

Rapport de dernière année de Magister

Personne Arnaud

17 octobre 2016

Table des matières

I	Présentation du parcours	3
1	Introduction	3
2	Remerciements	3
3	Mon Cursus	4
3.1	L3 MFA	4
3.2	M1 MFA	4
3.3	ENS Cachan	5
3.4	M2 MSV	5
3.5	Et maintenant ?	5
II	Présentation d'un sujet de recherche : Modèles mathématiques en écologie des communautés	7
4	Introduction	7
5	De la niche écologique aux modèles neutres.	7
6	Mathématiques des modèles neutres	8
6.1	Le modèle de Hubbell	8
6.1.1	Le modèle	8
6.1.2	Son équation	9
6.2	Les mathématiques pour ce modèle	10
6.2.1	Étude de l'abondance d'une espèce en temps long.	10
6.2.2	La loi du vecteur d'abondance.	11
7	Un modèle neutre avec variance environnementale	14
7.1	Mise en équation	14
7.2	Généalogie dans le cas avec fitness	15
III	Rapports de stages	20
8	Modélisation de la respiration.	20
9	Estimation de la probabilité de rencontre de deux animaux.	20
10	Modèles neutres en écologie.	20

Première partie

Présentation du parcours

1 Introduction

Après trois années de classes préparatoires (Mpsi-Mp) au Lycée Gay-Lussac de Limoges, mon projet était d'obtenir l'agrégation pour être enseignant. Sous les recommandations de mes professeurs, je décidais de m'orienter vers un magister de mathématiques. Ceux-ci étant réputés pour avoir de bons résultats d'admission à ce concours. Ayant postulé à plusieurs d'entre eux, alors que je m'apprêtais à aller passer les oraux de Rennes, j'appris que j'étais reçu au magister d'Orsay. Et c'est ainsi que début septembre 2012, je me suis retrouvé devant le bâtiment 425, un peu perdu mais prêt pour cette nouvelle aventure... Depuis ce jour, quatre années se sont écoulées. Quatre années pleines de souvenirs, de moments difficiles, d'heures de travail, et oui pour réussir il faut travailler ! Mais quatre années faites surtout de rencontres, d'amitiés, de découvertes et de bons moments.

2 Remerciements

Avant de vous parler de mon cursus et pire de mathématiques, j'aimerais prendre quelques lignes pour remercier les personnes sans lesquelles je n'aurais pu arriver jusque là. Comme elles sont très nombreuses, je m'excuse déjà au près de celles que je vais oublier.

Durant ces quatre années j'ai eu la chance de suivre une formation de qualité, et pour cela je remercie l'ensemble de mes enseignants à Orsay pour le temps qu'ils ont pris à m'expliquer et me ré-expliquer les mathématiques, pour les heures passées à corriger mes copies...mais aussi pour leurs disponibilités, leurs sympathies, et leur proximité avec les élèves.

Je tiens aussi à remercier les intervenants du M2 MSV et tout particulièrement Christophe Giraud pour tout le mal qu'il se donne afin que nous réussissions. Un grand merci aussi aux intervenants du magister, à M Paulin et à Nathalie Carrière; ainsi qu'aux professeurs de l'ENS de Cachan et aux membres de la FMJH.

J'ai aussi eu l'occasion comme vous le verrez par la suite d'effectuer des stages. Je remercie vraiment tous mes encadrants : Guy Henniart, Laurent Moonens, Alain Clergerie, Vincent Bensaye, Clément Callenge, Arnaud Guillin et Franck Jabot.

Ensuite, sans soutient, sans amitiés, sans complicité je n'aurais pas pu arriver au bout de ces quatre ans et pour cela, je remercie toute ma promo. Un grand merci à Timothée Tabouy, sans qui je ne ferais pas de thèse, à Justine Samson, Antoine Bera, Yvan Piorun, Vincent Bree, Toufic Boudinard, Mathilde Rale, Victor et Nadège Ruan, Lucile Morel, Margaux Bregere et tous les autres pour

leurs amitiés.

Merci aussi à ma famille pour leur soutien.

Pour terminer, je remercie l'ensemble du personnel de la faculté d'Orsay. Une pensée toute particulière pour Nathalie Carrière, qui, avec sourire et bonne humeur, m'a rendu d'innombrables services durant ces quatre années.

3 Mon Cursus

3.1 L3 MFA

Au premier semestre, j'ai suivi le tronc commun qui regroupait les enseignements d'Algèbre, d'Intégrale de Lebesgue, de Topologie et Calcul Différentiel. Pour compléter cette formation, j'ai choisi les options : Programmation, Algorithmique et Complexité et Mathématiques et Biologie. Étant depuis toujours été intéressé par la biologie, cette dernière option m'a permise de découvrir une partie des applications des mathématiques aux sciences du vivant.

J'ai tout particulièrement aimé au cours de ce premier semestre le cours d'algèbre qui traitait de théorie des groupes. C'est une des premières "belles" choses que j'ai vu en mathématiques.

Au second semestre, j'ai suivi les cours d'Analyse de Fourier, d'Équations Différentielles, de Fonctions Holomorphes et de Probabilités. Pour compléter cette formation, se sont ajoutés le cours d'Algèbre Effective et le Calcul Formel. Le cours spécifique de magistère traitant de la topologie générale est un complément très appréciable au cours de topologie et calcul différentiel du premier semestre. Il permet de renforcer les notions de topologies et celles d'analyses vues précédemment. Par ailleurs, ce cursus comprenait un TER que j'ai effectué avec Vincent Bree sous la direction de Guy Henniart. Il traitait de cas particuliers du théorème de Fermat et reprenait la démonstration théorème de Guy Terjanian .

Enfin au cours de cette première année, j'ai effectué un stage hors murs dans une classe de lycée. Cette expérience m'a conforté dans mon choix d'enseigner.

Cette année, orchestrée par M Hulin, fut avec du recul une année décisive. On ne se spécialise pas encore et on voit un large panel d'objets mathématiques importants pour la suite.

Pour ma part, c'est l'année où j'ai véritablement découvert les mathématiques. Je suis arrivé en L3 avec un niveau très médiocre et j'ai progressé tout au long de l'année jusqu'à obtenir en fin d'année la bourse FMJH.

3.2 M1 MFA

Au vu de mon projet qui était de passer l'agrégation, certains choix de matières étaient indispensables. Ainsi j'ai suivi au premier semestre les cours d'analyse et math-générales, j'ai ensuite choisi le cours de Distribution et le MAO calcul-scientifique. J'ai particulièrement aimé l'excellent cours de Laurent Moonens sur la construction de l'intégrale de Kurzweil et Henstock, ce dernier reprenait l'histoire de ces objets et soulignait les liens entre les différentes

intégrales connues.

Au second semestre, j'ai suivi encore le cours de mathématiques générales ainsi que le cours d'arithmétique de M. Fouvry. Se sont toujours rajoutés à cela les cours du magister à savoir les cours d'Introduction à la Théorie Spectrale et l'Analyse Harmonique et d'Introduction aux Systèmes Dynamiques. Enfin j'ai effectué un TER avec Antoine Bera sous la direction de Laurent Moonens traitant de la mesure de Hausdorff et du Cantor quatre coins.

Cette année fut très enrichissante, je soulignerai l'intérêt tout particulier que j'ai porté aux cours de Laurent Moonens et de M. Fouvry et je ne peux que les recommander. Par ailleurs au cours de cette année j'ai préparé et réussi le concours d'entrée à l'ENS. Je rajouterai que les cours du magister (surtout ceux de topologie et d'analyse harmonique) ne sont pas superflus pour passer cette épreuve.

3.3 ENS Cachan

Une année à Cachan qui se clôture par l'obtention de l'agrégation. Report de stage pour effectuer un M2 recherche.

3.4 M2 MSV

Il fallait faire à un moment donné un choix d'orientation sérieux, sachant que l'algèbre et l'arithmétique, mes matières de prédilection, ont très peu de débouchés. J'ai dû me tourner vers les mathématiques de l'aléatoire et pourquoi pas essayer de faire des mathématiques appliquées. C'est ainsi que je retournais à Orsay suivre le M2 MSV. Étant un peu dubitatif quant à ce choix, je me suis rapidement conforté dans l'idée que c'était le bon. En effet nous étions un petit effectif tous d'horizons différentes (ce qui est très enrichissant), et tous motivés pour réussir au mieux notre année. Par ailleurs les professeurs et le responsable de M2 Christophe Giraud étaient vraiment à notre écoute et prêt à nous aider à la moindre difficulté. En plus des cours, une fois par semaine des conférences sont organisées, ceci nous permet de rencontrer d'autres chercheurs et de voir différents thèmes que l'on peut aborder à la suite de ce M2.

Au cours de cette année j'ai effectué un projet avec Margaux Bregère, sous la direction de Vincent Bensaye et Clément Challenge. Celui-ci traitait du déplacement brownien de deux animaux et de leur probabilité de rencontre. Enfin l'année de M2 se clôturant par un stage de 4 mois, j'ai eu la chance de pouvoir effectuer ce dernier sous la direction de Arnaud Guillin et de Franck Jabot. Il traitait des modèles neutres en écologies.

Cette année fût passionnante, quel plaisir de voir des applications concrètes des mathématiques! Par ailleurs elle m'a permis de découvrir le monde de la recherche, de me confronter à des vrais problèmes de maths parfois sans solution.

3.5 Et maintenant ?

Je commence actuellement une thèse dans le prolongement de mon stage c'est l'occasion de se plonger vraiment dans la recherche, c'est sûrement une

expérience unique et très enrichissante. Je n'oublie pas pour autant mon objectif principal qui était d'enseigner...

En tout cas je suis très heureux d'avoir pu effectuer mes années d'études au sein du magister d'Orsay, sans lui je n'aurais pu obtenir la bourse Hadamard et je ne serais en aucun cas entré à l'ENS . J'ai eu tout simplement la chance de bénéficier d'une formation de qualité, avec des professeurs, des chercheurs, des camarades exceptionnelles. Merci à tous, je suis fier d'avoir choisi cette voie.

Deuxième partie

Présentation d'un sujet de recherche : Modèles mathématiques en écologie des communautés

4 Introduction

Prenez un instant, regardez par la fenêtre ou autour de vous et vous vous apercevrez rapidement de la complexité et de la multiplicité des formes de vie qui nous entourent. Au sein d'un même milieu, plusieurs espèces se disputent les mêmes ressources et chacune a acquis des stratégies différentes pour survivre. Comment peut-on expliquer la diversité et l'abondance des espèces qui nous entourent ?

Cette question intéresse depuis longtemps les hommes, nombreux sont les mythes religieux qui attribuent à chaque espèce une place au sein d'un ensemble harmonieux ; bien des scientifiques tels qu'Aristote ont tenté de classer les organismes selon leurs régimes alimentaires, leurs habitats, leurs comportements...

Aujourd'hui, l'homme a de plus en plus d'influences sur son environnement. Les activités humaines, en modifiant le climat, en introduisant de nouvelles espèces, modifient les écosystèmes. L'écologie (science qui étudie la distribution et l'abondance des organismes ainsi que toutes les interactions qu'ils ont avec leur milieu) devient une science prépondérante.

Dans la suite de cette partie nous nous intéresserons aux communautés. Une communauté est un ensemble d'espèces elles-mêmes constituées d'individus qui interagissent au sein d'un milieu homogène et borné. Nous reprendrons dans un premiers temps les concepts qui ont été au cœur des travaux d'écologie jusqu'à aujourd'hui. Ensuite, nous nous intéresserons aux modèles neutres et aux outils mathématiques que nécessite leur étude. Enfin, nous présenterons un modèle neutre avec variance environnementale qui sera l'objet de mes futures années d'études.

5 De la niche écologique aux modèles neutres.

Lorsque des espèces sont observées dans leur milieu de vie, elles semblent toutes être profondément adaptées à chacun de leur milieu. Darwin apporte la raison de cette observation qu'il nomme sélection naturelle. La théorie de la sélection naturelle permet d'expliquer et de comprendre comment l'environnement influe sur l'évolution des espèces et des populations en sélectionnant les individus les plus adaptés. Par la suite, les écologues ont tenté de chercher les critères environnementaux expliquant la répartition des individus dans une communauté. Ainsi est né le terme de niches. Une niche est un complexe de facteurs écologiques, une place, en raison de laquelle les espèces évoluent et s'excluent.

Par exemple, si on considère une espèce de fauvette, la niche occupée par cette espèce peut être assimilée à un ensemble de points dans un espace à plusieurs dimensions. Chaque dimension représentant une spécificité de la niche : heure d'activité, hauteur du nid, espèces d'insectes dont elle se nourrit, ses prédateurs etc... Chaque point a pour coordonnées la valeur de ces critères pour un individu donné.

Suite à des expériences visant à tester les équations de Lotka-Volterra, les écologues se sont aperçus qu'il y avait impossibilité de coexistence de deux espèces partageant la même niche. Au cours du temps, une espèce finit par exclure l'autre.

Par ce principe, ils expliquent le développement d'une structure de communauté par la sélection des espèces les mieux adaptées dans un environnement hétérogène en relation avec la complexité du milieu.

Si cette théorie largement développée au siècle dernier apporte des réponses probantes, il s'avère que dans certains cas elle ne puisse parvenir à justifier l'existence de populations. Par exemple, certaines forêts tropicales présentent plus de cent espèces d'arbres différentes sur un hectare. Et plusieurs de ces espèces ont clairement des exigences écologiques similaires (partagent la même niche). Par conséquent, la diversité des espèces présentes sur une même niche écologique ne serait pas due simplement à la niche elle-même, mais aussi à la dynamique des individus présents sur cette niche. C'est cette idée là qui est le point de départ des théories neutres en écologie, et en particulier de celle que Hubbell présente en 2001 dans son livre : "The unified neutral theory of biodiversity and biogeography". (cf biblio)

Les théories neutres décrivent une communauté comme un ensemble d'espèces partageant la même niche et où les individus sont en compétition symétrique (tous les individus ont les mêmes chances de prospérer indépendamment de l'espèce à laquelle ils appartiennent) les uns avec les autres. La compétition s'exerce par le maintien de la communauté à un effectif donné (somme nulle).

Cette théorie neutre propose donc en écologie une approche différente de celle suggérée par Darwin, la dynamique de la communauté ne dépend plus de sa composition mais de processus aléatoires. Nous allons maintenant nous pencher sur un modèle neutre particulier et comprendre quelles sont les objets mathématiques nécessaires à son étude.

6 Mathématiques des modèles neutres

6.1 Le modèle de Hubbell

6.1.1 Le modèle

Considérons une communauté dite régionale, composée de J individus d'espèces différentes. A chaque pas de temps (événement), un individu de cette population meurt au hasard et il est remplacé avec probabilité m par le fils d'un autre individu de cette communauté (naissance) ou avec probabilité $1 - m$ par le fils d'un individu provenant d'un réservoir, choisi au hasard de façon uniforme (immigration).

La communauté régionale peut être assimilée à une parcelle d'une forêt, celle que l'on étudie, et le réservoir au reste de la forêt.

On notera que ce réservoir est composé de S espèces (S grand) et n'évolue pas dans le temps. En effet, si sa taille est très grande par rapport à celle de la communauté, son évolution est peu significative par rapport à l'évolution de la communauté. Dans un premier temps, on considère le réservoir comme uniforme, c'est-à-dire que le taux d'immigration de chaque espèce est le même.

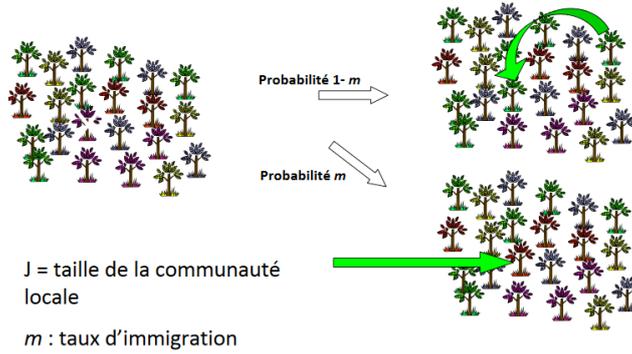


FIGURE 1 – Illustration du modèle.

Contrairement à une population complètement isolée dans laquelle la dérive écologique causerait inévitablement l'extinction de toutes les espèces sauf une (ce qui est le cas dans la théorie des niches), le taux d'immigration à l'échelle locale permet de compenser les extinctions et de maintenir de la diversité dans le système.

6.1.2 Son équation

On se concentre d'abord sur une espèce prise au hasard, on cherche la loi du nombre d'individus qui va la représenter au cours du temps. Cette loi est représentée par un vecteur de taille J ou chaque coordonnée x_j^t représente la probabilité que l'espèce considérée ait j individus au temps t . On va ici détailler les probabilités de transition.

Soit X_i^t la variable aléatoire du nombre d'individus de l'espèce i au temps t :

$$\begin{cases} P(X_i^{t+1} = N_i - 1 | X_i^t = N_i) = \left(\frac{N_i}{J}\right)(m(1 - P_i) + (1 - m)\left(\frac{J - N_i}{J - 1}\right)) \\ P(X_i^{t+1} = N_i + 1 | X_i^t = N_i) = \left(\frac{J - N_i}{J}\right)(mP_i + (1 - m)\left(\frac{N_i}{J - 1}\right)) \\ P(X_i^{t+1} = N_i | X_i^t = N_i) = 1 - P(N_i - 1 | N_i) - P(N_i + 1 | N_i) \end{cases}$$

où P_i est la proportion de l'espèce i dans le réservoir.

6.2 Les mathématiques pour ce modèle

6.2.1 Étude de l'abondance d'une espèce en temps long.

Nous avons ici une chaîne de Markov, X_n à valeurs dans $E = \{1 \dots J\}$, nous aimerions étudier son comportement en temps long. Nous aurons besoin des outils suivants :

Définition 1. On dit qu'un état x est récurrent pour la variable aléatoire X_n si partant de cet état, la probabilité que X_n retourne à cet état avant un temps infini est de 1. Il est dit transitoire dans le cas contraire. On note P la matrice de transition de X_n .

Définition 2. On dit que y est accessible à partir de x (noté $x \rightarrow y$) s'il existe $n \in \mathbb{N}$ tq $P_x(X_n = y) > 0$. On dit que les états communiquent si $x \rightarrow y$ et $y \rightarrow x$

Remarque 1. La relation $x \leftrightarrow y$ est une relation d'équivalence.

Lemme 1. L'état x est récurrent si et seulement si $\sum_{n=0}^{\infty} P_{xx}^n = \infty$

Theoreme 1. Soit $C \subset E$ une classe d'équivalence pour la relation \leftrightarrow . Alors tous les états de C sont soit récurrents soit transitoires.

Définition 3. Une chaîne de Markov est dite irréductible sur un espace E si elle est constituée d'une unique classe d'équivalence. Elle est dite récurrente irréductible si tous ses états sont récurrents et si elle est irréductible.

Proposition 1. Toute chaîne de Markov irréductible sur un espace fini est récurrente irréductible.

Définition 4. On appelle mesure sur E un vecteur ligne dont les coordonnées sont les $\gamma_x, x \in E$ tel que $0 \leq \gamma_x < \infty$ pour tout x . On peut la normaliser pour en faire une probabilité.

Une mesure γ est dite invariante par P si $\gamma P = \gamma$

Remarque 2. La matrice P admet toujours 1 pour valeur propre car c'est une matrice d'une chaîne de Markov. Alors comme la transposée d'une matrice a les mêmes valeurs propres que cette matrice, il existe toujours un vecteur γ tel que $\gamma P = \gamma$. Cependant un tel vecteur est il une mesure, c'est à dire ces coordonnées sont elles toujours positives ou nulles ?

Theoreme 2. Soit $(X_n)_{n \in \mathbb{N}}$ une chaîne de Markov de matrice de transition P récurrente irréductible. Alors il existe une mesure strictement positive invariante unique à une constante multiplicative près. (Cette constante pouvant être choisie pour en faire une probabilité) i.e il existe un unique γ tel que $\gamma P = \gamma$.

Remarque 3. Une telle mesure invariante est un point fixe pour l'application $\phi : \gamma \rightarrow \gamma P$. Si de plus P est une matrice de norme inférieure à 1, ϕ est contractante et le théorème de Picard nous assure la convergence de la suite $(\phi^n(X))$ vers γ , pour tout X

Application 1. Revenons à notre modèle neutre, la variable aléatoire X_n est une chaîne de Markov. Sa matrice de transition P est tridiagonale et les probabilités de transitions sont données dans la partie précédente.

Chaque état est accessible à partir de n'importe quel autre (effet du réservoir) donc P est irréductible et donc récurrente irréductible. Le fait que m ne soit pas nul nous assure qu'aucune de ces probabilités ne vaille 1. Ainsi la norme de P est strictement inférieure à 1. Et l'application $\phi : \gamma \rightarrow \gamma P$ est alors contractante. On a existence d'une unique mesure de probabilité invariante et convergence de la suite $(\phi^n(X))$ vers γ , pour tout X .

Autrement dit pour une espèce, au bout d'un temps long, sa probabilité d'abondance ne dépendra plus de la condition initiale et sera donnée par la mesure de probabilité invariante de P .

6.2.2 La loi du vecteur d'abondance.

Cette approche nous permet d'obtenir la loi d'abondance pour chaque espèce à l'équilibre. Cependant, on ne sait pas calculer la probabilité du vecteur d'abondance. En effet on a la contrainte supplémentaire suivante : $\sum_{i \in \text{especies}} X_n^i = J$.

Pour obtenir cette loi, il est nécessaire de regarder le problème sous un autre angle, celui du coalescent.

On va partir de l'abondance désirée remonter le temps et en déduire une probabilité pour cette abondance.

Partons d'une abondance donnée (c'est à dire une population de J individus connue), le dernier événement arrivé est soit la coalescence de deux individus soit une immigration. Ainsi le dernier individu arrivé vient du pool avec probabilité m ou est le fils d'un individu de la communauté avec probabilité $1 - m$. Pour remonter encore la généalogie, on considère un échantillon de $j - 1$ individus etc...

Plus généralement considérons un échantillon de k individus dans une population de taille J . Le dernier individu arrivé dans l'échantillon (ordre chronologique) peut venir du pool ou est le fils d'un individu de la communauté. Dans le second cas le père est un individu de l'échantillon ou pas. Si il ne l'est pas alors la généalogie à l'étape précédente est toujours composée de k individus, on a simplement remplacé le fils par le père et la situation est inchangée. Ceci peut être itéré jusqu'à ce que un ancêtre de l'individu considéré vienne du pool ou ait subi une coalescence avec un individu de l'échantillon. Ainsi un tel événement ne constitue pas une étape dans l'arbre généalogique. Et les seuls événements capables de modifier la nature de notre échantillon sont la coalescence de deux individus de l'échantillon ou une spéciation.

Proposition 2. Soit un échantillon de k individus parmi une population de J individus. La probabilité que le dernier événement qui se soit produit ait modifié cet échantillon est : $(m - 1) \frac{k(k-1)}{J(J-1)} + m \frac{k}{J}$.

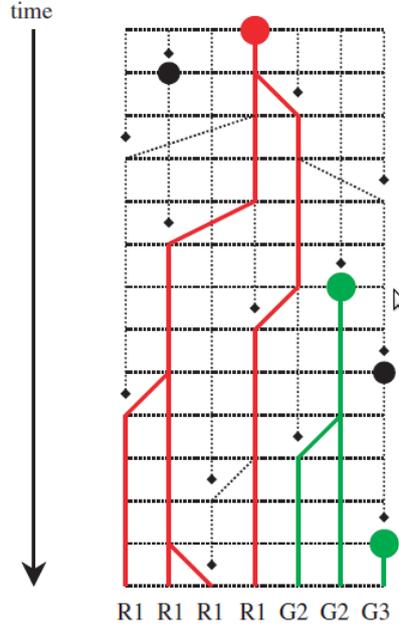


FIGURE 2 – Exemple de coalescent. (d'après, A novel genealogical approach to neutral biodiversity theory Rampal S. Etienne and Han Olf)

Démonstration. Comme on l'a vu précédemment les seuls événements capables de modifier notre échantillon sont une coalescence ou une spéciation (immigration).

- $(m-1) \frac{k(k-1)}{J(J-1)}$ représente la probabilité que le dernier événement soit une filiation et quelle affecte notre échantillon. En effet il faut choisir le père parmi les k individus puis le fils parmi les $k-1$ restants.
- $m \frac{k}{J}$ représente la probabilité que le dernier événement soit une spéciation et qu'elle affecte notre échantillon.

Il reste à sommer ces deux quantités. □

Remarque 4. On notera que la probabilité que le dernier événement qui se soit produit ait modifié notre échantillon correspond à la probabilité qu'un événement modifie l'échantillon. En effet les événements intermédiaires qui n'affectent pas l'échantillon, par neutralité, ne modifient pas cette probabilité.

Proposition 3. *Le dernier événement qui a touché notre échantillon est une coalescence avec probabilité :* $\frac{(m-1) \frac{k(k-1)}{J(J-1)}}{m \frac{k}{J} + (m-1) \frac{k(k-1)}{J(J-1)}} = \frac{\theta}{k-1+\theta}$ *ou une spéciation avec probabilité* $\frac{m \frac{k}{J}}{(m-1) \frac{k(k-1)}{J(J-1)} + m \frac{k}{J}} = \frac{k-1}{k-1+\theta}$ *où* $\theta = \frac{m(J-1)}{m-1}$

Démonstration. Ceci découle de la proposition précédente. □

Remarque 5. Si maintenant on part du plus vieil ancêtre pour aller vers le présent, sur cette lignée chaque événement sera une naissance, ainsi au bout de $k - 1$ étapes la population sera de taille k . Lors de la naissance d'un individu, celui ci sera comme son père avec probabilité $\frac{1}{k+\theta}$ ou de nature différente (viendra du pool) avec probabilité $\frac{\theta}{k+\theta}$ où k est la taille de la population. Le choix du père est uniforme parmi la taille de la population et ces deux événements sont indépendants. Ce processus est ni plus ni moins celui des restaurants chinois arrêté après J étapes.

La formule d'Ewens nous permet d'obtenir la loi du vecteur d'abondance :

Proposition 4. Soit J le nombre d'individus dans la communauté et n_i le nombre d'espèces représentées par i individus alors :

$$P[(n_1, \dots, n_J)] = \frac{J!}{\theta^J} \prod_{i=1}^J \frac{1}{n_i!} \left(\frac{\theta}{i}\right)^{n_i} \quad \text{où} \quad \theta_J = \prod_{i=1}^J (\theta + i - 1)$$

Démonstration. On se donne un vecteur (n_1, \dots, n_s) la probabilité recherchée correspond aux nombres de permutations d'un ensemble de J éléments ayant n_1 cycles de longueurs 1, ... etc multiplié par la probabilité d'avoir une permutation en particulier.

1. On sait par le principe des restaurants chinois que la probabilité d'obtenir une permutation en particulier est $\frac{\theta^{|p|}}{\theta^s}$ ou $|p| = \sum n_i$.
2. On multiplie cette probabilité par le nombre de permutations qui conduisent à la même permutation non ordonnée (on met en relation les permutations ayant les mêmes chiffres dans chacun des cycles de sorte à ce que l'ordre des chiffres dans les cycles n'ait pas d'importance). Ce nombre vaut alors $\prod (i - 1)!^{n_i}$.
Alors $\frac{\theta^{|p|}}{\theta^s} \prod (i - 1)!^{n_i}$ nous donne la probabilité d'obtenir une permutation non ordonnée.
3. Maintenant on calcule le nombre de permutations non ordonnées ayant n_1 cycles de longueurs 1, ... etc Pour cela on choisit n_1 nombres parmi J que l'on range dans n_1 cases puis $2n_2$ parmi $J - n_1$ éléments que l'on range dans n_2 cases de deux éléments sans tenir compte de l'ordre des n_2 cases, etc.... On aboutit au nombre $\frac{J!}{\prod_{j=1}^J j!^{n_j} n_j!}$
4. Enfin la multiplication des deux quantités donne le résultat.

□

Remarque 6. En considérant la vision par coalescent (backward) nous avons pu obtenir la probabilité d'une abondance, ceci n'était à priori pas possible directement (forward).

7 Un modèle neutre avec variance environnementale

On va ici tenter d'intégrer dans notre modèle de base ce phénomène de variance environnementale en attribuant à chaque espèce un poids (fitness) qui lui donne un avantage sélectif.

Par souci de neutralité, on réinitialisera fréquemment les fitness de chaque espèce, de façon aléatoire, de sorte qu'aucune ne soit favorisée ou défavorisée durant une longue période.

Ce modèle a été introduit par Michael Kalyuzhny, Ronen Kadmon and Nadav M. Shnerb dans l'article "A neutral theory with environmental stochasticity explains static and dynamic properties of ecological communities".

On conserve le modèle précédent, à savoir :

- Pool constant à l'équilibre dont l'abondance est celle donnée par les restaurants chinois. ($\theta = 60$, taille très grande)
- Communauté évoluant à taux I , taille J .

On prend en compte le fait que certaines espèces sont plus adaptées à leur environnement, ceci leur confère un avantage sélectif au moment de se reproduire et seulement à ce moment là, le mort étant toujours choisi uniformément :

- Maintenant les individus de la communauté ont une 'fitness'. Cette fitness se représente par un poids noté P_i (pour l'espèce i) propre à chaque espèce et identique pour tous les représentants de la même espèce. Quand le mort est remplacé par le fils d'un individu de la communauté, ce dernier est choisi avec probabilité $F_i^t = \frac{N_i^t P_i}{\sum_{j=1}^S N_j^t P_j}$ où N_i est le nombre d'individus de l'espèce i au temps t .
- Tous les P pas de temps les poids de chaque espèce sont retirés au hasard selon un log-normale(0,s) de façon indépendante. (neutralité en globalité)

7.1 Mise en équation

Donnons nous une espèce i et notons X_i^t la variable aléatoire du nombre d'individus qui représentent cette espèce dans la communauté au temps t , notons encore P_i^k la fitness de l'espèce i au k ème changement. Alors on a :

$$\begin{cases} P_k(X_i^{t+1} = N_i - 1 | X_j^t = N_j, \forall j) = \left(\frac{N_i}{J}\right)(m(1 - R_i) + (1 - m)(1 - F_i^t)) \\ P_k((X_i^{t+1} = N_i + 1 | X_j^t = N_j, \forall j) = \left(\frac{J - N_i}{J}\right)(mR_i + (1 - m)(F_i^t)) \\ P_k((X_i^{t+1} = N_i | X_j^t = N_j, \forall j) = 1 - P_k(X_i^{t+1} = N_i - 1 | X_i^t = N_i) - P_k(X_i^{t+1} = N_i + 1 | X_i^t = N_i) \end{cases}$$

Où R_i est la proportion de l'espèce i dans le pool et $F_i^t = \frac{N_i^t P_i}{\sum_{j=1}^S N_j^t P_j}$

Ce modèle est à première vue assez complexe, en effet les espèces n'étant plus toutes équivalentes, il est nécessaire de connaître l'état de toute la population

pour déterminer les probabilités de transition d'une espèce en particulier.

7.2 Généalogie dans le cas avec fitness

Ici, on suppose, qu'il n'y a plus de changement de fitness, les poids sont fixés pour chaque espèce et n'évoluent pas dans le temps. On cherche alors à calculer la probabilité d'une abondance particulière à l'équilibre.

Pour cela on va avoir, comme précédemment, besoin de reconstruire une généalogie.

Supposons une population de taille J , un échantillon E de taille k connu (k pouvant être égal à J).

Chaque individu i de l'échantillon à un poids P_i qui est le même pour tous les individus de la même espèce.

Le pool est supposé à l'équilibre, les proportions de chaque espèce sont connues.

Demandons nous maintenant quelle est la probabilité que le dernier événement arrivé ait affecté notre échantillon. Seulement deux types d'événements ont pu affecter notre échantillon, il s'agit d'une filiation, c'est à dire la coalescence de deux individus de l'échantillon ou alors d'une spéciation, c'est à dire que le dernier individu arrivé dans E provient d'un père du pool.

Proposition 5.

- La probabilité que le dernier événement soit une spéciation et qu'il ait affecté l'échantillon est :

$$m \times \frac{k}{J}$$

- La probabilité que le dernier événement arrivé soit une filiation de deux individus de l'échantillon (coalescence) est :

$$(1 - m) \times \frac{k}{J} \times \sum_{i \in 1 \dots k} \frac{1}{k} \times \frac{\sum_{j \in E \neq i} P_j}{P_{tot} - P_i}$$

où P_{tot} est le poids de toute la communauté.

Démonstration.

La probabilité de spéciation reste inchangée car les fitness n'influent pas ici. Pour la probabilité de filiation, commençons par choisir les fils potentiels, il y en a $\frac{k}{J}$. Ensuite, pour chaque fils i possible il faut choisir un père celui ci est choisi en fonction de sa fitness et donc avec probabilité $\frac{P_j}{P_{tot} - P_i}$ (pour le père j) où P_{tot} est le poids total de la communauté. Ainsi la probabilité de choisir les couples (fils i , père quelconque) est $\sum_{i \in 1 \dots k} \frac{1}{k} \times \frac{\sum_{j \in E \neq i} P_j}{P_{tot} - P_i}$

Il reste à sommer sur l'ensemble des fils possibles et à multiplier par $(1 - m) \times \frac{k}{J}$ pour obtenir la probabilité qu'il y ait eu filiation. \square

Le problème qui vient à nous maintenant est que l'on ne peut directement obtenir la probabilité du dernier événement affectant l'échantillon. Certains événements qui se produisent ne modifient pas l'échantillon mais le reste de la communauté. En ce sens ils modifient le poids total et donc la probabilité de la nature du dernier événement qui affectera l'échantillon. On doit donc tenir compte des événements intermédiaires qui font varier P_{tot} .

En revanche, si pour des raisons pratiques, on suppose que P_{tot} reste constant au cours du temps (ce qui n'est pas totalement faux si s est petit car les poids sont à peu près les mêmes pour toutes les espèces), on peut obtenir ces probabilités. On supposera que tous les poids sont proches de 1 de sorte que $P_{tot} = J$.

Proposition 6. *Supposons $P_{tot}=J$ constant, la probabilité que le dernier événement qui ait affecté E soit une filiation est $\frac{\frac{k-1}{k} \sum_{j \in E} P_j}{\theta + \frac{k-1}{k} \sum_{j \in E} P_j}$. Et la probabilité que le dernier événement qui ait affecté E soit une spéciation est :*

$$\frac{\theta}{\theta + \frac{k-1}{k} \sum_{j \in E} P_j}$$

Démonstration. Supposons le poids total constant. Considérons notre échantillon de taille k à un instant donné, le dernier événement qui est arrivé a ou non affecté notre échantillon. Si ce n'est pas le cas alors la situation avant cet événement est la même qu'après. En effet les probabilités du dernier événement affectant l'échantillon sont inchangées puisque le poids total reste le même.

La probabilité que le dernier événement ayant affecté l'échantillon soit une filiation est la probabilité que le dernier événement arrivé soit une filiation de deux individus de l'échantillon divisée par l'ensemble des probabilités des événements capables de modifier l'échantillon soit :

$$\begin{aligned} \frac{(1-m) \times \frac{k}{J} \times \sum_{i \in 1 \dots k} \frac{1}{k} \times \frac{\sum_{j \in E \neq i} P_j}{P_{tot} - P_i}}{(1-m) \times \frac{k}{J} \times \sum_{i \in 1 \dots k} \frac{1}{k} \times \frac{\sum_{j \in E \neq i} P_j}{P_{tot} - P_i} + m \times \frac{k}{J}} &= \frac{\frac{k-1}{k} \sum_{j \in E} P_j}{\frac{m(J-1)}{1-m} + \frac{k-1}{k} \sum_{j \in E} P_j} \\ &= \frac{\frac{k-1}{k} \sum_{j \in E} P_j}{\theta + \frac{k-1}{k} \sum_{j \in E} P_j} \end{aligned}$$

(de même pour la spéciation) □

Remarque 7.

— On notera que si les poids sont constants, égaux à 1, alors on retrouve les probabilités du modèle sans fitness.

- En supposant que le poids total soit constant, on suppose le poids du reste de la communauté connu. C'est en effet cette information qui est décisive ici : si le poids de notre échantillon est plus faible que celui du reste de la communauté alors, à chaque fois qu'il y aura filiation avec plus grande probabilité, elle touchera le reste de la communauté et n'affectera pas E. Mais les fitness n'intervenant pas dans la spéciation, l'échantillon E aura alors plus de chance d'être soumis à une spéciation qu'à une filiation. (et inversement)
- Remarquons que sur ce modèle le nombre d'étapes nécessaires pour que E soit affecté par une spéciation reste constant. En revanche le temps pour qu'il soit affecté par une filiation varie en fonction des poids de l'échantillon par rapport au reste de la communauté.

Exemple 1. Un exemple d'arbre :

Voici un exemple de généalogie. On part d'une population (ici 11122) où les individus sont assimilés à leurs poids et on regarde leur généalogie :

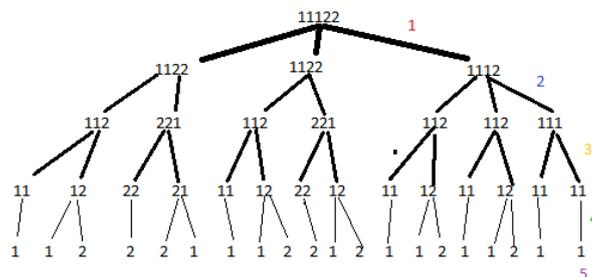


FIGURE 3 – Est représenté l'ensemble des événements qui ont pu conduire à la population 11122. Chaque événement arrive avec une certaine probabilité. Quelques exemples sont donnés ensuite

Valeurs correspondant aux probabilités des branches sur le dessin :

$$1 : \frac{\frac{4}{5} \times 7}{\theta + \frac{4}{5} \times 7} \times \frac{2}{5}, \quad 2 : \frac{\theta_2}{\theta + \frac{3}{4} \times 5}, \quad 3 : \frac{2}{\theta + 2} \times \frac{1}{2}, \quad 4 : \frac{1}{\theta + 1} \times 1, \quad 5 : \frac{\theta_1}{\theta}$$

Références

- Les mondes darwiniens 2011 Arnaud Pocheville.
- David Tilman, 2004, Niche tradeoffs, neutrality, and community structure : A stochastic theory of resource competition, invasion, and community assembly, PNAS vol. 101 no. 30 10854–10861
- Michael Kalyuzhny, Ronen Kadmon and Nadav M. Shnerb, neutral theory with environmental stochasticity explains static and dynamic properties of ecological communities (Ecology Letters,2015)
- Neutral theory and the species abundance distribution : recent developments and prospects for unifying niche and neutral perspectives Thomas J. Matthews and Robert J. Whittaker(Ecology and evolution,2014)
- Processus de Markov et applications 2007 Etienne Pardoux

Troisième partie

Rapports de stages

- 8 Modélisation de la respiration.
- 9 Estimation de la probabilité de rencontre de deux animaux.
- 10 Modèles neutres en écologie.

Modélisation du Poumon

Personne Arnaud

March 27, 2016

1 Introduction

En laboratoire, certaines expériences sur des souriceaux ont conduit aux résultats suivants:

-on observe que même si R (résistance au niveau de la trachée, cf loi de Poiseuille) est grand un petit volume d'air circule entre l'extérieur et les poumons.

-on remarque que dans certains cas il existe un déphasage entre la pression, le volume V du poumon, et l'air qui transite entre le poumon et l'extérieur.

Le modèle de base ne retranscrit pas ces phénomènes nous allons alors devoir complexifier le modèle initial.

2 Le premier modèle

2.1 Présentation du modèle

On va ici rajouter une dimension au problème, on va supposer que l'air est compressible dans le ballon. On introduit alors V_{at} le volume d'air occupé par l'air dans le ballon à la pression ambiante.

Il vient alors $P_{at}V_{at} = P_aV$, (conservation de la matière).

Ensuite la loi de Poiseuille pour V_{at} s'écrit: $RV'_{at} = P_{at} - P_a$. Il manque une relation qui relie la différence de pression, au volume:

$$E(V - V_0) = P_a - P(t)$$

2.2 Résolution

On a donc obtenu le système de trois équations à trois inconnues:

$$\begin{cases} P_{at}V_{at} = P_aV & (1) \\ RV'_{at} = P_{at} - P_a & (2) \\ E(V - V_0) = P_a - P(t) & (3) \end{cases}$$

Après simplifications ce système est équivalent à:

$$\left\{ \begin{array}{l} V' = \frac{P_{at}}{R}(V - V_0) + \frac{P_{at}^2 - P_{at}P}{2V - V_0 + \frac{P(t)}{E}} \quad (1) \\ V_{at}' = \frac{P_{at} - P_a}{R} \quad (2) \\ P_a = P(t) + E(V - V_0) \quad (3) \end{array} \right.$$

On notera que la solution $V(t) = V_0 + \frac{P_{at} - P(t)}{E}$ est solution particulière constante par morceaux. Ne pouvant résoudre directement cette équation on se propose de la résoudre numériquement:

2.3 Résultats numériques

On va tracer tout d'abord les solutions obtenues ainsi que la solution constante:

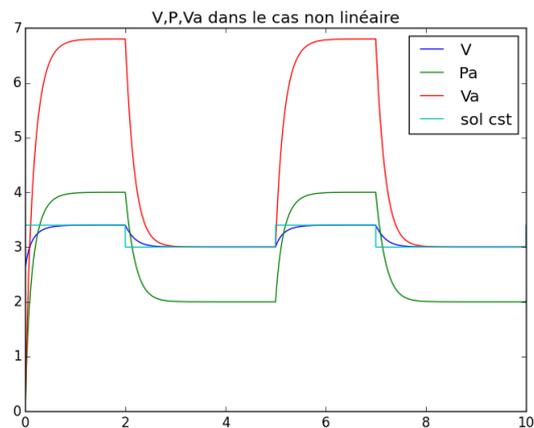


Figure 1: tracés des solutions

On va maintenant regarder l'influence de R sur les différentes variables ici R vaut 20:

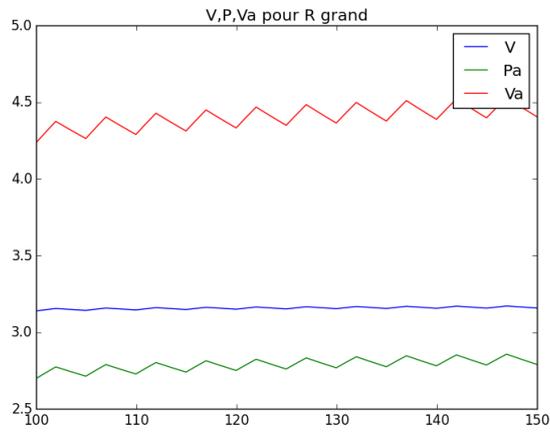


Figure 2: Pour R grand

On constate que le volume V reste quasiment constant alors Vat et Pa évoluent simultanément. Vat est non nul est oscille toujours. Cette variation se traduit par un flux d'air sortant et rentrant.

3 Vers un autre modèle

3.1 Observation

Des expériences menées sur des souris en laboratoire montrent que en régime forcé, quand R est grand on obtient un déphasage entre Vat et le Volume V et la pression.

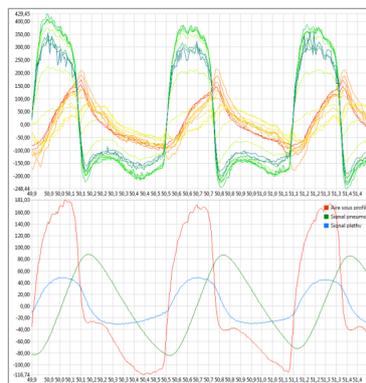


Figure 3: Valeurs mesurées chez un souriceau en temps réel

Or Dans notre modèle on ne met en évidence aucun déphasage quelques soient les valeurs de R.
 Il va falloir que nous considérons un nouveau modèle.

3.2 mise en évidence d'un déphasage

Pour commence, on va essayer de comprendre ce qui peut conduire à un déphasage. Pour cela on va considérer que notre volume V du ballon est imposé et évolue selon une sinusoïde.

La relation entre le volume et la pression via l'élastance n'a plus lieux d'être car le volume est imposé. Afin de simplifier le modèle on va supposer que comme R est grand on a à l'intérieur du ballon $P_a V = cst = P_{atm} V_0$ on conserve toujours la loi de Poiseuille, on suppose qu'un petit volume d'air s'introduit dans le ballon et que se volume est négligeable par rapport à V au cours du temps. Le nouveau système est:

$$\begin{cases} P_a V = cst = P_{atm} V_0 & (1) \\ V'_{at} = \frac{P_{at} - P_a}{R} & (2) \end{cases}$$

Via cette équation, un petit raisonnement va nous permettre d'intuiter le résultat:

$$\text{On a } P_a = \frac{cst}{V} = \frac{cst}{V - V_0 + V_0}$$

En supposant que les variations de volumes sont pas très grandes, ce que l'on peut éventuellement imposer, on obtient via un DL à l'ordre 1:

$$P_a = \frac{V - V_0}{R V_0}$$

Qui est une équation linéaire en V alors si V est sinusoïdale centrée en V_0 , on a P_a qui l'est aussi à un facteur multiplicatif près.

Ensuite l'équation: $V'_{at} = \frac{P_{at} - P_a}{R}$ nous dit que V_{at} va se comporter comme une sinusoïde déphasée par rapport à V et P_a

Pour vérifier cela, une étude numérique s'impose: Voici les résultats:

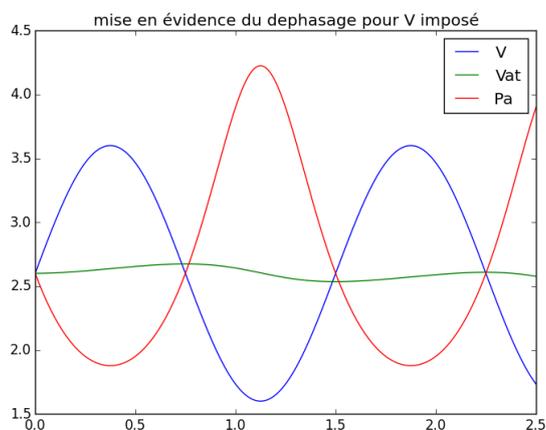


Figure 4: déphasage pour le modèle simplifié

Comme on le voit sur ce graphique, on a réussi à mettre en évidence un déphasage entre V_{at} et V , ceci est conforme à ce que nous avons imaginé.

Cependant ce modèle est trop simplifié en réalité, on ne peut supposer que $V P_a$ reste constant. Et l'influence de R ne peut se lire clairement sur ce modèle, on va devoir en construire un plus réaliste.

4 Un modèle plus réaliste

4.1 Mise en équation

On impose toujours V et on va supposer maintenant que notre quantité de matière dans le ballon varie au cours du temps, on introduit alors la variable n , on écrit la relation $PV = nRT$ abrégée ici en $Pv = n$ car T est supposée constante.

De plus la variation de matière, n' , dans le ballon est exactement $P_{at} V_{at}$. On obtient donc le système d'équations:

$$\begin{cases} P_a V = n & (1) \\ V_{at}' = \frac{P_{at} - P_a}{R} & (2) \\ P_{at} V_{at}' = n' & (3) \end{cases}$$

En simplifiant ce système, on se ramène après calculs au système suivant:

$$\begin{cases} P'_a + \frac{P_a(V'R + P_{at})}{VR} = \frac{P_{at}^2}{RV} & (1) \\ V'_{at} = \frac{P_{at} - P_a}{R} & (2) \\ P_a V = n & (3) \end{cases}$$

Il nous reste plus qu'à étudier ces équations. A partir de la première on calcule la pression P_a puis on en déduit V_{at} et n .

4.2 résultats

On a ici résolu les équations précédentes numériquement voici les résultats obtenus pour différentes valeurs de R :

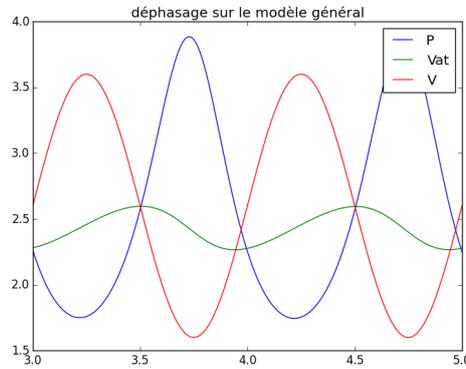


Figure 5: déphasage pour R grand

On constate que pour un R grand (2), les résultats obtenus sont similaires au modèle critique que nous avons considéré précédemment en faisant l'hypothèse que $P_a V$ reste constant. On a P_a qui est en phase avec V alors que V_{at} est déphasé. On note que c'est essentiellement la variation de pression qui compense

cette augmentation de volume, c'est ce qui explique que P et V sont en phase. D'autre part l'amplitude de V_{at} est relativement faible ce qui est normal puisque peu d'air circule entre l'extérieur et le ballon vu que R est élevé.

Remarquons aussi que même pour R grand les variations de V_{at} sont non nulles il y a toujours circulation d'une faible quantité d'air entre le ballon et l'extérieur.

Regardons ce qui se passe pour un R plus petit (0.6):

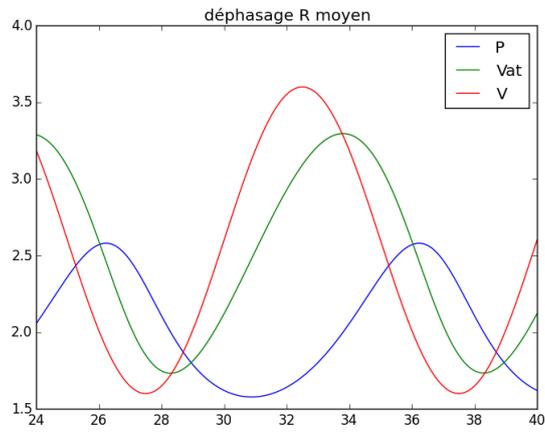


Figure 6: déphasage pour R intermédiaire

On observe ici une figure proche de celle que nous avons expérimentalement. Les trois variables sont toutes déphasées. Maintenant, une augmentation de volume entraîne non seulement une baisse de pression mais aussi une entrée d'air. Ces deux phénomènes sont liés pour compenser cette variation de volume et ceci entraîne bien un déphasage des trois variables.

Regardons ce qui se passe pour un R petit :

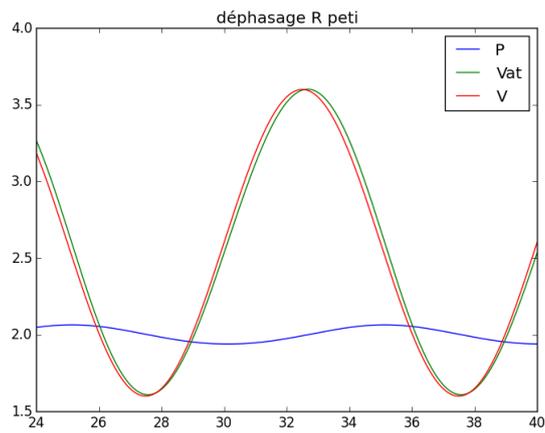


Figure 7: déphasage pour R petit

Enfin si on prend un R très petit (0.1), on observe que V et V_{at} sont en phase et que la variation de pression est très faible.

En effet si on impose très peu de résistance une augmentation de V va être compensée par une entrée d'air ie une augmentation de V_{at} . Et la pression n'aura pas de raison d'augmenter.

Ceci montre aussi qu'il est plus facile de déplacer des volumes d'air quand il y a peu de contraintes que de faire varier la pression.

5 Temps de relaxation et lien avec la période de la sinusoïde

Dans cette partie on cherche à trouver une condition sur R et la période de la sinusoïde V qui pourrait caractériser les régimes observés.

5.1 Temps de relaxation

On va estimer ici le temps caractéristique pour qu'une variation de Volume V soit compensée par une variation de pression sous les équations du modèle précédent ie $P'_a + \frac{P_a(V'_R + P_{at})}{VR} = \frac{P_{at}^2}{RV}$

Cette étude peut se faire directement à partir du modèle, on imposera ici une variation sinusoïdale du volume d'amplitude 1 ($V = \sin(g * t)$), puis quand le volume reste constant à 1 on regarde le temps nécessaire pour que la pression compense cette augmentation ie qu'on retrouve un équilibre.

Supposons que le volume ait subi une augmentation de 1 (qui est la variation moyenne qu'il subit car sinusoïde d'amplitude 1) sur un temps correspondant à un quart de la période. Puis qu'il reste constant égal à $V_0 + 1$. alors durant cette période la pression subit une variation estimée à $P_{at} - P_o$ où P_o est la nouvelle valeur de la pression au bout de ce temps.

Maintenant comme V est constant, l'équation qui régit la pression s'écrit

$$P'_a + \frac{P_a P_{at}}{VR} = \frac{P_{at}^2}{RV}$$

en notant $c = \frac{P_{at}}{VR}$, cette équation se résout et a pour solution:

$$(P_o - P_{at}) \exp(-ct) + P_{at}$$

La solution converge vers P_{at} , la tangente en 0 coupe l'asymptote au bout d'un temps $\tau = \frac{VR}{P_{at} - P_o}$

Pour calculer τ on aura besoin d'estimer P_o numériquement. (ceci est fait par un algorithme).

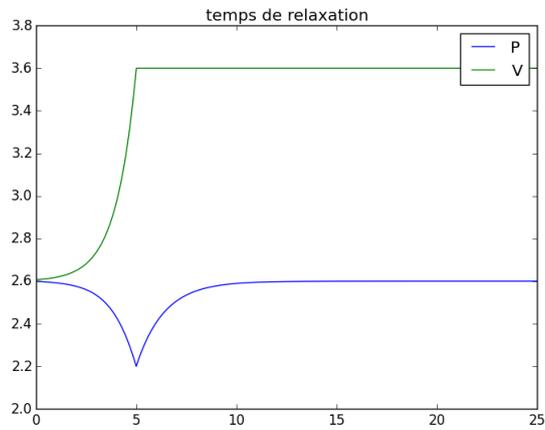


Figure 8: un exemple du temps de relaxation

On va maintenant faire varier R et comparer τ à la période.

5.2 Explication des différents états observés

Il est raisonnable de penser que si τ est grand devant la période alors ce n'est pas la pression qui compensera la variation de volume mais l'air qui rentrera dans le ballon, et l'inverse si τ est petit devant la période. Et si τ est environ de l'ordre de grandeur de V on a l'état intermédiaire décrit précédemment.

On va fixer la période à 10 et faire varier R . On a ici $P_{at} = 2.6 = V_o, P_o = 2, \tau = 27$

On observe:

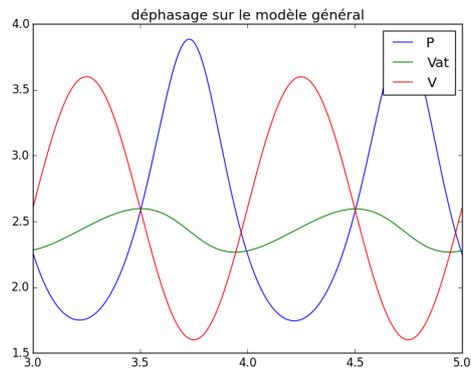


Figure 9: Taux est grand devant la période

On a ici $P_{at} = 2.6 = V_o, P_o = 2, \tau = 10.2$ On observe:

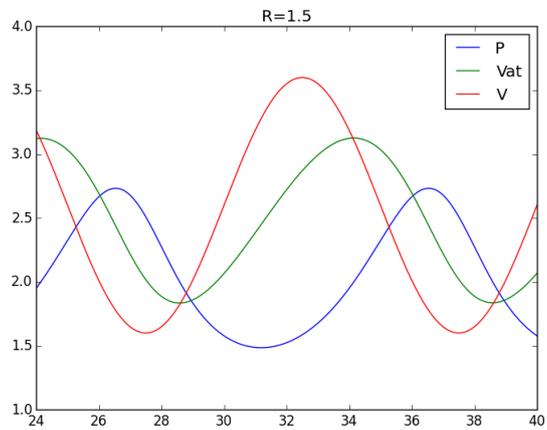


Figure 10: Taux est de l'ordre de la période

On a ici $P_{at} = 2.6 = V_o, P_o = 2.3, \tau = 5$ On observe:

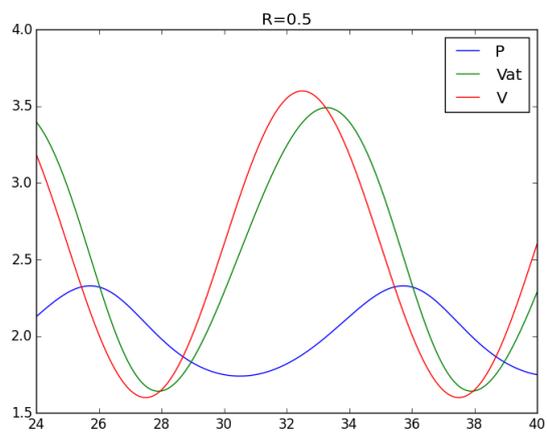


Figure 11: Taux est petit devant la période

Ainsi ce qui détermine l'état dans lequel on va se trouver est la valeur de τ qu'il faut comparer à la période. Notons que τ dépend linéairement de R .

6 Vers un modèle plus général

Maintenant que l'on a vu que ce modèle mettait en avant un déphasage, on peut faire en sorte de ne plus imposer le volume.

On garde alors le système précédent auquel on rajoute une équation liant V à P_a via l'élastance E . En imposant une pression $P(t)$ sur l'extérieur du poumon on obtient le système suivant:

$$\begin{cases} P'_a + \frac{P_a(V'R + P_{at})}{VR} = \frac{P_{at}^2}{RV} & (1) \\ V'_{at} = \frac{P_{at} - P_a}{R} & (2) \\ P_a V = n & (3) \\ P_a - P(t) = E(V - V_0) & (3) \end{cases}$$

Il resterait à résoudre numériquement ce problème et regarder si les résultats obtenus sont proches de ceux que l'on observe...

7 Conclusion

Nous sommes partis d'observations réalisées en laboratoire et nous avons tenté de construire un modèle capable d'expliquer ces observations. Nous avons construit un modèle très simplifié qui nous permettait de mettre en avant le déphasage puis nous l'avons complexifié pour nous rapprocher de la réalité. Ainsi nous avons mis en avant trois états différents et nous avons tenté d'expliquer quels étaient les facteurs qui nous permettaient de nous ramener à chacun de ces états. Nous avons vu que le fait de faire varier R influait linéairement sur τ et que ceci conduisait à différents états.

Enfin cette étude peut être approfondie il reste encore à étudier le modèle présenté en 6. On peut aussi essayer d'estimer P_0 par le calcul, il faut penser que P_0 dépend de R et donc τ n'est pas complètement linéaire en R .

ESTIMATION D'UNE PROBABILITÉ DE RENCONTRE ENTRE ANIMAUX
ENCADRÉ PAR VINCENT BANSAYE ET CLÉMENT CALENGE

Mars 2016

ARNAUD PERSONNE ET MARGAUX BRÉGÈRE

Table des matières

Introduction	2
1 Pont brownien en Dimension 1	3
1.1 Présentation du pont brownien	3
1.2 Calcul explicite et simulations en dimension 1	4
2 Confrontation au problème de dimension 2	6
2.1 Symétries du problème	6
2.1.1 Variables spatiales	6
2.1.2 Lien entre la longueur l et le coefficient de diffusion D	7
2.2 Calcul en dimension 2 non résolu	8
2.3 Estimateurs et autres approches	8
2.3.1 Approximations du temps de rencontre	8
2.3.2 Processus de Bessel de dimension 2	11
3 Algorithmes & Résultats	12
3.1 Simulation de ponts browniens	12
3.2 Algorithmes : approximations du temps de rencontre moyen et de la probabilité de rencontre	12
3.2.1 Probabilité de rencontre	12
3.2.2 Temps moyen de rencontre	13
3.3 Influence du paramètre de diffusion	14
3.4 Confrontation aux données	16
Conclusion	18
Références	18

Introduction

Les motivations de ce projet étaient le calcul ou l'approximation de probabilités de rencontre entre deux animaux. Plus précisément, nous nous sommes intéressés aux animaux de la forêt, tels que les chevreuils, les sangliers, les biches et les cerfs. Après avoir discuté avec Clément Calenge, à l'ONCFS (Organisation Nationale de la Chasse et de la Faune Sauvage), nous avons pu considérer un modèle adéquat à ce problème. Cette organisation dispose de données GPS grâce à des balises fixées sur certains animaux, qui émettent toutes les 20 minutes environ. Ainsi nous connaissons la position quasiment exacte de l'animal en ces instants.

Grâce à deux relevés GPS de deux animaux, comment approcher la probabilité que ces deux animaux se rencontrent ?

Entre deux positions GPS, nous allons modéliser le déplacement de l'animal par un mouvement brownien. C'est ce qu'on appelle un pont brownien, qui est caractérisé par une position de départ, une position d'arrivée, un temps, ou longueur l (temps au bout duquel l'animal arrive à sa position finale, ici on aurait donc $l \approx 20 \text{ min}$) et un coefficient de diffusion D . Ainsi, la trajectoire de l'animal suit une direction déterministe : la droite qui joint les points de départ et d'arrivée, et s'en éloigne de façon stochastique grâce au brownien de coefficient de diffusion D .

Astreindre le mouvement brownien à être au point d'arrivée au temps l est très contraignant, et comme nous avait prévenus Vincent Bansaye, qui nous a aidé sur les aspects théoriques du problème, les calculs sur le pont brownien sont très compliqués. Si le cas de la dimension 1 est compris et permet d'avoir des formules explicites, l'ajout d'une dimension complique énormément le problème. En effet, dans le cas de la dimension 1 si deux trajectoires de ponts browniens se coupent, alors forcément les animaux se rencontrent. Ce n'est plus du tout le cas en dimension supérieure et c'est ce qui fait toute la difficulté du projet.

Aussi, après avoir présenté plus en détails les propriétés du pont brownien et ainsi que le cas simple de la dimension 1, nous expliciterons les problèmes rencontrés en dimensions 2 et formulerons quelques alternatives. Enfin, nous présenterons les simulations qui nous ont permis de confronter théories, approximations et algorithmes.



1 Pont brownien en Dimension 1

Le pont brownien peut être défini de différentes manières, mais de la même manière que le mouvement brownien, c'est un processus gaussien caractérisé par son espérance et sa variance. Nous verrons différentes propriétés de ce processus qui permettront de calculer la probabilité que deux animaux ayant des trajectoires de ponts browniens se rencontrent, c'est à dire sont distants l'un de l'autre de moins de ϵ . Afin de bien comprendre les objets à manipuler et de se faire une idée des résultats, nous avons ainsi commencé par traiter le cas simple de la dimension 1. D'un point de vu biologique, on pourrait imaginer des poissons le long de la barrière de corail ou des araignées d'eau sur une rivière.

1.1 Présentation du pont brownien

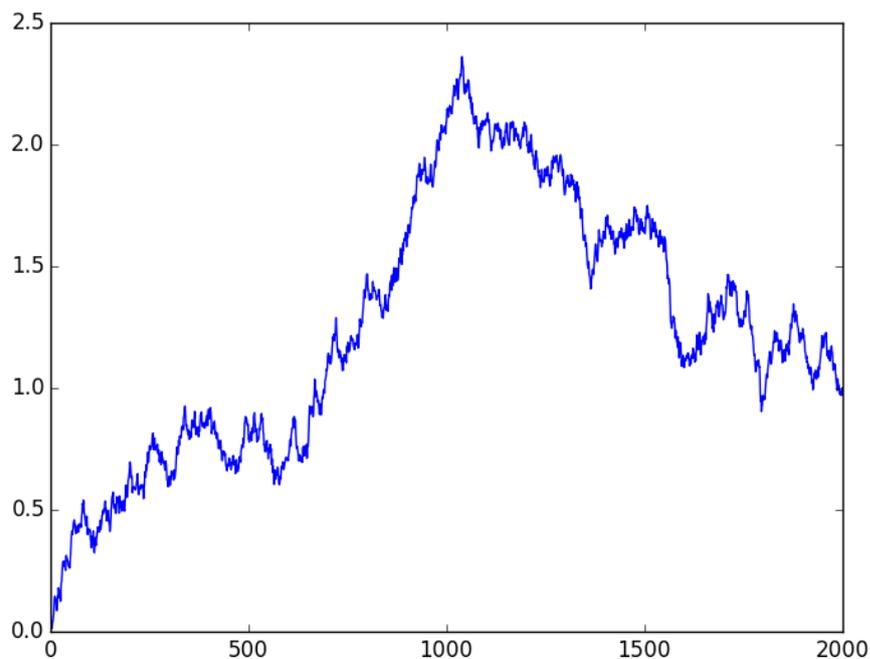
Un pont brownien issu de x , arrivant en y , de longueur l et de coefficient de diffusion D est un processus gaussien continu $B = \{B_t : 0 \leq t \leq l\}$, $B_0 = x$, tel que :

$$\mathbb{E}(B_t) = x + (y - x)\frac{t}{l}, \quad \mathbf{Cov}(B_t, B_s) = D^2[s \wedge t - \frac{st}{l}]$$

Il existe deux représentations équivalentes du pont brownien, que l'on définit à partir d'un mouvement brownien standard $(W_t)_t$:

$$B_t = x + (y - x)\frac{t}{l} + D(W_t - \frac{t}{l}W_l)$$

$$B_t = x + (y - x)\frac{t}{l} + \frac{l-t}{l}DW_{\frac{lt}{l-t}}$$



brownien.png

FIGURE 1 – Simulation d'une trajectoire de pont brownien en dimension 1 issu de 0 qui arrive en 1 au bout d'un temps 1

Un processus vérifiant l'équation différentielle stochastique suivante est aussi un pont brownien issu de x et arrivant en y au bout d'un temps l (Ceci est démontré dans le livre *Stochastic Differential Equations and Diffusion Processes*, page 229 [1]). Mais nous utiliserons peu cette dernière représentation.

$$\begin{cases} dB_t &= \frac{(y-B_t)l}{l-t}dt + DdW_t \\ B_0 &= x \end{cases}$$

Une propriété remarquable de ce processus est que la somme de deux ponts browniens indépendants de même longueur l est toujours un pont brownien. En effet si on considère $B_t^1 = x_1 + (y_1 - x_1)\frac{t}{l} + D(W_t^1 - \frac{t}{l}W_l^1)$ et $B_t^2 = x_2 + (y_2 - x_2)\frac{t}{l} + D(W_t^2 - \frac{t}{l}W_l^2)$, alors

$$B_t = B_t^1 + B_t^2 = (x_1 + x_2) + (y_1 + y_2 - (x_1 + x_2))\frac{t}{l} + D(W_t^1 + W_t^2 - \frac{t}{l}(W_l^1 + W_l^2))$$

Qui est bien un pont brownien issu de $(x_1 + x_2)$ et qui arrive en $(y_1 + y_2)$ au bout d'un temps l car $W_t = W_t^1 + W_t^2$ est bien un mouvement brownien.

1.2 Calcul explicite et simulations en dimension 1

Afin de calculer la probabilité que deux animaux se rencontrent, nous considérons la différence entre leur trajectoire : Ainsi, si le premier animal suit un pont brownien issu de x_1 , qui arrive en y_1 au bout d'un temps l , et en même temps un second animal qui part en x_2 et arrive en y_2 au bout du même temps l , et ce selon un second pont brownien indépendant du premier, nous posons $(B_t)_{0 \leq t \leq l}$ la différence entre les deux browniens. $(B_t)_{0 \leq t \leq l}$ est aussi un pont brownien issu de $x = x_1 - x_2$, qui arrive en $y = y_1 - y_2$ au bout d'un temps l . Quitte à intervertir les deux animaux, nous considérerons $y > 0$. Nous allons chercher la probabilité \mathbb{P}_R que les deux animaux se rencontrent, c'est à dire qu'il existe un temps t entre 0 et l tel que B_t soit inférieur à une distance fixée ϵ , qui caractérise la rencontre.

$$\mathbb{P}_R = 1 - \mathbb{P}(\inf_{0 \leq t \leq l} B_t \geq \epsilon).$$

Lemme 1. Soient $(a, b) \in \mathbb{R}^{+2}$, et $(W_t)_{t \geq 0}$ un mouvement brownien, alors

$$\mathbb{P}(W_t \geq -at - b, \forall 0 \leq t \leq +\infty) = 1 - e^{-2ab}$$

Preuve. Considérons le mouvement brownien $(X_t)_{t \geq 0}$ issu de 0 avec un drift de $-a$:

$$dX_t = adt + dW_t$$

Les fonctions f et g , C^2 bornées, sont telles que $Af = 0$ et $Ag = -1$, où A est le générateur du processus : $\forall h, Ah = +ah' + \frac{1}{2}h''$, avec :

$$f(x) = e^{-2ax} \text{ et } g(x) = -\frac{x}{a}.$$

L'existence de g assure que $\forall b \in \mathbb{R}^+, \forall n \in \mathbb{N}, \mathbb{P}(T_{-b} \wedge T_n \leq +\infty) = 1$. Et de plus,

$$\mathbb{P}(T_n < T_{-b}) = \frac{f(0) - f(-b)}{f(n) - f(-b)} = \frac{1 - e^{2ab}}{e^{-2an} - e^{2ab}} \xrightarrow{n \rightarrow \infty} 1 - e^{-2ab}$$

Lorsque $n \rightarrow \infty$, comme le processus est continu, nous avons forcément que $T_n \rightarrow \infty$, en réalité (X_t) suit le drift qui l'entraîne vers $-\infty$. Ainsi, cette probabilité correspond bien au fait que (X_t) reste à tout instant au dessus de b , i.e. $\mathbb{P}(W_t + at \leq -b, \forall 0 \leq t \leq +\infty) = 1 - e^{-2ab}$. \square

Considérons maintenant le pont brownien $(B_t)_{0 \leq t \leq l}$ issu de x qui arrive en y au bout d'un temps l , avec D le coefficient de diffusion :

Avec $(W_t)_{t \geq 0}$ un mouvement brownien standard, $B_t = x + D \frac{l-t}{l} W_{\frac{lt}{l-t}} + (y-x) \frac{t}{l}$

$$\begin{aligned} \{ \inf_{0 \leq t \leq l} B_t \geq \epsilon \} &= \{ W_{\frac{lt}{l-t}} \geq \frac{l}{D(l-t)} (\epsilon - x - (y-x) \frac{t}{l}), \forall 0 \leq t \leq l \} \\ &= \{ W_s \geq s (\frac{\epsilon - y}{Dl}) + \frac{\epsilon - x}{D}, \forall 0 \leq s \leq \infty \}, \text{ avec le changement } s = \frac{lt}{l-t} \xrightarrow{t \rightarrow l} \infty. \end{aligned}$$

En utilisant le lemme, avec $-a = \frac{\epsilon - y}{Dl}$, qui est bien négatif pour ϵ assez petit, et $-b = \frac{\epsilon - x}{D}$, négatif aussi pour le cas intéressant, nous avons immédiatement que :

$$\mathbb{P}(\inf_{0 \leq t \leq l} B_t \leq \epsilon) = 1 - e^{-2 \frac{(\epsilon - x)(\epsilon - y)}{lD^2}}$$

Ainsi la probabilité de rencontre est :

$$\mathbb{P}_R = \min(e^{-2 \frac{(\epsilon - x)(\epsilon - y)}{lD^2}}, 1).$$

En effet $e^{-2 \frac{(\epsilon - x)(\epsilon - y)}{lD^2}}$ est supérieur à 1 si $(\epsilon - x)$ et $(\epsilon - y)$ sont de signes opposés (i.e $b > 0$), c'est à dire (pour ϵ assez petit) que les trajectoires des ponts browniens, qui sont continuent se coupent presque sûrement. Effectivement, si x et y sont de signes opposés, ou si ϵ est trop grand et qu'il englobe déjà x ou y , la trajectoire de la différence entre les deux ponts browniens passe en 0, et il y a donc rencontre. Pour avoir la probabilité de rencontre dans le cas général, il faut donc prendre le minimum entre la probabilité calculée grâce au lemme et 1.

Notons que cette probabilité dépend de façon symétrique de x et de y , ce qui est évidemment tout à fait logique : le processus qui part de x pour arriver en y est le même que celui qui part de y pour arriver en x . Un résultat plus intéressant est la façon dont \mathbb{P}_R dépend du coefficient de diffusion. La probabilité de rencontre augmente avec D car les animaux ont des trajectoires beaucoup plus diffuses qui s'éloignent de la droite déterministe. Le temps l a le même effet que D^2 , nous verrons par la suite que ces deux paramètres sont très liés. Enfin, cette probabilité de rencontre sera très importante pour confronter nos simulations à la théorie.

2 Confrontation au problème de dimension 2

On cherche à adapter le modèle en dimension 2, c'est à dire à modéliser le mouvement d'une biche ou d'un sanglier allant de $D1[x_D, y_D]$ à $A1[x_A, y_A]$ (pour Départ et Arrivée). Il nous faut pour cela définir un objet mathématique capable de traduire cette trajectoire aléatoire. On choisit pour cela un pont Brownien en deux dimensions, c'est à dire un vecteur ayant pour coordonnées (X, Y) où X et Y sont deux ponts browniens de dimension 1 indépendants allant de x_D à x_A et de y_D à y_A . On s'intéresse alors à la rencontre de deux biches qui évolueraient dans un plan durant un temps T . Notons $P1$ et $P2$ ces deux animaux. Comme précédemment on considère deux ponts browniens $(B_t^1)_t$ et $(B_t^2)_t$ issus de deux points $D1$ et $D2$ et arrivant en deux autres points $A1$ et $A2$. On va tenter de calculer la probabilité de rencontre de ces deux ponts browniens c'est à dire la probabilité qu'ils soient à distance inférieure à ϵ l'un de l'autre.

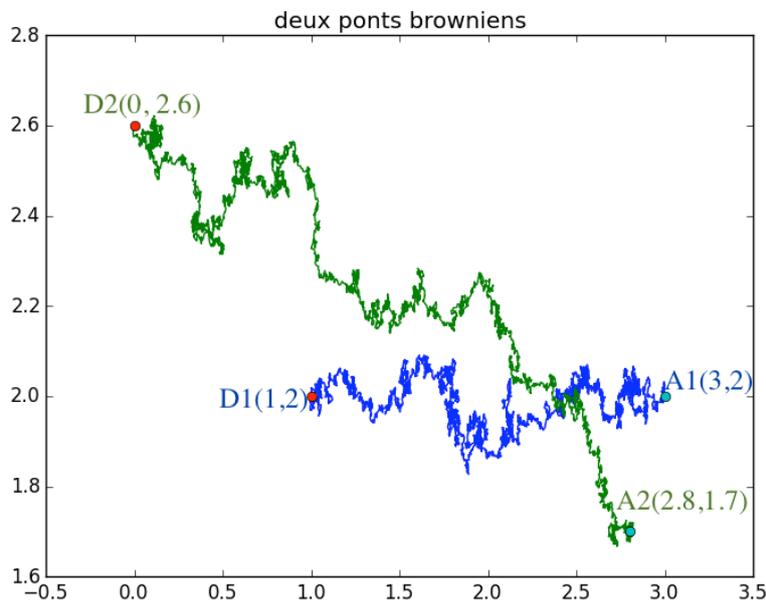


FIGURE 2 – Simulation de deux ponts issus de $D1$ et $D2$ et arrivant en $A1$ et $A2$.

2.1 Symétries du problème

2.1.1 Variables spatiales

On va chercher dans cette partie à réduire le nombre de paramètres notamment en se ramenant à 3 paramètres d'espaces au lieu de 8 : les coordonnées de $D1$, $D2$, $A1$, $A2$. Nous allons procéder comme suit :

- Le problème étant invariant par translation, nous allons fixer la position initiale de premier animal comme origine du repère et ainsi éliminer deux variables (les coordonnées de $D1$)
- Le problème est aussi invariant par rotation. On ne perd rien à imposer que $D2$, la position de départ du second animal soit sur l'axe des ordonnées. On élimine ainsi la seconde coordonnées de $D2$ en la fixant à 0 sans perte de généralité.

- Enfin, la différence entre deux ponts browniens reste un pont brownien. Le problème revient donc à considérer l'unique pont brownien de dimension 2 issu de $D(x_{D2}, 0)$ et qui arrive en $A(x_{A1} - x_{A2}, y_{A1} - y_{A2})$. On notera dorénavant les coordonnées de $D(x_d, 0)$ et de $A(x_a, y_a)$.

Nous nous sommes bien ramenés à 3 variables spatiales : x_d , x_a et y_a . Et la problématique est maintenant de calculer la probabilité que la norme du pont brownien issu de D et arrivant en A , devient plus petite que la distance de rencontre ϵ . Notons que cette diminution des variables simplifiera la complexité des algorithmes : au lieu de simuler deux ponts browniens on en simulera qu'un. Par la suite on considèrera un unique pont brownien.

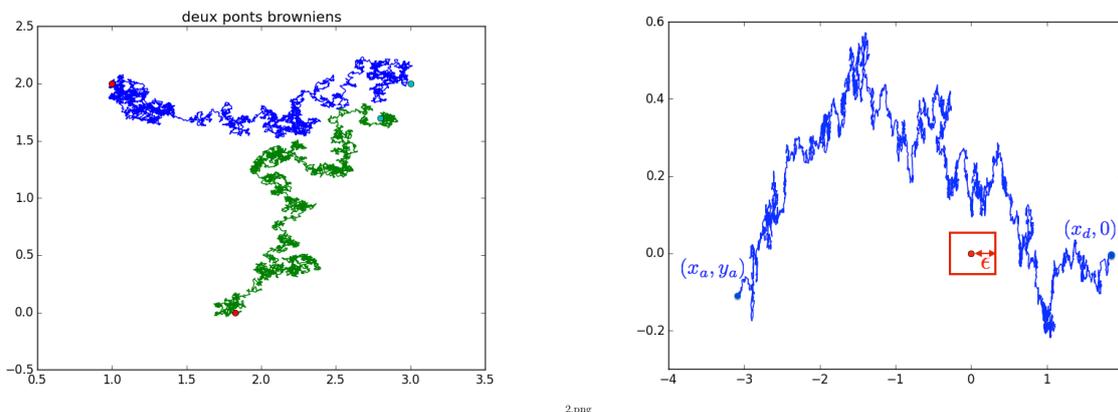


FIGURE 3 – Transformation du problème à deux ponts browniens à celui à un pont issu de $(x_d, 0)$ qui arrive en (x_a, y_a) . La rencontre se caractérise par la rentrée du pont dans le carré centré en $(0, 0)$ et de côté 2ϵ .

2.1.2 Lien entre la longueur l et le coefficient de diffusion D

Intuitivement, il est naturel de conjecturer un lien entre la longueur l et le coefficient de diffusion D . En effet pour garder le même pont brownien, en augmentant le temps, il faut diminuer la diffusion.

Considérons un pont brownien issu de x qui arrive en y au bout d'un temps l , et avec $D = 1$:

$$B_t = x + (y - x)\frac{t}{l} + (W_t - \frac{t}{l}W_l), \quad 0 \leq t \leq l$$

En faisant le changement de variable $t' = \frac{t}{l}$, on obtient le pont équivalent :

$$B_{t'} = x + (y - x)t' + (W_{lt'} - t'W_l), \quad 0 \leq t' \leq 1$$

Rappelons la propriété d'auto-similarité du mouvement brownien : Si $(W_t)_t$ est un mouvement brownien, alors $\forall c > 0$, $(\frac{1}{c}W_{c^2t})_t$ est aussi un mouvement brownien. On peut donc introduire le mouvement brownien $W'_t = \frac{1}{\sqrt{l}}W_{lt}$.

Ainsi

$$B_t = x + (y - x)t + \frac{1}{\sqrt{l}}(W'_t - tW'_1), \quad 0 \leq t \leq 1$$

est toujours le même pont brownien issu de x qui arrive en y en un temps l . Or on remarque, grâce à cette écriture que le pont brownien peut aussi être interprété comme un pont brownien issu de x qui arrive en y au bout d'un temps 1 mais avec un coefficient de diffusion $D = \frac{1}{\sqrt{l}}$. Ainsi il est équivalent de considérer un pont brownien de longueur 1 et de coefficient de diffusion D qui de considérer un pont de longueur D^2 et de coefficient de diffusion 1. Nous choisissons donc par la suite d'imposer $l = 1$, sans perte de généralité.

2.2 Calcul en dimension 2 non résolu

Notre problème se ramène donc à considérer un pont brownien $(B_t)_t$ partant de $(0, y_d)$ arrivant en (x_a, y_a) . Si $(W_t^1)_t$ et $(W_t^2)_t$ sont deux browniens standards indépendants on a :

$$B_t = \begin{pmatrix} X_t \\ Y_t \end{pmatrix} = \begin{pmatrix} W_t^1 - t(W_1^1 - x_a) \\ W_t^2 + y_d - t(W_1^2 - y_a + y_d) \end{pmatrix}$$

On cherche alors

$$\mathbb{P}(\inf_{0 \leq t \leq T} \|B_t\| \leq \epsilon).$$

A priori un tel calcul va dépendre de la norme choisie, la norme 2 serait la plus appropriée pour notre modélisation, puisqu'elle correspondrait à l'entrée du pont brownien dans la boule centrée de rayon ϵ . En revanche, la norme infinie présente le net intérêt de pouvoir utiliser l'indépendance entre les deux coordonnées à un temps t donné :

$$\mathbb{P}(\|B_t\|_\infty \leq \epsilon) = \mathbb{P}(|B_t^1| \leq \epsilon) \cdot \mathbb{P}(|B_t^2| \leq \epsilon)$$

Le problème majeur est que nous devons calculer :

$$\mathbb{P}(\inf_{0 \leq t \leq T} \max(|B_t^1|, |B_t^2|) \leq \epsilon)$$

Les deux coordonnées sont indépendantes mais pour qu'il y ait rencontre, les deux ponts browniens de dimension 1 doivent être dans $[-\epsilon, \epsilon]$ simultanément. Nous n'avons pas réussi à calculer cette probabilités et nous avons eu recours à des approximations.

2.3 Estimateurs et autres approches

2.3.1 Approximations du temps de rencontre

Pour simplifier les calculs, nous considérons ici que $l = 1$ et $D = 1$. Le pont brownien $(B_t)_t$ que nous considérons est issu de (d_1, d_2) et arrive en (a_1, a_2) .

Rappelons tous d'abord que $\forall t \in [0, 1], B_t \sim \mathcal{N}(d + t(a - d), t(1 - t))$, où $(B_t)_t$ est le pont brownien issu de d qui arrive en a au bout d'un temps 1.

On cherche à approcher la probabilité que les deux animaux se rencontrent c'est à dire que le pont brownien qui est défini comme la différence de leur trajectoire rentre dans une boule de rayon ϵ , et centre en 0. Comme nous l'avons vu précédemment, nous n'arrivons pas à résoudre ce problème en continu. Il est alors naturel de discrétiser le temps en N intervalles : $[t_i, t_{i+1}], i \in \llbracket 0, N - 1 \rrbracket$ avec $t_i = \frac{i}{N}$, on notera $h = \frac{1}{N}$ le pas de discrétisation. Naturellement, si deux animaux sont proches et se rencontrent sur un intervalle $[t_i, t_{i+1}]$, il y a de fortes chances qu'ils se rencontrent aussi sur l'intervalle suivant, la trajectoire du brownien étant continu. De ce fait, sommer de 0 à $N - 1$ les probabilités que les animaux se rencontrent sur $[t_i, t_{i+1}]$ est une majoration bien trop grossière : en effet plus le pas h sera petit, plus cette somme sera grande et divergera, car de plus en plus de rencontre seront comptabilisées. Une telle approximation donnerait le nombre de rencontres entre les deux animaux, ce qui n'a évidemment aucun sens en continu. Aussi, nous avons choisi d'estimer le temps de rencontre moyen entre les animaux, c'est à dire le temps qu'ils passent ensemble (*i.e.* à une distance inférieure à ϵ) sur toute leur trajectoire. Dès lors, on considère que si les animaux se rencontrent sur $[t_i, t_{i+1}]$, ils passeront un temps h ensemble. D'un point de vue modélisation, ce n'est pas absurde de considérer que si il y a rencontre à un endroit de l'intervalle, il y a rencontre sur tout l'intervalle.

Ainsi, on va approcher le temps de rencontre $T_{rencontre}$ ainsi :

$$\mathbb{E}\left[\int_{[0,1]} \mathbb{1}_{|B_t| \leq \epsilon} dt\right] = \mathbb{E}[T_{rencontre}] = \sum_{i=0}^{N-1} \frac{1}{N} \mathbb{P}(\exists \text{ une rencontre sur } [t_i, t_{i+1}])$$

On note

$$\begin{aligned}
P_i &= \mathbb{P}(\exists \text{ une rencontre sur } [t_i, t_{i+1}]) \\
&= \mathbb{P}(\exists \text{ une rencontre entre } [0, h] \text{ partant de } B_{t_i}) \\
&= \mathbb{E}[\mathbb{P}_{B_{t_i}}(\exists \text{ une rencontre entre } [0, h])]
\end{aligned}$$

Enfin, remarquons que nous pouvons nous ramener à la dimension 1. En effet, pour un pont brownien $B_t = (B_t^1, B_t^2)$ la probabilité de rencontre en t est $\mathbb{P}(\|B_t\| \leq \epsilon)$. Si on considère la norme infinie, ce qui revient à dire que le pont brownien rentre dans le carré de centre 0 et de côté 2ϵ (qui contient bien la boule centrée de rayon ϵ), comme les deux coordonnées du pont sont supposées indépendantes, nous avons la relation :

$$\mathbb{P}(\|B_{t_i}\| \leq \epsilon) = \mathbb{P}(|B_{t_i}^1| \leq \epsilon) \cdot \mathbb{P}(|B_{t_i}^2| \leq \epsilon)$$

Pour se faire, nous avons pensé à plusieurs approximations :

- Dans un premier temps, nous allons assimiler qu'il y a rencontre sur $[t_i, t_{i+1}]$ si en t_i nous sommes dans la boule centrée de rayon ϵ . Cette approximation paraît très grossière car elle considère les B_{t_i} indépendants.
- Nous pourrions aussi, assimiler la rencontre sur $[t_i, t_{i+1}]$ par la rencontre en t_{i+1} sachant la position B_{t_i} . Il s'agit déjà d'une approximation plus complexe, qui garde une notion de dépendance sur la trajectoire du pont.
- Enfin, grâce au calcul de dimension 1 nous allons pouvoir estimer la rencontre sur $[t_i, t_{i+1}]$ sachant les positions de B_{t_i} et $B_{t_{i+1}}$.

Première approximation : Heuristiquement, nous allons estimer $\mathbb{P}_{B_{t_i}}(\exists \text{ une rencontre entre } [0, h])$ brutalement par $\mathbb{1}_{|B_{t_i}| \leq \epsilon}$. On a alors bien

$$P_i = \mathbb{E}(\mathbb{1}_{|B_{t_i}| \leq \epsilon}) = \mathbb{P}(|B_{t_i}| \leq \epsilon) = \mathbb{P}(\text{il y a rencontre en } t_i)$$

Si $|B_{t_i}| \leq \epsilon$, alors il y a forcément rencontre sur l'intervalle, en revanche, si $|B_{t_i}| \geq \epsilon$, comme le mouvement brownien est continu, pour un h assez petit, (B_t) n'aura pas le temps de rentrer dans la boule de rayon ϵ en un temps $\frac{1}{N}$.

Si quand $N \rightarrow \infty$ il est intuitif que $\mathbb{P}_{B_{t_i}}(\exists \text{ une rencontre entre } [0, h]) \rightarrow \mathbb{1}_{|B_{t_i}| \leq \epsilon}$, mais nous ne connaissons pas le terme du seconde ordre et ne pouvons dire à quel point cette approximation est grossière.

En sommant ces probabilités, nous remarquons que nous avons complètement perdu la notion de dépendance dans la trajectoire : en effet nous allons sommer les probabilités sur chaque instant t_i sans se soucier de la position à l'instant t_{i-1} . Nous verrons par la suite, en confrontant ces résultats avec les simulations, si ces approximations sont admissibles.

On obtient donc :

$$P_i = \int_{-\epsilon}^{\epsilon} \frac{1}{\sqrt{2\pi t_i(1-t_i)}} e^{-\frac{(x-d_1-t_i(a_1-d_1))^2}{2t_i(1-t_i)}} dx \cdot \int_{-\epsilon}^{\epsilon} \frac{1}{\sqrt{2\pi t_i(1-t_i)}} e^{-\frac{(x-d_2-t_i(a_2-d_2))^2}{2t_i(1-t_i)}} dx$$

Soit, avec Φ la fonction de répartition de la loi normale centrée réduite :

$$\begin{aligned}
P_i &= \left[\Phi \left(\frac{\epsilon - d_1 - (a_1 - d_1)t_i}{\sqrt{t_i(1-t_i)}} \right) - \Phi \left(\frac{-\epsilon - d_1 - (a_1 - d_1)t_i}{\sqrt{t_i(1-t_i)}} \right) \right] \\
&\quad \left[\Phi \left(\frac{\epsilon - d_2 - (a_2 - d_2)t_i}{\sqrt{t_i(1-t_i)}} \right) - \Phi \left(\frac{-\epsilon - d_2 - (a_2 - d_2)t_i}{\sqrt{t_i(1-t_i)}} \right) \right]
\end{aligned}$$

En pondérant les probabilités P_i par le pas $\frac{1}{N}$, nous obtenons l'approximation du temps de rencontre suivante :

$$\mathbb{E}[T_{rencontre}] = \frac{1}{N-1} \sum_{i=1}^{N-1} \int_{-\epsilon}^{\epsilon} \frac{1}{\sqrt{2\pi \frac{i}{N} (1 - \frac{i}{N})}} e^{-\frac{(x-d_1 - \frac{i}{N}(a_1-d_1))^2}{2\frac{i}{N}(1-\frac{i}{N})}} dx \cdot \int_{-\epsilon}^{\epsilon} \frac{1}{\sqrt{2\pi \frac{i}{N} (1 - \frac{i}{N})}} e^{-\frac{(x-d_2 - \frac{i}{N}(a_1-d_2))^2}{2\frac{i}{N}(1-\frac{i}{N})}} dx.$$

Dans les deux approximations suivantes, nous ne considérerons qu'un seul pont brownien $(B_t)_t$ en dimension 1 issu de d et arrivant en a . En effet, les deux coordonnées étant indépendantes, il suffit comme dans la première approximation de prendre le produit des probabilités sur chaque intervalle $[t_i, t_{i+1}]$.

Seconde approximation : Nous allons à présent tenter de garder de la dépendance dans la trajectoire, c'est à le fait que B_{t_i} influence $B_{t_{i+1}}$. Supposons dans un premier temps que nous connaissons $B_{t_{i-1}}$, c'est à dire la position des animaux au temps t_i , disons $B_{t_i} = X$, donc le loi est connue. Alors, nous pouvons redéfinir un pont brownien issu de (X_1, X_2) qui arrive en (a_1, a_2) et alors la position après un temps h est telle que :

$$B_{i+1} \sim \mathcal{N}(X + h(a - X), h(1 - h))$$

Ainsi,

$$P_{i+1} = \mathbb{E}[\int_{-\epsilon}^{\epsilon} \frac{1}{\sqrt{2\pi h(1-h)}} e^{-\frac{(y-X-h(a-x))^2}{2h(1-h)}} dy.]$$

L'espérance est prise par rapport à la variable aléatoire X :

$$P_{i+1} = \int_{\mathbb{R}} \int_{-\epsilon}^{\epsilon} \frac{1}{\sqrt{2\pi h(1-h)}} e^{-\frac{(y-X-h(a-x))^2}{2h(1-h)}} dy \frac{1}{\sqrt{2\pi t_i(1-t_i)}} e^{-\frac{(x-d-(a-d)t_i)^2}{2t_i(1-t_i)}} dx$$

Dernière approximation : Cette dernière approximation utilise le résultat démontré en dimension 1. Nous notons $X = B_{t_i}$ et $Y = B_{t_{i+1}}$, les deux variables aléatoires dont les lois sont connues. En considérant le pont brownien de dimension 1 issu de $X = x$ qui arrive en $Y = y$ au bout d'un temps h , nous avons que,

$$\mathbb{P}_{X=x, Y=y} = \min(1, e^{-\frac{2(x-\epsilon)(y-\epsilon)}{h}}).$$

De la même manière que pour la seconde approximation, il faut ensuite intégrer le long des densités de X et de Y .

Ainsi, pour ces trois approximations, nous considérons trois estimations de la probabilité qu'il y ait rencontre sur l'intervalle $[t_i, t_{i+1}]$ sachant $B_{t_i} = x$, il reste ensuite à intégrer cette probabilité contre la loi de B_{t_i} .

- **Première approximation :**

$$\mathbb{P}_x(\exists \text{ une rencontre entre } [0, h]) \approx \mathbb{1}_{|x| \leq \epsilon}$$

- **Seconde approximation :**

$$\mathbb{P}_x(\exists \text{ une rencontre entre } [0, h]) \approx \int_{-\epsilon}^{\epsilon} \frac{1}{\sqrt{2\pi h(1-h)}} e^{-\frac{(y-x-h(a-x))^2}{2h(1-h)}} dy$$

- **Dernière approximation :**

$$\mathbb{P}_x(\exists \text{ une rencontre entre } [0, h]) \approx \int_{\mathbb{R}} \min(1, e^{-\frac{2(x-\epsilon)(y-\epsilon)}{h}}) \frac{1}{\sqrt{2\pi h(1-h)}} e^{-\frac{(y-x-h(a-x))^2}{2h(1-h)}} dy$$

2.3.2 Processus de Bessel de dimension 2

Étant donné que nous n'arrivons pas à calculer la probabilité de rencontre entre deux animaux, nous nous sommes tournés vers un modèle plus simple. L'idée est de ne pas contraindre la position finale des deux animaux. Cette approche est bien différente de la précédente mais cette alternative pourrait permettre des formules explicites : Pour se faire, nous modélisons le déplacement des animaux par deux mouvements browniens indépendants de dimension 2 issus de (x_1, y_1) et de (x_2, y_2) . La différence de ces deux mouvements brownien est toujours un mouvement brownien de dimension 2 issu de $(x_1 - x_2, y_1 - y_2)$. Un processus de Bessel est défini comme la norme du mouvement brownien qui lui est associé. Nous considérons un processus de Bessel de dimension 2 issu d'un rayon $r = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$, qui peut être caractérisé par l'équation différentielle stochastique suivante :

$$\begin{cases} dR_t &= dW_t + \frac{1}{2R_t} dt \\ R_0 &= r \end{cases}$$

Preuve. $R_t = \sqrt{(B_t^1)^2 + (B_t^2)^2}$, où B_t^1 et B_t^2 sont deux mouvements browniens de dimension 1 issus respectivement de $x_1 - x_2$ et de $y_1 - y_2$.

Or $dB_t^{i^2} = 2B_t^i dB_t^i + dt$, pour $i = 1, 2$ (en appliquant la formule d'Itô au mouvement brownien et à la fonction carré) et donc en dérivant R_t on obtient :

$$dR_t = \frac{1}{2R_t} (2B_t^1 dB_t^1 + 2B_t^2 dB_t^2 + 2dt) = dW_t + \frac{1}{2R_t} dt$$

où

$$dW_t = \frac{B_t^1 dB_t^1 + B_t^2 dB_t^2}{\sqrt{(B_t^1)^2 + (B_t^2)^2}}$$

et il reste à démontrer que W_t est bien un mouvement brownien pour conclure. En effet, c'est une martingale comme somme de martingales et son crochet vaut t , en appliquant la formule d'Itô en dimension 2 à la fonction carré sur (W_t) et en prenant l'espérance :

$$\begin{aligned} \mathbb{E}(W_t^2) &= \mathbb{E}(W_0^2) + \mathbb{E}\left(\int_0^t \left(\left(\frac{B_s^1}{\sqrt{(B_s^1)^2 + (B_s^2)^2}}\right)^2 + \left(\frac{B_s^2}{\sqrt{(B_s^1)^2 + (B_s^2)^2}}\right)^2\right) ds\right) \\ &= \mathbb{E}(W_0^2) + \mathbb{E}\left(\int_0^t ds\right) \\ &= \mathbb{E}(W_0^2) + t \end{aligned}$$

Donc $(W_t^2 - t)_t$ est une martingale. Ainsi par caractérisation de Lévy, $(W_t)_t$ est un mouvement brownien. \square

Le générateur de ce processus de Bessel est :

$$Ah(x) = h'(x) \frac{1}{2x} + \frac{1}{2} h''(x)$$

La fonction qui annule ce générateur est la fonction logarithme népérien. Si on considère une distance minimale ϵ qui caractérise la rencontre et une distance maximale n , nous pouvons calculer la probabilité que les animaux se rencontrent, c'est à dire que le processus de Bessel rentre dans la boule centrée en 0 de rayon ϵ avant qu'ils ne soient distance de plus de n :

$$\mathbb{P}(T_\epsilon < T_n) = \frac{\ln(r) - \ln(n)}{\ln(\epsilon) - \ln(n)}$$

Cette probabilité tend vers 1 quand $n \rightarrow \infty$. Il est difficile de conclure car, le processus finit toujours par revenir dans la boule centrée en 0 de rayon ϵ , en revanche on arrive à caractériser le fait qu'il revienne dans cette boule avant que la distance entre les deux animaux ne deviennent trop grande.

3 Algorithmes & Résultats

3.1 Simulation de ponts browniens

La première étape est de modéliser le pont brownien issu de a qui arrive en b au bout d'un temps T :

La formule $B_t = a + (b - a)\frac{t}{T} + D(W_t - \frac{t}{T}W_T)$ suggère qu'il suffit de savoir programmer un Brownien $(W_t)_t$. Pour cela, il est nécessaire de discrétiser $[0,1]$ en N intervalles de taille h , on note les temps $t_i = h.i, i \in \llbracket 0, N \rrbracket$

Ensuite on définit le mouvement brownien par récurrence :
 $W[i + 1] = W[i] + gauss$, où $gauss$ est la réalisation d'une variable aléatoire de loi $N(0, \sqrt{h})$.
Puis on déduit le pont brownien grâce à la formule ci-dessus. A partir de cela il est aisé de construire un pont brownien en deux dimensions en se rappelant que ces coordonnées sont des ponts browniens de dimension 1 indépendants.

```
def Brow(l, N, Di):
    t, h = np.linspace(0, T, N, retstep=True)
    X = np.zeros(N)
    for i in range(N-1):
        X[i+1] = X[i] + Di * rd.gauss(0, h**0.5)
    return t, X

def pontBrow(a, b, T, N, Di):
    u = Brow(T, N, Di)[1]
    t = np.linspace(0, T, N)
    return t, u + a - (u[-1] - b + a) * t / T
```

3.2 Algorithmes : approximations du temps de rencontre moyen et de la probabilité de rencontre

3.2.1 Probabilité de rencontre

Cet algorithme rend la probabilité de rencontre. On simule Nb mouvements browniens. Pour chaque simulation, nous initialisons une variable à 0, si la trajectoire de la simulation rencontre la boule $\mathcal{B}(0, \epsilon)$ sur un des temps, cette variable est mise à jour à 1. Enfin nous faisons la moyenne sur le nombre de simulations.

```
def rencontre(x1, x2, y1, y2, T, N, e, Di):
    B = 0
    U = diffpontbrow(x1, x2, y1, y2, T, N, Di)[1]
    i = 0
    while abs(U[i]) > e and i < N-1:
        i = i + 1
    if i == N-1:
        B = 0
    else:
```

```

    B=1
    return B

```

```

def proba_rencontre(x1, x2, y1, y2, T, N, e, Nb, Di):
    P=0
    for i in range(Nb):
        P=P+rencontre(x1, x2, y1, y2, T, N, e, Di)
    return P/Nb

```

Un des avantages de cet algorithme est qu'une fois que la rencontre s'est produite, nous n'avons plus besoin de parcourir le reste de la trajectoire : il s'agit d'une boucle *while* et non d'une boucle *for*. Il est un peu moins coûteux que l'algorithme de temps de rencontre. On s'en servira notamment pour estimer l'influence de la diffusion en dimension deux.

Nous avons aussi pu confronter le résultat théorique de la dimension 1 avec les simulations. Dès lors que N et Nb sont suffisamment grands, les résultats des simulations sont en partie en accord avec la théorie. Notons tout de même qu'il faut énormément discrétiser le temps pour avoir un brownien admissible (N de l'ordre de 10 000), ce qui est très coûteux. Les formules explicites et les approximations permettent de réduire énormément les temps de calculs des algorithmes, elles devraient être privilégiées lorsque l'on regarde la probabilité de rencontre sur tout un relevé GPS, c'est à dire un grand nombre de ponts browniens.

Voici, un tableau qui compare dans trois cas différents les probabilités de rencontre :

x_1	x_2	y_1	y_2	ϵ	D	N	Nb	valeur estimée	valeur exacte
0.5	0	0.2	0	0.0001	1	30 000	500	0.796	0.81884536703
1	0	0.5	0	0.0001	1.3	30 000	500	0.554	0.55347512265
0.5	0	0	0.5	0.0001	1	30 000	500	1.0	1

Si pour certains cas l'algorithme de probabilité de rencontre semble satisfaisant, il ne l'est pas pour d'autres et nous ne savons pas pourquoi. Il semblerait qu'il existe un lien entre ϵ , D et N mais nous n'arrivons pas à comprendre pourquoi l'algorithme marche remarquablement bien dans certains cas et donne des résultats aberrants dans d'autres. Nous obtenons parfois des probabilités plus petites que 1 alors que les points de départ et d'arrivée du pont brownien modélisant la distance relative entre les deux animaux passe par 0 en dimension 1, ou pire encore des probabilités beaucoup plus grande que la probabilité de rencontre exacte ce qui n'est absolument pas normal. En effet, comme le brownien simulé est discontinu, la probabilité calculée algorithmiquement devrait oublier des rencontres et donc être plus petite que la probabilité exacte...

3.2.2 Temps moyen de rencontre

En simulant des browniens : La première approche est relativement simple, il suffit de générer des ponts browniens représentés par des vecteurs. On récupère pour une simulation deux vecteurs, celui des abscisses et celui des ordonnées. Il ne reste plus qu'à parcourir simultanément ces deux vecteurs et compter un temps t de rencontre quand ces deux vecteurs sont simultanément plus petit que ϵ . Ensuite il suffit de faire une moyenne sur le nombre de simulations Nb . On note que pour avoir une variance faible par cette méthode, il faut considérer un N et un Nb assez grand, ce qui nécessite un temps de calcul assez long.

```

def difpontbrow(x1, x2, y1, y2, T, N, Di):
    U=pontBrow(x1, y1, T, N, Di)
    V=pontBrow(x2, y2, T, N, Di)
    t=np.linspace(0, T, N)
    A=U[1]-V[1]
    return t, A

```

```

def tpsrencontre(x1, x2, y1, y2, T, N, e, Di):
    U=difpontbrow(x1, x2, y1, y2, T, N, Di)[1]
    t=0
    for i in range(N):
        if abs(U[i])<e :
            t=t+1
    return t

```

Nb=550 *#nombre de simulations*

```

def tpsmoyren(x1, x2, y1, y2, T, N, e, Nb, Di):
    J=0
    for i in range(Nb):
        J=J+tpsrencontre(x1, x2, y1, y2, T, N, e, Di)
    return J/Nb
print('temps_moyen_de_rencontre:')
print(tpsmoyren(x1, x2, y1, y2, T, N, e, Nb, Di)*T/N)

```

Par approximation : On peut aussi programmer les approximations faites pour ce temps de rencontre. On a en fait programmé la formule définie dans la partie approximations. Ainsi, nous avons pu comparer les résultats des simulations et ceux donnés par la première approximation. Nous avons ainsi pu constater que les deux résultats étaient très proches. L'approximation est donc très bonne et présente le net avantage qu'elle n'est pas du tout couteuse.

```

def aproximation(N, T, d1, f1, d2, f2, e):
    t, h=np.linspace(0, T, N, retstep=True)
    P=0
    for Ti in t[1:-1]:
        def F1(x):
            return np.exp(-(x-d1-Ti*(f1-d1))**2/(2*Ti*(1-Ti)))
            /np.sqrt(2*np.pi*Ti*(1-Ti))
        def F2(x):
            return np.exp(-(x-d2-Ti*(f2-d2))**2/(2*Ti*(1-Ti)))
            /np.sqrt(2*np.pi*Ti*(1-Ti))
        intf1=int.quad(F1, -e, e)
        intf2=int.quad(F2, -e, e)
        P=P+intf1[0]*intf2[0]
    return P*h

```

3.3 Influence du paramètre de diffusion

Nous avons regardé l'effet de la diffusion sur la probabilité de rencontre en dimension 1 de façon algorithmique même si celui-ci ne nous paraît pas satisfaisant ? Nous avons donc regardé les cas où l'algorithme marchait bien en dimension 1 et ce pour des coefficients de diffusion par trop grands. Il semblerait alors que la diffusion influence de manière similaire la probabilité de rencontre théorique et celle prédite par l'algorithme. Nous avons tenté d'étendre ceci à la dimension 2 en traçant la probabilité de rencontre en fonction de la diffusion en conjecturant que la diffusion en 2D agissait de la même manière qu'en 1D.

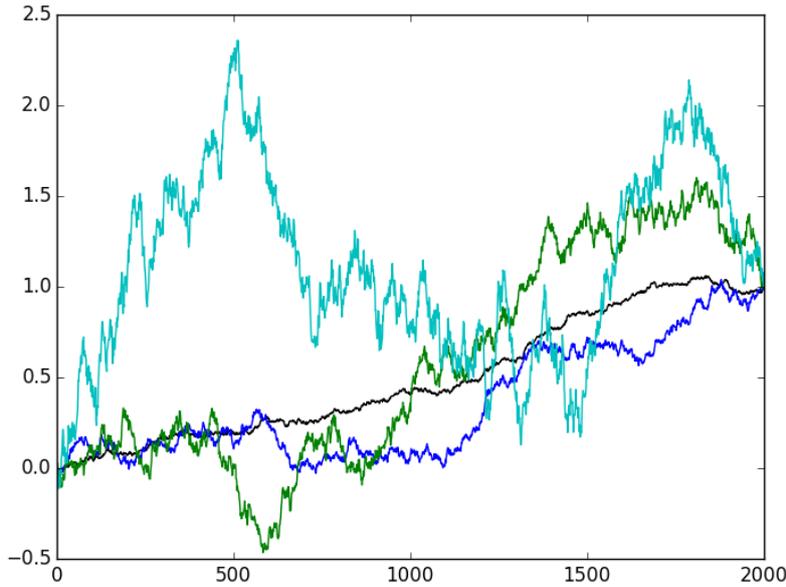


FIGURE 4 – Simulation de cinq ponts browniens en dimension 1 de coefficient de diffusion 0.2 (noir), 0.5 (bleu), 1 (vert) et 2(cyan).

Cet algorithme trace plusieurs ponts browniens avec une diffusion différente. Comprendre l'influence du paramètre de diffusion est essentiel.

En dimension 1 nous avons observé théoriquement la dépendance de la probabilité de rencontre en D , le coefficient de diffusions :

$$\mathbb{P}_R = C \frac{1}{D^2}$$

Nous avons présupposé qu'il existait cette même dépendance en dimension 2, et avons voulu observer ceci grâce aux simulations. Les résultats sont très satisfaisants, mais nous ne pouvons pas prendre des valeurs de D trop importantes, car, dans ce cas, il faudrait augmenter N drastiquement, ce qui est beaucoup trop coûteux.

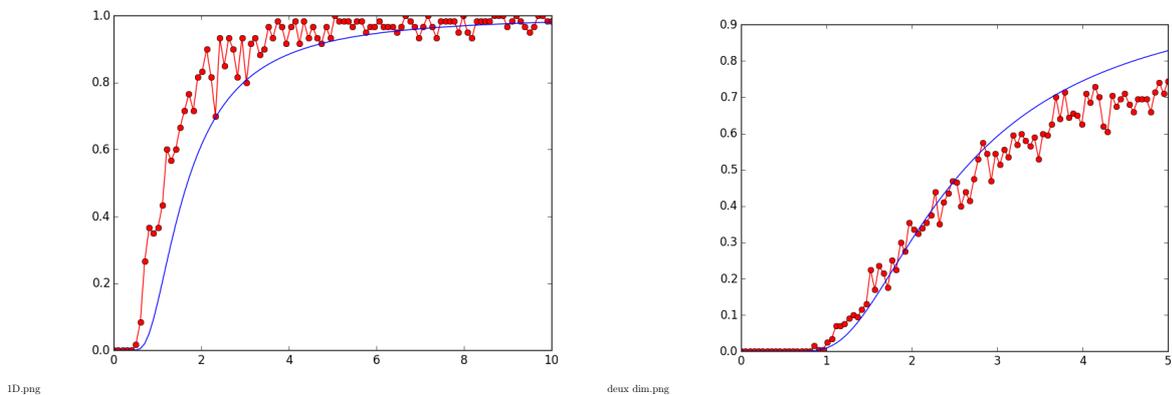


FIGURE 5 – Influence de la diffusion sur la probabilité de rencontre en dimensions 1 et 2. Comparaison entre la courbe théorique (en bleu) et les simulations (en rouge).

Pour la dimension 1, il est facile de tracer la courbe théorique car nous connaissons l'expression explicite de la probabilité de rencontre. En revanche pour la dimension 2 nous l'estimons avec les algorithmes précédents pour un coefficient de diffusion 1 et nous traçons : $\mathbb{P}_R(D^2) = \mathbb{P}_R(1) \frac{1}{D^2}$.

3.4 Confrontation aux données

Cet algorithme est un algorithme qui prend pour entrée plusieurs relevés de deux animaux à intervalles de temps constants et qui renvoient le temps que ces deux animaux ont passé ensemble, il trace aussi une simulation de ces mouvements en mettant en couleur les rencontres.

```
import numpy as np
import pylab as plt
import numpy.random as rdm
import random as rd

T=2000
N=1001
Nb=100
e=100
Di=0.7

X1=np.array([467,470,499,520,617,531,501,721,709,720,731,775,833,
711,724,807,952])
Y1=np.array([1390,1242,1159,1217,1221,1286,1338,1110,1092,1050,
1041,1002,1013,1057,946,941,940])
X2=np.array([188,233,294,301,287,249,403,315,499,371,496,461,530,
434,456,215,102])
Y2=np.array([1391,1224,1365,1255,1277,1385,1324,1235,1159,1297,
1268,1393,1316,1522,1380,1202,1269])

def suivianim(X1,Y1,X2,Y2,T,e,N,Nb,Di):
    Tp=0
    n=np.shape(X1)[0]
    plt.figure(1)
    plt.clf()

    for i in range(n-1):
        A=np.array([X1[i],Y1[i]])
        B=np.array([X1[i+1],Y1[i+1]])
        C=np.array([X2[i],Y2[i]])
        D=np.array([X2[i+1],Y2[i+1]])

        Tp=Tp+tpsmoyen(A,B,C,D,T,N,e,Nb,Di)

    U=pontBrow2(A,B,T,N,Di)
    K=pontBrow2(C,D,T,N,Di)
    plt.plot(U[0],U[1],'-b')
    plt.plot(K[0],K[1],'-g')
    plt.plot(A[0],A[1], 'ob')
    plt.plot(C[0],C[1], 'og')
    plt.plot(D[0],D[1], 'og')
    plt.plot(B[0],B[1], 'ob')
    plt.legend(('trajectoire1','trajectoire2'))
    for j in range(N):
        if abs(U[0][j]-K[0][j])<e and abs(U[1][j]-K[1][j])<e:
            plt.plot(U[0][j],U[1][j], 'or')
            plt.plot(K[0][j],K[1][j], 'mo')
```

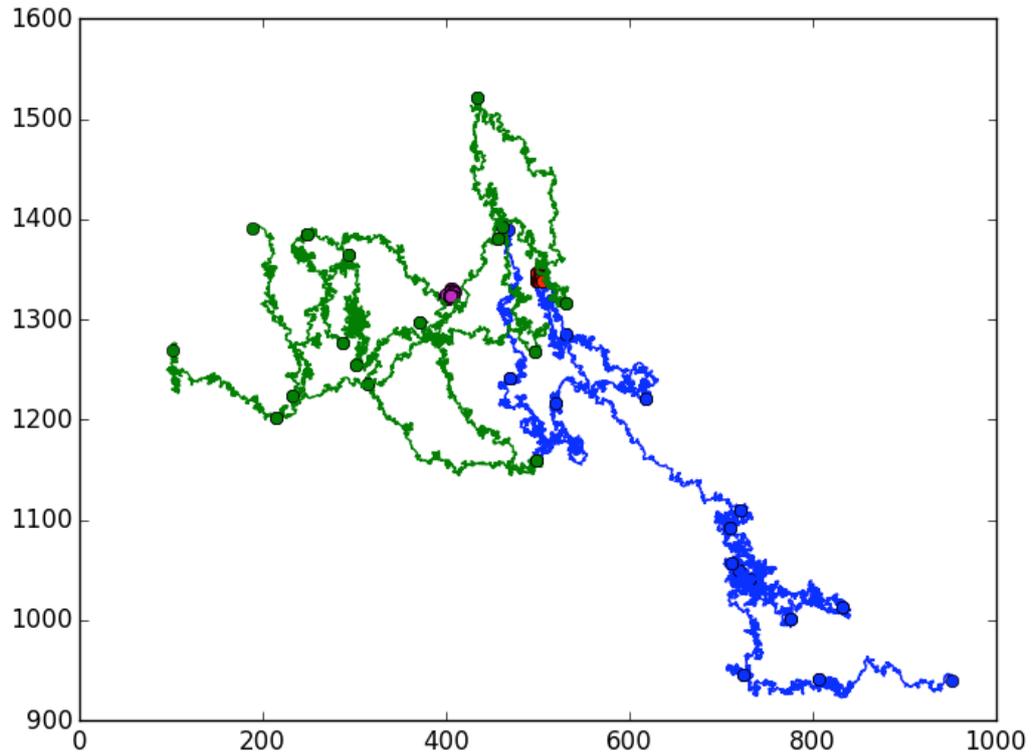
```

return(Tp/(n-1)*T/N)

print('temps_total_de_rencontre_sur_ce_trajet:')
print(suivianim(X1,Y1,X2,Y2,T,e,N,Nb,Di))
print('temps')
print(T)
#les points de rencontre apparaissent en rouge et violet

```

Ce programme simule le temps de rencontre entre deux animaux suivis sur une journée, nous avons tracé ceci à partir des vrais relevés fournis par l'ONCFS.



finales.png

FIGURE 6 – Simulations de ponts browniens entre les points GPS fournis. En rouge et mauve : les points de rencontre entre les animaux.

Conclusion

La première partie de notre travail a été de comprendre le problème et de le modéliser. Les choix étant nombreux, celui qui nous paraissait le plus adéquat était celui du pont brownien. Nous nous sommes tout d'abord penchés sur le problème en dimension 1, après avoir consulté certains ouvrages à ce sujet nous avons constaté que de nombreux résultats mathématiques étaient accessibles à ce sujet. Ainsi nous avons obtenu une formule exacte pour la probabilité de rencontre. Nous avons par la suite implémenté des algorithmes pour retrouver cette probabilité avec des simulations de trajectoires en pont brownien. Ces derniers n'étant pas entièrement satisfaisants, nous avons décidé d'estimer le temps de rencontre mien entre deux ponts browniens. Nous avons alors développé trois formules pour approcher ce temps de rencontre dont une que nous avons bien détaillée et implémentée. Les approximations théoriques n'étant pas si éloignées des résultats numériques nous avons choisi d'étendre ce modèle à la dimension deux. Encore une fois, les versions algorithmique et théorique semblent coïncider.

Actuellement notre projet permet de se faire une bonne idée des problématiques rencontrées et d'approcher très convenablement le temps moyen de rencontre entre deux trajectoires de ponts browniens. Cette étude mérite cependant d'être bien plus approfondie. Il reste notamment à comprendre pourquoi les résultats de probabilité de rencontre ne sont pas toujours en accord avec la théorie, il faudrait aussi étudier les autres formules d'approximation du temps de rencontre, afin de les comparer à celle que nous avons regardée en détail et qui est la plus grossière. Enfin nous pourrions aussi complexifier le modèle : la partie « déterministe » de notre pont brownien est ici une droite reliant le départ à l'arrivée mais on pourrait supposer qu'il s'agit d'une parabole (le cas du contournement d'un obstacle par exemple) ou de tout autre fonction? Voire même étendre le modèle à la dimension 3 pour pouvoir l'appliquer aux oiseaux ou aux poissons.

Nous tenons à remercier Vincent Bansaye et Clément Calenge pour l'aide et les conseils qu'ils nous ont apportés tout au long du projet.

Références

- [1] Ikeda and Watanabe : *Stochastic Differential Equations and Diffusion Processes* (1981)
- [2] J. L. Doob : *Heuristic approach to the Kolmogorov-Smirnov theorems*, Ann. Math. Stat., Vol.20 (1949), pp. 393-403
- [3] Borodin, Andrei, Salminen, Paavo : *Handbook of Brownian Motion*, (1969)

Rapport de stage de fin d'études.

Personne Arnaud, sous la direction de A.Guillin et F.Jabot

29 septembre 2016



FIGURE 1 – Diversité au sein des forêts

Table des matières

1	Mise en contexte	4
2	Le modèle de Hubbell avec pool régional constant.	5
2.1	Présentation du modèle	5
2.2	Approche probabiliste	6
2.2.1	Mise en équation	6
2.2.2	Loi du nombre d'individus de l'espèce i à l'équilibre	6
2.2.3	Premiers résultats numériques et les limites	7
2.3	La loi du vecteur d'abondance	8
2.4	Vers le coalescent et la formule d'Ewens	8
2.5	Indice de Simpson	11
2.5.1	Présentation	11
2.5.2	Indice de Simpson à l'équilibre	12
2.6	Résultats numériques	12
2.7	Conclusion	13
3	Un modèle plus général	14
3.1	Introduction	14
3.2	Présentation du modèle	14
3.3	Algorithme et résultats	15
3.3.1	Présentation de l'algorithme	15
3.3.2	Quelques résultats de simulations	15
3.4	Mise en équation	16
3.5	La loi du vecteur d'abondance	16
3.5.1	La vision en coalescent	17
3.5.2	La loi du vecteur d'abondance	18
3.6	Indice de Simpson pour ce modèle	21
3.6.1	En théorie	21
3.6.2	En pratique	21
3.7	Avantages et inconvénients de ce modèle	22
4	Un modèle neutre avec variance environnementale	23
4.1	Le modèle	23
4.2	Mise en équation	23
4.3	Premiers exemples	24
4.4	Le paramètre m	26
4.5	Le paramètre s	28
4.5.1	Influence de s sur la loi log normale	28
4.5.2	Influence de s sur la variance de l'indice de Simpson	30
4.5.3	Influence de s lors d'un changement de fitness.	30
4.6	Une approche déterministe	31
4.7	Comparaison avec le modèle	33
4.7.1	En régime asymptotique	33
4.7.2	Étude sur un temps court	33
4.7.3	Le nombre d'espèces	35
4.7.4	Influence des poids initiaux	36
4.7.5	Lors d'un changement de fitness	38

5	Vers des modèles plus simplifiés	39
5.1	Si $m=0$:	40
5.1.1	Présentation	40
5.1.2	Les résultats	40
5.1.3	Analyse	43
5.2	Tentative de calcul d'une abondance par coalescent en l'absence de changement de fitness	44
5.2.1	La généalogie	44
5.2.2	Quand les poids sont très proches de 1	46
6	Conclusion	49
7	Bibliographie	50
8	Annexes	50
8.1	algorithme	50

Résumé : Ce rapport traitera des modèles neutres en écologie. Nous reprendrons le modèle introduit par Hubbell en 2001, et nous y introduirons de la fitness. La dernière partie présentera nos premières études sur le modèle avec fitness.

J'aimerais par ailleurs remercier A.Guillin et F.Jabot pour leur aide et leur patience.

1 Mise en contexte

Depuis Darwin, en passant par Gause, Hutchinson et bien d'autres, les écologistes ont tenté d'expliquer la structure des communautés d'êtres vivants et l'abondance des espèces dans un milieu, en partant du principe selon lequel, chaque espèce occupe une niche qui lui est propre. Plus précisément, si deux espèces ont des exigences écologiques identiques et sont donc en compétition pour une même ressource (une même niche) l'une finit par s'approprier le milieu au détriment de l'autre.

Par ce principe, ils expliquent le développement d'une structure de communauté par la sélection des espèces les mieux adaptées dans un environnement hétérogène en relation avec la complexité du milieu.

Si cette théorie largement développée au siècle dernier apporte des réponses probantes, il s'avère que dans certains cas elle ne puisse parvenir à justifier l'existence de certaines populations. Par exemple, certaines forêts tropicales présentent plus de cent espèces d'arbres différentes sur un hectare. Et plusieurs de ces espèces ont clairement des exigences écologiques similaires (partagent la même niche).

Par conséquent, la diversité des espèces présentes sur une même niche écologique ne serait pas due simplement à la niche elle-même, mais aussi à la dynamique des individus présents sur cette niche. C'est cette idée là qui est le point de départ des théories neutres en écologie, et en particulier de celle que Hubbell présente en 2001 dans son livre : "The unified neutral theory of biodiversity and biogeography", sur lequel nous reviendrons par la suite.

Les théories neutres suggèrent plutôt que les espèces sont identiques et d'égaux compétiteurs, c'est à dire qu'elles ont des taux de naissance et de mort qui sont les mêmes. Par ailleurs, ces théories respectent un principe de somme-zéro (zero-sum), i.e. que la somme de tous les changements d'abondance des espèces est toujours nulle. Ceci suppose que les environnements sont saturés, ce qui est le cas des forêts tropicales par exemple. (cf Hubbell 2001). Toutes les espèces peuvent ainsi se développer sur un site donné, faisant en sorte que la composition de la communauté soit déterminée uniquement par la dispersion stochastique des individus.

Ainsi on espère, via ces modèles, pouvoir expliquer les abondances (nombre d'individus représentant une espèce) des espèces observées. Dans la suite nous présenterons différents modèles "neutres". Dans un premier temps nous verrons le modèle introduit par Hubbell avec un pool uniforme puis avec un pool qui

évolue dans le temps. Ensuite nous introduirons un modèle qui tient compte des variations environnementales, et nous nous familiariserons avec ce modèle en regardant l'effet des différents paramètres.

2 Le modèle de Hubbell avec pool régional constant.

2.1 Présentation du modèle

Considérons une communauté dite régionale, composée de J individus d'espèces différentes. A chaque pas de temps, on fait mourir au hasard un individu de cette population et il est remplacé avec probabilité m par le fils d'un autre individu de cette communauté (naissance) ou avec probabilité $1-m$ par le fils d'un individu provenant d'un réservoir, choisi au hasard. (immigration)

La communauté régionale pouvant être assimilée à une parcelle d'une forêt, celle que l'on étudie, et le réservoir au reste de la forêt.

On notera que ce réservoir est composé de S espèces (S grand) et n'évolue pas dans le temps. En effet, si sa taille est très grande par rapport à celle de la communauté, son évolution est peu significative par rapport à l'évolution de la communauté. Dans un premier temps, on considère le réservoir comme uniforme, c'est à dire que le taux d'immigration de chaque espèce est le même.

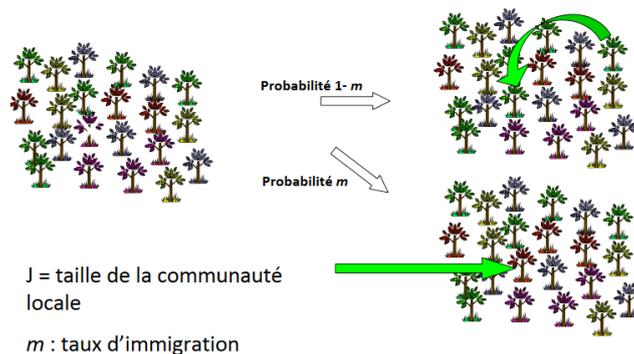


FIGURE 2 – illustration du modèle.

Contrairement à une population complètement isolée dans laquelle la dérive écologique causerait inévitablement l'extinction de toutes les espèces sauf une (ce qui est le cas dans la théorie des niches), le taux d'immigration à l'échelle locale permet de compenser les extinctions et de maintenir de la diversité dans le système.

2.2 Approche probabiliste

2.2.1 Mise en équation

On se concentre d'abord sur une espèce prise au hasard, on cherche la loi du nombre d'individus qui va la représenter au cours du temps. Cette loi est représentée par un vecteur de taille J ou chaque coordonnée x_j^t représente la probabilité que l'espèce considérée ait j individus au temps t . On va ici détailler les probabilités de transition :

Soit X_i^t la variable aléatoire du nombre d'individus de l'espèce j au temps t :

$$\begin{cases} P(X_i^{t+1} = N_i - 1 | X_i^t = N_i) = \left(\frac{N_i}{J}\right)(m(1 - P_i) + (1 - m)\left(\frac{J - N_i}{J - 1}\right)) \\ P(X_i^{t+1} = N_i + 1 | X_i^t = N_i) = \left(\frac{J - N_i}{J}\right)(mP_i + (1 - m)\left(\frac{N_i}{J - 1}\right)) \\ P(X_i^{t+1} = N_i | X_i^t = N_i) = 1 - P(N_i - 1 | N_i) - P(N_i + 1 | N_i) \end{cases}$$

où P_i est la proportion de l'espèce i dans le réservoir.

Ceci nous conduit à considérer une matrice de transition B , tridiagonale propre à chaque espèce, dont les puissances vont nous donner les différents états possibles au bout d'un nombre d'événements équivalent à la puissance.

2.2.2 Loi du nombre d'individus de l'espèce i à l'équilibre

Pour obtenir la loi du vecteur d'abondance au temps suivant, sachant que l'on part d'un vecteur initial de densité X , il suffit de calculer tXB .

Comme notre matrice B n'a pas d'états absorbants (cf nature du problème), la multiplication à gauche est une application contractante.

Par Picard elle admet un point fixe donné par la limite de ${}^tXB^k$ quand k tend vers l'infini pour tout x . Comme la matrice est apériodique, il est unique.

Ainsi ce point fixe va nous donner la loi du nombre d'individus de l'espèce i à l'équilibre.

Notons que la notion d'équilibre ici ne signifie pas que le système n'évolue plus mais simplement que les probabilités de transitions ne dépendent plus du temps.

Le calcul de ce vecteur semble délicat mais est néanmoins faisable (cf Hubbell 2001 pour plus de détails). Il est en revanche possible de programmer ce vecteur via la formule présente dans le livre de Hubbell cité en première partie :

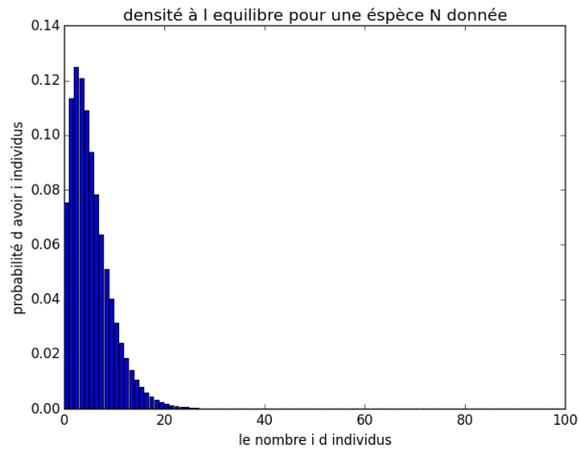


FIGURE 3 – Un exemple du vecteur de probabilité d’abondance pour une espèce.

2.2.3 Premiers résultats numériques et les limites

Nous avons dans un premier temps implémenté l’algorithme du modèle. Voici les résultats obtenus :

Le premier graphique représente le nombre d’individus par espèce. Le deuxième représente l’abondance, i.e le nombre d’espèces représentées par tant d’individus, on ne s’intéresse pas à une espèce en particulier.

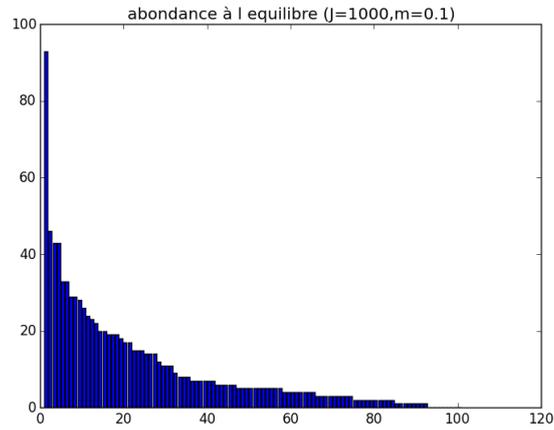


FIGURE 4 – Est représenté le nombre d’individus en fonction des espèce au bout de 500 000 étapes environ. La communauté initiale est formée de 4 espèces

Ensuite nous avons tiré pour chaque espèce un nombre aléatoire d'individus selon la loi du vecteur densité à l'équilibre sous la contrainte d'obtenir J individus au total :

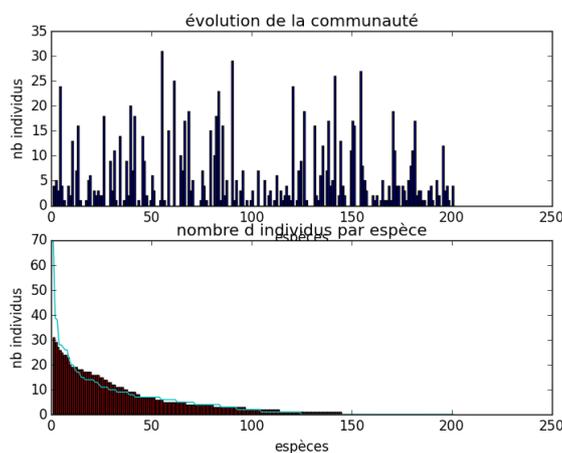


FIGURE 5 – Sont représentés le nombre d'individus en fonction des espèces au bout de 100 000 étapes environ (bar) ainsi que le résultat obtenu à partir du vecteur calculé précédemment(bleu clair).

Même si on obtient la loi d'une espèce à l'équilibre, on ne peut en déduire directement la loi du vecteur d'abondance à l'équilibre, c'est à dire la loi du vecteur qui a pour marginale i la quantité d'individus de l'espèce i à l'équilibre. Ceci équivaut à la loi du vecteur de taille S ayant pour marginale i la quantité d'espèces ayant i individus(puisque les individus sont identiques). La contrainte principale étant que la somme des abondances de chaque espèce vaille J .

2.3 La loi du vecteur d'abondance

Le but de cette partie va être de calculer la loi du vecteur d'abondance.

2.4 Vers le coalescent et la formule d'Ewens

Pour obtenir cette loi, il est nécessaire de regarder le problème sous un autre angle, celui du coalescent, c'est dire que l'on va remonter le temps. On part d'une population connue et on se demande quel est le dernier changement qui a conduit à cette population? C'est évidemment la naissance d'un individu de l'échantillon, celui ci vient du pool ou est le fils d'un individu du reste de l'échantillon. Avec quelles probabilités ces deux événements peuvent survenir?

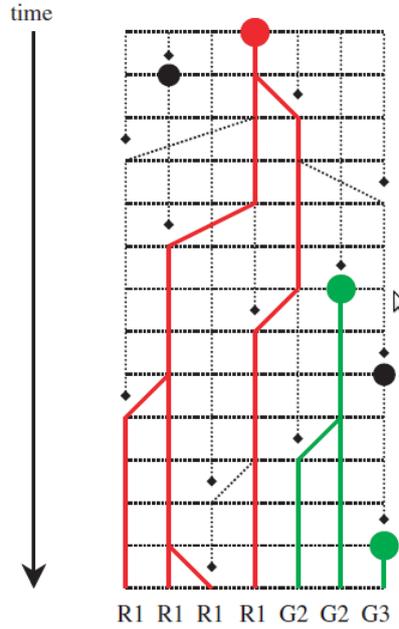


FIGURE 6 – Exemple de coalescent.(d'après, A novel genealogical approach to neutral biodiversity theory Rampal S. Etienne and Han Olf)

Considérons que l'on connaît une partie de la communauté composée de k individus (k pouvant être égal à J) parmi les J . A l'étape précédente, 2 cas ont pu se produire :

- soit deux individus coalescent
- soit un individu du pool remplace le mort

Les événements qui nous intéressent sont seulement ceux qui touchent notre échantillon : en effet tous les individus étant à priori les mêmes (principe de neutralité), les autres événements ne modifieront pas la situation dans laquelle nous nous trouvons.

Alors à l'étape précédente on a eu : -coalescence de deux individus de l'échantillon avec proba : $(m-1) \frac{k(k-1)}{J(J-1)}$
 -spéciation avec proba : $m \frac{k}{J}$

Ainsi le dernier évènement qui a touché notre échantillon est une coalescence avec probabilité : $\frac{(m-1) \frac{k(k-1)}{J(J-1)}}{m \frac{k}{J} + (m-1) \frac{k(k-1)}{J(J-1)}} = \frac{\theta}{k-1+\theta}$ ou une spéciation avec probabilité $\frac{m \frac{k}{J}}{(m-1) \frac{k(k-1)}{J(J-1)} + m \frac{k}{J}} = \frac{k-1}{k-1+\theta}$ où $\theta = \frac{m(J-1)}{m-1}$

Ainsi, pour une population donnée dans le présent, à condition que l'on puisse remonter assez loin dans le temps (ce qui est le cas à l'équilibre), il existe un individu dans la communauté qui est le plus "vieux", ie c'est le premier individu à être arrivé du pool et encore vivant au temps présent.

Si maintenant on part de cet ancêtre pour aller vers le présent, sur cette lignée chaque événement sera une naissance, (au bout de $k-1$ étapes la population sera de taille k); i.e chaque événement correspond à la naissance d'un individu celui ci sera comme son père avec proba $\frac{1}{k+\theta}$ ou de nature différente (viendra du pool) avec probabilité $\frac{\theta}{k+\theta}$ où k est la taille de la population, le choix du père étant uniforme parmi la taille de la population. (et ces deux événements sont indépendants).

Ce processus est ni plus ni moins celui des restaurants chinois arrêté après J étapes.

Alors la formule d'Ewens nous permet d'obtenir la loi du vecteur d'abondance :

Si on considère J le nombre d'individus dans la communauté et n_i le nombre d'espèces représentées par i individus alors :

$$P[(n_1, \dots, n_J)] = \frac{J!}{\theta_J} \prod_{i=1}^J \frac{1}{n_i!} \left(\frac{\theta}{i}\right)^{n_i} \quad \text{où} \quad \theta_J = \prod_{i=1}^J (\theta + i - 1)$$

Justification de cette formule : On se donne un vecteur (n_1, \dots, n_s) la probabilité recherchée correspond aux nombres de permutations d'un ensemble de J éléments ayant n_1 cycles de longueurs 1, ... etc multiplié par la probabilité d'avoir une permutation en particulier.

1. On sait par le principe des restaurants chinois que la probabilité d'obtenir une transposition en particulier est $\frac{\theta^{|p|}}{\theta_s}$ ou $|p| = \sum n_i$.
2. On multiplie cette probabilité par le nombre de transpositions qui conduisent à la même transposition non ordonnée (on met en relation les transpositions ayant les mêmes chiffres dans chacun des cycles de sorte à ce que l'ordre des chiffres dans les cycles n'ait pas d'importance). Ce nombre vaut alors $\prod (i-1)^{n_i}$
 enfin $\frac{\theta^{|p|}}{\theta_s} \prod (i-1)^{n_i}$ nous donne la probabilité d'obtenir une transposition non ordonnée.
3. Maintenant on calcule le nombre de transpositions non ordonnées ayant n_1 cycles de longueurs 1, ... etc Pour cela on choisit n_1 nombres parmi J que l'on range dans n_1 cases puis $2n_2$ parmi $J - n_1$ éléments que l'on range dans n_2 cases de deux éléments sans tenir compte de l'ordre des n_2 cases, etc.... On aboutit au nombre $\frac{J!}{\prod_{j=1}^J j^{n_j} n_j!}$
4. Enfin la multiplication des deux quantités donne le résultat.

Ensuite, on peut relier θ aux données du problème. Si l'on part de notre population actuelle de taille J , le dernier événement est l'arrivée d'un individu, il vient du pool avec probabilité m . Maintenant, si on considère la vision en coalescent, le dernier événement est aussi l'arrivée de cet individu, il vient du pool avec proba $\frac{\theta}{J-1+\theta}$.

Ainsi on en déduit $\frac{\theta}{J-1+\theta} = m$ et $\theta = \frac{(J-1)m}{1-m}$, ce qui est en accord avec le θ trouvé précédemment.

Remarques :

1. On notera avec la vision en coalescent que plus la population s'agrandit, plus l'arrivée de nouvelles espèces est rare.
2. La vision en coalescent fournit un algorithme pratique pour simuler le vecteur d'abondance à l'équilibre. (cf : principe des restaurants chinois)

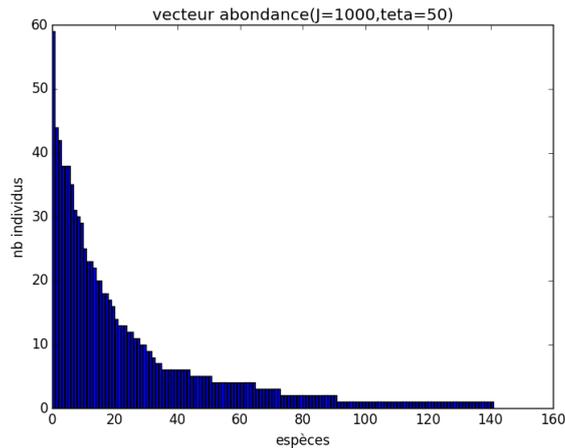


FIGURE 7 – simulation du vecteur d'abondance avec le principe des restaurants chinois ou urnes de polya.

2.5 Indice de Simpson

L'indice de Simpson correspond à la probabilité que deux individus pris au hasard soient de la même espèce. C'est la quantité qui nous servira par la suite pour mesurer et comparer les abondances entre elles.

2.5.1 Présentation

En se plaçant à l'équilibre, à partir du vecteur d'abondance, on souhaite calculer la probabilité que deux individus sélectionnés au hasard appartiennent à la même espèce. On notera D cette probabilité.

Cet indice est donné par la formule :

$$\sum_i \frac{X_i(X_i - 1)}{J(J - 1)}$$

où X_i est nombre d'individus de l'espèce i et J le nombre total d'individus.

Cet indice donne une indication sur la biodiversité, il aura une valeur de 0 pour indiquer le maximum de diversité (J espèces représentées par un individu

chacune), et une valeur de 1 pour indiquer le minimum de diversité (tous les individus sont de la même espèce). Dans le but d'obtenir des valeurs «plus intuitives», on préférera l'indice de diversité de Simpson représenté par $1-D$, le maximum de diversité étant représenté par la valeur 1, et le minimum de diversité par la valeur 0.

2.5.2 Indice de Simpson à l'équilibre

Comme dans notre modèle, le vecteur abondance se stabilise au bout d'un temps long, on peut calculer l'indice de Simpson pour la population à l'équilibre. On donnera deux façons pour obtenir ce résultat.

1) Première façon

On sait calculer la loi du vecteur d'abondance à l'équilibre (cf 2.4). Alors en notant $P(X_1, \dots, X_s)$ la probabilité d'obtenir le vecteur d'abondance (X_1, \dots, X_s) à l'équilibre :

On obtient

$$D = \sum_{\text{vecteurs}} P(X_1, \dots, X_s) \left(\sum_{i=1}^s \frac{X_i X_{i-1}}{J(J-1)} \right)$$

Cette formule n'est pas facile à simplifier ni à implémenter.

2) Seconde façon

On peut obtenir cet indice à partir du modèle de base qui évolue en temps. On se place au temps t , la probabilité que deux individus soient de la même espèce est la probabilité qu'ils aient été affectés par le dernier événement et que celui-ci soit une filiation, sinon il faut se référer au temps $t-1$.

Alors si on note D_t la probabilité qu'au temps t deux individus au hasard soient de la même espèce, on obtient :

$$\begin{aligned} D(t) &= \frac{2}{J(J-1)} \frac{j-1}{\theta + J + 1} + D(t-1) \left(\left(1 - \frac{2}{J(J-1)}\right) \frac{J-1}{J-1+\theta} + \frac{\theta}{\theta + J - 1} \left(1 - \frac{2}{J(J-1)}\right) \right) \\ &= \frac{2}{J(\theta + J - 1)} + D(t-1) \left(1 - \frac{2}{J(J-1+\theta)} - \frac{2\theta}{J(J-1+\theta)}\right) \end{aligned}$$

A l'équilibre, on a $D(t)=D(t-1)=D$ et en remplaçant dans l'équation précédente, on obtient :

$$D = \frac{1}{1+\theta}$$

On notera que cette valeur est la probabilité qu'il y ait eu filiation dans la première étape de simulation du coalescent.

2.6 Résultats numériques

On a implémenté la population à l'équilibre via un l'algorithme des restaurants Chinois, on a calculé l'indice de Simpson à partir du vecteur d'abondance

via la formule $\sum_{i=1}^s \frac{n_i n_{i-1}}{J(J-1)}$.

On compare cela avec le résultat théorique $\frac{1}{1+\theta}$ pour différentes valeurs de θ (en abscisse).

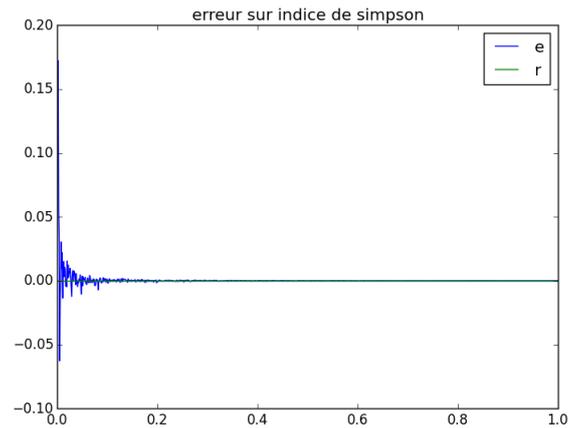


FIGURE 8 – Distance entre la valeur théorique et la valeur expérimentale.

Remarques :

1. On remarquera que la variance est plus faible si θ est grand i.e si m est grand .
2. En général θ est de l'ordre de 50 au moins dans la réalité, pour une telle valeur, l'erreur commise en utilisant le principe des restaurants chinois pour simuler le vecteur d'abondance est relativement faible. Ceci est donc un bon algorithme pour simuler le vecteur d'abondance si on s'intéresse à l'indice de Simpson.

2.7 Conclusion

On est parti d'un modèle composé d'une communauté alimentée par un pool uniforme. On a réussi à mettre en évidence un état d'équilibre pour le vecteur d'abondance, à calculer la loi de ce vecteur à l'équilibre et à trouver un algorithme capable de le simuler convenablement. Reprenons alors notre modèle de base et essayons de l'améliorer ...

3 Un modèle plus général

3.1 Introduction

Le modèle avec pool uniforme est limité, tout d'abord il ne représente pas bien la réalité. Le pool modélise l'environnement extérieur à la communauté étudiée, il est normal qu'il soit soumis aux mêmes règles environnementales que la communauté mais à une échelle différente, c'est à dire que des individus naissent et meurent, et que certaines espèces apparaissent ou disparaissent.

3.2 Présentation du modèle

Nous allons considérer maintenant un modèle un peu plus complexe. Dans ce modèle, le pool évolue de la même façon qu'évoluait la communauté dans le modèle précédent (sa taille sera R et le taux d'immigration θ) La communauté évolue aussi mais à taux I et est de taille J .

On suppose que :

- la taille du pool est très grande devant la communauté ce qui conduit à considérer R infini.
- le pool et la communauté évoluent avec deux échelles de temps différentes, le pool évoluant très lentement par rapport à la communauté. Ainsi, le pool est dans un état d'équilibre, celui que nous avons mis en évidence dans la partie précédente.
- les hypothèses de neutralité (au sein des individus d'une même espèce et entre les espèces) sont conservées

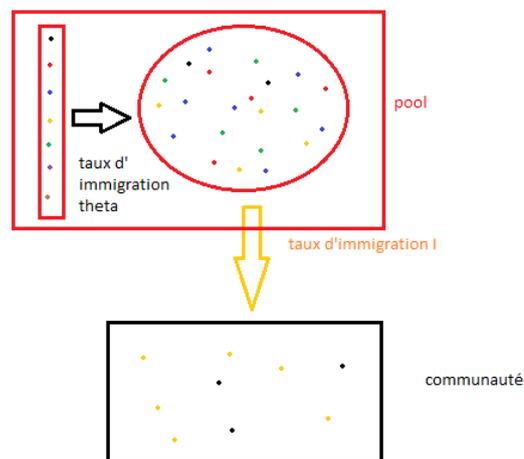


FIGURE 9 – Schéma résumant le modèle neutre.

3.3 Algorithme et résultats

3.3.1 Présentation de l'algorithme

Nous allons ici détailler l'algorithme permettant de simuler le modèle précédent :

1. Entrée :
Population de taille J , θ, I, T
2. Evolution :
On va effectuer T étapes.
A chaque étape :
 - un mort est choisi au hasard dans la population :
 - avec la probabilité m , le mort est remplacé par un individu du pool (cf pool) et avec probabilité $1-m$, il est remplacé par un individu de la communauté.
 - le pool se construit au fur et à mesure qu'on y pioche des éléments, selon le principe des restaurants chinois :
le k ème individu arrivant dans le pool est avec proba $\frac{\theta}{\theta+k-1}$ d'une nouvelle espèce ou avec $\frac{1}{\theta+k-1}$ le fils d'un individu donné du pool. Dans les deux cas, c'est cet individu qui arrivera dans la communauté, de sorte qu'au bout d'un grand nombre de tirages, l'abondance des individus injectés dans la communauté soit bien celle qu'on aurait obtenue via le principe des restaurants chinois.
3. Sortie : population de taille J

Dans la suite la valeur de J sera de 100 (il serait plus réaliste de prendre $J=1000$ mais pour des raisons de temps de calcul on prendra $J=100$), θ vaudra 60 et s 200.

3.3.2 Quelques résultats de simulations

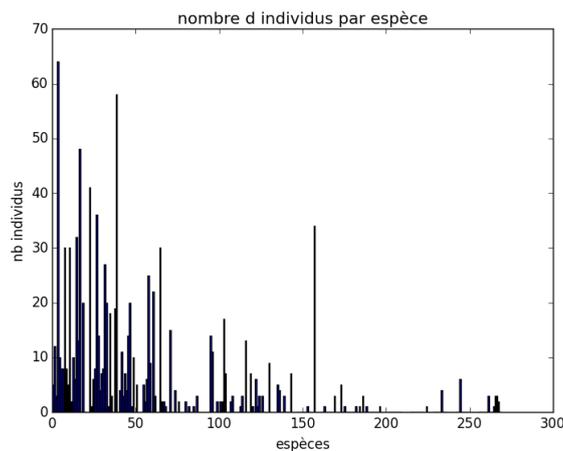


FIGURE 10 – abondance.

Voici le nombre d'individus en fonction des espèces, ou encore l'abondance de la population simulée via l'algorithme précédent.

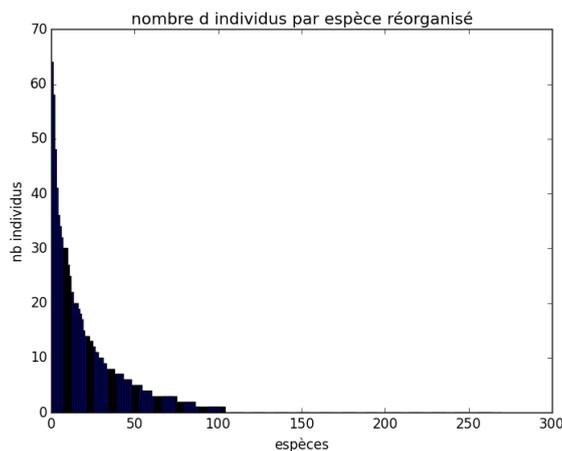


FIGURE 11 – abondance réorganisée

Ces résultats ne nous disent rien à priori il n'est pas évident de comparer ces abondances entre elles.

3.4 Mise en équation

Comme précédemment ce problème se met en équation. Rien ne change sinon les P_i qui sont maintenant une variable aléatoire de loi connue. (cf : restaurants chinois)

$$\begin{cases} P(N_i - 1|N_i) = \left(\frac{N_i}{J}\right)(m(1 - P_i) + (1 - m)\left(\frac{J - N_i}{J - 1}\right)) \\ P(N_i + 1|N_i) = \left(\frac{J - N_i}{J}\right)(mP_i + (1 - m)\left(\frac{N_i}{J - 1}\right)) \\ P(N_i|N_i) = 1 - P(N_i - 1|N_i) - P(N_i + 1|N_i) \end{cases}$$

On a encore existence d'un état d'équilibre. On va, par la suite, calculer la probabilité d'obtenir une abondance donnée.

3.5 La loi du vecteur d'abondance

Cette partie a été très largement inspirée de l'article "A novel genealogical approach to neutral biodiversity theory" (cf biblio)

3.5.1 La vision en coalescent

Comme dans le modèle précédent il est possible de reconstruire une généalogie à partir d'une population donnée.

Pour chaque individu, on remonte l'arbre généalogique jusqu'à obtenir un individu provenant du pool (un tel migrant existe si on considère le temps infini dans le passé).

Une fois que l'on a obtenu l'ensemble des ancêtres présents au nombre de A , on remonte la généalogie de ces individus dans le pool jusqu'à obtenir un ancêtre commun (qui existe si on considère encore le temps infini dans le passé).

Si on suppose, comme on l'a déjà dit, que l'évolution du pool est très lente devant l'évolution de la communauté alors, avec très grande probabilité, le pool n'aura pas évolué dans l'intervalle de temps existant entre l'arrivée du premier et l'arrivée du dernier ancêtre.

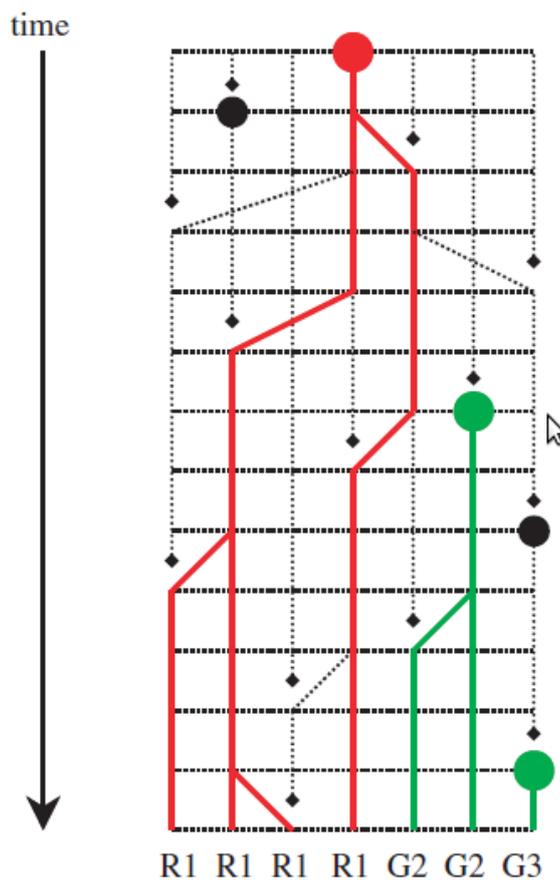


FIGURE 12 – Schéma résumant la coalescence.

Ainsi on se ramène à un coalescent formé de deux parties une évoluant dans la communauté et l'autre dans le pool.

3.5.2 La loi du vecteur d'abondance

Comme précédemment, la vision en coalescent va nous permettre d'établir la loi du vecteur d'abondance.

Supposons alors la population de taille J fixée. On va définir pour cela les quantités suivantes :

- A=nombre d'ancêtres
- S=nombre d'espèces (S est le même que l'on considère le pool ou la communauté)
- J_i le nombre d'individus de l'espèce i parmi les J
- a_i le nombre d'ancêtres de l'espèce i.
- $a_{i,j}$ le jème ancêtre de l'espèce i.
- $n_{i,j}$ le nombre de descendants de l'ancêtre j de l'espèce i
- ψ_l le nombre d'espèces dans la communauté représenté par l individus
- $\phi_{i,j}$ nombre d'ancêtres de l'espèce i ayant j descendants

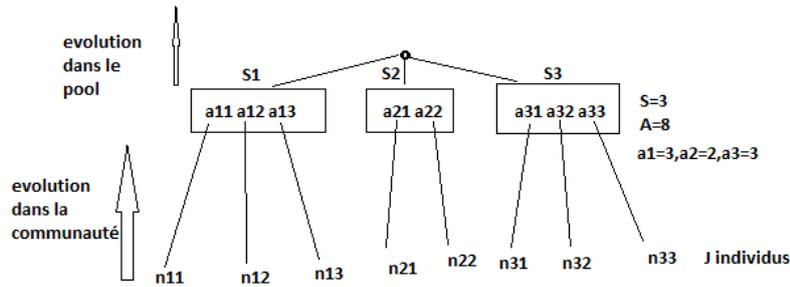


FIGURE 13 – Schéma résumant les données.

On cherche à calculer la probabilité d'obtenir une abondance particulière représentée sous la forme d'un vecteur j_S (nombre d'individus par espèce). Cependant pour obtenir cette probabilité il faut remonter la généalogie des individus qui représentent cette abondance. Comme elle est en deux parties, on a besoin de connaître le nombre d'ancêtres et combien chacun possède de descendants. Pour résumer cette information, on va considérer des distributions de la forme $\sigma = ((n_{1,1} \dots n_{1,a_1}), \dots, (n_{S,1} \dots n_{S,a_S}))$.

Notons qu'il est équivalent de considérer des vecteurs de la forme $n_S = ((\phi_{1,1}, \dots, \phi_{1,J_1}), \dots, (\phi_{S,j}, \dots, \phi_{S,J_S}))$. Cependant plusieurs de ces vecteurs mènent à la même abondance j_s . Il faut donc sommer leurs probabilités. On obtient alors :

$$P(j_S) = \sum_{n_S \text{ qui " conviennent "}} P(n_S)$$

La somme étant sur les σ menant à l'abondance désirée.

Par ex, si $j_s = (3, 2, 1)$, les σ qui mènent à cette abondance sont $\sigma = ((3), (2), (1)), \sigma = ((3), (1, 1), (1)), \sigma = ((2, 1), (2), (1)), \sigma = ((2, 1), (1, 1), (1)), \sigma = ((1, 1, 1), (2), (1))$ et $\sigma = ((1, 1, 1), (1, 1), (1))$

Il reste maintenant à calculer $P(n_S)$.

1. Pour cela, on va utiliser des transpositions de transpositions non ordonnées. C'est à dire un élément de la forme $((4, 2, 1)(5, 3), (6)(8, 9))$ par exemple, ce dernier mènerait au σ suivant $((3, 2), (1, 2))$ et à l'abondance $(5, 3)$.

Ainsi, $P(n_S) = P(\text{d'obtenir une transposition de transpositions non ordonnée donnée menant à cette (Nombre de transpositions de transpositions à J individus, non ordonnée menant à cette distribution$

2. Calcul du nombre de transpositions de transpositions non ordonnées d'un ensemble à J éléments menant au n_S désiré, noté NJ :

Cela revient à ranger ces individus dans des tiroirs (de taille $\phi_{i,j}$) de tiroirs (de taille J_i) sans tenir compte de l'ordre des tiroirs de même taille. On obtient :

$$NJ = C_J^{\psi_1 J_1} \frac{C_{\psi_1 J_1}^{J_1} (C_{J_1}^{1\phi_{1,1}} C_{J_1 - \phi_{1,1}}^{2\phi_{1,2}} [\frac{C_{2\phi_{1,2}}^{2\phi_{1,1,2}} C_{2\phi_{1,2} - 2 \dots 1}^{\dots 1}}{\phi_{1,2}!}] \dots 1)}{\psi_1!} \times C_{J - \psi_1 J_1}^{\psi_2 J_2} \dots \times \dots \times 1$$

On a choisi dans un premier temps les éléments à ranger dans les ψ_1 'grands' tiroirs de taille J_1 . Puis, on en choisit J_1 pour mettre dans le premier grand tiroir. Enfin, on répartit ces J_1 dans les $(\phi_{1,1}, \dots, \phi_{1,J_1})$ tiroirs qui sont de taille $1, \dots, J_1$ sans tenir compte de l'ordre. (cf formule d'Ewens)

On recommence pour le second grand tiroir jusqu'au ψ_1 ème. Enfin, comme on ne tient pas compte de l'ordre dans lequel on met les éléments dans les tiroirs de taille J_1 , on divise par $\psi_1!$

On recommence pour les 'grands' tiroirs de taille J_2 etc...

En simplifiant la formule précédente on aboutit à :

$$NJ = \frac{J!}{\prod_{i=1}^S [\prod_{j=1}^{J_i} (j!)^{\phi_{i,j}} \phi_{i,j}!] \prod_{l=1}^S \psi_l}$$

3. Maintenant on va calculer la probabilité d'obtenir une transposition de transposition non ordonnée donnée σ , notée $P\sigma$.

Alors, il faut considérer l'arbre en deux étapes et faire intervenir les ancêtres. Quand on remonte la généalogie dans la communauté, chacun des ancêtres intervient à des temps différents. Assimilons les distributions des ancêtres à une permutation non ordonnée où chaque cycle représente une espèce et le numéro de l'ancêtre son ordre d'arrivée dans la communauté.

Par ex $(1, 4, 5)(3)(2, 6)$ signifie que les 1er, 4ème et 5ème ancêtres arrivés sont de l'espèce 1 ; le 3ème ancêtre arrivé est de l'espèce 2 etc...

Si on note σ_A la variable aléatoire à valeurs dans l'ensemble de ces transpositions non ordonnées :

$$P\sigma = P(\sigma | \sigma_A) \times P(\sigma_A)$$

— Le calcul de $P(\sigma_A)$:

Cette probabilité est exactement la probabilité d'obtenir une transposition non ordonnée avec S cycles, composée de A éléments, via le principe des restaurants chinois avec un taux θ

$$P(\sigma_A) = \frac{\theta^S}{\theta_A} \prod (i-1)!^{\alpha_i} \quad \text{où } \alpha_i \text{ est le nombre d'espèces ayant } i \text{ ancêtres.}$$

— Le calcul de $P(\sigma|\sigma_A)$:

Cette probabilité est la probabilité d'obtenir σ connaissant l'ordre d'arrivée et l'espèce des ancêtres. Alors, si ces informations sont connues le coalescent devient un arbre simple, partant du premier ancêtre arrivé et où chaque ancêtre intervient dans la communauté comme une nouvelle espèce intervient dans le modèle avec pool constant.

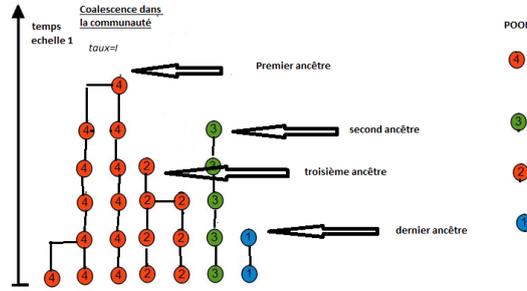


FIGURE 14 – Schéma de coalescence avec les ancêtres.

Alors $P(\sigma|\sigma_A)$ est la probabilité d'obtenir une transposition non ordonnée particulière ayant A cycles et composée de J éléments via le modèle des restaurants Chinois avec taux I. On a donc :

$$P(\sigma|\sigma_A) = \frac{I^A}{I_J} \prod (i-1)!^{\beta_i} \quad \text{où } \beta_i \text{ est le nombre d'ancêtres ayant } i \text{ descendants.}$$

4. Enfin si on réunit toutes ces étapes on obtient :

$$P(j_S) = \sum_{n_S \text{ qui "conviennent"}} \frac{J!}{\prod_{i=1}^S [\prod_{j=1}^{J_i} (j!)^{\phi_{i,j}} \phi_{i,j}!]} \times \frac{\theta^S}{\theta_A} \prod (i-1)!^{\alpha_i} \times \frac{I^A}{I_J} \prod (i-1)!^{\beta_i}$$

Cette expression n'est pas pratique, il est délicat de lister les ns qui conviennent dès que l'on a un nombre d'espèces assez grand.

3.6 Indice de Simpson pour ce modèle

3.6.1 En théorie

Rappelons que l'indice de Simpson à l'équilibre est la probabilité que deux individus, pris au hasard dans la population à l'équilibre, soient de la même espèce, on le note D .

Il est alors possible de calculer la probabilité que deux individus pris au hasard ne soient pas de la même espèce.

En effet, pour cela, il faut et il suffit que ces individus n'aient pas le même ancêtre et que leurs ancêtres respectifs ne soient pas de la même espèce. Ainsi comme ces deux événements sont supposés indépendants on obtient : $1 - D = \frac{\theta}{1+\theta} \times \frac{I}{1+I}$. Le calcul est analogue à celui exposé précédemment en considérant les deux coalescents différents (cf pool constant).

On notera la présence du facteur multiplicatif $\frac{\theta}{1+\theta}$ en plus par rapport au modèle avec pool uniforme. Quand θ est grand ce facteur vaut presque 1 et on doit être proche de la situation précédente. En effet θ grand signifie que l'on favorise l'apparition de nouvelles espèces et donc le pool est proche d'un pool uniforme. Regardons sur quelques exemples comment ceci influe :

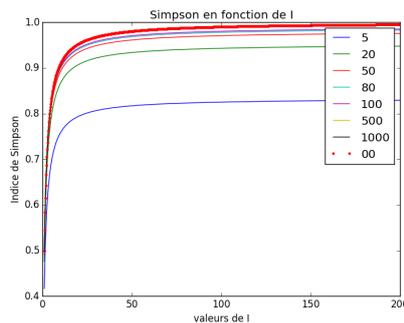


FIGURE 15 – Comparaison des indices de Simpson des deux modèles (théoriques). On a tracé ici 1-D pour différentes valeurs de θ

On notera que ce modèle a tendance à augmenter l'indice de Simpson par rapport au modèle avec pool uniforme. Ceci signifie que certaines espèces sont représentées avec un plus grand nombre d'individus. En effet, une espèce sera en proportion dominante dans le pool et elle aura alors tendance à envahir davantage la communauté.

3.6.2 En pratique

On a calculé l'indice de Simpson après avoir fait évoluer la population via l'algorithme présenté précédemment jusqu'à un temps $T=500000$. On a ensuite calculé la différence entre la valeur théorique et la valeur expérimentale pour différentes valeurs de θ . Voici les résultats obtenus :

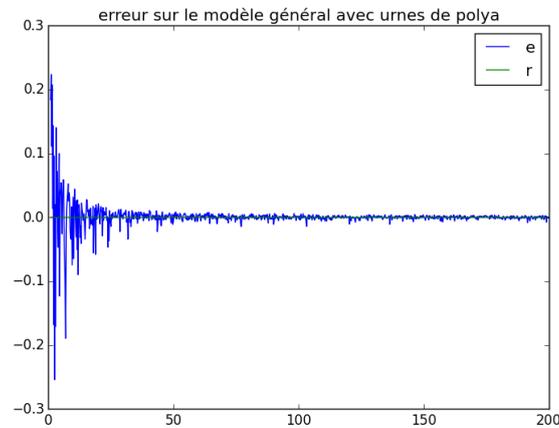


FIGURE 16 – Comparaison des 1-indices de Simpson des deux modèles (théoriques).

Remarques :

1. Pour obtenir une erreur plus faible, on aurait pu faire une moyenne sur un grand nombre de simulations cependant ceci est coûteux. Ici, on n'a effectué qu'une simulation pour chaque θ .
2. On remarquera que la variance est plus faible si θ est grand .
3. En général θ est de l'ordre de 50 au moins dans la réalité. Ainsi, pour une telle valeur, l'erreur commise en ne faisant qu'une simulation pour calculer l'indice de Simpson moyen est relativement faible.

3.7 Avantages et inconvénients de ce modèle

L'intérêt principal de ce modèle neutre est qu'il prédit des distributions d'abondance d'espèces assez proches de celles que l'on observe dans la réalité. (cf Volkov 2003, Nature)

En revanche, ces modèles prédisent des durées de vie d'espèces irréalistes. Par exemple, Ricklefs (2006) estime qu'une espèce de passereaux européens, composée de 14.43×10^8 individus, et ayant une espérance de vie de 3 ans mettrait (selon ce modèle neutre) 86×10^6 années pour s'éteindre. Ceci est clairement beaucoup trop long.

Il semblerait alors que des mécanismes autres que la dynamique stochastique décrite précédemment jouent un rôle prépondérants. Il s'agirait de prendre en compte dans notre modèle les variations environnementales comme les changements climatiques, l'apparition ou la disparition de prédateurs ou de maladies spécifiques (ou autres phénomènes extérieurs) favorisant ou défavorisant certaines espèces.

4 Un modèle neutre avec variance environnementale

On va ici tenter d'intégrer dans notre modèle de base ce phénomène de variance environnementale en attribuant à chaque espèce un poids (fitness) qui lui donne un avantage sélectif.

Par souci de neutralité, on réinitialisera fréquemment les fitness de chaque espèce, de façon aléatoire, de sorte qu'aucune ne soit favorisée ou défavorisée durant une longue période.

Ce modèle a été introduit par Michael Kalyuzhny, Ronen Kadmon and Nadav M. Shnerb dans l'article "A neutral theory with environmental stochasticity explains static and dynamic properties of ecological communities" (cf bibliographie).

On peut déjà imaginer que si des espèces ont un poids faible elles risquent de disparaître rapidement, à l'inverse si elles ont un poids élevé elle vont envahir la communauté plus vite. Ainsi en jouant sur les paramètres (s , m , θ , durée entre les réinitialisations ...) on pourrait obtenir des durées de vie plus réalistes. On va ici se concentrer sur les abondances : "quels sont les effets des paramètres sur l'indice de Simpson ?"

4.1 Le modèle

On conserve le modèle précédent, à savoir :

- Pool constant à l'équilibre dont l'abondance est celle donnée par les restaurants chinois. ($\theta = 60$, taille très grande)
- Communauté évoluant à taux I , taille J .

On prend en compte le fait que certaines espèces sont plus adaptées à leur environnement, ceci leur confère un avantage sélectif au moment de se reproduire et seulement à ce moment là, le mort étant toujours choisi uniformément :

- Maintenant les individus de la communauté ont une 'fitness'. Cette fitness se représente par un poids noté f_i (pour l'espèce i) propre à chaque espèce et identique pour tous les représentants de la même espèce. Quand le mort est remplacé par le fils d'un individu de la communauté, ce dernier est choisi avec probabilité $F_i^t = \frac{N_i^t f_i}{\sum_{j=1}^S N_j^t f_j}$ où N_i est le nombre d'individus de l'espèce i au temps t .
- Tous les P pas de temps les poids de chaque espèce sont retirés au hasard selon un log-normale(0, s) de façon indépendante. (neutralité en globalité) Le rôle de s est bien sûr important nous reviendrons sur ce point par la suite.

4.2 Mise en équation

Donnons nous une espèce i et notons X_i^t la variable aléatoire du nombre d'individus qui représentent cette espèce dans la communauté au temps t , notons encore f_i^k la fitness de l'espèce i au k ème changement. Alors on a :

$$\begin{cases} P_k(X_i^{t+1} = Ni - 1 | X_j^t = N_j, \forall j) = \left(\frac{N_i}{J}\right)(m(1 - R_i) + (1 - m)(1 - F_i^t)) \\ P_k((X_i^{t+1} = Ni + 1 | X_j^t = N_j, \forall j) = \left(\frac{J - N_i}{J}\right)(mR_i + (1 - m)(F_i^t)) \\ P_k((X_i^{t+1} = Ni | X_j^t = N_j, \forall j) = 1 - P_k(X_i^{t+1} = Ni - 1 | X_i^t = Ni) - P_k((X_i^{t+1} = Ni + 1 | X_i^t = Ni) \end{cases}$$

Où R_i est la proportion de l'espèce i dans le pool et $F_i^t = \frac{N_i^t f_i}{\sum_{j=1}^S N_j^t f_j}$

Ce modèle est à première vue assez complexe, en effet les espèces n'étant plus toutes équivalentes, il est nécessaire de connaître l'état de toute la population pour déterminer les probabilités de transition d'une espèce en particulier.

Néanmoins pour se donner une idée de l'influence des paramètres et de la pertinence de ce modèle on peut en faire un algorithme et l'implémenter sur machine, puis regarder ce que ça donne.

L'indice que l'on va utiliser pour analyser et comparer des abondances est toujours l'indice de Simpson, ainsi on pourra regarder l'influence des paramètres sur cette quantité.

4.3 Premiers exemples

Nous allons ici présenter des trajectoires et essayer de les analyser. On part sur la base de 5 espèces, représentées par 20 individus chacune. s vaut 2 (il est un peu plus grand que ce qu'il faut mais les changements sont plus marqués avec un s grand), m vaut 0.1.

On représente le poids total, l'indice de Simpson et le nombre d'espèces au cours du temps. Un changement de fitness est effectué au temps 4000.

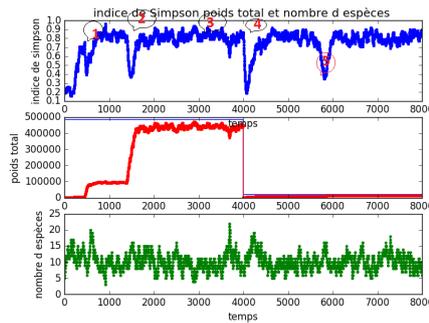


FIGURE 17 – Evolution de l'indice de Simpson, du poids et du nombre d'espèces au cours du temps.(une seule trajectoire)

- en 1, 2 et 5 : apparition d'une nouvelle espèce dans la communauté ayant un poids plus important que celles déjà présentes. Ceci se traduit par une chute de l'indice de Simpson et une augmentation du poids total. En effet l'espèce qui avait précédemment un poids dominant va toujours

être choisie pour mourir alors que celle qui arrive avec un poids supérieur sera choisie plus probablement pour remplacer le mort. On passe d'une population représentée en partie par une espèce (indice fort=1) à une population représentée par deux espèces (indice moyen=1/2) puis on retourne à un indice de 1, d'où la chute observée.

- en 3 : l' espèce qui a le poids le plus grand a envahi la communauté. Elle ne représente pas la totalité des individus car le pool entretient de la diversité. On est dans un régime où l'indice de Simpson se stabilise en moyenne
- en 4 : réinitialisation des fitness. On a une chute brutale de l'indice de Simpson. De nouvelles espèces vont prendre le dessus celles qui sont favorisées par le changement de fitness, en revanche leurs ascensions sont limitées par le paramètre m qui force à choisir les remplaçants des morts dans le pool.

Un changement de fitness entraîne une chute de l'indice de Simpson Que se passe t-il si les changements sont assez fréquents ?

On peut alors espérer obtenir une diversité plus importante c'est à dire un indice de Simpson plus faible. On va ici simuler plusieurs changements de fitness intervenant tous les 200 pas de temps c'est à dire tous les 2J événements ; puis on laisse relaxer le système.

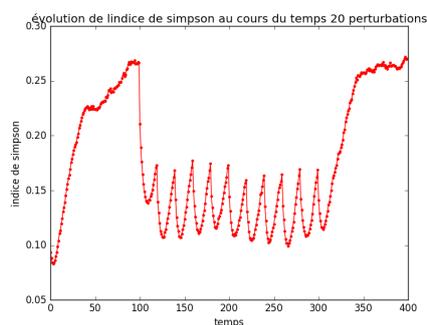


FIGURE 18 – Indice de Simpson en fonction du temps. On a simulé 100 trajectoires avec les mêmes paramètres et on a relevé l'Indice de Simpson tous les 10 pas de temps. Enfin on a fait une moyenne de l'indice de Simpson à chaque temps sur toutes les trajectoires.

Au vu de ces résultats on peut imaginer que des changements fréquents (de l'ordre de J pas de temps) de fitness peuvent entretenir un indice de Simpson faible. C'est à dire d'un point de vu biologique que des changements fréquents du milieu entretiennent de la biodiversité.

Il faudrait maintenant pouvoir estimer cette variation de l'indice de Simpson en fonction des paramètres.

4.4 Le paramètre m

Le paramètre m permet de conserver de la biodiversité. Son rôle est le même que dans le chapitre précédent.

Notons tout de même deux points :

- Premièrement, la probabilité qu'à chaque étape, l'individu remplaçant le mort, soit tiré dans le pool, n'est pas influencée par le poids des espèces de la communauté. Ainsi, des espèces avec des poids faibles seront toujours présentes simplement par immigration.
- Lors d'un changement de fitness, l'espèce avec le poids le plus grand n'est pas forcément présente dans la communauté. Si m est assez grand, des espèces avec des poids de plus en plus grands vont apparaître successivement après le changement de fitness. Ceci aura pour conséquence de faire chuter de nouveau l'indice de Simpson dans un voisinage proche du changement de fitness.

Si on fait, par la suite, une moyenne sur les trajectoires, cette moyenne aura une forme d'autant plus évasée que m est petit, c'est à dire que les espèces à fort poids arrivent tard. Ainsi, la moyenne n'est pas représentative du temps nécessaire à une espèce pour envahir la communauté.

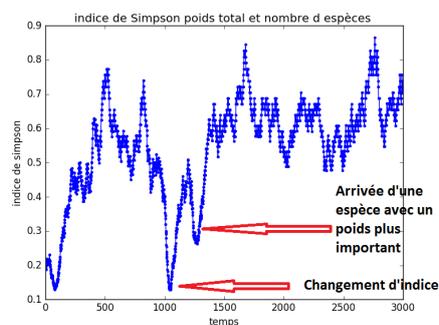


FIGURE 19 – Une trajectoire représentée pour $m=0.1$. On remarque une chute après le changement de fitness.

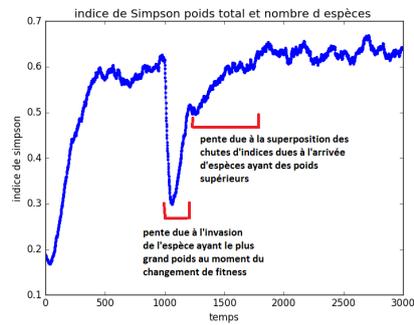


FIGURE 20 – On a ici fait une moyenne sur 100 simulations pour $m=0.1$. La forme "évasée apparaît clairement".

Pour se convaincre de l'influence de m , on va mesurer la pente de la courbe de relaxation, c'est à dire la taille de la perturbation divisée par le temps mis pour retrouver l'équilibre.

Pour $s=1$ et différentes valeurs de m , on fera une moyenne sur 100 trajectoires . Attention, cette pente n'est qu'une valeur indicative la courbe n'est pas linéaire ici !!

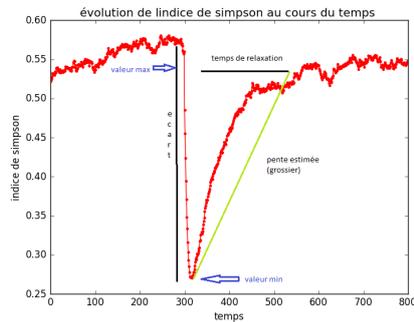


FIGURE 21 – figure représentant les quantités mesurées

m	val min	val max	temps	écart	pente*10000
0.001	*	*	*	*	*
0.01	0.78	0.94	540	0.16	3
0.03	0.66	0.87	480	0.21	4
0.05	0.59	0.85	420	0.26	6
0.08	0.46	0.79	480	0.37	7
0.1	0.45	0.75	375	0.30	8
0.2	0.23	0.57	380	0.34	7
0.3	0.16	0.42	340	0.26	7
0.4	0.10	0.27	270	0.17	6
0.5	0.06	0.18	280	0.12	4
0.6	0.04	0.10	140	0.06	4
0.7	0.02	0.05	100	0.03	3

TABLE 1 – évolution de la pente avec m, s=1, moyenne sur 100 trajectoires

Remarques : La pente de la courbe est faible, elle augmente puis diminue de nouveau au fur et à mesure que m augmente.

Pour m très petit, la pente faible est due à la superposition décrite précédemment. Comme prévu, elle augmente avec m, autrement dit l'espèce qui a le poids le plus grand met généralement moins de temps pour arriver dans la communauté. Puis la pente diminue de nouveau si m augmente. En effet pour m assez grand le fait d'aller chercher fréquemment des individus dans le pool ralentit l'invasion de l'espèce ayant le poids le plus important.

4.5 Le paramètre s

4.5.1 Influence de s sur la loi log normale

Rappelons que le paramètre s intervient dans l'attribution des fitness de chaque espèce.

C'est le "plus" par rapport au modèle précédent, il est alors normal de regarder son influence sur l'indice de Simpson.

On peut commencer par voir comment se comporte la loi log-normale avec s (notons que Y suit une log-N(u,s) ssi $X=\log(Y)$ suit une loi N(u,s)) :

On va tracer la densité et la fonction de répartition de la loi log normale pour différentes valeurs de s.

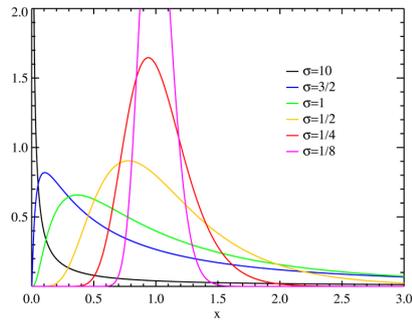


FIGURE 22 – figure représentant la densité de la loi log normale

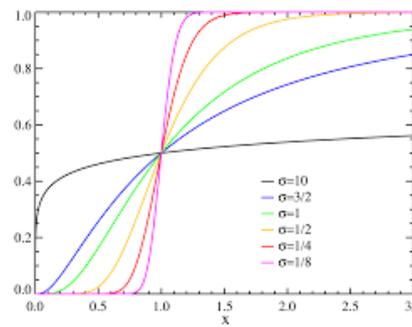


FIGURE 23 – figure représentant la fonction de répartition de la loi log normale pour différentes valeurs de s

Pour s petit (< 0.25) la variance de la loi est très faible et avec grande probabilité les valeurs obtenues pour les fitness seront proches les unes des autres. En revanche pour s grand (> 1) les valeurs de nos fitness seront très dispersées, ainsi une valeur dominera largement toutes les autres.

Le premier cas représente plutôt une population dans un état d'équilibre (cas qui nous intéresse) soumise à des changements environnementaux. Le second cas correspond à un phénomène d'invasion. On prendra alors des s de l'ordre de 0.1 dans la suite.

De plus, pour s très petit l'indice de Simpson du modèle précédent devrait être proche de celui sans variance environnementale à savoir $\frac{\theta}{1+\theta} \times \frac{I}{1+I}$:

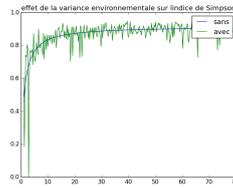


FIGURE 24 – Est représenté l'indice de Simpson du modèle sans fitness en bleu, et l'indice mesuré avec $s=0.1$ pour différentes valeurs de θ en vert au bout d'un grand nombre d'étapes.

Chaque point vert correspond à une simulation, on n'a pas fait de moyenne c'est pour cela que les perturbations sont si marquées.

4.5.2 Influence de s sur la variance de l'indice de Simpson

On vient de voir que plus s est grand plus les poids vont être dispersés, par exemple pour $s=1$ certaines espèces ont des poids 5 fois supérieurs aux autres. Alors si une de ces espèces fortement pondérée envahie la communauté son poids sera très grand et elle sera choisie presque assurément pour remplacer le mort à chaque étape. Les trajectoires seront alors plus proches de leur moyenne.

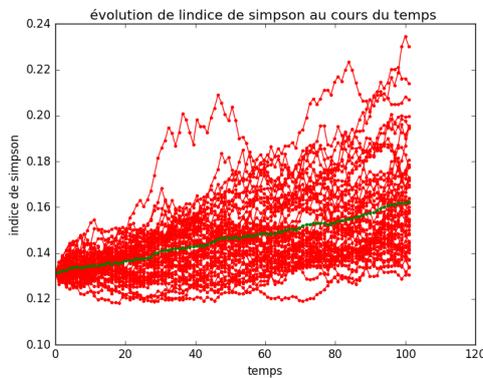


FIGURE 25 – $s=0.5$

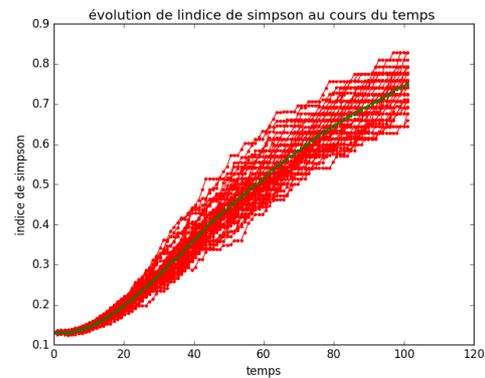


FIGURE 26 – $s=2$

Alors pour s grand, la variance de l'Indice de Simpson sera très probablement plus faible.

4.5.3 Influence de s lors d'un changement de fitness.

Nous allons regarder l'effet de s sur les trajectoires. Pour cela nous allons pour différentes valeurs de s tirer des fitness et simuler 20 trajectoires (pour chaque tirage). On fera une moyenne point par point sur ces trajectoires :

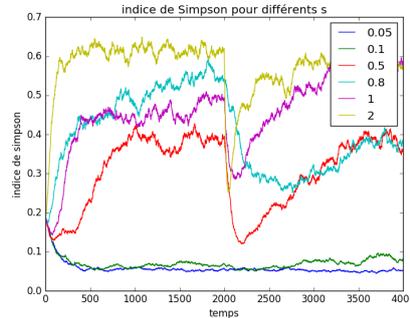


FIGURE 27 – est représenté l'indice de Simpson pour différents s . ici, $m=0.1, J=100, \theta=60$

Remarques :

- Tout d'abord remarquons que si s est très petit, (< 0.1) l'effet d'une perturbation est presque toujours négligeable. En effet, si la variance s est faible alors toutes les espèces ont probablement la même fitness. On est donc proche du modèle sans variance environnementale.
- Ensuite, la valeur du seuil est d'autant plus importante que s est grand, c'est à dire qu'un petit nombre d'espèces est représenté par un grand nombre d'individus. Une espèce ayant un poids plus grand aura tendance à envahir la communauté.
- L'amplitude de la chute est d'autant plus importante que s est grand. En effet, pour s grand, on passe d'un indice avoisinant 1, à un indice de $1/2$.
- La pente, lors de la chute d'indice, ne semble pas dépendre de s .
- En revanche, plus s est grand, plus la pente de la courbe après la chute semble importante. En d'autres termes, l'indice de Simpson retrouve sa valeur seuil plus rapidement. En effet, si s est grand, une espèce possède un poids qui domine tous les autres et elle sera choisie avec une plus grande probabilité pour remplacer le mort et donc son invasion est plus rapide.

Les simulations nous permettent de nous faire une idée sur le modèle, il faudrait maintenant essayer de trouver une approche plus théorique.

4.6 Une approche déterministe

Dans cette partie, on va tenter d'approcher, étape après étape, la moyenne de nos trajectoires en assimilant à chaque étape la valeur de l'indice de Simpson à sa valeur moyenne.

Supposons que l'on parte d'une condition initiale connue, c'est à dire que l'on connaît tous les $X_j^t, \forall j$, alors on est capable, pour chaque espèce, de calculer l'espérance de son nombre d'individus au temps suivant, c'est à dire $E[X_i^{t+1}|X_j^t, \forall j]$.

$$\text{En effet, } E[X_i^{t+1}|X_j^t, \forall j] = (X_i^t + 1) \times P[X_i^{t+1} = X_i^t + 1|X_j^t, \forall j] + (X_i^t - 1) \times$$

$$P[X_i^{t+1} = X_i^t - 1 | X_j^t, \forall j] + (X_i^t) \times P[X_i^{t+1} = X_i^t | X_j^t, \forall j]$$

Bien que l'on connaisse l'espérance du nombre d'individus au temps suivant, on ne peut savoir l'état de la population, puisque c'est aléatoire. On va alors supposer que l'état de la population, au temps suivant n'est pas très loin de son espérance, de sorte que l'on assimile $E[X_j^t]$ à $X_j^t, \forall j$.

On a alors :

$$\begin{aligned} E[X_i^{t+1} | X_i^t = E[X_j^t], \forall j] = \\ (E[X_i^t] + 1) \times P[X_i^{t+1} = E[X_i^t] + 1 | X_j^t = E[X_j^t], \forall j] \\ + (E[X_i^t] - 1) \times P[X_i^{t+1} = E[X_i^t] - 1 | X_j^t = E[X_j^t], \forall j] \\ + (X_i^t) \times P[X_i^{t+1} = X_i^t | X_j^t = E[X_j^t], \forall j] \end{aligned}$$

Remarquons que J n'est plus constant, à savoir $J_t = \sum_i E[X_i^t]$

En calculant on obtient :

$$E[X_i^{t+1} | X_i^t = E[X_j^t], \forall j] = (m-1) \times \frac{E[X_i^t] f_i}{\sum_j f_j E[X_j^t]} + E[X_i^t] \times \left(1 - \frac{2m-1}{\sum E[X_j^t]}\right) + m R_i$$

Ainsi on peut calculer une valeur $E[X_i^{t+1} | X_i^s = E[X_i^s], \forall s]$ par récurrence pour tout temps t .

Cette méthode à l'avantage d'être déterministe, le seul aléa possible est dans l'attribution des poids aux différentes espèces. On espère pouvoir comparer le résultat obtenu par cette méthode déterministe à la trajectoire moyenne de notre modèle, à condition que dans chaque cas les poids soient les mêmes pour chaque espèce.

On va comparer cette méthode aux trajectoires obtenues avec l'algorithme. Pour cela, on partira d'une situation initiale connue, on attribuera des poids identiques aux espèces et on les laissera évoluer dans le temps. On étudiera deux cas :

- Le régime asymptotique sans changement de fitness.
- Ce qui se passe sur quelques pas de temps. (temps court)

On espère que notre modèle déterministe pourra approcher la moyenne des trajectoires.

Remarque : On a vu dans notre modèle de base que l'espèce ou les espèces ayant les poids les plus importants n'étaient pas forcément présentes dans la communauté à l'état initial, ou après un changement de fitness. C'est le paramètre m qui va influencer sur leur vitesse d'apparition. Ainsi le comportement d'une trajectoire peut être très différent de celui de la moyenne de plusieurs trajectoires.

4.7 Comparaison avec le modèle

4.7.1 En régime asymptotique

On a vu que si on laisse évoluer notre population sans changements fréquents de fitness, la moyenne de plusieurs trajectoires a tendance à se stabiliser vers une valeur seuil, est ce le cas de notre approche par l'espérance ?

On peut calculer cette valeur à l'équilibre à partir des équations :

On reprend le calcul précédent

$$E[X_i^{t+1}|X_i^t = E[X_j]^t, \forall j] = (m-1) \times \frac{E[X_i^t]f_i}{\sum f_j E[X_j]^t} + E[X_i^t] \times \left(1 - \frac{2m-1}{\sum E[X_j]^t}\right) + mR_i$$

mais maintenant si l'espérance atteint un état d'équilibre et ne dépend plus du temps on a $E[X_i] = E[X_i^t] = E[X_i^{t+1}], \forall i$.

Et $E[X_i]$ est solution du système :

$$0 = (m-1) \times \frac{E[X_i^t]f_i}{\sum f_j E[X_j]^t} - E[X_i^t] \times \left(\frac{2m-1}{\sum E[X_j]^t}\right) + mR_i$$

Or, il s'agit d'un système de S équations à S inconnues non linéaire et donc difficile à résoudre explicitement.

Néanmoins on peut comparer les deux valeurs des seuils numériquement :

Pour cela il suffit de se référer aux courbes précédentes.

Si des changements de fitness arrivent fréquemment, ce régime ne sera jamais atteint et ceci ne nous est pas d'une grande utilité. Pour avoir une idée plus précise de la validité de notre modèle, il est nécessaire d'observer son comportement sur des temps relativement courts.

4.7.2 Étude sur un temps court

Supposons m pas trop grand, de l'ordre de 0.05. Sur une centaine de pas de temps, en moyenne, seulement 5 espèces ont immigré du pool. Pour peu que ces 5 espèces soient déjà présentes dans la communauté, l'effet du pool est négligeable. Dans cette partie, on va supposer, désormais, que m vaut 0 et on va se concentrer sur ce qu'il se passe sur une centaine de pas de temps, tout au plus.

Bien que le fait de considérer m=0 simplifie le problème, il n'est toujours pas évident d'étudier mathématiquement ce modèle. On fait des simulations. On part d'une situation initiale (poids et population) connue (J=100), identique pour les deux modèles. On regarde leur évolution en temps court, environ 100 pas de temps. On mesure l'indice de Simpson tous les 2 pas de temps. Sur chaque graphique est représenté

en rouge les trajectoires obtenues via le modèle de base
en vert la moyenne de ces trajectoires et
en bleu la trajectoire du modèle avec espérance.

Quelques exemples :

- Dans le cas trivial qui correspond à une condition initiale pour laquelle la population n'est composée que d'un individu, les deux courbes restent constantes et égales à 1.
- Cas où $s=0$, c'est à dire que les poids sont uniformes. Remarquons déjà que si de plus les populations initiales sont uniformes (même nombre d'individus), alors l'indice de Simpson n'évolue pas dans le temps sur le modèle avec espérance. En effet, à chaque étape, chaque espèce est représentée par la même quantité d'individus.

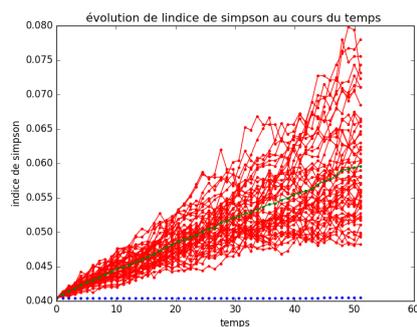


FIGURE 28 – Evolution de l'indice de Simpson, pour $s=0$, $m=0$, ci homogène

- un autre cas.

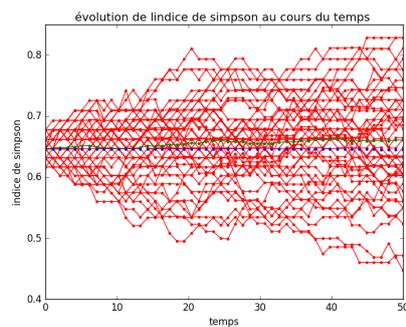


FIGURE 29 – Evolution de l'indice de Simpson, $s=0.1$, certaines espèces sont en plus grand nombre

Il est évident que notre approche avec l'espérance ne peut approcher correctement le modèle aléatoire dans certains cas. En effet, pour des espèces qui, a priori, ont les mêmes avantages sélectifs (fitness et nombre), c'est le hasard qui fait qu'une espèce va prendre le dessus sur les autres. Ainsi dans une situation proche de celle-ci (à savoir s petit et nombre d'individus à peu près identique) notre approche n'est plus valable.

On peut alors se demander dans quelles situations notre approche est appropriée ?

Plusieurs paramètres influencent le comportement de nos trajectoires : Comme m vaut 0, et comme on s'intéresse à des moyennes, l'aléa est contenu dans la situation initiale.

Alors deux informations sont à retenir, à savoir le nombre d'espèces représentées et l'état de la condition initiale. La condition initiale pouvant être homogène (toutes les espèces présentes ont le même poids) ou hétérogène (certaines espèces ont un poids bien supérieurs aux autres).

4.7.3 Le nombre d'espèces

On va maintenant faire varier le nombre d'espèces. Dans chaque cas, on présentera deux graphes, le premier pour une situation homogène et le second pour une situation hétérogène :

-graphes pour 20 espèces :

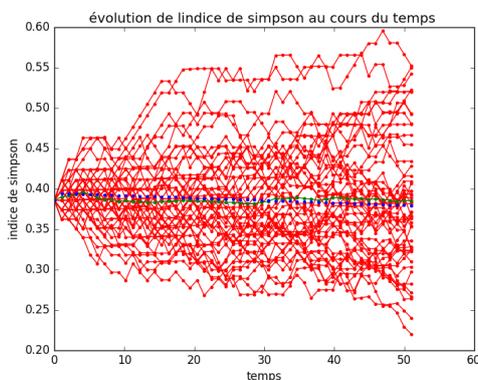


FIGURE 30 – $s=0$ avec une condition initiale hétérogène, 20 espèces représentées

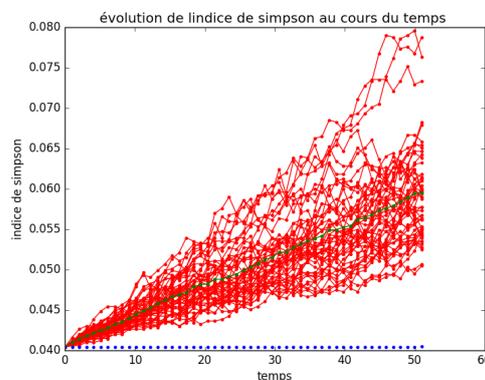


FIGURE 31 – $s=0$ avec une condition initiale homogène, 20 espèces représentées, $J=100$

-les graphes pour 10 espèces sont déjà présentés au dessus.

-graphes pour 5 espèces :

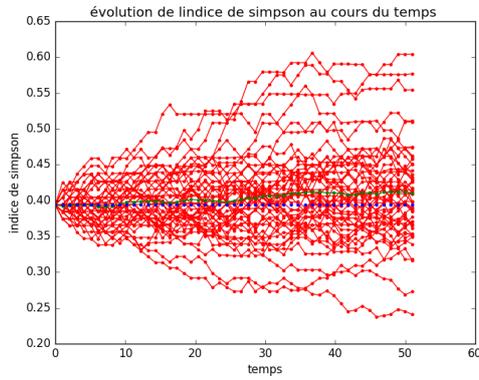


FIGURE 32 – $s=0$ avec une condition initiale hétérogène, 5 espèces représentées

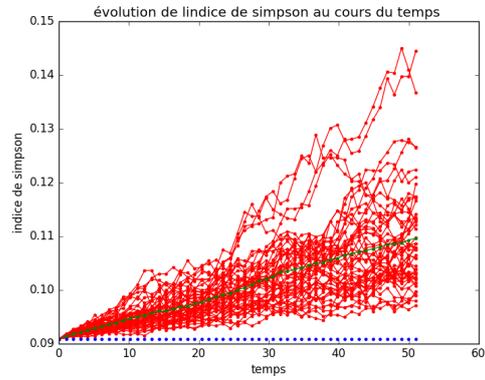


FIGURE 33 – $s=0$ avec une condition initiale homogène, 5 espèces représentées, $J=100$

-graphes pour 2 espèces :

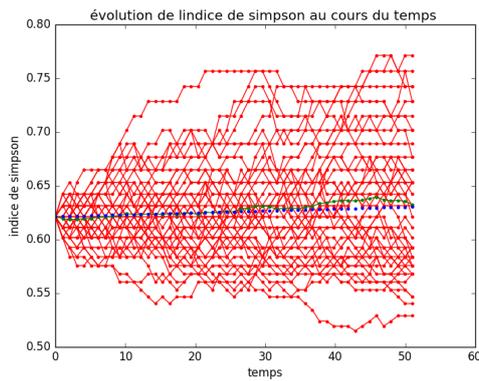


FIGURE 34 – $s=0$ avec une condition initiale hétérogène, 2 espèces représentées

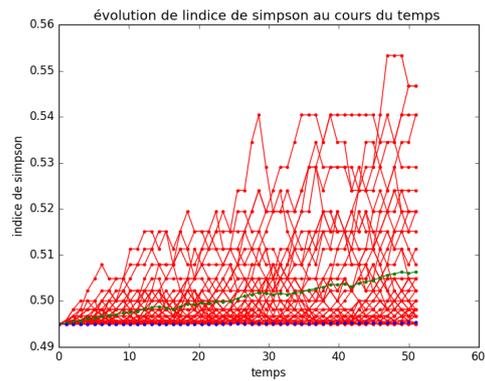


FIGURE 35 – $s=0$ avec une condition initiale homogène, 2 espèces représentées, $J=100$

Le nombre d'espèces présentes dans la communauté ne semble pas être un facteur déterminant au fait que les deux moyennes soient proches.

4.7.4 Influence des poids initiaux

On prend, par la suite, 5 espèces présentes à l'état initial.

Premier cas : on est proche de la situation $m=0, s=0$, condition initiale homogène. Le poids total de chaque espèce est presque le même à l'instant initial.

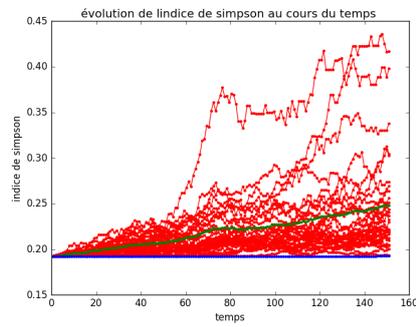


FIGURE 36 – ici $s=0.05, m=0, ci$: 5 espèces représentées par 20 individus

On rappelle que notre approche ne convient pas dans de pareils cas.

Regardons ce qui se passe si on perturbe un peu cette situation :

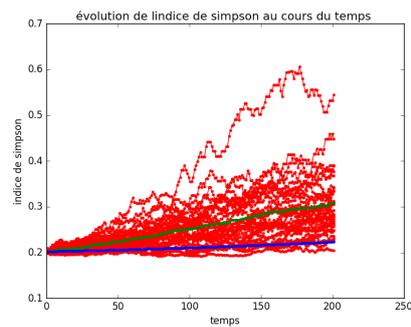


FIGURE 37 – ici $s=0.05, m=0, ci$: 2 espèces représentées par 25 individus, 2 par 15, 1 par 20

Maintenant augmentons le caractère hétérogène de la condition initiale :

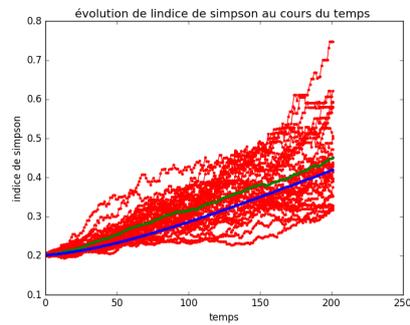


FIGURE 38 – ici $s=0.5, m=0, ci$: 2 espèces représentées par 25 individus, 2 par 15, 1 par 20

Prenons enfin une condition initiale vraiment hétérogène :

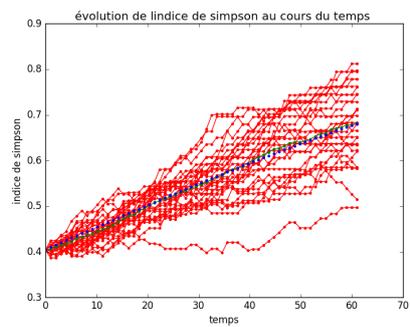


FIGURE 39 – ici $s=0.8, m=0, ci$: 2 espèces représentées par 5 individus, 2 par 15, 1 par 60

Ainsi le caractère d'homogénéité de la condition initiale est décisif pour que notre approche soit valable.

4.7.5 Lors d'un changement de fitness

A priori, lors d'un changement de fitness, on passe d'un état hétérogène à un autre. Alors notre approche devrait donner des résultats satisfaisants :

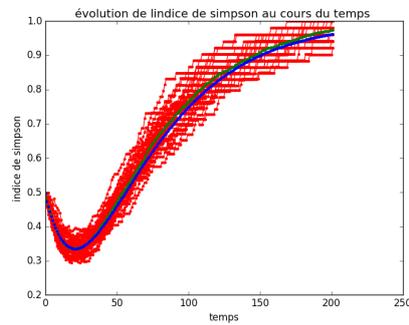


FIGURE 40 – changement de fitness, ici $s=1, m=0$

mais ce n'est pas toujours le cas...

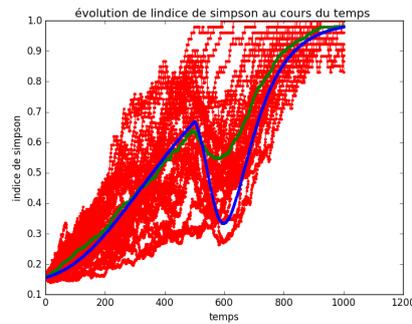


FIGURE 41 – changement de fitness, ici $s=1, m=0$

Cette approche n'est en effet pas toujours fiable. Pour comprendre, précisément pourquoi, dans certains cas, elle ne convient pas, il faudrait une étude plus approfondie.

5 Vers des modèles plus simplifiés

La dynamique du modèle précédent n'est pas évidente à étudier. Elle dépend de beaucoup de paramètres (s, m, θ, J , la période des changements de fitness etc)

On va alors simplifier ce modèle pour tenter de l'étudier. On peut par exemple conserver les fitness et enlever le pool.

5.1 Si $m=0$:

On suppose ici que m vaut 0. Comme il n'y a plus d'immigration, inévitablement, au bout d'un certain temps, une espèce va envahir la communauté.

On peut alors se demander quel est l'effet de changements fréquents de fitness sur la durée de vie des espèces ?

5.1.1 Présentation

On part ici d'une situation initiale représentée par 5 espèces de 20 individus chacune. On considère le même modèle que d'habitude avec $m=0$. Ainsi, au bout d'un certain temps une espèce va envahir toute la population. On va se focaliser sur le temps d'invasion. Est ce que des changements fréquents de fitness retardent l'invasion et permettent de conserver de la diversité plus longtemps ?

On va, pour cela, tirer au hasard selon un s donné, une fitness pour nos 5 espèces. Ensuite, on laissera d'un côté évoluer la population sans changement de fitness et de l'autre on réinitialisera les fitness tous les n pas de temps. (n dans 10,50,100) On trace ainsi 20 trajectoires (avec la même fitness initiale) et on fait une moyenne, point par point, de l'indice de Simpson sur ces trajectoires (courbe verte sur les graphes).

Les paramètres que nous ferons varier seront la valeur de s et le temps entre les changements de fitness.

Remarque : On ne fait pas ici une moyenne sur plusieurs tirages de fitness. De plus le nombre d'espèces pris en compte (5) étant relativement faible, l'échantillon de fitness tiré n'est pas forcément très représentatif de la valeur de s !

5.1.2 Les résultats

On représente à chaque fois pour 4 valeurs de s , quatre graphiques. Sur chaque graphique, une courbe avec changement de fitness, une autre sans fitness sont représentés.

Changements effectués tous les 10 pas de temps pour différentes valeurs de s :

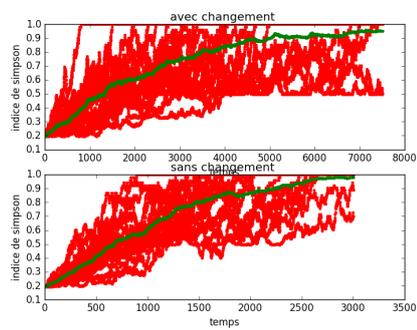


FIGURE 42 – $s=0.1$

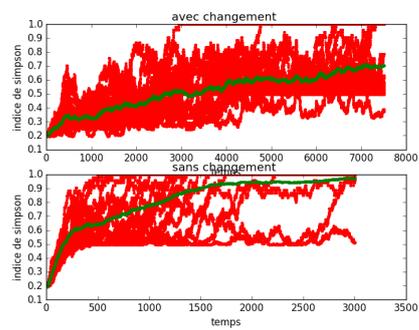


FIGURE 43 – $s=0.5$

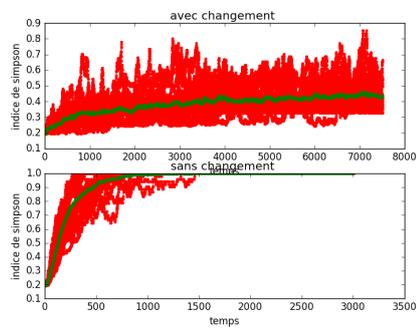


FIGURE 44 – $s=1$

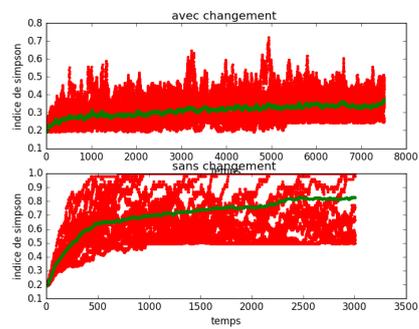


FIGURE 45 – $s=2$

Changements effectués tous les 50 pas de temps pour différentes valeurs de s :

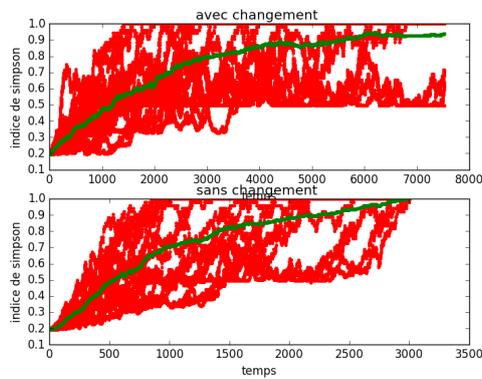


FIGURE 46 – $s=0.1$

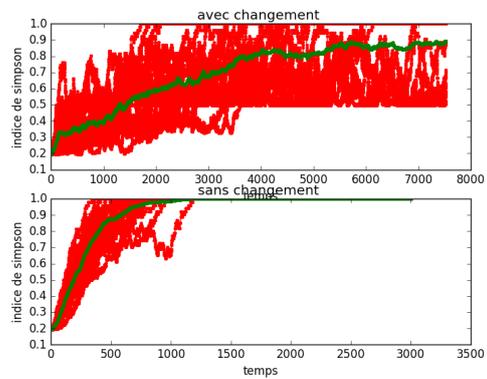


FIGURE 47 – $s=0.5$

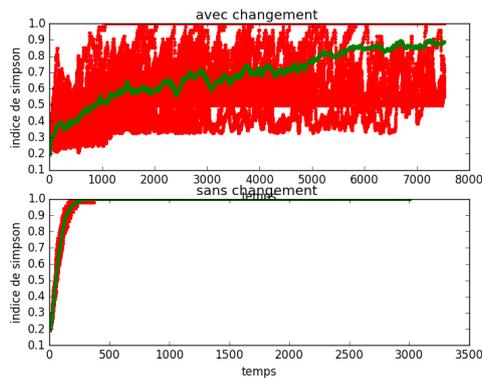


FIGURE 48 – $s=1$

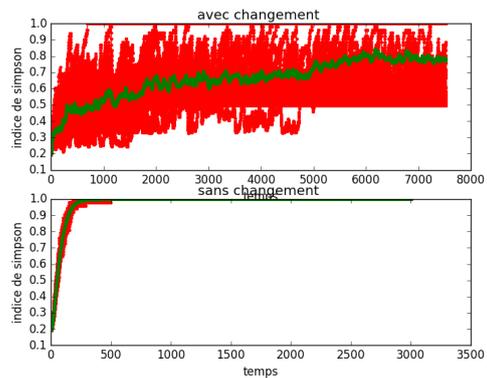


FIGURE 49 – $s=2$

Changements effectués tous les 100 pas de temps pour différentes valeurs de s :

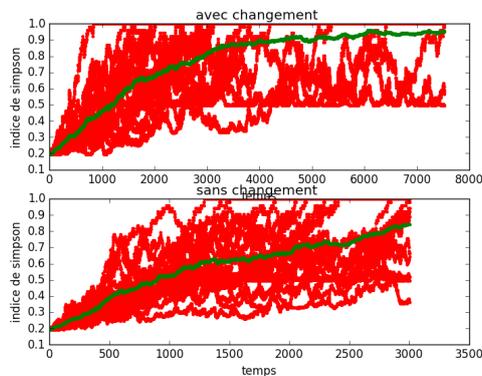


FIGURE 50 – $s=0.1$

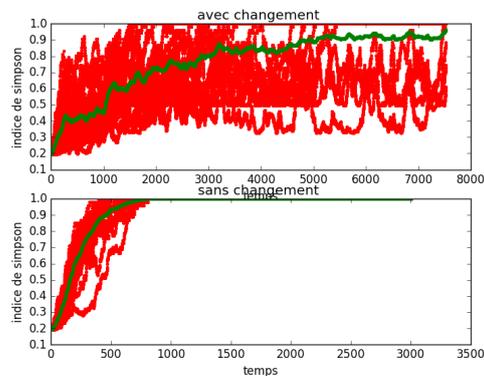


FIGURE 51 – $s=0.5$

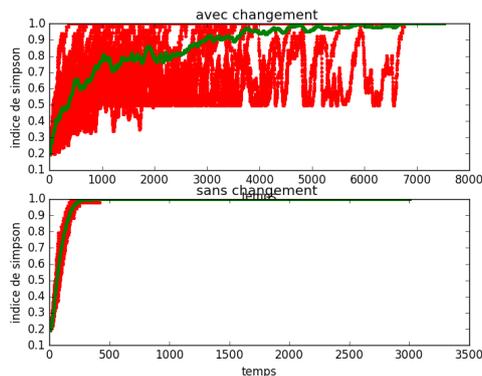


FIGURE 52 – $s=1$

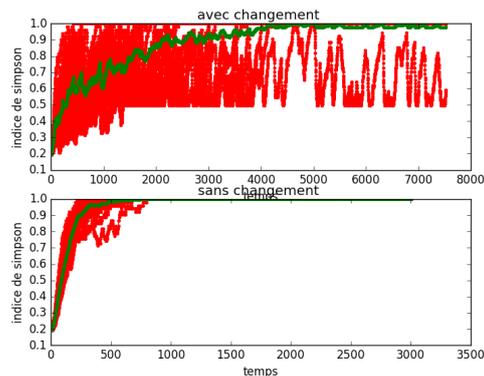


FIGURE 53 – $s=2$

5.1.3 Analyse

Au vu des résultats, des changements de fitness rallongent le temps nécessaire à l'invasion d'une espèce. Il semble, par ailleurs, que plus ces changements sont fréquents, plus le temps d'invasion est retardé. En effet, dans ce cas, aucune espèce n'est défavorisée assez longtemps pour s'éteindre complètement et on conserve ainsi une diversité d'espèces plus grande plus longtemps. En revanche, si les changements sont assez peu fréquents, certaines espèces vont s'éteindre entre ces changements et on aura une invasion d'autant plus rapide que s est

grand.

Quand s est petit, tous les individus ont un poids similaire et le nombre d'individus de chaque espèce est représentatif du poids de l'espèce. Ainsi, il est logique que les changements de fitness ne changent pas énormément la dynamique de la population. En revanche, si s est grand, l'espèce en plus grand nombre ne sera pas forcément celle qui aura le poids le plus grand et celle-ci sera choisie très probablement pour mourir (alors que celle qui aura le poids le plus grand sera choisie pour remplacer le mort). Ceci se traduit par des sauts dans l'indice de Simpson (cf courbes rouges pour s grand).

5.2 Tentative de calcul d'une abondance par coalescent en l'absence de changement de fitness

Ici, on suppose, qu'il n'y a plus de changement de fitness, les poids sont fixés pour chaque espèce et n'évoluent pas dans le temps. On cherche alors à calculer la probabilité d'une abondance particulière à l'équilibre. Pour cela on va avoir, comme précédemment, besoin de reconstruire une généalogie.

5.2.1 La généalogie

Supposons une population de taille J , un échantillon E de taille k connu (k pouvant être égal à J).

Chaque individu i de l'échantillon a un poids P_i qui est le même pour tous les individus de la même espèce.

Le pool est supposé à l'équilibre, les proportions de chaque espèce sont connues.

-Probabilités du dernier événement ayant affecté l'échantillon

Demandons nous maintenant quelle est la probabilité que le dernier événement arrivé ait affecté notre échantillon. Seulement deux types d'événements ont pu affecter notre échantillon, il s'agit d'une filiation, c'est à dire la coalescence de deux individus de l'échantillon ou alors d'une spéciation, c'est à dire que le dernier individu arrivé dans E provient d'un père du pool.

La probabilité qu'il y ait eu spéciation est :

$$m \times \frac{k}{J}$$

, en effet k/J correspond au choix du dernier arrivé dans E et m est la probabilité qu'il vienne du pool.

La probabilité qu'il y ait eu filiation est :

$$(1 - m) \times \frac{k}{J} \times \sum_{i=1 \dots k} \frac{1}{k} \times \frac{\sum_{j \in E \neq i} P_j}{P_{tot} - P_i}$$

où P_{tot} est le poids de toute la communauté.

Le problème qui vient à nous maintenant est que l'on ne peut directement obtenir la probabilité du dernier événement affectant l'échantillon sans tenir compte des événements intermédiaires car celle ci dépend de P_{tot} qui varie à chaque événement.

En revanche, si pour des raisons pratiques, on suppose que P_{tot} reste constant au cours du temps (ce qui n'est pas totalement faux si s est petit car les poids sont à peu près les mêmes pour toutes les espèces), on peut obtenir ces probabilités. On supposera que tous les poids sont proches de 1 de sorte que $P_{tot} = J$.

Alors, la probabilité que le dernier événement qui ait affecté E soit une filiation est :

$$\frac{(1-m) \times \frac{k}{J} \times \sum_{i \in 1 \dots k} \frac{1}{k} \times \frac{\sum_{j \in E \neq j} P_j}{P_{tot} - P_i}}{(1-m) \times \frac{k}{J} \times \sum_{i \in 1 \dots k} \frac{1}{k} \times \frac{\sum_{j \in E \neq j} P_j}{P_{tot} - P_i} + m \times \frac{k}{J}} = \frac{\frac{k-1}{k} \sum_{j \in E} P_j}{\frac{m(J-1)}{1-m} + \frac{k-1}{k} \sum_{j \in E} P_j} = \frac{\frac{k-1}{k} \sum_{j \in E} P_j}{\theta + \frac{k-1}{k} \sum_{j \in E} P_j}$$

Et la probabilité que le dernier événement qui ait affecté E soit une spéciation est :

$$\frac{\theta}{\theta + \frac{k-1}{k} \sum_{j \in E} P_j}$$

Remarques :

- On notera que si les poids sont constants, égaux à 1, alors on retrouve les probabilités du modèle sans fitness.
- En supposant que le poids total soit constant, on suppose le poids du reste de la communauté connu. C'est en effet cette information qui est décisive ici : si le poids de notre échantillon est plus faible que celui du reste de la communauté alors, à chaque fois qu'il y aura filiation avec plus grande probabilité, elle touchera le reste de la communauté et n'affectera pas E. Mais les fitness n'intervenant pas dans la spéciation, l'échantillon E aura alors plus de chance d'être soumis à une spéciation qu'à une filiation. (et inversement)
- Remarquons que sur ce modèle le nombre d'étapes nécessaires pour que E soit affecté par une spéciation reste constant. En revanche le temps pour qu'il soit affecté par une filiation varie en fonction des poids de l'échantillon par rapport au reste de la communauté.

-Un exemple d'arbre :

Voici un exemple de généalogie. On part d'une population (ici 11122) où les individus sont assimilés à leurs poids et on regarde leur généalogie :

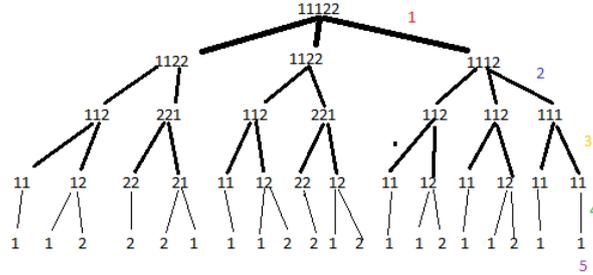


FIGURE 54 – Est représenté l'ensemble des événements qui ont pu conduire à la population 11122. Chaque événement arrive avec une certaine probabilité. Quelques exemples sont donnés ensuite

Valeurs correspondant aux probabilités des branches sur le dessin :

$$1 : \frac{\frac{4}{5} \times 7}{\theta + \frac{4}{5} \times 7} \times \frac{2}{5}, 2 : \frac{\theta_2}{\theta + \frac{3}{4} \times 5}, 3 : \frac{2}{\theta + 2} \times \frac{1}{2}, 4 : \frac{1}{\theta + 1} \times 1, 5 : \frac{\theta_1}{\theta}$$

5.2.2 Quand les poids sont très proches de 1

-Sur une branche de l'arbre

On se propose de regarder ici ce qu'il se passe quand on fait tendre les poids vers 1.

Posons $P_i = 1 + \epsilon \delta_i$ où les δ_i sont bornés dans leur ensemble. On va faire tendre ϵ vers 0.

Alors,

$$\frac{1}{\theta + \frac{k-1}{k} \sum_{j \in E} P_j} = \frac{1}{\theta + k - 1 + \frac{k-1}{k} \sum_{j \in E} \delta_j \epsilon} = \frac{1}{\theta + k - 1 + \Delta_k \epsilon}$$

Et,

$$\frac{1}{\theta + k - 1 + \Delta_k \epsilon} \stackrel{\epsilon \rightarrow 0}{=} \frac{1}{\theta + k - 1} \times \left(1 - \frac{\Delta_k \epsilon}{\theta + k - 1} + o(\epsilon)\right)$$

Alors la proba de spéciation devient :

$$\frac{\theta}{\theta + k - 1} - \frac{\theta \Delta_k \epsilon}{(\theta + k - 1)^2} + o(\epsilon)$$

et la proba de filiation :

$$\frac{k-1}{\theta+k-1} + \epsilon \Delta_k \times \left(\frac{1}{\theta+k-1} - \frac{k-1}{(\theta+k-1)^2} \right) + o(\epsilon)$$

Remarques :

- On retrouve bien au premier ordre les probabilités dans le cas sans fitness.
- En fonction du signe de Δ_k , la filiation ou la spéciation va être favorisée : si Δ_k est positif c'est à dire qu'en moyenne les poids des individus de l'échantillon sont supérieurs à ceux du cas neutre (qui correspondent à 1), alors par rapport au modèle neutre, une filiation aura plus de chance d'être le dernier événement à toucher l'échantillon. Et en accord avec la remarque précédente, la filiation sera favorisée par rapport à la spéciation.

-Sur une abondance

Fixons nous une population et supposons connu l'ordre d'arrivée des individus dans la communauté. On l'assimilera à un vecteur où les individus sont rangés par ordre d'arrivé dans la communauté du plus ancien au plus jeune., et notons N_s le nombre d'espèces présentes dans cette population.

Donnons nous $\sigma_1, \dots, \sigma_{N_s}$ les étapes de l'arbre auxquelles arrivent ces espèces. On fera l'hypothèse qu'à chaque fois qu'une espèce vient du pool, elle est de nature différente de celles présentes dans la communauté.

Enfin, à chaque fois qu'une filiation se produit il nous faut déterminer avec quelle probabilité un père donné est choisi.

S'il est choisi à l'étape k, il le sera avec la probabilité :

$$\begin{aligned} \frac{1 + \delta_k \epsilon}{\sum_{j=1}^{k-1} (1 + \delta_j \epsilon)} &= \frac{1 + \delta_k \epsilon}{k-1} \times \left(1 - \epsilon \sum_{j=1}^{k-1} \frac{\delta_j}{k-1} \right) + o(\epsilon) \\ &= \frac{1}{k-1} \times \left(1 - \epsilon \left(\sum_{j=1}^{k-1} \frac{\delta_j}{k-1} - \delta_k \right) \right) + o(\epsilon) = \frac{1}{k-1} \times (1 - \epsilon(\Delta'_k)) + o(\epsilon) \end{aligned}$$

Alors à l'étape k la probabilité qu'il y ait filiation et que le père fixé soit choisi est :

$$\begin{aligned} \left[\frac{k-1}{\theta+k-1} + \epsilon \Delta_k \times \left(\frac{1}{\theta+k-1} - \frac{k-1}{(\theta+k-1)^2} \right) \right] \times \frac{1}{k-1} \times (1 - \epsilon \Delta'_k) + o(\epsilon) \\ = \frac{1}{\theta+k-1} [1 + \epsilon C_k] + o(\epsilon) \quad \text{où} \quad C_k = \Delta_k \times \left(\frac{1}{k-1} - \frac{1}{\theta+k-1} \right) - \Delta'_k \end{aligned}$$

De même, la probabilité que l'espèce i arrive du pool à l'étape k est :

$$\left[\frac{\theta}{\theta+k-1} - \frac{\theta \Delta_k \epsilon}{(\theta+k-1)^2} \right] \times \frac{\theta_i}{\theta} + o(\epsilon) = \frac{\theta_i}{\theta+k-1} \times [1 + \epsilon D_k] + o(\epsilon)$$

où $D_k = -\frac{\Delta_k}{\theta+k-1}$ et $\frac{\theta_i}{\theta}$ la proportion de l'espèce i dans le pool.

Maintenant que l'on a les probabilités de chaque étape, il ne reste plus qu'à les multiplier pour obtenir la probabilité du vecteur de population, c'est à dire la probabilité que cette généalogie soit réalisée.

On obtient alors :

$$\begin{aligned} & \prod_{i=1}^{N_s} \theta_{\sigma_i} \times \prod_{j=1}^J \frac{1}{\theta + j - 1} \times \prod_{l=1}^{N_s} [1 + \epsilon D_{\sigma_l}] \times \prod_{k \leq J, k \neq \sigma_1 \dots \sigma_{N_s}} [1 + \epsilon C_k] + o(\epsilon) \\ &= \prod_{i=1}^{N_s} \theta_i \times \prod_{j=1}^J \frac{1}{\theta + j - 1} \times [1 + \epsilon (\sum_{i=1}^{N_s} D_{\sigma_i} + \sum_{k \leq J, k \neq \sigma_1 \dots \sigma_{N_s}} C_k)] + o(\epsilon) \end{aligned}$$

Remarque : le terme de premier ordre correspond à la probabilité d'obtenir une transposition précise dans le cas neutre.

Il reste alors à sommer sur tous les vecteurs conduisant à l'abondance désirée... On doit cependant tenir compte du fait que D_{σ} et C_k dépendent à chaque fois de l'ordre dans lequel les ancêtre apparaissent, autrement dit toutes les généalogies n'arrivent pas avec la même probabilité contrairement au cas neutre.

$$\begin{aligned} &= \prod_{i=1}^{N_s} \theta_i \times \prod_{j=1}^J \frac{1}{\theta + j - 1} \times \sum_{\text{toutes les branches de l'arbre}} [1 + \epsilon (\sum_{i=1}^{N_s} D_{\sigma_i} + \sum_{k \leq J, k \neq \sigma_1 \dots \sigma_{N_s}} C_k)] + o(\epsilon) \\ &= \prod_{i=1}^{N_s} \theta_i \times \prod_{j=1}^J \frac{1}{\theta + j - 1} \times (\text{nombre de branches}) + \epsilon \sum_{\text{branches}} [\sum_{i=1}^{N_s} D_{\sigma_i} + \sum_{k \leq J, k \neq \sigma_1 \dots \sigma_{N_s}} C_k] + o(\epsilon) \\ &= \prod_{i=1}^{N_s} \theta_i \times \prod_{j=1}^J \frac{1}{\theta + j - 1} \times NB + \epsilon \sum_{\text{branches}} [-\sum_{k=1}^J \frac{\Delta_k}{\theta + k - 1} + \sum_{l \leq J, l \neq \sigma_1 \dots \sigma_{N_s}} (\frac{\Delta_l}{l - 1} - \Delta'_l)] \end{aligned}$$

Remarque : le premier terme de cette somme est la probabilité d'obtenir l'abondance voulue dans le cas sans fitness. En effet le nombre de branches correspond au nombre de transpositions ayant N_s cycles de taille le nombre d'individus de chaque espèce.

6 Conclusion

Nous sommes partis du modèle suggéré par Hubbell en 2001 et nous l'avons complexifié pour essayer d'obtenir un modèle plus réaliste. Le fait de rajouter les fitness complique davantage le problème : les probabilités de transition pour passer d'un état à l'autre sont différentes pour chaque espèce et dépendent du tirage des fitness...

Pour commencer à se familiariser avec ce modèle on a essentiellement fait des simulations, ces dernières nous permettent de comprendre le rôle de chaque paramètre du modèle.

On a, par ailleurs, essayé de percevoir le comportement en moyenne de notre modèle via l'approche avec espérance. Si cette méthode semble convenir dans certains cas, dans d'autres elle n'est pas fiable. Une étude plus poussée est nécessaire pour bien comprendre pourquoi cette approche ne marche pas toujours.

Enfin pour pouvoir faire une étude mathématique, on a tenté de simplifier le problème. On a dans un premier temps supposé qu'il n'y avait pas de changement de fitness, le but étant de reconstruire une généalogie. Mais ceci n'est pas si simple, et les résultats peu probants pour l'instant. Une deuxième approche consiste à poser $m=0$, et à se demander si des changements fréquents de fitness modifient les temps de vie des espèces. On constate numériquement que oui, mais reste maintenant à quantifier cela en fonction des paramètres. Qu'en est il si on rajoute le paramètre m ?

On aimerait aussi estimer le temps d'invasion d'une espèce en fonction des paramètres, ceci nous permettrait d'avoir une idée sur la fréquence des changements de fitness nécessaire pour conserver de la biodiversité.

7 Bibliographie

- Hubbell, S. P. The Unified Neutral Theory of Biodiversity and Biogeography (Princeton Univ. Press, Princeton, NJ, 2001)
- Neutral theory and the species abundance distribution : recent developments and prospects for unifying niche and neutral perspectives Thomas J. Matthews and Robert J. Whittaker (Ecology and evolution, 2014)
- Etienne, R. S., and H. Olf. 2004. A novel genealogical approach to neutral biodiversity theory. *Ecol. Lett.* 7 :170–175.
- Etienne, R. S. 2005. A new sampling formula for neutral biodiversity. *Ecol. Lett.* 8 :253–260
- Ricklefs, R. E. 2003. A comment on Hubbell’s zero-sum ecological drift model. *Oikos* 100 :185–192.
- Ricklefs, R. E. 2006. The unified neutral theory of biodiversity : do the numbers add up? *Ecology* 87 :1424–1431.
- Michael Kalyuzhny, Ronen Kadmon and Nadav M. Shnerb, neutral theory with environmental stochasticity explains static and dynamic properties of ecological communities (Ecology Letters, 2015)
- Franck Jabot and Julien Pottier, A general modelling framework for resource-ratio and CSR theories of plant community dynamics, (Journal of Ecology 2012,)
- Rick Durrett, Essentials of Probability, Dec 1998

8 Annexes

Tous les algorithmes ont été écrits avec le langage python et exécutés via l’interface pyzo.

8.1 algorithme

Cet algorithme permet de modéliser l’évolution de la communauté, de l’indice de Simpson et du poids total au cours du temps.

```
import numpy as np
import pylab as plt
import numpy.random as rd
import scipy.integrate as int
import scipy.linalg as linalg
import matplotlib.pyplot as plt

plt.close('all')

def tirvectdens(dpop) :
    r=rd.rand() p=0 j=0 while p<r :
    p=p+dpop[j] j=j+1 return j-1

def transformevectenpop(vect) :
    j=1 compt=0 pop=np.zeros((J)) for i in vect :
    pop[compt :compt+i]=j compt=compt+i j=j+1
```

```

return pop

    m=0.1
    J=100 A=1 s=2
    S=200 (nb espèces) O=60 T=40
    T=2 t=2 Nb=1
Nb1=1000
Nb2=2000

    Rpoids1=np.zeros((S))
for j in range(S) :
Rpoids1[j]=rd.lognormal(1,s)

    max1=np.max(Rpoids1)

    Rpoids2=np.zeros((S))
for j in range(S) :
Rpoids2[j]=rd.lognormal(1,s)
    max2=np.max(Rpoids2)

    Création du pool et de sa proportion d'individus
    def chinois(O,J) :
j=0 r=0 R=np.array([])

    while j<J :
j=j+1
    a=rd.random() if a<O/(O+r) :(nouvelle espèce du pool) R=np.append(R,1)

    else :(espèce déjà présente dans le pool)
dR=R/r b=tirvectdens(dR) (donne le numéro de l'espèce tirée) R[b]=R[b]+1
r=r+1
return R

    R=np.array(chinois(O,10*J))

    print(np.shape(R))

    Rtot=0
for i in R : Rtot=Rtot+i
    R=R/Rtot (C'est un vecteur probabilisé)

    if np.shape(R)[0]>S : print('ERROR')

    Res=np.zeros((S)) a=np.shape(R)[0] Res[:a]=R
    def compteindiv(P,R) :
X=np.zeros((np.shape(R)[0]))
for i in range(np.shape(P)[0]) : X[P[i]-1]=X[P[i]-1]+1 return X

    def evotempsimpcomptepoids(Res,Rpoids,pop,T,m) :

```

```

    pop1=np.copy(pop)
    J=np.shape(pop)[0]
    jj=np.zeros((J))
for i in range(J) :
    jj[i]=Rpoids[pop1[i]-1]
    poidtot=0
for i in range(J) :
    poidtot=poidtot+jj[i]
    t=0

    while t<T :

        t=t+1

        mort=rd.randint(0,J)

        poidtot=poidtot-jj[mort]

        jj[mort]=0 a=rd.random()
        if a<m :

            born=tirvectdens(jj/poidtot)
            jj[mort]=jj[born] pop1[mort]=pop1[born] poidtot=poidtot+jj[born]
        else :

            born=tirvectdens(Res) jj[mort]=Rpoids[born] pop1[mort]=born+1 poidtot=poidtot+Rpoids[born]

    return pop1,poidtot
def evopoidsmoyenentemps(Res,Rpoids1,Rpoids2,pop,t,m,Nb1,Nb2,T) :

    Pmoy=0 MMoy=0
for l in range(Nb) : print(l)
    pop1,poidtot=evotempsimpcomptepoids(Res,Rpoids1,pop,T,m)

    Nesp=[] SimpsT10=[] poiT10=[]

    for i in range(Nb1) :

        pop1,poidtot=evotempsimpcomptepoids(Res,Rpoids1,pop1,t,m)

        z=compteindiv(pop1,Res) Po=0 nesp=0
for i in z :

        Po=Po+(i*(i-1))/(J*(J-1)) if i<0 :
        nesp=nesp+1 Nesp.append(nesp)
        SimpsT10.append(Po)

        poiT10.append(poidtot)

```

```

for i in range(Nb2) : on récupère les valeurs de l'indice de simpson tous les
10 pas de temps sans qu'il y ait la période sans renouvellement de fitness
pop1,poidtot=evotempsimpcomptepoids(Res,Rpoids2,pop1,t,m)

z=compteindiv(pop1,Res)
nesp=0 Po=0
for i in z :
Po=Po+(i*(i-1))/(J*(J-1)) if i<0 :
nesp=nesp+1 Nesp.append(nesp)
SimpsT10.append(Po) poiT10.append(poidtot)

Pmoy=Pmoy+np.array(poiT10) MMoy=MMoy+np.array(SimpsT10)

MMoy=MMoy/Nb Pmoy=Pmoy/Nb
return MMoy,Pmoy,Nesp

Max=[]
for i in range(Nb1) :
Max.append(max1*J)
for i in range(Nb2) :
Max.append(max2*J)

popo=np.zeros(J)
popo[0]=20 popo[1]=20 popo[2]=20 popo[3]=20 popo[4]=20 popo[5]=0 popo[6]=0
popo[7]=0 popo[8]=0 popo[9]=0 popo[10]=0
pop=transformevectenpop(popo)

plt.figure(1)
plt.clf()

for s in [0.05,0.1,0.5,0.8,1,2] :
Rpoids1=np.zeros((S))
for j in range(S) :
Rpoids1[j]=rd.lognormal(1,s)
Rpoids2=np.zeros((S))
for j in range(S) :
Rpoids2[j]=rd.lognormal(1,s)

Y,X,Z=evopoidsmoyenentemps(Res,Rpoids1,Rpoids2,pop,t,m,Nb1,Nb2,T)
a=np.shape(X)[0]
x=np.linspace(0,a,a)
plt.plot(x,Y)
plt.legend(('0.05','0.1','0.5','0.8','1','2'))
plt.title('indice de Simpson pour différents s')
plt.xlabel('temps')
plt.ylabel('indice de simpson')
plt.show()
Y,X,Z=evopoidsmoyenentemps(Res,Rpoids1,Rpoids2,pop,t,m,Nb1,Nb2,T)

```

```

a=np.shape(X)[0]
x=np.linspace(0,a,a)

plt.close('all')
plt.figure(2)
plt.clf()
plt.subplot(3,1,2)

plt.plot(x,Max)
plt.plot(x,X,'-r.')
plt.title('poids total indice de simpson ')
plt.xlabel('temps')
plt.ylabel('poids total')

plt.subplot(3,1,1)

plt.plot(x,Y,'-b.')
plt.title('indice de Simpson poids total et nombre d espèces ')
plt.xlabel('temps')
plt.ylabel('indice de simpson')

plt.subplot(3,1,3)

plt.plot(x,Z,'-g.')
plt.title('nb d espèces ')
plt.xlabel('temps')
plt.ylabel('nombre d espèces')
plt.show()

```