Devoir Surveillé 6 - BCPST 2

durée : 2 heures

Documents et calculatrices non autorisés.

Aucune réponse non justifiée ne sera prise en compte dans la notation. Il sera tenu compte de la clarté de la rédaction.

Ce sujet a pour objectif l'étude de quelques propriétés mathématiques de deux modèles en génétique des populations : le modèle d'évolution de Wright-Fisher et l'équilibre de Hardy-Weinberg.
Les deux parties du problème sont indépendantes.

Partie I : Modèle de Wright-Fisher

Soit $N \in \mathbf{N}^\star$. On considère un variant génétique biallélique dont les allèles sont notés A et a. Dans le cadre du modèle de Wright-Fisher, pour tout entier naturel n, on note X_n la variable aléatoire égale au nombre d'allèles de type A à la génération n dans une population finie de taille N.
On se donne une variable aléatoire X_0 à valeurs dans $[\![0, 2N]\!]$ et on considère que la suite $(X_n)_{n\in\mathbf{N}}$ de variables aléatoires à valeurs dans $[\![0, 2N]\!]$ vérifie pour tout entier n :

$$\forall (i,j) \in [\![0, 2N]\!]^2, \quad \mathrm{P}\left(X_{n+1} = j \,|X_n = i\,\right) = \binom{2N}{j}\left(\frac{i}{2N}\right)^j\left(1 - \frac{i}{2N}\right)^{2N-j}.$$

On rappelle que la notation $P(A|B)$ désigne $P_B(A)$.

1) Étude d'un cas particulier

On suppose ici que $N = 1$ et on note, pour tout entier n, $V_n = \begin{pmatrix} \mathrm{P}(X_n = 0) \\ \mathrm{P}(X_n = 1) \\ \mathrm{P}(X_n = 2) \end{pmatrix}$.

1. En utilisant la formule des probabilités totales, montrer que pour tout entier n, on a $V_{n+1} = MV_n$ où

$$M = \begin{pmatrix} 1 & 1/4 & 0 \\ 0 & 1/2 & 0 \\ 0 & 1/4 & 1 \end{pmatrix}.$$

   On rappelle que par convention $0^0 = 1$.

2. Prouver que M est diagonalisable et déterminer une matrice inversible P et une matrice diagonale D telles que

$$M = PDP^{-1}.$$

3. Pour tout entier naturel n, exprimer V_n à l'aide de M^n et V_0.
4. On admet que pour tout entier naturel n, $M^n = \begin{pmatrix} 1 & \frac{1}{2} - \frac{1}{2^{n+1}} & 0 \\ 0 & \frac{1}{2^n} & 0 \\ 0 & \frac{1}{2} - \frac{1}{2^{n+1}} & 1 \end{pmatrix}$. En déduire que :
   (a) Pour tout entier n, on a $E(X_n) = E(X_0)$. Interpréter ce résultat en termes de nombre d'allèles.
   (b) $\lim\limits_{n\to\infty} \mathrm{P}\left((X_n = 0) \cup (X_n = 2)\right) = 1$. Interpréter ce résultat en termes de génotypes.

2) Cas général

On suppose désormais que $N \geqslant 1$ et on cherche à montrer que les résultats de la question **4.** sont encore vérifiés.

5. (a) Soit $i \in [\![0, 2N]\!]$. On note

$$S_i = \sum_{j=0}^{2N} j\binom{2N}{j}\left(\frac{i}{2N}\right)^j\left(1 - \frac{i}{2N}\right)^{2N-j}$$

   En donnant une interprétation probabiliste de S_i, justifier que $S_i = i$.
   (b) Soit $n \in \mathbf{N}$ et $j \in [\![0, 2N]\!]$. Écrire $\mathrm{P}(X_{n+1} = j)$ sous forme d'une somme qu'on ne cherchera pas à calculer.
   (c) Soit $n \in \mathbf{N}$. Déduire des questions précédentes que l'on a $E(X_{n+1}) = E(X_n)$.
6. On considère la suite u de terme général $u_n = \mathrm{P}\left(H_n\right)$ où $H_n = (X_n = 0) \cup (X_n = 2N)$.
   (a) Soit $n \in \mathbf{N}$. Calculer $\mathrm{P}(H_{n+1}|X_n = 0)$ et $\mathrm{P}(H_{n+1}|X_n = 2N)$.

(b) Soit $n \in \mathbf{N}$ et $k \in [\![1, 2N-1]\!]$. Montrer que $\mathrm{P}\left(H_{n+1} | X_n = k\right) \geqslant 2\left(\dfrac{1}{2N}\right)^{2N}$.

(c) En utilisant la formule des probabilités totales, montrer alors que pour tout entier n, on a

$$u_{n+1} \geqslant u_n + 2(1-u_n)\left(\frac{1}{2N}\right)^{2N}.$$

(d) Soit $\alpha \in]0,1[$. On considère la suite w définie par :

$$w_0 = u_0 \quad \text{et} \quad \forall n \in \mathbf{N}, \quad w_{n+1} = w_n + \alpha(1-w_n) = (1-\alpha)w_n + \alpha.$$

Justifier que la suite w est convergente et donner sa limite.

(e) Montrer que, pour un α à déterminer : $\forall\, n \in \mathbf{N}, w_n \leqslant u_n \leqslant 1$.

(f) En déduire la limite de la suite u et conclure.

Partie II : Vers l'équilibre de Hardy-Weinberg

Soit $N \in \mathbf{N}^\star$. On considère une population de N individus. Chaque individu est de type AA (type 1), aa (type 2), ou Aa (type 3) en fonction de son génotype à un locus génétique donné.
Pour $k \in [\![1, N]\!]$, on note T_k la variable aléatoire égale au numéro du type (1, 2 ou 3) de l'individu k. On suppose que les variables aléatoires $T_1, \ldots, T_N$ sont mutuellement indépendantes et de même loi donnée par :

$$\forall\, k \in [\![1, N]\!], \quad T_k(\Omega) = \{1,2,3\} \quad \text{et} \quad P(T_k = 1) = p_1, \quad P(T_k = 2) = p_2, \quad P(T_k = 3) = p_3.$$

où p_1, p_2, p_3 sont trois réels strictement positifs tels que $p_1 + p_2 + p_3 = 1$.
Pour tout $i \in [\![1, 3]\!]$, on note N_i la variable aléatoire égale au nombre d'individus de type i. On a donc $N_1 + N_2 + N_3 = N$.

7. (a) Déterminer la loi de N_1.
   (b) Donner, sans justification, l'espérance et la variance de la variable N_1.
   (c) Déterminer la variance de $N_1 + N_2$. En déduire $\mathrm{Cov}(N_1, N_2)$.
8. (a) Soit $a \in \mathbf{R}$. Exprimer la quantité :

$$\lim_{N \to \infty} P\left(-a \leqslant \frac{N_1 - Np_1}{\sqrt{Np_1(1-p_1)}} \leqslant a\right)$$

   à l'aide de la fonction de répartition Φ d'une variable qui suit la loi normale centrée réduite.
   (b) Déterminer alors un intervalle I_N tel que

$$\lim_{N \to +\infty} \mathrm{P}(N_1 \in I) = 0,95.$$

   On rappelle que $\Phi(1,96) = 0,975$.
9. On considère la matrice

$$W = \begin{pmatrix} V(N_1) & \mathrm{Cov}\,(N_1, N_2) \\ \mathrm{Cov}\,(N_2, N_1) & V(N_2) \end{pmatrix}.$$

   (a) Justifier que W est une matrice diagonalisable.
   (b) Soit $(a, b) \in \mathbf{R}^2$. Prouver que :

$$V(aN_1 + bN_2) = \begin{pmatrix} a & b \end{pmatrix} \times W \times \begin{pmatrix} a \\ b \end{pmatrix}$$

   où l'on a identifié $\mathscr{M}_1(\mathbf{R})$ à $\mathbf{R}$.
   (c) En déduire que les valeurs propres de W sont strictement positives.
   On pourra considérer λ une valeur propre et $\begin{pmatrix} a \\ b \end{pmatrix}$ un vecteur propre associé à λ.
   (d) En déduire l'existence de deux matrices P et D appartenant à $\mathscr{M}_2(\mathbf{R})$ telles que :
   - D soit diagonale et inversible
   - P soit inversible et d'inverse $P^\top$
   - $W = PD^2P^{-1}$.
10. On note $A = D^{-1}P^{-1}$ et on considère les variables aléatoires Y_1, Y_2 telles que :

$$\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} = A \times \begin{pmatrix} N_1 - Np_1 \\ N_2 - Np_2 \end{pmatrix}.$$

   On admet que l'on a :

$$\begin{pmatrix} V(Y_1) & \mathrm{Cov}\,(Y_1, Y_2) \\ \mathrm{Cov}\,(Y_2, Y_1) & V(Y_2) \end{pmatrix} = AWA^\top.$$

   En déduire $V(Y_1)$, $V(Y_2)$, et $\mathrm{Cov}\,(Y_1, Y_2)$.