Université Denis Diderot Paris 7

27 juin 2014

Probabilités et Simulations U1PS36

Examen de rattrapage

durée : 3 heures
Les documents et calculatrices ne sont pas autorisés.
On prendra soin de bien justifier les réponses.
Le barême est provisoire, et n'est donné qu'à titre d'indication de l'importance relative des exercices.

Questions de cours — Application directe du cours

Les questions qui suivent sont indépendantes.

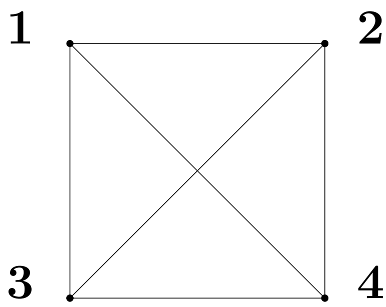
1. *(1.5 point)* Enoncer la loi forte des grands nombres, puis le théorème de la limite centrale.
2. *(1.5 point)* Quand dit-on d'une chaîne de Markov sur un espace d'état E fini qu'elle est irréductible, apériodique ? Enoncer le théorème de convergence pour une telle chaîne.
3. *(2 points)* Donner un exemple de chaîne de Markov réversible à 4 états, et préciser sa mesure stationnaire (justifier).
   Donner un exemple de chaîne de Markov non réversible à 4 états (justifier).
4. *(1 point)* Soit $n \geq 1$, et $X_1, ...X_n$ des variables i.i.d. suivant une loi exponentielle de paramètre 1. Quelle est la densité de $S_n = \sum_{i=1}^n X_i$?
5. *(2.5 points)* Soient $X_1, X_2, ...$ des variables i.i.d. suivant une loi de Cauchy de paramètre 1. On rappelle que $\Phi_{X_1}(t) = \exp(-|t|)$. Pour $n \geq 1$, on note $S_n = \sum_{i=1}^n X_i$, et $Z_n = \frac{S_n}{n}$. quelle est la limite en loi de $(Z_n)_{n\geq 1}$ (justifier) ? Comment vérifier le résultat par une simulation ?

1. LFGN : Soient $\{X_i, i \geq 1\}$ des variables i.i.d. avec $\mathbb{E}[|X_1|] < \infty$. Presque sûrement, lorsque $n \to \infty$,
$$\frac{1}{n}\sum_{i=1}^{n} X_k \longrightarrow \mathbb{E}[X_1].$$
   TCL : Sous les mêmes hypothèses, et si de plus $\mathrm{Var}[X_1] = \sigma^2 \in (0, \infty)$, alors
$$\frac{\sum_{k=1}^n X_k - n\mathbb{E}[X_1]}{\sigma\sqrt{n}} \xrightarrow[n\to\infty]{\text{(loi)}} Z \sim \mathcal{N}(0,1).$$
2. Une chaîne de Markov sur E est irréductible ssi pour toute paire d'états i, j, il existe $k \in \mathbb{N}$ tel que $p_{ij}^{(k)} > 0$.
   Elle est de plus apériodique ssi, pour un $x \in E$, $\mathrm{pgcd}\{k \in \mathbb{N} : p_{xx}^{(k)} > 0\} = 1$.

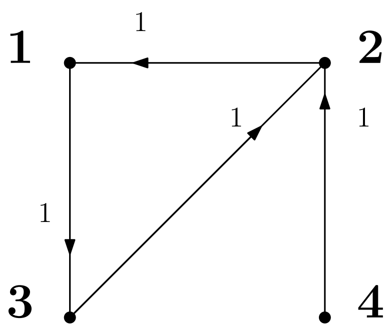
Une telle chaîne possède une unique distribution stationnaire π (puisque l'espace d'état est fini il y en a au moins une, et puisque la chaîne est irréductible elle est unique), et vérifie le théorème de convergence, à savoir que, quelle que soit la mesure initiale ν, la loi de X_n converge vers π lorsque $n \to \infty$. Plus précisément, quelle que soit ν mesure sur E,

$$\mathbb{P}_\nu(X_n = k) \underset{n\to\infty}{\longrightarrow} \pi(k), \quad \forall\, k \in E.$$

3. La marche aléatoire sur le graphe complet à 4 sommets (voir le schéma ci-dessous) est réversible, et bien évidemment sa distribution stationnaire est uniforme sur $\{1, ..., 4\}$.



En revanche, le diagramme suivant est celui d'une chaîne non réversible :



En effet, il est facile de voir que pour cette chaîne la seule classe récurrente est $\{1, 2, 3\}$, et l'unique distribution invariante vérifie $\pi(1) = \pi(2) = \pi(3) = 1/3$. Mais

$$1/6 = \pi(1)p_{12} \neq \pi(2)p_{21} = 0,$$

de sorte que les équations de balance détaillée ne sont pas satisfaites, et la chaîne n'est pas réversible.

4. D'après le cours $S_n \sim \Gamma(n, 1)$ (on peut le vérifier par exemple à l'aide des fonctions caractéristiques). Pour $t \in \mathbb{R}$,

$$f_{S_n}(t) = \frac{1}{(n-1)!} t^{n-1} \exp(-t) \mathbf{1}_{\{t\geq 0\}}.$$

5. En utilisant le rappel, et le fait que les $X_k, k \geq 1$ sont i.i.d.,

$$\begin{aligned}
\Phi_{Z_n}(t) &= \mathbb{E}\left[\exp\left(i\frac{t}{n}\sum_{k=1}^{n} X_i\right)\right] \\
&= \prod_{k=1}^{n} \mathbb{E}[\exp(i\frac{t}{n}X_k)] \\
&= (\exp(-\frac{|t|}{n}))^n \\
&= (1-\frac{|t|}{n}+o(1/n))^n \underset{n\to\infty}{\longrightarrow} \exp(-|t|)
\end{aligned}$$

La convergence ponctuelle des fonctions caractéristiques est équivalente à la convergence en loi, on en déduit que

$$Z_n \underset{n\to\infty}{\overset{(\text{loi})}{\longrightarrow}} X_1.$$

On notera que les variables $X_k, k \geq 1$ ne satisfont pas la loi des grands nombres, mais cela ne contredit pas le théorème puisque $\mathbb{E}[|X_1|] = +\infty$.
Pour vérifier le résultat par une simulation, rappelons que si $U \sim \text{Unif}(0,1)$, $\tan(\pi(U-1/2)) \sim \text{Cauchy}(1)$.
Il est facile de programmer la fonction $f : x \to \tan(\pi(x-1/2))$:

```
function f=forcauchy(x)
 f=tan(%pi*(x-1/2));
endfunction
```

Il est alors aisé de tracer un histogramme de N réalisations d'une loi de Cauchy :

```
function histCauchy(N)
 x=rand(1:N);
 y=feval(x,forcauchy);
 histplot(-10:0.5:10,y);
endfunction
```

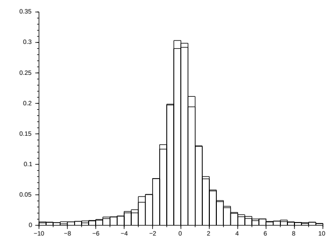
On peut d'autre part tracer un histogramme de N réalisations de Z_n :

```
 function verif(N,n)
 x=rand(N,n);
 y=feval(x,forcauchy);
 z=sum(y,'c')/n;
 histplot(-10:0.5:10,z);
 endfunction
```

Par exemple, pour $N = 5000, n = 5000$ on obtient les 2 histogrammes suivants (on les a représentés sur un même graphique pour une comparaison plus directe) :

Exercice 1 *(4 points)* On considère le couple (X,Y) de densité :

$$f(x,y) = \frac{x^2}{2}\exp(-xy)\mathbf{1}_{\{0\leq x\leq 2, y\geq 0\}}.$$



1. Calculer la densité de X, puis celle de Y. Les deux variables sont-elles indépendantes ?
2. Calculer $\mathbb{P}(X \leq Y)$.
3. Déterminer la loi du couple $(U, V) := (X, 1 - \exp(-XY))$.

1. La densité de X est
$$f_X(x) = \int_{\mathbb{R}} f(x,y)dy = \mathbf{1}_{\{0\leq x\leq 2\}}\frac{x}{2}\int_0^{+\infty} x\exp(-xy)dy = \frac{x}{2}\mathbf{1}_{\{0\leq x\leq 2\}}.$$
La densité de Y se calcule de même, en utilisant une double intégration par parties :
$$\begin{aligned} f_Y(y) &= \int_{\mathbb{R}} f(x,y)dx = \mathbf{1}_{\{y\geq 0\}}\int_0^2 \frac{x^2}{2}\exp(-xy)dx \\ &= \mathbf{1}_{\{y\geq 0\}}\left(\left[-\frac{x^2}{2y}\exp(-xy)\right]_0^2 + \int_0^2 \frac{x}{y}\exp(-xy)dx\right) \\ &= \mathbf{1}_{\{y\geq 0\}}\left(\left[(-\frac{x^2}{2y} - \frac{x}{y^2})\exp(-xy)\right]_0^2 + \int_0^2 \frac{1}{y^2}\exp(-xy)dx\right) \\ &= \mathbf{1}_{\{y\geq 0\}}\left(\left[(-\frac{x^2}{2y} - \frac{x}{y^2} - \frac{1}{y^3})\exp(-xy)\right]_0^2\right) \\ &= \left(\frac{1}{y^3} - \left(\frac{2}{y} + \frac{2}{y^2} + \frac{1}{y^3}\right)\exp(-2y)\right)\mathbf{1}_{\{y\geq 0\}}. \end{aligned}$$
Note : Pour s'assurer que f_Y reste une fonction intégrable sur $\mathbb{R}_+$, on peut voir que malgré la présence de fractions rationnelles en y cette densité est en fait prolongeable par continuité en 0, un DL permet en effet d'assurer que $\lim_{y\searrow 0} f_Y(y) = 4/3$.
2. On a (en utilisant Fubini-Tonelli)
$$\begin{aligned} \mathbb{P}(X \leq Y) &= \mathbb{E}[\mathbf{1}_{\{X\leq Y\}}] = \int_{\mathbb{R}^2} \mathbf{1}_{\{x\leq y\}} f(x,y)dxdy \\ &= \int_0^2 \frac{x}{2}\left(\int_x^{+\infty} x\exp(-xy)dy\right)dx \\ &= \int_0^2 \frac{x}{2}\left[-\exp(-xy)\right]_0^{+\infty} dx \\ &= \int_0^2 \frac{x}{2}\exp(-x^2)dx \\ &= \left[\frac{1}{4}\exp(-x^2)\right]_0^2 = \frac{1}{4}(1-\exp(-4)). \end{aligned}$$

3. Soit $\phi : \mathbb{R}^2 \to \mathbb{R}$ bornée mesurable, on a

$$\begin{aligned}
\mathbb{E}[\phi(U,V)] &= \int_{\mathbb{R}^2} \phi(x, 1-\exp(-xy)) f(x,y) dx dy \\
&= \int_{[0,2]\times\mathbb{R}_+} \phi(x, 1-\exp(-xy)) \frac{x^2}{2} \exp(-xy) dx dy \\
&= \int_{[0,2]\times[0,1]} \phi(u,v) \frac{u}{2} du dv,
\end{aligned}$$

où à la dernière ligne on a effectué le changement de variables

$$\left\{ \begin{array}{l} [0,2]\times\mathbb{R}_+ \to [0,2]\times[0,1] \\ (x,y) \to (u,v) = (x, 1-\exp(-xy)) \end{array} \right. ,$$

de jacobien $x\exp(-xy)$.
Comme ceci est valable pour toute fonction ϕ bornée mesurable, on en déduit que

$$\begin{aligned}
f_{(U,V)}(u,v) &= \frac{u}{2} \mathbf{1}_{\{u\in[0,2], v\in[0,1]\}} \\
&= \frac{u}{2} \mathbf{1}_{\{u\in[0,2]\}} \mathbf{1}_{\{v\in[0,1]\}} =: f_U(u) f_V(v),
\end{aligned}$$

autrement dit, les variables $U = X$ et $V = 1-\exp(-XY)$ sont indépendantes, on a donné la densité de $U = X$ à la question 1, tandis que $V = 1-\exp(-XY) \sim \mathrm{Unif}[0,1]$.

Exercice 2 *(4 points)* On considère une chaîne de Markov $(X_n, n \geq 0)$, dont le diagramme est représenté par la figure suivante :

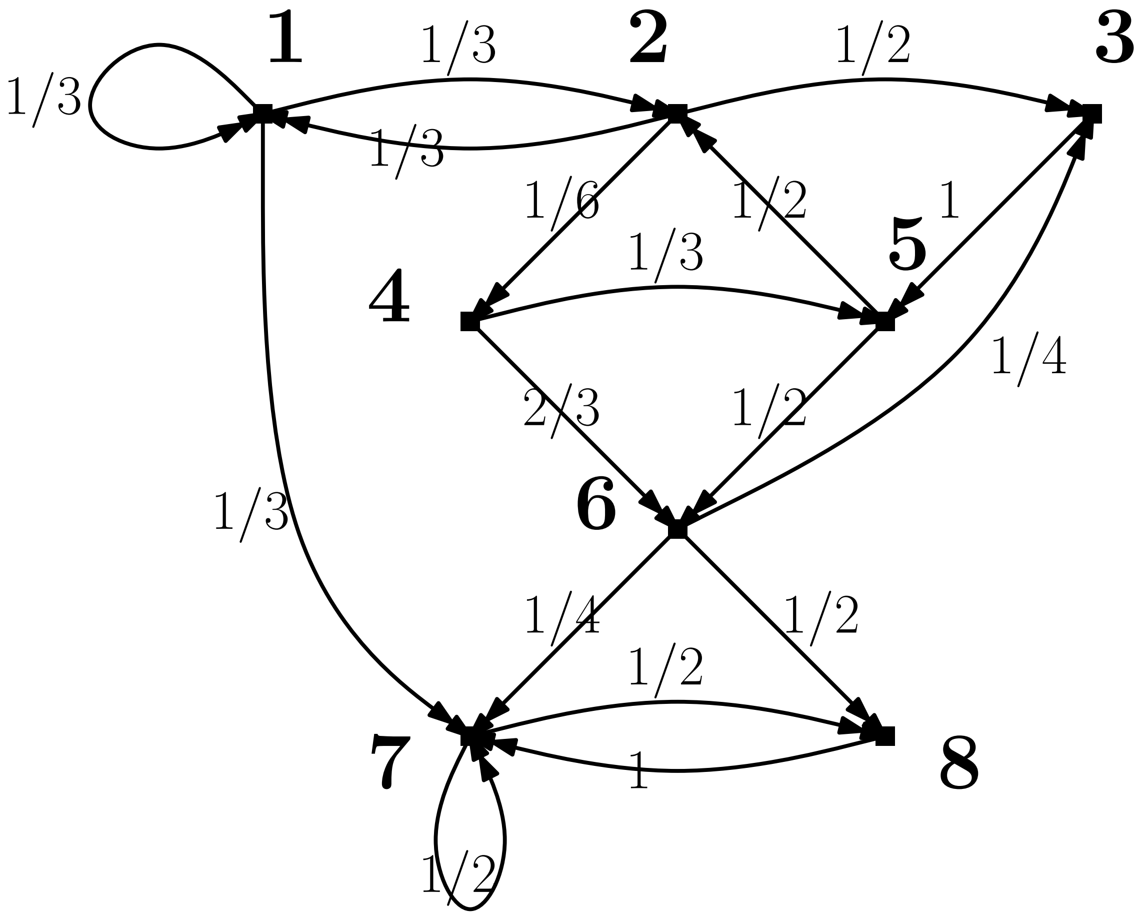


FIGURE 1. Le diagramme de la chaîne X

1. Ecrire la matrice de transition P de la chaîne X.
2. Décomposer l'espace d'état $\{1, 2, ..., 8\}$ en classes de communication. Lesquelles parmi ces classes sont transientes ? récurrentes ? On justifiera les réponses à ces deux questions.
3. Soit ν la mesure uniforme sur $\{1, ..., 8\}$. Calculer $\mathbb{P}_\nu(X_1 = k), k \in \{1, ..., 8\}$.
4. Trouver l'ensemble des distributions invariantes de la chaîne.
5. Calculer $\lim_{n\to\infty} \mathbb{P}_\nu(X_n = k)$ pour $k \in \{1, ..., 8\}$.

1. Par inspection du diagramme, on voit que

$$P = \begin{pmatrix} 1/3 & 1/3 & 0 & 0 & 0 & 0 & 1/3 & 0 \\ 1/3 & 0 & 1/2 & 1/6 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1/3 & 2/3 & 0 & 0 \\ 0 & 1/2 & 0 & 0 & 0 & 1/2 & 0 & 0 \\ 0 & 0 & 1/4 & 0 & 0 & 0 & 1/4 & 1/2 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1/2 & 1/2 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \end{pmatrix}.$$

2. On a $1 \to 2 \to 4 \to 6 \to 3 \to 5 \to 2 \to 1$, les états $\{1, 2, 3, 4, 5, 6\}$ appartiennent donc clairement à une même classe.

   On a $7 \to 8 \to 7$, donc $\{7, 8\}$ appartiennent également à une même classe. Enfin de 7 et 8 on ne peut parvenir à aucun état parmi $\{1, 2, 3, 4, 5, 6\}$.

   On conclut finalement qu'il y a deux classes :

   $$\{1, 2, 3, 4, 5, 6\}, \{7, 8\}.$$

   La classe $\{7, 8\}$ est la seule classe fermée, et donc récurrente, tandis que $\{1, 2, 3, 4, 5, 6\}$ est transientes puisque non fermée (cf, e.g. $1 \to 7$).

3. En notant $\nu = (1/8 \ldots 1/8)$ on a d'après le cours $\nu P = (P_\nu(X_1 = k))_{k=1,\ldots,8}$. On trouve

   $$\nu P = (1/12\ 5/48\ 3/32\ 1/48\ 1/6\ 7/48\ 25/96\ 1/8) = \frac{1}{96}(8\ 10\ 9\ 2\ 16\ 14\ 25\ 12).$$

4. Il y a une unique classe fermée et donc une unique distribution invariante, qui ne charge que les états de cette classe. On a $\pi(7) = 1/2\pi(7) + \pi(8)$ et donc $\pi(7) = 2\pi(8)$, et comme $\pi(7) + \pi(8) = 1$, on conclut finalement que $\pi(7) = 2/3, \pi(8) = 1/3$.

5. Le temps d'atteinte de la classe $\{7, 8\}$ est fini presque sûrement, il suffit donc de considérer la chaîne restreinte aux états $\{7, 8\}$. Cette chaîne restreinte est évidemment irréductible, et comme $p_{77} > 0$, elle est également apériodique. La chaîne X vérifie donc le théorème de convergence, et on a donc, pour tout $k \in \{1, ..., 8\}$,

   $$\mathbb{P}(X_n = k) \underset{n\to\infty}{\longrightarrow} \pi(k),$$

   où π est l'unique distribution stationnaire trouvée en 2.

Exercice 3 *(4 points)* On considère $(X_n, n \geq 0)$ la marche aléatoire simple sur le graphe représenté sur la figure suivante.
On dit que ce graphe est un arbre 2-régulier de hauteur 4 : chaque noeud u de hauteur $n \in \{0, 1, 2, 3\}$ possède en effet 2 "descendants" $u1, u2$ à la hauteur $n + 1$, tandis que les noeuds de hauteur 4 n'ont pas de descendants : ce sont les feuille de l'arbre.
On suppose que sous $\mathbb{P}$, X est issue de la racine $\emptyset$ de l'arbre (la racine est le seul noeud du graphe à hauteur 0).

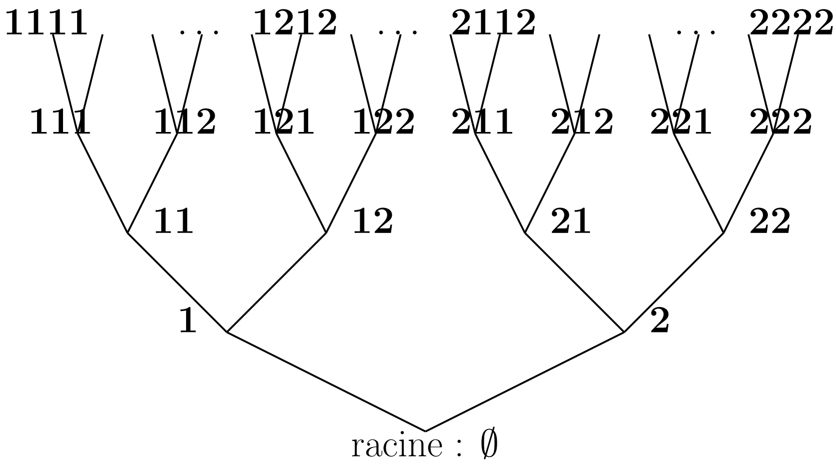


FIGURE 2. Faute de place, on n'a pas numéroté tous les noeuds du graphe à la hauteur 4, mais par exemple, les deux descendants de 112 sont, de gauche à droite, notés 1121 et 1122.

1. Calculer $\mathbb{P}(X_1 = 1)$, $\mathbb{P}(X_2 = \emptyset)$, $\mathbb{P}(X_4 = 1111)$, $\mathbb{P}(X_7 = 12)$, et enfin $\mathbb{P}(X_{2k} = \emptyset\ \forall k \geq 0)$.
2. La chaîne X est-elle irréductible ? réversible ? apériodique ?
3. Montrer que la chaîne $(X_n, n \geq 0)$ possède une unique distribution invariante π, que l'on déterminera.
4. Pour un noeud u de l'arbre, on note $h(u)$ la hauteur dans l'arbre du noeud u. Par exemple $h(1211) = 4, h(21) = 2$, etc... Que vaut $\mathbb{P}(h(X_2) = 0)$? $\mathbb{P}(h(X_4) = 0)$? Calculer
$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} h(X_n).$$

1. On a $\mathbb{P}(X_1 = 1) = 1/2$. Par ailleurs
$$\begin{aligned} \mathbb{P}(X_2 = \emptyset) &= \mathbb{P}(X_1 = 1, X_2 = \emptyset) + \mathbb{P}(X_1 = 2, X_2 = \emptyset) \\ &= 1/2 * 1/3 + 1/2 * 1/3 = 1/3. \end{aligned}$$
D'autre part (il n'y a qu'une trajectoire possible pour se rendre en une feuille de l'arbre au temps 4),
$$\mathbb{P}(X_4 = 1111) = \mathbb{P}(X_1 = 1, X_2 = 11, X_3 = 111, X_4 = 1111) = 1/2 * 1/3 * 1/3 * 1/3 = 1/54.$$

Remarquons qu'en un temps pair, la marche issue de la racine se trouve forcément en une hauteur paire de l'arbre, tandis qu'en un temps impair elle se trouve forcément en une hauteur impaire de l'arbre, on a donc $\mathbb{P}(X_7 = 12) = 0$.
Enfin, pour tout $n \geq 1$,

$$\mathbb{P}(X_{2k} = \emptyset \ \forall k \geq 0) \leq \mathbb{P}(X_{2k} = \emptyset, k = 1, ..., n) = (1/3)^n,$$

et donc $\mathbb{P}(X_{2k} = \emptyset \ \forall k \geq 0) = 0$.

2. Le graphe est connexe, et donc la marche simple sur ce graphe est irréductible, d'après le cours elle est réversible avec une distribution stationnaire proportionnelle au degré. Finalement, on a déjà mentionné à la question précédente que la chaîne issue de la racine ne peut être qu'à une hauteur impaire aux temps impairs, et donc

$$\text{pgcd}\{k \in \mathbb{N} : p_{\emptyset\emptyset}^{(k)} > 0\} = \text{pgcd}(2\mathbb{N}) = 2,$$

de sorte que la chaîne X est périodique de période 2.

3. Le degré de la racine est 2, celui des feuilles est 1, et celui de tous les noeuds intermédiaires (aux hauteurs $1, 2, 3$ est 3. Le degré total est donc $2 + 3 * (2 + 4 + 8) + 1 * 16 = 50$, et on a

$$\begin{aligned}
&\pi(\emptyset) = 1/25, \\
&\pi(1) = \pi(2) = \pi(11) = \cdots = \pi(222) = 3/50, \\
&\pi(1111) = \pi(1112) = \pi(1121) = \cdots = \pi(2222) = 1/50.
\end{aligned}$$

Note sur la question 2 : Puisque pour tous noeuds (x, y) du graphe $p_{xy} = \frac{1}{\deg(x)}$, on a bien

$$\pi_x p_{xy} = 1/50, \ \forall x, y,$$

ce qui assure que les équations de balance détaillée sont satisfaites, et la chaîne est donc bien réversible.

4. La chaîne est irréductible sur un espace d'état fini E (et l'unique distribution stationnaire a été calculée à la question précédente), le théorème ergodique s'applique et donc presque sûrement

$$\begin{aligned}
\lim_{n\to\infty} \frac{1}{n} \sum_{k=0}^{n-1} h(X_n) &= \sum_{x \in E} h(x)\pi(x) \\
&= 0 * \frac{1}{25} + 1 * 2 * \frac{3}{50} + 2 * 4 * \frac{3}{50} + 3 * 8 * \frac{3}{50} + 4 * 16 * \frac{1}{50} \\
&= \frac{166}{50} = 3.32
\end{aligned}$$