Feuille 4,
Chaînes de Markov et modèles spatiaux

Exercice 1:
Un modèle de migration.
On considère un modèle très simple d'une population animale dans un habitat ouvert. A chaque génération $n \in \mathbb{N}$, chaque animal y vivant le quitte, indépendamment des autres, avec une probabilité p, et y demeure avec une probabilité $1-p$.
Simultanément et indépendamment, un nombre poissonnien (de paramètre $a > 0$) d'animaux immigrent depuis le monde extérieur dans cet habitat.
On rappelle qu'une loi de Poisson N de paramètre a est définie par :

$$\forall k \in \mathbb{N}, \qquad \mathbb{P}(N = k) = e^{-a} a^k / k!.$$

1. Expliquer pourquoi une loi d'immigration suivant une loi de Poisson peut être pertinente (par exemple en utilisant la limite de lois binomiales).
2. Décrire le nombre X_n d'animaux vivant dans l'habitat par une formule de récurrence, puis donner la matrice de transition P de cette chaîne de Markov.
3. Déterminer $\mathbb{E}(X_n)$.
4. Calculer la loi de X_n quand la loi initiale est une loi de Poisson de paramètre $\lambda > 0$.

Exercice 2:
Chaîne de naissance et mort.
On considère une chaîne de naissance et mort $(X_n : n \in \mathbb{N})$ sur $\mathbb{N}$, c'est à dire une chaîne de Markov dont la matrice de transition Q est définie par

$$Q(0,0) = r_0, \quad Q(0,1) = p_0 > 0, \quad p_0 + r_0 = 1$$

et pour tout $i \geq 1$

$$Q(i,i-1) = q_i > 0, \quad Q(i,i) = r_i, \quad Q(i,i+1) = p_i > 0, \quad q_i + r_i + p_i = 1.$$

On définit $\tau_i := \inf\{n \in \mathbb{N} : X_n = i\}$ (on pose $\inf \varnothing = \infty$) comme le temps d'atteinte de l'état i par la chaîne.
On définit également

$$\gamma_0 = 1, \quad \forall i \geq 1,\ \gamma_i = \frac{q_1 \dots q_i}{p_1 \dots p_i}.$$

et pour $a, b \in \mathbb{N}$ tels que $a + 1 < b$, on considère $\tau = \min(\tau_a, \tau_b)$.

1. Soit u une fonction telle que $Qu(i) = u(i)$ pour tout $a+1 \leq i \leq b-1$, où l'on rappelle que $Qu(i) = \sum_j Q(i,j)u(j)$.
   Montrer que
   $$u(i) = u(a) + (u(a+1) - u(a))\gamma_a^{-1} \sum_{k=a}^{i-1} \gamma_k$$
   pour tout $a \leq i \leq b$ et que u est monotone sur cet intervalle.

2. Montrer que pour tout $a \le i \le b$, $\mathbb{P}_i(\tau < \infty) = 1$ et que
$$\mathbb{P}_i(X_\tau = b) = \frac{\sum_{k=a}^{i-1} \gamma_k}{\sum_{k=a}^{b-1} \gamma_k}.$$
3. Calculer $\mathbb{P}_1(\tau_0 = \infty)$ et déterminer quand la chaîne est récurrente. Que fait-elle sinon ?
4. Exemples : étudier la récurrence de la chaîne dans les cas suivants :
   (a) Le cas homogène : $\forall i \ge 1$, $p_i = p$, $q_i = q$, $r_i = r$.
   (b) $p_i = \frac{i+2}{2i+3}$, $\quad q_i = \frac{i+1}{2i+3}$.
   (c) $p_i = \frac{i+3}{2i+3}$, $\quad q_i = \frac{i}{2i+3}$.

Exercice 3:
Grandes déviations

On considère une suite de variables aléatoires $(X_i : i \in \mathbb{N})$ i.i.d de même loi que la v.a. X. On suppose que X admet des moments exponentiels de tout ordre et on définit pour $\lambda, x \in \mathbb{R}$:
$$m := \mathbb{E}(X), \quad \Lambda(\lambda) := \log\left(\mathbb{E}(e^{\lambda X})\right), \quad \Psi(x) := \sup_{\lambda \in \mathbb{R}}\{\lambda x - \Lambda(\lambda)\} = \sup_{\lambda \ge 0}\{\lambda x - \Lambda(\lambda)\}.$$

On veut montrer que pour tout $x \ge m$,
$$\lim_{n\to\infty} -\frac{1}{n} \log \mathbb{P}\left(\sum_{i=1}^n X_i \ge nx\right) = \Psi(x).$$

a) Montrer la borne inférieure en utilisant l'inégalité de Markov $\mathbb{P}(X \ge 0) \le \mathbb{E}(e^X)$.

b) En utilisant le réel τ et le changement de loi associés à x par
$$x = \Lambda'(\tau) = \frac{\mathbb{E}\left(Xe^{\tau X}\right)}{\mathbb{E}\left(e^{\tau X}\right)}, \quad \mathbb{P}(\widetilde{X} \in \mathrm{d}u) = \frac{e^{\tau u}}{\mathbb{E}(e^{\tau X})}\, \mathbb{P}(X \in \mathrm{d}u),$$
montrer que pour tout $m \le x < y$,
$$\limsup_{n\to\infty} -\frac{1}{n} \log \mathbb{P}\left(nx \le \sum_{i=1}^n X_i \le ny\right) \le \tau y - \Lambda(\tau).$$

c) Conclure.

Exercice 3.I. Application au branchement (ramification) spatial.

On considère le modèle de reproduction spatial suivant. À chaque génération $n \in \mathbb{N}$, chaque individu meurt après avoir donné naissance indépendamment à un nombre aléatoire N d'enfants qui se dispersent, indépendamment les uns des autres et indépendamment du passé. Le déplacement est donné par la variable X, c'est à dire que si la mère est à la position x, la position d'un enfant est distribuée comme la variable $x + X$.
On note Z_n la taille de la population à la génération n, $(X_n^i : 1 \le i \le Z_n)$ les positions des individus à la génération n et
$$u_n(I) = \mathbb{E}(\#\{i : X_n^i \in I\})$$
le nombre moyen d'individus dans l'intervalle I à la génération n.

a) Où les individus se trouvent-t-ils en majorité au temps n ?

b) Montrer que pour tout $x \ge \mathbb{E}(X)$,
$$\lim_{n\to\infty} -\frac{1}{n} \log(u_n([nx, \infty])) = -\log(\mathbb{E}(N)) + \Psi(x).$$

où Ψ est la transformée de Fenchel Legendre de X définie à l'exercice précédent.

Exercice 3.II. Application à la stochasticité environnementale.

On regarde une population se reproduisant en environnement aléatoire. C'est à dire qu'à chaque génération $n \in \mathbb{N}$, on tire un environnement (noté $\mathcal{E}_n$) de manière i.i.d. On note $N(e)$ la v.a. de reproduction dans l'environnement e et $m(e) = \mathbb{E}(N(e))$ le nombre moyen d'enfants par individu dans cet environnement. On note Z_n la taille de la population à l'instant n : on part d'un individu ($Z_0 = 1$) et à chaque génération, conditionnellement au fait que l'environnement soit égal à e, les individus se reproduisent indépendamment suivant la v.a. $N(e)$.

a) Rappeler pourquoi le modèle est décrit par la récurrence suivante : Pour tout $n \in \mathbb{N}$

$$Z_{n+1} = \sum_{i=1}^{Z_n} N_n(i),$$

où conditionnellement à $\mathcal{E}_n = e$, les $N_n(i)$ sont des v.a. i.i.d. distribuées comme $N(e)$ et indépendantes de Z_n.
Retrouver également que

$$m(\mathcal{E}_0, ..., \mathcal{E}_{n-1}) = \mathbb{E}(Z_n \mid \mathcal{E}_0, ..., \mathcal{E}_{n-1}) = \exp(S_n)$$

où $S_n = \sum_{i=0}^{n-1} \log(m(\mathcal{E}_i))$ ($S_0 = 0$).

b) Montrer que pour tout $x \geq \mathbb{E}(\log(m(\mathcal{E})))$,

$$\lim_{n\to\infty} -\frac{1}{n} \log \mathbb{P}\left(m(\mathcal{E}_0, ..., \mathcal{E}_{n-1}) \geq \exp(nx)\right) = \Psi(x).$$

où Ψ est la transformée de Fenchel Legendre de X définie à l'exercice précédent.

c) Interprétez d'un point de vue de la modélisation en écologie ce résultat.

Exercice 4:
Modèle "faucon-colombe" On considère le modèle "faucon-colombe" suivant, où deux espèces sont réparties en des communautés discrètes dans lesquelles les interactions sont dictées par la proportion de chacune des espèces. Ces modèles correspondent plutôt à des stratégies de coopération/tricherie et nous renvoyons à un bon ouvrage sur la théorie des jeux pour plus de détails. On appelle $F_t(x)$ (resp $C_t(x)$) le nombre de faucons (resp. colombes) au temps t à l'emplacement x et

$$P_t(x) = \frac{F_t(x)}{F_t(x) + C_t(x)}$$

la fraction de faucons au temps t à l'emplacement x. Le modèle est décrit de la manière suivante :
- Migration : chaque individu migre au taux m vers l'un de ses plus proches voisins (choisi uniformément au hasard) ;
- Mort : chaque individu du site x meurt au taux $\kappa(F_t(x) + C_t(x))$;
- De surcroît, pour chaque individu, le taux de naissance (ou de mort, si le paramètre correspondant est négatif) des faucons est de la forme $aP_t(x) + b(1 - P_t(x))$ et celui des colombes s'écrit $cP_t(x) + d(1 - P_t(x))$.

1) Commenter le modèle en expliquant les différents termes.

2) Expliquer pourquoi l'analogue déterministe est donné par

$$\begin{cases} u' = u\big(a\frac{u}{u+v} + b\frac{v}{u+v} - \kappa(u+v)\big) \\ v' = v\big(c\frac{u}{u+v} + d\frac{v}{u+v} - \kappa(u+v)\big), \end{cases}$$

Dans quel cadre le modèle déterministe serait-il une bonne approximation du modèle précédent ?

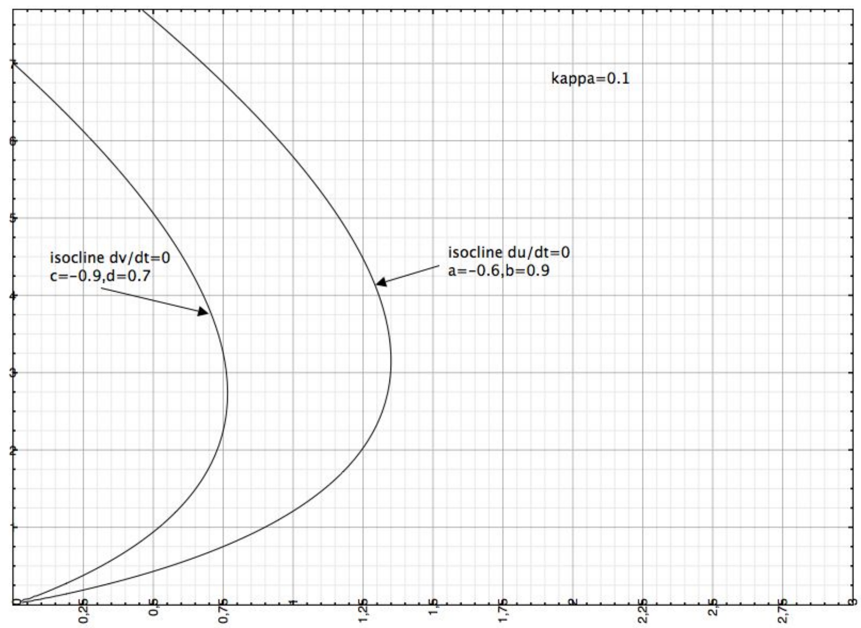

FIG. 1 – Isoclines du modèle faucon colombe quand $\kappa = 0,1$, $a = -0.6, b = 0.9, c = -0.9, d = 0.7$.

3) Dans le cas où $\kappa = 0,1$, $a = -0.6, b = 0.9, c = -0.9, d = 0.7$, prédire le comportement des solutions du modèle déterministe, en utilisant les isoclines tracées ci-dessous.

Exercice 5:
Marquage de territoire.
Le but de cet exercice est de modéliser le déplacement et ainsi la répartition d'animaux, en particulier de carnivores comme les loups, coyotes ou lynx (en dimension 1 ici, la dimension 2 étant plus technique).

1) Commencons par établir des EDP qui apparaissent naturellement pour modéliser le mouvement d'individus.
On note $u(x,t)$ la densité de présence à la position x au temps t. Si l'on note X_t la position aléatoire d'un animal au temps t, on a ainsi

$$\mathbb{P}(a \leq X_t < b) = \int_a^b u(x,t)\mathrm{d}x.$$

On note $L_h(x,t)$ (resp. $R_h(x,t)$) la probabilité d'aller de x en $x-h$ (resp. $x+h$) entre le temps t et le temps $t+h^2$.
a) Justifier que la densité u est solution de

$$\frac{\partial}{\partial t}u = -\frac{\partial}{\partial x}(cu) + \frac{\partial^2}{\partial x^2}(du)$$

où $(x,t) \in \mathbb{R}^2 \to c(x,t), d(x,t)$ sont les coefficients d'advection et de diffusion à préciser.
b) Déterminer la solution stationnaire dans le cas où $c(x,t) = -c.sign(x)$ et $d(x,t) = d$, où $c, d \in \mathbb{R}^+$, en précisant pourquoi ce cas est naturel pour la modélisation.
Note : Ici $sign(x) = -1, 0, 1$ suivant que x est strictement négatif, nul ou strictement positif.
c) Critiquer le modèle.

2) Nous considérons maintenant le cas où il y a deux groupes d'animaux qui ont chacun un territoire, avec un coefficient de diffusion commun constant $d \in \mathbb{R}_+$. La région considérée est le

segment $[0,1]$, le premier groupe a son foyer en 0 et sa densité de présence en x au temps t est donnée par $u(x,t)$, tandis que le deuxième a son foyer en 1 et sa densité de présence est donnée par $v(x,t)$.
Chaque groupe marque le territoire et réagit aux marques de l'autre espèce en tendant à revenir vers son foyer quand il les rencontre. On note $p(x,t)$ (resp. $q(x,t)$) la densité de marques en x au temps t du premier (resp. deuxième) groupe. Les animaux produisent des marques à taux 1 en l'absence de l'autre groupe, et le marquage croît linéairement en fonction des marques de l'autre groupe. Enfin les marques disparaissent à taux 1.

a) Justifiez que le modèle peut maintenant être décrit par

$$\begin{aligned}
\frac{\partial}{\partial t}u &= -c\frac{\partial}{\partial x}(qu) + d\frac{\partial^2}{\partial x^2}u \\
\frac{\partial}{\partial t}v &= c\frac{\partial}{\partial x}(pv) + d\frac{\partial^2}{\partial x^2}v \\
\frac{\partial}{\partial t}p &= u(1+mq) - p \\
\frac{\partial}{\partial t}q &= v(1+mp) - q
\end{aligned}$$

avec les conditions aux bords pour $x = 0, 1$

$$\frac{\partial}{\partial t}u = -cqu + d\frac{\partial}{\partial x}u, \quad \frac{\partial}{\partial t}v = cqv + d\frac{\partial}{\partial x}v$$

b) Déterminer la solution stationnaire quand $m = 0$ et tracer les courbes.
c) On suppose maintenant $m > 0$. En utilisant les conditions de conservation pour les densités, prouver que pour la solution stationnaire

$$(1+mu)(1+mv) = k$$

où k est un réel positif.
Expliquer alors la méthode pour trouver la solution stationnaire, conjecturer les courbes et faire les calculs pour $k = 2$.

Pour aller plus loin (avec des données...), *Mechanistic Home Range Analysis*, de Paul R. Moorcroft et Maura A. Lewis.

Pour s'entraîner

Exercice 6:
Modèle de Wright Fisher
Nous décrivons ici un modèle haploïde de reproduction aléatoire, sans mutation ni sélection ni immigration. On s'intéresse à une population génétique de $2N$ gènes pouvant être de deux types A et a. Le type de chaque gène à la génération $n+1$ est obtenu en prenant le type d'un individu pris au hasard (uniformément et avec remise) dans la population de gènes à la génération n.
On note X_n le nombre de gènes de type A à la génération n et on suppose $0 < X_0 < 2N$.

1) Donner la matrice de transition de la chaine de Markov $(X_n)_{n\in\mathbb{N}}$.

2) Calculer l'espérance et la variance de X_n.

3) Montrer que p.s. $(X_n)_{n\in\mathbb{N}}$ est absorbée en 0 ou en $2N$ et calculer la probabilité de chacun de ces deux événements.

Exercice 7:
Marche aléatoire sur un polygone. On indexe par $i = 1...N$ les sommets d'un polygone (dans

le sens des aiguilles d'une montre par exemple) et on considère la chaîne de Markov correspondant à la marche aléatoire sur ces sommets avec une probabilité p de sauter au sommet suivant, q de sauter au sommet précédent et r de ne pas bouger, avec $p+q+r=1$.

1. Quand la chaîne est-elle irréductible ? Quelle est alors sa loi invariante ?
2. On suppose $p, q > 0$. On rappelle que la période d'un état i est

$$\mathrm{PGCD}(n \in \mathbb{N} : P^n(i,i) > 0),$$

où P est la matrice de transition de la chaîne.

   (a) Retrouver le fait qu'ici la période ne dépend pas de l'état : on l'appelle la période de la chaîne. Pourquoi est-ce le cas pour une chaîne irréductible en général ?

   (b) Montrer que si $r > 0$, la chaîne est de période 1 (apériodique).

   (c) Montrer que si $r = 0$, la chaîne est de période 2 quand N pair et 1 quand N impair, sauf dans les cas dégérés où p ou q sont également nuls.

   (d) Discuter alors de la convergence en distribution de la chaîne.

Exercice 8:
La domestication du millet
On considère trois variétés de millet qui se différencient par un couple d'allèles : la variété domestique correspond à (a,a), la variété sauvage à (A,A) et la variété hybride à (A,a). Les variétés cultivées sont les variétés domestiques et hybrides qui sont sélectionnées par les cultivateurs car leurs grains restent mieux accrochés aux plants.
On cherche à comprendre comment cette caractéristique peut se fixer dans la population cultivée malgré les croisements constants avec la variété sauvage.
On note X_n la proportion de la variété domestique dans le champ et Y_n la proportion de la variété hybride.

1. La méthode de culture est la suivante. D'une part, afin d'éviter les reproductions entre hybrides pouvant conduire à la création de millet de type (A,A) (indésirable), le cultivateur ne tente de reproduire que des fleurs de type domestique (a,a). Pour ce qui est du pollen utilisé, le cultivateur commence par éliminer avant la floraison une proportion $v \in [0,1]$ d'hybrides, puis il récolte le pollen des plants de millet (domestiques et hybrides) restants. Ensuite,

   (a) une proportion α des fleurs à reproduire sont choisies dans les domestiques et subissent une auto-fécondation (elles se reproduisent avec du pollen issu de leur propre plant)

   (b) pour les autres, la reproduction se fait par allo-fécondation :

   (b.1) pour une proportion m avec la population sauvage

   (b.2) pour les autres, par panmixie parmi les plantes cultivées (c-à-d qu'on choisit au hasard une plante dans le champ).

   Montrer que

$$\begin{aligned} X_{n+1} &= \alpha + (1-\alpha)(1-m)\left(\frac{X_n}{1-vY_n} + \frac{1}{2}\frac{(1-v)Y_n}{1-vY_n}\right) \\ Y_{n+1} &= (1-\alpha)m + (1-\alpha)(1-m)\frac{1}{2}\frac{(1-v)Y_n}{1-vY_n}. \end{aligned}$$

   En déduire que Y_n converge vers l'unique point fixe donné par

$$Y_\infty = F(m, v, \alpha, Y_\infty)$$

   où

$$F(m,v,\alpha,y) := (1-\alpha)\left(m + \frac{(1-v)(1-m)y}{2(1-vy)}\right).$$

   Vérifier que Y_∞ est une fonction décroissante de α et de v, et une fonction croissante de m.

2. La population est maintenant supposée d'effectif constant égal à N et on est amené à introduire le modèle stochastique suivant. La composition des variétés de millet à partir d'une population contenant i hybrides est obtenue par échantillonnage aléatoire (avec remise). En procédant comme dans 1) aléatoirement pour chaque reproduction, on est amené à introduire la fréquence (probabilité)

$$p_i = p_i(N) = F\big(m, v, \alpha, \frac{i}{N}\big)$$

qui donne la probabilité d'obtenir un hybride pour chaque reproduction à chaque échantillonnage.

a) On note $U_n^N := NY_n^N$ le nombre d'hybrides à la génération n. Déterminer la matrice de transition de la chaîne de Markov (U_n^N).
b) Montrer que

$$\mathbb{E}(U_n^N | U_{n-1}^N = i) = Np_i.$$

En moyenne, on retrouve donc le modèle déterministe.

c) Montrer qu'indépendamment de l'état initial, la chaîne de Markov (U_n^N) converge vers une unique loi stationnaire.

d) Montrer que $\mathrm{Var}(Y_{n+1}^N - Y_n^N | Y_n^N = y)$ tend vers zéro quand $N \to \infty$.

e) En déduire que Y_n^N converge quand $N \to \infty$ vers la suite déterministe Y_n de la première partie.

f) Afin d'étudier les fluctuations de Y^N autour de sa limite Y, introduisons la quantité

$$V_n^N := \sqrt{N}\left(Y_n^N - Y_n\right).$$

Montrer que

$$\lim_{N\to\infty} \mathbb{E}\big(V_{n+1}^N - V_n^N | V_n^N = v\big) = -(1 - F'(Y_n))\, v$$

et que

$$\lim_{N\to\infty} \mathrm{Var}\big(V_{n+1}^N - V_n^N | V_n^N = v\big) = F(Y_n)(1 - F(Y_n)).$$

Ceci permet de montrer que le processus V_n^N est approximé quand $N \to \infty$ par la chaîne de Markov autorégressive $(V_n,\, n \geq 0)$ définie par

$$V_{n+1} = -(1 - F'(Y_n))V_n + E_n,$$

où E_n est indépendante de V_n et de loi normale $\mathcal{N}(0, F(Y_n)(1 - F(Y_n)))$.