Dynamique des populations : algorithme de Wright-Fisher

Louis Bouvier, Régis Santet

École Nationale des Ponts et Chaussées

Introduction

Objectifs du projet :

- Compréhension du modèle biologique proposé par le modèle de Wright-Fisher : évolution de la proportion des allèles d'un gène au fur et à mesure des générations
- Compréhension des outils mathématiques nécessaires à son étude :
  - martingales
  - chaînes de Markov
- Illustration numérique des résultats théoriques de [1, 2]
- Familiarisation avec les limites diffusives

Modèle de Wright-Fisher : cadre théorique

On se place sous les hypothèses formulées dans [1] :

- La population est de taille finie constante, hermaphrodite
- Les générations sont séparées
- Il n'y a pas de migration
- Les unions sont indépendantes du caractère étudié

Nous pouvons alors mettre en place le modèle :

- Une première génération est créée de taille N et nous fixons une proportion initiale d'une version du gène possible -allèle notée **A**, l'autre étant notée **B**. Cela donne une proportion $X_0 := \frac{\#A}{N}$ où $\#A$ est le nombre d'invidus portant l'allèle **A**.
- Pour une génération n ayant comme proportion X_n de personnes portant l'allèle **A**, une génération $n+1$ de même taille est créée où chaque personne a une probabilité $\Psi(X_n)$ d'avoir l'allèle **A**, où Ψ est une fonction permettant d'inclure des phénomènes de *mutations* et de *sélection*. Nous obtenons alors une proportion X_{n+1}.

Les variables aléatoires $(X_n)_{n\in\mathbb{N}}$ forment une chaîne de Markov.

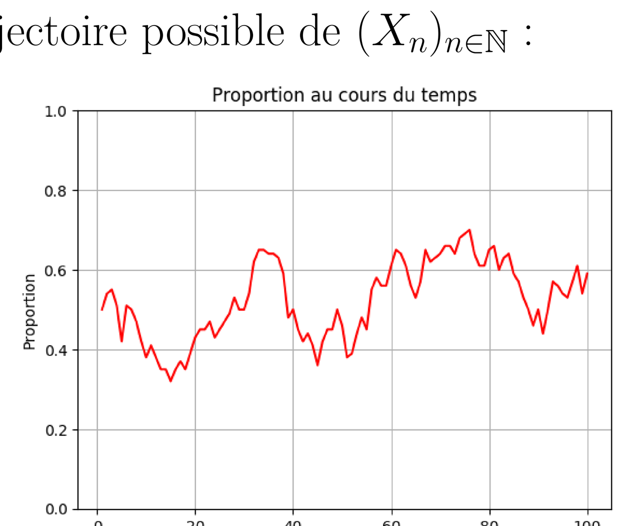
Il y a *fixation* à la génération n lorsque $X_n \in \{0,1\}$ (disparition de l'allèle A ou B respectivement). S'il n'y a pas de mutations, elle est alors constante après ce temps n, ces états étant absorbants.

Les phénomènes de *sélection* et *mutations* sont introduits mathématiquement de la façon suivante :

- Il y a *sélection* de paramètre $s \in \mathbb{R}^+$ lorsque
$$\Psi(X_n) = \frac{X_n \times (1+s)}{X_n \times (1+s) + 1 - X_n}$$
- Nous parlons de *mutations* de paramètres $u \in [0,1]$ et $v \in [0,1]$ lorsque
$$\Psi(X_n) = X_n \times (1-u) + (1-X_n) \times v$$

Cas simple sans mutation ni sélection

Dans ce cas, nous avons $\Psi = id$ et $(X_n)_{n\in\mathbb{N}}$ est une *martingale*.
Voici une trajectoire possible de $(X_n)_{n\in\mathbb{N}}$:



$N = 100,\ X_0 = 0.5$

$(X_n)_{n\in\mathbb{N}}$ a pour support $\{0, \frac{1}{N}, \dots, 1\}$. Les états 0 et 1 sont absorbants, on peut définir :

$$\tau = \inf\{n \in \mathbb{N}, X_n \in \{0,1\}\}$$

Nous avons :

$$\mathbb{P}(\tau < +\infty) = 1 \quad (1)$$

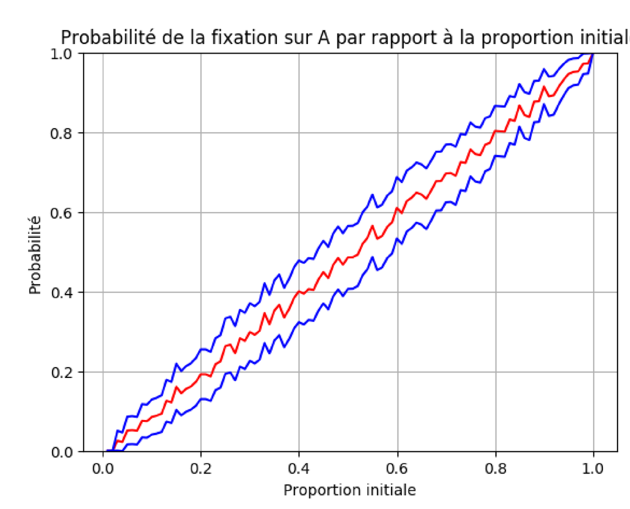
En attendant donc suffisamment longtemps, il y aura fixation.

- Il faut choisir un temps final (production de générations) assez grand dans nos simulations !

Dans ce cas simple, nous pouvons même prédire la probabilité de fixation sur l'allèle **A** : $(X_n)_{n\in\mathbb{N}}$ converge *presque sûrement* vers X_τ et nous avons donc :

$$\mathbb{P}_x(X_\tau = 1) = x \quad (2)$$

Nous ne dépendons donc que de la condition initiale, ce qui est illustré par la figure suivante, qui montre une tendance linéaire :



$N = 40,\ T_{\text{final}} = 3000,\ N_{\text{itérations}} = 2000$

Limite diffusive

Nous pouvons obtenir des résultats très intéressants et plus faciles à obtenir que dans le cas discret en passant au processus continu.

- Hypothèse : (Ns, Nu, Nv) converge vers $(\alpha, \beta_u, \beta_v) \in (\mathbb{R}^+)^3$ quand $N \longrightarrow +\infty$
- On prendra typiquement N grand et un facteur de sélection dépendant de N : $s_N \sim \frac{\alpha}{N}$ avec α fixé.

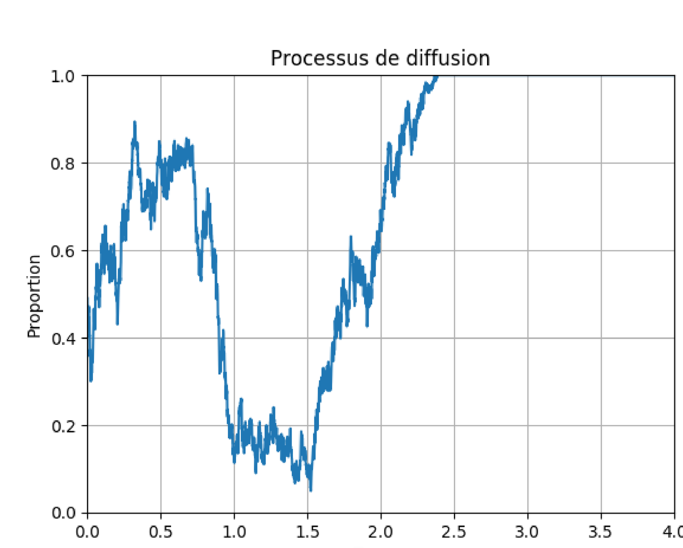
Nous notons X_n^N la proportion de la génération n de taille N ayant l'allèle **A** et nous posons

$$Z_t^N := X_{\lfloor Nt \rfloor}^N$$

- La suite $(Z_\cdot^N)_{N>0}$ converge en loi vers un processus de diffusion $(Z_t)_{t\in\mathbb{R}^+}$.
- Le processus de diffusion est caractérisé par l'équation différentielle stochastique :

$$dZ_t = \sigma(Z_t)dB_t + b(Z_t)dt \quad (3)$$

avec $\sigma^2(z) = z(1-z)$, $b(z) = \alpha z(1-z) - \beta_A z + \beta_B(1-z)$



$N = 4000,\ t \in [0,4]$

Dans le cas sans mutation ni sélection, soit $(\alpha, \beta_u, \beta_v) = (0,0,0)$, le processus de diffusion est à nouveau une martingale, et en définissant

$$T_{0,1} = \inf\{t \in \mathbb{R}^+, Z_t \in \{0,1\}\} \quad (4)$$

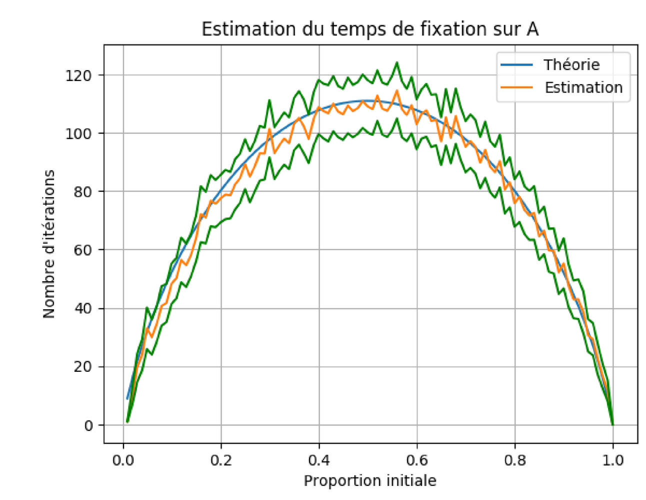
Nous avons alors le résultat similaire au cas simple étudié précédemment :

$$\mathbb{P}_x(Z_{T_{0,1}} = 1) = x \quad (5)$$

De plus, nous avons la formule explicite pour le temps de fixation donnée par :

$$\mathbb{E}_x(T_{0,1}) = -2[x\ln(x) + (1-x)\ln(1-x)] \quad (6)$$

Dans le cas discret, obtenir une expression analytique aurait été impossible, ce qui prouve la puissance du passage à la limite diffusive pour le modèle de Wright-Fisher.



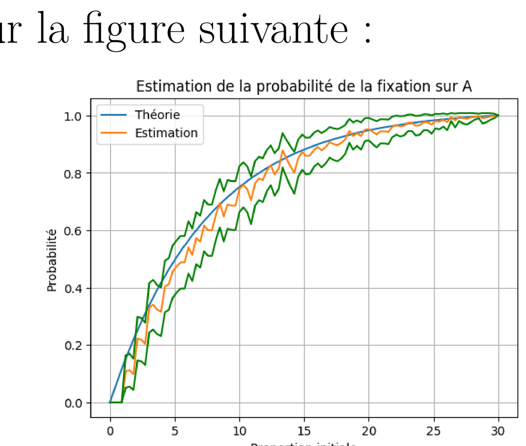
$N = 80$

Limite diffusive avec sélection

Dans le cas où nous ajoutons de la sélection, même un facteur infime influe très grandement sur la fixation d'un allèle. En effet, nous avons dans ce cas [1] :

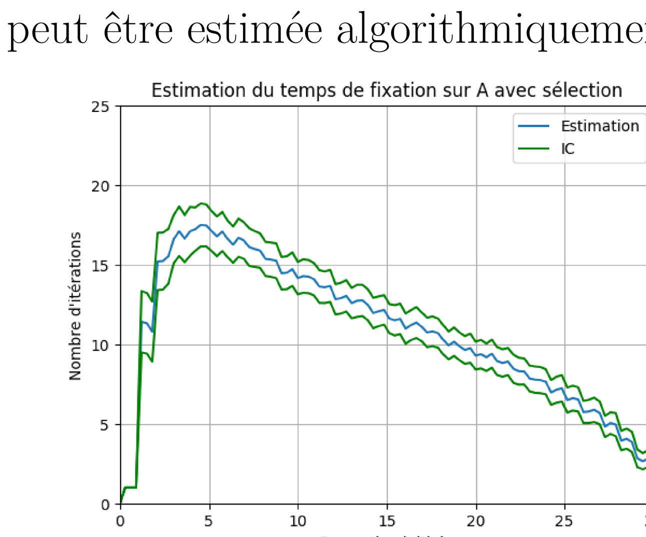
$$\mathbb{P}_x(Z_{T_{0,1}} = 1) = \frac{1 - \exp(-2\alpha x)}{1 - \exp(-2\alpha)} \quad (7)$$

C'est illustré sur la figure suivante :



$N = 30, \alpha = 2$

Avec un facteur de sélection très faible et une population suffisamment grande, par exemple $N = 10^9$ et $s = 2.10^{-8}$, on a $\alpha = 20$ et pour une proportion initiale de $0,5$, nous obtenons une probabilité de fixation égale à 0,99999999 ! Certains facteurs exogènes peuvent donc influencer radicalement sur la disparition d'un allèle !
L'espérance du temps de fixation n'a pas d'expression analytique mais peut être estimée algorithmiquement :



$N = 30,\ \alpha = 10$

Conclusion

Ce projet montre l'intérêt de l'étude des processus aléatoires : à partir d'un modèle simple, nous avons pu tirer de nombreux résultats en termes de loi et de comportement asymptotique. Ce dernier peut être complexifié afin de représenter au mieux la réalité en revenant sur certaines hypothèses (population de taille fixe, indépendance des tirages, etc.). Une ouverture possible serait par exemple de le coupler avec celui de Galton-Watson.

Références

[1] Djalil Chafaï and Florent Malrieu. *Recueil de Modèles Aléatoires*, volume 78 of *Mathématiques et Applications*. Springer Berlin Heidelberg, Berlin, Heidelberg, 2016.

[2] Jean-François Delmas. *Stochastic Processes and Applications*. 2018.

Remerciements

Nous tenons à remercier Pierre-André Zitt, notre professeur de *Processus Stochastiques* et encadrant pour ce projet, avec qui il a été très intéressant de travailler et d'apprendre.