
Evolution of communities under selection

Rapport de stage de M1

Gaspard Dousson Lys,
encadré par Cornelia Pokalyuk

Table des matières

Déroulement du stage et introduction mathématique	2
1 Modèle mathématique et simulation numérique d'une communauté écologique	3
2 Distribution des abondances dans une large communauté écologique	5
3 Évolution des interactions écologiques pour une pression sélective collective	12
Références	16

Remerciements

Je tiens tout d'abord à remercier Cornelia Pokalyuk pour son encadrement, son soutien, son aide et tout le temps que cela représente. Plus généralement, je remercie l'ensemble des membres de l'institut de mathématiques de Lübeck pour leur accueil.

Je remercie également Silvia De Monte pour ses réponses à nos nombreuses questions sur le modèle étudié au cours de ce stage. Enfin, je remercie Amaury Lambert pour ses précieux conseils pour ma formation, et pour m'avoir mis en contact avec Cornelia Pokalyuk en premier lieu.

Notations

Étant donnée une grandeur N , on note \dot{N} sa dérivée temporelle $\frac{dN}{dt}$.

Étant donnée une variable aléatoire X , on note $E[X]$ son espérance et $V[X]$ sa variance.

Étant données deux variables aléatoires X et Y , on note $\langle X|Y \rangle$ leur covariance.

Étant donné un réel x , on note $(x)_+$ sa partie positive $\max(0, x)$.

Déroulement du stage

Mon stage s'est déroulé à l'institut de mathématiques de l'université de Lübeck sous la direction de Cornelia Pokalyuk. Il portait sur l'évolution de communautés écologiques soumises à une pression sélective à l'échelle collective. En se basant sur un modèle étudié dans [1], simulé dans [2] et présenté ci-après, la majeure partie du travail consistait à conjecturer et démontrer certaines propriétés, notamment asymptotiques en le nombre d'espèces considérées.

Dans ce contexte, les premières semaines ont été dédiées à la programmation du modèle afin de le simuler, d'y mesurer diverses statistiques et d'établir des conjectures. La suite du stage a consisté en un travail mathématique de documentation et de démonstration assisté par les simulations initiales. Parallèlement, de nombreuses conférences m'ont permis de découvrir divers champs de recherche en mathématiques appliquées à la biologie. En particulier, certains séminaires du Max Planck Institute de Plön, spécialisé en biologie évolutive, ont été déterminants dans mes choix d'orientation, en plus d'avoir été particulièrement enrichissants.

Introduction mathématique

L'écologie est une discipline de la biologie s'intéressant non pas aux êtres vivants directement mais à leurs interactions. Une communauté écologique est naturellement définie comme un ensemble d'espèces en interaction. Une bonne modélisation est essentielle pour bien comprendre le rôle et l'importance de chaque partie d'une telle communauté. Pour autant, celle-ci dépend largement de notre connaissance préalable des constituants d'un écosystème, alors même qu'il peut s'avérer fondamental de décrire le plus précisément possible des interactions dont les acteurs sont parfois méconnus.

Dans ce cadre, on représente les interactions écologiques parmi S espèces via une matrice α de taille $S \times S$. Le nombre élevé de paramètres d'un tel modèle est réduit en choisissant cette matrice à l'aide d'une loi normale paramétrée par une espérance μ_α , un écart-type σ_α et un coefficient de corrélation γ_α . On cherche alors à décrire la distribution des abondances de chaque espèce en fonction de ces quelques paramètres – par exemple en vue de produire des outils statistiques –. On montre qu'une telle distribution est composée d'une gaussienne tronquée en 0 et d'un atome isolé représentant le taux d'extinction.

On s'intéresse également à l'évolution des interactions écologiques lorsqu'elles sont soumises à des mutations aléatoires η d'ordre de grandeur $O(\epsilon)$ petit. On montre d'abord que le choix précis de la loi de probabilité des mutations n'a pas d'incidence sur la trajectoire évolutive. À l'aide de simulations numériques, on constate ensuite qu'une structure non-aléatoire est imprimée par la sélection à l'échelle collective dans les valeurs propres de la matrice α .

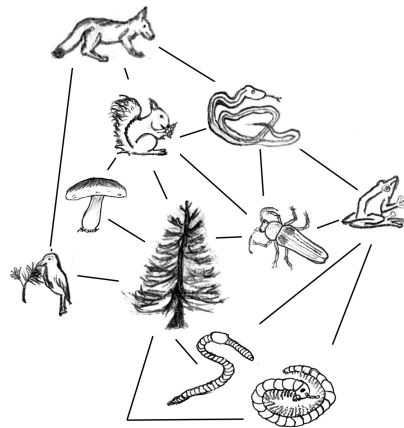


FIGURE 1 – Le dessin *Trophic web* de Mark David Thompson. L'écologie correspond à l'étude des arêtes du graphe, et non de ses nœuds.

1 Modèle mathématique et simulation numérique d'une communauté écologique

Une modélisation mathématique fréquemment utilisée pour les communautés écologiques est le système de Lotka-Volterra. Cette conjonction d'équations différentielles combine les interactions intraspécifiques et interspécifiques pour décrire la dynamique démographique des différentes espèces d'une même communauté.

Dans la sous-section 1.1, on construit progressivement un tel modèle de communauté écologique. On réduit son nombre élevé de paramètres via l'introduction de variables aléatoires. Dans la sous-section 1.2, on présente une méthode d'intégration spécifiquement conçue pour simuler ce modèle. Les résultats numériques qui en découlent permettent de mieux comprendre la distribution des abondances, étudiée formellement dans la section 2.

1.1 Le modèle de communauté écologique

Un champ de la biologie particulièrement prolifique en modèles mathématiques est la dynamique des populations. Il s'agit essentiellement de décrire l'évolution démographique d'une ou plusieurs populations. Le modèle le plus simple pour étudier l'*abondance* N d'une espèce isolée, caractérisée par un *taux de croissance* r , est l'équation de Malthus

$$\dot{N} = rN. \quad (1)$$

Cependant, la croissance exponentielle prédite par l'équation (1) est inéluctablement limitée par les ressources de l'environnement et la compétition intraspécifique pour celles-ci. Le modèle de Verhulst prend en compte ce phénomène en définissant une *capacité porteuse* K , correspondant à l'abondance de l'espèce à l'équilibre, hors extinction. L'équation correspondante, dite logistique, s'écrit

$$\dot{N} = r \frac{N}{K} (K - N). \quad (2)$$

Si on s'intéresse maintenant à une communauté écologique composée de S espèces, un simple système d'équations (2) découplées n'est pas satisfaisant. Il convient de prendre en compte les interactions interspécifiques, décrites via une *matrice d'interaction* α de diagonale nulle. Le modèle est alors constitué des équations de Lotka-Volterra généralisées¹

$$\dot{N}_i = r_i \frac{N_i}{K_i} \left(K_i - N_i + \sum_{j \neq i} \alpha_{i,j} N_j \right). \quad (3)$$

On dira qu'une communauté écologique donnée *admet un équilibre* si, lorsque $t \rightarrow \infty$, les abondances (N_j) vérifiant le système (3) convergent vers un même équilibre stable fini quelle que soit la condi-

1. Les équations de Lotka-Volterra à proprement parler décrivent une communauté de deux espèces dans deux cas : la compétition (interactions négatives) et la prédation (interactions corrélées négativement). Lorsque plus de deux espèces interagissent, les équations sont donc dites « généralisées ».

tion initiale, sous réserve que $N_i(0) > 0$ pour toute espèce i . Dès lors, cet équilibre vérifie les équations

$$N_i \left(K_i - N_i + \sum_{j \neq i} \alpha_{i,j} N_j \right) = 0, \quad (4)$$

où les taux de croissance (r_i) – strictement positifs par définition – ont été simplifiés. Par la suite, on ne s'intéressera plus qu'à des communautés admettant un équilibre, pour lesquels il est donc inutile de préciser ces taux.

Malgré des équations peu sophistiquées, le nombre de paramètres augmente quadratiquement en le nombre d'espèces considérées. Pour de grandes communautés dont les espèces et les interactions sont souvent méconnues, le modèle s'avère fatalement sous-déterminé. Dans ce cadre, on décide de réduire le nombre de paramètres en les fixant à l'aide de lois de probabilités contrôlées par une poignée de valeurs seulement.

Les capacités porteuses (K_i) sont indépendantes et identiquement distribuées suivant une loi normale $\mathcal{N}(\mu_K, \sigma_K)$ de moyenne μ_K et d'écart-type σ_K . De même, les coefficients non-diagonaux de la matrice d'interaction α suivent la loi $\mathcal{N}(\mu_\alpha, \sigma_\alpha)$. Ils sont indépendants, à l'exception des interactions réciproques $\alpha_{i,j}$ et $\alpha_{j,i}$ de coefficient de corrélation γ_α indépendant de i et j . On réduit ainsi le modèle à cinq paramètres.

On dira que les paramètres $(\mu_K, \sigma_K, \mu_\alpha, \sigma_\alpha, \gamma_\alpha)$ *admettent un équilibre* si la probabilité qu'une communauté aléatoire de S espèces admette un équilibre tend vers 1 quand $S \rightarrow \infty$. Par la suite, on fera toujours l'hypothèse que les paramètres choisis admettent un équilibre. Le diagramme de phase associé est donné dans [1]. À titre d'exemple, il est nécessaire de choisir $\mu_\alpha < 1$, sans quoi il est toujours probable que deux espèces adoptent une relation mutualiste compensant leurs compétitions intraspécifiques respectives, permettant alors à leurs abondances de diverger.

On s'intéresse à la distribution de l'abondance d'une espèce à l'équilibre en fonction des paramètres $(\mu_K, \sigma_K, \mu_\alpha, \sigma_\alpha, \gamma_\alpha)$, et en particulier dans le cas d'une communauté large ($S \rightarrow \infty$). Pour commencer, on présente une méthode d'intégration utilisée pour simuler le modèle décrit jusqu'à présent.

1.2 La méthode d'intégration des équations de Lotka-Volterra généralisées

Pour simuler le modèle décrit précédemment, on utilise une méthode d'intégration spécifique aux équations de Lotka-Volterra généralisées. En effet, la méthode d'Euler² peut produire des abondances négatives. Dans notre cas, on préfère considérer que, pour chaque espèce i , les abondances $(N_j)_{j \neq i}$ sont constantes sur l'intervalle de temps $[t, t + \tau[$, de sorte à découpler localement l'équation associée à l'espèce i . On peut alors la résoudre et en déduire une valeur approchée de $N_i(t + \tau)$.

Étant donné les taux de croissance (r_i) et les capacités porteuses (K_i), on définit leurs *équivalents effectifs* (\tilde{r}_i) et (\tilde{K}_i) par

$$\begin{cases} \tilde{K}_i = K_i + \sum_{j \neq i} \alpha_{i,j} N_j(t) \\ \tilde{r}_i = r_i \frac{\tilde{K}_i(t)}{K_i} \end{cases}$$

2. La méthode d'Euler consiste à supposer la dérivée \dot{N}_i « suffisamment constante » sur un court intervalle de temps $[t, t + \tau[$ pour approcher la valeur $N_i(t + \tau)$ par le développement au premier ordre $N_i(t) + \dot{N}_i(t)\tau$.

Une équation du système (3) se réduit alors, sur l'intervalle $[t, t + \tau[$, à une équation de la forme (2)

$$\dot{N}_i = \tilde{r}_i \frac{N_i}{\tilde{K}_i} (\tilde{K}_i - N_i),$$

d'où on peut tirer l'équation d'intégration, à l'instant t et pour le pas τ ,

$$N_i(t + \tau) \approx \frac{N_i(t) \tilde{K}_i(t)}{N_i(t) + (\tilde{K}_i(t) - N_i(t)) \exp(-\tilde{r}_i(t)\tau)}. \quad (5)$$

Cette équation est obtenue comme solution de l'équation logistique³. Elle permet d'éviter d'atteindre des abondances négatives. Remarquons enfin que son développement au premier ordre en τ donne exactement la méthode d'Euler.

À partir de cette méthode d'intégration, il est possible de réaliser des simulations. On commence par choisir les paramètres $(\mu_K, \sigma_K, \mu_\alpha, \sigma_\alpha, \gamma_\alpha)$. On tire les capacités porteuses (K_i) et la matrice d'interaction α en fonction. On réalise enfin une intégration numérique jusqu'à la convergence vers le point d'équilibre stable. On peut mesurer la loi de probabilité des abondances, représentée sous la forme d'un histogramme sur la figure 2. Celle-ci se décompose en une distribution continue sur $]0, +\infty[$ et en un atome en 0, correspondant au *taux d'extinction*. La section 2 présente le calcul d'une distribution approchée de l'abondance pour une communauté écologique large ($S \rightarrow \infty$).

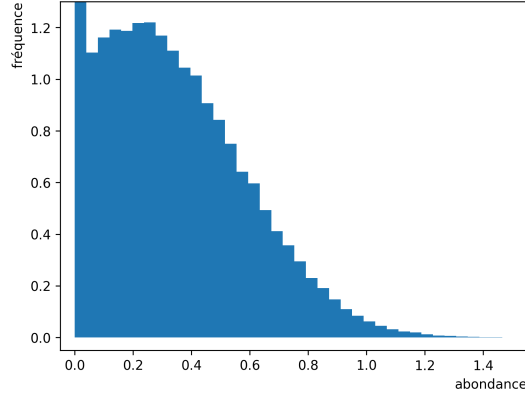


FIGURE 2 – La distribution empirique des abondances dans une communauté écologique à l'équilibre.

2 Distribution des abondances dans une large communauté écologique

La distribution des abondances dans de larges communautés a déjà été décrite à l'aide de la *méthode de la cavité* dans [1]. Issue de la physique statistique, cette méthode permet d'étudier le comportement stochastique d'une particule dans un large système. On commence par ajouter une nouvelle particule à l'ensemble. En découle une perturbation, qu'on résume à son premier ordre non nul. Il est alors possible de déterminer le comportement approché de cette nouvelle particule via un théorème central limite. Si la particule supplémentaire avait été choisie avec des caractéristiques identiques à celles du système initial, alors ce comportement est en fait le même pour toute particule du système. Dans notre cadre – où les espèces remplacent les particules –, l'usage de cette méthode n'a jamais donné lieu, à notre connaissance,

3. On peut résoudre l'équation logistique (2) en utilisant le changement de variable $M = \frac{1}{N}$ pour se ramener à une équation différentielle linéaire (inhomogène) du premier ordre.

à une démonstration mathématique complète. Ce stage a été l'occasion de clarifier les étapes de cette méthode à l'aide de résultats rigoureux.

Dans la sous-section 2.1, on calcule une expression fractionnaire pour l'abondance à l'aide de la méthode de la cavité. Dans la sous-section 2.2, on présente un théorème central limite récent pour déterminer la loi du numérateur. Dans la sous-section 2.3, des calculs fastidieux suffisent à déterminer la limite du dénominateur. Dans la sous-section 2.4, le théorème 1 ci-contre est démontré, puis comparé aux résultats numériques de la section 1.

Théorème 1. *Étant donnée une suite – indexée par $S \in \mathbf{N}$ – de communautés écologiques de S espèces – indexées par $i \in \llbracket 1, S \rrbracket$ – tirées aléatoirement et indépendamment avec les paramètres $(\mu_K, \sigma_K, \mu_\alpha, \sigma_\alpha, \gamma_\alpha)$ indépendants de S et admettant un équilibre,*

$$N_i^{(S)} \underset{S \rightarrow \infty}{\text{en loi}} \left(\frac{\mu_K + \mu_\alpha \mathbf{E}[N] + \sqrt{\sigma_K^2 + \sigma_\alpha^2 \mathbf{E}[N^2]} X}{1 - \gamma_\alpha \sigma_\alpha^2} \right) \text{ pour toute espèce } i,$$

où $(N_i^{(S)})$ désigne l'équilibre de la communauté S , X une variable aléatoire réelle suivant la loi normale centrée réduite $\mathcal{N}(0, 1)$, $\mathbf{E}[N]$ l'espérance limite $\lim \mathbf{E}[N^{(S)}]$ et $\mathbf{E}[N^2]$ le moment d'ordre 2 limite $\lim \mathbf{E}[(N^{(S)})^2]$.

2.1 Une expression fractionnaire pour l'abondance

Indiquons tout de suite que les limites $\mathbf{E}[N]$ et $\mathbf{E}[N^2]$ existent et sont finies. Pour obtenir un tel résultat, on peut exprimer les quantités⁴ $\mathbf{E}[N^{(S)}]$ et $\mathbf{E}[(N^{(S)})^2]$ à l'aide du taux d'extinction $\mathbf{P}[N^{(S)} = 0]$. Or cette suite de taux, évidemment bornée, s'avère être monotone. Elle est donc convergente. En particulier, soulignons que $N_i^{(S)} = \mathbf{E}[N^{(S)}] + O(\sqrt{\mathbf{V}[N^{(S)}]}) = O(1)$.

Dans le cadre de la méthode de la cavité, on considère désormais le passage de la communauté d'indice S à celle d'indice $S+1$. Pour cela, on réécrit l'équation (4) de la communauté $S+1$ à l'équilibre via les perturbations $\xi_i^{(S+1)} = \frac{S}{S+1} \alpha_{i,S+1}^{(S)} N_{S+1}^{(S+1)}$, quitte à supposer que $\alpha^{(S+1)} = \frac{S}{S+1} \alpha^{(S)}$ en prolongeant la matrice $\alpha^{(S)}$. On obtient ainsi

$$N_i^{(S+1)} \left(K_i - N_i^{(S+1)} + \sum_{j \neq i} \frac{S}{S+1} \alpha_{i,j}^{(S)} N_j^{(S+1)} + \xi_i^{(S+1)} \right) = 0.$$

D'une part, remarquons que $\alpha^{(S)} = \frac{\mu_\alpha}{S} + O\left(\frac{\sigma_\alpha}{\sqrt{S}}\right) = O\left(\frac{1}{\sqrt{S}}\right)$. D'autre part, rappelons que $N_{S+1}^{(S+1)} = O(1)$. On constate ainsi que la perturbation $\xi_i^{(S+1)}$ est de l'ordre de $\frac{1}{\sqrt{S}}$. En notant $\beta_{i,j}^{(S+1)} = \frac{\partial N_i^{(S+1)}}{\partial \xi_j^{(S+1)}}$, on peut procéder au développement limité

$$\begin{aligned} N_i^{(S+1)} &= N_i^{(S)} + \sum_j \frac{\partial N_i^{(S+1)}}{\partial \xi_j^{(S+1)}} \xi_j^{(S+1)} + o\left(\xi_j^{(S+1)}\right) \\ &= N_i^{(S)} + N_{S+1}^{(S+1)} \sum_j \frac{S}{S+1} \beta_{i,j}^{(S+1)} \alpha_{j,S+1}^{(S)} + o\left(\frac{1}{\sqrt{S}}\right). \end{aligned} \tag{6}$$

4. Les moments des abondances $(N_i^{(S)})$ ne dépendent pas de l'indice i , qu'on ne prend donc pas la peine de spécifier.

Lorsque l'espèce introduite $S + 1$ n'est pas poussée vers l'extinction, on peut simplifier l'écriture de son équilibre (4) et utiliser le développement limité (6) pour obtenir l'expression

$$N_{S+1}^{(S+1)} = \frac{K_{S+1} + \frac{S}{S+1} \sum_i \alpha_{S+1,i}^{(S)} N_i^{(S)} + o\left(\frac{1}{\sqrt{S}}\right)}{1 - \frac{S}{S+1} \sum_{i,j} \alpha_{S+1,i}^{(S)} \beta_{i,j}^{(S+1)} \alpha_{j,S+1}^{(S)}}. \quad (7)$$

De plus, lorsque l'espèce introduite $S + 1$ s'est éteinte à l'équilibre, l'expression fractionnaire (7) est négative. En effet, dans le cas contraire, les équations de Lotka-Volterra généralisées (3) ne seraient pas dans un équilibre stable⁵, ce qui constitue une contradiction avec l'hypothèse d'admission d'un équilibre. En conséquence, que l'espèce $S + 1$ soit éteinte ou non,

$$N_{S+1}^{(S+1)} = \left(\frac{K_{S+1} + \frac{S}{S+1} \sum_i \alpha_{S+1,i}^{(S)} N_i^{(S)} + o\left(\frac{1}{\sqrt{S}}\right)}{1 - \frac{S}{S+1} \sum_{i,j} \alpha_{S+1,i}^{(S)} \beta_{i,j}^{(S+1)} \alpha_{j,S+1}^{(S)}} \right)_+. \quad (8)$$

Or les espèces sont indifférenciables, et l'expression (8) vaut donc pour toutes les espèces de la communauté $S + 1$. Désormais, il s'agit d'étudier et de décrire le numérateur et le dénominateur de (8) lorsque $S \rightarrow \infty$ pour obtenir un résultat général sur la loi limite des abondances $(N_i^{(S)})$. L'intérêt de la méthode de la cavité est d'avoir remplacé les produits de variables dépendantes $(\alpha_{S+1,i}^{(S+1)} N_i^{(S+1)})$ par les produits de variables indépendantes $(\alpha_{S+1,i}^{(S)} N_i^{(S)})$.

2.2 Une limite aléatoire pour le numérateur

On s'intéresse à la loi limite de la somme $\sum_i \alpha_{S+1,i}^{(S)} N_i^{(S)}$, qui suffira à caractériser la totalité du numérateur de l'expression (8). On constate que les variables $(\alpha_{S+1,i}^{(S)})$ sont mutuellement indépendantes par définition. De plus, elles sont indépendantes des abondances $(N_i^{(S)})$, puisque celles-ci ne dépendent que de la communauté d'indice S . Pour autant, ces mêmes abondances sont couplées via les équations (4), ce qui empêche d'utiliser le théorème central limite directement.

Malgré tout, on constate que plus la communauté est large ($S \rightarrow \infty$), plus l'effet d'une espèce sur une autre est négligeable : $\alpha_{i,j}^{(S)} N_j^{(S)} \rightarrow 0$. En conséquence, deux espèces tendent à « devenir indépendantes », laissant entrevoir l'espoir d'un théorème central limite exploitant une « indépendance asymptotique ». Énonçons et appliquons un tel théorème, démontré dans [3]. Soulignons que c'est un tel théorème central limite qui manque généralement aux utilisations rigoureuses de la méthode de la cavité.

Théorème 2 (Dependent Lindeberg central limit theorem). *Étant donnée une famille $(X_m^{(n)})_{m \leq n}$ de variables aléatoires réelles d'espérance nulle, si :*

1. $\exists \delta > 0, \forall m \in \mathbf{N}, \forall n \in \mathbf{N}, \mathbf{E} \left[\left| X_m^{(n)} \right|^{2+\delta} \right] < \infty;$
2. $A_n = \sum_{m=1}^n \mathbf{E} \left[\left| X_m^{(n)} \right|^{2+\delta} \right] \xrightarrow{n \rightarrow \infty} 0;$
3. $\exists \sigma > 0, \sigma_n^2 = \sum_{m=1}^n \mathbf{V} [X_m^{(n)}] \xrightarrow{n \rightarrow \infty} \sigma^2;$

5. Pour une condition initiale avec $N_i > 0$, les équations de Lotka-Volterra généralisées (3) et une expression fractionnaire (7) strictement positive impliquent que $\lim_{N_i \rightarrow 0} \frac{1}{N_i} \dot{N}_i > 0$, d'où, inévitablement, $\lim_{t \rightarrow \infty} N_i > 0$.

$$4. \forall t \in \mathbf{R}, T_n(t) = \sum_{m=1}^n \left| \left\langle e^{it \sum_{k=1}^{m-1} X_k^{(n)}} \mid e^{it X_m^{(n)}} \right\rangle \right| \xrightarrow{n \rightarrow \infty} 0;$$

alors

$$S_n = \sum_{m=1}^n X_m^{(n)} \xrightarrow[n \rightarrow \infty]{\text{en loi}} \mathcal{N}(0, \sigma^2).$$

En l'occurrence, on s'intéresse à la famille $(X_i^{(S)} = \alpha_{S+1,i}^{(S)} N_i^{(S)} - \frac{\mu_\alpha}{S} \mathbf{E}[N^{(S)}])$. Il s'agit bien de variables aléatoires réelles et, par définition, d'espérance nulle. Par récurrence, on constate que ces variables sont dans L^3 – entre autres –. Ainsi, la première condition du théorème 2 est vérifiée pour $\delta = 1$.

En particulier, la somme A_S est bien définie. Une inégalité triangulaire et un développement binomial permettent de vérifier la seconde condition du théorème 2.

$$\begin{aligned} A_S &= \sum_{i=1}^S \mathbf{E} \left[\left| X_i^{(S)} \right|^3 \right] \leq \sum_{i=1}^S \mathbf{E} \left[\left(\left| \alpha_{S+1,i}^{(S)} N_i^{(S)} \right| + \left| \frac{\mu_\alpha}{S} \mathbf{E}[N_i^{(S)}] \right| \right)^3 \right] \\ &= \sum_{i=1}^S \sum_{k=0}^3 \binom{3}{k} \mathbf{E} \left[\left| \alpha_{S+1,i}^{(S)} N_i^{(S)} \right|^k \right] \mathbf{E} \left[\left| \frac{\mu_\alpha}{S} \mathbf{E}[N_i^{(S)}] \right| \right]^{3-k} \\ &= \sum_{i=1}^S \sum_{k=0}^3 O(1) O\left(\frac{1}{\sqrt{S}^k}\right) O\left(\frac{1}{S^{3-k}}\right) = O(S) O\left(\frac{1}{S\sqrt{S}}\right) \xrightarrow{S \rightarrow \infty} 0. \end{aligned}$$

La troisième condition du théorème 2 découle de l'existence de la limite $\mathbf{E}[N^2]$, expliquée dans la sous-section 2.1. En effet,

$$\begin{aligned} \sigma_S^2 &= \sum_{i=1}^S \mathbf{V} \left[\alpha_{S+1,i}^{(S)} N_i^{(S)} \right] \\ &= \sum_{i=1}^S \left(\mathbf{V} \left[\alpha_{S+1,i}^{(S)} \right] \mathbf{V} \left[N^{(S)} \right] + \mathbf{V} \left[\alpha_{S+1,i}^{(S)} \right] \mathbf{E} \left[N_i^{(S)} \right]^2 + \mathbf{V} \left[N^{(S)} \right] \mathbf{E} \left[\alpha_{S+1,i}^{(S)} \right]^2 \right) \\ &= \sigma_\alpha^2 \mathbf{E} \left[\left(N^{(S)} \right)^2 \right] + \frac{\mu_\alpha^2}{S} \mathbf{V} \left[N^{(S)} \right], \end{aligned}$$

d'où on déduit que $\sigma^2 = \lim \sigma_S^2 = \sigma_\alpha^2 \mathbf{E}[N^2]$.

Il ne reste plus qu'à vérifier la quatrième condition du théorème 2, la seule différant des conditions de Lindeberg et intégrant l'idée d'une « dépendance négligeable » ou d'une « indépendance asymptotique ». On commence par réexprimer $T_S(t)$ avec la factorisation

$$\sum_{j=1}^S \left| \left\langle e^{it \sum_{k=1}^{j-1} X_k^{(S)}} \mid e^{it X_j^{(S)}} \right\rangle \right| = \sum_{j=1}^S \left| e^{-it \frac{\mu_\alpha}{S} j \mathbf{E}[N^{(S)}]} \left\langle e^{it \sum_{k=1}^{j-1} \alpha_{S+1,k}^{(S)} N_k^{(S)}} \mid e^{it \alpha_{S+1,j}^{(S)} N_j^{(S)}} \right\rangle \right|.$$

On peut alors utiliser le développement limité de l'exponentielle au second ordre

$$\exp\left(it \alpha_{S+1,j}^{(S)} N_j^{(S)}\right) = 1 + it \alpha_{S+1,j}^{(S)} N_j^{(S)} + it \alpha_{S+1,j}^{(S)} \sum_{k \neq j} \alpha_{j,k}^{(S)} N_k^{(S)} - t^2 \left(\alpha_{S+1,j}^{(S)}\right)^2 N_j^{(S)2} + o\left(\frac{1}{S}\right)$$

pour calculer la covariance

$$\left\langle e^{it \sum_{k=1}^{j-1} \alpha_{S+1,k}^{(S)} N_k^{(S)}} \mid e^{it \alpha_{S+1,j}^{(S)} N_j^{(S)}} \right\rangle = -t^2 (j-1) \frac{\mu_\alpha^3}{S^3} \mathbf{V} \left[N^{(S)} \right] - i 2 t^2 (j-1)^2 \frac{\mu_\alpha^4}{S^4} \mathbf{E} \left[N^{(S)} \right] \mathbf{V} \left[N^{(S)} \right] + o\left(\frac{1}{S}\right).$$

Finalement, on peut appliquer les convergences de $\mathbf{E}[N^{(S)}]$ et $\mathbf{V}[N^{(S)}]$ pour obtenir la limite

$$T_S(t) = \sum_{j=1}^S \left| e^{-it \frac{\mu_\alpha}{S} j \mathbf{E}[N^{(S)}]} \left\langle e^{it \sum_{k=1}^{j-1} \alpha_{S+1,k}^{(S)} N_k^{(S)}} \left| e^{it \alpha_{S+1,j}^{(S)} N_j^{(S)}} \right. \right\rangle \right| = \sum_{j=1}^S \left| O(1) o\left(\frac{1}{S}\right) \right| \xrightarrow{S \rightarrow \infty} 0.$$

En conclusion, le théorème 2 s'applique, et la famille $(X_i^{(S)})$ converge en loi vers $\mathcal{N}(0, \sigma_\alpha^2 \mathbf{E}[N^2])$. Sachant que la suite $(\mu_\alpha \mathbf{E}[N^{(S)}])$ converge vers $\mu_\alpha \mathbf{E}[N]$, on obtient la loi limite de la somme $\sum_i \alpha_{S+1,i}^{(S)} N_i^{(S)}$ par linéarité sous la forme du lemme 1.

Lemme 1. $\sum_i \alpha_{S+1,i}^{(S)} N_i^{(S)} \xrightarrow[S \rightarrow \infty]{\text{en loi}} \mathcal{N}(\mu_\alpha \mathbf{E}[N], \sigma_\alpha^2 \mathbf{E}[N^2])$.

2.3 Une limite déterminée pour le dénominateur

On s'intéresse à la valeur limite de la somme $\sum_{i,j} \alpha_{S+1,i}^{(S)} \beta_{i,j}^{(S+1)} \alpha_{j,S+1}^{(S)}$, qui suffira à caractériser la totalité du dénominateur de l'expression (8). On décompose cette somme en différenciant la diagonale $i = j$ des autres termes $i \neq j$, l'idée étant que les coefficients $\beta_{i,j}^{(S+1)} = \partial N_i^{(S+1)} / \partial \xi_j^{(S+1)}$ correspondent à des effets directs et indirects selon ces deux cas. Vérifions cette intuition puis étudions l'espérance et la variance de la somme. Pour simplifier, les différentes covariances apparaissant dans les calculs sont dissimulées dans un ordre de grandeur commun $O(S^p)$, avec $p < -4$ ⁶.

Dans cette sous-section, et sauf indication contraire, la notation « i, j » correspond au cas non-diagonal $i \neq j$ tandis qu'on utilise les notations « i, i » et « j, j » dans le cas diagonal. En dérivant l'équation (4) dans le cas $N_i^{(S)} > 0$ – sans quoi $\beta_{i,i}^{(S+1)} = 0$ et $\beta_{i,j}^{(S+1)} = 0$ –, on a les équations

$$\beta_{i,i}^{(S+1)} = 1 + \frac{S}{S+1} \sum_{j \neq i} \alpha_{i,j}^{(S)} \beta_{j,i}^{(S+1)}$$

et

$$\beta_{i,j}^{(S+1)} = \frac{\partial \xi_i^{(S+1)}}{\partial \xi_j^{(S+1)}} + \frac{S}{S+1} \alpha_{i,j}^{(S)} \beta_{j,j}^{(S+1)} + \frac{S}{S+1} \sum_{k \neq i,j} \alpha_{i,k}^{(S)} \beta_{k,j}^{(S+1)}.$$

De là, on peut calculer les ordres de grandeur des espérances de $\beta_{i,i}^{(S+1)}$ et $\beta_{i,j}^{(S+1)}$. Ainsi, dans le cas diagonal,

$$\mathbf{E}[\beta_{i,i}^{(S+1)}] = 1 + \frac{S-1}{S+1} \mu_\alpha \mathbf{E}[\beta_{i,j}^{(S+1)}] + O(S^{p+1})$$

et, dans le cas non-diagonal,

$$\mathbf{E}[\beta_{i,j}^{(S+1)}] = \frac{1}{S+1} \mu_\alpha \mathbf{E}[\beta_{i,i}^{(S+1)}] + \frac{S-1}{S+1} \mu_\alpha \mathbf{E}[\beta_{i,j}^{(S+1)}] + O(S^{p+1}).$$

En rassemblant ces deux résultats, on obtient

$$\mathbf{E}[\beta_{i,j}^{(S+1)}] = \frac{1}{S+1} \mu_\alpha + \left(\frac{S-1}{S+1} + \frac{S-1}{(S+1)^2} \right) \mu_\alpha \mathbf{E}[\beta_{i,j}^{(S+1)}] + O(S^{p+1}),$$

⁶. Ce résultat peut être établi par de longs calculs mais ceux-ci ne permettent pas de mieux comprendre les communautés écologiques ou la distribution des abondances. Il est donc admis ici sans démonstration.

d'où $\mathbf{E} \left[\beta_{i,j}^{(S+1)} \right] = O\left(\frac{1}{S}\right)$ et $\mathbf{E} \left[\beta_{i,i}^{(S+1)} \right] = 1 + O\left(\frac{1}{S}\right)$.

Le même raisonnement peut être utilisé pour calculer les variances et obtenir

$$\mathbf{V} \left[\beta_{i,i}^{(S+1)} \right] = \left(\frac{S}{S+1} \right)^2 (S-1) \left[\left(\frac{\sigma_\alpha^2}{S} + \frac{\mu_\alpha^2}{S^2} \right) \mathbf{V} \left[\beta_{i,j}^{(S+1)} \right] + \frac{\sigma_\alpha^2}{S} \left(\mathbf{V} \left[\beta_{i,j}^{(S+1)} \right] + \mathbf{E} \left[\beta_{i,j}^{(S+1)} \right]^2 \right) \right] + O(S^p),$$

puis

$$\mathbf{V} \left[\beta_{i,j}^{(S+1)} \right] = \left(\frac{S}{S+1} \right)^2 \frac{\left(\frac{S}{S+1} \right)^2 \frac{S-1}{S} \frac{\sigma_\alpha^2}{S} \sigma_\alpha^2 \mathbf{E} \left[\beta_{i,j}^{(S+1)} \right]^2 + \frac{\sigma_\alpha^2}{S} \mathbf{E} \left[\beta_{i,i}^{(S+1)} \right]^2 + O\left(\frac{1}{S^2}\right) + \frac{S-1}{S} \sigma_\alpha^2 \mathbf{E} \left[\beta_{i,j}^{(S+1)} \right]^2 + O(S^{p+2})}{1 - \left(\frac{S}{S+1} \right)^2 \left[4 \left(\frac{S}{S+1} \right)^2 (S-1) \left(\frac{\sigma_\alpha^2}{S} + O\left(\frac{1}{S^2}\right) \right)^2 + 2 \left(\frac{S}{S+1} \right)^2 (S-1) \left(\frac{\sigma_\alpha^2}{S} + O\left(\frac{1}{S^2}\right) \right) \right]},$$

pour finalement conclure que $\mathbf{V} \left[\beta_{i,j}^{(S+1)} \right] = O\left(\frac{1}{S}\right)$ et $\mathbf{V} \left[\beta_{i,i}^{(S+1)} \right] = O\left(\frac{1}{S}\right)$. En guise de résumé, les coefficients diagonaux $\beta_{i,i}^{(S+1)}$ correspondent à des effets directs de la forme $1 + O\left(\frac{1}{\sqrt{S}}\right)$, tandis que les autres représentent des effets indirects négligeables de la forme $\beta_{i,j}^{(S+1)} = O\left(\frac{1}{\sqrt{S}}\right)$.

Pour déterminer l'espérance de la somme, la linéarité nous permet de considérer indépendamment les termes $\alpha_{S+1,i}^{(S)} \beta_{i,j}^{(S+1)} \alpha_{j,S+1}^{(S)}$. Lorsque $i \neq j$, les coefficients $\alpha_{S+1,i}^{(S)}$ et $\alpha_{j,S+1}^{(S)}$ sont indépendants, d'où

$$\mathbf{E} \left[\alpha_{S+1,i}^{(S)} \beta_{i,j}^{(S+1)} \alpha_{j,S+1}^{(S)} \right] = \frac{\mu_\alpha^2}{S^2} O\left(\frac{1}{S}\right) + O(S^p).$$

Lorsque $i = j$, le même calcul laisse émerger un terme de corrélation supplémentaire, et

$$\mathbf{E} \left[\alpha_{S+1,i}^{(S)} \beta_{i,i}^{(S+1)} \alpha_{j,S+1}^{(S)} \right] = \left(\frac{\mu_\alpha^2}{S^2} + \gamma_\alpha \frac{\sigma_\alpha^2}{S} \right) \left(1 + O\left(\frac{1}{S}\right) \right) + O(S^p).$$

En rassemblant ces résultats, on constate que la diagonale domine les autres termes. En effet, l'espérance totale est

$$\mathbf{E} \left[\sum_{i,j} \alpha_{S+1,i}^{(S)} \beta_{i,j}^{(S+1)} \alpha_{j,S+1}^{(S)} \right] = \gamma_\alpha \sigma_\alpha^2 + O\left(\frac{1}{S}\right) + O(S^{p+2}). \quad (9)$$

L'étude de la variance est un calcul similaire, quoique plus laborieux. On commence par souligner que les termes peuvent toujours être étudiés un à un, puis rassemblés via

$$\mathbf{V} \left[\sum_{i,j} \alpha_{S+1,i}^{(S)} \beta_{i,j}^{(S+1)} \alpha_{j,S+1}^{(S)} \right] = \sum_{i,j} \mathbf{V} \left[\alpha_{S+1,i}^{(S)} \beta_{i,j}^{(S+1)} \alpha_{j,S+1}^{(S)} \right] + O(S^{p+4}).$$

On réécrit un terme $\mathbf{V} \left[\alpha_{S+1,i}^{(S)} \beta_{i,j}^{(S+1)} \alpha_{j,S+1}^{(S)} \right]$ sous la forme

$$\mathbf{V} \left[\beta_{i,j}^{(S+1)} \right] \mathbf{V} \left[\alpha_{S+1,i}^{(S)} \alpha_{j,S+1}^{(S)} \right] + \mathbf{V} \left[\beta_{i,j}^{(S+1)} \right] \mathbf{E} \left[\left(\alpha_{S+1,i}^{(S)} \alpha_{j,S+1}^{(S)} \right)^2 \right] + \mathbf{V} \left[\alpha_{S+1,i}^{(S)} \alpha_{j,S+1}^{(S)} \right] \mathbf{E} \left[\left(\beta_{i,j}^{(S+1)} \right)^2 \right] + O(S^p) + O(S^{2p}).$$

Lorsque $i \neq j$, l'indépendance entre $\alpha_{S+1,i}^{(S)}$ et $\alpha_{j,S+1}^{(S)}$ permet d'obtenir que

$$\mathbf{V} \left[\alpha_{S+1,i}^{(S)} \beta_{i,j}^{(S+1)} \alpha_{j,S+1}^{(S)} \right] = \frac{\mu_\alpha^4}{S^4} O\left(\frac{1}{S}\right) + \left(\frac{\sigma_\alpha^4}{S^2} + 2 \frac{\mu_\alpha^2 \sigma_\alpha^2}{S^2} \right) O\left(\frac{1}{S}\right) = O\left(\frac{1}{S^3}\right),$$

tandis que, lorsque $i = j$, un terme de corrélation s'ajoute à nouveau pour donner

$$\mathbf{V} \left[\alpha_{S+1,i}^{(S)} \beta_{i,j}^{(S+1)} \alpha_{j,S+1}^{(S)} \right] = \left(\frac{\mu_\alpha^4}{S^4} + 2\gamma_\alpha \frac{\sigma_\alpha^2}{S} \right) O\left(\frac{1}{S}\right) + \left(\frac{\sigma_\alpha^4}{S^2} + 2\frac{\mu_\alpha^2 \sigma_\alpha^2}{S^2} + O\left(\frac{1}{S}\right) \right) \left(1 + O\left(\frac{1}{S}\right) \right) = O\left(\frac{1}{S^2}\right).$$

En combinant ces résultats, on est en mesure de montrer que la variance recherchée est négligeable. De fait,

$$\mathbf{V} \left[\sum_{i,j} \alpha_{S+1,i}^{(S)} \beta_{i,j}^{(S+1)} \alpha_{j,S+1}^{(S)} \right] = O(S) O\left(\frac{1}{S^2}\right) + O(S^2) O\left(\frac{1}{S^3}\right) + O(S^{p+4}) \xrightarrow{S \rightarrow \infty} 0. \quad (10)$$

En rassemblant les résultats (9) et (10), on obtient que la somme $\sum_{i,j} \alpha_{S+1,i}^{(S)} \beta_{i,j}^{(S+1)} \alpha_{j,S+1}^{(S)}$ est de la forme $\gamma_\alpha \sigma_\alpha^2 + o(1)$. Autrement dit, on dispose du lemme 2.

Lemme 2. $\sum_{i,j} \alpha_{S+1,i}^{(S)} \beta_{i,j}^{(S+1)} \alpha_{j,S+1}^{(S)} \xrightarrow{S \rightarrow \infty} \gamma_\alpha \sigma_\alpha^2.$

2.4 Une distribution approchée dans une large communauté

En appliquant les lemmes 1 et 2 sur l'expression (8) et en utilisant la continuité de la partie positive, on obtient que

$$N_{S+1}^{(S+1)} \xrightarrow[S \rightarrow \infty]{\text{en loi}} \left(\frac{\mu_K + \mu_\alpha \mathbf{E}[N] + \sqrt{\sigma_K^2 + \sigma_\alpha^2 \mathbf{E}[N^2]} X}{1 - \gamma_\alpha \sigma_\alpha^2} \right)_+, \quad (11)$$

où $X \sim \mathcal{N}(0, 1)$. Il s'agit exactement de la convergence annoncée en début de section 2.

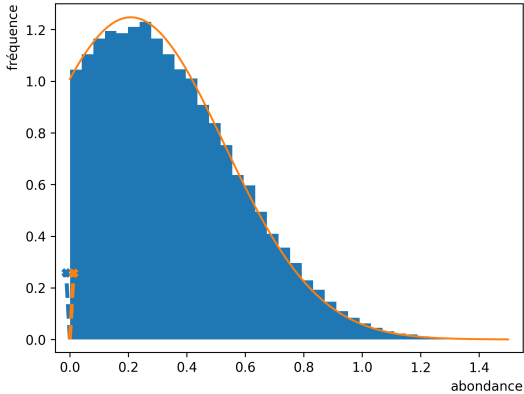


FIGURE 3 – Les distributions empirique (en bleu) et théorique (en orange) des abondances dans une communauté écologique à l'équilibre.

Décrivons désormais ce résultat théorique, et comparons-le aux résultats empiriques de la section 1. La distribution limite des abondances est une loi composée d'une partie continue – une gaussienne tronquée en 0 – et d'une partie discrète – un atome représentant le taux d'extinction en 0 –. En reprenant les mesures de la figure 2, et en considérant toutes les abondances inférieures à 10^{-3} comme associées à des espèces éteintes, on obtient un histogramme et un atome comparables à la distribution prédite⁷ par le théorème 1. On constate que les résultats numériques semblent être très bien expliqués par la distribution limite calculée dans cette section 2.

On est parvenu à relativement bien décrire ce modèle de communauté écologique. Cependant, les interactions écologiques réelles ne sont pas aléatoires mais dictées, entre autres, par l'évolution des espèces

7. Pour dessiner la figure 3, les paramètres $\mathbf{E}[N]$ et $\mathbf{E}[N^2]$ de la distribution théorique sont remplacés par leurs estimateurs par la « moyenne empirique » $\frac{1}{n} \Sigma$.

impliquées. L'article [2] utilise justement ce modèle de communauté écologique pour étudier l'évolution des interactions au sein d'une communauté écologique soumise à une pression sélective à l'échelle collective. La section 3 présente une modélisation et une description d'une telle évolution.

3 Évolution des interactions écologiques pour une pression sélective collective

L'évolution repose a minima sur deux phénomènes fondamentaux : la mutation et la sélection. Un modèle d'évolution des communautés écologiques est décrit dans [2], qui a été une source d'inspiration pour le sujet de ce stage. Cependant, ce modèle fait un certain nombre de choix, notamment à propos des mutations, sans préciser les conséquences de chacun d'entre eux.

Dans la sous-section 3.1, on décrit les règles de mutation et de sélection appliquées à la matrice d'interaction. Dans la sous-section 3.2, on explique en quoi l'évolution ne dépend en réalité pas du choix de la loi de probabilité des mutations. Dans la sous-section 3.3, on présente brièvement des résultats de cette évolution obtenue par simulation pour une sélection agissant à l'échelle collective.

3.1 Évolution de la matrice d'interaction par mutation et sélection

La sélection est généralement interprétée comme une question d'optimisation. Chaque individu possède une *valeur sélective* propre et, plus celle-ci est élevée, plus les individus sont « reproduits ». Dans notre cas, l'individu considéré est la communauté écologique, et la valeur sélective est une fonction donnée f à valeurs réelles. Par exemple, on s'intéresse par la suite à la sélection collective par la biomasse, c'est-à-dire à la somme des abondances $\sum_i N_i$ une fois à l'équilibre.

Pour réaliser une sélection, il est nécessaire de disposer de différentes communautés écologiques distinctes. Pour cela, partant d'une communauté donnée, on en produit un certain nombre n identiques, à une mutation η aléatoire de loi μ fixée près. On choisit alors la communauté maximisant la valeur sélective f , et on peut recommencer le processus – sous réserve que la nouvelle communauté soit dans le domaine de définition de f –.

Précisons la méthode de mutation choisie dans [2] et utilisée pendant ce stage. On fixe une matrice d'interaction α . Pour produire une mutation de α , on tire aléatoirement une matrice η de diagonale nulle. Ses coefficients suivent la loi normale centrée réduite $\mathcal{N}(0, 1)$. De plus, les coefficients réciproques $\eta_{i,j}$ et $\eta_{j,i}$ présentent un coefficient de corrélation identique à celui de la matrice α , de sorte à n'entraîner aucun biais. On fixe préalablement une force de mutation $\epsilon > 0$ et on réduit α à une matrice β telle que

$$\alpha = \frac{\mu}{S} + \frac{\sigma}{\sqrt{S}}\beta,$$

avec

$$\begin{cases} \frac{\mu}{S} = \frac{1}{S^2} \sum_{i,j} \alpha_{i,j} \\ \frac{\sigma}{\sqrt{S}} = \sqrt{\frac{1}{S^2} \sum_{i,j} (\alpha_{i,j} - \frac{\mu}{S})^2}. \end{cases}$$

On applique alors la mutation η sur la matrice réduite β . En procédant ainsi, l'impact de la mutation ne dépend que de la force ϵ choisie et pas de l'ordre de grandeur de la matrice α . On pose donc

$$\tilde{\beta} = \frac{1}{\sqrt{1+\epsilon^2}} (\beta + \epsilon\eta),$$

d'où on peut définir la matrice d'interaction mutée

$$\tilde{\alpha} = \frac{\mu}{S} + \frac{\sigma}{\sqrt{S}} \tilde{\beta}.$$

En réalisant n mutations différentes partant toutes de la même matrice α , on obtient une collection de matrices mutées $\tilde{\alpha}_1, \dots, \tilde{\alpha}_n$. On peut alors définir la matrice sélectionnée α^* comme le maximiseur⁸, sur cet ensemble fini, de la valeur sélective f .

Une fois ces étapes de mutation et de sélection définies, on peut simuler et étudier l'évolution de la matrice d'interaction. Partant d'une matrice $\alpha(0)$ arbitraire, une génération consiste simplement à passer d'une matrice $\alpha(\tau)$ à une matrice sélectionnée $\alpha(\tau+1) = \alpha(\tau)^*$. Un tel procédé définit une suite de matrices aléatoires, et ce tant que celles-ci sont dans le domaine de définition de la valeur sélective. Par exemple, dans le cas de la sélection par la biomasse présentée précédemment, l'évolution peut se poursuivre jusqu'à obtenir une communauté n'admettant plus d'équilibre. La figure 4 révèle par exemple qu'après plusieurs générations certaines abondances de la communauté peuvent diverger.

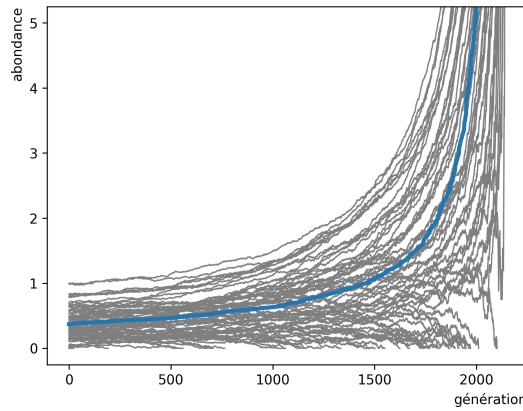


FIGURE 4 – Les abondances à l'équilibre et leur moyenne (en bleu) en fonction de la génération.

3.2 Conséquences sur la trajectoire évolutive du choix des mutations

La loi de probabilités des mutations choisie dans [2] et décrite dans la sous-section 3.1 peut paraître arbitraire. En réalité, on peut montrer que la trajectoire évolutive ne dépend pas de ce choix. On commence par énoncer et interpréter un résultat sur l'espérance de la mutation sélectionnée, avant d'en donner une démonstration.

On fixe des capacités porteuses $(K_i) \in \mathbf{R}^S$. On note \mathbf{I} l'espace vectoriel des matrices de $\mathbf{R}^{S \times S}$ de diagonale nulle. On y choisit une matrice d'interaction $\alpha \in \mathbf{I}$.

On fixe une fonction de sélection f bien définie au voisinage de α dans \mathbf{I}^9 et différentiable en α . On

8. Pour la plupart des lois de probabilité choisies pour la mutation et pour la plupart des fonctions choisies pour la sélection, un tel maximiseur est presque sûrement unique. Si ce n'est pas le cas, il est toujours possible d'introduire une relation d'ordre strict arbitraire sur les matrices pour définir correctement la matrice sélectionnée.

9. Par exemple, dans le cas d'une sélection sur la biomasse, il faut supposer que, pour toute matrice d'interaction $\tilde{\alpha}$ au voisinage de α , la communauté écologique associée admette un équilibre $(N_i(\tilde{\alpha}))$.

note g le gradient de f en α . On pose $g^\perp = \{\eta \in \mathbf{I} \mid \langle \eta \mid g \rangle = 0\}$ le sous-espace orthogonal au gradient g . On considère également $\pi_g : \eta \mapsto \langle \eta \mid g \rangle g$ le projecteur orthogonal le long du gradient g et $\pi_{g^\perp} : \eta \mapsto \eta - \pi_g(\eta)$ le projecteur orthogonal sur le sous-espace supplémentaire g^\perp .

On choisit une loi de probabilité μ sur l'ensemble des matrices de $\mathbf{R}^{S \times S}$ à diagonale nulle. On suppose que la sélection est toujours clairement définie, c'est-à-dire que deux mutations indépendantes de loi μ ont presque sûrement des images différentes par $\eta \mapsto f(\alpha + \epsilon\eta)$ pour tout $\epsilon > 0$.

On suppose que le nombre n de mutations parmi lesquelles la sélection est réalisée est au moins égal à 2. Étant données n mutations $\eta_1, \dots, \eta_n \sim \mu$, on définit $\mathbf{E}_+ = \mathbf{E}[\max_k \pi_g(\eta_k) \mid \max_k \pi_g(\eta_k) > 0]$ l'espérance de progrès suivant le gradient sachant qu'un progrès est fait et $\mathbf{E}_- = -\mathbf{E}[\max_k \pi_g(\eta_k) \mid \max_k \pi_g(\eta_k) \leq 0]$ la régression moyenne, en valeur absolue, sachant qu'une régression a lieu. Enfin, on pose η_ϵ^* la mutation de force $\epsilon > 0$ sélectionnée, c'est-à-dire le maximiseur de $\eta \mapsto f(\alpha + \epsilon\eta)$ parmi η_1, \dots, η_n .

Théorème 3. *Pour $\eta \sim \mu$, si :*

1. $\pi_g(\eta) \perp \pi_{g^\perp}(\eta)$; *(les composantes sélectionnée et non-sélectionnée sont indépendantes)*
2. $\mathbf{E}[\pi_{g^\perp}(\eta)] = 0$; *(la composante non-sélectionnée est non-biaisée)*
3. $\mathbf{P}[\langle \eta \mid g \rangle \leq 0] < \sqrt{\frac{\mathbf{E}_+}{\mathbf{E}_+ + \mathbf{E}_-}}$; *(la composante sélectionnée est peu biaisée¹⁰)*

alors $\mathbf{E}[\eta_\epsilon^*] \xrightarrow{\epsilon \rightarrow 0} \lambda g$, avec $\lambda > 0$.

Le gradient g ne dépend pas de la loi μ . En conséquence, tant que les mutations ne sont pas trop fortes, suffisamment symétriques et « de composantes indépendantes », le choix de leur loi de probabilité n'a pas d'effet sur la trajectoire évolutive moyenne, qui est tangente au gradient de sélection en tout point. Seule la vitesse évolutive dépend du choix de la loi. Précisons également que les mutations présentées dans la sous-section 3.1 vérifient les hypothèses du théorème 3 lorsque $\gamma_\alpha = 0$.

Pour démontrer un tel résultat, commençons par écrire le développement limité au premier ordre $f(\alpha + \epsilon\eta_k) = f(\alpha) + \epsilon \langle \eta_k \mid g \rangle + o(\epsilon)$, pour $k \in [1, n]$. Par unicité presque sûre du maximiseur, il existe une force seuil $\epsilon_0 > 0$ en dessous de laquelle $\eta_\epsilon^* = \eta_{\epsilon_0}^*$ ne dépend plus de ϵ . La sélection peut alors être caractérisée par la maximisation de $\eta \mapsto \epsilon \langle \eta \mid g \rangle$, puisque maximiser $\eta \mapsto f(\alpha + \epsilon\eta)$ équivaut exactement à maximiser $\eta \mapsto \langle \eta \mid g \rangle + \frac{o(\epsilon)}{\epsilon} \xrightarrow{\epsilon \rightarrow 0} \langle \eta \mid g \rangle$.

Dans un premier temps, on peut montrer que $\mathbf{E}[\pi_{g^\perp}(\eta_\epsilon^*)] \xrightarrow{\epsilon \rightarrow 0} 0$. En effet, on peut utiliser la caractérisation de $\eta_{\epsilon_0}^*$ par $\pi_g : \eta \mapsto \langle \eta \mid g \rangle g$ et la condition 1 pour identifier la mutation sélectionnée dans le théorème de l'espérance totale, puis lui appliquer la condition 2.

$$\begin{aligned}
\mathbf{E}[\pi_{g^\perp}(\eta_\epsilon^*)] &= \sum_{k=1}^n \mathbf{P}[\pi_g(\eta_\epsilon^*) = \pi_g(\eta_k)] \mathbf{E}[\pi_{g^\perp}(\eta_\epsilon^*) \mid \pi_g(\eta_\epsilon^*) = \pi_g(\eta_k)] \\
&\xrightarrow{\epsilon \rightarrow 0} \sum_{k=1}^n \mathbf{P}[\pi_g(\eta_{\epsilon_0}^*) = \pi_g(\eta_k)] \mathbf{E}[\pi_{g^\perp}(\eta_{\epsilon_0}^*) \mid \pi_g(\eta_{\epsilon_0}^*) = \pi_g(\eta_k)] \\
&= \sum_{k=1}^n \mathbf{P}[\pi_g(\eta_{\epsilon_0}^*) = \pi_g(\eta_k)] \mathbf{E}[\pi_{g^\perp}(\eta_k) \mid \pi_g(\eta_{\epsilon_0}^*) = \pi_g(\eta_k)] \\
&= \sum_{k=1}^n \mathbf{P}[\pi_g(\eta_{\epsilon_0}^*) = \pi_g(\eta_k)] \mathbf{E}[\pi_{g^\perp}(\eta_k)] \\
&= 0.
\end{aligned}$$

10. Si $\pi_g(\eta)$ est symétrique (ie $\pi_g(\eta)$ et $-\pi_g(\eta)$ suivent la même loi) et si $\mathbf{P}[\pi_g(\eta) = 0] = 0$, alors la condition 3 est vérifiée. La condition 3 énoncée ne fait que souligner qu'un biais contre la sélection peut être combattu, avec une force dépendant de n .

Par linéarité, il en vient que $\mathbf{E}[\eta_\epsilon^*] \xrightarrow{\epsilon \rightarrow 0} \mathbf{E}[\pi_g(\eta_\epsilon^*)]$. Or la projection π_g est à valeurs dans la droite vectorielle $\mathbf{R}g$, de sorte que $\mathbf{E}[\eta_\epsilon^*] \xrightarrow{\epsilon \rightarrow 0} \lambda g$, avec $\lambda = \lim_{\epsilon \rightarrow 0} \mathbf{E}[\langle \eta_\epsilon^* | g \rangle]$.

Dans un second temps, constatons que $\mathbf{E}[\langle \eta_\epsilon^* | g \rangle] \xrightarrow{\epsilon \rightarrow 0} \mathbf{E}[\langle \eta_{\epsilon_0}^* | g \rangle] = \mathbf{E}[\max_k \langle \eta_k | g \rangle]$. Dès lors, pour établir définitivement le théorème 3, une simple reformulation de l'inégalité $\lambda > 0$ suffit. De fait,

$$\begin{aligned} \lambda > 0 &\iff \mathbf{E} \left[\max_k \langle \eta_k | g \rangle \right] > 0 \\ &\iff \mathbf{P}[\exists k, \langle \eta_k | g \rangle > 0] \mathbf{E}_+ - \mathbf{P}[\forall k, \langle \eta_k | g \rangle \leq 0] \mathbf{E}_- > 0 \\ &\iff (1 - \mathbf{P}[\langle \eta_k | g \rangle \leq 0]^n) \mathbf{E}_+ - \mathbf{P}[\langle \eta_k | g \rangle \leq 0]^n \mathbf{E}_- > 0 \\ &\iff \mathbf{E}_+ - \mathbf{P}[\langle \eta_k | g \rangle \leq 0]^n (\mathbf{E}_+ + \mathbf{E}_-) > 0, \end{aligned}$$

ce qui correspond exactement à la condition 3.

3.3 Description de la trajectoire évolutive pour une sélection collective

A donc été établi que la trajectoire évolutive ne dépend que de la pression sélective imposée à la communauté écologique. On a choisi une pression d'échelle collective en désignant l'abondance totale comme valeur sélective. On présente ici brièvement l'évolution produite par une telle sélection. Indiquons que les résultats obtenus sont similaires lorsque les différentes abondances sont pondérées par des poids strictement positifs dans le calcul de la valeur sélective de la communauté. Enfin, d'une simulation à l'autre, la trajectoire évolutive est identique et le système n'apparaît pas comme chaotique a priori.

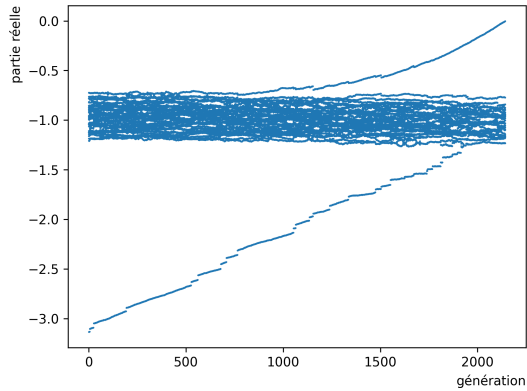


FIGURE 5 – Les parties réelles des valeurs propres de la matrice d'interaction en fonction de la génération.

mutations aléatoires sont appliquées sans sélection, cette valeur propre rejoint à nouveau le disque central. De tels résultats sont étudiés et exposés dans [2] plus précisément.

On remarque l'émergence d'une structure dans la répartition des valeurs propres de la matrice d'interaction sur le plan complexe. Sur la figure 5, on constate que la matrice aléatoire initiale présente une valeur propre réelle correspondant à la moyenne μ_α et un disque de rayon σ_α composé de valeurs propres complexes aléatoires. Au fur et à mesure des sélections, la valeur propre négative isolée croît, favorisant l'essor des abondances à l'équilibre, comme sur la figure 4. En parallèle, une valeur propre réelle positive se sépare du disque. Cette structure nouvelle est imprimée par la sélection et s'avère propre à la trajectoire évolutive. Lorsque des

Note sur les figures

Les figures 2, 3, 4 et 5 sont issues de simulations réalisées pendant le stage avec les paramètres suivants :

- une taille de communauté écologique $S = 100$;
- des capacités porteuses de moyenne $\mu_K = 1$ et d'écart-type $\sigma_K = 0.3$;
- une matrice d'interaction de moyenne $\mu_\alpha = -3$, d'écart-type $\sigma_\alpha = 0.3$ et de corrélation $\gamma_\alpha = 0$;
- des mutations tirées conformément aux explications de la sous-section 3.1;
- une valeur sélective $f = N_i \mapsto \sum_i N_i$, avec (N_i) les abondances à l'équilibre.

Références

- [1] Guy Bunin. Ecological communities with Lotka-Volterra dynamics. *Physical Review E*, 2017.
- [2] Jules Fraboul, Giulio Biroli, and Silvia De Monte. Artificial selection of communities drives the emergence of structured interactions. *Journal of Theoretical Biology*, 2023.
- [3] Jean-Marc Bardet, Paul Doukhan, Gabriel Lang, and Nicolas Ragache. Dependent Lindeberg central limit theorem and some applications. *ESAIM : Probability and Statistics*, 2008.