

# Étude de l'évolution de la distribution de la fitness dans un modèle de sélection et mutation

**Stage de M2 Mathématiques pour les Sciences du  
vivant**

Unité de recherche: Mathématiques et informatique appliquées  
(AgroParisTech), UMR 518 INRAE  
Réfèrent: Camille Coron

par

**Taliesin ROMANET**

**Encadrement du stage :**

Camille CORON  
Olivier HENARD

Co-encadrante  
Co-encadrant

# Table des matières

---

<b>Table des matières</b>	<b>i</b>
<b>Remerciements</b>	<b>ii</b>
<b>1 Modèle House of Cards</b>	<b>1</b>
1.1 Définition du Modèle . . . . .	1
1.2 Convergence du Modèle . . . . .	3
<b>2 Modèle de Mutation Locale</b>	<b>5</b>
2.1 Modèle de Mutation Délétère Basé sur la Fitness du Parent . . . . .	5
2.2 Définition du Modèle . . . . .	5
2.3 Modèle Simplifié . . . . .	7
2.3.1 Existence d'une Limite . . . . .	8
2.3.2 Expression de la limite . . . . .	12
2.4 Modèle en dimension finie . . . . .	13
2.4.1 Définition du modèle . . . . .	13
<b>3 Programmation</b>	<b>16</b>
<b>Table des figures</b>	<b>19</b>

# Remerciements

---

Je souhaitais remercier Camille CORON et Olivier HENARD d’avoir encadré mon stage et m’avoir guidé dans mon travail.

De plus, je souhaitais remercier Alexandre PRADIGNAC, Alizée GEOFFROY et Louis LACOSTE de m’avoir conseillé dans mon travail.

Enfin, je remercie toute l’équipe de l’INRAE de m’avoir accueilli au sein de l’équipe.

# Introduction

---

L'évolution des populations, composées de divers types d'individus d'une même espèce, est influencée par deux principaux mécanismes : la sélection naturelle et les mutations génétiques. Comprendre comment ces mécanismes interagissent pour façonner la diversité génétique et la répartition des types d'individus est un objectif central de la génétique des populations.

Pour chaque individu, nous définissons un type  $x$  et une fonction  $f$  telle que  $f(x)$  représente la fitness de cet individu. La fitness, un concept fondamental en génétique des populations, mesure la capacité d'un individu à se reproduire et à transmettre ses gènes à la génération suivante. Elle est généralement calculée en multipliant la viabilité de l'individu par sa fertilité, offrant ainsi une évaluation de son succès reproductif.

Cette étude vise à analyser l'évolution des types d'individus au sein d'une population en considérant la sélection des types et la mutation aléatoire. Un individu engendre un nombre d'enfants proportionnel à sa fitness, et ces enfants héritent initialement de la même fitness que leur parent. Toutefois, il existe une probabilité à chaque naissance que l'enfant subisse une mutation, altérant ainsi sa fitness par rapport à celle de son parent. Cette probabilité de mutation introduit de la variabilité et constitue un moteur important de la diversité génétique.

La sélection naturelle agit en faveur des individus possédant une fitness élevée, tendant ainsi à réduire la diversité génétique en éliminant les types moins adaptés. En revanche, les mutations contribuent à augmenter cette diversité en introduisant de nouvelles variations aléatoires parmi les types des individus mutants. L'interaction entre sélection et mutation est complexe et joue un rôle crucial dans la dynamique évolutive des populations.

Dans le cadre de cette étude, nous considérons des générations non-chevauchantes, ce qui signifie que les individus de la  $n$ -ième génération donnent naissance à ceux de la  $(n + 1)$ -ième génération, sans qu'il y ait d'individu commun entre ces deux générations. Ce modèle simplifié permet de mieux comprendre les mécanismes sous-jacents aux changements de la répartition des types d'une génération à l'autre.

L'objectif de cette recherche est de déterminer l'existence d'un équilibre entre la sélection et la mutation. Autrement dit, nous cherchons à savoir si, sous certaines conditions initiales, la répartition des types au sein de la population converge vers un état stable. La réponse à cette question pourrait apporter des éclaircissements significatifs sur la manière dont les forces évolutives façonnent la diversité génétique à long terme.

# 1

CHAPITRE

## Modèle House of Cards

Le modèle de House of Cards, décrit par Kingman en 1978 ([KINGMAN, 1978](#)), est particulièrement adapté à des populations où la fitness des mutants est indépendante de celle du parent. Ce modèle est souvent utilisé pour décrire des populations telles que celle des drosophiles ([HODGINS-DAVIS et al., 2015](#)). Dans ce contexte, la fitness de chaque individu, notée  $x$ , est un paramètre compris dans l'intervalle  $[0, 1]$ . La fitness est représentée par l'identité  $f(x) = x$ , où les types d'individus varient entre 0 et 1. L'objectif de ce modèle est de suivre l'évolution de la distribution de la fitness au fil du temps dans une population infinie. Ainsi, la distribution de la fitness est modélisée par une mesure de probabilité sur l'intervalle  $[0, 1]$ .

Ce modèle a été décrit en temps continu dans ([CLOEZ et GABRIEL, 2022](#)). Dans notre cas, nous sommes en temps discret.

### 1.1 Définition du Modèle

Dans ce modèle, la mutation est telle qu'un individu mutant possède une fitness aléatoire, indépendante de celle de son parent. Soit  $p_n$  la distribution de la fitness dans la population au temps  $n$ . À partir de cette distribution initiale  $p_0$ , Kingman décrit l'évolution de la distribution de fitness par l'équation de récurrence suivante :

$$p_{n+1}(dx) = \beta q(dx) + (1 - \beta) \frac{x p_n(dx)}{w_n} \quad (1.1)$$

où  $w_n := \int_{[0,1]} y p_n(dy)$ ,  $\beta \in [0, 1]$  est la probabilité de mutation, et  $q$  est une distribution sur  $[0, 1]$  indépendante de toutes les mesures  $p_n$ .

L'équation (1.1) modélise le passage de la génération  $n$  à la génération  $n + 1$ . Selon ce modèle, une fraction  $\beta$  de la population mute, et la fitness des individus mutants suit la distribution  $q$ . L'autre fraction,  $1 - \beta$ , est sélectionnée de manière à ce que les individus aient un nombre d'enfants proportionnel à leur fitness ( $x p_n(dx)$ ).  $w_n$  est une normalisation assurant que si  $p_n$  est une mesure de probabilité, alors  $p_{n+1}$  en est également une.

Les trois paramètres du modèle sont :

- $p_0$  : la distribution de la fitness à la génération initiale.
- $\beta \in [0, 1]$  : la probabilité de mutation.

- $q$  : la distribution de la fitness en cas de mutation.

En simulant les distributions successives, il est possible d'observer l'apparition ou non d'une condensation. Par exemple, la figure (1.1) montre que, bien que les deux distributions de mutation soient presque identiques, les limites peuvent différer par la présence ou l'absence d'un Dirac.

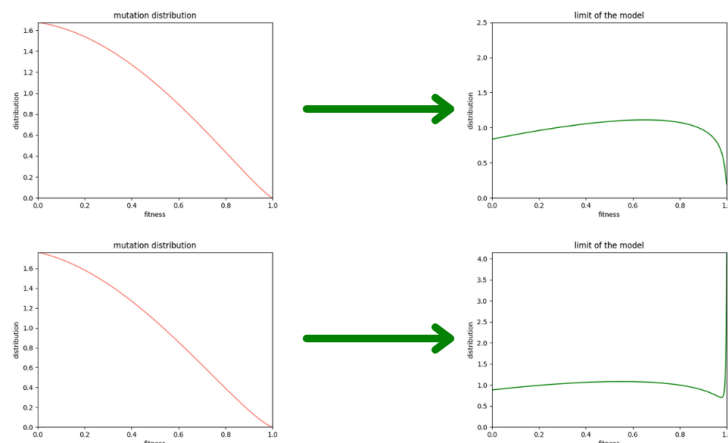


FIGURE 1.1 : Présence ou absence de condensation dans la limite en fonction de la distribution de mutation  $q$ .

Dans la figure (1.1), pour  $\beta = 0.5$  et  $p_0$  uniforme sur  $[0, 1]$  (c'est-à-dire  $p_0(dx) = dx$ ), nous examinons deux distributions de mutation presque identiques qui mènent à des limites différentes :

- La première distribution est  $q(dx) = C(1-x)^{1.3}(1+x)dx$ , avec  $C$  normalisant  $q$  en une mesure de probabilité. On observe une condensation en  $x = 1$ .
- La deuxième distribution est  $q(dx) = C(1-x)^{1.2}(1+x)dx$ , avec  $C$  également normalisant  $q$ . Il n'y a pas de condensation dans la limite.

La condensation ici est une répartition de la fitness sur un point. En effet, dans le deuxième cas nous avons dans la limite une proportion non nulle de la fitness qui vaut 1. Cette condensation s'explique biologiquement par le fait que les individus les plus avantageux représentent asymptotiquement une proportion positive de la population.

De plus, la variation de  $p_0$  influe sur l'emplacement de la condensation dans la limite. Pour  $\beta = 0.5$  et  $q(dx) = 2\mathbb{1}_{\{x \leq 0.5\}}dx$ , en faisant varier  $p_0$ , nous observons :

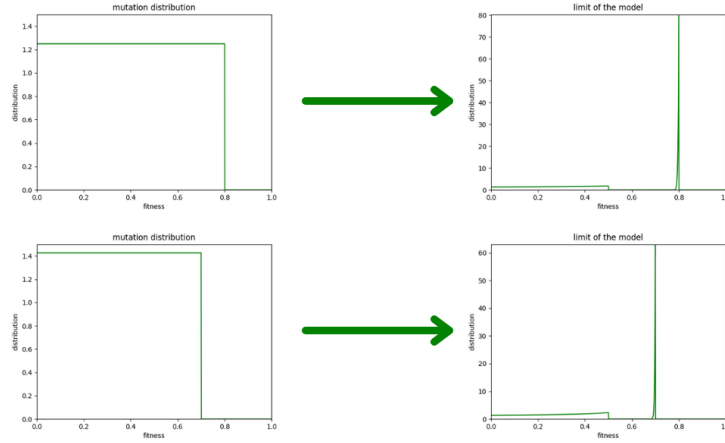


FIGURE 1.2 : Emplacement de la condensation dans la limite en fonction de  $p_0$ .

Dans la figure (1.2), les choix de  $p_0$  sont respectivement  $p_0(dx) = \frac{\mathbf{1}_{\{x \leq 0.8\}} dx}{0.8}$  et  $p_0(dx) = \frac{\mathbf{1}_{\{x \leq 0.7\}} dx}{0.7}$ . Nous pouvons observer que, selon le maximum du support de  $p_0$ , la condensation dans la limite ne se situe pas au même endroit.

La suite de ce chapitre traite de l'apparition de la condensation dans la limite et de son emplacement.

## 1.2 Convergence du Modèle

Définissons  $\eta_0 := \max(\text{Supp}(p_0))$  et  $\eta_q := \max(\text{Supp}(q))$ . Sans perte de généralité, on suppose  $\eta_0 \geq \eta_q$ . En effet, dans le cas contraire, d'après l'équation (1.1), le maximum du support de  $p_1$  sera égal à  $\eta_q$  donc l'hypothèse sera vérifiée à partir du rang 1.

Nous introduisons également la quantité suivante :

$$y_0 := \inf \left\{ x' \in [\eta_0, \infty) : \int_0^{\eta_q} \frac{\beta q(dx)}{1 - \frac{x}{x'}} \leq 1 \right\}$$

**Théorème 1.2.1.** *La séquence  $(p_n)_{n \in \mathbb{N}}$  converge vers la mesure de probabilité  $\pi$  définie par :*

$$\pi(dx) = \frac{\beta q(dx)}{1 - \frac{x}{y_0}} + \pi_0 \delta_{\eta_0} \quad (1.2)$$

où  $\pi_0$  est donné par :

$$\pi_0 = 1 - \int_0^{\eta_q} \frac{\beta q(dx)}{1 - \frac{x}{y_0}}$$

Dans l'expression (1.2), nous observons la continuité sur le support de  $q$  et un éventuel Dirac sur le maximum du support de  $p_0$ .

La fonction  $f : y \mapsto \int_0^{\eta_q} \frac{\beta q(dx)}{1 - \frac{x}{y}}$  étant décroissante et continue, si  $f(\eta_0) \geq 1$ , alors  $f(y_0) = 1$  et donc  $\pi_0 = 0$  et il n'y aura pas de Dirac dans l'expression de la limite.

Ainsi, la condition nécessaire et suffisante pour l'apparition d'une condensation est :

$$\int_0^{\eta_q} \frac{\beta q(dx)}{1 - \frac{x}{\eta_0}} < 1 \quad (1.3)$$

Et la condensation aura lieu en  $\eta_0$ .



# 2

CHAPITRE

## Modèle de Mutation Locale

### 2.1 Modèle de Mutation Délétère Basé sur la Fitness du Parent

Dans cette section, nous examinons un modèle prenant en compte la fitness du parent lors de la mutation. Plus précisément, la fitness d'un mutant est aléatoire mais reste proche de celle de son parent. Ce modèle, ainsi que ses résultats, sont nouveaux. C'est pourquoi ils peuvent être incomplets.

Dans ce modèle, chaque individu possède un type  $x \in \mathbb{R}^+$ . La fitness  $s$  d'un individu est déterminée par la fonction  $s = f(x) = e^{-x}$ . Ainsi, la sélection favorise les individus dont le type est proche de 0. Pour équilibrer la mutation et la sélection, la mutation est délétère, c'est-à-dire qu'elle réduit la fitness en augmentant le type de l'individu mutant par rapport à son parent.

Pour modéliser cette mutation, nous introduisons une variable aléatoire positive de densité  $j$ . Cette variable aléatoire est ajoutée au type du parent, ce qui permet d'obtenir une mutation aléatoire tout en restant dépendante du parent. La convolution entre la densité de l'individu parent et la densité de la variable aléatoire modélise cette somme.

Ce modèle a été décrit en temps continu dans (GIL et al., 2016), ici comme précédemment, nous sommes en temps discret.

### 2.2 Définition du Modèle

Lors d'une mutation, une variable aléatoire positive est ajoutée au type du parent, entraînant une diminution de la fitness pour l'individu mutant par rapport à son parent.

Soit  $j$  la fonction de densité de la variable aléatoire ajoutée au type du parent lors d'une mutation. Pour intégrer cette variable aléatoire au type du parent, nous utilisons la convolution  $*$ . La formule de récurrence pour passer de la génération  $n$  à la génération  $n + 1$  est donnée par :

$$p_{n+1}(dx) = \beta p_n * j(dx) + (1 - \beta) \frac{e^{-x} p_n(dx)}{w_n} \quad (2.1)$$

où  $w_n := \int_{\mathbb{R}^+} e^{-x} p_n(dx)$  est la normalisation de la distribution de fitness afin de garantir que  $p_{n+1}$  est bien une mesure de probabilité.

**Remarque.** Si  $p_0(\{0\}) > 0$  alors  $\forall n, p_n(\{0\}) > 0$  et on a :

$$\begin{aligned} p_{n+1}(\{0\}) &= (1 - \beta) \frac{e^0 p_n(\{0\})}{w_n} \\ &= (1 - \beta) \frac{p_n(\{0\})}{w_n} \end{aligned}$$

d'où :

$$w_n = (1 - \beta) \frac{p_n(\{0\})}{p_{n+1}(\{0\})} \quad (2.2)$$

La Figure (2.1) illustre les distributions limites simulées du modèle selon les itérations décrites par (2.1) avec différentes fonctions de densité  $j$ .

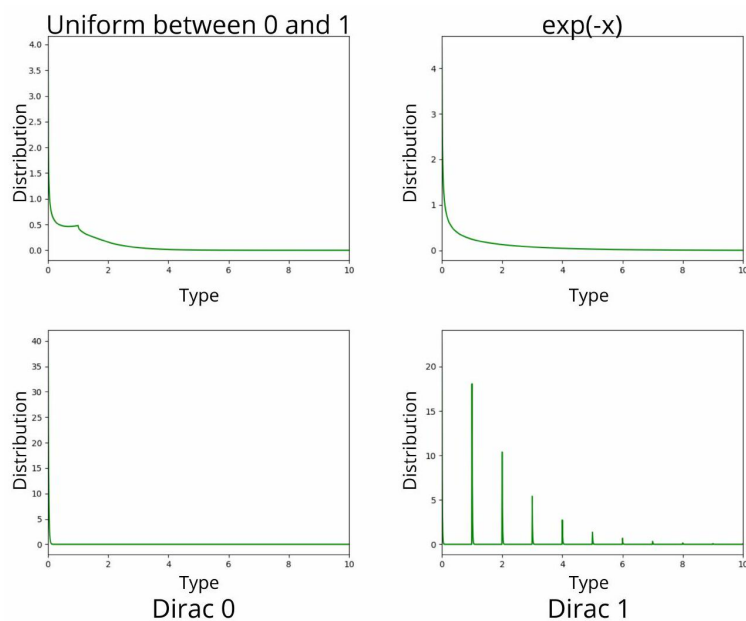


FIGURE 2.1 : Distribution limite de la fitness avec respectivement  $j(dx) = \mathbb{1}_{\{0 \leq x \leq 1\}} dx$ ,  $j(dx) = \frac{e^{-x} \mathbb{1}_{\{0 \leq x \leq 10\}}}{\int_0^{10} e^{-y} dy} dx$ ,  $j(dx) = \delta_0(dx)$  et  $j(dx) = \delta_1(dx)$ .

Le but ici est de déterminer s'il existe une distribution limite suivant l'algorithme et si possible de déterminer cette distribution en fonction de  $j$ . Pour simplifier, nous allons nous restreindre au cas où  $j(dx) = \delta_1(dx)$ .

## 2.3 Modèle Simplifié

Nous considérons le cas où  $j = \delta_1$ , ce qui correspond à la densité d'une variable aléatoire qui vaut 1 presque sûrement. Cela signifie qu'à chaque mutation, le type de l'enfant est le type du parent augmenté de 1.

L'équation (2.1) se simplifie alors en :

$$p_{n+1}(dx) = \beta p_n * \delta_1(dx) + (1 - \beta) \frac{e^{-x} p_n(dx)}{w_n} \quad (2.3)$$

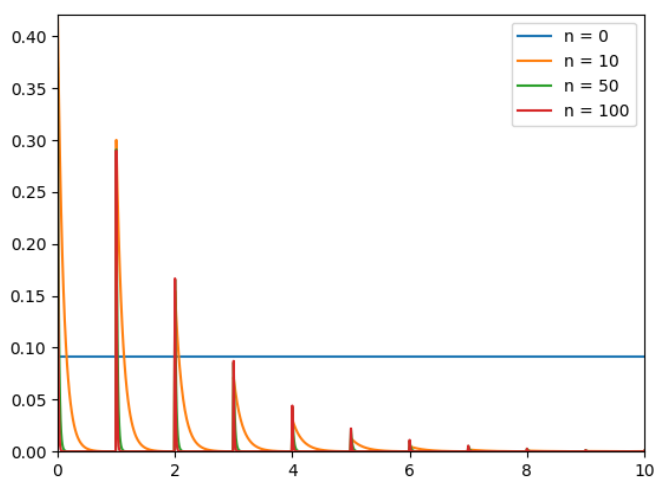


FIGURE 2.2 : Convergence du modèle pour  $\beta = 0.5$ .

La Figure (2.2) révèle la convergence du modèle simplifié par simulation. En effet, le graphique affiche la distribution  $p_n$  pour  $n = 0, 10, 50$  et  $100$  avec une renormalisation afin de pouvoir comparer les itérations entre elles. On remarque une convergence vers une somme de Diracs. La Figure (2.3) montre la limite simulée du modèle simplifié pour différentes valeurs de  $\beta$ .

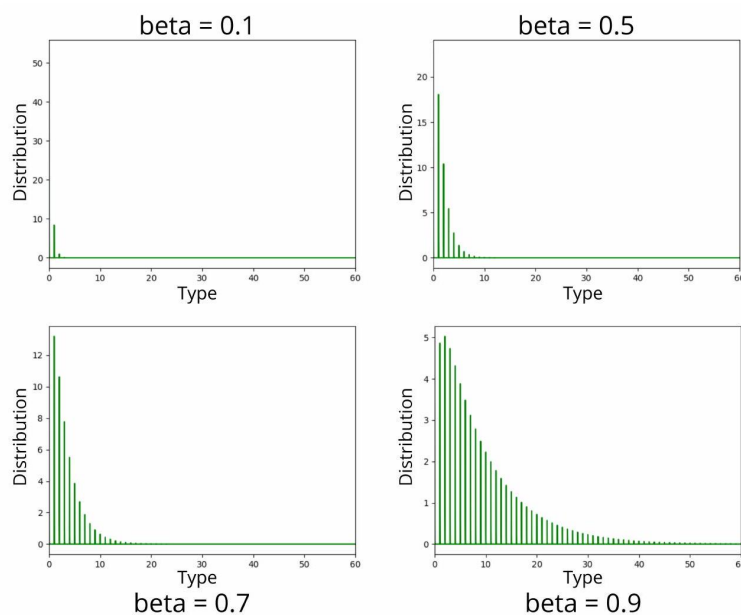


FIGURE 2.3 : Limites du modèle simplifié pour  $\beta = 0.1, 0.5, 0.7$  et  $0.9$ .

La Figure (2.3) révèle que les limites varient mais semblent toutes converger vers une mesure constituée de sommes de Dirac. Lorsque  $\beta$  augmente, la diversité biologique dans la population limite augmente. Il n'y a pas nécessairement une décroissance uniforme des Dirac dans la mesure limite ; par exemple, pour  $\beta = 0.9$ , la limite montre une évolution croissante puis décroissante en termes de Dirac.

### 2.3.1 Existence d'une Limite

Nous cherchons à déterminer si la suite  $(p_n)_{n \in \mathbb{N}}$  converge, et si oui, à identifier la limite.

Considérons la transformée de Laplace de  $p_n$  et notons-la  $\varphi_n$ . Pour tout  $\lambda \in \mathbb{R}^+$ ,  $\varphi_n(\lambda)$  est définie par :

$$\varphi_n(\lambda) = \int_{\mathbb{R}^+} e^{-\lambda x} p_n(dx) \quad (2.4)$$

**Remarque.** On a  $\varphi_n(1) = \int_{\mathbb{R}^+} e^{-x} p_n(dx) = w_n$ .

Grâce à l'équation (2.3), nous pouvons déterminer une formule de récurrence pour cette transformée de Laplace :

$$\begin{aligned}
\varphi_{n+1}(\lambda) &= \int_{\mathbb{R}^+} e^{-\lambda x} p_{n+1}(dx) \\
&= \beta \int_{\mathbb{R}^+} e^{-\lambda x} (p_n * j)(dx) + \frac{1-\beta}{w_n} \int_{\mathbb{R}^+} e^{-\lambda x} e^{-x} p_n(dx) \\
&= \beta \int_{\mathbb{R}^+} e^{-\lambda(x+1)} p_n(dx) + \frac{1-\beta}{\varphi_n(1)} \int_{\mathbb{R}^+} e^{-(\lambda+1)x} p_n(dx) \\
&= \beta e^{-\lambda} \varphi_n(\lambda) + (1-\beta) \frac{\varphi_n(\lambda+1)}{\varphi_n(1)}
\end{aligned}$$

Ainsi, pour tout  $n$  et pour tout  $\lambda \in \mathbb{R}^+$ , nous avons :

$$\varphi_n(\lambda+1) = (\varphi_{n+1}(\lambda) - \beta e^{-\lambda} \varphi_n(\lambda)) \frac{\varphi_n(1)}{1-\beta} \quad (2.5)$$

**Proposition 2.3.1.** Si  $(\varphi_n(1))_{n \in \mathbb{N}}$  converge alors  $\forall \lambda \in \mathbb{N}$ ,  $(\varphi_n(\lambda))_{n \in \mathbb{N}}$  converge également.

*Démonstration.* Supposons que  $\varphi_n(1)$  converge. Supposons que pour un certain  $\lambda \in \mathbb{N}$  fixé,  $\varphi_n(\lambda)$  converge. Alors d'après l'équation (2.5),  $\varphi_n(\lambda+1)$  converge. Donc si  $\varphi_n(1)$  converge alors  $\varphi_n(\lambda)$  converge pour tout entier  $\lambda$ .  $\square$

Or, si  $\varphi_n(\lambda)$  converge pour tout  $\lambda$ , alors en posant  $X_n$  la variable aléatoire de loi  $p_n$ . On a que  $\mathbb{E}[e^{-\lambda X_n}]$  converge vers un  $\mathbb{E}[e^{-\lambda X}]$  et que donc les  $X_n$  convergent en loi vers  $X$ . Et enfin,  $p_n$  converge.

**Proposition 2.3.2.** Pour  $p_0(\{0\}) > 0$ , si la suite  $(p_n)_{n \in \mathbb{N}}$  converge, alors la suite  $(w_n)_{n \in \mathbb{N}}$  converge vers  $1-\beta$ .

*Démonstration.* Supposons que  $(p_n)_{n \in \mathbb{N}}$  converge. Alors  $(p_n(\{0\}))_{n \in \mathbb{N}} \in \mathbb{R}^{\mathbb{N}}$  converge.

Et d'après l'équation (2.2), si  $p_n(\{0\})$  converge alors  $w_n$  converge également et vers  $1-\beta$ .  $\square$

**Proposition 2.3.3.** Pour  $p_0(\{0\}) > 0$ ,  $\forall A \subset \mathbb{R}_+ \setminus \mathbb{N}$ ,  $\lim_{n \rightarrow \infty} \frac{p_n(A)}{p_n(\{0\})} = 0$ . C'est-à-dire que toute la masse entre les entiers va disparaître avec le temps.

*Démonstration.* Pour commencer, nous savons que  $\forall \epsilon \in ]0; 1[, \exists 0 < \epsilon' < 1$ , tel que  $\forall x \in [\epsilon, +\infty[, e^{-x} < 1 - \epsilon' \in ]0; 1[$  et  $\forall n \in \mathbb{N}, w_n \leq 1$

Commençons avec  $A = ]\epsilon; 1[$ .

$$\begin{aligned}
\frac{p_{n+1}(A)}{p_{n+1}(\{0\})} &= \frac{\beta \int_A p_n * j(dx) + \frac{1-\beta}{w_n} \int_A e^{-x} p_n(dx)}{\beta p_n * j(\{0\}) + \frac{1-\beta}{w_n} p_n(\{0\})} \\
&= \frac{\frac{1-\beta}{w_n} \int_A e^{-x} p_n(dx)}{\frac{1-\beta}{w_n} p_n(\{0\})} \\
&= \frac{\frac{1-\beta}{w_n} \int_A e^{-x} p_n(dx)}{\frac{1-\beta}{w_n} p_n(\{0\})} \\
&= \frac{\int_A e^{-x} p_n(dx)}{p_n(\{0\})} \\
&\leq \frac{\int_A (1 - \epsilon') p_n(dx)}{p_n(\{0\})} \\
&= (1 - \epsilon') \frac{p_n(A)}{p_n(\{0\})}
\end{aligned}$$

Et donc

$$\frac{p_n(A)}{p_n(\{0\})} \leq (1 - \epsilon')^n \frac{p_0(A)}{p_0(\{0\})} \xrightarrow{n \rightarrow \infty} 0$$

Maintenant posons  $A_m = ]m + \epsilon; m + 1[$  et supposons que pour un  $m \in \mathbb{N}$ ,  $\lim_{n \rightarrow \infty} \frac{p_n(A_m)}{p_n(\{0\})} = 0$ .

$$\forall \delta' > 0, \exists k_0 \in \mathbb{N} \text{ tel que } \forall k \geq k_0, \frac{p_k(A_m)}{p_k(\{0\})} \leq \delta' \quad (2.6)$$

Comme  $x \mapsto x \frac{\epsilon'(1-\beta)}{\beta(1-\epsilon')}$  est une bijection, il existe  $\delta > 0$  tel que  $\delta' = \delta \frac{\epsilon'(1-\beta)}{\beta(1-\epsilon')}$ .  
Donc nous avons

$$\forall \delta > 0, \exists k_0 \in \mathbb{N} \text{ tel que } \forall k \geq k_0, \frac{p_k(A_m)}{p_k(\{0\})} \leq \delta \frac{\epsilon'(1-\beta)}{\beta(1-\epsilon')}$$

Maintenant, on calcule pour  $m + 1$  :

$$\begin{aligned}
\frac{p_{n+1}(A_{m+1})}{p_{n+1}(\{0\})} &= \frac{\beta \int_{A_{m+1}} p_n * j(dx) + \frac{1-\beta}{w_n} \int_{A_{m+1}} e^{-x} p_n(dx)}{\frac{1-\beta}{w_n} p_n(\{0\})} \\
&= \frac{\beta w_n}{1 - \beta} \frac{\int_{A_m} p_n(dx)}{p_n(\{0\})} + \frac{\int_{A_{m+1}} e^{-x} p_n(dx)}{p_n(\{0\})} \\
&\leq \frac{\beta w_n}{1 - \beta} \frac{p_n(A_m)}{p_n(\{0\})} + (1 - \epsilon') \frac{p_n(A_{m+1})}{p_n(\{0\})}
\end{aligned}$$

Et donc par récurrence,

$$\frac{p_n(A_{m+1})}{p_n(\{0\})} \leq (1 - \epsilon')^n \frac{p_0(A_{m+1})}{p_0(\{0\})} + \sum_{k=1}^n (1 - \epsilon')^{k-1} \frac{\beta w_{n-k} p_{n-k}(A_m)}{1 - \beta} \frac{p_{n-k}(\{0\})}{p_{n-k}(\{0\})}$$

En prenant que le deuxième terme :

$$\begin{aligned} & \sum_{k=1}^n (1 - \epsilon')^{k-1} \frac{\beta w_{n-k} p_{n-k}(A_m)}{1 - \beta} \frac{p_{n-k}(\{0\})}{p_{n-k}(\{0\})} \\ & \leq \frac{\beta}{1 - \beta} \sum_{k=0}^{n-1} (1 - \epsilon')^{n-k-1} \frac{p_k(A_m)}{p_k(\{0\})} \\ & = \frac{\beta}{1 - \beta} \sum_{k=0}^{k_0-1} (1 - \epsilon')^{n-k-1} \frac{p_k(A_m)}{p_k(\{0\})} + \frac{\beta}{1 - \beta} \sum_{k=k_0}^{n-1} (1 - \epsilon')^{n-k-1} \frac{p_k(A_m)}{p_k(\{0\})} \\ & \leq (1 - \epsilon')^n \frac{\beta}{1 - \beta} \sum_{k=0}^{k_0-1} (1 - \epsilon')^{-k-1} \frac{p_k(A_m)}{p_k(\{0\})} + \frac{\beta}{1 - \beta} \sum_{k=k_0}^{n-1} (1 - \epsilon')^{n-k-1} \delta \frac{\epsilon'(1 - \beta)}{\beta(1 - \epsilon')} \quad \text{d'après (2.6)} \\ & \leq (1 - \epsilon')^n \frac{\beta}{1 - \beta} C_{k_0} + \delta \frac{\epsilon'}{(1 - \epsilon')} \sum_{k=k_0}^{n-1} (1 - \epsilon')^{n-k-1} \\ & = (1 - \epsilon')^n \frac{\beta}{1 - \beta} C_{k_0} + \delta \frac{\epsilon'}{(1 - \epsilon')} (1 - \epsilon')^{n-1-k_0} \sum_{k=0}^{n-1-k_0} (1 - \epsilon')^{-k} \\ & = (1 - \epsilon')^n \frac{\beta}{1 - \beta} C_{k_0} + \delta \epsilon' (1 - \epsilon')^{n-2+k_0} \frac{1 - (1 - \epsilon')^{-n+1+k_0}}{1 - (1 - \epsilon')^{-1}} \\ & = (1 - \epsilon')^n \frac{\beta}{1 - \beta} C_{k_0} + \delta \frac{\epsilon' (1 - \epsilon')^{n-2+k_0} - (1 - \epsilon')^{-1}}{1 - (1 - \epsilon')^{-1}} \\ & \frac{p_n(A_{m+1})}{p_n(\{0\})} \leq (1 - \epsilon')^n \frac{p_0(A_{m+1})}{p_0(\{0\})} + (1 - \epsilon')^n \frac{\beta}{1 - \beta} C_{k_0} + \delta \frac{\epsilon' (1 - \epsilon')^{n-2+k_0} - (1 - \epsilon')^{-1}}{1 - (1 - \epsilon')^{-1}} \\ & \xrightarrow{n \rightarrow \infty} \epsilon' \delta \frac{(1 - \epsilon')^{-1}}{1 - (1 - \epsilon')^{-1}} = \delta \end{aligned}$$

Donc nous avons pour tout  $m, \delta$  et pour  $n$  suffisamment grand,  $\frac{p_n(A_m)}{p_n(\{0\})} \leq \delta$  c'est-à-dire  $\frac{p_n(A_m)}{p_n(\{0\})} \xrightarrow{n \rightarrow \infty} 0$ .

Enfin,  $\frac{p_n(\mathbb{R}_+ \setminus \mathbb{N})}{p_n(\{0\})} = \lim_{\epsilon \rightarrow 0} \sum_{m \geq 0} \frac{p_n(A_m)}{p_n(\{0\})} \xrightarrow{n \rightarrow \infty} 0$ . C'est-à-dire que toute la masse sera située sur les entiers naturels.

□

La proposition (2.3.3) implique que si une limite existe, alors c'est une somme de Dirac sur  $\mathbb{N}$ .

### 2.3.2 Expression de la limite

Cette section s'appuie sur l'hypothèse où la limite existe.

**Proposition 2.3.4.** *Supposons que  $(p_n(dx))_{n \in \mathbb{N}}$  converge vers  $\pi(dx) = \sum_{k \in \mathbb{N}} \pi_k \delta_k(dx)$ , alors :*

$$\pi_k = \frac{\beta^k}{\prod_{i=1}^k 1 - e^{-i}} \frac{1}{\sum_{i \geq 0} \frac{\beta^i}{\prod_{j=1}^i 1 - e^{-j}}} \in \mathbb{R}^+ \quad (2.8)$$

*Démonstration.* Admettons qu'une limite  $\pi$  existe pour le modèle simplifié. Alors,  $\pi$  est un point fixe de l'équation (2.3).

$$\pi(dx) = \beta \pi * \delta_1(dx) + (1 - \beta) \frac{e^{-x} \pi(dx)}{w}$$

Or d'après la proposition (2.3.2),  $w_n$  converge vers  $1 - \beta$ . Alors  $w = 1 - \beta$ . Donc

$$\pi(dx) = \beta \pi * \delta_1(dx) + e^{-x} \pi(dx)$$

Ainsi,  $\forall k \in \mathbb{N}^*$ ,

$$\begin{aligned} \pi_k &= \beta \pi_{k-1} + e^{-k} \pi_k \\ \iff \pi_k &= \frac{\beta}{1 - e^{-k}} \pi_{k-1} \end{aligned}$$

Et donc par récurrence,

$$\begin{aligned} \pi_k &= \prod_{j=1}^k \frac{\beta}{1 - e^{-j}} \pi_0 \\ \pi_k &= \frac{\beta^k}{\prod_{j=1}^k 1 - e^{-j}} \pi_0 \end{aligned}$$

Déterminons maintenant  $\pi_0$ . On sait que  $\pi$  est une mesure de probabilité donc son intégrale vaut 1 donc la somme des  $\pi_k$  vaut 1 .



$$\begin{aligned}
1 &= \sum_{k \in \mathbb{N}} \pi_k \\
1 &= \sum_{k \in \mathbb{N}} \frac{\beta^k}{\prod_{j=1}^k (1 - e^{-j})} \pi_0 \\
\iff \pi_0 &= \frac{1}{\sum_{k \in \mathbb{N}} \frac{\beta^k}{\prod_{j=1}^k (1 - e^{-j})}}
\end{aligned}$$

Donc

$$\pi_k = \frac{\frac{\beta^k}{\prod_{i=1}^k (1 - e^{-i})}}{\sum_{l \in \mathbb{N}} \frac{\beta^l}{\prod_{j=1}^l (1 - e^{-j})}}$$

□

## 2.4 Modèle en dimension finie

Ce modèle est également nouveau. Le but de celui-ci est de partir avec un nombre fini d'individu et d'augmenter ce nombre afin d'observer une convergence potentiellement similaire au modèle précédent. Nous modélisons le processus simplifié en dimension finie en utilisant un processus de naissance et de mort en temps discret. Nous commençons avec un individu de type 0. Au temps suivant, un nouvel individu naît. Le choix du parent est aléatoire, chaque individu étant choisi avec une probabilité proportionnelle à sa fitness. Une fois le parent choisi, l'enfant a le type de son parent avec une probabilité de  $1 - \beta$ , ou le type du parent augmenté de 1 avec une probabilité de  $\beta$ .

Par exemple, au temps 1, nous avons un individu unique de type 0 presque sûrement. Au temps 2, un individu de type 0 est choisi comme parent. L'enfant sera de type 0 avec une probabilité de  $1 - \beta$  et de type 1 avec une probabilité de  $\beta$ .

### 2.4.1 Définition du modèle

Soit  $X_n$  le vecteur du nombre d'individus de chaque type au temps  $n$ . Pour tout  $n$ ,  $X_n = (X_{n,0}, X_{n,1}, \dots)$  où  $X_{n,k}$  représente le nombre d'individus de type  $k$  au temps  $n$ . Pour chaque  $k \in \mathbb{N}$ , définissons  $e_k$  comme le vecteur  $(0, 0, \dots, 0, 1, 0, \dots)$  avec un 1 à la  $k$ -ième position.

Nous pouvons alors écrire :

$$X_{n+1} = X_n + e_{K_n} \quad (2.7)$$

où  $(K_n)_{n \in \mathbb{N}}$  est une suite de variables aléatoires telles que

$$\begin{aligned} \mathbb{P}(K_n = 0 \mid X_n) &= (1 - \beta) \frac{X_{n,0}}{w_n} \\ \mathbb{P}(K_n = k \mid X_n) &= (1 - \beta) \frac{X_{n,k} e^{-k}}{w_n} + \beta \frac{X_{n,k-1} e^{-k+1}}{w_n} \quad \text{pour } k > 0 \end{aligned}$$

où  $w_n = \sum_{i \in \mathbb{N}} X_{n,i} e^{-i}$ . Ainsi, un individu de type  $k$  naît au temps  $n + 1$  si la variable aléatoire  $K_n$  vaut  $k$ . Cette naissance peut se produire soit par sélection du type  $k$ , soit par mutation à partir du type  $k - 1$ .

**Remarque.**

$$\sum_{k \in \mathbb{N}} X_{n,k} = n$$

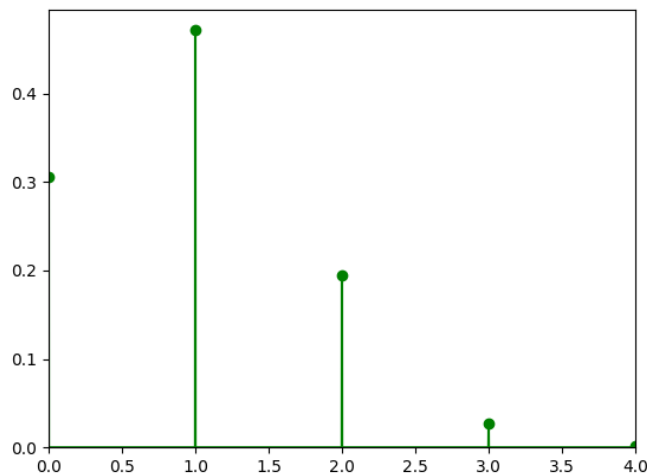


FIGURE 2.4 : Répartition limite simulée du modèle en dimension finie avec  $\beta = 0.5$ .

La figure (2.4) représente la répartition limite  $\frac{X_n}{n}$  simulée par un processus de naissance à temps discret pour  $\beta = 0.5$ . On observe que la répartition limite diffère de celle du modèle simplifié avec  $\beta = 0.5$ , comme montré sur la figure (2.3). Cela est probablement dû au fait que, dans le modèle simplifié, les générations sont disjointes et les individus meurent tous à la fin de leur génération, tandis que dans le modèle en dimension finie, les individus restent en vie.

**Proposition 2.4.1.** Pour tout  $n \in \mathbb{N}^*$  et tout  $k \in \mathbb{N}$ ,

$$\mathbb{E}[X_{n+1,k}] = \mathbb{E}[X_{n,k}] + (1 - \beta)e^{-k} \mathbb{E}\left[\frac{X_{n,k}}{w_n}\right] + \beta e^{-k+1} \mathbb{E}\left[\frac{X_{n,k-1}}{w_n}\right]$$

Nous considérons par convention que pour tout  $k < 0$ ,  $X_{n,k} = 0$  presque sûrement.

*Démonstration.* Soit  $n \in \mathbb{N}^*$  et  $k \in \mathbb{N}$ . Nous avons :

$$\begin{aligned} \mathbb{E}[X_{n+1,k}] &= \mathbb{E}[\mathbb{E}[X_{n+1,k} \mid X_n]] \\ &= \mathbb{E}[X_{n,k} + 1 \cdot \mathbb{P}(K_n = k \mid X_n) + (X_{n,k} - X_{n,k} \cdot \mathbb{P}(K_n = k \mid X_n))] \\ &= \mathbb{E}[X_{n,k} \cdot \mathbb{P}(K_n = k \mid X_n) + \mathbb{P}(K_n = k \mid X_n)] \\ &= \mathbb{E}[X_{n,k}] + (1 - \beta)e^{-k} \mathbb{E}\left[\frac{X_{n,k}}{w_n}\right] + \beta e^{-k+1} \mathbb{E}\left[\frac{X_{n,k-1}}{w_n}\right]. \end{aligned}$$

□

La limite du modèle en dimension finie varie de celle en dimension infinie de part l'absence de mort. Il serait intéressant de déterminer une expression de cette limite pour pouvoir la comparer avec celle du modèle précédent. La simulation des modèles a été mise en avant dans cette partie pour observer des potentielles limites et se donner une idée de leurs expressions. De plus, la programmation a été faite de sorte à pouvoir simuler des modèles généraux. Par exemple, la simulation peut être faite avec n'importe quelle valeur de  $j$ . Dans la suite, nous verrons les difficultés de programmation qu'il y a eu pour simuler les différents modèles.

3  
CHAPITRE

## Programmation

Cette section se concentre sur la mise en œuvre informatique des concepts et objets théoriques abordés précédemment, avec un accent particulier sur les mesures et leurs manipulations.

En informatique, les mesures continues telles que  $\mu(dx) = 2x dx$  sont modélisées comme des fonctions simples. Par exemple, une fonction de mesure continue peut être définie pour calculer des intégrales en utilisant des méthodes numériques, comme la méthode des rectangles, ou pour afficher la fonction.

Cependant, la modélisation des mesures de Dirac pose des défis uniques. En théorie des mesures, un Dirac est une mesure singulière concentrée en un point, mais en pratique, le concept d'infini ne peut être directement manipulé en informatique. En effet, on ne peut définir le Dirac comme une mesure à densité.

Pour contourner cette difficulté, un Dirac est défini en informatique par une paire composée d'un point  $a$  et d'une fonction  $f$ . Ainsi, un Dirac  $\delta_a$  en  $a$  est représenté comme  $f(x)\delta_a(dx)$ . Lorsque l'on intègre ce Dirac, le programme évalue simplement la fonction  $f$  au point  $a$ , car  $\int f(x)\delta_a(dx) = f(a)$ .

Cependant, la situation se complique lorsqu'on considère la somme de plusieurs Dirac. Si l'on combine deux Dirac distincts,  $\delta_a(dx) + \delta_b(dx)$ , la somme des Dirac ne se réduit pas à une simple évaluation de la fonction en un seul point. En effet, l'intégration d'une somme de Dirac nécessite l'évaluation de la fonction en plusieurs points. Pour traiter ce problème, nous définissons une nouvelle classe d'objets appelée somme de Dirac.

Un objet de type somme de Dirac est caractérisé par une liste de Diracs, ainsi que potentiellement une mesure continue supplémentaire. Lors de l'intégration de cet objet, le programme calcule l'intégrale en sommant les intégrales de chaque mesure contenue dans la liste.

Un autre défi rencontré est le temps d'exécution, notamment lorsqu'on additionne des mesures. La définition d'une mesure résultant de la somme de deux mesures implique la somme des fonctions correspondantes. Dans certains algorithmes, cette récursivité peut entraîner des coûts computationnels élevés. Pour optimiser cette opération, nous utilisons une technique appelée mémoïsation.

La mémoïsation consiste à stocker les résultats des calculs dans une structure de données, typiquement un dictionnaire. À chaque fois qu'une valeur de mesure doit être calculée, le programme vérifie d'abord si le résultat a déjà été stocké dans le dictionnaire. Si c'est le cas, le programme récupère le résultat à partir du

dictionnaire. Sinon, le programme calcule la valeur, la stocke dans le dictionnaire, puis la retourne.

Voici un résumé des principaux concepts abordés :

- Mesures continues : Représentées par des fonctions simples permettant le calcul et l’affichage.
- Mesures de Dirac : Définies par une paire (point, fonction) et intégrées en évaluant la fonction en ce point.
- Somme de Dirac : Représentée comme une liste de Diracs, avec une méthode d’intégration par sommation des intégrales.
- Mémoïsation : Utilisation d’un dictionnaire pour stocker les résultats de calculs et éviter les recalculs coûteux.

Ces techniques permettent une gestion efficace des mesures dans les simulations informatiques et optimisent les performances des algorithmes impliqués.

# Bibliographie

---

- KINGMAN, J. F. C. (1978). A Simple Model for the Balance between Selection and Mutation. *Journal of Applied Probability*, 15.
- HODGINS-DAVIS, A., RICE, D. P., & TOWNSEND, J. P. (2015). Gene Expression Evolves under a House-of-Cards Model of Stabilizing Selection.
- CLOEZ, B., & GABRIEL, P. (2022). Fast, Slow Convergence, And Concentration In The House Of Cards Replicator-Mutator Model.
- YUAN, L. (2020). Kingman's Model with Random Mutation Probabilities : Convergence and Condensation.
- GIL, M.-E., HAMEL, F., MARTIN, G., & ROQUES, L. (2016). Mathematical properties of a class of integro-differential models from population genetics.
- DEREICH, S., & MÖRTERS, P. (2013). Emergence of condensation in Kingman's model of selection and mutation.
- CERF, R., & DALMAU, J. (2016). The quasispecies distribution.

# Table des figures

---

1.1	Présence ou absence de condensation dans la limite en fonction de la distribution de mutation $q$ . . . . .	2
1.2	Emplacement de la condensation dans la limite en fonction de $p_0$ . . . . .	3
2.1	Distribution limite de la fitness avec respectivement $j(dx) = \mathbb{1}_{\{0 \leq x \leq 1\}}dx$ , $j(dx) = \frac{e^{-x} \mathbb{1}_{\{0 \leq x \leq 10\}}}{\int_0^{10} e^{-y} dy}dx$ , $j(dx) = \delta_0(dx)$ et $j(dx) = \delta_1(dx)$ . . . . .	6
2.2	Convergence du modèle pour $\beta = 0.5$ . . . . .	7
2.3	Limites du modèle simplifié pour $\beta = 0.1, 0.5, 0.7$ et $0.9$ . . . . .	8
2.4	Répartition limite simulée du modèle en dimension finie avec $\beta = 0.5$ . . . . .	14