MT330 - Probabilités et modèles stochastiques

TP N°2 : chaînes de Markov à temps discret

1 Introduction, définitions et notations

Nous allons considérer des processus stochastiques (aléatoires) à espace d'état discret $E \subset \mathbb{N}$ et à temps discret *sans mémoire*. On notera $X_n \in E$ l'état du processus au temps $n \in \mathbb{N}$. Dans un *processus markovien*, l'état futur du système X_{n+1} ne peut dépendre que de l'état présent X_n, mais pas des valeurs passées X_k pour $k < n$. De plus, nous ne regarderons que des *processus homogènes* : la loi de transition qui établit les probabilités de passage d'un état $X_n$au temps n à un autre état au temps X_{n+1} au temps $n+1$ ne variera pas au cours du temps.

Définition 1 (chaîne de Markov homogène à temps discret (CMTD)) Une suite $(X_n)_{n\in\mathbb{N}}$ de variables aléatoires à valeurs dans $E \subset \mathbb{N}$ est une *chaîne de Markov* si et seulement si

$$\begin{gathered} \mathbb{P}\left[X_{n+1} = x_{n+1} \middle| \, X_n = x_n; \ldots; X_0 = x_0\right] = \\ \mathbb{P}\left[X_{n+1} = x_{n+1} \middle| X_n = x_n\right] \\ \forall n \in \mathbb{N},\, \forall x_{n+1} \in E,\, \forall (x_0, x_1, \ldots, x_n) \in E^n \end{gathered} \tag{1}$$

Elle est dite *homogène* si

$$\mathbb{P}\left[X_{n+1} = x_j \middle| X_n = x_i\right] = p_{x_i,x_j} \ ,\ \forall x_i, x_j \in E \tag{2}$$

On notera

$$p_{i,j} := \mathbb{P}\left[X_{n+1} = j \middle| \, X_n = i\right] \ \forall n \in \mathbb{N},\, \forall i, j \in E \tag{3}$$

les probabilités de transition de l'état i vers l'état j, et $\mathbf{P}$ la *matrice de transition* du processus définie par :

$$(\mathbf{P})_{i,j} := p_{i,j} \tag{4}$$

Exemple 1 (labyrinthe) Dans un labyrinthe représenté schématiquement à la figure 1 (partie gauche), des rats doivent emprunter les corridors dans le sens indiqué par les double-flèches. Ce labyrinthe comporte 5 pièces,

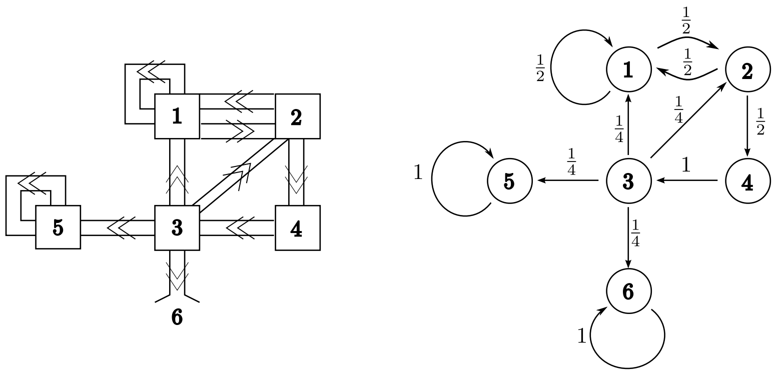

Fig. 1: Représentation schématique du labyrinthe de l'exemple 1 (à gauche) et graphe de la chaîne de Markov associée (à droite). Les noeuds représentent les états et les poids sur les arrêtes représentent les probabilités de transition correspondantes.

numérotées 1 à 5, ainsi qu'une sortie, numérotée 6. Quand un rat est dans une pièce, il emprunte aléatoirement un des corridors qui s'ouvrent devant lui avec une distribution de probabilité uniforme. Par exemple, un rat dans la pièce 3 ira au hasard vers les pièces 1, 2, 5 ou 6 avec une probabilité $\frac{1}{4}$ pour chacune de ces pièces. L'hypothèse markovienne se traduit ici par le fait que le rat choisit parmi les corridors qui s'ouvrent devant lui toujours avec la même distribution de probabilités, quel que soit son parcours antérieur dans le labyrinthe : il n'a littéralement aucune mémoire ! Nous pouvons alors représenter la marche aléatoire des rats dans le labyrinthe par une CMTD. L'état au temps n sera le numéro de la pièce dans laquelle se trouve le rat au temps n. L'espace d'état est donc fini $E := \{1, \ldots, 6\} \subset \mathbb{N}$. Il est fréquent de représenter une CMTD par un *graphe associé*. Les noeuds sont les états et les arrêtes représentent les transitions entre états dont la probabilité de transition est non nulle (voir figure 1, partie droite). La *matrice de transition* est donnée dans cet exemple par :

$$\mathbf{P} = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & 0 \\ \frac{1}{2} & 0 & 0 & \frac{1}{2} & 0 & 0 \\ \frac{1}{4} & \frac{1}{4} & 0 & 0 & \frac{1}{4} & \frac{1}{4} \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \tag{5}$$

La CMTD qui modélise la marche aléatoire du rat dans le labyrinthe est complètement définie, soit par la matrice de transition (5), soit par le graphe associé décrit à la figure 1. Toutes les probabilités relatives à ce problème peuvent alors être calculées à l'aide de ce modèle et de la *distribution initiale de probabilité* (i.e. la loi de probabilité de la variable aléatoire X_0).

2 Probabilités d'états et probabilités de chemins

Définition 2 (Probabilité d'état) Soit $\{X_n\}_{n\in\mathbb{N}}$ une CMTD homogène sur un espace d'état E discret. Le vecteur ligne $\boldsymbol{\pi}^{(n)}$ des *probabilités d'état* au temps n ($n \geqslant 0$) est défini composante par composante par :

$$\boldsymbol{\pi}_j^{(n)} := I\!P\left[X_n = j\right] \ , \forall j \in E$$

On appelle *distribution initiale* de probabilités le vecteur des probabilités d'état à l'instant $n = 0$:

$$\boldsymbol{\pi}_j^{(0)} := I\!P\left[X_0 = j\right] \ , \forall j \in E$$

La distribution $\boldsymbol{\pi}^{(0)}$ représente notre connaissance de la configuration du système au temps $n = 0$. Si l'état initial $X_0 = i$ est connu, on choisira $\boldsymbol{\pi}_j^{(0)} = \delta_{i,j}$. Une fois connue la distribution initiale de probabilités $\boldsymbol{\pi}^{(0)}$ et les probabilités de transition $p_{i,j}$, toutes les valeurs ultérieures du vecteur des probabilités d'état $\boldsymbol{\pi}^{(n)}$ peuvent être calculées grâce à l'hypothèse markovienne et à la formule des probabilités totales. On a en effet :

$$\begin{aligned}
\boldsymbol{\pi}_j^{(n)} &= I\!P\left[X_n = j\right] \\
&= \sum_{i \in E} I\!P\left[X_n = j \,|X_{n-1} = i\right] I\!P\left[X_{n-1} = i\right] \\
&= \sum_{i \in E} p_{i,j} \boldsymbol{\pi}_i^{(n-1)}
\end{aligned}$$

Sous forme matricielle, on obtient pour le vecteur des probabilités d'état :

$$\boldsymbol{\pi}^{(n)} = \boldsymbol{\pi}^{(n-1)} \mathbf{P} \tag{6}$$

ou encore, par récurrence sur n :

$$\boldsymbol{\pi}^{(n)} = \boldsymbol{\pi}^{(0)} \mathbf{P}^n \tag{7}$$

L'équation (7) est parfois appelée relation de Chapman-Kolmogorov.

Définition 3 (Probabilité d'un chemin) Soit $\{X_n\}_{n\in\mathbb{N}}$ une CMTD homogène sur un espace d'état E discret. Soit $(x_0, x_1, \ldots, x_n) \in E^n$ un chemin quelconque de longueur n dans l'espace des états. La *probabilité du chemin* $(x_0, x_1, \ldots, x_n)$ est définie par

$$p_{x_0,x_1,\ldots,x_n} := I\!P\left[X_n = x_n, \ldots, X_1 = x_1, X_0 = x_0\right] \tag{8}$$

On a :

$$\begin{aligned}
p_{x_0,x_1,\ldots,x_n} &:= I\!P\left[X_n = x_n, \ldots, X_1 = x_1, X_0 = x_0\right] \\
&= I\!P\left[X_n = x_n, \ldots, X_1 = x_1 \,|X_0 = x_0\right] \pi_{x_0}^{(0)} \\
&= I\!P\left[X_n = x_n, \ldots, X_2 = x_2 \,|X_1 = x_1\right] p_{x_0,x_1} \pi_{x_0}^{(0)} \\
&= p_{x_{n-1},x_n} p_{x_{n-2},x_{n-1}} \cdots p_{x_1,x_2} p_{x_0,x_1} \pi_{x_0}^{(0)}
\end{aligned} \tag{9}$$

La probabilité de suivre un chemin quelconque du graphe associé à une CMTD homogène est donc le produit des probabilités de transition entre les noeuds du graphe (pris deux à deux) qui composent ce chemin.

Exercice 1 (Le labyrinthe ... suite)
On considère le modèle markovien du labyrinthe défini dans l'exemple 1.

1. Calculer la probabilité qu'un rat parte de la pièce numéro 2 et y revienne après quatre déplacements (i.e. $P(X_4 = 2 \,|\, X_0 = 2))$ à l'aide de la relation de Chapman-Kolmogorov (7)
2. Calculer cette même probabilité à l'aide de la formule de la probabilité d'un chemin (9)
3. Vérifier le résultat obtenu par échantillonnage (programme **R**)

3 Classes d'équivalence, états récurrents et transitoires

Définition 4 (classes d'équivalence) Soit $\{X_n\}_{n\in\mathbb{N}}$ une CMTD homogène sur un espace d'état E discret. L'état $j \in E$ est dit *atteignable* depuis l'état $i \in E$ (on notera $i \rightsquigarrow j$) si et seulement si

$$\exists n \in \mathbb{N} : p_{i,j}^{(n)} > 0$$

où $p_{i,j}^{(n)} := (\mathbf{P}^n)_{i,j} = P\,(X_n = j \,|\, X_0 = i\,)$. Les état $i, j \in E$ *communiquent* (on notera $i \leftrightsquigarrow j$) si et seulement si $i \rightsquigarrow j$ et $j \rightsquigarrow i$. La relation de communication est une relation d'équivalence entre les états de E, c'est-à-dire une relation binaire sur E réflexive, symétrique et transitive. On définit la *classe d'équivalence* de $i \in E$ par $[i] := \{j \in E \,|\, i \leftrightsquigarrow j\,\}$. On appelle alors i un *représentant* de la classe $[i]$. L'espace d'état E se *partitionne* en l'ensemble des classes d'équivalence de E qui sont toutes non vides et disjointes deux à deux. La chaîne de Markov $\{X_n\}_{n\in\mathbb{N}}$ est dite *irréductible* si et seulement si tous les états de E communiquent entre eux, c'est-à-dire si il n'y a qu'une seule classe d'équivalence dans E

Dans l'exemple 1 (labyrinthe), on a $E := \{1, \ldots, 7\}$. Les états 1, 2, 3 et 4 communiquent. Par contre, ils ne sont pas atteignables depuis les états 5 ou 6. Les classes d'équivalence sont donc $[1] = [2] = [3] = [4] = \{1, 2, 3, 4\}$, $[5]$ et $[6]$. L'espace d'état E se partitionne donc en trois classes d'équivalence disjointes : $E = \{1, 2, 3, 4\} \cup \{5\} \cup \{6\}$. Parmi ces trois classes, $[5]$ et $[6]$ sont deux sous-chaînes *absorbantes* : dès qu'on y est entré, on n'en sort plus. La classe $[1]$ par contre est une sous-chaîne *transitoire*. On est sûr de la quitter à un moment ou un autre (pour peu qu'on démarre dans cette classe) et on n'y reviendra ensuite plus. Les états de la classe $[1]$ sont dits *transitoires*.

Définition 5 (Probabilités de transition en n étapes exactement) Soit $\{X_n\}_{n\in\mathbb{N}}$, une CMTD homogène sur un espace d'état E discret. La probabilité de transition de l'état i à l'état j en n étapes *exactement* est :

$$f_{i,j}^{(n)} := I\!P\left[X_n = j; X_{n-1} \neq j; \ldots; X_1 \neq j \,|X_0 = i\right] \tag{10}$$

C'est la probabilité que le premier retour en $i \in E$ ait lieu en n étapes exactement. Le temps moyen de premier retour en $i \in E$ s'écrit :

$$M_i := \sum_{n\in\mathbb{N}} n f_{i,i}^{(n)} \tag{11}$$

La probabilité de transition de l'état $i \in E$ à l'état $j \in E$ en un nombre *quelconque* d'étapes vaut :

$$f_{i,j} := \sum_{n\in\mathbb{N}} f_{i,j}^{(n)} \tag{12}$$

La probabilité revenir en $i \in E$ s'écrit alors $f_{i,i}$.

Le calcul direct des probabilités $f_{i,j}^{(n)}$ et $f_{i,j}$ peut s'avérer compliqué et dans certains cas inextricable. Heureusement, une approche récursive de ce problème permet d'établir une récurrence pour les $f_{i,j}^{(n)}$ et un système d'équations linéaires pour obtenir les $f_{i,j}$. L'idée générale est de décomposer le trajet de i à j en n étapes comme un trajet de i à k en une étape, puis un trajet de k à j en $(n-1)$ étapes. On obtient alors, grâce au caractère markovien :

$$\begin{aligned}
f_{i,j}^{(1)} &= p_{i,j} \\
f_{i,j}^{(n)} &:= \sum_{k\neq j} P\left(X_1 = k\,|X_0 = i\right) f_{k,j}^{(n-1)}\text{ , pour } n \geqslant 1 \\
&= \sum_{k\neq j} p_{i,k} f_{k,j}^{(n-1)}\text{ , pour } n \geqslant 1
\end{aligned} \tag{13}$$

pour tout $i, j \in E$. En appliquant la même idée aux probabilités de transition en un nombre quelconque d'étapes, on obtient :

$$\begin{aligned}
f_{i,j} &:= \sum_{n\in\mathbb{N}} f_{i,j}^{(n)} = f_{i,j}^{(1)} + \sum_{n\geqslant 2} f_{i,j}^{(n)} = f_{i,j}^{(1)} + \sum_{n\geqslant 2}\sum_{k\neq j} p_{i,k} f_{k,j}^{(n-1)} \\
&= p_{i,j} + \sum_{k\neq j} p_{i,k} \sum_{n\geqslant 2} f_{k,j}^{(n-1)} = p_{i,j} + \sum_{k\neq j} p_{i,k} \sum_{m\geqslant 1} f_{k,j}^{(m)} \\
&= p_{i,j} + \sum_{k\neq j} p_{i,k} f_{k,j}
\end{aligned} \tag{14}$$

pour tout $i, j \in E$. Les équations (14) constituent un système linéaire dont les inconnues sont les probabilités $f_{i,j}$, au nombre de N^2 si E est de dimension finie N. Nous pouvons maintenant définir les notions d'états récurrents et transitoires, pour des chaînes finies ou non.

Définition 6 (états récurrents et transitoires) Soit $\{X_n\}_{n\in\mathbb{N}}$, une CMTD homogène sur un espace d'état E discret. Un état $j \in E$ est *transitoire* si et seulement si

$$f_{j,j} < 1$$

Il est dit *récurrent* dans le cas contraire, c'est-à-dire lorsque

$$f_{j,j} = 1$$

Un état récurrent j est dit *récurrent non nul* si

$$M_j < \infty$$

Dans le cas contraire, il est dit *récurrent nul*

Un état récurrent est donc tel que la chaîne y revient presque sûrement (i.e. avec un probabilité 1). Dans le cas d'une chaîne finie, le temps moyen de retour est alors nécessairement fini (états récurrents non nuls) et le nombre moyen de retours est infini. Par contre, dans le cas d'une chaîne infinie, on peut imaginer des situations où la probabilité de retour est égale à 1, mais où l'espérance du temps de retour est infinie. On montre facilement les propriétés suivantes :

(i) Dans une CMTD irréductible (ou dans une classe d'équivalence), tous les états sont de même nature : ils sont tous, simultanément, transitoires, récurrents nuls ou récurrents non nuls

(ii) Dans une CMTD finie (i.e. lorsque l'espace d'état E est fini), si j est un état récurrent, alors (partant de j) la chaîne repassera nécessairement une infinité de fois en j. A contrario, si j est transitoire, elle n'y repassera qu'un nombre fini de fois.

(iii) Tous les états d'une CMTD irréductible finie sont récurrents non nuls

(iv) Dans une CMTD finie, il y a au moins un état récurrent (donc au moins une classe dont tous les états sont récurrents

(v) Dans une CMTD finie, la chaîne se retrouve nécessairement, après un nombre fini d'étapes, dans l'ensemble des états récurrents

Exercice 2 (Le labyrinthe ... suite)
On considère le modèle markovien du labyrinthe (voir figure 1).

1. Nous voulons répondre à la question suivante : *Supposons que la souris est dans la pièce 2. Combien de déplacement fera-t-elle en moyenne pour revenir à la pièce 2 ?*
   (a) Calculer les probabilités $f_{2,2}^{(n)}$ à l'aide de la formule de récurrence (13)
   (b) Calculer ces mêmes probabilités $f_{2,2}^{(n)}$ de manière plus directe, en raisonnant à partir de la topologie du le graphe de la figure 1

(c) Calculer le temps moyen de retour en 2 :

$$M_2 = \sum_{n \geqslant 1} n f_{2,2}^{(n)}$$

(d) Vérifier le résultat obtenu par échantillonnage (programme **R**)

2. Nous voulons répondre à la question suivante : *Supposons que la souris est au départ dans la pièce 2. Quelle est la probabilité qu'elle trouve la sortie du labyrinthe ?*

   (a) Calculer la probabilité $f_{2,6}$ d'aller de l'état 2 à l'état 6 en un nombre quelconque d'étape à l'aide de la formule (14)

   (b) Vérifier le résultat obtenu par échantillonnage (programme **R**)

3. Nous voulons répondre à la question suivante : *Sachant que la souris est initialement dans la pièce 2, quel est le nombre moyen de passages qu'elle fera dans la pièce 3, avant de sortir en 6 ou d'être piégée dans la pièce 5 ?*

   (a) Parmi les états de la CMTD associée à l'exemple 1, lesquels sont récurrents et lesquels sont transitoires ?

   (b) Soient i, j deux états transitoires. Soit

   $$R_{i,j} := \sum_{n \geqslant 1} n \cdot P\left(n \text{ passages par } j \,|X_0 = i\right)$$

   le nombre moyen de passages à un état j, partant d'un état i. En décomposant l'évènement $\{n \text{ passages par } j\}$ en trois évènements succesifs :

   $$\{\text{aller de } i \text{ à } j\} \cap \{\text{revenir } n \text{ fois en } j\} \cap \{\text{ne plus revenir en } j\}$$

   montrer que

   $$R_{i,j} = \frac{f_{i,j}}{1 - f_{j,j}} \tag{15}$$

   (c) Calculer $R_{2,3}$, le nombre moyen de passage par la pièce 3 en partant de la pièce 2. Vérifier le résultat obtenu par échantillonnage (programme **R**)

4. Nous voulons répondre à la question suivante : *Quel est le temps moyen de séjour d'un rat dans le labyrinthe, lorsqu'il est initialement dans la pièce 2 ?* Il s'agit d'un problème de calcul de temps moyen de séjour dans la sous-chaîne $\bar{F} := \{1, 2, 3, 4\}$ ou, de manière équivalente, de temps d'entrée dans la sous-chaîne absorbante $F := \{5, 6\}$. Ce temps moyen dépend de l'état initial dans $\bar{F}$. On notera :

   $$\mathbf{T}_i := E\left(T_F^{\{X_0=i\}}\right) = E\left(T_F \,|X_0 = i\right) \;, i \in \bar{F} := E \backslash F$$

   l'*espérance du temps d'entrée* dans F (ou du temps de séjour dans $\bar{F}$), *partant de l'état initial* $X_0 = i \in \bar{F}$.

(a) Ecrire l'équation liant $\mathbf{T}_1$ et $\mathbf{T}_2$ qui traduit l'observation suivante (voir sur la figure 1) :

Partant de la pièce 1, le rat va dans la pièce 1 avec une probabilité $\frac{1}{2}$ ou dans la pièce 2 avec une probabilité $\frac{1}{2}$ également. Dans les deux cas, le déplacement aura pris un pas de temps. Une fois arrivé dans la pièce 1, il faudra encore au rat un temps moyen $\mathbf{T}_1$ pour sortir du labyrinthe, alors que dans la pièce 2, il lui faudra un temps moyen $\mathbf{T}_2$.

(b) Répéter le même raisonnement en supposant un départ dans chacun des quatre états de $\bar{F} = \{1, 2, 3, 4\}$. Montrer que les équations obtenues s'écrivent sous la forme

$$\begin{pmatrix} \mathbf{T}_1 \\ \mathbf{T}_2 \\ \mathbf{T}_3 \\ \mathbf{T}_4 \end{pmatrix} = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} + \begin{pmatrix} p_{1,1} & p_{1,2} & p_{1,3} & p_{1,4} \\ p_{2,1} & p_{2,2} & p_{2,3} & p_{2,4} \\ p_{3,1} & p_{3,2} & p_{3,3} & p_{3,4} \\ p_{4,1} & p_{4,2} & p_{4,3} & p_{4,4} \end{pmatrix} \begin{pmatrix} \mathbf{T}_1 \\ \mathbf{T}_2 \\ \mathbf{T}_3 \\ \mathbf{T}_4 \end{pmatrix} \tag{16}$$

(c) Calculer les temps moyens de séjour solution de (16). En déduire le temps moyen de séjour d'un rat dans le labyrinthe en supposant une distribution initiale de probabilité uniforme

$$\boldsymbol{\pi}_i^{(0)} = \frac{1}{4}, \ \forall i \in \bar{F} = \{1, 2, 3, 4\}$$

Vérifier le résultat obtenu par échantillonnage (programme **R**)

Exercice 3 On étudie la propagation d'une maladie dans une population. Une personne peut être dans 4 états : saine et non vaccinée (état 0), vaccinée ou immunisée (état 1), inféctée (état 2) et morte (état 3). On suppose que :

(i) si une personne est malade, elle a 10 % de "chance" d'être morte la semaine d'après. Le tiers des survivants sera vacciné ou guéri et immunisé la semaine suivante. Les deux tiers restant seront toujours malades (i.e. infectés) et non vaccinés la semaine suivante.

(ii) Si une personne est vaccinée, elle le reste à vie, et elle ne peut tomber malade ou mourir.

(iii) Si une personne est saine une semaine, la semaine d'après, elle tombe malade dans 15 % des cas, va se faire vacciner dans 20% des cas et reste saine dans les autres cas.

1. Modéliser cette situation par une chaine de Markov dont vous donnerez la matrice de transition et le graphe associé. Vous serez précis sur la définition des grandeurs considérées en précisant notamment, le cas échéant, à quels instants elles sont calculées.
2. Supposons qu'à la semaine 0, la population comporte 100% de personnes saines, non vaccinées. Un mois plus tard (i.e. cinq semaines),

on se demande quelle sont les proportions de personnes vaccinées, de malades, saines et de mortes dans la population. Calculer la distribution de probabilité correspondante. Vérifier ensuite le résultat obtenu par échantillonnage (programme **R**)

3. Montrez qu'au bout d'un certain temps, toute personne est nécessairement morte ou vaccinée. Donnez la valeur théorique du temps moyen nécessaire pour qu'une personne soit morte ou vaccinée. Vérifier le résultat obtenu par échantillonnage (programme **R**)
4. Robert est sain. Quelle est la probabilité qu'il meure de la maladie ? Faites le calcul théorique, puis vérifiez le résultat obtenu par échantillonnage (programme **R**)

4 Comportement sur le long terme d'une CMTD

4.1 Périodicité dans une CMTD

Définition 7 (Périodicité) Soit $\{X_n\}_{n\in\mathbb{N}}$ une CMTD homogène sur un espace d'état E discret. L'état $j \in E$ est *périodique* si et seulement si, partant de cet état j, il n'est possible d'y revenir qu'après un nombre d'étapes multiple de $k > 1$, c'est-à-dire si et seulement si

$$\exists k > 1 : p_{j,j}^{(m)} = 0 \,,\, \forall m \notin \{kp \,|p \in \mathbb{N}\}$$

Le nombre k est appelé la *période* de l'état j. La période d'une CMTD est le plus grand commun diviseur (PGCD) des périodes de chacun de ses états. Une CMTD est dite périodique si sa période est strictement supérieure à 1.

Tous les états d'une même classe d'équivalence ont nécessairement la même période. Par exemple, dans la chaîne de gauche sur la figure 2, l'état 3 ne peut être périodique puisqu'il y a un terme de rebouclage direct (cycle de longueur 1). Donc aucun des états de cette chaîne ne peut être périodique puisqu'il n'y a qu'une classe d'équivalence. Les états de la chaîne de droite ne peuvent pas non plus être périodiques. En effet, il existe des cycles de longueurs 2, 3, 5, 7, 8, etc. permettant de revenir à l'état 3. Aucun des états de cette chaîne irréductible ne peut donc être périodique. Dans la chaîne du milieu, l'état 1 est périodique de période 3 (on peut y revenir en 3, 6 , 9 , etc. étapes). Donc tous les états cette chaîne sont périodiques de période 3. La chaîne elle même est donc irréductible et périodique de période 3. Lorsqu'on observe une trajectoire (réalisation) particulère de la chaîne 3, par exemple

$$1, 3, 4, 1, 3, 4, 2, 3, 4, 1, 3, 4, 2, 3, 4, 2, 3, 4, \ldots$$

on ne reviendra à un état donné qu'après un nombre de pas de temps multiple de 3. Par contre, la suite des réalisations n'est pas une suite périodique.

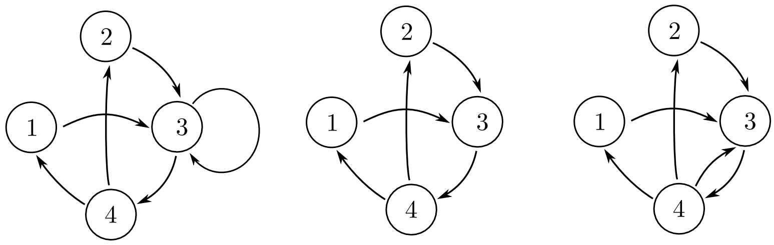

FIG. 2: Trois exemples de CMTD pour illustrer la notion de péridocité

4.2 Régime stationnaire et distribution limite

On s'intéresse au comportement asymptotique (lorsque $n \to \infty$) du vecteur des probabilités d'état $\boldsymbol{\pi}^{(n)}$.

Proposition 1 (Distribution limite et distribution stationnaire) Soit $\{X_n\}_{n\in\mathbb{N}}$, une CMTD ***irréductible*** et ***apériodique*** définie sur un espace d'état discret E. Alors la distribution limite

$$\boldsymbol{\pi}^{(\infty)} := \lim_{n\to\infty} \boldsymbol{\pi}^{(n)}$$

existe. Elle est indépendante de la distribution initiale de probabilités (i.e. elle est la même, quelle que soit $\boldsymbol{\pi}^{(0)}$). Deux cas sont possibles :

(i) *tous les états sont transitoires ou récurrents nuls.* On a alors

$$\boldsymbol{\pi}_i^{(\infty)} = 0\,,\ \forall i \in E$$

(ii) *tous les états sont récurrents non nuls.* On a alors

$$\boldsymbol{\pi}^{(\infty)} = \boldsymbol{\pi}$$

où $\boldsymbol{\pi}$ est la *distribution stationnaire* de probabilités définie par :

$$\boldsymbol{\pi} = \boldsymbol{\pi}\mathbf{P} \tag{17}$$

avec la condition de normalisation

$$\|\boldsymbol{\pi}\|_1 := \sum_{i\in E} |\boldsymbol{\pi}_i| = \sum_{i\in E} \boldsymbol{\pi}_i = 1 \tag{18}$$

Nous admettrons ce résultat, en ce qui concerne l'existence. La valeur de la limite, lorsqu'elle existe, se montre facilement. Comme la chaîne est supposée irréductible, tous les états sont de même nature. Dans le cas où ils sont tous transitoires ou récurrents nuls, la chaîne ne reviendra plus dans un état $j \in E$, à partir d'un certain rang. La suite $\boldsymbol{\pi}_j^{(n)}$ est donc nulle à partir de ce rang. Dans le cas où tous les états sont récurrents non nuls, la récurrence (6) donne, par passsage à la limite ($n \to \infty$) à gauche et à droite, l'équation (17).

Remarque 1 Le vecteur $\boldsymbol{\pi}$ est le vecteur propre à gauche de $\mathbf{P}$ associé à la valeur propre 1, de norme $\|\boldsymbol{\pi}\|_1 = 1$. Lorsqu'il existe (il doit alors être non nul), on dit que la chaîne admet un *régime stationnaire*. La distribution stationnaire décrit le comportement sur le long terme de la CMTD. Elle ne dépend pas de la distribution initiale de probabilité. Dans le cas d'une chaîne périodique, si la distribution stationnaire existe, $\boldsymbol{\pi}_j$ est la proportion du temps que la chaîne passe dans l'état j, en moyenne, lorsque $n \to \infty$, pour une distribution uniforme de probabilités initiales.

Exercice 4 On considère la CMTD irréductible et apériodique représentée à la figure 3.

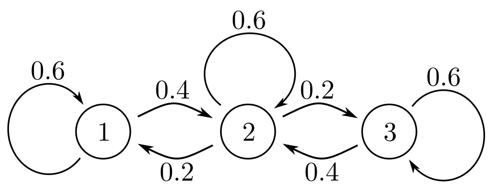

Fig. 3: CMTD irréductible et apériodique considérée dans l'exercice 4

1. Calculer la distribution stationnaire de probabilités (avec (17) et (18))
2. La limite $\boldsymbol{\pi}^{(\infty)}$ existe-t-elle ? Si oui, que vaut-elle ?
3. Calculer les probabilités d'état au temps n, pour $n \in \{0, \dots, 20\}$, pour les trois conditions initiales $X_0 = 1$, $X_0 = 2$ et $X_0 = 3$
4. Commenter l'influence des conditions initiales lorsque n augmente

Exercice 5 (Pagerank)
On examine dans cet exercice le fonctionnement de l'algorithme *Pagerank* de Google. On note N le nombre de pages internet existantes, on note r_i le nombre de liens existant sur la page i. On modélise l'attitude d'un surfeur de la façon suivante : une fois visitée la page i, soit il choisit une adresse web au hasard (avec la probabilité p), soit il décide de visiter une des pages référencées sur la page i (avec la probabilité $1 - p$). Dans le cas où il choisit une adresse au hasard, on considère que chaque page à la même probabilité $\frac{1}{N}$ d'être choisie. Dans le cas où le surfeur choisit un des liens de la page i, il le fait également avec une distribution uniforme. Dans ce deuxième cas, chaque page a donc une probabilité $\frac{1}{r_i}$ d'être choisie. Le web considéré dans cet exercice est un web miniature constitué de 5 pages et représenté à la figure 4.

1. Modéliser la promenade du surfeur par une CMTD dont on donnera le graphe et la matrice de transition.
2. Pourquoi existe-t-il nécessairement une loi limite $\boldsymbol{\pi}^{(\infty)}$? Comment peut-on la calculer ?

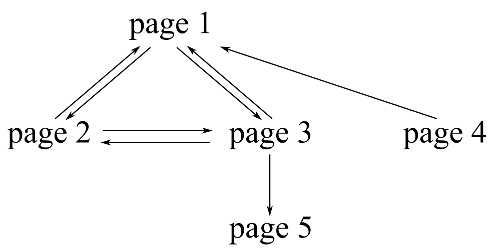

Fig. 4: Le Web simplifié considére à l'exercice 5

3. Soit $\boldsymbol{\pi} = (\boldsymbol{\pi}_1, \boldsymbol{\pi}_2, \ldots, \boldsymbol{\pi}_N)$ la distribution stationnaire de probabilité. La mesure *Pagerank* de la page i est alors simplement définie par

$$\text{Pagerank}(i) := \boldsymbol{\pi}_i$$

Calculer le *Pagerank* pour toutes les pages du Web de la figure 4, successivement pour $p = 0.15$, $p = 10^{-2}$ et $p = 0.99$.

4. Vérifier les résultats obtenus par échantillonnage (programme **R**)

Remarque 2 Dans le cas du www, les principes de modélisation et la définition de la mesure *pagerank* restent les mêmes. Les méthodes de modélisation, d'identification et de calcul, par contre, sont adaptées à des CMTD de très grande taille (en l'occurrences des dizaines de milliards de pages). On pourra pour une introduction rapide au sujet se reporter par exemple à la page Wikipedia en anglais sur le Pagerank.