $K \to +\infty$,

★ **Asexual reproduction:** birth rate of 1. The individual at position $x \in \mathbb{R}$ gives birth to a new offspring at the same location.

★ **Motion/mutations:** during its life, the individual moves according to a Brownian motion with diffusion coefficient $\sigma > 0$ **or** according to a pure jump process with jump measure $\gamma m(x, y)dy$ (nonlocal mutation operator).

★ **Natural death:** an individual at position $x$ at time $t$ dies with the natural death rate

$$\frac{1}{2}\left(x - \sigma ct\right)^2.$$

The optimal location is $\sigma ct$, which moves linearly with time.

★ **Competition:** each individual dies with the extra competition rate $N_t/K$.

# Stochastic differential equation with jumps

★ The population is represented by:

$$Z_t^K(dx) = \frac{1}{K}\sum_{i=1}^{N_t^K} \delta_{x_i(t)} \in \mathcal{M}_F(\mathbb{R})$$

Notation: $\langle Z_t^K, f\rangle = \int_{\mathbb{R}} f(x) Z_t^K(dx) = \frac{1}{K}\sum_{i=1}^{N_t^K} f\big(x_i(t)\big)$.

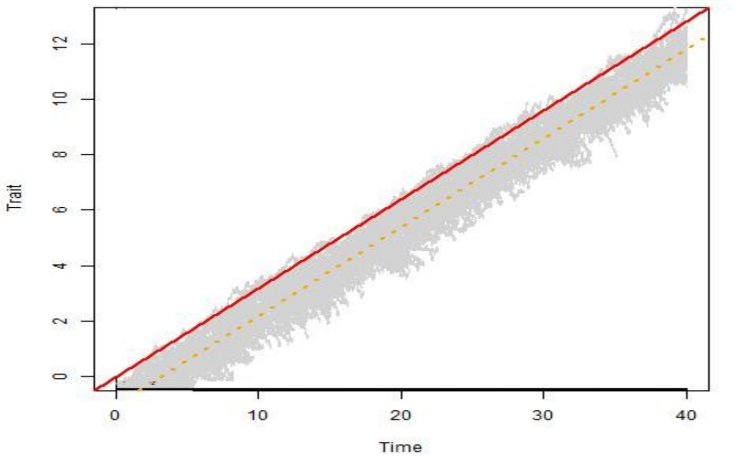
★ If $\sup_K \mathbb{E}(\langle Z_0^K, 1\rangle^2) < +\infty$, the evolution of $(Z_t^K, t \geq 0)$ can be described by a SDE and for all $f \in \mathcal{C}_b^2(\mathbb{R},\mathbb{R})$:

$$\langle Z_t^K, f\rangle = \langle Z_0^K, f\rangle$$
$$+ \int_0^t \int_{\mathbb{R}} \left[\Big(1 - \frac{1}{2}(x - \sigma cs)^2 - \langle Z_s^K, 1\rangle\Big) f(x) + \frac{\sigma^2}{2} f''(x)\right] Z_s^K(dx)\, ds + M_t^{K,f}$$

where $M_t^K$ is a square integrable martingale with:

$$\langle M^{K,f}\rangle_t = \frac{1}{K}\int_0^t \int_{\mathbb{R}} \left[\Big(1 + \frac{1}{2}(x - \sigma cs)^2 + \langle Z_s^K, 1\rangle\Big) f^2(x) + \sigma^2 (f'(x))^2\right] Z_s^K(dx)\, ds.$$

# Simulation (1)



# Large population limit

★ **Prop:** If $\sup_K \mathbb{E}(\langle Z_0^K, 1\rangle^{2+\varepsilon}) < +\infty$ and $\lim_{K\to+\infty} Z_0^K(dx) = u_0(x)dx$, then when $K \to +\infty$, $(Z_t^K, t \geq 0)$ converges in $\mathbb{D}(\mathbb{R}_+, \mathcal{M}_F(\mathbb{R}))$ to $(\xi_t, t \geq 0) = (u(t,x)dx,\ t \geq 0)$ where:

$$\partial_t u(t,x) = \big(1 + \frac{1}{2}(x - \sigma ct)^2 + \int_{\mathbb{R}} u(t,x')dx'\big)u(t,x) + \frac{\sigma^2}{2}\partial^2_{xx}u(t,x).$$

There exists a unique non negative solution for this PDE.

★ **Change of variable:** $f(t,y) = u(t, y + \sigma ct)$. Then:

$$\partial_t f(t,y) = \big(1 + \frac{1}{2}y^2 + \int_{\mathbb{R}} f(t,y)dy\big)f(t,y) + \sigma c\partial_y f(t,y) + \frac{\sigma^2}{2}\partial^2_{yy}f(t,y).$$

1. Fournier Méléard, *Annals of Applied Probability*, (2004)

# Large population limit

★ **Prop:** If $\sup_K \mathbb{E}(\langle Z_0^K, 1\rangle^{2+\varepsilon}) < +\infty$ and $\lim_{K\to+\infty} Z_0^K(dx) = u_0(x)dx$, then when $K \to +\infty$, $(Z_t^K, t \geq 0)$ converges in $\mathbb{D}(\mathbb{R}_+, \mathcal{M}_F(\mathbb{R}))$ to $(\xi_t, t \geq 0) = (u(t,x)dx,\ t \geq 0)$ where:

$$\partial_t u(t,x) = \Big(1 + \frac{1}{2}(x - \sigma ct)^2 + \int_{\mathbb{R}} u(t,x')dx'\Big) u(t,x) + \frac{\sigma^2}{2}\partial^2_{xx} u(t,x).$$

There exists a unique non negative solution for this PDE.

★ **Change of variable:** $f(t,y) = u(t, y + \sigma ct)$. Then:

$$\partial_t f(t,y) = \Big(1 + \frac{1}{2}y^2 + \int_{\mathbb{R}} f(t,y)dy\Big) f(t,y) + \sigma c \partial_y f(t,y) + \frac{\sigma^2}{2}\partial^2_{yy} f(t,y).$$

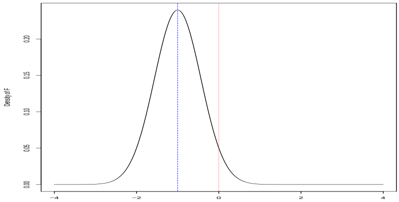
★ For the nonlocal mutation operator:

$$\partial_t f(t,y) = \Big(1 + \frac{1}{2}y^2 + \int_{\mathbb{R}} f(t,y)dy\Big) f(t,y) + \sigma c \partial_y f(t,y) + \gamma \int_{\mathbb{R}} \big(f_t(y) - f_t(x)\big) m(y,x)dy.$$

1. Fournier Méléard, *Annals of Applied Probability*, (2004)

# Stationary solution

$$\partial_t f(t,y) = \Big(1 + \frac{1}{2}y^2 + \int_{\mathbb{R}} f(t,y)dy\Big) f(t,y) + \sigma c \partial_y f(t,y) + \frac{\sigma^2}{2}\partial^2_{yy} f(t,y).$$




★ When $(c^2 + \sigma)/2 < 1$, there exists a unique non trivial stationary solution:

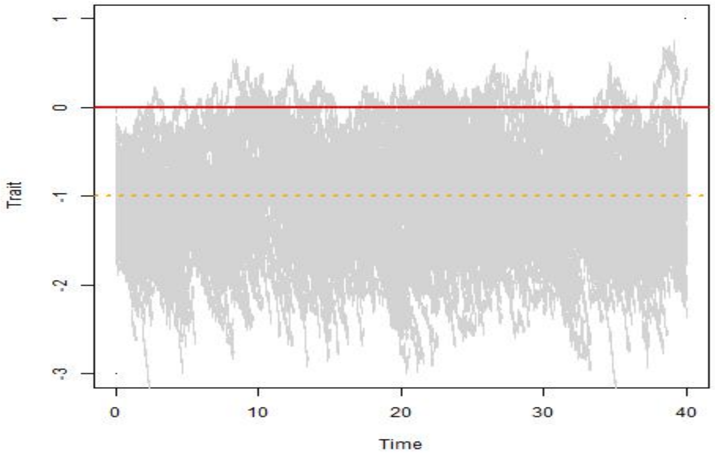
$$F(y) = \frac{\lambda}{\sqrt{2\pi\sigma}} \exp\Big( - \frac{(y+c)^2}{2\sigma}\Big),$$

$$\text{with} \quad \|F\|_1 = \lambda = 1 - \frac{c^2}{2} - \frac{\sigma}{2}.$$

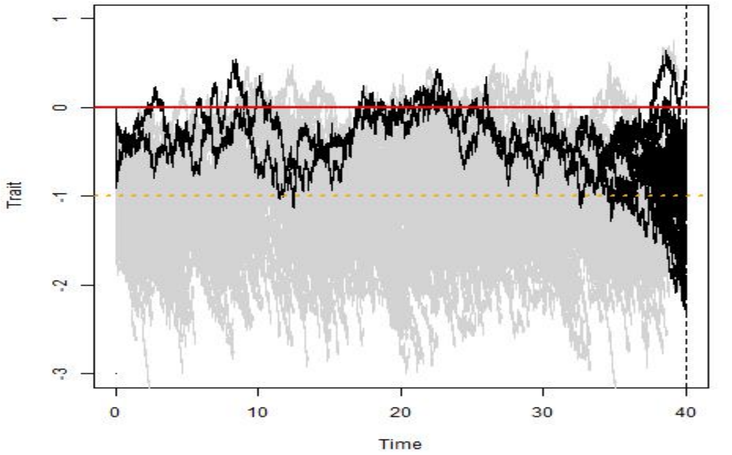
★ For the non-local mutation operator: existence and uniqueness as well.

1. Cloez and Gabriel, *CRAS*, (2020)

# A simulation: who are the ancestors?



# Historical process



A toy model with global warming

Genealogies and ancestral paths

# Historical process (2)

★ We consider the ancestral path or lineage:

$$y_t = \text{trait of the ancestor living at time } t$$

$y \in \mathbb{D}_{\mathbb{R}} = \mathbb{D}(\mathbb{R}_+, \mathbb{R})$ embedded with the Skohorod topology.

**Notation:** $y_t$, $y^t = y_{.\wedge t}$, $(y|s|w)$

1. Dawson Perkins, *Memoirs of the AMS*, (1991)
2. Méléard Tran, *EJP*, (2012)

# Historical process (2)

★ We consider the ancestral path or lineage:

$$y_t = \text{trait of the ancestor living at time } t$$

$y \in \mathbb{D}_{\mathbb{R}} = \mathbb{D}(\mathbb{R}_+, \mathbb{R})$ embedded with the Skohorod topology.

**Notation:** $y_t$, $y^t = y_{.\wedge t}$, $(y|s|w)$

★ Population:

$$H_t^K(dy) = \frac{1}{K} \sum_{i=1}^{N_t^K} \delta_{y^i_{.\wedge t}}(dy)$$

in $\mathcal{M}(\mathbb{D}_{\mathbb{R}})$ embedded with the weak convergence topology. Thus $H^K \in \mathbb{D}(\mathbb{R}_+, \mathcal{M}_F(\mathbb{D}_{\mathbb{R}}))$, embedded with the Skorohod topology.

1. Dawson Perkins, *Memoirs of the AMS*, (1991)
2. Méléard Tran, *EJP*, (2012)

# Test functions for historical processes

★ Usual class of test functions:

$$\varphi(y) = \prod_{j=1}^{m} g_j(y_{t_j})$$

for $m \in \mathbb{N}^*$, $0 \leq t_1 < \cdots < t_m$ and $\forall j \in [\![1, m]\!]$ , $g_j \in \mathcal{C}_b^2(\mathbb{R}, (0, +\infty))$.
**However these functions are not continuous for discontinuous $y$'s.**

1. Dawson Perkins, *Memoirs of the AMS*, (1991)
2. Méléard Tran, *EJP*, (2012)

# Test functions for historical processes

★ Usual class of test functions:

$$\varphi(y) = \prod_{j=1}^{m} g_j(y_{t_j})$$

for $m \in \mathbb{N}^*$, $0 \leq t_1 < \cdots < t_m$ and $\forall j \in [\![1, m]\!]$, $g_j \in \mathcal{C}_b^2(\mathbb{R}, (0, +\infty))$.
**However these functions are not continuous for discontinuous $y$'s.**

★ For a real $\mathcal{C}_b^2$-function $g$ on $\mathbb{R}_+ \times \mathbb{R}$ and a real $\mathcal{C}_b^2$-function $G$ on $\mathbb{R}$, we define the continuous function $G_g$ as

$$G_g(y) = G\Big(\int_0^T g(s, y_s) ds\Big).$$

1. Dawson Perkins, *Memoirs of the AMS*, (1991)
2. Méléard Tran, *EJP*, (2012)

# Test functions for historical processes

★ Usual class of test functions:

$$\varphi(y) = \prod_{j=1}^{m} g_j(y_{t_j})$$

for $m \in \mathbb{N}^*$, $0 \leq t_1 < \cdots < t_m$ and $\forall j \in [\![1, m]\!]$, $g_j \in \mathcal{C}_b^2(\mathbb{R}, (0, +\infty))$.
**However these functions are not continuous for discontinuous $y$'s.**

★ For a real $\mathcal{C}_b^2$-function $g$ on $\mathbb{R}_+ \times \mathbb{R}$ and a real $\mathcal{C}_b^2$-function $G$ on $\mathbb{R}$, we define the continuous function $G_g$ as

$$G_g(y) = G\Big(\int_0^T g(s, y_s)ds\Big).$$

★ **Lemma:** Let $\varphi$ be a test function of the form proposed by Dawson. Then, there exists a sequence of test functions of the second form $(\varphi_q)_{q\in\mathbb{N}^*}$ such that for every $y \in \mathbb{D}_\mathbb{R}$ and every $t \in \mathbb{R}_+$ at which $y$ is continuous,

$$\lim_{q\to+\infty} \varphi_q(y) = \varphi(y).$$

(choose $G(x) = e^x$ and $g_q(s, y_s) = \sum_{j=1}^{m} \log g_j(y_s) k^q(t_j - s)$)

1. Dawson Perkins, *Memoirs of the AMS*, (1991)
2. Méléard Tran, *EJP*, (2012)

# Evolution equation for $H^K$

★ With the same initial conditions as before, we have:

$$\langle H_t^K, \varphi \rangle = \langle H_0^K, \varphi \rangle + \int_0^t \int_{\mathcal{C}(\mathbb{R}_+,\mathbb{R})} \Big( \frac{\sigma^2}{2} \widetilde{\Delta}\varphi(s,y) - \sigma c \widetilde{D}\varphi(s,y) + (1 - \frac{y_s^2}{2} - \langle H_s^K, 1 \rangle)\varphi(y) \Big) H_s^K(dy)\, ds + \mathcal{M}_t^{K,\varphi},$$

where $\mathcal{M}_t^{K,\varphi}$ is a square integrable martingale with predictable quadratic variation process:

$$\langle \mathcal{M}^{K,\varphi} \rangle_t = \frac{1}{K} \int_0^t \int_{\mathcal{C}(\mathbb{R}_+,\mathbb{R})} \Big( (1 + \frac{y_s^2}{2} + \langle H_s^K, 1 \rangle)\varphi^2(s,y) + \sigma^2 (\widetilde{D}\varphi(s,y))^2 \Big) H_s^K(dy)\, ds.$$

★ Let $U_T^K$ be a uniform random variable on the set of living individuals at time $T$. and let

$$Y_t^K = X_t^{U_T^K}, \qquad \text{for } t \in [0,T].$$

Then,

$$\mathbb{E}_x \left[ \Phi \left( Y_t^K, \ t \in [0,T] \right) \right] = \mathbb{E}_{\delta_x} \left[ \frac{\langle H_T^K, \Phi \rangle}{\langle H_T^K, 1 \rangle} \right].$$

# Coupling with a branching process

★ We now assume that $\sup_K \mathbb{E}(\langle Z_0^K, 1\rangle^3) < +\infty$ and that $\lim_{K\to+\infty} Z_0^K = F$.

★ Then,

$$\lim_{k\to+\infty} \mathbb{E}\big(\sup_{t\leq T} |\langle Z_t^K, f\rangle - \langle F, f\rangle|\big) = 0.$$

★ Let us freeze the competition term $\langle Z_t^K, 1\rangle$ to $\|F\|_1 = \lambda$. We obtain:

$$\langle \tilde{Z}_t^K, f\rangle = \langle Z_0^K, f\rangle + \int_0^t \int_{\mathbb{R}} \tilde{Z}_s^K(dy)\Big[\big(1 - \frac{1}{2}y^2 - \lambda\big)f(y) - c\sigma f'(y) + \frac{\sigma^2}{2}f''(y)\Big] + \tilde{M}_t^{K,f},$$

where

$$\langle \tilde{M}^{K,f}\rangle_t = \frac{1}{K}\int_0^t \int_{\mathbb{R}} \Big[\big(1 + \frac{y^2}{2} + \lambda\big)f^2(y) + \sigma^2(f')^2(y)\Big]\tilde{Z}_s^K(dy)ds.$$

This process satisfies the branching property (independence between individuals)!
Similar equation for $\widetilde{H}^K$.

# Approximation by the branching process

★ **Prop:** If $Z_0^K \xrightarrow[K\to\infty]{w} F$ . Then for any continuous and bounded function $\varphi$ on $\mathbb{D}$,

$$\lim_{K\to+\infty} \mathbb{E}(\sup_{t\leq T} |\langle H_t^K, \varphi\rangle - \langle \widetilde{H}_t^K, \varphi\rangle|^2) = 0$$

$$\text{and} \qquad \lim_{k\to+\infty} \mathbb{E}\Big(\sup_{t\leq T} |\langle Z_t^K, f\rangle - \langle \tilde{Z}_t^K, f\rangle|^2\Big) = 0.$$

★ We have a toolbox for dealing with branching processes.

$(\tilde{Z}_t, t \geq 0)$ is the branching process started with one particle.

# Many-to-one formula

★ Based on the branching property, we can replace expectation over the tree by expectation along 1 branch!

$$\mathbb{E}_x\left[\langle \tilde{Z}_t, f\rangle\right] = \mathbb{E}_x\left[\exp\left(\int_0^t \left(1-\frac{1}{2}Y_s^2-\lambda\right)ds\right)f(Y_t)\right] =: \widehat{P}_t f(x),$$

where $Y$ is the drifted motion process, for instance:

$$dY_t = \sigma(dB_t - cdt).$$

★ This can be generalized in:

$$\mathbb{E}_x\left[\langle \tilde{H}_t, \Phi\rangle\right] = \mathbb{E}_x\left[\exp\left(\int_0^t \left(1-\frac{1}{2}Y_s^2-\lambda\right)ds\right)\Phi(Y_s, s\leq t)\right].$$

★ The expected population size $m_t(x) = \mathbb{E}_x(\langle \tilde{Z}_t, 1\rangle)$ satisfies:

$$m_t(x) = \mathbb{E}_x\left[\exp\left(\int_0^t \left(1-\frac{1}{2}Y_s^2-\lambda\right)ds\right)\right]$$

from which we deduce that $m \in C_b^{1,\infty}([0,T]\times\mathbb{R})$.
★ We can thus define a probability measure on path space by renormalizing the intensity measure of $\tilde{H}_t$ by $m_t(x)$.

1. Hardy Harris, *Séminaires de probabilité*, (2009)

# Many-to-one formula (2)

★ **Th:** We have

$$\frac{1}{m_T(x)}\mathbb{E}_x\left[\langle \tilde{H}_T,\Phi\rangle\right]=\mathbb{E}_x\left[\Phi(\widetilde{Y}_t,\ t\le T)\right]$$

for the inhomogeneous Markov process $\widetilde{Y}_t$ (depending on $T$) with infinitesimal generator

$$\mathcal{G}_t f(x)=\frac{L(m_{T-t}f)(x)-f(x)Lm_{T-t}(x)}{m_{T-t}(x)},$$

$L$ being the generator of $Y$, for example $Lf(x)=\frac{\sigma^2}{2}f''(x)-\sigma c f'(x)$ in the Brownian case.

★ **Th:** Returning to the original process: recall $Y_t^K=X_t^{U_T^K}$.

$$\begin{aligned}\lim_{K\to+\infty}\mathbb{E}_F\left[\Phi\left(Y_t^K,\ t\in[0,T]\right)\right]&=\lim_{K\to+\infty}\mathbb{E}_{Z_0^K}\left[\frac{\langle H_T^K,\Phi\rangle}{\langle H_T^K,1\rangle}\right]\\&=\lim_{K\to+\infty}\mathbb{E}_{Z_0^K}\left[\frac{\langle \tilde{H}_T^K,\Phi\rangle}{\langle \tilde{H}_T^K,1\rangle}\right]\\&=\int_{\mathbb{R}}\mathbb{E}_x\left[\Phi(\tilde{Y}_s,s\le T)\right]\frac{m_T(x)F(dx)}{\lambda}\end{aligned}$$

1. Marguet, (2018)

# Spine of the process (Brownian case)

★ Using Feynman-Kac's formula, $m$ is the solution of

$$\partial_t m = \frac{\sigma^2}{2}\partial_{yy} m - \sigma c \partial_y m + (1 - \frac{y^2}{2} - \lambda)m, \qquad m_0(y) = 1.$$

Following Fitzsimmons-Pitman-Yor arguments using Girsanov's, we obtain that:

$$m_t(y) = \sqrt{1 + \tanh(\sigma t)} \exp\left(-\frac{\left(y + e^{-\sigma t} c\right)^2}{2\sigma}(1 + \tanh(\sigma t)) + \frac{(y+c)^2}{2\sigma}\right).$$

★ Using the explicit value of $m_T(x)$, we obtain that:

$$\widetilde{Y}_t = \frac{\cosh(\sigma(T-t))}{\cosh(\sigma T)}\widetilde{Y}_0 + c\cosh(\sigma(T-t))\big(\tanh(\sigma(T-t)) - \tanh(\sigma T)\big)$$
$$+ \sigma \cosh(\sigma(T-t)) \int_0^t \frac{dB_s}{\cosh(\sigma(T-s))}.$$

1. Wenocur, *J. Appl. Probab.*, (1986)
2. Fitzsimmons Pitman Yor, Springer, (1993)

# Backward spine (Brownian case)

★ For any $t \geq 0$, $\widetilde{Y}_t$ is a Gaussian r.v.

Computation gives that:

$$\widetilde{Y}_t \sim \mathcal{N}\left(-ce^{-\sigma(T-t)},\ \frac{\sigma}{1+\tanh(\sigma(T-t))}\right)$$

and thus, the density $p(t, y)$ of $\widetilde{Y}_t$ is:

$$\partial_y \log p(t, y) = -\frac{y + ce^{-\sigma(T-t)}}{\sigma}\left(1 + \tanh(\sigma(T-t))\right).$$

★ Using a formula by Haussmann-Pardoux, the time-reverse diffusion process of $\widetilde{Y}$ is the Ornstein-Uhlenbeck process:

$$dX_t = -\sigma\big(cX_t dt + dB_t\big).$$

1. Haussmann Pardoux, *Annals of Probability*, (1986)

# Spine for non-local diffusions (1)

★ Recall the generator of $\tilde{Y}$:

$$\mathcal{G}_t f(y) = \frac{L(m_{T-t} f)(y) - f(y) L m_{T-t}(y)}{m_{T-t}(y)},$$

★ We can compute the semi-group of the inhomogeneous Markov process $\tilde{Y}$:

$$\tilde{P}_{s,t+s} f(x) = \frac{\widehat{P}_t (f m_{T-t-s})(x)}{m_{T-s}(x)},$$

where for $Y$ of generator $Lf(x) = \gamma \int_{\mathbb{R}} (f(y) - f(x)) m(x,y) dy - \sigma c f'(x)$,

$$\widehat{P}_t f(x) = \mathbb{E}_x \left[ \exp\left( \int_0^t \left( 1 - \frac{1}{2} Y_s^2 - \lambda \right) ds \right) f(Y_t) \right].$$

★ Define for $L^* f(x) = \gamma \int_{\mathbb{R}} (f(y) - f(x)) m(y,x) dy - \sigma c f'(x)$,

$$\widehat{P}_t^* f(x) = \mathbb{E}_x \left[ \exp\left( \int_0^t \left( 1 - \frac{1}{2} (Y_s^*)^2 - \lambda \right) ds \right) f(Y_t^*) \right].$$

# Spine for non-local diffusions (2)

★ For $f, g$ measurable and bounded, $\langle g, \widehat{P}_t f\rangle = \langle \widehat{P}_t^* g, f\rangle$.

★ $\int_{\mathbb{R}} m_t(x)F(x)dx = \langle \widehat{P}_t 1, F\rangle = \langle 1, \widehat{P}_t^* F\rangle = \langle 1, F\rangle = \lambda$.

★ As a consequence, if we start from $m_T F$,

$$\mathbb{E}_{m_T F}\left[f(\tilde{Y}_t)\right] = \int_{\mathbb{R}} f(x) m_{T-t}(x) F(x) dx$$

and $\tilde{Y}_t \rightsquigarrow m_{T-t}(x)F(x)dx$.

Proof:

$$\begin{aligned}\mathbb{E}_{m_T F}\left[f(\tilde{Y}_t)\right] =& \langle m_T F, \frac{\widehat{P}_t(f m_{T-t})}{m_T}\rangle = \langle F, \widehat{P}_t(f m_{T-t})\rangle \\ =& \langle \widehat{P}_t^* F, f m_{T-t}\rangle = \langle F, f m_{T-t}\rangle.\end{aligned}$$

# Backward spine

$$\tilde{P}_{s,t+s}f(x) = \frac{\widehat{P}_t(fm_{T-t-s})(x)}{m_{T-s}(x)}.$$

★ Returning the time for $\tilde{Y}$ can be done without computing explicitly $m_t(x)$. The backward spin is a homogeneous Markov process with semigroup:

$$P_t^R f(x) = \frac{\widehat{P}_t^*(fF)}{F},$$

where $\widehat{P}^*$ is the dual of $\widehat{P}$. For the nonlocal mutations, this yields the generator:

$$\begin{aligned} L^R f(x) =& \frac{L^*(fF)(x)}{F(x)} + \left(1 - \frac{x^2}{2} - \lambda\right) f(x) \\ =& -\sigma c f'(x) + \gamma \int_{\mathbb{R}} \left(f(y) - f(x)\right) \frac{F(y)}{F(x)} m(y,x) dy. \end{aligned}$$

1. Nagasawa, *Nagoya Math. J.*, (1964)
2. Reinhard Roynette, *AIHP*, (1970)
3. Dellacherie Meyer, Hermann, (1987)

## Thank You