

Modèles markoviens en biologie

Examen du 20 avril 2007¹

Règles du jeu. Les trois exercices sont indépendants. Vous pouvez admettre certains résultats donnés dans le sujet pour répondre aux questions suivantes. Le sujet est volontairement long, aussi, ne vous angoissez pas de n'en traiter qu'une partie. Il n'y a pas de pièges. Tous les documents sont autorisés. Vous savez tout, et vous avez le temps (**quatre heures**), alors bon travail!

Exercice 1 (Mort ou vif...). On souhaite comprendre le comportement d'une population de bactéries. Chaque individu meurt avec un taux $\mu \geq 0$ et se dédouble avec un taux $\lambda \geq 0$ indépendamment des autres. On désigne par X_t la taille de la population à l'instant $t \geq 0$, par $(X_t)_{t \geq 0}$ le processus d'espace d'état \mathbb{N} associé, et par $(P_t)_{t \geq 0}$ le semi-groupe de Markov correspondant.

1. Sachant que le processus est en i au temps t , qu'elle est la loi du temps de saut suivant et la loi spatiale de ce saut? En déduire que le générateur infinitésimal L du processus $(X_t)_{t \geq 0}$ est donné pour tous $i, j \in \mathbb{N}$ par

$$L(i, j) = \begin{cases} i\lambda & \text{si } j = i + 1, \\ i\mu & \text{si } j = i - 1, \\ -i(\lambda + \mu) & \text{si } j = i, \\ 0 & \text{sinon.} \end{cases}$$

2. Qu'elle est la nature de l'état 0? Quel est le comportement qualitatif de $(X_t)_{t \geq 0}$ lorsque $\mu = 0$ et $\lambda > 0$? Idem pour $\lambda = 0$ et $\mu > 0$? Qu'elle est l'allure des trajectoires du processus $(X_t)_{t \geq 0}$?
3. On souhaite étudier le cas spécial où les taux de duplication et de mort sont égaux. **On suppose donc dans toute la suite de l'exercice que $\lambda = \mu$.** On note f la fonction $n \mapsto n$ et g la fonction $n \mapsto n^2$. Exprimer Lf et Lg en fonction de f et g . En déduire que $t \mapsto P_t f$ et $t \mapsto P_t g$ sont solutions d'équations différentielles linéaires que l'on résoudra. En déduire que $t \mapsto \mathbb{E}(X_t)$ est constante et que $\text{Var}(X_t) = 2\lambda t \mathbb{E}(X_0) + \text{Var}(X_0)$ pour tout $t \geq 0$.

On considère à présent une suite $(\lambda_N)_{N \geq 1}$ de réels positifs. Pour tout entier $N \in \mathbb{N}^*$ fixé, on note $(X^N)_{t \geq 0}$ le processus décrit précédemment associé aux taux $\lambda = \mu = \lambda_N$. On désigne par Y^N le processus $(Y_t^N)_{t \geq 0} = (N^{-1} X_{Nt}^N)_{t \geq 0}$ à valeurs dans $N^{-1}\mathbb{N}$.

4. Déterminer le générateur infinitésimal A_N de Y^N ;
5. Soit $a > 0$ un réel. Donner une condition sous laquelle la suite de processus $(Y^N)_N$ converge vers le processus de diffusion Y sur \mathbb{R}_+ de générateur infinitésimal A donné pour toute fonction $f \in \mathcal{C}^2(\mathbb{R}_+, \mathbb{R})$ par $Af(y) = ayf''(y)$. De quelle équation différentielle stochastique Y est-il solution?
6. Montrer que $t \mapsto \mathbb{E}(Y_t)$ est constante tandis que $\text{Var}(Y_t) = at\mathbb{E}(X_0) + \text{Var}(Y_0)$ pour tout $t \geq 0$;
7. Soit I un intervalle de la forme $]l, r[\subset \mathbb{R}_*^+$ et soit $y \in I$. Calculer la probabilité que partant de y , le processus Y sorte de I en r . Quelle est la limite de cette probabilité quand $r \rightarrow +\infty$?

¹Il se trouve qu'il s'agit du vendredi précédant le premier tour des élections présidentielles françaises! Pour l'heure, il s'agit d'élire les étudiants qui obtiendront le module B2 du M2R. . . Que les candidats se rassurent, cette « élection » est beaucoup moins aléatoire que les présidentielles. . . Mais votez quand même dimanche!

Exercice 2 (Développement et déclin d'une population de plantes). On considère une espèce particulière de plantes qui fleurissent chaque année, puis **meurent après la floraison**. Chaque fleur produit des graines qui sont emportées par le vent. Les graines ont très peu de chance de trouver des circonstances favorables à leur germination puis à leur développement. Cependant, il arrive malgré tout que quelques unes parviennent à donner naissance l'année suivante à une nouvelle plante, situé à une distance aléatoire de leur défunte « mère ». Ce procédé se poursuit indéfiniment, au fil des années, et peut s'arrêter en cas d'extinction. On souhaite décrire la **répartition spatiale au fil des générations** des plantes **obtenues à partir d'une population initiale**. On appelle **plante fille** d'une plante donnée les plantes qu'elle est parvenue à produire l'année d'après. Pour simplifier, on fait les hypothèses suivantes :

- Les plantes sont situés sur le plan $E = \mathbb{R}^2$ muni de la mesure de Lebesgue dx ;
- Les plantes de la génération 0 sont positionnées selon un processus ponctuel de Poisson sur \mathbb{R}^2 de mesure d'intensité Λ_0 fixée vérifiant $\Lambda_0(\mathbb{R}^2) < \infty$;
- À la génération n , les plantes font des filles de manière indépendante puis meurent toutes. L'ensemble de ces plantes filles constitue la génération $n + 1$;
- Pour chaque plante, le nombre de filles est indépendant des positions spatiales des filles.

On modélise le nombre de plantes filles d'une plante donnée par une réalisation de loi de Poisson $P = \mathcal{P}(m)$. Soit S une loi de probabilité sur \mathbb{R}^2 . On modélise la position spatiale des plantes filles d'une plante située en $x \in \mathbb{R}^2$ par des réalisations i.i.d. de loi $\delta_x * S$. Les lois P et S sont fixées une bonne fois pour toutes et constituent des caractéristiques propres à l'espèce de plantes considérée.

1. Justifier l'usage de la loi $\delta_x * S$ plutôt que de la loi S . Expliquer également pourquoi la loi de Poisson P est appropriée à la situation ;
2. Pour tout entier $n \geq 0$, soit X_n la mesure ponctuelle aléatoire représentant la position spatiale des plantes de la génération n . Fixons $n \geq 0$. Conditionnellement à X_n , montrer que l'ensemble des positions spatiales des filles d'une plante de la génération n située en x (un atome de X_n) constituent un processus ponctuel de Poisson H_x sur \mathbb{R}^2 de mesure d'intensité $m(\delta_x * S)$. Qu'elle est la loi du nombre de filles de cette plante ? Le nombre moyen de filles de cette plante ?
3. On note encore X_n l'ensemble des positions spatiales des plantes de la génération n . On note enfin $|X_n|$ le nombre d'individus de la génération n . Expliquer pourquoi pour tout $n \geq 0$ fixé, et conditionnellement à X_n , on a $X_{n+1} = \emptyset$ si $|X_n| = 0$ et sinon $X_{n+1} = \sum_{x \in X_n} H_x$ où $(H_x)_{x \in X_n}$ constituent des processus ponctuels de Poisson indépendants, avec H_x comme dans la question précédente pour chaque x ;
4. Dédire de la question précédente que pour tout $n \geq 0$, conditionnellement à X_n , la mesure atomique aléatoire X_{n+1} constitue un processus ponctuel de Poisson sur \mathbb{R}^2 dont on précisera la mesure d'intensité. En déduire que le nombre d'individus de la génération n suit la loi de Poisson $\mathcal{P}(|X_n| m)$. En déduire l'évolution du nombre moyen de plantes au fil des générations, en fonction de la valeur de m ;
5. La probabilité d'extinction de la population de plantes est définie par $\mathbb{P}(\lim_{n \rightarrow \infty} |X_n| = 0)$. Montrer que $(|X_n|)_{n \geq 0}$ constitue un processus de branchement (de Galton-Watson) de loi de reproduction $\mathcal{P}(m)$ et de mesure initiale $|X_0| \sim \mathcal{P}(\Lambda_0(\mathbb{R}^2))$. En déduire le comportement de la suite $(|X_n|)_{n \geq 0}$ en fonction de m . **Passer à la question suivante en cas de difficulté** ;
6. Calculer la transformée de Laplace de X_{n+1} conditionnellement à X_n , et en déduire que X_n ne peut pas être un processus ponctuel de Poisson pour tout $n > 0$;
7. Peut-on remplacer le plan \mathbb{R}^2 par un espace mesurable plus raisonnable ? Les résultats obtenus dépendent-ils de la nature de cet espace ?
8. Considérons une plante donnée, et modélisons la position spatiale des graines qu'elle produit par un processus ponctuel de Poisson X sur \mathbb{R}^2 de mesure d'intensité Θ . Supposons que

conditionnellement à X , chaque graine (i.e. chaque atome de X) produit une plante indépendamment des autres, avec probabilité $p(x)$ où x désigne la position de la graine. Montrer que les plantes filles ainsi produites sont réparties dans l'espace selon un processus ponctuel de Poisson sur \mathbb{R}^2 dont on précisera la mesure d'intensité en fonction de Θ et p ;

- Supposons que le nombre de plantes filles d'une plante donnée dépende d'une part de la qualité de son positionnement spatial (nature du sol par exemple) ainsi que de l'année (intempéries, précipitations). Comment en tenir compte le plus simplement possible avec un modèle ?

Exercice 3 (Gréganisme et individualisme des oiseaux du pacifique). On considère une population d'oiseaux migrateurs qui, chaque année, traverse l'océan pacifique pour passer l'hiver dans l'archipel des Nouvelles-Hébrides. Au début de la période hivernale, le premier oiseau arrivé s'installe sur un îlot vide. Les oiseaux suivants ont le choix de rejoindre un îlot déjà occupé ou de s'installer sur un îlot inoccupé. Pour simplifier, on adopte les hypothèses suivantes :

- Cette espèce d'oiseaux ne change pas d'îlot une fois celui-ci choisi ;
- Le nombre d'îlots est si considérable qu'on suppose qu'il est infini ;
- Chaque îlot est suffisamment vaste pour héberger tous les oiseaux migrateurs.

On appelle *taille d'un îlot* le nombre d'oiseaux qu'il héberge. Cette taille ne peut qu'augmenter au fil du temps. On souhaite modéliser l'évolution de la taille et du nombre des îlots occupés au fil de l'arrivée des oiseaux. À l'instant $n \geq 1$, il y a n oiseaux déjà installés. Si ces oiseaux sont numérotés de 1 à n , les îlots qu'ils occupent sont modélisés par une **partition aléatoire** B_n de l'ensemble fini $\{1, \dots, n\}$ en $|B_n|$ îlots. À l'instant $n + 1$, un nouvel oiseau numéroté $n + 1$ arrive de sa longue traversée et décide **soit** de s'installer sur l'un des n îlots déjà occupés avec une probabilité proportionnelle au nombre d'occupants de l'îlot (gréganisme), **soit** de s'installer sur un îlot inoccupé (individualisme). On modélise par $\theta/(n + \theta)$ la probabilité qu'a l'oiseau $n + 1$ de s'installer sur un îlot inoccupé, où $\theta > 0$ est un paramètre réel fixé et inconnu (**le même pour tous les oiseaux**).

Soit \mathcal{B}_n l'ensemble des partitions de $\{1, \dots, n\}$. La suite $(B_n)_{n \geq 1}$ constitue un processus à temps discret tel que pour tout $n \geq 1$, B_n est une variable aléatoire à valeurs dans \mathcal{B}_n . La partition B_n est constituée de $|B_n|$ sous-ensembles non vides de $\{1, \dots, n\}$. On pose $B_1 = \{\{1\}\}$.

- Pour tout $B \in \mathcal{B}_n$, la notation $b \in B$ signifie que b est un sous-ensemble non vide de $\{1, \dots, n\}$ de cardinal $|b|$ faisant partie de la partition B . On dit que b est un *bloc* de taille $|b|$ de B . Montrer que pour tout $n \in \mathbb{N}^*$ et tout $B \in \mathcal{B}_n$,

$$\sum_{b \in B} \frac{|b|}{\theta + n} = \frac{n}{\theta + n};$$

- Pour tout $B \in \mathcal{B}_n$, tout $B' \in \mathcal{B}_{n+1}$, et tout $b \in B$, on note $B \rightarrow_b B'$ lorsque B' s'obtient à partir de B en remplaçant le bloc b par $b \cup \{n + 1\}$. Notons que si $B \rightarrow_b B'$ alors $|B| = |B'|$. Montrer que pour tout $n \in \mathbb{N}^*$ et tout $B \in \mathcal{B}_n$,

$$\mathbb{P}(B_{n+1} = B' \mid B_n = B) = \begin{cases} \frac{|b|}{\theta + n} & \text{si } B \rightarrow_b B', \\ \frac{\theta}{\theta + n} & \text{si } B' = B \cup \{\{n + 1\}\}, \\ 0 & \text{sinon;} \end{cases}$$

- $(B_n)_{n \geq 1}$ est-elle une chaîne de Markov (homogène?) d'espace d'état $\cup_{n=1}^{\infty} \mathcal{B}_n$?
- Que se passe-t-il dans le cas extrême où $\theta = 0$? Expliquer pourquoi plus θ est grand, plus les oiseaux ont tendance à s'installer sur des îlots non occupés ;

5. Établir que pour tout $n \geq 1$ et tout $B \in \mathcal{B}_n$,

$$\mathbb{P}(B_n = B) = \frac{\theta^{|B|}}{(\theta + 0)(\theta + 1) \cdots (\theta + n - 1)} \prod_{b \in B} (|b| - 1)!;$$

6. Soit B une partition de $\{1, \dots, n\}$ en blocs $b_1, \dots, b_{|B|}$ et a_1, \dots, a_n la suite des nombres de blocs de B de tailles respectives $1, \dots, n$. Montrer que $a_1 + 2a_2 + \dots + na_n = n$. Montrer que le nombre de partitions $B \in \mathcal{B}_n$ de taille $|B| = n$ à nombres a_1, \dots, a_n prescrits vaut

$$\frac{n!}{\prod_{k=1}^n (k!)^{a_k} a_k!};$$

7. Pour tout temps $n \geq 1$ et tout $1 \leq k \leq n$, soit $C_{n,k}$ le nombre d'îlots de taille k au temps n , de sorte que $n = C_{n,1} + 2C_{n,2} + \dots + nC_{n,n}$. L'entier $C_{n,k}$ correspond au nombre de blocs de taille k dans la partition aléatoire B_n . Dédire des deux questions précédentes que pour tous entiers $0 \leq a_1, \dots, a_n \leq n$ tels que $a_1 + 2a_2 + \dots + na_n = n$, on a

$$\mathbb{P}(C_{n,1} = a_1, \dots, C_{n,n} = a_n) = \frac{n!}{(\theta + 0)(\theta + 1) \cdots (\theta + n - 1)} \prod_{k=1}^n \frac{1}{a_k!} \left(\frac{\theta}{k}\right)^{a_k};$$

8. Pour tout $n \geq 1$, on pose $C_n = (C_{n,1}, \dots, C_{n,n})$ où $C_{n,k}$ est comme dans la question précédente. La suite de vecteurs aléatoires $(C_n)_{n \geq 1}$ est-elle une chaîne de Markov ?

9. Montrer que pour tout $n \geq 1$, $|B_n|$ s'écrit $|B_n| = \xi_1 + \dots + \xi_n$ où ξ_1, \dots, ξ_n sont des variables aléatoires indépendantes de lois de Bernoulli vérifiant pour tout $1 \leq k \leq n$

$$\mathbb{P}(\xi_k = 1) = 1 - \mathbb{P}(\xi_k = 0) = \frac{\theta}{\theta + k - 1},$$

et en déduire que pour tout $n \geq 1$, la moyenne et la variance de $|B_n|$ sont données par

$$\mathbb{E}(|B_n|) = \sum_{k=0}^{n-1} \frac{\theta}{\theta + k} \quad \text{et} \quad \text{Var}(|B_n|) = \sum_{k=1}^{n-1} \frac{\theta k}{(\theta + k)^2};$$

10. Qu'elle est l'allure de la suite $(|B_n|)_{n \geq 1}$. Est-elle une chaîne de Markov ? homogène ? Donner un équivalent de la moyenne et de la variance de $|B_n|$ lorsque $n \rightarrow \infty$;

11. Pour tout $n \geq 1$ fixé, on estime $\mathbb{E}(|B_n|)$ à partir d'un échantillon de la loi $\mathcal{L}(|B_n|)$. Dédire de la formule pour $\mathbb{E}(|B_n|)$ une équation vérifiée par un estimateur $\hat{\theta}$ de θ . Cette équation est-elle résoluble explicitement en $\hat{\theta}$?

12. Pour tout $n \geq 1$ fixé, proposer un estimateur de la quantité $\theta_n = (1 + n\theta^{-1})^{-1}$ à partir de l'expression de $\mathbb{E}(|B_{n+1}|) - \mathbb{E}(|B_n|)$. Préciser le biais et la variance de cet estimateur ;

13. On s'intéresse à présent au nombre $C_{n,1}$ qui représente le nombre d'îlots hébergeant un seul oiseau au temps n . Il s'agit du nombre de blocs de taille 1 dans la partition B_n . Montrer que pour tout $n \geq 1$, tout $0 \leq a \leq n$ et tout $0 \leq a' \leq n + 1$, on a

$$\mathbb{P}(C_{n+1,1} = a' | C_{n,1} = a) = \begin{cases} \frac{\theta}{\theta + n} & \text{si } a' = a + 1, \\ \frac{a}{\theta + n} & \text{si } a' = a - 1, \\ 1 - \frac{\theta}{\theta + n} - \frac{a}{\theta + n} & \text{si } a' = a, \\ 0 & \text{sinon;} \end{cases}$$

La suite $(C_{n,1})_{n \geq 1}$ est-elle une chaîne de Markov (homogène?) d'espace d'état \mathbb{N} ?

14. En déduire que pour tout $n \geq 1$,

$$\mathbb{E}(C_{n+1,1} | C_{n,1} = a) = \frac{a(\theta + n - 1) + \theta}{\theta + n}$$

puis que pour tout $n \geq 1$,

$$\mathbb{E}(C_{n,1}) = \frac{n\theta}{n + \theta - 1} \quad \text{et} \quad \text{Var}(C_{n,1}) = \frac{n(n-1)(n-2+2\theta)\theta}{(n+\theta-2)(n+\theta-1)^2};$$

15. En déduire qu'en temps n grand, il y a en moyenne un unique îlot occupé par un seul oiseau et que la variance du nombre d'îlots de ce type est égale au coefficient d'individualisme θ ;

16. On s'intéresse à présent au nombre $C_{n,n}$ d'îlots de taille n au temps n . Il s'agit du nombre de blocs de taille n dans la partition B_n . Montrer que $(C_{n,n})_{n \geq 1}$ est une chaîne de Markov et préciser son espace d'état. Montrer que la suite $(C_{n,n})_{n \geq 1}$ est presque-sûrement décroissante, et que pour tout $n \geq 2$,

$$\mathbb{P}(C_{n,n} = 1) = \prod_{k=1}^{n-1} \frac{k}{\theta + k};$$

17. En déduire que $(C_{n,n})_{n \geq 1}$ converge presque-sûrement vers 0 lorsque $n \rightarrow \infty$;

18. Le modèle $(B_n)_{n \geq 1}$ utilisé pour l'occupation des îlots par les oiseaux favorise-t-il l'apparition d'îlots peu occupés? le maintient de tels îlots? Proposer un modèle dans lequel les oiseaux déjà installés peuvent changer d'îlot. Expliquer quelques aspects qualitatifs de ce nouveau modèle par rapport au précédent;

19. Soit $(W_n)_{n \geq 1}$ une suite de v.a. réelles i.i.d. de loi Beta(1, θ) de densité par rapport à la mesure de Lebesgue $x \mapsto \theta(1-x)^{\theta-1} \mathbf{1}_{[0,1]}(x)$. On définit la suite de réels aléatoires $(P_n)_{n \geq 1}$ par

$$P_1 = W_1, \quad P_2 = (1 - W_1)W_2, \quad P_3 = (1 - W_1)(1 - W_2)W_3, \quad \dots$$

Établir que presque-sûrement, la mesure discrète $\sum_{n=1}^{\infty} P_n \delta_n$ constitue une loi de probabilité sur \mathbb{N}^* . On pourra exprimer $1 - P_1 - \dots - P_n$ en fonction de $1 - W_1, \dots, 1 - W_n$ puis utiliser le lemme de Borel-Cantelli;

On désigne par \mathcal{P}_θ la loi de la mesure atomique aléatoire $\sum_{n=1}^{\infty} P_n \delta_n$. Fixons $n \geq 1$, puis tirons une réalisation P de loi \mathcal{P}_θ , puis, conditionnellement à ce tirage, tirons n réalisations i.i.d. M_1, \dots, M_n de loi P . Cette expérience fournit donc une suite de n entiers aléatoires dans \mathbb{N}^* . Pour tout $k \geq 1$, on note $N_{n,k}$ le nombre d'entiers apparaissant exactement k fois dans cette suite. Il se trouve que $(N_k)_{1 \leq k \leq n}$ suit la même loi que $(C_{n,k})_{1 \leq k \leq n}$. Mais ce résultat dépasse le cadre de cet examen... Auriez-vous toutefois une idée de la manière de procéder?

— fin du sujet —