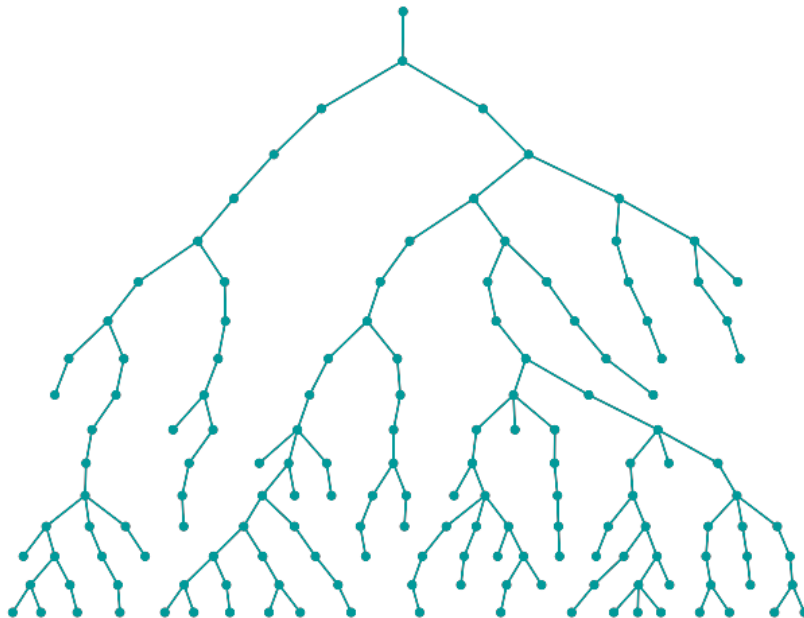


# Modal : Simulation numérique aléatoire autour des événements rares

Ismail Atourki  
Ilyas Malik

5 juin 2018

## Rapport de Projet : Dynamique de population, processus de Galton-Watson



# Table des matières

<b>1</b>	<b>Introduction</b>	<b>3</b>
1.1	Méthodes exactes . . . . .	3
1.2	Monte-Carlo . . . . .	4
1.3	Sélection-Mutation . . . . .	4
1.4	Échantillonnage préférentiel . . . . .	4
<b>2</b>	<b>Loi de reproduction géométrique</b>	<b>5</b>
2.1	Le cas sous-critique $m < 1$ . . . . .	5
2.1.1	Probabilité de non extinction $P(A_n)$ . . . . .	5
2.1.2	Loi de $X_n$ conditionnellement à la survie $A_n$ . . . . .	10
2.1.3	Scénarios typiques où les populations ne sont pas éteintes . . . . .	11
2.1.4	Problème des ressources consommées . . . . .	12
2.2	Le cas sur-critique $m > 1$ . . . . .	14
<b>3</b>	<b>Extension à une loi de reproduction aléatoire</b>	<b>15</b>
3.1	Cas d'un environnement à deux habitats . . . . .	17

# 1 Introduction

Le but de ce projet est la simulation des événements rares dans le modèle de la population dynamique, modèle de Galton-Watson. Nous nous intéresserons tout d'abord au calcul de la probabilité de non extinction de la population dans le cas sous critique, ainsi qu'au problème des ressources consommées. Ensuite, nous traiterons le cas de la croissance dans le régime sur-critique. Avant de faire les simulations et les calculs des probabilités dans le régime rare, nous présentons les différentes méthodes utilisées dans ce projet.

Dans un premier temps, on modélise la loi des naissances par une loi géométriques  $\nu = G(m)$  de paramètre  $m$ , c'est-à-dire

$$P(Y = k) = \frac{1}{1+m} \left(\frac{m}{1+m}\right)^k, \text{ pour } k \geq 0.$$

Ainsi le nombre d'individu par génération est une chaîne de Markov, et est donnée par :

$$X_{n+1} = \sum_{k=1}^{X_n} Y_{n,k} \text{ où } Y_{n,k} \text{ sont iid de loi } \nu$$

Dans la suite nous noterons  $p = \frac{1}{1+m}$ .

## 1.1 Méthodes exactes

### Fonction génératrice

On note  $f$  la fonction génératrice de la loi  $\nu$ , un calcul immédiat montre que

$$f(x) = \frac{1}{1+m(1-x)}.$$

On peut trouver dans [Mél16] par exemple, une étude détaillée sur la loi de  $X_n$ . En particulier si  $X_0 = 1$  on trouve que la probabilité d'extinction vérifie

$$P(X_{n+1} = 0) = f(P(X_n = 0)).$$

Ce qui nous permet d'affirmer que

$$P(A_{n+1}) = P(X_{n+1} > 0) = 1 - P(X_{n+1} = 0) = 1 - f(P(X_n = 0))$$

Et dans le cas général si la population est constituée de  $q$  individus à l'instant initial, la probabilité d'extinction est  $P(X_n = 0)^q$  où  $X_n$  est la même chaîne de Markov défini précédemment avec  $X_0 = 1$ .

### Calcul de la matrice de transition

Cette méthode de calcul (quasi)exacte consiste à calculer la matrice de transition de la chaîne de Markov  $(X_n)_n$ . En effet,  $X_{n+1} = \sum_{k=1}^{X_n} Y_{n,k}$

Donc,

$$\begin{aligned} P(X_{n+1} = y | X_n = x) &= P\left(\sum_{k=1}^x Y_{n,k} = y | X_n = x\right) \\ &= \binom{x+y-1}{x-1} (1-p)^y p^x \end{aligned} \quad (1)$$

En effet, il suffit de voir la somme comme une liste de descendants et de cloisons qui délimitent le nombre de fils de chaque  $Y_{n,k}$ . Ensuite, il est facile de voir que pour marquer le dernier enfant dans une loi géométrique, on multiplie par  $p$ , pour  $x$  parents cela donne bien  $p^x$ . Enfin  $(1-p)^y$  correspond au nombre total d'enfants.

Nous avons calculé la matrice de transition de  $X_n$  avec un nombre fini d'états  $\{0, 1, \dots, 500\}$ . En élevant à la puissance  $n$ , on obtient  $P(X_n = j | X_0 = i)_{i,j}$  avec une précision plus que satisfaisante pour des  $i, j$  ne dépassant pas 100. En guise d'exemple, si  $m = 0.8$  :

$|P(X_{40} = 0 | X_0 = 20) - p_{calculée}| \approx 10^{-15}$ , pour une précision relative de  $10^{-11}$ .

## 1.2 Monte-Carlo

On considère une variable aléatoire  $Y$  à valeur dans  $\mathbf{R}$ , et une fonction  $f : \mathbf{R} \rightarrow \mathbf{R}$ , nous allons estimer la quantité  $E(f(Y))$ . La méthode de Monte-Carlo consiste à considérer une suite de variable aléatoire  $(Y_n)_n$  i.i.d de même loi que  $X$ , ensuite à utiliser la loi forte des grands nombres pour estimer  $E(f(Y))$ .

Ainsi on peut assurer pour  $n$  assez grand que :

$$E(f(Y)) \approx m_n \text{ où } m_n = \frac{1}{n} \sum_{i=0}^n f(Y_i)$$

Dans notre cas, chaque  $Y_i$  est une trajectoire  $X_{0,1,\dots,n}$ .

La valeur donnée par cet estimateur n'est pas exacte, il faut donc être sûr avec probabilité  $1 - \alpha$  que la vraie valeur se trouve dans l'intervalle de confiance IC. Celui-ci se calcule de la manière suivante :

$$IC = \left[ m_n \pm \frac{\sigma_n q_{1-\alpha/2}}{\sqrt{n}} \right]$$

où  $\sigma_n$  est l'écart-type empirique des variables  $Y_i$ .

## 1.3 Sélection-Mutation

Comme vu en cours, le principe de cette méthode est toujours de simuler un grand nombre de trajectoires, mais en sélectionnant cette fois les trajectoires les plus avantageuses c.à.d celles qui sont les plus aptes à vérifier une condition -surtout lorsque cette condition ne se produit pas souvent-.

Pour cela, on pondère à chaque pas de temps les  $X_n$  par une fonction de pondération  $G$  -Sélection- qui est grande si la trajectoire est avantageuse et petite sinon, ensuite on tire aléatoirement et avec probabilité proportionnelle à  $G$  -Mutation- les trajectoires (qui peuvent alors être dupliquées, supprimées ...).

Les nouvelles variables aléatoire ont alors une nouvelle densité jointe (celle-ci est calculée en cours), on peut alors calculer la probabilité d'un évènement rare en le favorisant puis en effectuant un simple changement de probabilité.

En reprenant les notations du cours, on a la formule suivante :

$$\mathbb{E}[f(X_{0:n})] \approx \left( \frac{1}{M} \sum_{m=1}^M \frac{f(X_{0:n}^{(n,m)})}{\prod_{p=0}^{n-1} G_p(X_{0:p}^{(n,m)})} \right) \times \left( \prod_{p=0}^{n-1} \frac{1}{M} \sum_{m=1}^M G_p(X_{0:p}^{(p,m)}) \right).$$

Remarquons qu'il n'est pas possible de donner un intervalle de confiance à partir de cette simulation car les trajectoires ne sont pas indépendantes. Nous allons donc essayer de donner l'intervalle de confiance empirique de la valeur exacte  $p$ , en faisant apparaître la moyenne empirique de  $L$  valeurs calculée à l'aide de l'algorithme de sélection mutation  $(p_i)_{0 \leq i \leq L}$ . Ainsi on peut assurer avec une probabilité  $1 - \alpha$ , que la valeur exacte appartient à l'intervalle de confiance suivant :

$$IC = \left[ \frac{1}{L} \sum_{l=1}^L p_l \pm \frac{t_{(\alpha, L-1)} \sigma_L}{\sqrt{L}} \right]$$

où  $t_{(\alpha, L-1)}$  est le quantile de la loi Student à  $L-1$  degré et au niveau  $\alpha$  et  $\sigma_L$  et l'écart type empirique de  $(p_i)_{0 \leq i \leq L}$ .

## 1.4 Échantillonnage préférentiel

Cette méthode consiste à effectuer un changement de probabilité puis à lancer des simulations de trajectoires sous cette nouvelle loi de probabilité. Celle ci doit permettre d'avoir un nombre non négligeable de réalisations qui vérifient la condition dont on veut calculer la probabilité.

En reprenant les notations du cours :

$$\mathbb{E}[g(X)] = \mathbb{E} \left[ \frac{g(X)}{L(X)} L(X) \right] = \mathbb{E}_{\mathbb{Q}} \left[ \frac{g(X)}{L(X)} \right].$$

Sachant que  $L$  vérifie :  $E(L(X)) = 1$ . Les trajectoires étant indépendantes, l'intervalle de confiance se retrouve de la même façon que pour celui d'un Monte-Carlo naïf grâce au TCL.

## 2 Loi de reproduction géométrique

On suppose ici que  $\nu = G(m)$ ,  $m$  s'interprète comme la moyenne du nombre de descendants d'un individu donné.

### 2.1 Le cas sous-critique $m < 1$

Dans ce cas, p.s la population s'éteint en temps fini. L'évènement de non-extinction  $A_n = \{X_n > 0\}$  à la génération  $n$  devient rare si  $n$  est grand, si  $m$  est trop petit ou si la population initiale est trop faible.

#### 2.1.1 Probabilité de non extinction $P(A_n)$

Nous avons estimé la probabilité de non extinction de la population par les méthodes suivantes :

**Méthode exacte :** Les figures 1 et 2 représentent l'influence de la taille de la population initiale  $X_0$ , ainsi que l'influence du paramètre de loi  $\nu$ .

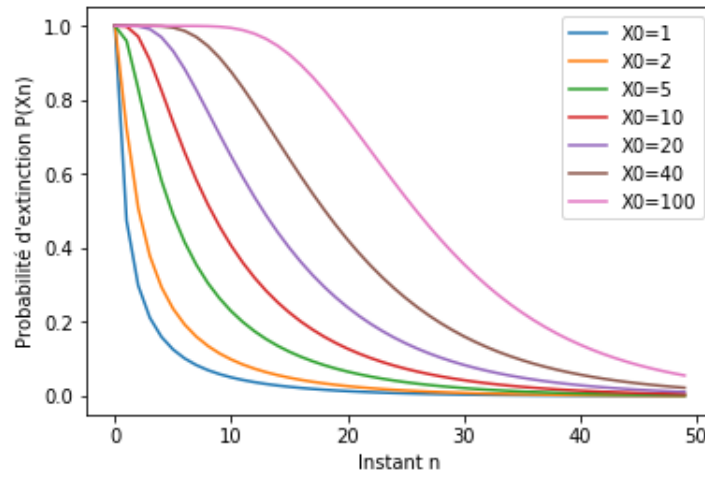


FIGURE 1 –  $P(X_n > 0 | X_0 = X_0)$  pour différentes valeurs de  $X_0$  avec  $m = 0.9$

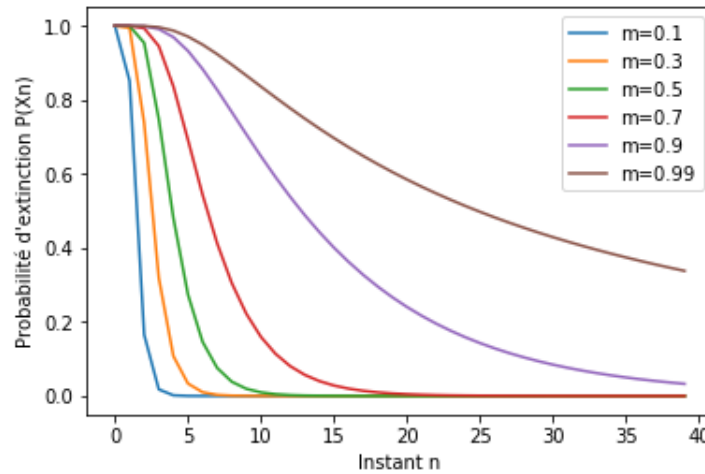


FIGURE 2 –  $P(X_n > 0 | X_0 = 20)$  pour différentes valeurs de  $m$

La figure 1 représente l'impact de la taille de la population initiale sur la probabilité d'extinction, nous avons choisi  $m=0.9$ . Ainsi on peut remarquer que la population s'éteint très rapidement quand la taille de la population initiale devient petite (pour  $X_0 = 1$  par exemple). Et s'éteint lentement si la taille initiale est grande.

La figure 2 représente l'impact du paramètre de la loi  $\nu$  sur la taille de la population. Nous avons choisi une population initiale de 20 individus. Ainsi on peut remarquer que la population s'éteint très rapidement quand le paramètre de la loi  $\nu$  devient petit. Ceci est intuitif parce que  $m$  représente la moyenne du nombre d'enfant d'un individu donné. Ainsi si  $m$  est proche de 1 alors l'individu a tendance à donner naissance à un enfant (On voit que pour  $m = 0.99$  la probabilité est non négligeable même à la génération  $n = 40$ ). En revanche si  $m \ll 1$ , chaque individu a une probabilité très faible d'avoir un descendant, ce qui explique la convergence rapide vers 0.

Par exemple, nous avons calculé la probabilité de l'évènement de non-extinction  $A_n = \{X_n > 0\}$  à la génération  $n = 40$  et pour un paramètre  $m = 0.7$ , et la taille initiale  $X_0 = 20$ . Ainsi la valeur exacte est :  $P(X_n > 0) = 3.82007 \times 10^{-6}$ .

Un deuxième exemple de calcul de probabilité de non-extinction de la génération  $n = 40$  et pour un paramètre  $m = 0.8$ , et la taille initiale  $X_0 = 20$ . Ainsi la valeur exacte est :  $P(X_n > 0) = 5.3161 \times 10^{-4}$ .

Ces deux exemples montre que le paramètre  $m$  a un effet non négligeable sur la taille de la population.

**Monte-Carlo naïf** Nous avons utilisé la méthode de Monte-Carlo classique avec un grand nombre de simulations de trajectoires indépendantes. Ensuite, nous avons utilisé le théorème centrale limite afin de trouver un intervalle de confiance à 95%.

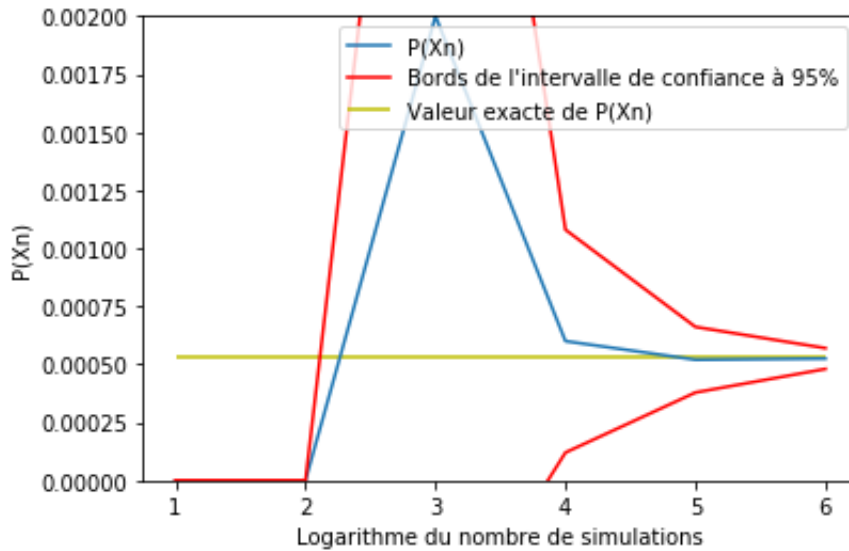


FIGURE 3 –  $P(X_n > 0 | X_0 = X_0)$  avec  $X_0 = 20, m = 0.8$  et  $10^x$  simulations

La valeur de  $P(X_n)$  représentée en bleu est encadrée par deux courbes en rouges qui représentent les bords de l'intervalle de confiance obtenus à partir du TCL. On remarque en particulier pour un petit nombre de simulations (inférieur à 100), Monte-Carlo ne donne pas une bonne estimation de la probabilité  $P(X_n)$  car aucune des trajectoires ne vérifie  $X_n > 0$ , de même l'intervalle de confiance à 95% ne donne aucune information sur cette probabilité. Mais une fois que le nombre de simulation devient grand (supérieur à 1000) la valeur estimée avec Monte-Carlo (la courbe en bleu) s'approche de la valeur exacte (courbe en jaune). D'autre part les deux bornes de l'intervalle de confiance (courbes en rouge) tendent vers la solution exacte.

Nous avons utilisé cette méthode pour calculer la probabilité de l'évènement de non-extinction  $A_n = \{X_n > 0\}$  à la génération  $n = 40$ . Nous avons choisi  $m = 0.7$  et  $X_0 = 20$  et finalement le nombre de simulation est  $10^4$ . Ainsi la valeur estimée avec cette méthode est :  $3 \times 10^{-4}$ , et

l'intervalle de confiance au niveau 95% dans ce cas est :  $[-0.00003942, 0.00063942]$ , ce qui n'est pas très satisfaisant.

**Sélection mutation** Nous avons utilisé la fonction  $G : x \mapsto \exp \lambda x$  pour pondérer les trajectoires, on sélectionne ainsi les trajectoires hautes. Le choix de  $\lambda$  a été crucial pour ce calcul.

En effet, il faut éviter des valeurs trop petites car on tomberait sur des simulations similaires à celles de Monte-Carlo (les trajectoires hautes ne seraient pas privilégiées) et les trajectoires risquent de toutes s'annuler à la générations  $n$ .

Il faut aussi éviter des valeurs trop grandes de  $\lambda$  car cela conduirait à une seule et unique trajectoire dupliquée partout qui croît à l'infini (l'algorithme prend alors trop de temps car il faut simuler un nombre gigantesque de lois géométriques), le résultat est alors biaisé.

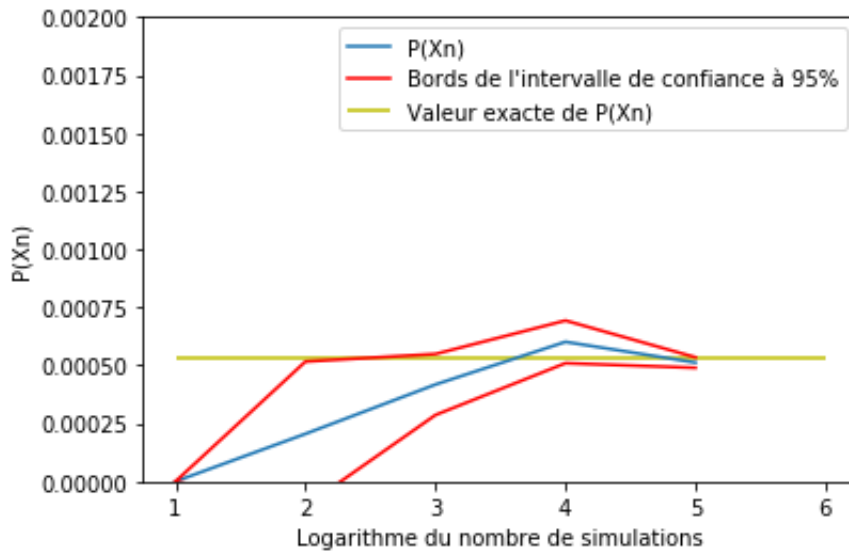


FIGURE 4 –  $P(X_n > 0 | X_0 = X_0)$  avec  $X_0 = 20, m = 0.8, \lambda = 0.01$  et  $10^x$  simulations

Nous avons utilisé cette méthode pour calculer la probabilité de l'évènement de non-extinction  $A_n = \{X_n > 0\}$  à la génération  $n = 40$ . Nous avons choisi  $m = 0.7, X_0 = 20$  et le paramètre de la fonction de poids  $G$  est choisi  $\lambda = 0.01$ . Finalement le nombre de simulations est  $20 \times 10^4$  ( $L \times M$ ). Ainsi la valeur estimée avec cette méthode est :  $6.7266 \times 10^{-4}$ , et l'intervalle de confiance au niveau 95% dans ce cas est :  $[0.00051713, 0.00082819]$

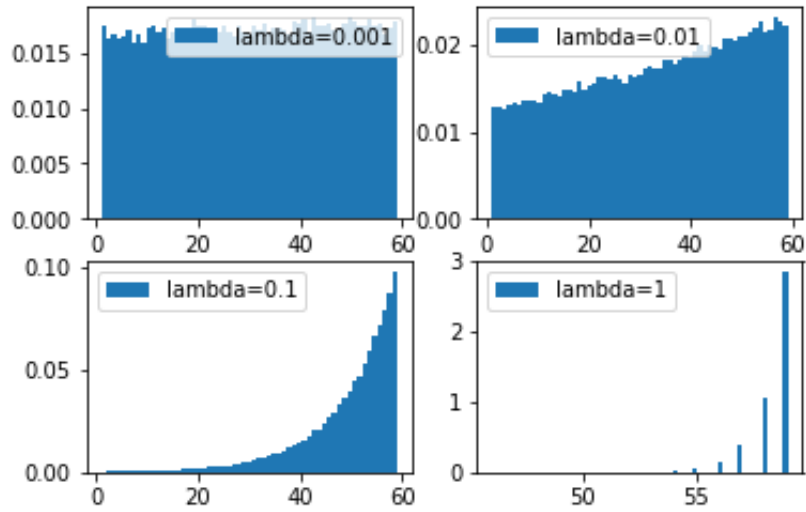


FIGURE 5 – Histogrammes de la fonction de pondération  $G$  pour différentes valeurs de  $\lambda$

Comme on peut le remarquer sur les figures ci-dessus (l'espace d'états considéré  $[0, 60]$  est fini), un bon choix de  $\lambda$  serait tel que :  $\lambda \approx 0.01$ . Remarquons que pour  $\lambda = 1$ ,  $G(i) \approx 0$  pour  $i \leq 55$ , si bien que ces valeurs ne sont même pas représentées.

**Échantillonnage préférentiel** Nous avons estimé la probabilité de non extinction de la population à l'aide d'un changement de probabilité :

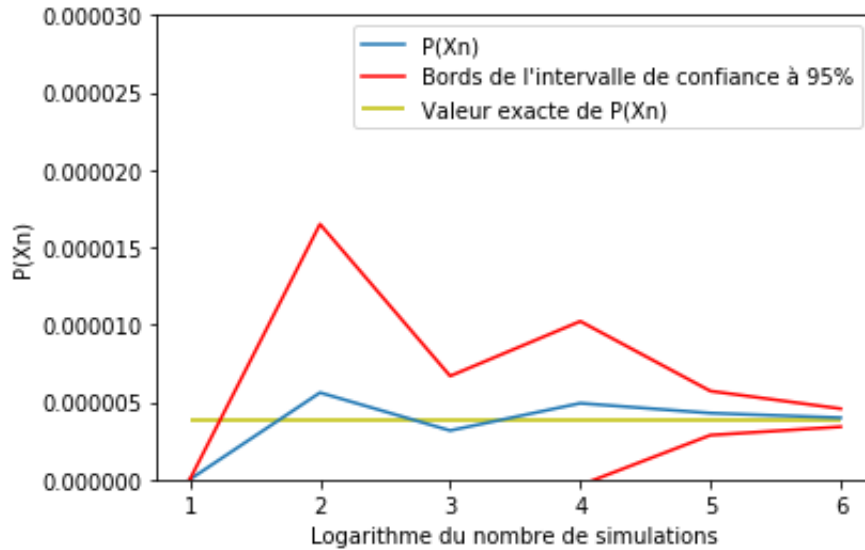


FIGURE 6 –  $P(X_n > 0 | X_0 = X_0)$  avec  $X_0 = 20, m = 0.7, m' = 0.97$  et  $10^x$  simulations

Le problème avec la méthode de Monte-Carlo est que les événements  $A_n$  se produisent très rarement, nous avons alors effectué un changement de probabilité qui consiste simplement à modifier la moyenne de descendants par individu  $m$ , en augmentant celui-ci (il doit tout de même vérifier  $m < 1$  pour éviter la divergence des trajectoires et le blocage de l'algorithme), on augmente nettement les chances de non extinction à la génération  $n$ , bien sûr ceci est compensé par la multiplication par la densité  $L = \frac{P}{Q}$  comme vu dans le cours.

Il est important de remarquer que cette méthode reste extrêmement précise même pour un petit nombre de simulations  $\approx 10^3$ , ceci n'est pas si anormal car comme on l'a vu, sous la nouvelle loi de probabilité  $Q$ , l'événement  $A_n$  n'est plus rare et la formule vue en cours donne un résultat assez exact.



Nous avons utilisé cette méthode pour calculer la probabilité de l'évènement de non-extinction  $A_n = \{X_n > 0\}$  à la génération  $n = 40$ . Nous avons choisi  $m = 0.7, m' = 0.97$  et  $X_0 = 20$  et finalement le nombre de simulation est  $10^4$ . Ainsi la valeur estimée avec cette méthode est :  $5.6315 \times 10^{-4}$ , et l'intervalle de confiance au niveau 95% dans ce cas est :  $[0.00042970, 0.00069660]$

**Comparaison** Dans cette partie, nous comparerons les différentes méthodes utilisées précédemment. Tout d'abord, on peut remarquer que la méthode d'échantillonnage préférentiel est particulièrement efficace, surtout pour des  $m$  petit ( $m < 0.5$ ) dans quel cas, les autres méthodes sont très inefficace. On peut aussi voir que la méthode de Sélection mutation est plus efficace que Monte-Carlo (évidemment), mais pas aussi efficace que l'échantillonnage préférentiel lorsque  $m$  est petit :

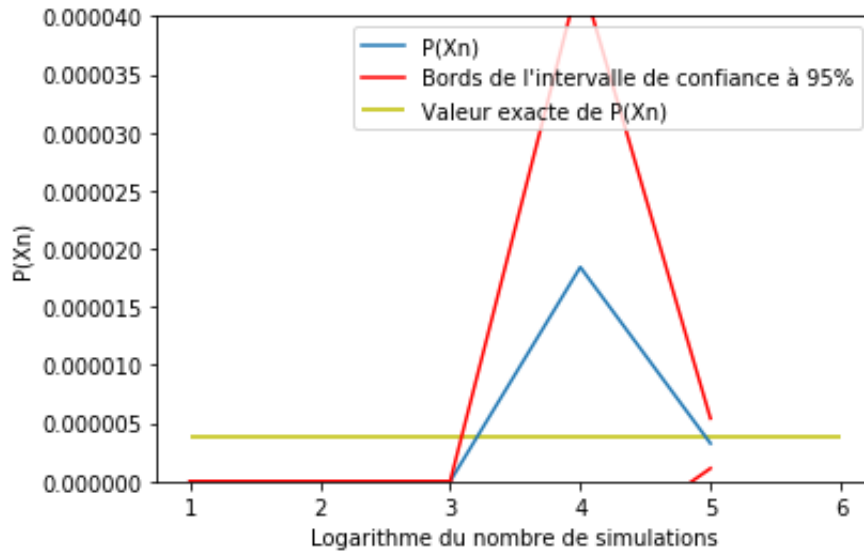


FIGURE 7 –  $P(X_n > 0 | X_0 = X_0)$  avec  $X_0 = 20, m = 0.7$  et  $10^x$  simulations (Sélection-Mutation)

Voici un exemple de ce que retourne notre algorithme pour  $m = 0.8, n = 40, X_0 = 20$  :

Valeur exacte: 0.00000382

Monte-Carlo naif: 1 000 000 simulations

Valeur: 3e-06

Erreur relative: 21.47 %

Intervalle de confiance à 95% : [-0.00000039, 0.00000639]

Largeur relative: 113.16 %

Sélection mutation: 20 fois avec 100 000 simulations

Valeur: 0.00000334

Erreur relative: 12.56 %

Intervalle de confiance à 95%: [0.00000165, 0.00000503]

Largeur relative: 50.63 %

Échantillonnage préférentiel: 1 000 000 simulations

Valeur: 3.677406262409005e-06

Erreur relative: 3.73 %

Intervalle de confiance à 95% : [0.00000330,0.00000405]

Largeur relative: 10.15 %

Remarquons que la méthode de Monte-Carlo donne un résultat peu précis (erreur relative de l'ordre 21.47%), ce qui montre que la méthode de Monte-Carlo n'est pas valable dans le régime rare (avec ce nombre de simulations). Pour cette raison, nous avons testé d'autres méthodes :  $2 \times 10^6$  simulations par la méthode de sélection-mutation donnent une valeur proche de la valeur exacte, mais avec une erreur relative de l'ordre de 12.56%, qui est plus petite que celle de Monte-Carlo. Le même nombre de simulations par échantillonnage préférentiel donne une très bonne valeur de  $P(\{X_n > 0\})$ , l'erreur relative dans ce cas est 3.73%. Ce qui montre que la méthode d'échantillonnage préférentiel est la meilleure méthode pour estimer la probabilité de non extinction. Remarquons aussi que la valeur exacte se trouve toujours à l'intérieur des intervalles de confiance au niveau 95% et que la largeur relative des IC est décroissante selon la méthode utilisée, ce qui donne un classement grossier de la qualité des méthodes utilisées.

### 2.1.2 Loi de $X_n$ conditionnellement à la survie $A_n$

À partir des simulations précédentes (Monte-Carlo et sélection-mutation), nous avons tracé deux histogrammes de  $X_n$  conditionnellement à la survie  $A_n$  pour chaque méthode et nous avons tracé des lois exponentielles qui collent bien aux histogrammes en choisissant les bons paramètres.

Les paramètres choisis dans ce cas sont :  $m = 0.8$ ,  $n = 30$  et  $X_0 = 20$ .

— **Monte-Carlo**

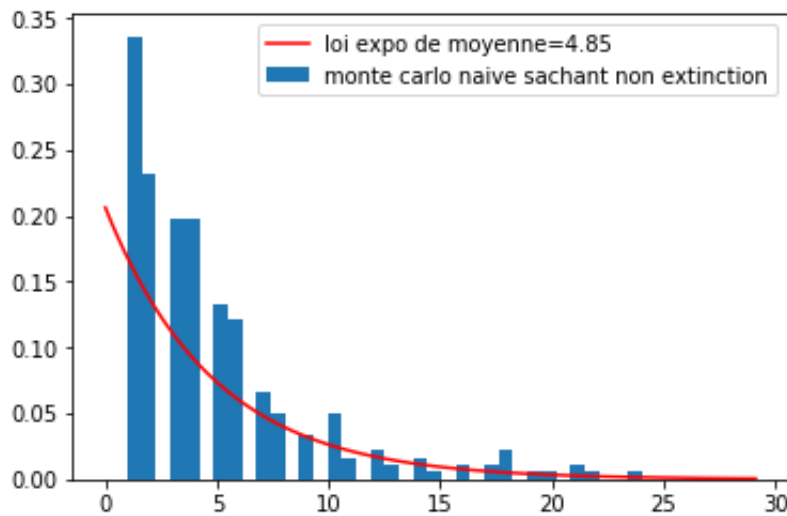


FIGURE 8 – Scénario où la population n'est pas éteinte par la méthode de Monte-Carlo

— **Sélection mutation**

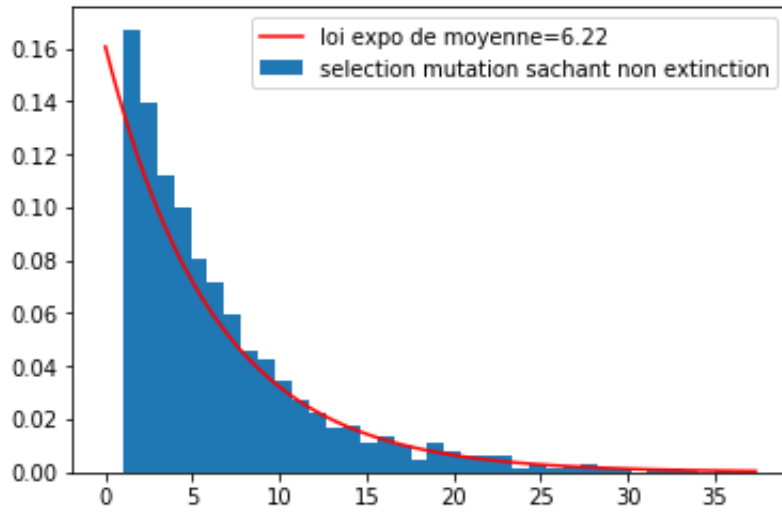


FIGURE 9 – Scénario où la population n’est pas éteinte par la sélection mutation pour  $\lambda = 0.01$

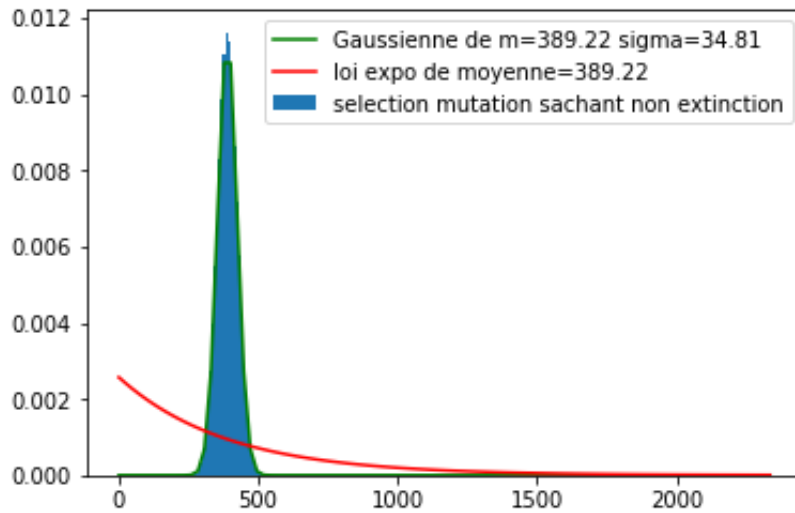


FIGURE 10 – Scénario où la population n’est pas éteinte par la sélection mutation pour  $\lambda = 0.1$

Le fait d’avoir une gaussienne dans l’histogramme est naturel dans le cas où les simulations sont indépendantes (Théorème centrale limite). Dans notre cas,  $\lambda$  est tellement grande que les trajectoires qui ne se sont pas éteintes ont forcément été sélectionnées pour grandir beaucoup, on peut imaginer que la gaussienne fait une translations vers la droite lorsque la génération finale  $n$  augmente.

### 2.1.3 Scénarios typiques où les populations ne sont pas éteintes

Nous avons tracé ceux-ci en sélectionnant les populations non éteintes. Les résultats sont présentés sur les deux figures 11 et 12. Remarquons que pour  $X_0 = 10$  et  $m = 0.8$  il ne reste que trois familles non éteintes au temps  $n = 30$ . En revanche si  $X_0 = 20$  et  $m = 0.9$  il reste beaucoup de familles non éteintes au temps  $n = 40$ . D’autre part, les familles non éteintes donnent naissance un moment donné à un grand nombre d’enfants ce qui explique les piques pour  $n = 5, 15$  présentés dans la figure 12, c’est ce grand nombre d’enfants qui leur permet de survivre pendant une longue période.

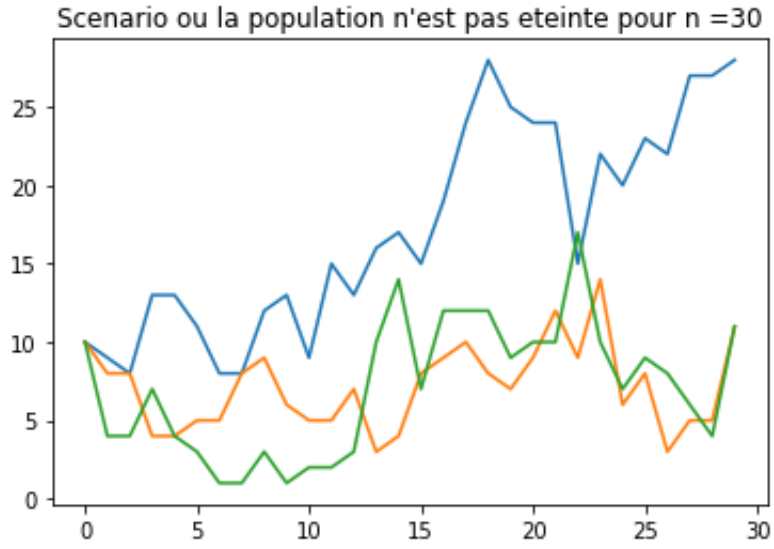


FIGURE 11 – Scénario où la population n'est pas éteinte pour  $m = 0.8$  et  $X_0 = 10$

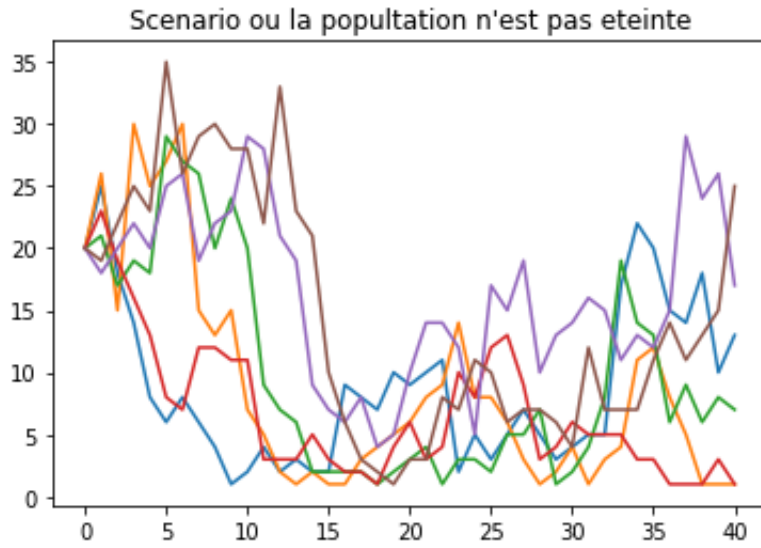


FIGURE 12 – Scénario où la population n'est pas éteinte pour  $m = 0.9$  et  $X_0 = 20$

#### 2.1.4 Problème des ressources consommées

On étudie dans cette partie le problème des ressources consommées par la population, en évaluant  $P(\max_{i=0,\dots,n} X_i > x)$  et  $P(\sum_{i=0}^n X_i > x)$  pour  $x$  grand.

**Estimation de  $P(\sum_{i=0}^n X_i > x)$**

Dans ce cas nous avons simulé  $10^4$  simulation de Monte-Carlo, avec les paramètres  $m = 0.6$  et  $x = 200$  et  $n = 40$ , la probabilité estimée dans ce cas est :  $P(\sum_{i=0}^n X_i > x) \approx 1 \times 10^{-5}$  et l'intervalle de confiance au niveau 95% est  $[9.6 \times 10^{-6}, 2.9 \times 10^{-5}]$ .

La figure 13 représente différents scénarios de la population, remarquons en particulier que seule

une famille dépasse 100 individus sur 50 ans. On remarque que la somme  $\sum_{i=0}^n X_i$  reste stationnaire presque pour toutes les trajectoires à partir du moments  $n = 10$ , car les populations éteintes à partir de ce moment là.

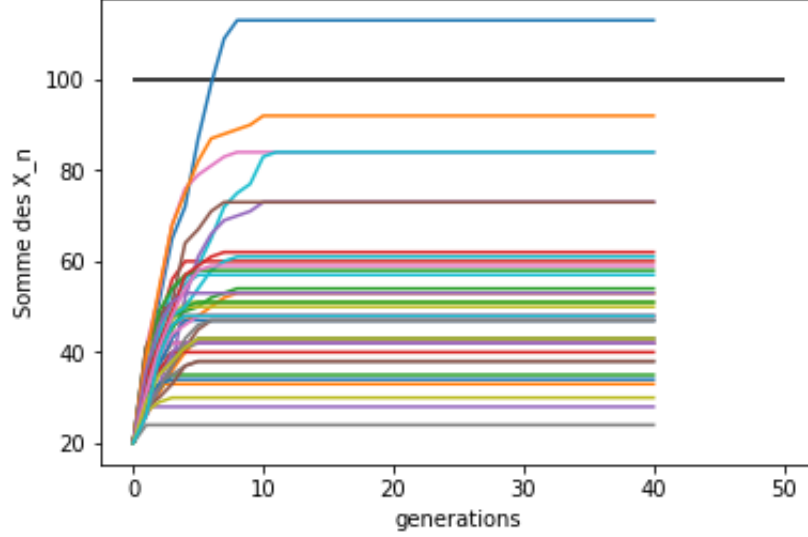


FIGURE 13 – Croissance de la population en fonction des paramètres  $c$  et  $m$

#### Estimation de $P(\max_{i=0,\dots,n} X_i > x)$

Dans ce cas nous avons effectué  $10^4$  simulations de Monte-Carlo, avec les paramètres  $m = 0.6$  et  $x = 150$ , la probabilité estimée dans ce cas est :  $P(\max_{i=0,\dots,n} X_i > x) \approx 5.8 \times 10^{-5}$  et l'intervalle de confiance au niveau 95% est  $[4.307 \times 10^{-5}, 7.2 \times 10^{-5}]$

La figure 14 représente différents scénarios pour ce cas. Remarquons que seulement 3 scénarios ont pu donner naissance à plus de 150 enfants comme record sur 40 générations.

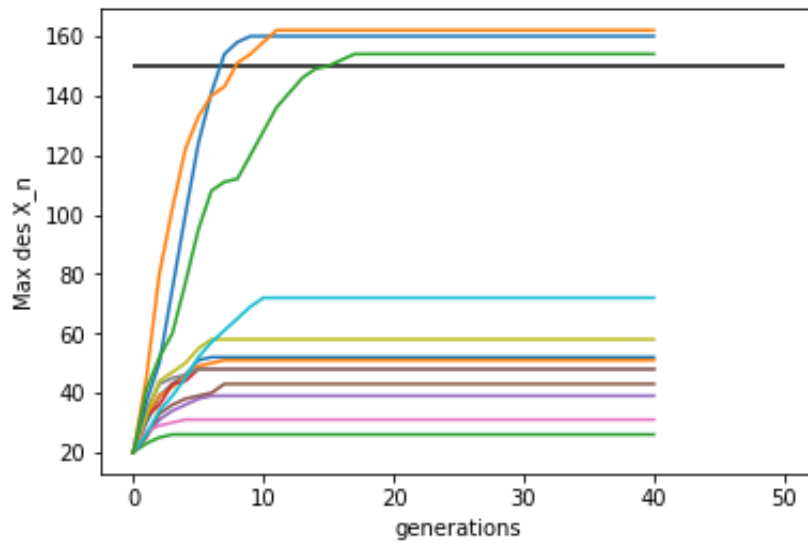


FIGURE 14 – Croissance de la population en fonction des paramètres  $c$  et  $m$

## 2.2 Le cas sur-critique $m > 1$

Dans ce paragraphe on suppose que  $m > 1$ , on s'intéressera à l'étude de la croissance de la population dans ce cas.

### Estimation de la probabilité $P(0 < X_n \leq c^n)$

Nous avons estimé la probabilité  $P(0 < X_n \leq c^n)$  à l'aide de Monte-Carlo et de la méthode sélection-mutation. Les paramètres du problème utilisés dans ce cas sont :  $X_0 = 2$ ,  $m = 1.2$ ,  $c = 1.08$ .

$10^5$  simulations par Monte Carlo donnent une valeur estimée :  $P(0 < X_n \leq c^n) = 5.4 \times 10^{-4}$ . L'intervalle de confiance au niveau 95% dans ce cas est :  $[4 \times 10^{-4}, 6 \times 10^{-4}]$ . Donc avec une probabilité de 0.95 on peut assurer que la valeur exacte est dans cette intervalle.

Par exemple, la figure 15 représente la convergence vers la valeur exacte pour les paramètres ( $m = 1.2, c = 1.1$ ). La courbe en bleu est la valeur estimée par Monte-Carlo et les deux courbes rouges sont l'intervalle de confiance au niveau 95%. Remarquons en particulier que l'amplitude des oscillations des trois courbes diminue, ce qui montre bien qu'il s'agit d'une convergence vers la valeur exacte.

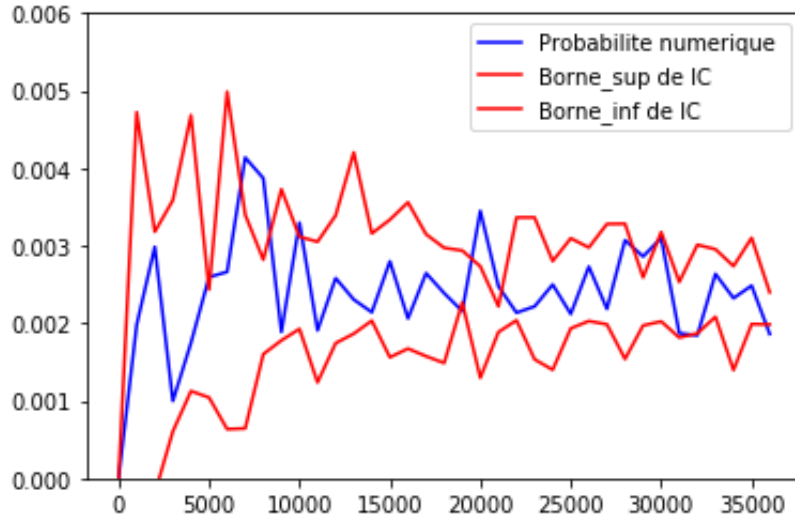


FIGURE 15 – Croissance de la population en fonction des paramètres  $c$  et  $m$

La probabilité estimée sort parfois de l'intervalle de confiance à 95%. Ceci est normal puisque la probabilité estimée ne suis pas exactement une loi gaussienne mais l'approche plus ou moins bien selon le nombre de simulations. Mais surtout, l'IC est un IC à 95%, ce qui signifie qu'à 5% de chance, la probabilité estimée sort de l'IC.

**Scénarios de la population :** Nous avons tracé ces scénarios à l'aide de la méthode de Monte-Carlo, pour un nombre de simulations  $M = 10^5$  et pour différentes valeurs de  $m$  et de  $c$ . Les résultats des scénarios sont présentés dans la figure 16. Une étude détaillée de la chaîne de Markov  $(X_n)_n$  dans [Mél16], montre que  $E(X_n) = m^n$ , ainsi on peut déjà dire, qu'en moyenne pour  $n$  assez grand  $X_n \sim m^n$ , on a alors 3 cas pour  $n$  assez grand :

- $P(0 < X_n \leq c^n) \approx 0$  si  $c \leq 1$ , car on sait déjà que la population explose à l'infini. Alors presque sûrement toutes les valeurs de  $X_n$  sont supérieures 1. Les deux courbes de la figure 12 montre ce phénomène. Pour  $c = 1$  et  $m = 1.2$ , seules deux valeur de  $X_n$  parmi  $10^6$  sont inférieures à 1 comme le montre la courbe pour  $c = 1$  et  $m = 1.2$ . Mais pour  $c = 0.8$  aucune valeur de  $X_n$  n'est inférieure à  $c^n$ .
- $P(0 < X_n \leq c^n) \approx 1$  si  $c \geq m$ , car le nombre moyen d'individus dans la population est de l'ordre de  $m^n$ . C'est le cas de la courbe ci-dessous  $(m, c) = (1.2, 2)$ , toutes les trajectoires

$(X_n)_n$  sont en dessous de la fonction  $c^n$ . le cas  $m = c$  est un cas critique et il est représenté pour  $m = c = 1.2$  ci-dessous.

- Dans le cas où  $1 < c < m$  la probabilité  $P(0 < X_n \leq c^n)$  dépend à priori de  $m$  et  $c$ , mais on peut être sûr que celle-ci tend vers 0 à l'infini.

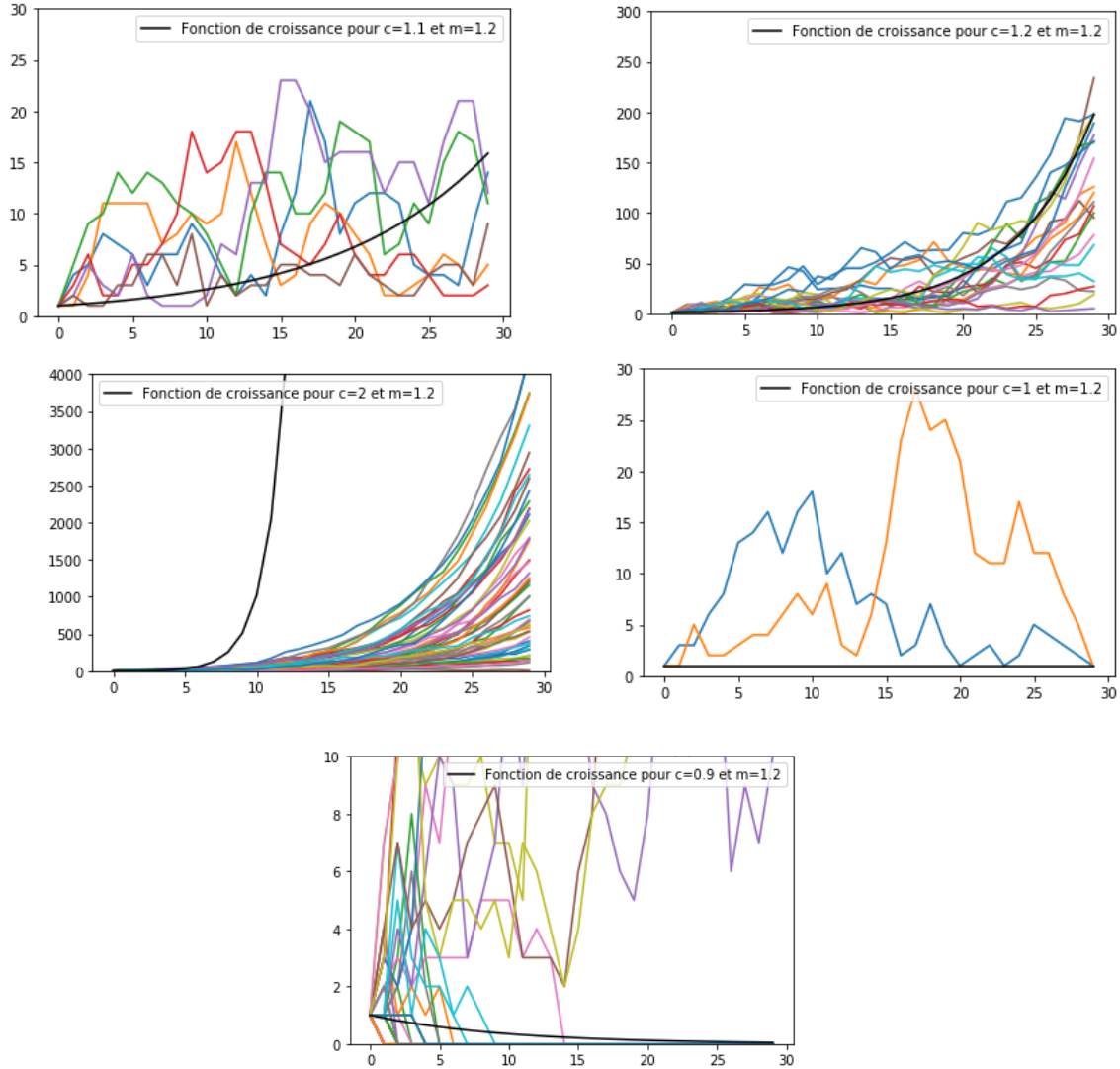


FIGURE 16 – Croissance de la population en fonction des paramètres  $c$  et  $m$

### 3 Extension à une loi de reproduction aléatoire

Dans toute la suite, on prendra  $X_0 = 20$ .

Dans cette partie, on prend en compte le fait que la reproduction de la population dépend de son environnement, qui change aléatoirement à chaque génération. À la génération  $n$ , la loi du nombre d'enfants est encore de type géométrique avec un paramètre  $m_n$ , qui est lui-même une variable aléatoire prenant deux valeurs  $m_h$  (environnement favorable) et  $m_b$  (défavorable). Les v.a.  $(m_n)_n$  sont supposées être i.i.d..

Nous considérerons que  $m_n$  suit une loi de Bernoulli :  $P(m_n = m_b) = p_b$ , et  $P(m_n = m_h) = 1 - p_b$ .

La figure suivante montre la dépendance de  $P(X_n > 0)$  par rapport à  $m_b, m_h$  et  $p_b$  :

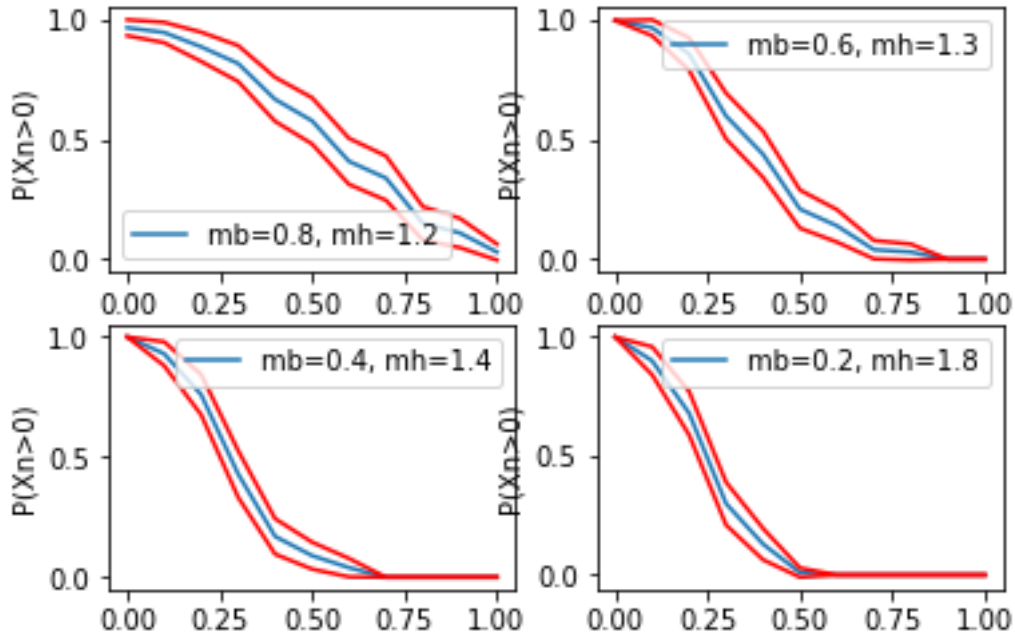


FIGURE 17 –  $P(X_n > 0)$  en fonction de  $p_b$  avec  $n = 20$

On voit bien qu'à la limite  $p_b = 0$ ,  $P(X_n > 0) = 1$  puisque  $m_n = m_h > 1$  avec probabilité 1. D'autre part, si  $p_b = 1$  on retombe sur le problème initial avec  $m = m_b$ . Les 4 courbes de la figure ont à peu près la même allure, la décroissance s'interprétant comme l'effet nocif de la probabilité d'occurrence d'une catastrophe (qui ferait baisser  $m_n$  à  $m_b$ ). On voit que plus  $m_b$  est petit, plus  $P(X_n > 0)$  a une décroissance accentuée en  $p_b = 0$ . On peut aussi remarquer que l'intervalle de confiance est plus large pour des probabilités  $p_b \approx \frac{1}{2}$ , en effet ceci est plus clair sur la figure suivante :

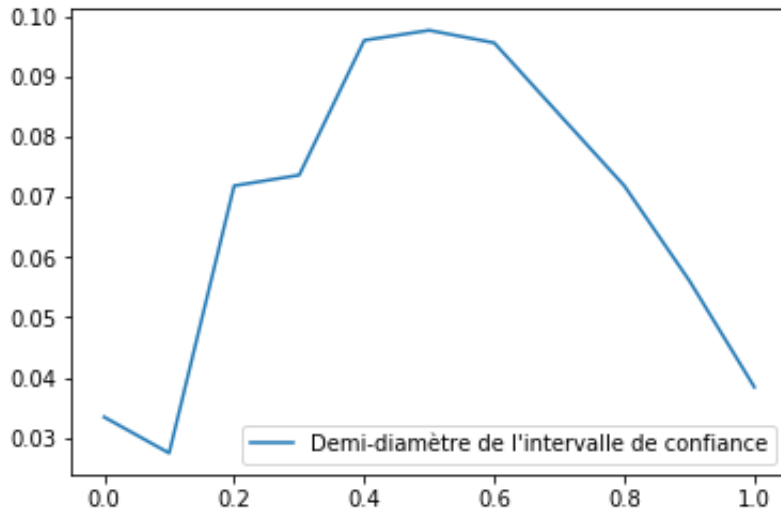


FIGURE 18 – Demi-diamètre de l'IC en fonction de  $p_b$  avec  $m_b = 0.8, m_h = 1.2, n = 20$  et 100 simulations

Moralement, on pourrait interpréter cela comme l'effet de la variance de la loi  $m_n$  sur la variance de notre résultat. En effet, la variance d'une loi  $\mathcal{B}(p)$  est égale à  $p(1-p)$  qui est maximale en  $p = \frac{1}{2}$ . Nous pouvons observer le comportement de quelques trajectoires non éteintes à l'étape  $n = 37$  avec  $m_b = 0.1, m_h = 1.5, p_b = 0.4$ . En utilisant la méthode de Monte-Carlo, on a  $P(X_n > 0) \in [0.000413, 0.000707]$  avec une probabilité 95% et voici quelques trajectoires non éteintes à la génération  $n$  :



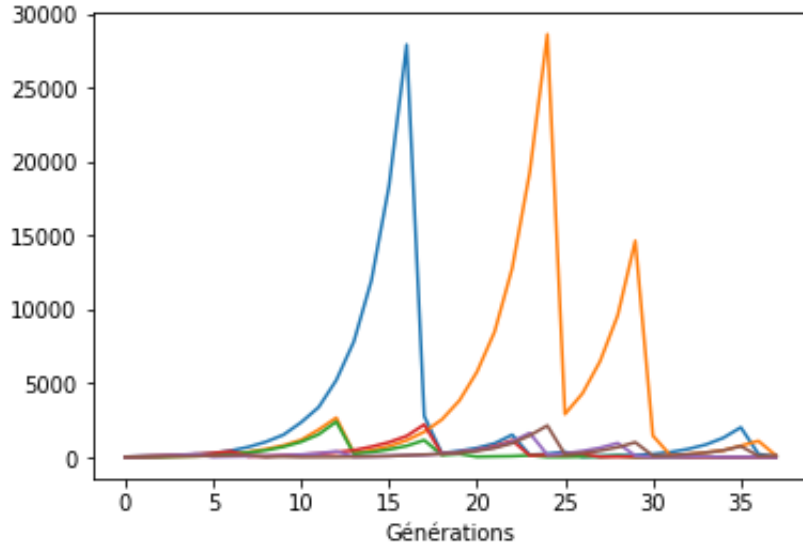


FIGURE 19 – Trajectoires non éteintes à l'étape 37 avec la loi de reproduction aléatoire

Nous remarquons d'après la figure 19 que souvent, les populations qui persistent ont connu des phases de grandes fortunes (qui correspondent aux pics) : elles ont eu la chance d'avoir de bonnes conditions  $m_i = m_h$  pour des générations successives et ont atteint de grandes démographies, elles ont aussi failli s'éteindre pendant quelques générations.

### 3.1 Cas d'un environnement à deux habitats

Nous allons maintenant considérer deux habitats 1 et 2. Le premier habitat est de bonne qualité  $m^{(1)} > 1$  mais il s'y produit parfois des catastrophes  $m^{(1)} \ll 1$ . Le deuxième habitat est de moins bonne qualité  $m^{(2)} < 1$ , néanmoins celle-ci est stable.

Nous avons alors :

$$m_n^{(1)} = \begin{cases} m_b < 1 & \text{MOYENNE BASSE AVEC PROBABILITÉ } p \\ m_h > 1 & \text{MOYENNE HAUTE AVEC PROBABILITÉ } 1 - p \end{cases}$$

et  $m_n(2) = m_s < 1$  la moyenne est stable dans le deuxième habitat.

Si on prend  $p = 0.1, m_b = 0.00001, m_h = 1.5, m_s = 0.9$ , chacun des habitats séparément est très mal adapté à nos populations, il conduit à l'extinction rapide de celles-ci.

La figure suivante compare  $P(X_n > 0)$  dans les cas où la population est confinée dans l'un des habitats et dans le cas où l'environnement contient les deux habitats 1 et 2 et que la fraction  $f$  de la population est mise dans l'habitat 1 et  $1 - f$  est dans l'habitat 2 :

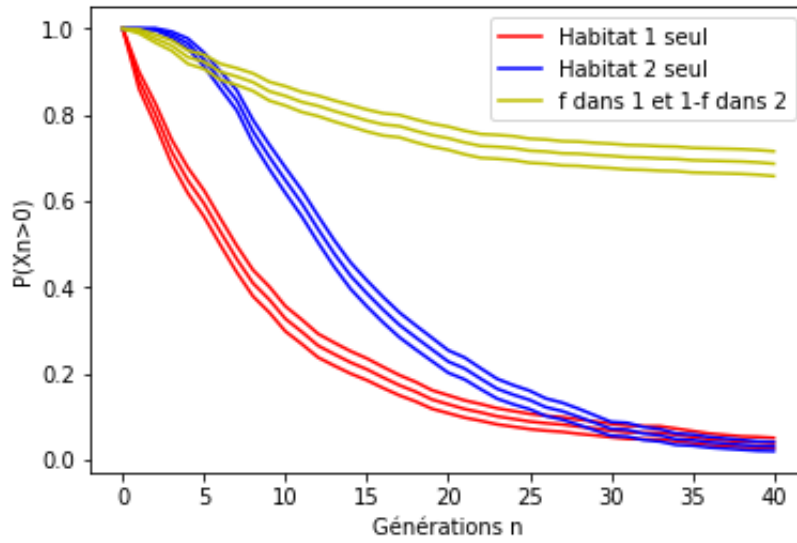


FIGURE 20 –  $P(X_n > 0)$  avec les bords de l'IC à 95%

Dans la figure ci-dessus, nous avons pris  $f \approx 75\%$ . En effet, selon l'étude du taux de croissance de la population (qui correspond à la moyenne géométrique du taux de croissance par génération) cela correspond à la valeur optimale pour assurer la prospérité de notre population :  $f_{optimale} = m_s \frac{(m_h - m_s) - p(m_h - m_b)}{(m_s - m_b)(m_h - m_s)}$ . Cette valeur maximise alors le taux de croissance et si celui-ci est supérieur à 1, la population peut alors persister dans l'environnement [AY98].

Voici une figure qui montre la dépendance du taux de croissance en fonction de la fraction  $f$  mise dans l'habitat 1 :

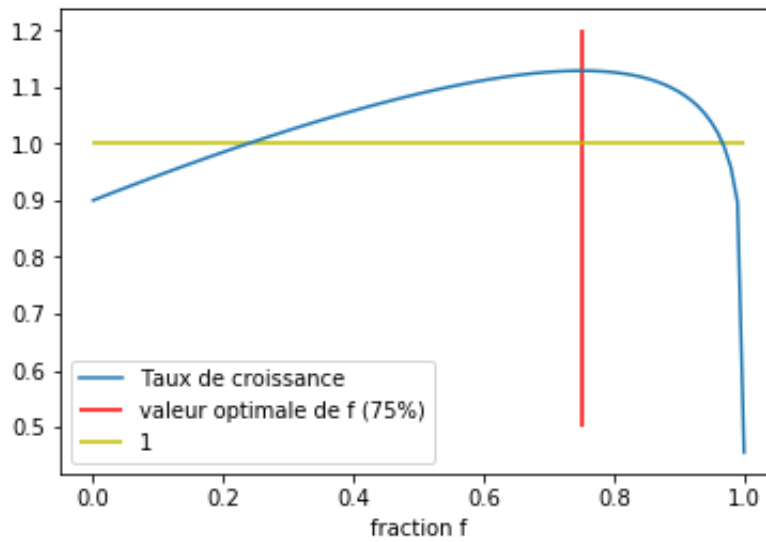


FIGURE 21 – Taux de croissance en fonction de  $f$

On voit bien que la fraction utilisée dans la figure 20 est la fraction optimale, comme le taux de croissance est supérieur à 1, la population persiste.

Pour nous assurer que la probabilité de non extinction reste importante, on peut voir l'évolution de  $P(X_n > 0)$  pour  $n$  allant jusqu'à 60 pour l'environnement composé de deux habitats :

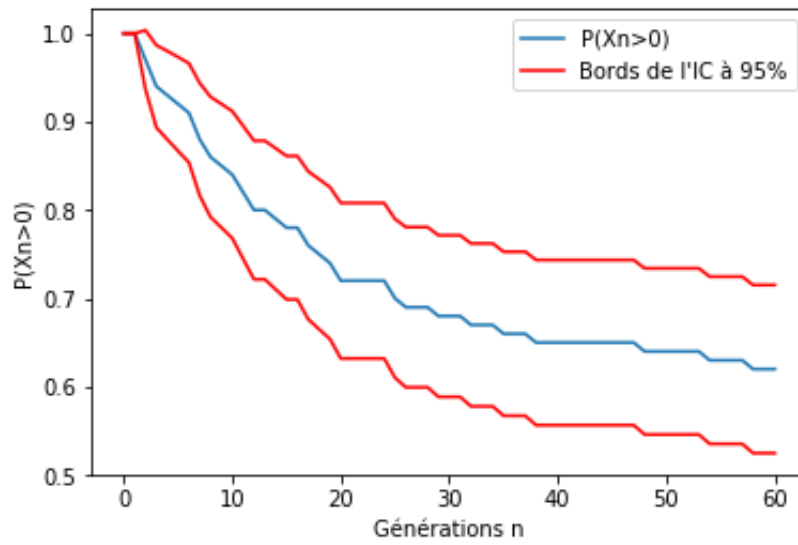


FIGURE 22 –  $P(X_n > 0)$  avec  $f = f_{optimale}$  et 100 simulations

Dans la figure ci-dessus, nous avons pris  $X_0 = 60$  :

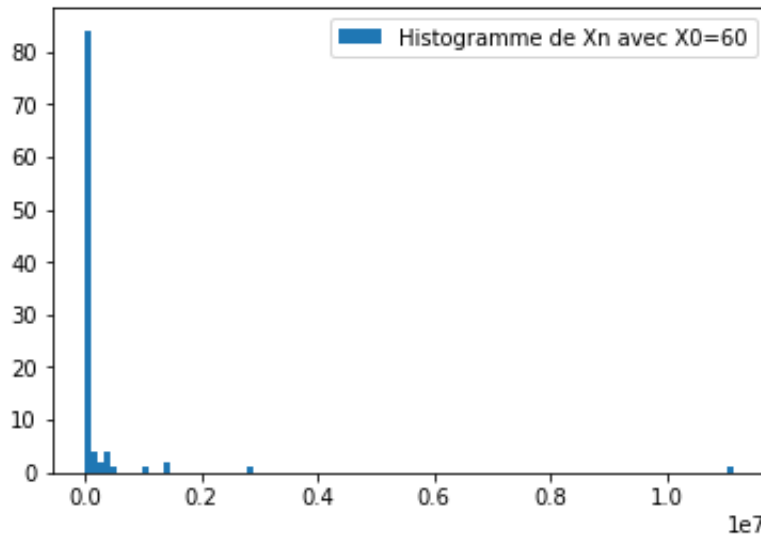


FIGURE 23 – Histogramme de  $X_n$  dans le cas du l'environnement optimal à deux habitats

Remarquons que la population atteint des démographies gigantesque (les abscisses sont en unité de  $10^7$ ), ce qui prouve bien que l'environnement est très prospère.

La figure suivante montre 6 trajectoires issues du même modèle mais avec  $X_0 = 20$  :

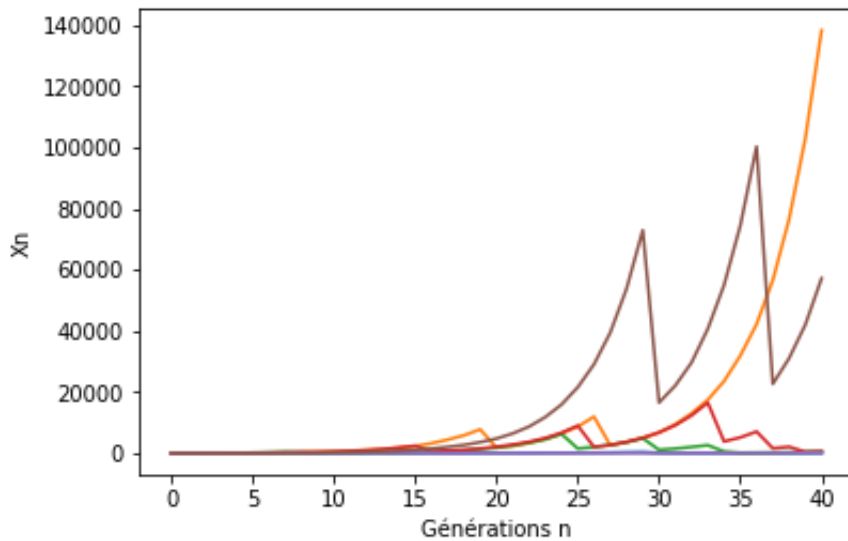


FIGURE 24 – 6 trajectoires où  $f = f_{optimale}$

Les trajectoires semblent croître globalement, ce qui est en harmonie avec les dernières courbes de cette partie : La population augmente avec probabilité non négligeable et atteint un grand nombre.

L'environnement est prospère, mais on peut voir que par moments (avec probabilité  $p = 0.1$ ), une catastrophe se produit dans l'habitat 1 (ce qui correspond aux pics) -qui contient  $f \approx 75\%$  de la population- et cause la mort d'à peu près 75% de la population.

L'habitat 2 agit comme un habitat de stockage temporaire. En effet, tant que l'habitat 1 est prospère, il produit de plus en plus d'individus qui sont alors stockés dans l'habitat 2 (afin de garder la fraction  $f$  constante). Quand l'habitat 1 est détruit, la population n'est alors détruite qu'à  $f = 75\%$  et les individus dans l'habitat stable (2) repartent alors dans l'habitat de grande productivité (1) pour augmenter la démographie.

## Références

- [AY98] Vincent A.A.JANSEN et Jin YOSHIMURA. "Populations can persist in an environment consisting of sink habitats only". In : (1998).
- [Mél16] Sylvie MÉLÉARD. *Modèles aléatoires en Ecologie et Evolution*. Ecole polytechnique, 2016.