Contrôle de Modèles et algorithmes stochastiques 1 et 2

Durée 1h30

Les documents, les calculatrices et les téléphones portables sont interdits.
Le barème sur 20 est approximatif.

Exercice 1 (Variante du modèle de Moran avec mutations) (12 points)
On considère une population de taille constante $N \geq 2$, pour laquelle la transmission des gènes d'une génération à une autre se fait suivant une version modifiée du modèle de Moran, qui peut être décrite comme suit.
À chaque génération, on choisit au hasard deux individus distincts de la population. Le premier individu se réproduit, en donnant naissance à exactement un enfant, et continue lui-même de vivre, alors que le second individu meurt. Tous les autres individus continuent de vivre dans la génération suivante.

1. On considère un gène qui peut s'exprimer sous la forme de deux allèles différentes, A et a. On suppose qu'à chaque événement de réproduction il y a une probabilité $\theta \in]0,1[$ de mutation d'un allèle vers l'autre allèle. Plus précisément, si le parent est de type A (respectivement a), alors il y a une probabilité θ pour que l'enfant soit de type a (respectivement A).
   Pour $n \geq 0$, on appelle X_n le nombre d'individus portant l'allèle A dans la population à la génération n. On note $P = (p_{ij})_{0 \leq i,j \leq N}$ la matrice de transition de la chaîne de Markov $(X_n)_{n \geq 0}$.
   (a) Pour tout $1 \leq i \leq N-1$, donner $p_{i,i+1}$, $p_{i,i-1}$ et $p_{i,i}$. Donner ensuite toutes les autres entrées de la matrice P.
   (b) Montrer que la chaîne $(X_n)_n$ admet une unique probabilité invariante $\pi > 0$ sur $\{0, \dots, N\}$. Montrer ensuite que, pour toute loi initiale, la chaîne $(X_n)_n$ converge en loi vers π quand $n \to \infty$.
   (c) Montrer que pour tout $n \in \mathbb{N}$ et pour tout $i = 0, \dots, N$ on a
   $$\mathbb{E}(X_{n+1} | X_n = i) = \left(1 - \frac{2\theta}{N}\right) i + \theta.$$
   (d) Déterminer une relation de récurrence reliant $\mathbb{E}(X_{n+1})$ et $\mathbb{E}(X_n)$, pour tout $n \in \mathbb{N}$.
   (e) Pour tout $n \in \mathbb{N}$, montrer la relation suivante :
   $$\mathbb{E}(X_n) = \left(1 - \frac{2\theta}{N}\right)^n \mathbb{E}(X_0) + \frac{N}{2}\left\{1 - \left(1 - \frac{2\theta}{N}\right)^n\right\}.$$
   (*Indication* : vous pourriez faire une démonstration par récurrence.)
   (f) Déterminer la moyenne de la probabilité invariante π. Pourriez-vous justifier d'une autre façon le résultat obtenu ?
2. On s'intéresse maintenant à l'arbre généalogique d'un échantillon de k individus de la population à l'instant présent, avec $2 \leq k \leq N$. On note τ_k l'instant du premier événement de coalescence quand on rémonte le temps, i.e. le nombre de générations qu'il faut rémonter en arrière, à partir de l'instant présent, jusqu'à ce que deux individus parmi les k trouvent un ancêtre commun.
   (a) Déterminer la loi de la variable aléatoire τ_k.
   (b) Déterminer le nombre moyen de générations qu'il faut rémonter en arrière pour trouver le plus récent ancêtre commun de tous les k individus de l'échantillon.

Exercice 2 (Prédiction généralisée) (2 points)
On considère une chaîne de Markov cachée (X_n, Y_n) à valeurs dans $E \times F$ fini de noyau de transition P et noyau d'observation Q. Montrer que pour tout $n \geq 0$, pour tout $k \geq 1$,

$$\mathbb{P}(X_{n+k} = x_{n+k} | Y_{0:n} = y_{0:n}) = \sum_{z \in E} P^k(z, x_{n+k}) \Pi_n(y_{0:n}, z)$$

où

$$\Pi_n(y_{0:n}, z) = \mathbb{P}(X_n = z | Y_{0:n} = y_{0:n}).$$

Exercice 3 (Distance entre particules) (6 points)
On considère $(S_n)_{n\geq 0}$ une marche aléatoire sur $\mathbb{Z}$. Plus précisément, on suppose que S_0 suit la loi uniforme sur $\{-1, 0, 1\}$ ($\mathbb{P}(S_0 = -1) = \mathbb{P}(S_0 = 0) = \mathbb{P}(S_0 = 1) = 1/3$) et que pour tout $n \geq 0$, $S_{n+1} = S_n + U_{n+1}$ où $(U_n)_{n\geq 1}$ est une suite de variables aléatoires i.i.d. à valeurs dans $\{-1, 1\}$. On note $p = \mathbb{P}(U_1 = 1)$ et $q = \mathbb{P}(U_1 = -1)$ ($p + q = 1$).
On note ensuite $(\tilde{S}_n)_{n\geq 0}$ une marche aléatoire indépendante de $(S_n)_{n\geq 0}$ et de même loi. On s'intéresse à la distance algébrique entre S_n et $\tilde{S}_n$. On note donc $X_n = S_n - \tilde{S}_n$. $(X_n)_{n\geq 0}$ est une chaîne de Markov sur $\mathbb{Z}$. On note P sa "matrice" de transition.

1. (a) Déterminer la loi initiale ν de la chaîne $(X_n)_{n\geq 0}$. (On pourra par exemple montrer que $\nu(\{0\}) = 1/3$).
   (b) Calculer $P_{i,j}$ pour tout $i, j \in \mathbb{Z}$.
2. On suppose dans la suite que l'on a une observation bruitée des positions des marches et que l'on cherche à estimer la distance algébrique X_n. Plus précisément, on pose $V_n = S_n + \varepsilon_n$ et $\tilde{V}_n = \tilde{S}_n + \tilde{\varepsilon}_n$ où (ε_n) et $(\tilde{\varepsilon}_n)$ sont deux suites indépendantes et de même loi de variables aléatoires i.i.d à valeurs dans $\{-1, 0, 1\}$. On suppose que $\mathbb{P}(\varepsilon_1 = 1) = \mathbb{P}(\varepsilon_1 = -1) = \delta \in (0, 1/2)$ et que $\mathbb{P}(\varepsilon_1 = 0) = 1 - 2\delta$. En vue de l'estimation de la distance, on conserve la donnée de $Y_n = V_n - \tilde{V}_n$.
   (a) On note Q, le noyau d'observation ($Q_{i,j} = \mathbb{P}(Y_n = j | X_n = i)$). Calculer $Q_{i,j}$ pour tout i, j.
   (b) Calculer $\mathbb{P}(X_0 = 0 | Y_0 = 0)$ en fonction de P, Q et ν.
   (c) On souhaite dans cette dernière question étudier la probabilité que la trajectoire observée corresponde à la trajectoire réelle. On suppose dans cette question connues les probabilités de prédiction $\Pi_{n/n-1}(y_{0:n-1}, z)$ pour tout z et $n \geq 1$. A l'aide de ces quantités et de P et Q, proposer un algorithme récursif permettant de calculer $\mathbb{P}(X_{0:n} = y_{0:n} | Y_{0:n} = y_{0:n})$.