

# Échantillonnage dans une population structurée branchante

Aline Marguet

8 juin 2016



# Un modèle pour décrire l'évolution d'une population de cellules

Population de cellules caractérisées par un trait (âge, nombre de parasites, taille, ...)

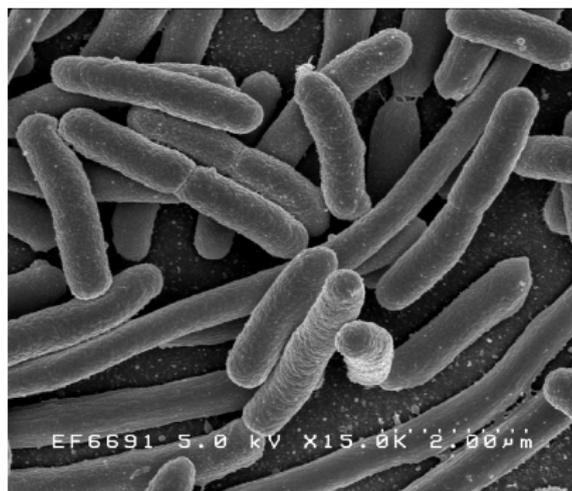


Image : NIAD

## Description du modèle

- Le trait  $(X_t^u)_{t \geq 0}$  de chaque individu  $u$  évolue suivant un processus de Markov de générateur  $\mathcal{G}$ .

# Description du modèle

- Le trait  $(X_t^u)_{t \geq 0}$  de chaque individu  $u$  évolue suivant un processus de Markov de générateur  $\mathcal{G}$ .
- Chaque individu  $u$  se divise à un taux  $B(X_t^u)$ , i.e.

$$\mathbb{P}(\beta(u) > t | (X_s^u)_{s \leq t}) = \exp\left(-\int_0^t B(X_s^u) ds\right),$$

avec  $\beta(u)$  le temps de division de la cellule  $u$ .

## Description du modèle

- Le trait  $(X_t^u)_{t \geq 0}$  de chaque individu  $u$  évolue suivant un processus de Markov de générateur  $\mathcal{G}$ .
- Chaque individu  $u$  se divise à un taux  $B(X_t^u)$ , i.e.

$$\mathbb{P}(\beta(u) > t | (X_s^u)_{s \leq t}) = \exp\left(-\int_0^t B(X_s^u) ds\right),$$

avec  $\beta(u)$  le temps de division de la cellule  $u$ .

- À sa mort, un individu de trait  $x$  est remplacé par  $A_u$  descendants. Les traits à la naissance sont distribués suivant une loi qui dépend de  $x$  et  $A_u$ .

## Description du modèle

- Le trait  $(X_t^u)_{t \geq 0}$  de chaque individu  $u$  évolue suivant un processus de Markov de générateur  $\mathcal{G}$ .
- Chaque individu  $u$  se divise à un taux  $B(X_t^u)$ , i.e.

$$\mathbb{P}(\beta(u) > t | (X_s^u)_{s \leq t}) = \exp\left(-\int_0^t B(X_s^u) ds\right),$$

avec  $\beta(u)$  le temps de division de la cellule  $u$ .

- À sa mort, un individu de trait  $x$  est remplacé par  $A_u$  descendants. Les traits à la naissance sont distribués suivant une loi qui dépend de  $x$  et  $A_u$ .
- Conditionnellement au trait de leur ancêtre, les descendants évoluent indépendamment les uns des autres.

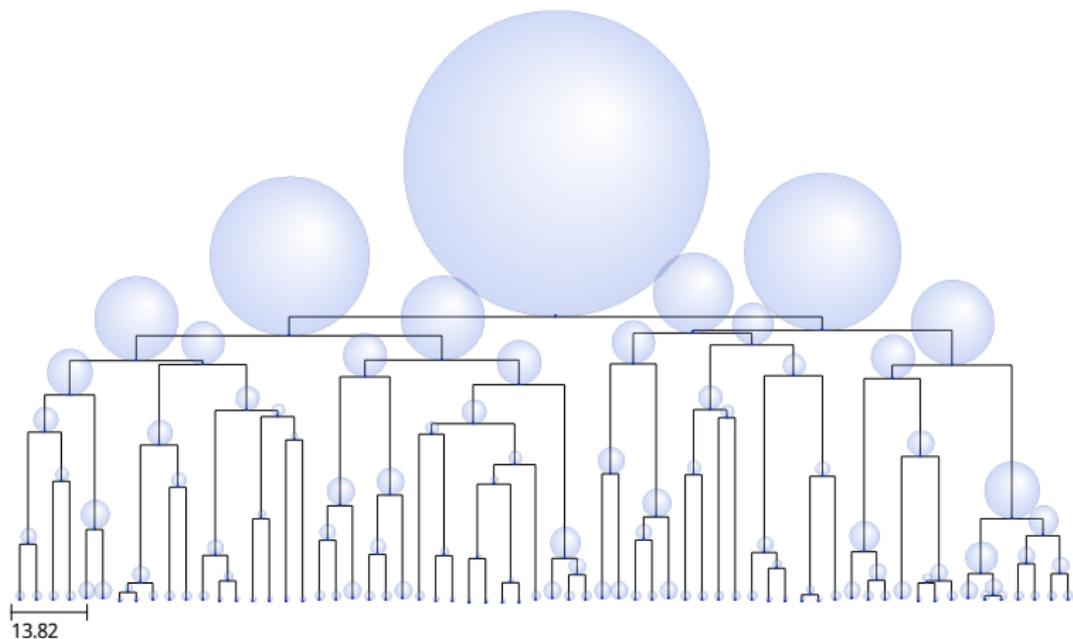


Figure – Généalogie descendante jusqu'au temps  $T = 50$  d'une population de cellules structurée par la taille issue d'une cellule de taille 1. Chaque cellule croît exponentiellement à taux 0.01 et se divise à taux  $B(x) = x$ .

Suivant Fournier et Méléard (2004), on pose :

$$Z_t = \sum_{u \in V_t} \delta_{X_t^u},$$

- $V_t$  : population en vie au temps  $t$ ,

# Description à l'aide d'un processus à valeurs mesures

Suivant Fournier et Méléard (2004), on pose :

$$Z_t = \sum_{u \in V_t} \delta_{X_t^u},$$

- $V_t$  : population en vie au temps  $t$ ,
- $N_t = \# V_t$ .

# Description à l'aide d'un processus à valeurs mesures

Suivant Fournier et Méléard (2004), on pose :

$$Z_t = \sum_{u \in V_t} \delta_{X_t^u},$$

- $V_t$  : population en vie au temps  $t$ ,
- $N_t = \# V_t$ .

Pour la bonne définition du processus, deux jeux d'hypothèses sont considérés.

## Hypothèse A

① Il existe  $b_1, b_2 \geq 0$  et  $\gamma \geq 0$  tels que pour tout  $x \in \mathcal{X}$ ,

$$B(x) \leq b_1 |x|^\gamma + b_2.$$

## Hypothèse A

- ① Il existe  $b_1, b_2 \geq 0$  et  $\gamma \geq 0$  tels que pour tout  $x \in \mathcal{X}$ ,

$$B(x) \leq b_1 |x|^\gamma + b_2.$$

- ② Pour tout  $t \geq 0$ , il existe  $\underline{x}(t) \in \mathcal{X}$  croissant en temps tel que pour tout  $x \in \mathcal{X}$ , si  $A_u = k$  et  $a$  uniformément distribué sur  $[0, 1]$ . :

$$\sum_{i=1}^k F_i^{(k)}(x, a) \leq x \vee \underline{x}(t), \text{ p.s.}$$

## Hypothèse A

- ① Il existe  $b_1, b_2 \geq 0$  et  $\gamma \geq 0$  tels que pour tout  $x \in \mathcal{X}$ ,

$$B(x) \leq b_1 |x|^\gamma + b_2.$$

- ② Pour tout  $t \geq 0$ , il existe  $\underline{x}(t) \in \mathcal{X}$  croissant en temps tel que pour tout  $x \in \mathcal{X}$ , si  $A_u = k$  et  $a$  uniformément distribué sur  $[0, 1]$ . :

$$\sum_{i=1}^k F_i^{(k)}(x, a) \leq x \vee \underline{x}(t), \text{ p.s.}$$

- ③ Il existe  $\bar{m} \geq 0$  tel que pour tout  $x \in \mathcal{X}$ ,

$$m(x) = \sum_k k p_k(x) \leq \bar{m}.$$

## Hypothèse B

Pour tout  $T > 0$ , il existe  $c_1, c_2 \geq 0$  tels que pour tout  $t \in [0, T]$  et  $x \in \mathcal{X}$  :

$$\mathcal{G}h_\gamma(x) \leq c_1 h_\gamma(x) + c_2,$$

avec  $\gamma$  défini précédemment et  $h_\gamma(x) = |x|^\gamma = \left(\sum_{i=1}^d x_i\right)^\gamma$ , pour  $x \in \mathbb{R}_+^d$ .

## Hypothèse B

Pour tout  $T > 0$ , il existe  $c_1, c_2 \geq 0$  tels que pour tout  $t \in [0, T]$  et  $x \in \mathcal{X}$  :

$$\mathcal{G}h_\gamma(x) \leq c_1 h_\gamma(x) + c_2,$$

avec  $\gamma$  défini précédemment et  $h_\gamma(x) = |x|^\gamma = \left(\sum_{i=1}^d x_i\right)^\gamma$ , pour  $x \in \mathbb{R}_+^d$ .

En particulier, si on a :

$$\mathcal{G}f(x) = r(x, t)f'(x) + \sigma^2(x, t)f''(x),$$

alors l'hypothèse précédente est vérifiée si :

$$r(x, t) \leq a_r x + b_r \text{ et } \sigma(x, t) \leq a_\sigma x + b_\sigma.$$

## Trait d'un individu échantillonné

On note  $U(t)$  la variable aléatoire correspondant à un individu échantillonné dans la population au temps  $t$ . On cherche à décrire la loi de

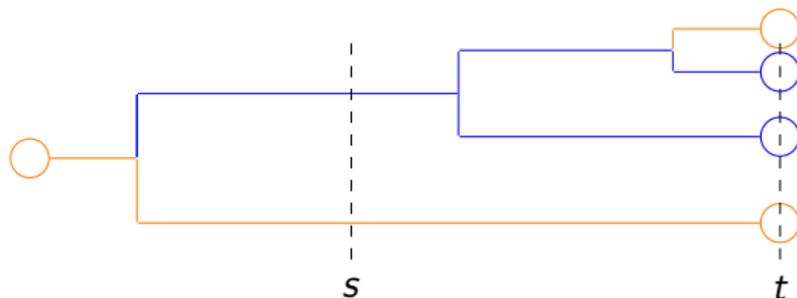
$$\left( X_s^{U(t)} \right)_{s \leq t}$$

## Trait d'un individu échantillonné

On note  $U(t)$  la variable aléatoire correspondant à un individu échantillonné dans la population au temps  $t$ . On cherche à décrire la loi de

$$\left( X_s^{U(t)} \right)_{s \leq t}$$

Ce processus est différent de celui de la cellule étiquetée.



- Arbre biaisé et épine
  - Chauvin, Rouault (1988), Chauvin, Rouault, Wakolbinger (1991), Gorostiza, Roelly, Wakolbinger (1991) : arbre biaisé par la taille,
  - Lyons, Peres, Pemantle (1995) : théorème de Kesten-Stigum,
  - Kurtz, Lyons, Peres Pemantle (1997), Athreya (2000) : extensions multitypes,
  - Georgii et Baake (2003) : étude du trait des ancêtres d'un individu typique dans un cadre spectral,
  - Hardy et Harris (2006) : nouvelle formalisation de l'épine

- Arbre biaisé et épine
  - Chauvin, Rouault (1988), Chauvin, Rouault, Wakolbinger (1991), Gorostiza, Roelly, Wakolbinger (1991) : arbre biaisé par la taille,
  - Lyons, Peres, Pemantle (1995) : théorème de Kesten-Stigum,
  - Kurtz, Lyons, Peres Pemantle (1997), Athreya (2000) : extensions multitypes,
  - Georgii et Baake (2003) : étude du trait des ancêtres d'un individu typique dans un cadre spectral,
  - Hardy et Harris (2006) : nouvelle formalisation de l'épine
- Approche avec un processus auxiliaire ergodique
  - Bansaye, Delmas, Marsalle, Tran (2011) : modèle continu dans le cas neutre, loi des grands nombres,
  - Cloez (2011) : extension au cas non-neutre du modèle de Bansaye et al. à l'aide de la théorie spectrale.

# Formule Many-to-One

## Théorème

Pour tout  $t > 0$ ,  $x \in \mathcal{X}$ , pour toute fonction mesurable positive  $F : \mathbb{D}([0, t], \mathcal{X}) \rightarrow \mathbb{R}$  on a :

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

avec  $(Y_s^{(t)})_{s \leq t}$  processus de Markov inhomogène en temps de générateurs infinitésimaux  $(\mathcal{A}_s^{(t)})_{s \leq t}$  donnés par :

$$\mathcal{A}_s^{(t)} f(x) = \widehat{\mathcal{G}}_{s,t} f(x) + \widehat{B}_{s,t}(x) \left( \int_{\mathcal{X}} (f(y) - f(x)) \widehat{P}_{s,t}(x, dy) \right).$$

## Le processus auxiliaire

$$m(x, dy) = \sum_{k \geq 0} p_k(x) \sum_{j=1}^k P_i^{(k)}(x, dy), \quad m(t, s, x) = \mathbb{E}(N_t | Z_s = \delta_x),$$

## Le processus auxiliaire

$$m(x, dy) = \sum_{k \geq 0} p_k(x) \sum_{j=1}^k P_i^{(k)}(x, dy), \quad m(t, s, x) = \mathbb{E}(N_t | Z_s = \delta_x),$$

- le processus saute à un taux accéléré :

$$\widehat{B}_{s,t}(x) = B(x) \int_{\mathcal{X}} \frac{m(t, s, y)}{m(t, s, x)} m(x, dy),$$

## Le processus auxiliaire

$$m(x, dy) = \sum_{k \geq 0} p_k(x) \sum_{j=1}^k P_i^{(k)}(x, dy), \quad m(t, s, x) = \mathbb{E}(N_t | Z_s = \delta_x),$$

- le processus saute à un taux accéléré :

$$\widehat{B}_{s,t}(x) = B(x) \int_{\mathcal{X}} \frac{m(t, s, y)}{m(t, s, x)} m(x, dy),$$

- le trait de l'unique descendant est choisi selon le noyau de probabilité :

$$\widehat{P}_{s,t}(x, dy) = \frac{m(t, s, y)}{m(t, s, x)} m(x, dy) \left( \int_{\mathcal{X}} \frac{m(t, s, y)}{m(t, s, x)} m(x, dy) \right)^{-1}.$$

## Le processus auxiliaire

$$m(x, dy) = \sum_{k \geq 0} p_k(x) \sum_{j=1}^k P_i^{(k)}(x, dy), \quad m(t, s, x) = \mathbb{E}(N_t | Z_s = \delta_x),$$

- le processus saute à un taux accéléré :

$$\widehat{B}_{s,t}(x) = B(x) \int_{\mathcal{X}} \frac{m(t, s, y)}{m(t, s, x)} m(x, dy),$$

- le trait de l'unique descendant est choisi selon le noyau de probabilité :

$$\widehat{P}_{s,t}(x, dy) = \frac{m(t, s, y)}{m(t, s, x)} m(x, dy) \left( \int_{\mathcal{X}} \frac{m(t, s, y)}{m(t, s, x)} m(x, dy) \right)^{-1}.$$

- la dynamique du trait est donnée par :

$$\widehat{\mathcal{G}}_{s,t} f(x) = \frac{\mathcal{G}(m(t, s, \cdot) f)(x) - f(x) \mathcal{G}(m(t, s, \cdot))(x)}{m(t, s, x)}$$

## Exemples

- Dans le cas neutre,  $B \equiv b$ , on retrouve les résultats de Bansaye et al. :

$$\mathcal{A}f(x) = \mathcal{G}f(x) + 2b \int_x (f(y) - f(x)) \frac{m(x, dy)}{2}$$

## Exemples

- Dans le cas neutre,  $B \equiv b$ , on retrouve les résultats de Bansaye et al. :

$$\mathcal{A}f(x) = \mathcal{G}f(x) + 2b \int_x (f(y) - f(x)) \frac{m(x, dy)}{2}$$

- Dans le cas d'une croissance des cellules exponentielle à taux  $a$  et d'une division binaire à taux  $B(x) = x$ , on a :

$$m(t, s, x) = 1 + \frac{x}{a} \left( e^{a(t-s)} - 1 \right),$$

et :

$$\mathcal{A}_s^{(t)} f(x) = axf'(x) + x \left( 1 + \frac{1}{1 + \frac{x}{a} (e^{a(t-s)} - 1)} \right) \left( f\left(\frac{x}{2}\right) - f(x) \right).$$

# Approximation en grande population

## Hypothèse C

Il existe  $x \in \mathcal{X}$  tel que pour tout  $t \geq 0$  :  $\mathbb{P}_{\delta_x}(N(t) > 0) > 0$ .

Soit  $K$  un compact inclus dans  $\mathcal{X}$  tel que pour tout  $t \geq 0$  :

$$\inf_{x \in K} \mathbb{P}_{\delta_x}(N_t > 0) > 0$$

## Théorème

Sous les hypothèses A, B et C, si  $\nu_n = \sum_{i=1}^n \delta_{X_i}$ , avec  $(X_i)_{1 \leq i \leq n}$  i.i.d de loi  $\nu$  à support dans  $K$  on a en loi dans  $\mathbb{D}([0, t], \mathcal{X})$  :

$$X_{[0,t]}^{U(t), \nu_n} \mathbf{1}_{A_n^\nu(t)} \xrightarrow{n \rightarrow +\infty} Y_{[0,t]}^{(t), \pi_t}, \text{ avec } \pi_t(dx) = \frac{m(t, 0, x)\nu(dx)}{m(t, 0, \nu)}.$$

## Perspectives

- Estimation statistique du taux de division,
- Application à un modèle de switch génétique,
- Application à un modèle de vieillissement cellulaire.