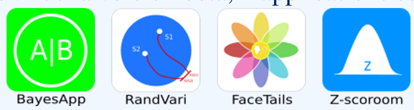
Chaines de Markov

Diana Nurbakova

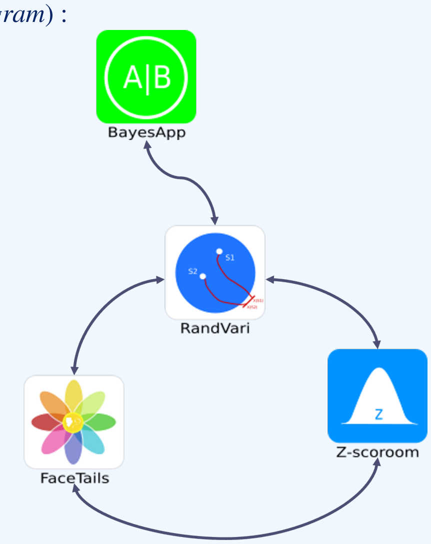
Contents

Scénario : Applications mobiles

Sur un téléphone mobile “*markovien*” de la version beta, 4 applications sont installées :



Il existe des connexions entre les applications: depuis une application, une autre peut être lancée (e.g. partage d’une photo via un messanger). Ces connexions peuvent être présentées sous la forme suivante (**diagramme de transition**, en. *transition diagram*) :



Notons que pas toutes les applications sont liées directement : *BayesApp* est liée avec *RandVari* et cette dernière est liée avec toutes les applications.

On suppose que l’utilisation des applications est soumise aux contraintes suivantes :

1. Le lancement d’une application à partir d’une autre ne peut se faire que dans le sens de connexion existante
2. L’historique des lancements n’est pas gardée. Une fois qu’une application est lancée, une autre peut être lancée à partir de cette dernière d’une manière aléatoire. On est au courant que de l’application en cours, pas les précédentes.

Par example, si on commence par *BayesApp*, la seule application qui peut être lancée est *RandVari*. Une fois *RandVari* est lancée, le choix se fait d’une manière aléatoire parmi 3 applications car il existe des connexions avec toutes les autres applications. Imaginons que c’est *FaceTails* qui était lancée par la suite. Maintenant, le choix est entre *RandVari* et *Z-scoroom*.

Ainsi à chaque étape, le système “décide” de sa prochaine étape d’une manière aléatoire en fonction de son état en cours. Il est donc possible de prédire quelle application va être lancée après.

Quelle application va être la plus utilisée ?

Chaînes de Markov : définitions de base

L’exemple décrit dans le scénario présente l’idée principale de *chaines de Markov*.

Une **chaîne de Markov** (en. *Markov chain*) est un modèle stochastique qui décrit une séquence d’évènements où la probabilité de l’évènement suivant ne dépend que de l’évènement en cours (présent).

Plus formellement :

Soit $(\Omega, \mathcal{A}, \mathbb{P})$ un espace probabilisé, $E = \{1, 2, \ldots, N\}$ un ensemble fini appelé l'**espace d'états**, $\mathcal{E}$ l'ensemble des parties de E. Une **chaîne de Markov discrète** (en. *discrete Markov chain*) est une séquences $(X_k)_{k\in\mathbb{N}}$ de v.a.r. à valeurs dans l'espace de l'états $(E, \mathcal{E})$ telle qu'elle vérifie $\forall n \in \mathbb{N},\ \forall (e_0, \ldots, e_n, e_{n+1}) \in E^{n+2}$ la *propriété de Markov* (en. *Markov property*) :

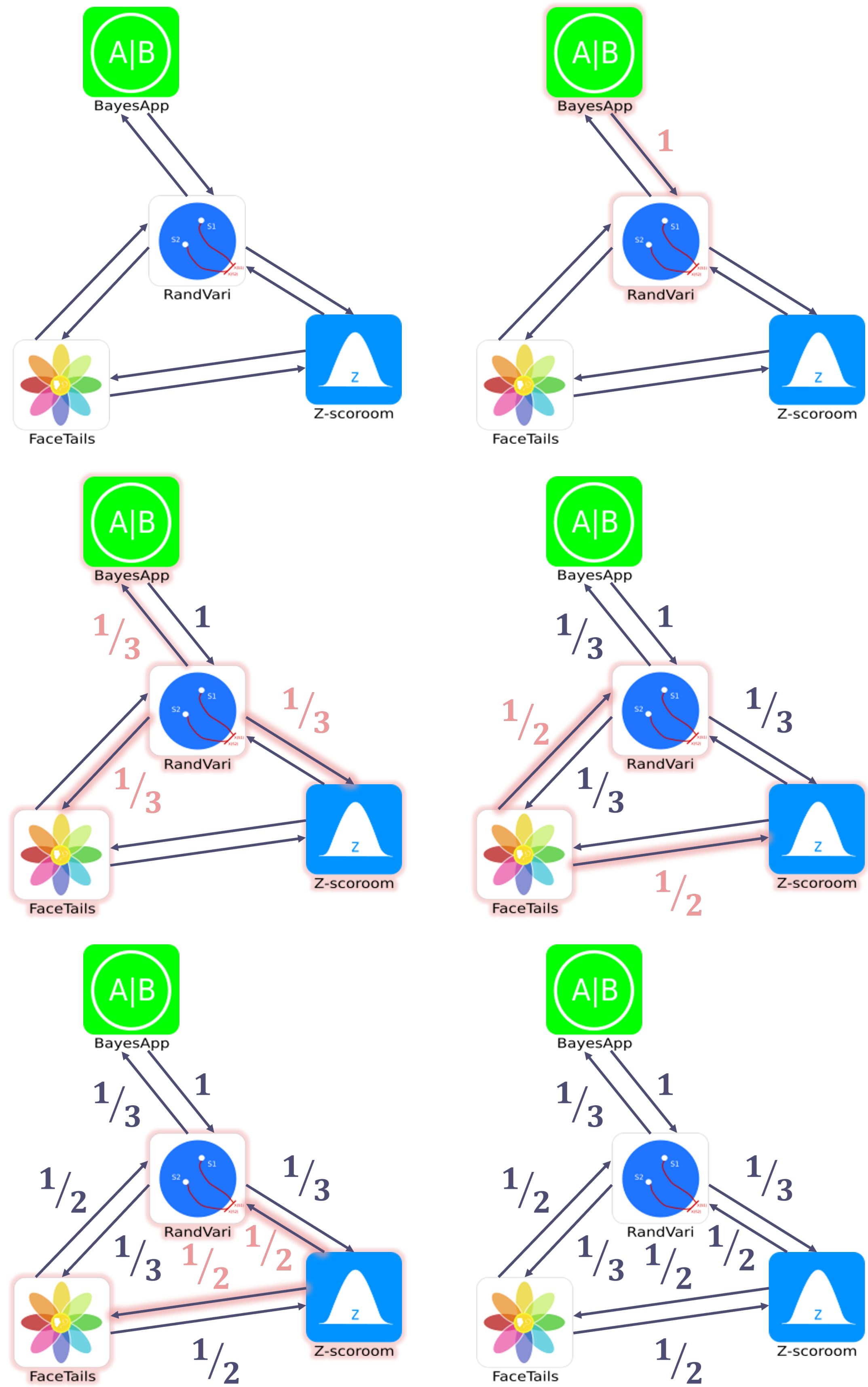
$$\mathbb{P}\left(X_{n+1} = e_{n+1} | X_n = e_n, \ldots, X_0 = e_0\right) = \mathbb{P}\left(X_{n+1} = e_{n+1} | X_n = e_n\right)$$

Les termes de *perte de mémoire* et *sans mémoire* (en. *memorylessness*) sont parfois utilisés pour faire références à la propriété de Markov.
La valeur $X_k,\ \forall k \in \mathbb{N}$ est *l'état de la chaîne de Markov à l'instant (étape) k*. Le passage d'un état à l'autre est appelé une **transition** (en. *transition*).

On suppose qu'à chaque instant, le système (le processus) se trouve dans un des états possibles. Par convention, on suppose que le processus de ne termine pas.

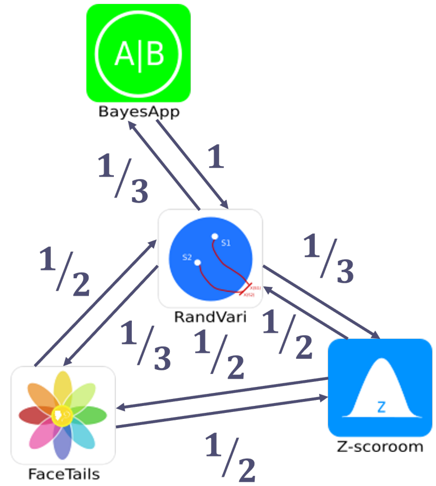
Transitions

Reprenons l'exemple initiale. Les connexions entre deux applications (états de la chaîne de Markov) ont une direction exprimée par une flèche. Estimons les probabilités de passage par chaque flèche, en supposant que les options sont équiprobables.



A|B
BayesApp
RandVari
Z
Z-scoroom
FaceTails
A|B
BayesApp
1
RandVari
Z
Z-scoroom
FaceTails
A|B
BayesApp
1/3
1
1/3
RandVari
1/3
Z
Z-scoroom
FaceTails
A|B
BayesApp
1/3
1
1/2
1/3
RandVari
1/3
Z
Z-scoroom
FaceTails
1/2
A|B
BayesApp
1/3
1
1/2
1/3
RandVari
1/3
1/2
1/2
Z
Z-scoroom
FaceTails
1/2
A|B
BayesApp
1/3
1
1/2
1/3
RandVari
1/3
1/2
1/2
Z
Z-scoroom
FaceTails
1/2

Ainsi, nous obtenons :



Commençons par *BayesApp*. Il existe qu'une seule option de transition : *BayesApp* → *RandVari*. Donc la probabilité de cette direction est 1. L'état *RandVari* a 3 directions sortantes. Supposant l'équiprobabilité de chacune des direction, la probabilité de chacune est alors de $1/3$. Pour *FaceTails*, il existe 2 options avec la probabilité $1/2$ chacune. Pareil pour *Z-scoroom*.

Le diagramme construit ainsi est appelé le **diagramme de transitions** ou **diagramme sagittal** (en. *transition diagram*). Il liste tous les états possibles de la chaîne de Markov aussi que les probabilités de transition entre ces états. Notons que les transitions possibles sont affichées, i.e. celles à probabilité de transition strictement positive.

D'une manière plus formelle :

Soit $(X_k)_{k\in\mathbb{N}}$ une chaîne de Markov sur $(\Omega, \mathcal{A}, \mathbb{P})$ à valeurs dans $(E, \mathcal{E})$. Pour $(m, n) \in \mathbb{N}^2$, $n > m$, on appelle la **probabilité de transition** (en. *transition probability*) ou **probabilité de passage** de l'état e_m à l'état e_n en $n - m$ étapes, la probabilité conditionnelle :

$$\mathbb{P}(X_n = e_n | X_m = e_m)$$

Parfois, la notation $p_{m,n}$ est utilisée : $p_{m,n} = \mathbb{P}(X_n = e_n|X_m = e_m)$. Pour indiquer le nombre d'étapes (de pas), un indice exposant peut être utilisé, i.e. $p_{m,n}^{(k)}$ pour désigner la probabilité de transition entre l'état m et n en k étapes.
Au lieu de nombre d'étapes $(n - m)$, le terme *l'intervalle de temps de* m *à* n peut être utilisé.
Afin de simplifier les notations, souvent les évènements sont numérotés : $E = \{1, 2, \dots, n\}$. On peut réécrire la définition comme suit : soit $(i, j) \in E^2$, alors la probabilité de transition de i à j en $n - m$ étapes est : $p_{i,j} = \mathbb{P}(X_n = j|X_m = i)$.

Selon les conditions imposées aux probabilités de transition, on peut distinguer certains types de chaîne de Markov.

Une chaîne de Markov $X = (X_k)_{k\in\mathbb{N}}$ est dite **homogène** (en. *time-homogeneous*), si pour $\forall n \in \mathbb{N},\ \forall (i, j) \in E^2$ les probabilités de transition sont indépendantes de l'instant n, i.e. :

$$\mathbb{P}(X_{n+1} = j|X_n = i) = \mathbb{P}(X_n = j|X_{n-1} = i)$$

Ou dans le cas plus général, $\forall l \in \mathbb{N},\ \forall n \in \mathbb{N},\ l < n,\ \forall (i, j) \in E^2$:

$$\mathbb{P}(X_n = j|X_{n-l} = i) = \mathbb{P}(X_l = j|X_0 = i)$$

Une chaîne de Markov $X = (X_k)_{k\in\mathbb{N}}$ est dite **stationnaire** (en. *stationary*), si $\forall n \in \mathbb{N},\ \forall k \in \mathbb{N}$:

$$\mathbb{P}(X_0 = e_0, X_1 = e_1, \ldots, X_k = e_k) = \mathbb{P}(X_n = e_0, X_{n+1} = e_1, \ldots, X_{n+k} = e_k)$$

Notons que toute chaîne de Markov stationnaire est homogène.

En sachant les probabilités de transition, on peut faire des prédictions.

Exemple Considérons l'exemple suivant.

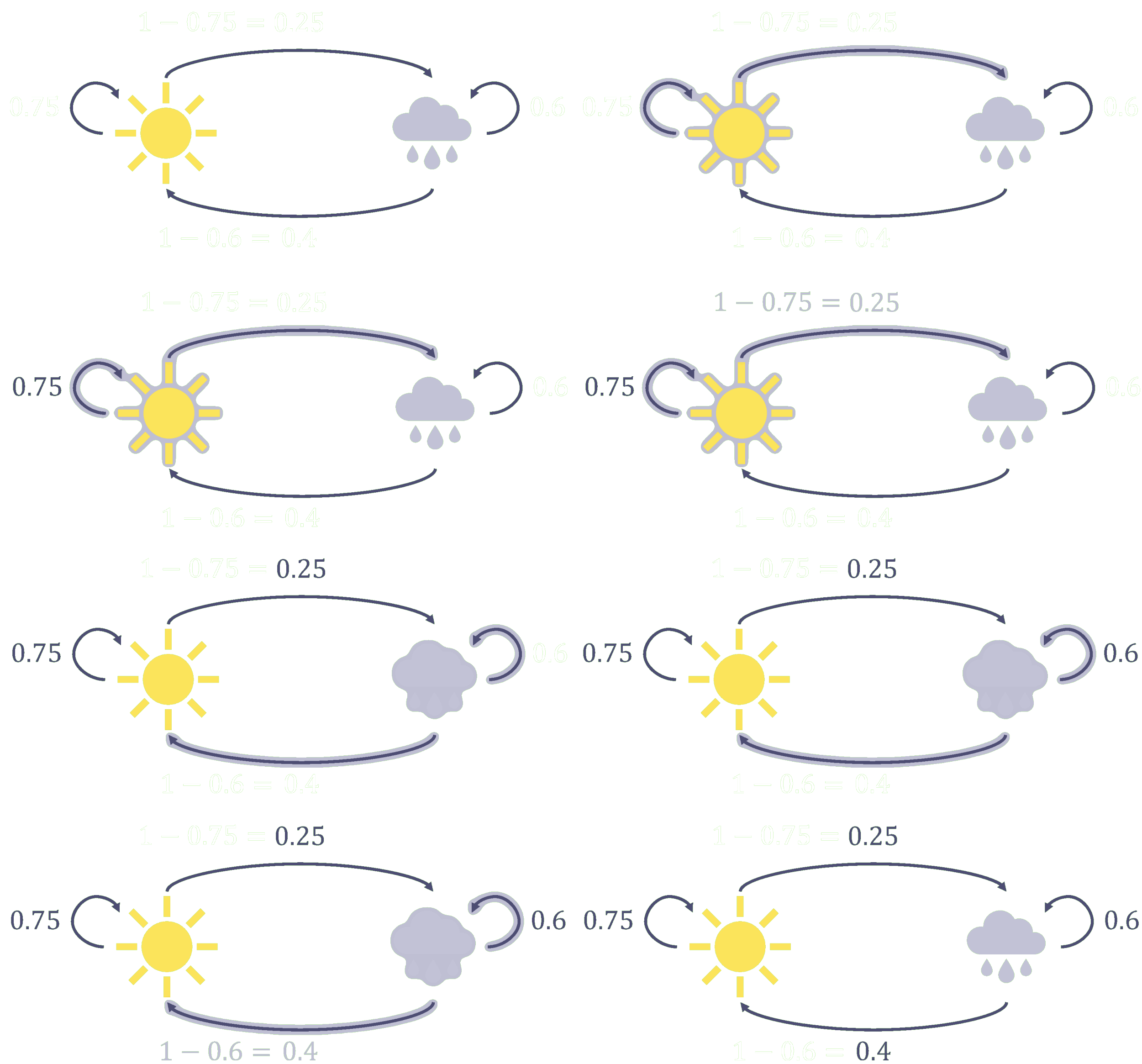
A Sunville, il fait soit beau, soit il pleut. Nous savons que :

- dans 75% de cas, un jour ensoleillé est suivi par un autre jour ensoleillé
- dans 60% de cas, un jour pluvieux est suivi par un autre jour pluvieux

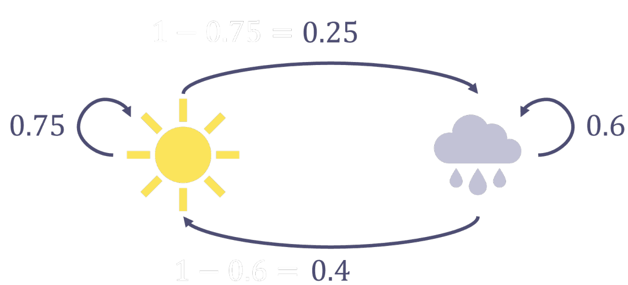
S'il fait beau aujourd'hui, quelle est la probabilité qu'il fera beau dans 2 jours ?

Solution.

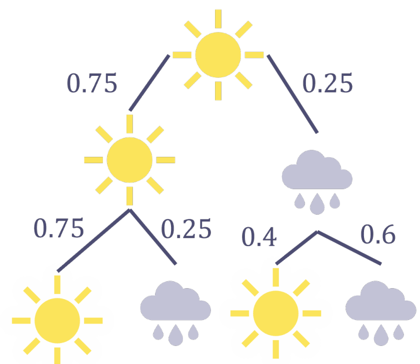
Commençons par construire le diagramme de transistion.



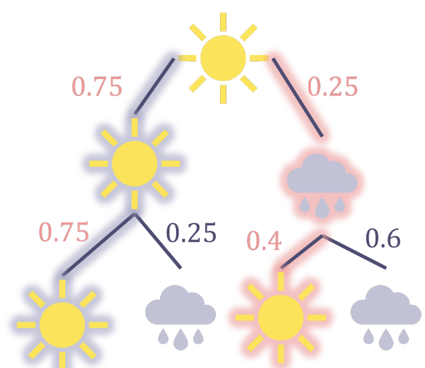
Ainsi, nous obtenons :



Maintenant, on peut faire des prédictions. Pour cela, on va redessiner le diagramme de transisiton sous forme du diagramme en arbre pour plus de visibilité :



Remarquons qu'il existe 2 chemins ramenant au jour ensoleilé dans 2 jours :



1. jour ensoleilé → jour ensoleilé → jour ensoleilé
2. jour ensoleilé → jour pluvieux → jour ensoleilé

Donc la probabilité d'avoir un jour ensoléilé dans 2 jours, sachant qu'aujourd'hui il fait beau :
$\mathbb{P}(X_2 = soleil|X_0 = soleil) = 0.75 \times 0.75 + 0.25 \times 0.4 = 0.5625 + 0.1 = 0.6625$

Comment gérer les transition en plusieurs étapes sans passer par le diagramme en arbre ?

Matrice de transition

Si l'espace d'états est fini, la distribution de la probabilité de transition peut être présentée sous forme d'une *matrice de transition*.

On appelle **matrice de transition** (en. *transition matrix*) d'une chaîne de Markov homogène $X = (X_k)_{k\in\mathbb{N}}$ la matrice $G = (p_{i,j})_{i,j\in E} \in \mathcal{M}_N(\mathbb{R})$, où $p_{i,j}$ est la probabilité de transition entre l'état i à l'état j, i.e. :

$$p_{i,j} = \mathbb{P}(X_1 = j|X_0 = i)$$

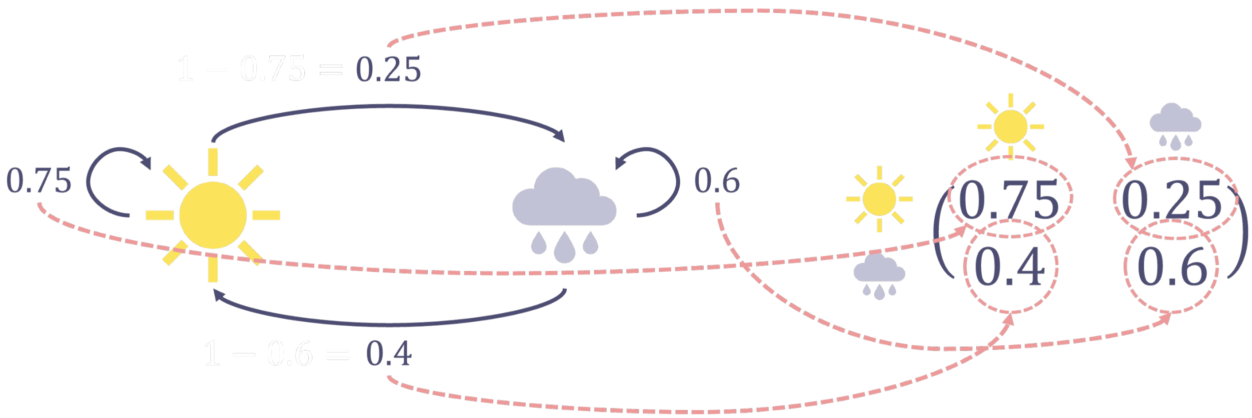
Notons que la matrice G est une matrice stochastique (en. *right stochastic matrix*) vérifiant :
1. une matrice carrée
2. chaque élément représente une probabilité $\forall(i,j) \in E^2,\ p_{ij} \geq 0$
3. la somme des éléments de chaque ligne vaut 1 : $\forall i \in E,\ \sum_{j\in E} p_{ij} = 1$

On peut dire que les lignes de la matrice de transition correspondent au *présent* et les colonnes au *futur*.

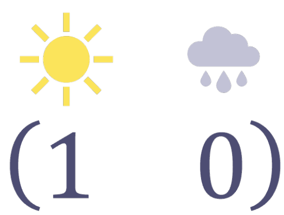
Reprenons l'exemple précédent. Remarquons qu'il s'agit de deux états : jour ensoleillé et jour pluvieux.

La matrice de transition est donnée par :

$$G = \begin{pmatrix} 0.75 & 0.25 \\ 0.4 & 0.6 \end{pmatrix}$$



Selon l'énoncé, "on part" du jour ensoleillé *s'il fait beau aujourd'hui,. . . .* On peut l'interpréter en termes d'états de la manière suivante : l'état initiale du système consiste à *jour ensoleill* : *Oui*, *jour pluvieux* : *Non*.



On peut l'écrire sous forme d'un vecteur ligne stochastique (en. *stochastic vector*) dont tous les éléments sont non-négatifs et leur somme vaut 1 :

$$S_0 = \begin{pmatrix} 1 & 0 \end{pmatrix}$$

Trouvons l'étape 1 :

$$S_1 = S_0 \cdot G = \begin{pmatrix} 1 & 0 \end{pmatrix} \cdot \begin{pmatrix} 0.75 & 0.25 \\ 0.4 & 0.6 \end{pmatrix} =$$

$$= \begin{pmatrix} 1 \times 0.75 + 0 \times 0.4 & 1 \times 0.25 + 0 \times 0.6 \end{pmatrix} = \begin{pmatrix} 0.75 & 0.25 \end{pmatrix}$$

Dans le R :

```
# état initial
s0 <- c(1, 0); s0
```

```
## [1] 1 0
```

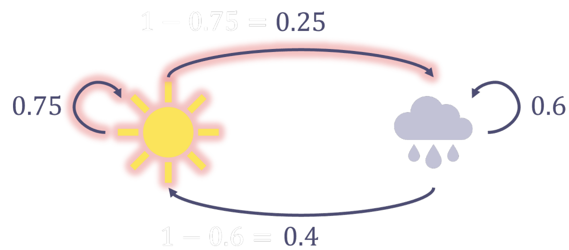
```
# matrice de transition
G <- matrix(c(0.75, 0.25, 0.4, 0.6), nrow=2, ncol=2, byrow=TRUE); G
```

```
##      [,1] [,2]
## [1,] 0.75 0.25
## [2,] 0.40 0.60
```

```
# multiplication
s1 <- s0 %*% G; s1
```

```
##      [,1] [,2]
## [1,] 0.75 0.25
```

On se retrouve dans l'état 1 avec la probabilité de 0.75 de se retouver avec un jour ensoleillé contre la probabilité de 0.25 d'un jour pluvieux :



Maintenant, si on veut savoir l'étape 2, on peut procéder d'une manière similaire :

$$S_2 = S_1 \cdot G = \begin{pmatrix} 0.75 & 0.25 \end{pmatrix} \cdot \begin{pmatrix} 0.75 & 0.25 \\ 0.4 & 0.6 \end{pmatrix} = \begin{pmatrix} 0.75 \times 0.75 + 0.25 \times 0.4 & 0.75 \times 0.25 + 0.25 \times 0.6 \end{pmatrix} =$$

$$= \begin{pmatrix} 0.5625 + 0.1 & 0.1875 + 0.15 \end{pmatrix} = \begin{pmatrix} 0.6625 & 0.3375 \end{pmatrix}$$

Vérifions avec R :

```r
# étape 2
s2 <- s1 %*% G; s2
```

```
##        [,1]   [,2]
## [1,] 0.6625 0.3375
```

Si ce qui nous intéresse est la probabilité d'avoir un jour ensoleillé deux jours après un jour ensoleillé, elle vaut 0.6625.

Notons qu'on a trouvé la même valeur que passant par le diagramme en arbre. Cependant, la méthode matricielle est mieux généralisable.

Remarquons également que

$$S_2 = S_1 \cdot G = S_0 \cdot G \cdot G = S_0 \cdot G^2$$

Vérifions sur l'exemple :

$$G^2 = \begin{pmatrix} 0.75 & 0.25 \\ 0.4 & 0.6 \end{pmatrix} \cdot \begin{pmatrix} 0.75 & 0.25 \\ 0.4 & 0.6 \end{pmatrix} = \begin{pmatrix} 0.75 \times 0.75 + 0.25 \times 0.4 & 0.75 \times 0.25 + 0.25 \times 0.6 \\ 0.4 \times 0.75 + 0.6 \times 0.4 & 0.4 \times 0.25 + 0.6 \times 0.6 \end{pmatrix} =$$

$$= \begin{pmatrix} 0.5625 + 0.1 & 0.1875 + 0.15 \\ 0.3 + 0.24 & 0.1 + 0.36 \end{pmatrix} = \begin{pmatrix} 0.6625 & 0.3375 \\ 0.54 & 0.46 \end{pmatrix}$$

Alors :

$$S_0 \cdot G^2 = \begin{pmatrix} 1 & 0 \end{pmatrix} \cdot \begin{pmatrix} 0.6625 & 0.3375 \\ 0.54 & 0.46 \end{pmatrix} = \begin{pmatrix} 0.6625 & 0.3375 \end{pmatrix}$$

Dans le R :

```r
# G^2
G2 <- G %*% G
# étape 2
s0 %*% G2
```

```
##         [,1]   [,2]
## [1,] 0.6625 0.3375
```

Résumons : Afin de définir (caractériser) une chaîne de Markov, on va avoir besoin de :

- ensemble d'états
- matrice de transition
- distribution initiale (ou état initial) donné ou choisi d'une manière aléatoire.

Loi initiale d'une chaîne de Markov

Supposons qu'on connait une loi initiale de la chaîne de Markov qui correspond à la loi de la v.a.r. X_0 (aussi appelée *condition initiale de la chaîne de Markov*).

> Comment obtenir les distributions de X_1, X_2, etc. sachant la distribution de X_0 et la matrice de transition ?

Une chaîne de Markov homogène est caractérisée par sa matrice de transition et sa distribution initiale, i.e. la loi de X_0. Cette loi initiale peut être interprétée comme les probabilités de chaque état d'être l'état initial.

D'une manière formelle [1] :

Soit $X = (X_n)_{n\in\mathbb{N}}$ une chaîne de Markov homogène de matrice de transition G.
On désigne par Π la fonction de masse de la v.a.r. discrète X_0, appelée **distribution initiale** ou **loi initiale** (en. *initial distribution*) :

$$\Pi : \begin{array}{l} E \to [0,1] \\ k \to \pi_k = \mathbb{P}(X_0 = k) \end{array}$$

Ainsi, on définit un vecteur ligne $\pi = (\pi_1,\ \pi_2, \dots, \pi_N)$.
$\forall n \in \mathbb{N}$, la fonction de masse de la v.a.r. discrète X_n est une application :

$$\Pi^{(n)} : \begin{array}{l} E \to [0,1] \\ k \to \pi_k^{(n)} = \mathbb{P}(X_n = k) \end{array}$$

La distribution de X_n est donc donnée par un vecteur ligne $\pi^{(n)} = (\pi_1^{(n)}, \pi_2^{(n)}, \dots, \pi_N^{(n)})$ obtenu comme suit :

$$\begin{array}{l} \pi^{(n)} = \pi G^n \\ \pi^{(0)} = \pi \end{array}.$$

Notons que $\forall n \in \mathbb{N}$ et $(e_0, e_1, \dots, e_n) \in E^{n+1}$, le suivant est vrai :

$$\mathbb{P}(X_0 = e_0, \dots, X_n = e_n) = \pi_{e_0} G_{e_0,e_1} G_{e_2,e_2} \dots G_{e_{n-1},e_n}$$

Pour les démonstrations, voir [1].

Ainsi, on peut trouver les futurs états de la chaîne de Markov homogène via la puissance correspondante de la matrice de transition.

Soit $X = (X_k)_{k \in \mathbb{N}}$ une chaîne de Markov homogène de matrice de transition G [1]. Pour $n \in \mathbb{N}$, on désigne par $G^n = \underbrace{G \cdot G \cdot \ldots \cdot G}_{n \text{ fois}}$ la puissance n de la matrice G. La probabilité de transition en n étapes de l'état i à l'état j est donc donnée par :

$$\mathbb{P}(X_n = j | X_0 = i) = G^n_{i,j}$$

où $G^n_{i,j}$ est l'élément d'indice (i, j) de la matrice G^n.
Pour la démonstration voir [1].

Pour les chaîne de Markov homogènes l'**équation de Chapman-Kolmogorov** est vraie [1] :

Soit $X = (X_k)_{k \in \mathbb{N}}$ une chaîne de Markov homogène d'espace d'états E.
$\forall i, j \in E$, $\forall n, m \in \mathbb{N}$, le suivant est vrai :

$$\mathbb{P}(X_{m+n} = j | X_0 = i) = \sum_{k \in E} \mathbb{P}(X_m = j | X_0 = k) \times \mathbb{P}(X_n = k | X_0 = i)$$

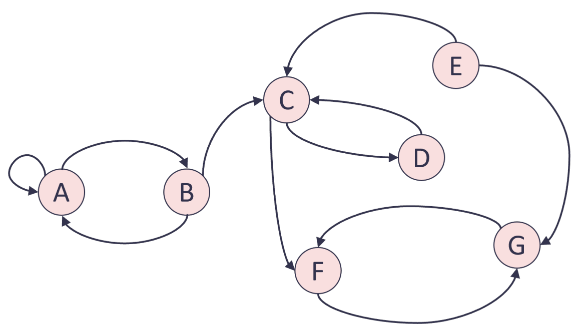
Pour la démonstration voir [1].

Classification des états d'une chaîne de Markov

Nous avons introduit les notions de *matrice de transition* et de *diagramme de transition*.

Une transition *possible* de l'état i à l'état j, notée $i \to j$ et illustrée par une flèche sur le diagramme de transition, signifies que la probabilité de transition entre i et j est positive pour un certain $n \in \mathbb{N}$, i.e. $\exists n \in \mathbb{N} : \ G^{(n)}_{ij} > 0$.

Considérons l'exemple suivant :



Pour chaque état, regardons à partir de quel état il est possible d'y "venir" en une transition (les flèches entrantes) :

- A : à partir des états A et B
- B : à partir de l'état A
- C : à partir des états B, D, E
- D : à partir de l'état C
- E : à partir d'aucun d'autre état
- F : à partir des états C et G
- G : à partir des états F et E

Soit $X = (X_n)_{n\in\mathbb{N}}$ une chaîne de Markov homogène sur l'espace E.
On dit que l'état $j \in E$ est **accessible** à partir de l'état $i \in E$ (en. *accessible from*), noté $i \to j$, si la probabilité de transition est positive pour un certain $n \in \mathbb{N}$, i.e.

$$\exists n \in \mathbb{N} : \; G_{ij}^{(n)} > 0$$

On considère que chaque état est accessible à partir de lui-même.

Notons que dans notre example il existe des couples d'états qui sont accessibles les uns depuis les autres :

- A et B : $A \to B$ et $B \to A$
- C et D : $C \to D$ et $D \to C$
- F et G : $F \to G$ et $G \to F$

On dit que l'état i et l'état j **communiquent** (en. *communicate*), noté $i \leftrightarrow j$, si chacun d'eux est accessible à partir de l'autre : $i \to j$ et $j \to i$.

La relation de communication $i \leftrightarrow j$ est une relation d'**équivalence** (en. *equivalence*), c'est-à-dire :

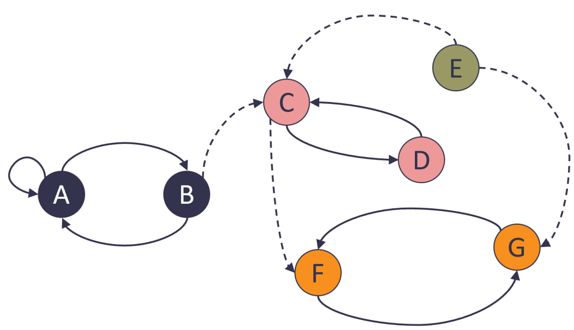
- *réflexive* : $i \leftrightarrow i$
- *symétrique* : si $i \leftrightarrow j$, alors $j \leftrightarrow i$
- *transitivité* : si $i \leftrightarrow j$ et $j \leftrightarrow k$, alors $i \leftrightarrow k$

Ceci permet de regrouper les états d'une chaine de Markov, en construisant ainsi les partitions des états de la chaîne de Markov, appelées les **classes d'équivalence** (en. *communicating classes*). Deux états i et j appartiennent à la même classe si et seulement si $i \leftrightarrow j$, et deux états appartenant à deux classes différentes ne communiquent pas.

Notons que la relation d'accessibilité (être accessible) s'étend aux classes d'équivalence.

Trouvons les classes d'équivalence dans notre exemple :

- $\{A, B\}$
- $\{C, D\}$
- $\{E\}$
- $\{F, G\}$



Une chaîne de Markov est dite **irréductible** (en. *irreducible*), si pour elle existe qu'une seule classe (tous les états communiquent entre eux), c'est-à-dire tous les états communiquent :

$$\forall (i, j) \in E^2, \; \exists n \in \mathbb{N} : G_{ij}^{n} > 0$$

En regardant notre exemple, nous remarquons qu'on peut distinguer deux types de classes :

1. en rentrant dans une classe, on reste dans cette classe sans possibilité d'en sortir : e.g. la classe $\{F, G\}$

2. il est possible de sortir d'une classe mais une fois sorti, il n'est plus possible d'y retourner : les classes $\{A, B\}$, $\{E\}$, $\{C, D\}$.

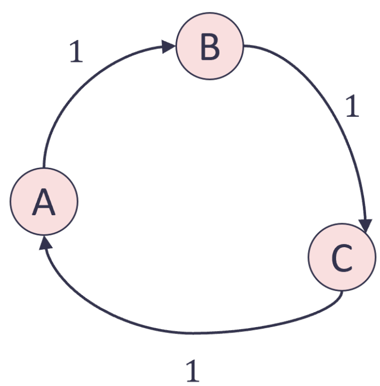
Plus formellement :

Une classe est dite **récurrente** ou **finale** (en. *recurrent state*) si elle ne conduit à aucune d'autre, autrement dit, il est impossible de la quitter. Les états d'une classe récurrente sont appelés *récurrents*.
Un état i est dit **absorbant** (en. *absorbing state*), s'il est impossible de le quitter (une classe recurrente composée d'un seul état), i.e. :

$$\forall k \neq i,\ G_{ik} = 0,\ G_{ii} = 1$$

Une classe est dite **transitoire** ou **transiente** (en. *transient*), si la probabilité d'y retourner est inférieur à 1. Les états d'une classe transitoire sont appelés *transitoires*.

Considérons un exemple suivant :



On remarque qu'il existe un pattern (un motif) cyclique : ainsi, on sort d'un état et on y retourne d'une manière obligatoire après un certain nombre d'étapes (période).

Supposons qu'on commence par l'état A (moment 0). D'abord on passe à l'état B ($n = 1$), d'où obligatoirement on passe à l'état C ($n = 2$) pour retourner dans l'état A ($n = 3$) et recommencer. On peut remarquer qu'on ne retourne à l'état A qu'aux moments $n = 3, 6, \ldots$, c'est-à-dire multiples de 3. On dit que A est de la période 3.

La **période** d'un état i (en. *period*), notée $d(i)$, est un entier tel que :

$$d(i) = PGCD(\{n \in \mathbb{N} \mid G_{ii}^n > 0\})$$

Si :
- $d(i) > 1$, l'état i est appelé **périodique** (en. *periodic state*)
- $d(i) = 1$, l'état i est appelé **apériodique** (en. *aperiodic state*)

Les états d'une même classe ont la même période :

$$i \leftrightarrow j \Rightarrow d(i) = d(j)$$

On peut donc parler de la période d'une classe d'états. Si la période vaut 1, la classe est dite **apériodique**, sinon **périodique**.

Une chaîne de Markov est dite **apériodique** (en. *aperiodic*) si tous les états ont la période 1, i.e. :

$$\forall i \in E,\ PGDC(\{n \in \mathbb{N} \mid\ G_{ii}^n > 0\}) = 1$$

Comment définir si une chaîne de Markov est apériodique ?

Considérons une chaîne de Markov $X = (X_n)_{n\in\mathbb{N}}$ irréductible.

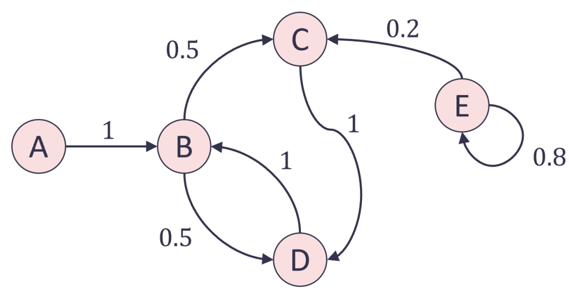
1. S'il existe une auto-transition dans la chaîne ($G_{ii} > 0$), alors la chaîne de Markov est apériodique.
2. Supposons qu'il est possible d'aller de l'état i à l'état j en $l \in \mathbb{N}$ étapes, i.e. $G^l_{ij} > 0$. Supposons également qu'il existe $m \in \mathbb{N},\ m \neq l,\ G^m_{ii} > 0$. Si $PGDC(l, m) = 1$, alors l'état i est apériodique.
3. La chaîne de Markov est apériodique si et seulement si $\exists n \in \mathbb{N},\ n > 0$ tel que tous les élements de la matrice de transition G^n sont strictement positifs, i.e. :

$$\forall (i, j) \in E^2,\ G^n_{ij} > 0$$

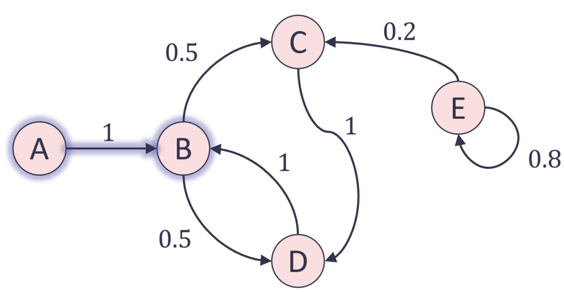
Une chaîne de Markov est dite **ergodique** (en. *ergodic*) si elle est irréductible et apériodique.

Probabilité stationnaire (invariante) d'une Chaîne de Markov

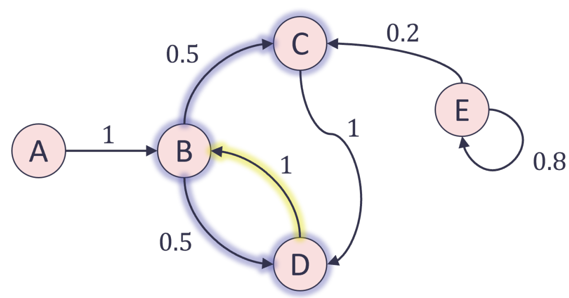
Considérons une chaîne de Markov suivante (inspirée par lien utile 2) :



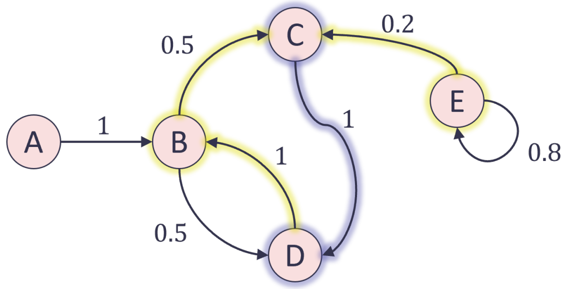
On peut observer qu'à partir de l'état A on passe à l'état B avec la probabilité 1. Notons qu'il n'y a aucune flèche qui amène à l'état A. C'est-à-dire, une fois sortie de l'état A, il n'est plus possible d'y retourner.



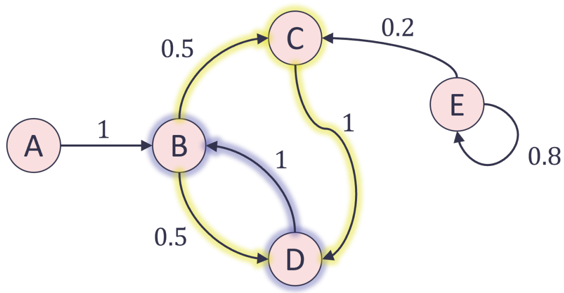
En arrivant à l'état B, on peut passer à l'état C ou l'état D avec les probabilités de 0.5. Notons qu'on peut retourner à l'état B à partir de l'état D.



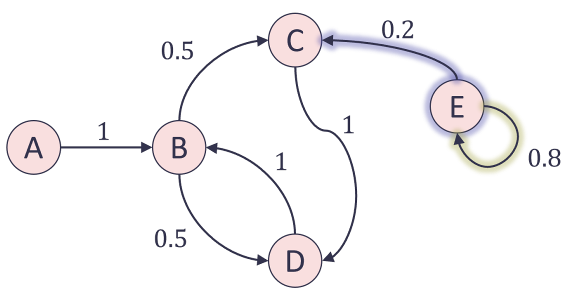
A partir de l'état C, on passe à l'état D avec la probabilité de 1, mais il reste possible de retourner à l'état C en passant par l'état B. Il est aussi possible de venir à l'état C depuis l'état E.



Une fois dans l'état D, on passe à l'état B avec la probabilité de 1. Il est possible de retourner dans l'état D soit depuis l'état B, soit en passant par l'état C.



Un cas particulier est l'état E. On remarque que la chance qu'on reste dans l'état E lui-même est de 0.8. Il est néanmoins possible de sortir de l'état E avec la probabilité 0.2. Cependant, une fois sorti de l'état E à l'état C, il n'est plus possible de retourner à l'état E.



La matrice de transition qui correspond à cette chaîne de Markov est la suivante :

$$\begin{array}{c c} & \begin{array}{ccccc} A & B & C & D & E \end{array} \\ \begin{array}{c} A \\ B \\ C \\ D \\ E \end{array} & \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0.5 & 0.5 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0.2 & 0 & 0.8 \end{pmatrix} \end{array}$$

Est-ce qu'il existe un vecteur ligne des probabilités initiales :

$$\pi = (\pi_A,\ \pi_B,\ \pi_C,\ \pi_D,\ \pi_E)$$

telles que après une étape (transition) ces probabilités restent les mêmes :

$$\mathbb{P}(X_{n+1} = i) = \mathbb{P}(X_n = i),\ \forall i \in \{A, B, C, D, E\}$$

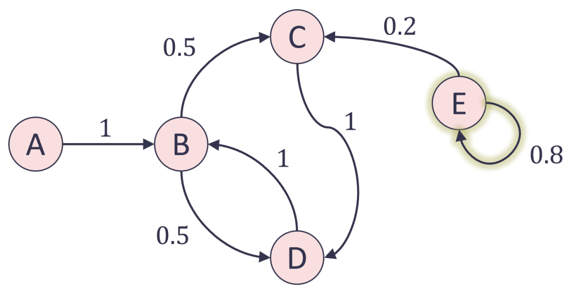
> *Autrement dit* : en sachant qu'il y a une chance de π_A que la chaîne de Markov est dans l'état A, une chance de π_B qu'elle est dans l'état B, ... une chance π_E qu'elle est dans l'état E, comment à partir de la matrice de transition G de prédire les probabilités que la chaîne de Markov sera dans ces états à l'étape suivante ?

Ainsi, pour chaque état (A, B, C, D, E), on veut que la probabilité que la chaîne de Markov est dans cet état au moment $n + 1$ soit la même qu'au moment précédent n.

Reprenons la chaîne de Markov. Comme c'est dit avant, si on est dans l'état A, la probabilité qu'on passe à l'état B est de 1, et il n'y a pas de possibilité de retourner à l'étape A. Donc, la probabilité qu'on reste à l'état A à la prochaine étape est 0 :

$$\pi_A = 0$$

Considérons l'état E. Quelle est la probabilité qu'on se retrouve dans le même état à l'étape suivante ? La seule possibilité est si à l'étape actuelle, la chaîne soit à l'état E.



La probabilité dans ce cas-là est donnée par :

$$\pi_E^{(n)} \times 0.8 = \pi_E^{(n+1)} \Rightarrow \pi_E = 0$$

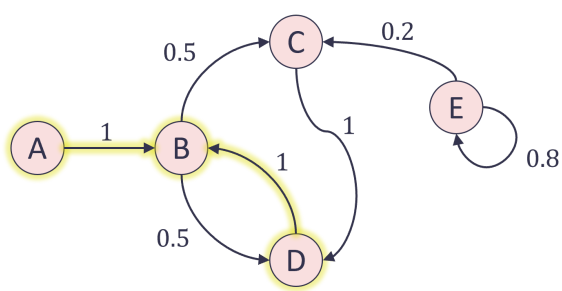
Cette valeur n'est pas surprenante car une fois sorti de l'état E, il est impossible d'y retourner.

Passons aux états B, C, D.

Pour l'état B la réfléxion peut être la suivante :

Comment est-ce qu'on peut arriver à l'état B ?

- depuis l'état A avec la probabilité 1
- depuis l'état D avec la probabilité 1

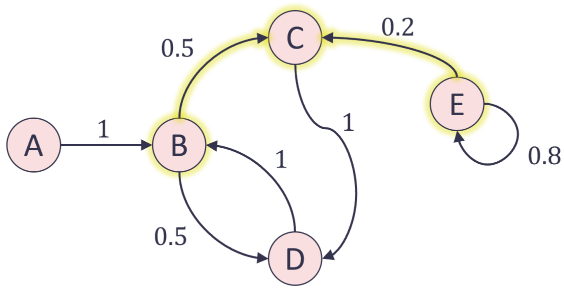


Donc :

$$\pi_B = \pi_A \times 1 + \pi_D \times 1 = 0 \times 1 + \pi_D \times 1 = \pi_D$$

Comment est-ce qu'on peut arriver à l'état C ?

- depuis l'état B avec la probabilité 0.5
- depuis l'état E avec la probabilité 0.2



Donc :

$$\pi_C = \pi_B \times 0.5 + \pi_E \times 0.2 = \pi_B \times 0.5 + 0 \times 0.2 = \pi_B \times 0.5$$

Alors, on obtient :

$$\begin{cases} \pi_A = 0 \\ \pi_B = \pi_D \\ \pi_C = \pi_B \times 0.5 \\ \pi_E = 0 \\ \pi_A + \pi_B + \pi_C + \pi_D + \pi_E = 1 \end{cases}$$

$$\pi_A + \pi_B + \pi_C + \pi_D + \pi_E = 0 + \pi_B + 0.5 \times \pi_B + \pi_B + 0 = 2.5 \times \pi_B = 1$$

Donc,

$$\pi_B = \frac{2}{5} = 0.4$$

Alors,

$$\pi_D = \pi_B = 0.4$$

Et

$$\pi_C = 0.5 \times 0.4 = 0.2$$

On obtient, donc :

$$\pi = (0,\ 0.4,\ 0.2,\ 0.4,\ 0)$$

Vérifions que $\pi G = \pi$:

$$(0,\ 0.4,\ 0.2,\ 0.4,\ 0) \times \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0.5 & 0.5 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0.2 & 0 & 0.8 \end{pmatrix} =?$$

1. $\pi \times$ 1ère colonne :

$$0 \times 0 + 0.4 \times 0 + 0.2 \times 0 + 0.4 \times 0 + 0 \times 0 = 0$$

2. $\pi \times$ 2ème colonne :

$$0 \times 1 + 0.4 \times 0 + 0.2 \times 0 + 0.4 \times 1 + 0 \times 0 = 0.4$$

3. $\pi \times$ 3ème colonne :

$$0 \times 0 + 0.4 \times 0.5 + 0.2 \times 0 + 0.4 \times 0 + 0 \times 0.2 = 0.2$$

4. $\pi \times$ 4ème colonne :

$$0 \times 0 + 0.4 \times 0.5 + 0.2 \times 1 + 0.4 \times 0 + 0 \times 0 = 0.2 + 0.2 = 0.4$$

5. $\pi \times$ 5ème colonne :

$$0 \times 0 + 0.4 \times 0 + 0.2 \times 0 + 0.4 \times 0 + 0 \times 0.8 = 0$$

On obtient donc :

$$(0,\ 0.4,\ 0.2,\ 0.4,\ 0) \times \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0.5 & 0.5 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0.2 & 0 & 0.8 \end{pmatrix} = (0,\ 0.4,\ 0.2,\ 0.4,\ 0)$$

Dans le R :

```
# matrice de transition
G <- matrix(data = c(0, 1, 0, 0, 0,
                     0, 0, 0.5, 0.5, 0,
                     0, 0, 0, 1, 0,
                     0, 1, 0, 0, 0,
                     0, 0, 0.2, 0, 0.8), nrow = 5, ncol = 5, byrow = TRUE); G
```

```
##      [,1] [,2] [,3] [,4] [,5]
## [1,]    0    1  0.0  0.0  0.0
## [2,]    0    0  0.5  0.5  0.0
## [3,]    0    0  0.0  1.0  0.0
## [4,]    0    1  0.0  0.0  0.0
## [5,]    0    0  0.2  0.0  0.8
```

```
# vérification
p <- c(0, 0.4, 0.2, 0.4, 0) # vecteur pi

# multiplication matricielle
test <- p %*% G; test
```

```
##      [,1] [,2] [,3] [,4] [,5]
## [1,]    0  0.4  0.2  0.4    0
```

```
# test
all(test == p)
```

```
## [1] TRUE
```

Il est possible d'obtenir ce résultat en résolvant l'équation suivante :

$$\pi G = \pi$$

Notons que cette équation ressemble à la recherche des vecteurs propres avec $\lambda = 1$:

$$Ax = \lambda x$$

Dans ce cas là, il s'agit de vecteur propre gauche (*left eigenvector*) correspondant à la valeur propre $\lambda = 1$. Notons ce vecteur propre comme **e** (on prend la matrice transposée G^T pour le trouver). La relation entre π et le vecteur **e** est la suivante étant donné la normalisation ($\sum_i \pi_i = 1$) :

$$\pi = \frac{e}{\sum_i e_i}$$

Dans le R :

```
# transposer la matrice de transition
tG <- t(G); tG
```

```
##      [,1] [,2] [,3] [,4] [,5]
## [1,]    0  0.0    0    0  0.0
## [2,]    1  0.0    0    1  0.0
## [3,]    0  0.5    0    0  0.2
## [4,]    0  0.5    1    0  0.0
## [5,]    0  0.0    0    0  0.8
```

```
# trouver les vecteurs propres
e <- eigen(tG); e
```

```
## eigen() decomposition
## $values
## [1]  1.0+0.0i  0.8+0.0i -0.5+0.5i -0.5-0.5i  0.0+0.0i
##
## $vectors
##                [,1]           [,2]                  [,3]                  [,4]            [,5]
## [1,]  0.0000000+0i  0.0000000+0i  0.0000000+0.0000000i  0.0000000+0.0000000i  7.071068e-01+0i
## [2,] -0.6666667+0i -0.4294100+0i -0.7071068+0.0000000i -0.7071068+0.0000000i  4.612147e-16+0i
## [3,] -0.3333333+0i -0.0601174+0i  0.3535534+0.3535534i  0.3535534-0.3535534i  2.747662e-16+0i
## [4,] -0.6666667+0i -0.3435280+0i  0.3535534-0.3535534i  0.3535534+0.3535534i -7.071068e-01+0i
## [5,]  0.0000000+0i  0.8330555+0i  0.0000000+0.0000000i  0.0000000+0.0000000i  0.000000e+00+0i
```

```
# la valeur propre qui nous intéresse est 1
# prendre l'indice de la valeur propre = 1
ind <- which(e$values == 1); ind
```

```
## [1] 1
```

```
# predre le vecteur propre qui correspond à la valeur propre 1
# dans la matrice renvoyée par eigen() les colonnes correspondent
# aux valeurs propres
ee <- e$vectors[, ind]; ee
```

```
## [1]  0.0000000+0i -0.6666667+0i -0.3333333+0i -0.6666667+0i  0.0000000+0i
```

```r
# normaliser ce vecteur
ee <- ee / sum(ee); ee
```

```
## [1] 0.0+0i 0.4+0i 0.2+0i 0.4+0i 0.0+0i
```

```r
# la partie imaginaire de tous les éléments dans cet exemple = 0
# on peut donc laisser que la partie réelle
ee <- Re(ee)
```

Dans le R vous pouvez également utiliser une librairie spéciale pour le traitement de chaîne de Markov `markovchain`.

```r
library(markovchain)
# création de l'objet chaîne de Markov
mcstates <- new("markovchain", states = c("A", "B", "C", "D", "E"),
                transitionMatrix = G, name = "states"); mcstates
```

```
## states
##  A  5 - dimensional discrete Markov Chain defined by the following states:
##  A, B, C, D, E
##  The transition matrix  (by rows)  is defined as follows:
##   A B   C   D   E
## A 0 1 0.0 0.0 0.0
## B 0 0 0.5 0.5 0.0
## C 0 0 0.0 1.0 0.0
## D 0 1 0.0 0.0 0.0
## E 0 0 0.2 0.0 0.8
```

```r
# calcul du vecteur ligne pi
steadyStates(mcstates)
```

```
##      A   B   C   D E
## [1,] 0 0.4 0.2 0.4 0
```

On peut procéder par la résolution de l'équation :

$$\pi(G - 1 \cdot I) = 0$$

$$G - I = \begin{pmatrix} -1 & 1 & 0 & 0 & 0 \\ 0 & -1 & 0.5 & 0.5 & 0 \\ 0 & 0 & -1 & 1 & 0 \\ 0 & 1 & 0 & -1 & 0 \\ 0 & 0 & 0.2 & 0 & -0.2 \end{pmatrix}$$

On multiplie le vecteur π par cette matrice :

$$\begin{cases} \pi_A \times (-1) + \pi_B \times 0 + \pi_C \times 0 + \pi_D \times 0 = 0 \\ \pi_A \times 1 + \pi_B \times (-1) + \pi_C \times 0 + \pi_D \times 1 + \pi_E \times 0 = 0 \\ \pi_A \times 0 + \pi_B \times 0.5 + \pi_C \times (-1) + \pi_D \times 0 + \pi_E \times 0.2 = 0 \\ \pi_A \times 0 + \pi_B \times 0.5 + \pi_C \times 1 + \pi_D \times (-1) + \pi_E \times 0 = 0 \\ \pi_A \times 0 + \pi_B \times 0 + \pi_C \times 0 + \pi_D \times 0 + \pi_E \times (-0.2) = 0 \\ \pi_A + \pi_B + \pi_C + \pi_D + \pi_E = 1 \end{cases}$$

$$\begin{cases} -\pi_A = 0 \\ \pi_A - \pi_B + \pi_D = 0 \\ 0.5\pi_B - \pi_C + 0.2\pi_E = 0 \\ 0.5\pi_B + \pi_C - \pi_D = 0 \\ -0.2\pi_E = 0 \\ \pi_A + \pi_B + \pi_C + \pi_D + \pi_E = 1 \end{cases}$$

$$\begin{cases} \pi_A = 0 \\ 0 - \pi_B + \pi_D = 0 \\ 0.5\pi_B - \pi_C + 0.2 \times 0 = 0 \\ 0.5\pi_B + \pi_C - \pi_D = 0 \\ \pi_E = 0 \\ \pi_A + \pi_B + \pi_C + \pi_D + \pi_E = 1 \end{cases}$$

$$\begin{cases} \pi_A = 0 \\ \pi_B = \pi_D \\ 0.5\pi_B = \pi_C \\ 0.5\pi_B + \pi_C - \pi_B = 0 \\ \pi_E = 0 \\ 0 + \pi_B + \pi_C + \pi_B + 0 = 1 \end{cases}$$

$$\begin{cases} \pi_A = 0 \\ \pi_B = \pi_D \\ 0.5\pi_B = \pi_C \\ \pi_E = 0 \\ \pi_B + 0.5\pi_B + \pi_B = 1 \end{cases}$$

$$\begin{cases} \pi_A = 0 \\ \pi_B = \pi_D \\ 0.5\pi_B = \pi_C \\ \pi_E = 0 \\ 2.5\pi_B = 1 \end{cases}$$

$$\begin{cases} \pi_A = 0 \\ \pi_B = \pi_D = 0.4 \\ \pi_C = 0.5\pi_B = 0.2 \\ \pi_E = 0 \end{cases}$$

On peut aussi interpréter *la distribution stationnaire* d'une chaîne de Markov comme la fraction de temps passé en chaque état de cette chaîne de Markov, asymptotiquement.

D'une manière plus formelle :

Soit $X = (X_n)_{n \in \mathbb{N}}$ une chaîne de Markov homogène sur un espace d'états E, de matrice de transition G et de condition initiale X_0 ayant pour fonction de masse Π. La loi de la v.a.r. discrète X_n s'obtient à partir du vecteur π et de matrice de transition G par la relation de récurrence.
Loi de probabilité invariante (stationnaire) (en. *stationary distribution*, *steady state distribution*, *invariant distribution*) de $X = (X_n)_{n \in \mathbb{N}}$ est une fonction de masse $\Xi :\ k \in E \to \xi_k \in [0, 1]$ où le vecteur $\xi = (\xi_1, \ldots, \xi_N)$ est une solution du système linéaire :

$$\xi = \xi G$$

Lorsqu'une loi de probabilité invariante est choisie comme loi pour la variable X_0, la loi pour X_n ($n \in \mathbb{N}$) reste la même fonction de masse :

$$\xi^{(1)} = \xi G = \xi$$

$$\xi^{(2)} = \xi^{(1)} G = \xi G = \xi$$
$$\xi^{(n)} = \xi^{(n-1)} G = \xi G = \xi$$

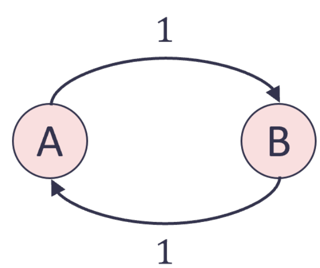
Remarque : une telle loi existe toujours.

Remarque : en pratique, souvent on note cette distribution avec π.

Est-ce qu'il n'existe qu'une seule distribution stationnaire d'une chaîne de Markov ?

Considérons un cas suivant :

$$G = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$$



On remarque qu'il y a une dépendance de l'état X_0. Ainsi :

$$X_n = \begin{cases} X_0 \text{ si } n \text{ est pair} \\ X_1 \text{ si } n \text{ est impair} \end{cases}$$

Les probabilités stationnaires dans ce cas-là vont être $(0,\ 1)$ et $(1,\ 0)$.

Dans le R :

```
# matrice de transition
G2 = matrix(c(1, 0, 0, 1), nrow = 2, ncol = 2, byrow = TRUE); G2
```

```
##      [,1] [,2]
## [1,]    1    0
## [2,]    0    1
```

```
# objet chaîne de Markov
mcstates2 <- new("markovchain", states = c("A", "B"),
                 transitionMatrix = G2, name = "states"); mcstates2
```

```
## states
##  A  2 - dimensional discrete Markov Chain defined by the following states:
##  A, B
##  The transition matrix  (by rows)  is defined as follows:
##   A B
## A 1 0
## B 0 1
```

```r
# distribution stationnaire
steadyStates(mcstates2)
```

```
##      A B
## [1,] 0 1
## [2,] 1 0
```

Quel est le comportement asymptotique d'une chaîne de Markov $n \to \infty$?

On appelle **distribution limite** (en. *limiting distribution*) d'une chaîne de Markov $X = (X_n)_{n\in\mathbb{N}}$ sur un espace d'états E, de matrice de transition G, une distribution donnée par un vecteur ligne $\pi = (\pi_1, \pi_2, \dots, \pi_N)$ telle que :

$$\forall i, j \in E : \ \pi_j = \lim_{n\to\infty} \mathbb{P}(X_n = j|X_0 = i) = \lim_{n\to\infty} G^n_{ij}$$

et :

$$\sum_{j\in E} \pi_j = 1$$

Théorème :
Soit $X = (X_n)_{n\in\mathbb{N}}$ une chaîne de Markov homogène, ergodique (irréductible et apériodique) sur un espace d'états E, de matrice de transition G.
Alors :

1. Il existe une distirbution stationnaire unique $\pi = (\pi_1, \pi_2, \dots, \pi_N)$ qui est une solution de l'ensemble d'équations :

$$\begin{array}{l} \pi G = \pi \\ \sum_{j=1}^{N} \pi_j = 1 \end{array}$$

2. Cette distribution stationnaire est une distribution limite de cette chaîne de Markov, i.e. $\forall (i, j) \in E$:

$$\pi_j = \lim_{n\to\infty} \mathbb{P}(X_n = j|X_0 = i) = \lim_{n\to\infty} G^n_{ij}$$

Notons que si une distribution limite existe, elle ne dépend pas de l'état initial ($X_0 = i$). Alors :

$$\pi_j = \lim_{n\to\infty} \mathbb{P}(X_n = j|X_0 = i) \approx \mathbb{P}(X_n = j)$$

Ce fait peut être interpréter comme l'indépendance de deux états de la chaîne de Markov séparés par une longue histoire.

Liens utiles

1. Introducing Markov Chains by HarvardX
2. Markov Chain Stationary Distribution : Data Science Concepts

Références

[1] Stéphane Balac and Olivier Mazet. *Introduction aux Probabilités*. URL: https://perso.univ-rennes1.fr/stephane.balac/publis/polypbs.pdf (visited on 05/13/2021).