Générations de petits pois-Modèle de Wright Fisher

On cultive des petits pois (N plants) numérotés de 1 à N. De chaque génération de rang k on conserve N petits pois pour cultiver la génération suivante de rang $k+1$. Les N plants d'une génération sont supposés produire le même nombre de petits pois et ont la même chance d'être sélectionnés pour engendrer la génération suivante. Ainsi, par exemple, le plant de numéro 1 peut être le géniteur de tous les plants de la génération suivante ou bien ces derniers peuvent provenir chacun de plants différents.

Etant donné un plant d'une génération k donné, on nommera ascendant de rang 1 le plant de la génération $k-1$ l'ayant engendré, ascendant de rang 2 le plant de la génération $k-2$ ayant engendré ce dernier et ainsi de suite. On supposera le nombre d'ascendant infini.

1) On s'intéresse à une génération k donnée et dans cette génération, à deux plants donnés $P_{1.}$ et $P_{2.}$, par exemple, pour fixer les idées, aux plants de numéro 1 et 2.

Soit les familles de variables aléatoires $(X_n)_{n\in\mathbb{N}^*}, (Y_n)_{n\in\mathbb{N}^*}$ définies par :

$X_n =$ numéro du plant ascendant de rang n (donc dans la génération $k-n$) de $P_{1.}$

$Y_n =$ numéro du plant ascendant de rang n de $P_{2.}$

a) calculer :

$$P(X_n = i \cap Y_n = j)$$

$$P_{(X_n=1 \cap Y_n=2)}(X_{n+1} = i \cap Y_{n+1} = j)$$

$$P_{(X_n=1 \cap Y_n=1)}(X_{n+1} = i \cap Y_{n+1} = j)$$

Que dire quant à l'indépendance des évènements $(X_n = r \cap Y_n = s)$ et

$(X_{n+1} = i \cap Y_{n+1} = j)$ pour (i,j,r,s) dans $\{1,2,\dots,N\}^4$?

Déterminer pour tout couple (i,j) de $\{1,2,\dots,N\}^2$, pour un évènement B formé d'une intersection non vide d'évènements de type $(X_t = r \cap Y_t = s)$ avec $1 \leq t \leq n$ et $r \neq s$:

$$P_B(X_{n+1} = i \cap Y_{n+1} = j)$$

b) Déterminer la probabilité de l'évènement $S_{n,i,j} =$ "P_1 a pour ascendant de rang n le plant i et P_2 le plant j"

c) En déduire celle de l'évènement $S_n =$ "P_1 et P_2 ont même ascendant de rang n" puis la loi de la variable aléatoire $T =$ le rang de la génération dans laquelle P_1 et P_2 ont un premier ascendant commun.

2) On s'intéresse maintenant dans la génération k précédente à trois plants donnés P_1, $P_{2.}$, P_3, par exemple, pour fixer les idées, aux plants de numéro 1, 2 et 3.

a) Déterminer la probabilité de l'évènement $S_{n,i,j,k} =$ "P_1 a pour ascendant de rang n le plant , P_2 le plant j et P_3 le plant k"

b) En déduire la probabilité des évènements suivants :

$S'_n =$ "P_1, P_2 et P_3ont un même ascendant de rang n"

$S''_n =$ "Parmi P_1, P_2et P_3 deux ont le même ascendant de rang n mais pas les trois "

$S'''_n =$ "P_1, P_2 et P_3ont des ascendants de rang n distincts 2 à 2"

c) Soit V la variable aléatoire égale au rang de la génération pour laquelle au moins deux plants parmi P_1, P_2 et P_3 ont un même ascendant. Déterminer la loi de V

d) Soit U la variable aléatoire égale au rang de la génération pour laquelle les trois plants P_1, P_2 et P_3 ont un même ascendant et $W = U - V$. Déterminer pour $n \in \mathbb{N}^*$:

$$P_{V=n}(W = 0)$$

$$P_{V=n}(W > m)$$

En déduire la loi de W

3) On s'intéresse maintenant aux descendants d'un plant de la génération k par exemple le plant P_1 de numéro 1.

a) Soit la variable $D =$ nombre de descendants de P_1 dans la génération $k+1$. Déterminer la loi de D.

b) Déterminer la probabilité de l'évènement $A =$ "P_1 n'a aucun descendant"

c) En déduire le nombre moyen de plants de la génération k n'ayant aucun descendant dans la génération $k+1$

4) On s'intéresse maintenant à r plants $P_1, P_2, \dots, P_r$ de la génération k par exemple ceux de numéros $1,2, \dots r$. Soit $T_r =$ le rang de la première génération ascendante telle que parmi $P_1, P_2, \dots, P_r$ deux au moins aient même ascendant. Déterminer la loi de T_r, son espérance et sa variance.

Solution

1)

a) La sélection au hasard parmi la production identique des N plants d'une génération de N petits pois fait que les évènements $(X_n = i \cap Y_n = j)$ qui sont au nombre de N^2 sont équiprobables donc :

$$P(X_n = i \cap Y_n = j) = \frac{1}{N^2}$$

Par un raisonnement analogue :

$$P_{(X_n=1 \cap Y_n=2)}(X_{n+1} = i \cap Y_{n+1} = j) = \frac{1}{N^2}$$

En revanche pour $i \neq j$:

$$P_{(X_n=1 \cap Y_n=1)}(X_{n+1} = i \cap Y_{n+1} = j) = 0$$

Car si P_1 et P_2 ont même ascendant à la n -ième génération, ils ont forcément même ascendant à la $(n+1)$ −ième génération et

$$P_{(X_n=1 \cap Y_n=1)}(X_{n+1} = i \cap Y_{n+1} = i) = \frac{1}{N}$$

$(X_n = r \cap Y_n = s)$ et $(X_{n+1} = i \cap Y_{n+1} = j)$ sont indépendants si $r \neq s$ et ne le sont pas sinon.

$$P_B(X_{n+1} = i \cap Y_{n+1} = j) = \frac{1}{N^2}$$

b) On a :

$$S_{n,i,j} = (X_n = i \cap Y_n = j)$$

$$P(S_{n,i,j}) = \frac{1}{N^2}$$

c) On a

$$S_n = \cup_{1\le i\le N}\ S_{n,i,i}$$

$$P(S_n) = \sum_{i=1}^{N} P\left(S_{n,i,i}\right) = N \times \frac{1}{N^2} = \frac{1}{N}$$

on en déduit pour $n \in \mathbb{N}^*$:

$$P(T = n) = P\left(\overline{S_1} \cap \overline{S_2} .. \cap \overline{S_{n-1}} \cap S_n\right) = P\left(\overline{S_1}\right) P\left(\overline{S_2}\right) \dots P\left(\overline{S_{n-1}}\right) P(S_n)$$

$$= \left(1 - \frac{1}{N}\right)^{N-1} \times \frac{1}{N}$$

En posant :

$$q = 1 - \frac{1}{N}$$

On a :

$$\boxed{P(T = n) = q^{N-1}\,(1 - q)}$$

Donc T suit une loi géométrique de paramètre q

2)

a) De façon analogue au cas précédent on a :

$$P\left(S_{n,i,j,k}\right) = \frac{1}{N^3}$$

b) On en déduit :

$$P(S'_n) = \sum_{i=1}^{N} P\left(S_{n,i,i,i}\right) = N \times \frac{1}{N^3} = \frac{1}{N^2}$$

Et notant $\mathcal{A}_N$ l'ensemble des arrangements de trois éléments pris dans un ensemble à N éléments

$$P(S'''_n) = \sum_{(i,j,k)\in\mathcal{A}_N}^{N} P\left(S_{n,i,j,k}\right) = A_N^3 \times \frac{1}{N^3} = \frac{N\,(N-1)\,(N-2)}{N^3}$$

$$= \left(1 - \frac{1}{N}\right)\left(1 - \frac{2}{N}\right)$$

Et :

$$P(S''_n) = 1 - P(S'_n) - P(S'''_n) = 1 - \frac{1}{N^2} - \left(1 - \frac{1}{N}\right)\left(1 - \frac{2}{N}\right)$$

$$= \frac{3}{N} - \frac{3}{N^2} = \frac{3\,N - 3}{N^2}$$

c) Posons :

$$\boxed{p = \left(1 - \frac{1}{N}\right)\left(1 - \frac{2}{N}\right)}$$

On a :

$$P(V = n) = P\left(S'''_1 \cap S'''_2 .. \cap S'''_{n-1} \cap \overline{S'''_n}\right) = P(S'''_1)\,P(S'''_2) \dots P(S'''_{n-1})\,P\left(\overline{S'''_n}\right)$$

$$= p^{n-1} \times (1 - p)$$

Donc V suit une loi géométrique de paramètre p avec :

$$1 - p = 1 - \left(1 - \frac{1}{N}\right)\left(1 - \frac{2}{N}\right) = \frac{3}{N} - \frac{2}{N^2}$$

d) On a :

$$P_{V=n}(W = 0) = P_{V=n}(U = n) = \frac{P\big((U = n) \cap (V = n)\big)}{P(V = n)}$$

$$= \frac{P(S'''_1 \cap S'''_2 .. \cap S'''_{n-1} \cap S'_n)}{P(V = n)}$$

$$= \frac{p^{n-1} \times \frac{1}{N^2}}{p^{n-1} \times (1 - p)} = \frac{1}{N^2\left(\frac{3}{N} - \frac{2}{N^2}\right)} = \frac{1}{3\,N - 2}$$

Et pour $m \in \mathbb{N}$:

$$P_{V=n}(W > m) = \frac{P\big((U > n + m) \cap (V = n)\big)}{P(V = n)}$$

$$= \frac{P(S'''_1 \cap S'''_2 .. \cap S'''_{n-1} \cap S''_n \cap S''_{n+1} \dots \cap S''_{n+m})}{P(V = n)}$$

Soit en posant :

$$\boxed{s = \frac{3\,N - 3}{N^2}}$$

$$= \frac{p^{n-1} \times s^{m+1}}{p^{n-1} \times (1 - p)}$$

$$= \frac{s^{m+1}}{1-p}$$

On en déduit :

$$\boxed{P(W=0) = \sum_{n=1}^{+\infty} P_{V=n}(W=0)\, P(V=n) = \frac{1}{3\,N-2}}$$

Et pour $m \in \mathbb{N}^*$:

$$P_{V=n}(W=m) = P_{V=n}(W>m-1) - P_{V=n}(W>m)$$

$$= \frac{s^m\,(1-s)}{1-p}$$

$$\boxed{P(W=m) = \sum_{n=1}^{+\infty} P_{V=n}(W=m)\, P(V=n) = \frac{s^m\,(1-s)}{1-p}}$$

On vérifie au passage :

$$\sum_{m=0}^{+\infty} P(W=m) = \frac{1}{3\,N-2} + \sum_{m=1}^{+\infty} \frac{s^m\,(1-s)}{1-p}$$

$$= \frac{1}{3\,N-2} + \frac{(1-s)}{1-p} \times \frac{s}{1-s}$$

$$= \frac{1}{3\,N-2} + \frac{s}{1-p} = \frac{1}{3\,N-2} + \frac{3\,N-3}{N^2} \times \frac{N^2}{3\,N-2} = 1$$

3)

a) Considérons une génération donnée k. La génération $k+1$ peut être vue comme un échantillon de taille N prélevé dans une population de petits pois de la génération k pour laquelle il y a une proportion $p = 1/N$ de petits pois issus du même plant P_1. En supposant le nombre de petits pois produits par la génération k de plants suffisamment grand, la variable aléaoire D qui compte le nombre de descendants de P_1 dans la génération $k+1$ suit une loi binomiale de paramètres N et p. Ainsi pour $m \in \{1,2,..N\}$:

$$P(D=m) = \binom{N}{m}\, p^m\,(1-p)^{N-m}$$

b) En particulier :

$$\boxed{P(D=0)=\left(1-\frac{1}{N}\right)^N}$$

c) Définissons N variables aléatoires de Bernoulli A_i par :

$A_i = 1$ si le plant P_i n'a aucun descendant dans la génération suivante, 0 sinon et

$$A = A_1 + A_2 + \cdots + A_N$$

Alors on a :

$$E(A_i) = P(A_i = 1) = \left(1-\frac{1}{N}\right)^N$$

Et :

$$\boxed{E(A) = E(A_1) + E(A_2) + \cdots + E(A_N) = N\,\left(1-\frac{1}{N}\right)^N}$$

4) Considérons les évènements $S_{i_1,i_2,\ldots,i_{k+1}}$ = les plants $P_1, P_2, \ldots, P_{k+1}$ de la génération considérée ont pour ascendant à la génération précédente les plants de numéros respectifs $i_1, i_2, \ldots, i_{k+1}$. Alors, en désignant pas $\mathcal{A}_{k+1,N}$, l'ensemble des arrangements à $k+1$ éléments de $\{1,2,\ldots,N\}$ on a :

$$P(T_r > 1) = \sum_{(i_1,i_2,\ldots,i_{k+1})\in\mathcal{A}_{k+1,N}} P\left(S_{i_1,i_2,\ldots,i_{k+1}}\right) = \frac{A_N^{k+1}}{N^{k+1}}$$

$$= \frac{N\,(N-1)\,(N-2)\ldots(N-k)}{N^{k+1}}$$

$$= \left(1-\frac{1}{N}\right)\left(1-\frac{2}{N}\right)\ldots\left(1-\frac{N}{N}\right) = \prod_{i=1}^{N}\left(1-\frac{i}{N}\right) = q$$

De la même façon pour $n \in \mathbb{N}^*$:

$$P(T_r > n) = q^n$$

Donc pour $n \in \mathbb{N}^*$

$$P(T_r = n) = q^{n-1}(1-q)$$

T_r suit donc une loi géométrique de paramètre q et ainsi en posant $p = 1-q$:

$$\boxed{E(T_r) = \frac{1}{1-q} = \frac{1}{p}}$$

$$\boxed{V(T_r) = \frac{q}{p^2}}$$

Redémontrons ces deux dernières propriétés à partir de l'identité pour $-1 < x < 1$:

$$\sum_{n=0}^{+\infty} x^n = \frac{1}{1-x}$$

Qui par dérivation donne :

$$\sum_{n=1}^{+\infty} n\, x^{n-1} = \frac{1}{(1-x)^2}$$

$$\sum_{n=2}^{+\infty} n\,(n-1)\, x^{n-2} = \frac{2}{(1-x)^3}$$

$$E(T_r) = \sum_{n=0}^{+\infty} n\, q^{n-1}(1-q) = (1-q)\,\frac{1}{(1-q)^2} = \frac{1}{1-q} = \frac{1}{p}$$

$$E\left({T_r}^2\right) = \sum_{n=0}^{+\infty} n^2 q^{n-1}(1-q) = (1-q)\sum_{n=0}^{+\infty}(n(n-1)+n)q^{n-1} =$$

$$= (1-q)\left(\sum_{n=0}^{+\infty} n(n-1)\, q^{n-1} + \sum_{n=0}^{+\infty} n\, q^{n-1}\right)$$

$$= (1-q)\left(q\sum_{n=2}^{+\infty} n(n-1)\, q^{n-2} + \sum_{n=1}^{+\infty} n\, q^{n-1}\right)$$

$$= (1-q)\left(q\,\frac{2}{(1-q)^3} + \frac{1}{(1-q)^2}\right)$$

$$= \frac{2\,q}{p^2} + \frac{1}{p}$$

Donc :

$$V(T_r) = \frac{2\,q}{p^2} + \frac{1}{p} - \frac{1}{p^2} = \frac{2\,q + p - 1}{p^2} = \frac{q}{p^2}$$