Plaquette 5 – Modéliser l'aléatoire : chaînes, génétique, stratégies

Probabilités finies — Prépa MPSI / MP2I (1re année)

Corrections détaillées

Boîte à outils

Modélisation.

— **Chaîne à deux états** $\{1,2\}$: $\mathbb{P}(1\to 2)=a$, $\mathbb{P}(2\to 1)=b$, $0<a+b<2$; si $p_n=\mathbb{P}(\text{état 1 à l'étape } n)$:

$$p_{n+1}=(1-a)p_n+b(1-p_n),\qquad p_n \underset{n\to+\infty}{\longrightarrow} \frac{b}{a+b}$$

— **Forme matricielle** : $X_{n+1}=MX_n$, donc $X_n=M^nX_0$; une loi **stationnaire** vérifie $MX=X$.

— **Conditionner sur le premier pas** (ou sur un composant) pour obtenir une équation.

— **Bayes séquentiel** : la loi a posteriori après une observation sert d'a priori pour la suivante.

— **Génétique** (Hardy-Weinberg) : allèles A, a de fréquences p, $q=1-p$; en panmixie, génotypes AA, Aa, aa en proportions p^2, $2pq$, q^2.

Correction de l'exercice 1 – Une machine qui tombe en panne

a) Probabilités totales selon l'état au jour n :

$$p_{n+1}=0{,}9p_n+0{,}6(1-p_n)=0{,}3p_n+0{,}6$$

b) Point fixe $\ell=0{,}3\ell+0{,}6$, soit $\ell=\frac{6}{7}$. Donc $p_n=\frac{6}{7}+0{,}3^n\left(1-\frac{6}{7}\right)=\frac{6}{7}+\frac{0{,}3^n}{7}\to\frac{6}{7}\approx 0{,}857$.

c) À long terme, la machine fonctionne environ 86 % du temps, indépendamment de l'état initial. On retrouve la formule $\frac{b}{a+b}=\frac{0{,}6}{0{,}1+0{,}6}$.

Réponse. $p_{n+1}=0{,}3p_n+0{,}6$; $p_n=\frac{6}{7}+\frac{0{,}3^n}{7}\to\frac{6}{7}$.

Correction de l'exercice 2 – Marche aléatoire sur un triangle

a) Le pion est en A à l'étape $n+1$ si et seulement s'il était en B ou C à l'étape n (probabilité $1-a_n$) et qu'il choisit A (probabilité $\frac{1}{2}$) :

$$a_{n+1}=\frac{1}{2}(1-a_n)$$

b) Point fixe $\ell=\frac{1}{2}(1-\ell)$, soit $\ell=\frac{1}{3}$; $a_{n+1}-\frac{1}{3}=-\frac{1}{2}\left(a_n-\frac{1}{3}\right)$, et $a_0=1$: $a_n=\frac{1}{3}+\frac{2}{3}\left(-\frac{1}{2}\right)^n$.

c) $a_1=0$, $a_2=\frac{1}{3}+\frac{1}{6}=\frac{1}{2}$, $a_3=\frac{1}{3}-\frac{1}{12}=\frac{1}{4}$. Énumération pour $n=2$: 4 chemins équiprobables ABA, ABC, ACA, ACB, dont 2 reviennent en A : $\frac{1}{2}$.

Réponse. $a_{n+1}=\frac{1}{2}(1-a_n)$; $a_n=\frac{1}{3}+\frac{2}{3}(-\frac{1}{2})^n$; $a_2=\frac{1}{2}$, $a_3=\frac{1}{4}$.

Correction de l'exercice 3 – La loi de Hardy-Weinberg

a) Probabilités totales selon le génotype du parent : AA transmet A avec probabilité 1, Aa avec probabilité $\frac{1}{2}$, aa jamais :

mathlvl.fr — Tous droits réservés. Toute reproduction, diffusion, revente ou exploitation commerciale, même partielle, est strictement interdite.
Tout contrevenant s'expose à des poursuites en vertu du Code de la propriété intellectuelle (art. L.335-2 et suivants).

$$p = u \times 1 + v \times \frac{1}{2} + w \times 0 = u + \frac{v}{2}$$

b) Les deux allèles de l'enfant sont indépendants (parents choisis indépendamment) : AA avec probabilité p^2, aa avec q^2, et Aa avec $2pq$ (deux ordres).

c) À la génération suivante, la fréquence de l'allèle A transmis vaut $p^2 + \frac{2pq}{2} = p(p+q) = p$: les proportions restent p^2, $2pq$, q^2.

d) $p = 0{,}5 + 0{,}1 = 0{,}6$, $q = 0{,}4$: génération suivante AA : 0,36, Aa : 0,48, aa : 0,16, puis stable.

Réponse. $p = u + \frac{v}{2}$; $p^2, 2pq, q^2$, stables ; d) 0,36, 0,48, 0,16.

Correction de l'exercice 4 – Le modèle des urnes d'Ehrenfest

a) Depuis 0 ou 2, on passe à coup sûr en 1. Depuis 1, on va en 0 ou en 2 avec probabilité $\frac{1}{2}$ chacun.

b) $X_1 = 1$; $X_2 \in \{0, 2\}$ avec probabilités $\frac{1}{2}$, $\frac{1}{2}$; $X_3 = 1$. Par récurrence : si n est impair, $X_n = 1$; si $n \geq 2$ est pair, $\mathbb{P}(X_n = 0) = \mathbb{P}(X_n = 2) = \frac{1}{2}$.

c) Si X_n suit la loi $\left(\frac{1}{4}, \frac{1}{2}, \frac{1}{4}\right)$:

$$\mathbb{P}(X_{n+1} = 1) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2}, \qquad \mathbb{P}(X_{n+1} = 0) = \mathbb{P}(X_{n+1} = 2) = \frac{1}{2} \times \frac{1}{2} = \frac{1}{4}$$

— La loi est stationnaire (c'est la loi du nombre de boules dans l'urne 1 quand chacune est placée au hasard). Mais la loi de X_n ne converge pas : elle oscille selon la parité de n (chaîne **périodique**).

Réponse. $X_n = 1$ si n impair ; $\frac{1}{2}$, $\frac{1}{2}$ en 0, 2 si n pair ; stationnaire $\left(\frac{1}{4}, \frac{1}{2}, \frac{1}{4}\right)$, pas de convergence.

Correction de l'exercice 5 – Pile-Pile contre Face-Pile

a) Si les deux premiers lancers sont PP, PP gagne immédiatement. Sinon, un F apparaît parmi les deux premiers lancers ; avant le premier PP éventuel, il y a nécessairement un F suivi d'un P (le premier P du premier PP est précédé d'un F), donc FP apparaît strictement avant.

b) $\mathbb{P}(PP \text{ gagne}) = \frac{1}{4}$, et FP gagne avec probabilité $\frac{3}{4}$.

c) PP contre PF : on attend le premier P ; le lancer suivant décide (P : PP gagne, F : PF gagne), probabilité $\frac{1}{2}$ chacune. L'asymétrie de b) vient de ce que FP « guette » la fin de n'importe quel F, alors que PP doit recommencer à zéro après chaque F.

Réponse. PP contre FP : $\frac{1}{4}$; PP contre PF : $\frac{1}{2}$.

Correction de l'exercice 6 – Collisions dans une table de hachage

a) Sans collision, les clés occupent des cases distinctes (arrangements) :

$$\frac{m(m-1)\cdots(m-n+1)}{m^n} = \prod_{k=0}^{n-1}\left(1 - \frac{k}{m}\right)$$

b) $\prod_{k=0}^{9}\left(1 - \frac{k}{100}\right) \approx 0{,}628$: il y a déjà 37 % de chances de collision avec une table remplie à 10 %.

mathlvl.fr — Tous droits réservés. Toute reproduction, diffusion, revente ou exploitation commerciale, même partielle, est strictement interdite.
Tout contrevenant s'expose à des poursuites en vertu du Code de la propriété intellectuelle (art. L.335-2 et suivants).

c) $1 - \frac{k}{m} \le \mathrm{e}^{-k/m}$, donc le produit est majoré par $\exp\left(-\sum_{k<n} \frac{k}{m}\right) = \exp\left(-\frac{n(n-1)}{2m}\right)$. Pour $n = 30$, $m = 1\,000$: $\exp(-0{,}435) \approx 0{,}65$; plus d'un tiers de risque de collision. C'est le paradoxe des anniversaires : les collisions apparaissent dès n de l'ordre de $\sqrt{m}$, d'où la nécessité de gérer les collisions.

Réponse. $\prod_{k<n}(1 - \frac{k}{m})$; $\approx 0{,}628$; majorant $\mathrm{e}^{-n(n-1)/(2m)}$.

Correction de l'exercice 7 – Le problème du recrutement

a) S_0 accepte le premier candidat, qui est le meilleur avec probabilité $\frac{1}{4}$.

b) Si le meilleur est en position $j \ge 2$, S_1 le retient si et seulement si aucun candidat en positions $2, \dots, j-1$ ne bat tous ses prédécesseurs, c'est-à-dire si le meilleur des $j-1$ premiers est le premier. Par symétrie, chacun des $j-1$ premiers a la même probabilité d'être leur meilleur : $\frac{1}{j-1}$.

c) Le meilleur est en position j avec probabilité $\frac{1}{4}$:

$$\mathbb{P}(S_1) = \frac{1}{4}\left(1 + \frac{1}{2} + \frac{1}{3}\right) = \frac{1}{4} \times \frac{11}{6} = \frac{11}{24} \approx 0{,}458$$

— S_2 : $\frac{1}{4}\left(\frac{2}{2} + \frac{2}{3}\right) = \frac{1}{4} \times \frac{5}{3} = \frac{5}{12} \approx 0{,}417$. La meilleure stratégie est S_1. (Pour n grand, on refuse environ $\frac{n}{\mathrm{e}}$ candidats et l'on réussit avec probabilité proche de $\frac{1}{\mathrm{e}}$.)

Réponse. S_0 : $\frac{1}{4}$; S_1 : $\frac{11}{24}$; S_2 : $\frac{5}{12}$; meilleure : S_1.

Correction de l'exercice 8 – Le problème des partis

a) A a besoin d'une victoire, B de deux : au plus 2 parties décident. On complète fictivement à 2 parties, 4 issues équiprobables : AA, AB, BA, BB (même si la série s'arrête plus tôt, cela ne change pas le vainqueur).

b) B ne gagne que sur BB : $\mathbb{P}(A \text{ gagne}) = \frac{3}{4}$. Partage équitable : $\frac{3}{4}$ de la mise pour A, $\frac{1}{4}$ pour B (solution de Pascal et Fermat, 1654).

c) A a besoin de 2 victoires, B de 4 : au plus 5 parties. B gagne s'il remporte au moins 4 des 5 parties fictives :

$$\mathbb{P}(B) = \frac{\binom{5}{4} + \binom{5}{5}}{2^5} = \frac{6}{32} = \frac{3}{16}$$

— A reçoit $\frac{13}{16}$ de la mise.

Réponse. b) $\frac{3}{4}$ / $\frac{1}{4}$; c) $\frac{13}{16}$ / $\frac{3}{16}$.

Correction de l'exercice 9 – Un gène lié au chromosome X

a) C : « elle est porteuse », N_1 : « le fils n'est pas daltonien ». $\mathbb{P}(N_1 \mid C) = \frac{1}{2}$, $\mathbb{P}(N_1 \mid \overline{C}) = 1$. Bayes :

$$\mathbb{P}(C \mid N_1) = \frac{\frac{1}{2} \times \frac{1}{2}}{\frac{1}{2} \times \frac{1}{2} + \frac{1}{2} \times 1} = \frac{1/4}{3/4} = \frac{1}{3}$$

b) $\mathbb{P}(N_1 \cap N_2 \mid C) = \frac{1}{4}$: $\mathbb{P}(C \mid N_1 \cap N_2) = \frac{\frac{1}{8}}{\frac{1}{8} + \frac{1}{2}} = \frac{1}{5}$.

mathlvl.fr — Tous droits réservés. Toute reproduction, diffusion, revente ou exploitation commerciale, même partielle, est strictement interdite.
Tout contrevenant s'expose à des poursuites en vertu du Code de la propriété intellectuelle (art. L.335-2 et suivants).

c) $\mathbb{P}(C \mid k \text{ fils sains}) = \dfrac{\frac{1}{2}\times 2^{-k}}{\frac{1}{2}\times 2^{-k}+\frac{1}{2}} = \dfrac{1}{1+2^k}$: chaque fils sain rend l'hypothèse « porteuse » moins probable, sans jamais l'exclure.

Réponse. $\frac{1}{3}$; $\frac{1}{5}$; $\dfrac{1}{1+2^k}$.

Correction de l'exercice 10 – Le pont de Wheatstone

a) Si c_5 marche, il faut une liaison de E au nœud $\{M, N\}$ (c_1 ou c_2) et du nœud à S (c_3 ou c_4) : deux blocs parallèles en série,

$$R_1 = (1-(1-p)^2)^2$$

b) Si c_5 est en panne, il reste deux chemins $E-M-S$ (c_1, c_3 en série, fiabilité p^2) et $E-N-S$ (p^2), en parallèle : $R_0 = 1-(1-p^2)^2$.

c) Probabilités totales selon l'état de c_5 :

— $R = p(1-(1-p)^2)^2 + (1-p)(1-(1-p^2)^2)$.

— $p = 0{,}9$: $0{,}9\times 0{,}99^2 + 0{,}1\times(1-0{,}19^2) = 0{,}88209 + 0{,}09639 = 0{,}97848$.

Réponse. $R_1 = (1-(1-p)^2)^2$, $R_0 = 1-(1-p^2)^2$; $R \approx 0{,}978$ pour $p = 0{,}9$.

Correction de l'exercice 11 – Quelle pièce a-t-on choisie ?

a) Les lancers sont indépendants sachant la pièce. Bayes, avec les vraisemblances p^2 :

$$\mathbb{P}(p \mid PP) = \frac{\frac{1}{3}p^2}{\frac{1}{3}\left(\frac{1}{16}+\frac{4}{16}+\frac{9}{16}\right)} = \frac{16p^2}{14}$$

— soit $\frac{1}{14}$, $\frac{4}{14}$, $\frac{9}{14}$ pour $p = \frac{1}{4}, \frac{1}{2}, \frac{3}{4}$.

b) Probabilités totales (sous la probabilité conditionnelle) :

— $\frac{1}{14}\times\frac{1}{4} + \frac{4}{14}\times\frac{1}{2} + \frac{9}{14}\times\frac{3}{4} = \frac{0{,}25+2+6{,}75}{14} = \frac{9}{14} \approx 0{,}643$.

c) Avant observation, $\mathbb{P}(P_1) = \frac{1}{3}\left(\frac{1}{4}+\frac{1}{2}+\frac{3}{4}\right) = \frac{1}{2}$. Les deux Piles observés rendent plus probable la pièce favorable à Pile : la prédiction passe de $\frac{1}{2}$ à $\frac{9}{14}$ (apprentissage bayésien).

Réponse. A posteriori $\frac{1}{14}$, $\frac{4}{14}$, $\frac{9}{14}$; troisième Pile : $\frac{9}{14}$ (contre $\frac{1}{2}$ a priori).

Correction de l'exercice 12 – Le paradoxe de Simpson

a) Petits : A : $\frac{81}{87} \approx 93\,\%$, B : $\frac{234}{270} \approx 87\,\%$. Gros : A : $\frac{192}{263} \approx 73\,\%$, B : $\frac{55}{80} \approx 69\,\%$. A est meilleur **dans chaque groupe**.

b) Globalement : A : $\frac{273}{350} = 78\,\%$; B : $\frac{289}{350} \approx 83\,\%$. B paraît meilleur !

c) Probabilités totales selon la taille du calcul :

$$\mathbb{P}(\text{succès} \mid A) = \mathbb{P}(\text{succès} \mid A, \text{petit})\,\mathbb{P}(\text{petit} \mid A) + \mathbb{P}(\text{succès} \mid A, \text{gros})\,\mathbb{P}(\text{gros} \mid A)$$

— Les poids diffèrent : A a surtout été donné pour de gros calculs ($\frac{263}{350} \approx 75\,\%$), cas difficiles, et B pour de petits ($\frac{270}{350} \approx 77\,\%$). La moyenne globale mélange des populations différentes ; la bonne comparaison est par groupe (la taille du calcul est un **facteur de confusion**).

Réponse. A meilleur dans chaque groupe (93 vs 87 %, 73 vs 69 %) mais global 78 vs 83 %

mathlvl.fr — Tous droits réservés. Toute reproduction, diffusion, revente ou exploitation commerciale, même partielle, est strictement interdite.
Tout contrevenant s'expose à des poursuites en vertu du Code de la propriété intellectuelle (art. L.335-2 et suivants).

: effet des poids.

Correction de l’exercice 13 – Contamination dans un réseau d’ordinateurs

a) $\mathbb{P}(O_2) = 0{,}3$. Par indépendance : les deux, 0,09 ; aucun, $0{,}7^2 = 0{,}49$; exactement un, $2 \times 0{,}3 \times 0{,}7 = 0{,}42$.

b) L’ordinateur sain reste sain s’il échappe aux deux contaminations (indépendantes) :

$$\mathbb{P}(\text{infecté}) = 1 - 0{,}7^2 = 0{,}51$$

c) Probabilités totales selon l’état après le premier jour :

— déjà deux infectés : 0,09 ;
— exactement un : $0{,}42 \times 0{,}51 = 0{,}2142$;
— aucun : il faut que O_1 contamine les deux le second jour, $0{,}49 \times 0{,}09 = 0{,}0441$ (seul O_1 est contagieux au début du second jour).
— Total : $0{,}09 + 0{,}2142 + 0{,}0441 = 0{,}3483$.

Réponse. a) 0,3 ; 0,09 ; 0,49 ; b) 0,51 ; c) $\approx 0{,}348$.

Correction de l’exercice 14 – Sondage et échantillon

a) Majorité favorable : 2 ou 3 favorables. Loi hypergéométrique :

$$\frac{\binom{4}{2}\binom{6}{1} + \binom{4}{3}\binom{6}{0}}{\binom{10}{3}} = \frac{36 + 4}{120} = \frac{1}{3}$$

b) La vraie proportion 0,4 est plus proche de $\frac{1}{3}$ que de $\frac{2}{3}$: l’image est exacte si exactement 1 membre sur 3 est favorable, $\frac{\binom{4}{1}\binom{6}{2}}{120} = \frac{60}{120} = \frac{1}{2}$.

c) Avec remise, le nombre de favorables suit une loi binomiale de paramètres 3 et 0,4 : $\mathbb{P}(\geq 2) = 3 \times 0{,}4^2 \times 0{,}6 + 0{,}4^3 = 0{,}288 + 0{,}064 = 0{,}352$, un peu plus que $\frac{1}{3}$.

Réponse. $\frac{1}{3}$; $\frac{1}{2}$; avec remise 0,352.

Correction de l’exercice 15 – Synthèse : chaîne à deux états sous forme matricielle

a) Probabilités totales : $p_{n+1} = (1-a)p_n + bq_n$ et $q_{n+1} = ap_n + (1-b)q_n$, soit $X_{n+1} = MX_n$.

b) $M(b,a)^{\mathsf{T}} = ((1-a)b + ba,\ ab + (1-b)a) = (b,a)$.
$M(1,-1)^{\mathsf{T}} = (1-a-b,\ a-1+b) = (1-a-b)(1,-1)$.

c) $X_0 = \frac{1}{a+b}\begin{pmatrix} b \\ a \end{pmatrix} + c\begin{pmatrix} 1 \\ -1 \end{pmatrix}$, où c est déterminé par X_0 (la somme des coordonnées de X_0 vaut 1, celle du second vecteur 0). Par linéarité :

$$X_n = M^n X_0 = \frac{1}{a+b}\begin{pmatrix} b \\ a \end{pmatrix} + c(1-a-b)^n \begin{pmatrix} 1 \\ -1 \end{pmatrix} \xrightarrow[n \to +\infty]{} \frac{1}{a+b}\begin{pmatrix} b \\ a \end{pmatrix}$$

— car $|1-a-b| < 1$.

d) $a + b = 0{,}7$; $X_0 = (1,0) = \frac{1}{0{,}7}(0{,}6, 0{,}1) + c(1,-1)$ donne $c = 1 - \frac{6}{7} = \frac{1}{7}$. Donc $p_n = \frac{6}{7} + \frac{0{,}3^n}{7}$: c’est la formule trouvée pour la machine.

Réponse. $X_n = \frac{1}{a+b}(b,a) + c(1-a-b)^n(1,-1) \to \frac{1}{a+b}(b,a)$; d) $p_n = \frac{6}{7} + \frac{0{,}3^n}{7}$.

mathlvl.fr — Tous droits réservés. Toute reproduction, diffusion, revente ou exploitation commerciale, même partielle, est strictement interdite.
Tout contrevenant s’expose à des poursuites en vertu du Code de la propriété intellectuelle (art. L.335-2 et suivants).