



Introduction

Yule avait étudié la répartition des espèces entre les différents genres dans la classification des plantes et des animaux. Les données qu'il avait compilées montraient que la plupart des genres ne contenaient qu'une seule espèce, que de moins en moins de genres contenaient un plus grand nombre d'espèces et qu'il y avait simplement quelques genres contenant un grand nombre d'espèces. [1]

Ainsi, ce projet de recherche vise à **comprendre, critiquer et améliorer** le modèle mathématique de Yule dans une démarche mathématique et informatique.

Premièrement, nous avons implémenté le modèle de Yule sur python afin de comparer les résultats obtenus via informatique et les données compilées de Yule.

| Number of species | Chrysomelidae | Number of genera | Snakes | Lizards |
|-------------------|---------------|------------------|--------|---------|
| 1 | 215 | 469 | 131 | 105 |
| 2 | 90 | 152 | 35 | 44 |
| 3 | 38 | 82 | 28 | 23 |
| 4 | 35 | 61 | 17 | 14 |
| 5 | 21 | 33 | 16 | 12 |
| 6 | 16 | 36 | 9 | 7 |
| 7 | 15 | 18 | 8 | 6 |
| 8 | 14 | 17 | 8 | 4 |
| 9 | 5 | 14 | 9 | 5 |
| 10 | 15 | 11 | 4 | 5 |
| 11-20 | 58 | 74 | 10 | 17 |
| 21-30 | 32 | 21 | 12 | 9 |
| 31-40 | 13 | 15 | 3 | 3 |
| 41-50 | 14 | 8 | 1 | 2 |
| 51-60 | 5 | 4 | 0 | 0 |
| 61-70 | 8 | 3 | 0 | 1 |
| 71-80 | 7 | 0 | 1 | 0 |
| 81-90 | 7 | 1 | 0 | 0 |
| 91-100 | 3 | 1 | 1 | 0 |
| 101- | 16 | 4 | 0 | 2 |
| total | 627 | 1024 | 293 | 259 |

Figure 1: Données compilées de Yule

Ensuite, à partir de la brève nous avons cherché des points d'amélioration comme:

- La prise en compte du temps
- Ajouter un poids à la mutation selon sa capacité à générer ou non une nouvelle espèce
- Une espèce peut engendrer plusieurs espèces
- L'extinction d'une espèce

Modélisation

Ensuite, nous avons modifié l'équation de Yule en prenant appui sur le processus de branchement de Galton-Watson. Ainsi, selon un processus de branchement simple qui suit une loi de probabilité individuelle, nous cherchons à obtenir le nombre de successeurs directs que nous appliquons par récurrence afin d'obtenir la taille d'une population, modélisée par un arbre.

Processus stochastique: $\begin{cases} Z_0 = 1 \\ Z_{n+1} = f(Z_n) \end{cases}$

- Modèle probabiliste permettant d'étudier un phénomène aléatoire au cours du temps.
- f une règle d'évolution qui contient des aléas.

Modèle:

- **Modèle (1) : processus de branchement de Galton-Watson**

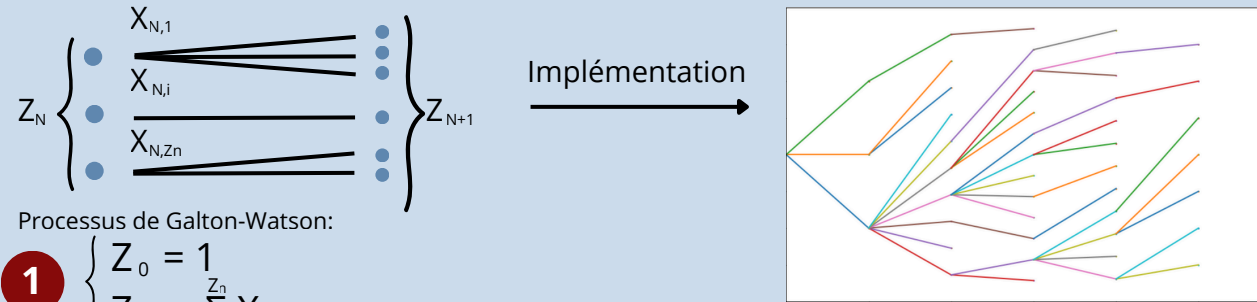


Figure 2: Processus de branchement selon le modèle 1

- **Modèle (2) : Notre proposition**

Ce modèle repose sur une loi individuelle qui génère le nombre de descendants ainsi qu'une loi de Binomiale déterminant si un descendant est viable ou non.

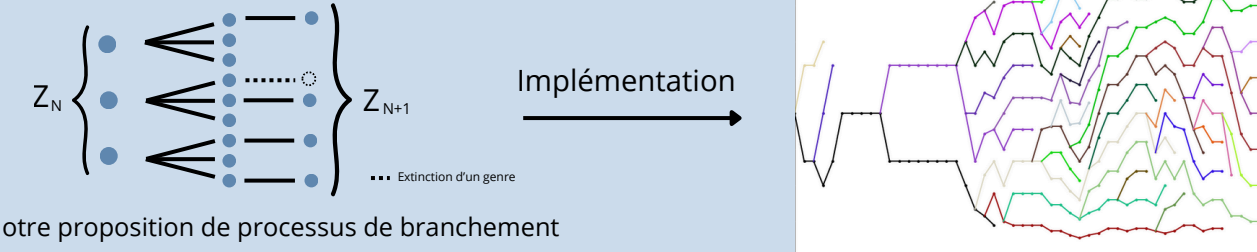


Figure 3: Processus de branchement selon le modèle 2

L'arbre généré semble alors plus plausible car une espèce ne va pas fréquemment donner une ou plusieurs nouvelles espèces.

Loi individuelle: Il s'agit de la loi de $X_{n,i}$

- **Loi de Poisson:** **A**

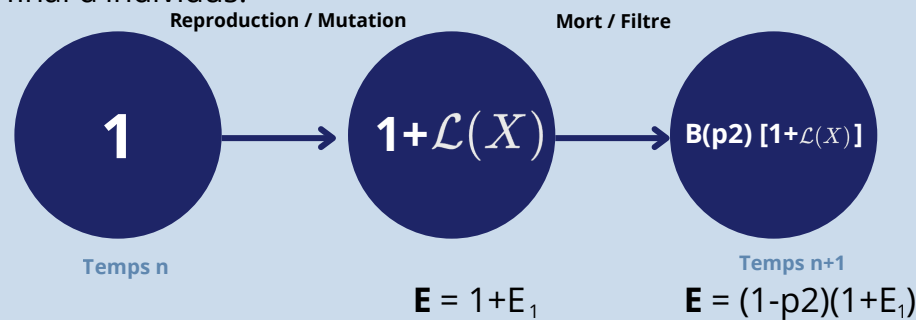
Génération de n descendants selon une loi de Poisson
Paramètre: λ

- **Loi Géométrique:** **B**

Génération de n descendants selon une loi Géométrique
Paramètre: p_1

Code/ Simulations

Pour tester statistiquement notre modèle, nous avons supprimé toutes les fonctionnalités d'affichage, en nous concentrant uniquement sur le nombre final d'individus.



Notations:

- E1: Espérance de la loi individuelle
- p2: Paramètre de la loi de Bernoulli

Avec autant de paramètres, on peut supposer qu'il est possible d'adapter n'importe quel arbre; quels sont donc les cas particuliers les plus judicieux? Voici une approche conceptuelle. L'opinion générale parmi les biologistes de l'évolution est que, à l'exception des extinctions massives, le nombre global d'espèces n'a pas tendance à augmenter ou à diminuer exponentiellement [2].

Peut-on rencontrer ces augmentations ou diminutions exponentielles en faisant varier nos paramètres?

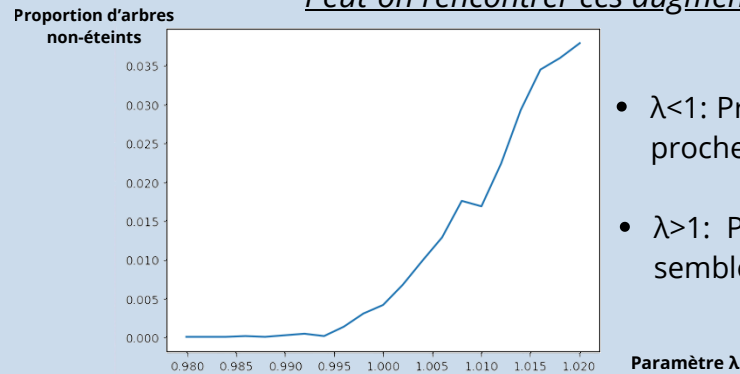


Figure 4 : Représentation graphique de la proportion d'arbres non-éteints après 500 générations en fonction du paramètre λ selon le modèle 1A

- $\lambda < 1$: Proportion faible, proche de 0
- $\lambda > 1$: Proportion élevée, semble exploser

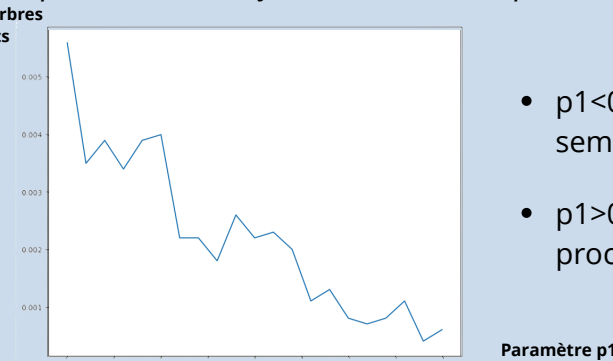


Figure 5: Représentation graphique de la proportion d'arbres non-éteints après 500 générations en fonction du paramètre p_1 selon le modèle 1B

- $p_1 < 0.5$: Proportion élevée, semble exploser
- $p_1 > 0.5$: Proportion faible, proche de 0

Ce qui veut dire que le processus de branchement devrait typiquement être proche du cas critique.

Nécessité de tester la robustesse ainsi que la fiabilité des deux modèles en testant les cas critiques, sous-critiques et sur-critiques.

- **Criticité Modèle 1A:**

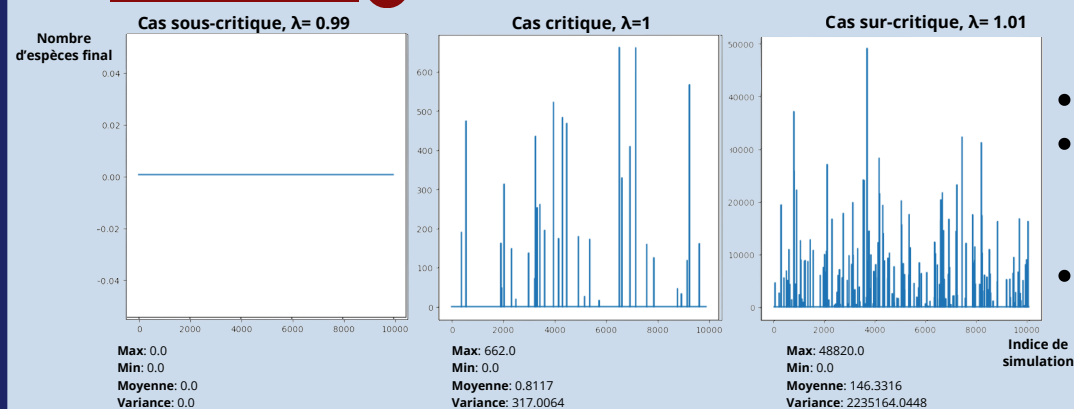


Figure 6: Représentation graphique du modèle 1A selon un cas sous-critique, cas critique, cas sur-critique avec 10 000 simulations de 500 générations.

- Sous-critique: Extinction de la population
- Sur-critique: La probabilité de survie est non-nulle et en cas de survie la taille de la population connaît une croissance exponentielle
- Critique: Cas intermédiaire

- **Paramétrage Modèle 2B:**

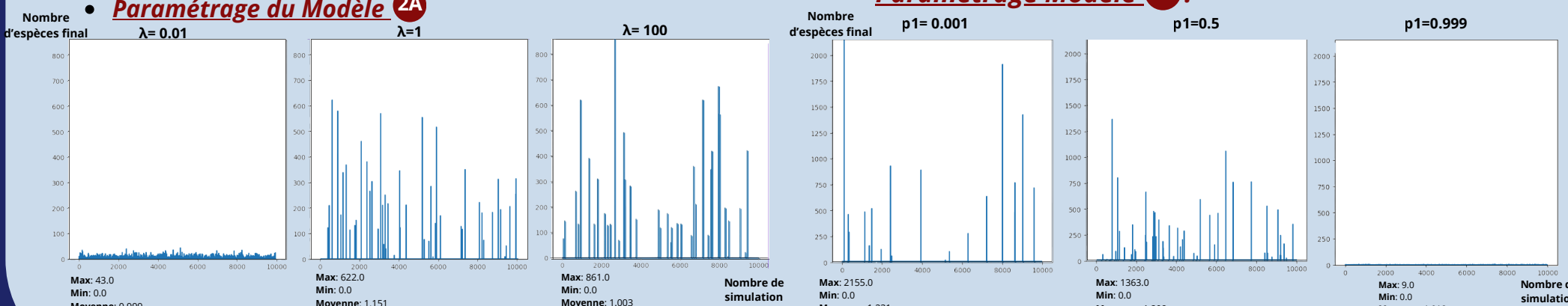


Figure 7: Représentation graphique du modèle 2A pour différents cas critiques avec 10 000 simulations de 500 générations.

Figure 8: Représentation graphique du modèle 2B pour différents cas critiques avec 10 000 simulations de 500 générations.

Discussion

- La prise en compte du temps marque la différence entre notre modèle et celui de Yule. Cela nous impose de définir un pas de temps et un temps de simulation. Cette contrainte est l'une des limitations du paramétrage de notre modèle.
- La distinction du processus de branchement que nous proposons réside dans l'ajout d'une loi qui modélise la mortalité. Cela implique l'existence d'un deuxième paramètre et d'une infinité de cas critiques possibles.
- La comparaison d'une loi de Poisson et d'une loi géométrique pour la génération de nouvelles espèces met en évidence une plus grande instabilité des modèles basés sur une loi géométrique.

Conclusion

- Hors période d'extinction, la diversité des espèces tend à croître légèrement. Un processus basé sur un cas critique n'est pas idéal pour la modélisation sachant que l'apparition de nouvelles espèces est plus fréquente que l'extinction. A noter: Une période d'extinction des espèces peut se modéliser par un cas sous critique.
- Les paramètres optimaux pour notre modèle doivent impliquer une variance assez faible, de manière à générer beaucoup de genres contenant peu d'espèces
- A l'échelle de l'évolution, la diversité des espèces est un équilibre dynamique entre cas sous-critiques et sur-critiques. Il serait donc intéressant de modéliser une alternance de cas sous-critiques et sur-critiques pour mieux approximer la diversification des espèces.

Références/ Liens

- [1] A short history of mathematical population dynamics, Nicolas Bacaër, chapitre 15 : Yule and Evolution
- [2] Stochastic Models and Descriptive Statistics for Phylogenetic Trees, from Yule to Today, 2001, David J. Aldous



Scannez moi !