Chaînes de Markov

Cours de probabilités et statistiques

Master 1 MEEF - Mathématiques
Sorbonne Université

Séance (annulée) du 16 mars 2020

Chaînes de Markov : définition formelle

Définition

Une chaîne de Markov est une suite de variables aléatoires à valeurs dans un ensemble fini $E = \{1, \ldots, M\}$ vérifiant la propriété (pour tous $i_0, \ldots, i_{n+1}$ de E) :

$$\mathbf{P}(X_{n+1} = i_{n+1} | X_0 = i_0, \ldots, X_n = i_n) = \mathbf{P}(X_{n+1} = i_{n+1} | X_n = i_n).$$

Signification

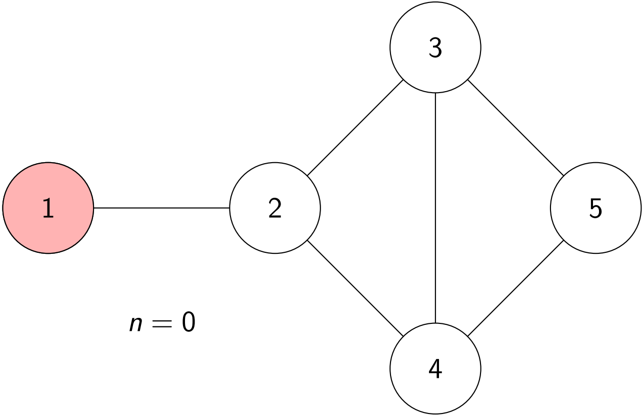
Définissons $T_0 = \{X_n = i_n\}$ (le “présent”), $T_+ = \{X_{n+1} = i_{n+1}\}$ (le “futur”) et $T_- = \{X_0 = i_0, \ldots, X_{n-1} = i_n\}$ (le “passé”).
Alors la définition se récrit

$$\mathbf{P}_{T_0}(T_+ | T_-) = \mathbf{P}_{T_0}(T_+),$$

ce qui peut s'exprimer comme “conditionnellement au présent, le passé et le futur sont indépendants”.

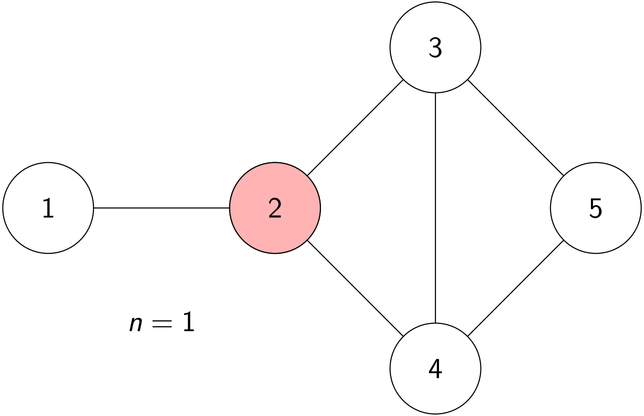
Chaînes de Markov : explication informelle

Donnons un exemple : en rouge, la valeur d'une réalisation de X_n.



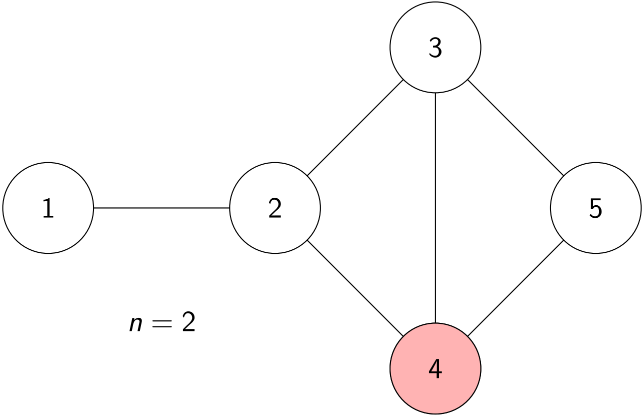
Chaînes de Markov : explication informelle

Donnons un exemple : en rouge, la valeur d'une réalisation de X_n.



Chaînes de Markov : explication informelle

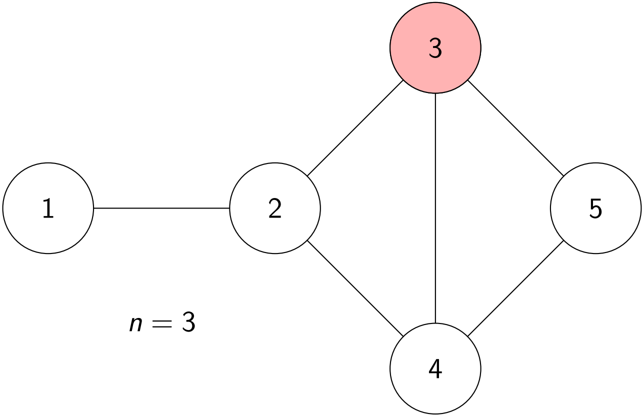
Donnons un exemple : en rouge, la valeur d'une réalisation de X_n.



$n = 2$

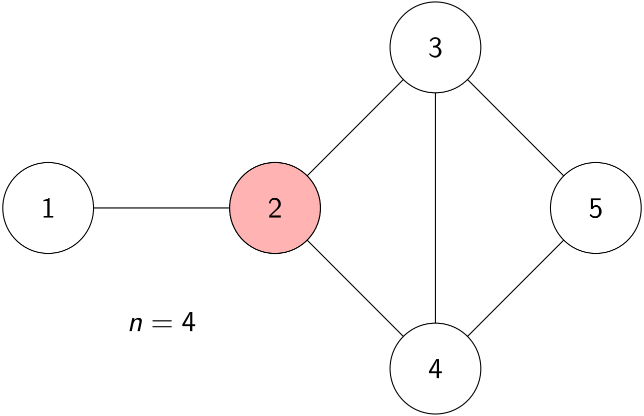
Chaînes de Markov : explication informelle

Donnons un exemple : en rouge, la valeur d'une réalisation de X_n.



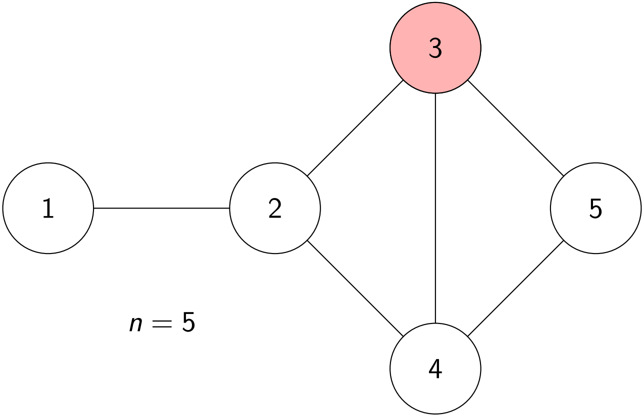
Chaînes de Markov : explication informelle

Donnons un exemple : en rouge, la valeur d'une réalisation de X_n.



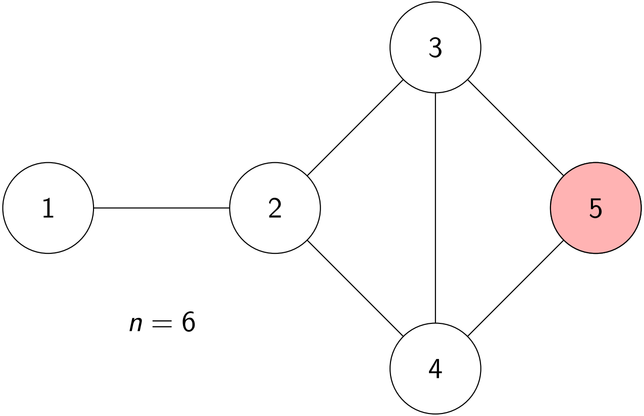
Chaînes de Markov : explication informelle

Donnons un exemple : en rouge, la valeur d'une réalisation de X_n.



Chaînes de Markov : explication informelle

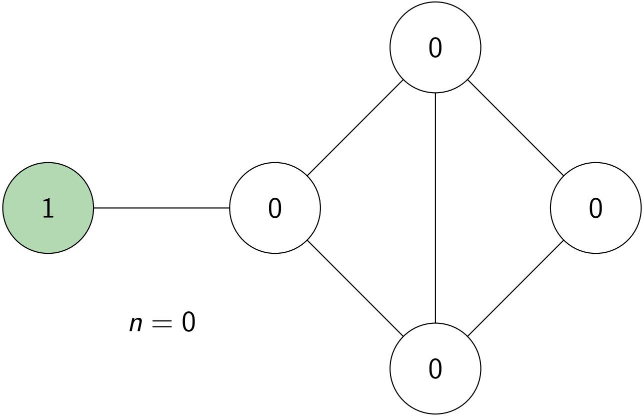
Donnons un exemple : en rouge, la valeur d'une réalisation de X_n.



$n = 6$

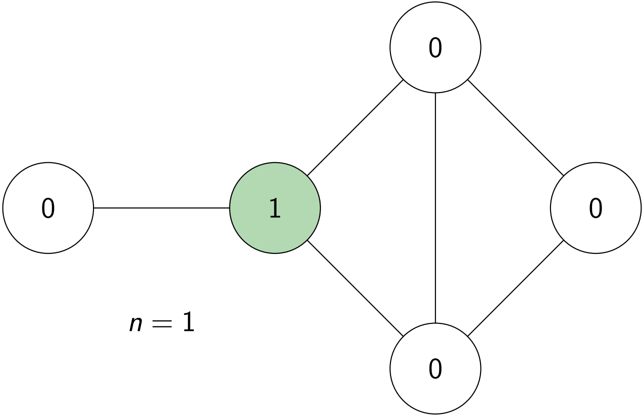
Chaînes de Markov : explication informelle

Donnons un exemple : en vert, la loi de X_n.



Chaînes de Markov : explication informelle

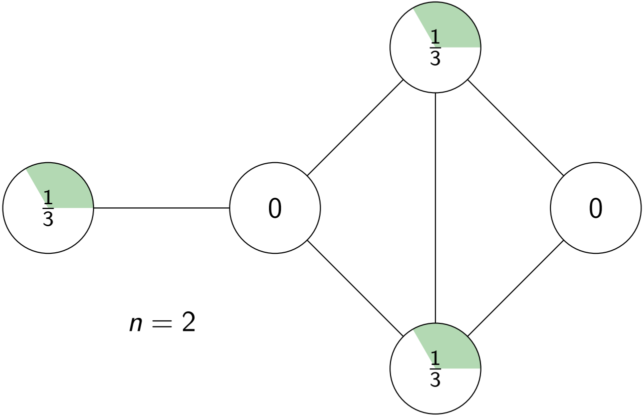
Donnons un exemple : en vert, la loi de X_n.



$n = 1$

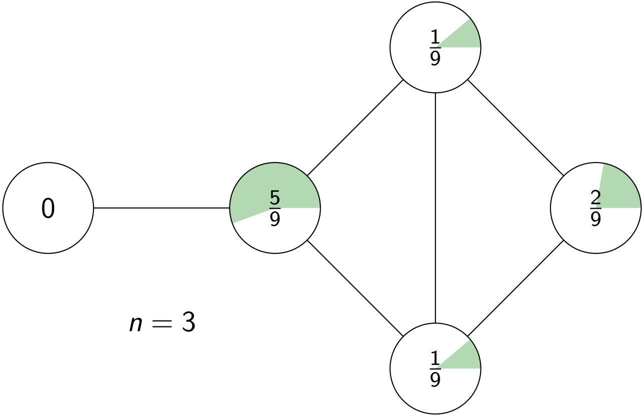
Chaînes de Markov : explication informelle

Donnons un exemple : en vert, la loi de X_n.



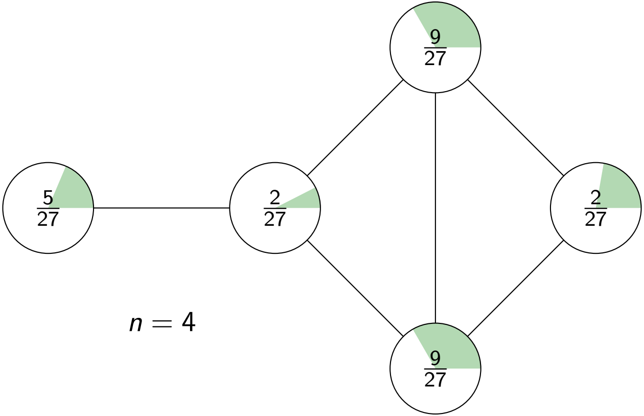
Chaînes de Markov : explication informelle

Donnons un exemple : en vert, la loi de X_n.



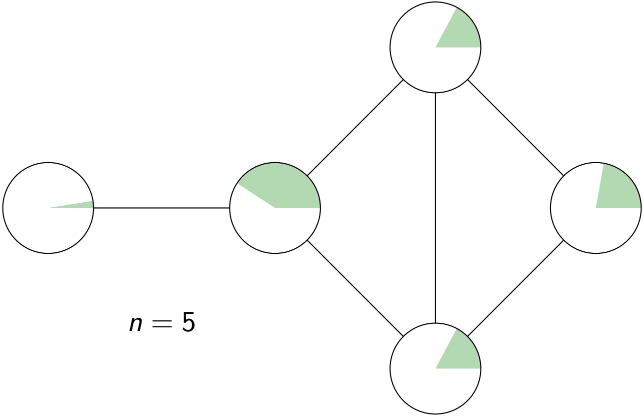
Chaînes de Markov : explication informelle

Donnons un exemple : en vert, la loi de X_n.



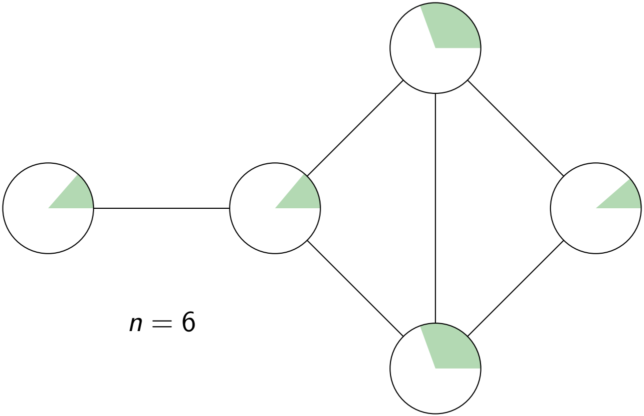
Chaînes de Markov : explication informelle

Donnons un exemple : en vert, la loi de X_n.



Chaînes de Markov : explication informelle

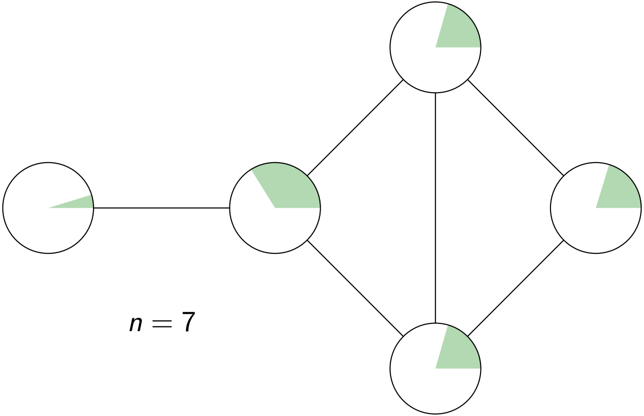
Donnons un exemple : en vert, la loi de X_n.



$n = 6$

Chaînes de Markov : explication informelle

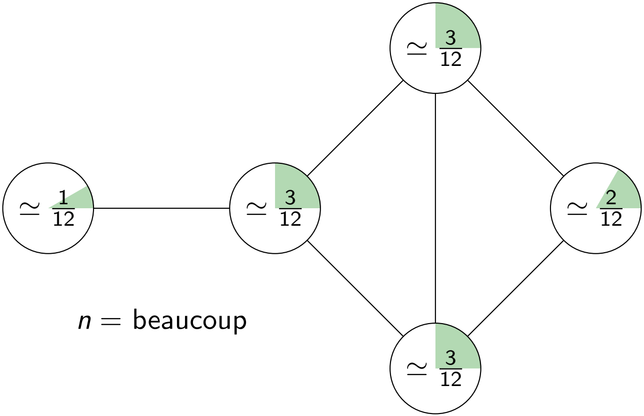
Donnons un exemple : en vert, la loi de X_n.



$n = 7$

Chaînes de Markov : explication informelle

Donnons un exemple : en vert, la loi de X_n.



Sur cet exemple, la loi de X_n converge vers une limite.

Comment calculer la loi de X_n ?

On note $P_{i,j} = \mathbf{P}(X_1 = j | X_0 = i)$.

Comment s'exprime $\mathbf{P}(X_2 = j | X_0 = i)$ (= probabilité de passer de i à j en deux pas) ?
"Passer de i à j en deux pas" = "Passer de i à 1, puis de 1 à j, ou, de i à 2, puis de 2 à j, ou...". Autrement dit :

$$\{X_2 = j, X_0 = i\} = \bigcup_{k=1}^{M} \{X_2 = j, X_1 = k, X_0 = i\}$$

Comment calculer la loi de X_n ?

On note $P_{i,j} = \mathbf{P}(X_1 = j | X_0 = i)$.

Comment s'exprime $\mathbf{P}(X_2 = j | X_0 = i)$ ($=$ probabilité de passer de i à j en deux pas) ?
"Passer de i à j en deux pas" $=$ "Passer de i à 1, puis de 1 à j, ou, de i à 2, puis de 2 à j, ou...". Autrement dit :

$$\mathbf{P}(X_2 = j, X_0 = i) = \sum_{k=1}^{M} \mathbf{P}(X_2 = j, X_1 = k, X_0 = i)$$

Comment calculer la loi de X_n ?

On note $P_{i,j} = \mathbf{P}(X_1 = j | X_0 = i)$.

Comment s'exprime $\mathbf{P}(X_2 = j | X_0 = i)$ (= probabilité de passer de i à j en deux pas) ?
"Passer de i à j en deux pas" = "Passer de i à 1, puis de 1 à j, ou, de i à 2, puis de 2 à j, ou...". Autrement dit :

$$\mathbf{P}(X_2 = j | X_0 = i) = \sum_{k=1}^{M} \mathbf{P}(X_2 = j, X_1 = k | X_0 = i)$$

Comment calculer la loi de X_n ?

On note $P_{i,j} = \mathbf{P}(X_1 = j | X_0 = i)$.

Comment s'exprime $\mathbf{P}(X_2 = j | X_0 = i)$ (= probabilité de passer de i à j en deux pas) ?
"Passer de i à j en deux pas" = "Passer de i à 1, puis de 1 à j, ou, de i à 2, puis de 2 à j, ou...". Autrement dit :

$$\mathbf{P}(X_2 = j | X_0 = i) = \sum_{k=1}^{M} \mathbf{P}(X_2 = j, X_1 = k | X_0 = i)$$

Comment calculer la loi de X_n ?

On note $P_{i,j} = \mathbf{P}(X_1 = j | X_0 = i)$.

Comment s'exprime $\mathbf{P}(X_2 = j | X_0 = i)$ (= probabilité de passer de i à j en deux pas) ?
"Passer de i à j en deux pas" = "Passer de i à 1, puis de 1 à j, ou, de i à 2, puis de 2 à j, ou...". Autrement dit :

$$\mathbf{P}(X_2 = j | X_0 = i) = \sum_{k=1}^{M} \mathbf{P}(X_1 = k | X_0 = i)\mathbf{P}(X_2 = j | X_1 = k, X_0 = i)$$

Comment calculer la loi de X_n ?

On note $P_{i,j} = \mathbf{P}(X_1 = j | X_0 = i)$.

Comment s'exprime $\mathbf{P}(X_2 = j | X_0 = i)$ (= probabilité de passer de i à j en deux pas) ?
“Passer de i à j en deux pas” = “Passer de i à 1, puis de 1 à j, ou, de i à 2, puis de 2 à j, ou...”. Autrement dit :

$$\mathbf{P}(X_2 = j | X_0 = i) = \sum_{k=1}^{M} \mathbf{P}(X_1 = k | X_0 = i) \mathbf{P}(X_2 = j | X_1 = k, X_0 = i)$$

Comment calculer la loi de X_n ?

On note $P_{i,j} = \mathbf{P}(X_1 = j | X_0 = i)$.

Comment s'exprime $\mathbf{P}(X_2 = j | X_0 = i)$ (= probabilité de passer de i à j en deux pas) ?
“Passer de i à j en deux pas” = “Passer de i à 1, puis de 1 à j, ou, de i à 2, puis de 2 à j, ou...”. Autrement dit :

$$\mathbf{P}(X_2 = j | X_0 = i) = \sum_{k=1}^{M} \mathbf{P}(X_1 = k | X_0 = i) \mathbf{P}(X_2 = j | X_1 = k)$$

Comment calculer la loi de X_n ?

On note $P_{i,j} = \mathbf{P}(X_1 = j | X_0 = i)$.

Comment s'exprime $\mathbf{P}(X_2 = j | X_0 = i)$ (= probabilité de passer de i à j en deux pas) ?
"Passer de i à j en deux pas" = "Passer de i à 1, puis de 1 à j, ou, de i à 2, puis de 2 à j, ou...". Autrement dit :

$$\mathbf{P}(X_2 = j | X_0 = i) = \sum_{k=1}^{M} P_{i,k} P_{k,j}$$

Comment calculer la loi de X_n ?

On note $P_{i,j} = \mathbf{P}(X_1 = j | X_0 = i)$.

Comment s'exprime $\mathbf{P}(X_2 = j | X_0 = i)$ (= probabilité de passer de i à j en deux pas) ?
"Passer de i à j en deux pas" = "Passer de i à 1, puis de 1 à j, ou, de i à 2, puis de 2 à j, ou...". Autrement dit :

$$\mathbf{P}(X_2 = j | X_0 = i) = \sum_{k=1}^{M} P_{i,k} P_{k,j}$$

On reconnaît le (i,j)ème terme du produit matriciel $P \times P$.

Comment calculer la loi de X_n ?

Par récurrence, on montrerait :

Théorème

Si P est la matrice $M \times M$ dont les coefficients sont les $P_{i,j}$, alors

$$\mathbf{P}(X_n = j | X_0 = i) = {P^n}_{i,j},$$

où $P^n_{i,j}$ est le (i,j)ème coefficient de la matrice P^n.

Comment calculer la loi de X_n ?

Comment s'exprime $\mathbf{P}(X_1 = j)$ (X_0 est maintenant aléatoire) ?
On note μ la loi de X_0 ($\mu_i = \mathbf{P}(X_0 = i)$).

$$\{X_1 = j\} = \bigcup_{i=1}^{M} \{X_1 = j, X_0 = i\}$$

Comment calculer la loi de X_n ?

Comment s'exprime $\mathbf{P}(X_1 = j)$ (X_0 est maintenant aléatoire) ?
On note μ la loi de X_0 ($\mu_i = \mathbf{P}(X_0 = i)$).

$$\mathbf{P}(X_1 = j) = \sum_{i=1}^{M} \mathbf{P}(X_1 = j, X_0 = i)$$

Comment calculer la loi de X_n ?

Comment s'exprime $\mathbf{P}(X_1 = j)$ (X_0 est maintenant aléatoire) ?
On note μ la loi de X_0 ($\mu_i = \mathbf{P}(X_0 = i)$).

$$\mathbf{P}(X_1 = j) = \sum_{i=1}^{M} \mathbf{P}(X_0 = i)\mathbf{P}(X_1 = j | X_0 = i)$$

Comment calculer la loi de X_n ?

Comment s'exprime $\mathbf{P}(X_1 = j)$ (X_0 est maintenant aléatoire) ?
On note μ la loi de X_0 ($\mu_i = \mathbf{P}(X_0 = i)$).

$$\mathbf{P}(X_1 = j) = \sum_{i=1}^{M} \mu_i P_{i,j}$$

Comment calculer la loi de X_n ?

Comment s'exprime $\mathbf{P}(X_1 = j)$ (X_0 est maintenant aléatoire) ?
On note μ la loi de X_0 ($\mu_i = \mathbf{P}(X_0 = i)$).

$$\mathbf{P}(X_1 = j) = \sum_{i=1}^{M} \mu_i P_{i,j}$$

On reconnaît le $j$ème terme du produit $\mu \times P$, si μ est vu comme un vecteur ligne.

Comment calculer la loi de X_n ?

Comment s'exprime $\mathbf{P}(X_1 = j)$ (X_0 est maintenant aléatoire) ?
On note μ la loi de X_0 ($\mu_i = \mathbf{P}(X_0 = i)$).

$$\mathbf{P}(X_1 = j) = \sum_{i=1}^{M} \mu_i P_{i,j}$$

On reconnaît le $j$ème terme du produit $\mu \times P$, si μ est vu comme un vecteur ligne.

Théorème

Si X_0 suit la loi μ, alors

$$\mathbf{P}(X_1 = j) = (\mu P)_j \,.$$

Comment calculer la loi de X_n ?

Comment s'exprime $\mathbf{P}(X_1 = j)$ (X_0 est maintenant aléatoire) ?
On note μ la loi de X_0 ($\mu_i = \mathbf{P}(X_0 = i)$).

$$\mathbf{P}(X_1 = j) = \sum_{i=1}^{M} \mu_i P_{i,j}$$

On reconnaît le $j$ème terme du produit $\mu \times P$, si μ est vu comme un vecteur ligne.

Théorème

Si X_0 suit la loi μ, alors

$$\mathbf{P}(X_n = j) = (\mu P^n)_j \, .$$

Retour sur l'exemple

Sur l'exemple de tout à l'heure, la matrice P est donnée par :

$$P = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 1/3 & 0 & 1/3 & 1/3 & 0 \\ 0 & 1/3 & 0 & 1/3 & 1/3 \\ 0 & 1/3 & 1/3 & 0 & 1/3 \\ 0 & 0 & 1/2 & 1/2 & 0 \end{pmatrix},$$

et on peut vérifier les égalités :

$$\begin{pmatrix}1 & 0 & 0 & 0 & 0\end{pmatrix} P = \begin{pmatrix}0 & 1 & 0 & 0 & 0\end{pmatrix},$$
$$\begin{pmatrix}1 & 0 & 0 & 0 & 0\end{pmatrix} P^2 = \begin{pmatrix}1/3 & 0 & 1/3 & 1/3 & 0\end{pmatrix},$$
$$\begin{pmatrix}1 & 0 & 0 & 0 & 0\end{pmatrix} P^3 = \begin{pmatrix}0 & 5/9 & 1/9 & 1/9 & 2/9\end{pmatrix},$$

etc.

Loi invariante

Théorème

Pour tout P, il existe une loi de probabilité μ invariante, c'est-à-dire telle que si X_0 est de loi μ, alors X_n est de loi μ pour tout $n \geq 0$.

Remarque

Une loi de probabilité est invariante si et seulement si $\mu P = \mu$. En effet, dans ce cas, on a $\mu P^n = \mu$ pour tout n. Si X_0 suit la loi μ, la loi de X_n (donnée par μP^n) est donc égale à celle de X_0.

Loi invariante

Théorème

Pour tout P, il existe une loi de probabilité μ invariante, c'est-à-dire telle que si X_0 est de loi μ, alors X_n est de loi μ pour tout $n \geq 0$.

Preuve

Soit ν une probabilité quelconque (vue comme vecteur ligne). On pose

$$\mu^n = \frac{1}{n} \sum_{k=0}^{n-1} \nu P^k.$$

$(\mu^n)_{n \in \mathbf{N}}$ est une suite de probabilités, donc il existe une sous-suite $(\mu^{n_k})_{k \in \mathbf{N}}$ convergente (vers μ) ("de toute suite bornée..."). Or

$$\mu^n P - \mu^n = \frac{1}{n}(\nu P^n - \nu) \to 0.$$

Donc $\mu P - \mu = \lim_k \mu^{n_k} P - \mu^{n_k} = 0$. μ est invariante.

Retour sur l'exemple

Toujours avec la matrice P donnée par :

$$P = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 \\ 1/3 & 0 & 1/3 & 1/3 & 0 \\ 0 & 1/3 & 0 & 1/3 & 1/3 \\ 0 & 1/3 & 1/3 & 0 & 1/3 \\ 0 & 0 & 1/2 & 1/2 & 0 \end{pmatrix},$$

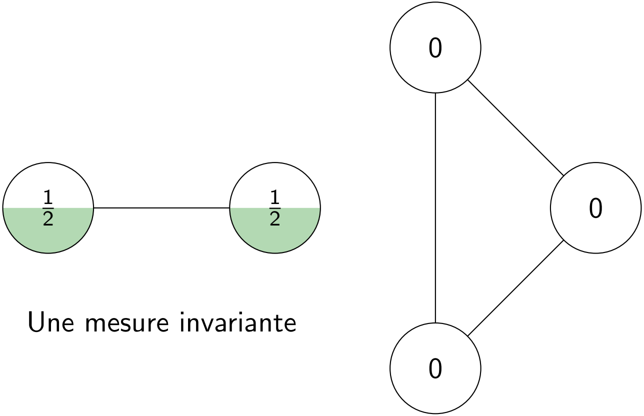
on peut vérifier que le vecteur $\mu = \frac{1}{12}\begin{pmatrix} 1 & 3 & 3 & 3 & 2 \end{pmatrix}$ vérifie bien $\mu P = \mu$.

Ce qu'on n'a (pas) montré

On a montré qu'une chaîne de Markov admettait une loi invariante.

Ce qu'on n'a (pas) montré

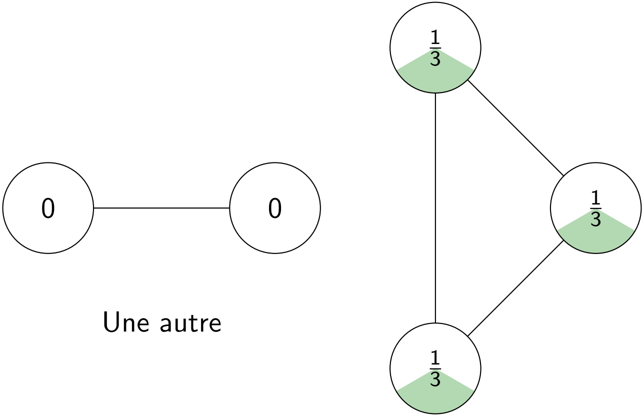
On a montré qu'une chaîne de Markov admettait une loi invariante.
On n'a pas montré que cette mesure invariante était unique.



Une mesure invariante

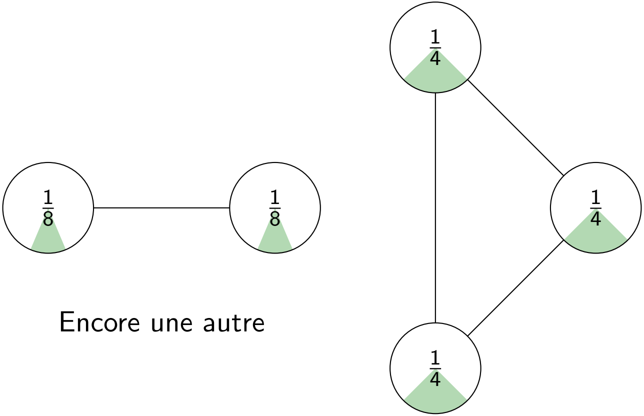
Ce qu'on n'a (pas) montré

On a montré qu'une chaîne de Markov admettait une loi invariante.
On n'a pas montré que cette mesure invariante était unique.



Ce qu'on n'a (pas) montré

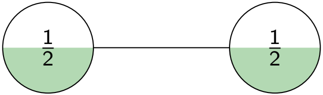
On a montré qu'une chaîne de Markov admettait une loi invariante.
On n'a pas montré que cette mesure invariante était unique.



Encore une autre

Ce qu'on n'a (pas) montré

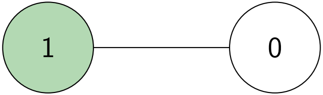
On a montré qu'une chaîne de Markov admettait une loi invariante.
On n'a pas montré que cette mesure invariante était unique.
Même si la mesure invariante est unique, on n'a pas montré que la loi de X_n converge vers cette mesure invariante.



L'unique mesure invariante

Ce qu'on n'a (pas) montré

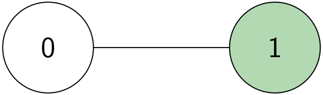
On a montré qu'une chaîne de Markov admettait une loi invariante.
On n'a pas montré que cette mesure invariante était unique.
Même si la mesure invariante est unique, on n'a pas montré que la loi de X_n converge vers cette mesure invariante.



$$n = 0$$

Ce qu'on n'a (pas) montré

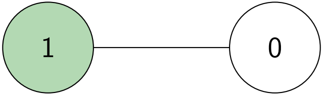
On a montré qu'une chaîne de Markov admettait une loi invariante.
On n'a pas montré que cette mesure invariante était unique.
Même si la mesure invariante est unique, on n'a pas montré que la loi de X_n converge vers cette mesure invariante.



$n = 1$

Ce qu'on n'a (pas) montré

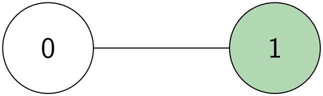
On a montré qu'une chaîne de Markov admettait une loi invariante.
On n'a pas montré que cette mesure invariante était unique.
Même si la mesure invariante est unique, on n'a pas montré que la loi de X_n converge vers cette mesure invariante.



$n = 2$

Ce qu'on n'a (pas) montré

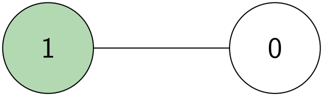
On a montré qu'une chaîne de Markov admettait une loi invariante.
On n'a pas montré que cette mesure invariante était unique.
Même si la mesure invariante est unique, on n'a pas montré que la loi de X_n converge vers cette mesure invariante.



$$n = 3$$

Ce qu'on n'a (pas) montré

On a montré qu'une chaîne de Markov admettait une loi invariante.
On n'a pas montré que cette mesure invariante était unique.
Même si la mesure invariante est unique, on n'a pas montré que la loi de X_n converge vers cette mesure invariante.



$n = 4$