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 (13 points)
On considère une population de taille constante N, pour laquelle la transmission des gènes d'une génération à une autre se fait suivant le modèle de Moran, qui peut être décrit comme suit.
À chaque génération, deux individus de la population sont choisis au hasard, avec remise.

- Si les deux individus choisis sont distincts, alors le premier individu a deux descendants, le second individu n'a aucun descendant, et tous les autres individus ont exactement un descendant dans la génération suivante.
- Si les individus choisis sont les mêmes, alors tous les individus de la population ont un seul descendant 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 que l'allèle A a un avantage sélectif $s > 0$ sur l'allèle a, dans le sens suivant : à chaque génération, quand on choisit l'individu destiné à avoir deux descendants, chaque individu portant l'allèle A a une probabilité proportionnelle à $1 + s$ d'être choisi, alors que chaque individu portant l'allèle a a une probabilité proportionnelle à 1 d'être choisi.
   Pour $n \geq 0$, soit 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 $p_{i,i+1}$ et $p_{i,i-1}$ et montrer ensuite la relation :
   $$p_{i,i+1} = (1 + s)p_{i,i-1}.$$
   (b) Donner la matrice de transition P. Y-a-t-il des états absorbants ?
   (c) Montrer que $(X_n)_{n\geq 0}$ est une sous-martingale par rapport à la filtration naturelle.
   (d) Montrer que $(X_n)_n$ converge presque surement, quand $n \to \infty$, vers une variable aléatoire X_∞. Quelles valeurs peut prendre X_∞ ?

   On cherche à déterminer $q_i := \mathbb{P}(X_\infty = N \mid X_0 = i)$, la probabilité de fixation de l'allèle A sachant que dans la population initiale on a i individus qui portent l'allèle A.
   (e) Donner q_0 et q_N. Montrer ensuite que pour tout $i = 1, ..., N - 1$ on a une relation de récurrence du type :
   $$q_i = a_i q_{i-1} + b_i q_i + c_i q_{i+1},$$
   avec des suites a_i, b_i et c_i qu'il faudra préciser.
   (f) Montrer que pour tout $i = 1, ..., N - 1$ on a : $q_{i+1} - q_i = (1 + s)^{-1}(q_i - q_{i-1})$.
   (g) En déduire que pour tout $i = 0, ..., N$ on a
   $$q_i = \frac{1 - (1 + s)^{-i}}{1 - (1 + s)^{-N}}.$$

2. On s'intéresse maintenant à l'arbre généalogique d'un échantillon de $m \leq N$ individus de la population présente. On note T'_m l'instant du premier événement de coalescence dans l'arbre généalogique des m individus, i.e. le nombre de générations qu'il faut rémonter en arrière jusqu'à ce que deux individus parmi les m trouvent leur ancêtre commun.
   (a) Montrer que la probabilité que deux individus donnés aient le même parent dans la génération précédente est égale à $\frac{2}{N^2}$. Donner ensuite la loi de T'_2.

(b) Pour m quelconque, déterminer la probabilité que, dans un échantillon de m individus, il y ait deux individus avec le même parent dans la génération précédente. Donner ensuite la loi de T'_m.

(c) Montrer que, pour m fixé, si la taille de la population $N \to \infty$, alors $\frac{2T'_m}{N^2}$ converge en loi vers une variable aléatoire T_m de loi exponentielle dont on précisera le paramètre.

(d) Si on mesure le temps en unités de $\frac{N^2}{2}$ générations et la taille de la population est très grande ($N \to \infty$), déterminer la hauteur moyenne de l'arbre généalogique de m individus.

Exercice 2 (Observations dépendant du futur) (7 points)
On considère $(X_n)_{n\geq 0}$ une chaîne de Markov à valeurs dans E fini de matrice de transition P et de loi initiale ν. On suppose qu'à l'instant n, on observe une variable aléatoire Y_n définie par :

$$Y_n = \frac{(X_n + X_{n+1})U_{n+1}}{2} \quad \forall n \in \mathbb{N}.$$

$(U_n)_{n\geq 1}$ est une suite indépendante de la suite (X_n) de variables aléatoires *i.i.d.* On suppose que U_1 est à valeurs dans $\{-1, 0, 1\}$ et que $\mathbb{P}(U_1 = 1) = \mathbb{P}(U_1 = -1) = p \in (0, 1/2)$.

1. Déterminer le noyau d'observation Q associé à ce modèle, *i.e.* déterminer pour tous x, x',

$$Q(x, x', y) = \mathbb{P}(Y_n = y | X_n = x, X_{n+1} = x')$$

en fonction de y.

2. Montrer que, conditionnellement à X_n et X_{n+1}, Y_n est indépendant de $Y_{0:n-1}$, *i.e.* montrer que pour tous x, x', y et $y_{0:n-1}$,

$$\mathbb{P}(Y_n = y | X_n = x, X_{n+1} = x') = \mathbb{P}(Y_n = y | X_n = x, X_{n+1} = x', Y_{0:n-1} = y_{0:n-1}).$$

3. Soit $n \geq 1$. On cherche à prédire la valeur de X_n connaissant les observations $Y_0, \ldots, Y_{n-1}$: calculer

$$\pi_{n/n-1}(y_{0:n}, x) = \mathbb{P}(X_n = x | Y_{0:n} = y_{0:n})$$

en fonction de P, Q et

$$\pi_{n-1}(y_{0:n}, .) = \mathbb{P}(X_{n-1} = . | Y_{0:n-1} = y_{0:n-1}).$$

4. Exprimer $\mathbb{P}(Y_0 = y_0)$ en fonction de P, Q et ν. En déduire $\pi_0(y_0, x)$.
5. Pour $n \geq 1$, calculer $\pi_n(y_{0:n}, x)$ en fonction de P, Q et $\pi_{n/n-1}$.