

# Calcul stochastique

**Mohamed Essaied Hamrita**

**2021**



---

# TABLE DES MATIÈRES

<b>1</b>	<b>Rappel sur le calcul des probabilités</b>	<b>5</b>
1.1	Notion de probabilité . . . . .	5
1.2	Variables aléatoires . . . . .	6
1.3	Espérance mathématique . . . . .	7
1.4	Fonctions génératrices des moments . . . . .	9
1.5	Fonctions caractéristiques . . . . .	11
1.6	Espérance conditionnelle - Variance conditionnelle . . . . .	12
1.7	Processus stochastique . . . . .	19
	Exercices . . . . .	19
<b>2</b>	<b>Chaînes de Markov</b>	<b>21</b>
2.1	Définitions et exemples . . . . .	21
2.2	Loi des $X_n$ . . . . .	27
2.3	Classification des états . . . . .	31
2.4	Temps d'absorption . . . . .	32
2.5	Simulation d'une chaîne de Markov discrète . . . . .	33
<b>3</b>	<b>Processus de poisson</b>	<b>35</b>
<b>4</b>	<b>Processus Gaussien</b>	<b>37</b>
<b>5</b>	<b>Applications</b>	<b>39</b>
5.1	Exemple one . . . . .	39
5.2	Exemple two . . . . .	39
	<b>Bibliographie</b>	<b>41</b>



# RAPPEL SUR LE CALCUL DES PROBABILITÉS

---

## 1.1 Notion de probabilité

Une notion basique dans la théorie des probabilités est l'**expérience aléatoire** dont on ne connaît pas son résultat en avance. L'ensemble de tous les résultats de l'expérience est l'**ensemble des possibles** ou encore l'**univers** et noté  $\Omega$ .

Un **évènement** est un sous ensemble de l'univers.

On donne quelques exemples :

1. Si l'expérience consiste à jeter une pièce de monnaie, alors  $\Omega = \{P, F\}$ .
2. Si l'expérience consiste à jeter un dé cubique dont ses faces sont numérotées de 1 à 6, alors  $\Omega = \{1, 2, 3, 4, 5, 6\}$ .

Pour chaque évènement  $E$  de l'univers  $\Omega$ , on définit un nombre  $P(E)$  qui satisfait les axiomes suivants :

- **Axiome 1** :  $0 \leq P(E) \leq 1$  ;
- **Axiome 2** :  $P(\Omega) = 1$  ;
- **Axiome 3** : Pour toute séquence des évènements  $E_1, E_2, \dots, E_n$  qui sont mutuellement exclusifs ( $E_i \cap E_j = \emptyset, \forall i \neq j$  et  $\bigcup_{i=1}^n E_i = \Omega$ ), on a

$$P\left(\bigcup_{i=1}^n E_i\right) = \sum_{i=1}^n P(E_i)$$

Quelques conséquences de ces axiomes sont tirées :

- Si  $E \subset F$ , alors  $P(E) \leq P(F)$ .
- $P(\bar{E}) = 1 - P(E)$  où  $\bar{E}$  est le complémentaire de  $E$ .
- $P\left(\bigcup_{i=1}^n E_i\right) = \sum_{i=1}^n P(E_i)$ , lorsque  $E_i$  sont mutuellement exclusifs.
- $P\left(\bigcup_{i=1}^n E_i\right) \leq \sum_{i=1}^n P(E_i)$  (Inégalité de Boole).

**Exemple 1.1.** L'expérience consiste à lancer une pièce de monnaie équilibrée. Donc  $P(\{P\}) = P(\{F\}) = 0.5$

**Exemple 1.2.** On lance un dé cubique équilibré dont ses faces sont numérotées de 1 à 6.

$$P(\{i\}) = \frac{1}{6}, \forall i = 1, 2, \dots, 6.$$

La probabilité d'obtenir un nombre pair est

$$P(\{2, 4, 6\}) = P(\{2\}) + P(\{4\}) + P(\{6\}) = \frac{1}{2}.$$

### 1.1.1 Probabilité conditionnelle

**Définition 1.1.** Soient  $A$  et  $B$  deux évènements tels que  $P(B) \neq 0$ . La probabilité de **A sachant B** est le nombre

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

**Exemple 1.3.** Une urne contient 10 boules numérotées de 1 à 10 indiscernables au toucher. On tire au hasard une boule. Sachant que le numéro de la boule tirée est au moins égale à 5, quelle est la probabilité qu'il soit égale à 10 ?

Soit  $A$  l'évènement d'avoir une boule portant le numéro 10. Soit  $B$  l'évènement d'avoir une boule portant un numéro supérieur ou égale à 5. La probabilité demandée est  $P(A|B)$

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(A)}{P(B)} = \frac{1/10}{6/10} = 1/6$$

### 1.1.2 Indépendance

**Définition 1.2.** On dit que deux évènements,  $A$  et  $B$ , sont **indépendants** si  $P(A \cap B) = P(A) \times P(B)$

Si  $A$  et  $B$  deux évènements indépendants, alors  $P(A|B) = P(A)$  et  $P(B|A) = P(B)$ .

**Exemple 1.4.** On lance deux dés cubiques équilibrés numérotés de 1 à 6. Soit  $A$  l'évènement d'obtenir une somme égale six et  $B$  désigne l'évènement où le premier dé est égal à quatre.

Vérifier que les deux évènements  $A$  et  $B$  sont indépendants.

$P(A \cap B) = p(\{4, 2\}) = 1/36$  et  $P(A) = 1/6$ ,  $P(B) = 1/6$ . Donc  $P(A) \times P(B) = 1/6 \times 1/6 = 1/36 = P(A \cap B)$ . Ainsi, les évènements  $A$  et  $B$  sont indépendants.

## 1.2 Variables aléatoires

**Définition 1.3.** Soit une expérience aléatoire d'univers  $\Omega$ .

Une variable aléatoire  $X$  est une application de l'ensemble  $\Omega$  vers un ensemble de réalisations.

Pour tout évènement  $A$ , on  $P(X \in A) = P(X^{-1}(A))$  où  $X^{-1}(A)$  est l'évènement comprenant tous les éléments  $\omega \in \Omega$  tels que  $X(\omega) \in A$ .

La **fonction de répartition**  $F$  d'une variable aléatoire  $X$  est définie par

$$F(x) = P(X \leq x) = P(X \in ]-\infty, x]), \forall x \in \mathbb{R}$$

Une variable aléatoire  $X$  est dite **discrète** si son ensemble des valeurs possibles est dénombrables. Dans ce cas, on a

$$F(x) = \sum_{k \leq x} P(X = k)$$

Une variable aléatoire  $X$  est dite **continue** s'il existe une fonction  $f(x)$ , appelée *densité de probabilité*, telle que

$$P(X \in B) = \int_B f(x)dx \text{ pour tout ensemble } B$$

Puisque  $F(x) = \int_{-\infty}^x f(x)dx$ , alors

$$f(x) = \frac{d}{dx}F(x)$$

La fonction de répartition **jointe** d'un couple aléatoire  $X$  et  $Y$  est  $F(x, y) = P(X \leq x, Y \leq y)$ .

Les fonctions de répartition de  $X$  et  $Y$ ,

$$F_X(x) = P(X \leq x) \text{ et } F_Y(y) = P(Y \leq y)$$

peuvent être déduites de  $F(x, y)$ . En effet,

$$F_X(x) = \lim_{y \rightarrow \infty} F(x, y) \text{ et } F_Y(y) = \lim_{x \rightarrow \infty} F(x, y)$$

Les variables aléatoires  $X$  et  $Y$  sont **indépendantes** si

$$F(x, y) = F_X(x)F_Y(y)$$

$X$  et  $Y$  sont **continues** s'il existe une fonction  $f(x, y)$ , dite densité de probabilité *jointe*, telle que

$$P(X \in A, Y \in B) = \int_A \int_B f(x, y) dx dy \quad \forall A, B$$

La fonction de répartition d'une suite de  $n$  variables aléatoires  $X_1, X_2, \dots, X_n$  est définie par :

$$F(X_1, X_2, \dots, X_n) = P(X_1 \leq x_1, X_2 \leq x_2, \dots, X_n \leq x_n)$$

et sont indépendantes si

$$F(X_1, X_2, \dots, X_n) = F_{X_1}(x_1)F_{X_2}(x_2) \dots F_{X_n}(x_n)$$

## 1.3 Espérance mathématique

**Définition 1.4.** L'espérance mathématique ou **moyenne** d'une variable aléatoire  $X$ , notée  $\mathbb{E}(X)$ , est définie par :

$$\begin{aligned} \mathbb{E}(X) &= \int_{\mathbb{R}} x dF(x) \\ &= \begin{cases} \sum_x x P(X = x) & \text{si } X \text{ est discrète} \\ \int_{\mathbb{R}} x f(x) dx & \text{si } X \text{ est continue} \end{cases} \end{aligned}$$

De même, on définit l'espérance d'une fonction de  $X$ ,  $g(X)$ , par :

$$\mathbb{E}[g(X)] = \int_{\mathbb{R}} x dF_g(x) = \int_{\mathbb{R}} g(x) dF(x)$$

L'espérance d'une somme de variables aléatoires est la somme des espérances :

$$\mathbb{E} \left[ \sum_{i=1}^n X_i \right] = \sum_{i=1}^n \mathbb{E}(X_i)$$

La **variance** d'une variable aléatoire,  $X$ , est définie par

$$\mathbb{V}(X) = \mathbb{E}[(X - \mathbb{E}(X))^2] = \mathbb{E}(X^2) - \mathbb{E}(X)^2$$

Deux variables aléatoires,  $X$  et  $Y$ , sont dites **non corrélées** si leur covariance, définie par :

$$\text{Cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}(X))(Y - \mathbb{E}(Y))] = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$$

est **nulle**. Noter que si  $X$  et  $Y$  sont indépendantes, alors elles sont non corrélées ( $\text{Cov}(X, Y) = 0$ ).

## Propriétés de la covariance

Pour toutes variables aléatoires  $X, Y, Z$  et  $a \in \mathbb{R}$ , on a :

1.  $\text{Cov}(X, X) = \mathbb{V}(X)$  et  $\text{Cov}(X, Y) = \text{Cov}(Y, X)$
2.  $\text{Cov}(aX, Y) = a\text{Cov}(X, Y)$ .
3.  $\text{Cov}(X, Y + Z) = \text{Cov}(X, Y) + \text{Cov}(X, Z)$ .

Une généralisation de la troisième propriété est donnée par :

$$\text{Cov}\left(\sum_{i=1}^n X_i, \sum_{j=1}^m Y_j\right) = \sum_{i=1}^n \sum_{j=1}^m \text{Cov}(X_i, Y_j)$$

Une expression utile pour la variance de la somme des variables aléatoires peut être déduite comme suit :

$$\begin{aligned} \mathbb{V}\left(\sum_{i=1}^n X_i\right) &= \text{Cov}\left(\sum_{i=1}^n X_i, \sum_{i=1}^n X_i\right) \\ &= \sum_{i=1}^n \sum_{j=1}^n \text{Cov}(X_i, X_j) \\ &= \sum_{i=1}^n \text{Cov}(X_i, X_i) + \sum_{i=1}^n \sum_{i \neq j} \text{Cov}(X_i, X_j) \\ &= \sum_{i=1}^n \mathbb{V}(X_i) + 2 \sum_{i=1}^n \sum_{j < i} \text{Cov}(X_i, X_j) \end{aligned}$$

**Définition 1.5.** Si  $X_1, X_2, \dots, X_n$  sont *indépendantes et identiquement distribuées*, noté  $X_i \sim iid$ , d'espérance  $m$  et de variance  $\sigma^2$ , alors :

- $\bar{X} = \frac{1}{n} \sum_{i=1}^n X_i$  est appelée moyenne empirique.
- $\mathbb{E}(\bar{X}) = m$  et  $\mathbb{V}(\bar{X}) = \frac{\sigma^2}{n}$ .
- $\text{Cov}(\bar{X}, X_i - \bar{X}) = 0, i = 1, 2, \dots, n$ .

**Exemple 1.5.** Calculer la variance d'une variable aléatoire  $X$  suivant une loi binomiale de paramètres  $n$  et  $p$ .

Puisqu'une telle variable aléatoire représente le nombre de succès dans  $n$  essais indépendants lorsque chaque essai a une probabilité commune  $p$  d'être un succès, nous pouvons écrire

$$X = X_1 + X_2 + \dots + X_n$$

où  $X_i \stackrel{iid}{\sim} B(p)$  telle que

$$X_i = \begin{cases} 1 & \text{si le } i\text{ème essai est un succès} \\ 0 & \text{sinon} \end{cases}$$

Par conséquent, on aura  $\mathbb{V}(X) = \sum_{i=1}^n \mathbb{V}(X_i)$ , Or

$$\begin{aligned} \mathbb{V}(X_i) &= \mathbb{E}(X_i^2) - \mathbb{E}(X_i)^2 \\ &= \mathbb{E}(X_i) - \mathbb{E}(X_i)^2 \text{ car } X_i^2 = X_i \\ &= p - p^2 = p(1 - p) \end{aligned}$$

Donc  $\mathbb{V}(X) = np(1 - p)$ .



## 1.4 Fonctions génératrices des moments

**Définition 1.6.** La fonction génératrice des moments  $\phi(t)$  de la variable aléatoire  $X$  est définie pour tout  $t \in \mathbb{R}$  par

$$\begin{aligned}\phi(t) &= \mathbb{E} [e^{tX}] \\ &= \begin{cases} \sum_x e^{tx} P(X=x) & \text{si } X \text{ est discrète} \\ \int_{\mathbb{R}} e^{tx} f(x) dx & \text{si } X \text{ est continue} \end{cases}\end{aligned}$$

$\phi(t)$  est appelée fonction génératrice des moments car tous les moments de  $X$  peuvent être obtenues par les dérivées successives de  $\phi(t)$ . Par exemple

$$\begin{aligned}\phi'(t) &= \frac{d}{dt} \mathbb{E} [e^{tX}] \\ &= \mathbb{E} \left[ \frac{d}{dt} (e^{tX}) \right] = \mathbb{E} [X e^{tX}]\end{aligned}$$

Par conséquent  $\phi'(0) = \mathbb{E}(X)$ .

D'une manière plus générale,  $\phi^n(0) = \mathbb{E}(X^n)$ ,  $n \geq 1$ .

Une propriété importante des fonctions génératrices des moments est que la fonction génératrice des moments de la somme des variables aléatoires indépendantes est simplement le produit des fonctions génératrices des moments individuelles. Pour voir cela, supposons que  $X$  et  $Y$  sont indépendantes et ont respectivement des fonctions génératrices des moments  $\phi_X(t)$  et  $\phi_Y(t)$ . Alors la fonction génératrice des moments de  $X + Y$  est donnée par :

$$\begin{aligned}\phi_{X+Y}(t) &= \mathbb{E} (e^{t(X+Y)}) = \mathbb{E} (e^{tX} e^{tY}) \\ &= \mathbb{E} (e^{tX}) \mathbb{E} (e^{tY}) = \phi_X(t) \phi_Y(t)\end{aligned}$$

**Exemple 1.6** (Loi Binomiale de paramètres  $n$  et  $p$ ).

$$\begin{aligned}\phi(t) &= \mathbb{E} [e^{tX}] = \sum_{k=0}^n e^{tk} C_n^k p^k (1-p)^{n-k} \\ &= \sum_{k=0}^n C_n^k (pe^t)^k (1-p)^{n-k}\end{aligned}$$

Or, d'après la formule de Binôme, on a

$$(a+b)^n = \sum_{k=0}^n C_n^k a^k b^{n-k}$$

d'où  $\phi(t) = (pe^t + (1-p))^n$  et par conséquent

$$\phi'(t) = n(pe^t + 1 - p)^{n-1} pe^t$$

D'où  $\mathbb{E}(X) = \phi'(0) = np$ .

Dérivons une deuxième fois la fonction  $\phi(t)$ , on obtient

$$\phi''(t) = n(n-1)(pe^t + 1 - p)^{n-2} (pe^t)^2 + n(pe^t + 1 - p)^{n-1} pe^t$$

En déduit, alors

$$\mathbb{E}(X^2) = \phi''(0) = n(n-1)p^2 + np$$

Donc, on peut déduire la variance de  $X$ .

$$\mathbb{V}(X) = \mathbb{E}(X^2) - \mathbb{E}(X)^2 = \phi''(0) - (\phi'(0))^2 = np(1-p)$$

**Exemple 1.7** (Loi Normale standard).

$$\begin{aligned}
 \mathbb{E}(e^{tX}) &= \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{tx-x^2/2} dx \\
 &= \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-(x^2-2tx)/2} dx \\
 &= e^{t^2/2} \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-(x-t)^2/2} dx \\
 &= e^{t^2/2}
 \end{aligned}$$

Si  $Y \sim N(m, \sigma^2)$ , alors

$$\phi_Y(t) = \mathbb{E}[e^{t(\sigma X+m)}] = \exp\left[\frac{\sigma^2 t^2}{2} + m\right]$$

Sous R, on peut déterminer et évaluer les fonctions génératrices des moments en utilisant l'extension **MGF** qui est téléchargeable depuis l'adresse suivante : <https://github.com/alexandernel14/MGF>.

Une description de l'installation et l'utilisation de cette extension est donnée comme suit :

```
# installer 'devtools': install.packages("devtools")
# installer 'MGF' depuis github
# devtools::install_github("alexandernel14/MGF", force = T)
# charger l'extension 'MGF'
library(MGF)
# fonction génératrice de la loi binomiale
mgf("Binomial")
```

```
[1] "(1-p+p*exp(t))^n"
```

```
# t=0, ordre 1
MGF_evaluator("Binomial", t=0, order_of_moment=1, n=10, p=0.4)
```

The Moment Generating Function for the Binomial distribution is:

```
[1] "(1-p+p*exp(t))^n"
```

The formula for the 1st moment is:

```
(1 - p + p * exp(t))^(n - 1) * (n * (p * exp(t)))
```

The value of the 1st moment is:

```
[1] 4
```

The formula of the 1st centralized moment is:

```
(1 - p + p * exp(t))^(n - 1) * (n * (p * exp(t)))/(1 - p + p *
exp(t))^n
```

The 1st centralized moment's value is:

```
[1] 4
```

```
# t=0, ordre 2
MGF_evaluator("Binomial", t=0, order_of_moment=2, n=10, p=0.4)
```

The Moment Generating Function for the Binomial distribution is:

```
[1] "(1-p+p*exp(t))^n"
```

The formula for the 2nd moment is:

```
(1 - p + p * exp(t))^(n - 1) * ((n - 1) * (p * exp(t))) *
(n * (p * exp(t))) + (1 - p + p * exp(t))^(n - 1) * (n *
(p * exp(t)))
```

The value of the 2nd moment is:

```
[1] 18.4
```

The formula of the 2nd centralized moment is:

```
((1 - p + p * exp(t))^(n - 1) - 1) * ((n - 1) * (p * exp(t))) *
  (n * (p * exp(t))) + (1 - p + p * exp(t))^(n - 1) * (n *
  (p * exp(t)))/((1 - p + p * exp(t))^n - (1 - p + p * exp(t))^(n -
  1) * (n * (p * exp(t))) * ((1 - p + p * exp(t))^(n - 1) *
  (n * (p * exp(t))))/((1 - p + p * exp(t))^n)^2
```

The 2nd centralized moment's value is:

```
[1] 2.4
```

```
# fonction génératrice de la loi normale
mgf("Normal")
```

```
[1] "exp(mu*t + 0.5*sigma^2*t^2)"
```

```
# t=0, ordre 1
MGF_evaluator("Normal",t=0, order_of_moment=1)
```

The Moment Generating Function for the Normal distribution is:

```
[1] "exp(mu*t + 0.5*sigma^2*t^2)"
```

The formula for the 1st moment is:

```
exp(mu * t + 0.5 * sigma^2 * t^2) * (mu + 0.5 * sigma^2 * (2 *
  t))
```

The value of the 1st moment is:

```
[1] 0
```

The formula of the 1st centralized moment is:

```
mu + 0.5 * sigma^2 * (2 * t)
```

The 1st centralized moment's value is:

```
[1] 0
```

```
# t=0, ordre 2
MGF_evaluator("Normal",t=0, order_of_moment=2)
```

The Moment Generating Function for the Normal distribution is:

```
[1] "exp(mu*t + 0.5*sigma^2*t^2)"
```

The formula for the 2nd moment is:

```
exp(mu * t + 0.5 * sigma^2 * t^2) * (mu + 0.5 * sigma^2 * (2 *
  t)) * (mu + 0.5 * sigma^2 * (2 * t)) + exp(mu * t + 0.5 *
  sigma^2 * t^2) * (0.5 * sigma^2 * 2)
```

The value of the 2nd moment is:

```
[1] 1
```

The formula of the 2nd centralized moment is:

```
0.5 * sigma^2 * 2
```

The 2nd centralized moment's value is:

```
[1] 1
```

## 1.5 Fonctions caractéristiques

Il existe des variables aléatoires pour lesquelles la fonction génératrice des moments n'existe pas. Dans ce cas, on peut faire recours à la fonction caractéristique définie ci-dessous.

**Définition 1.7.** La **fonction caractéristique** d'une variable aléatoire  $X$  est la fonction à valeurs complexes définie sur  $\mathbb{R}$  par

$$\begin{aligned}\psi_X(t) &= \mathbb{E}[e^{itX}] \\ &= \begin{cases} \sum_{i=1}^n e^{itx_i} P(X = x_i) & \text{si } X \text{ est discrète} \\ \int_{\mathbb{R}} e^{itx} f(x) dx & \text{si } X \text{ est continue} \end{cases}\end{aligned}$$

où  $i = \sqrt{-1}$ .

**Proposition 1.1.** Soit  $X$  une variable aléatoire,  $a$  et  $b$  deux réels. Alors les propriétés suivantes sont toujours vraies :

1. Pour tout  $t \in \mathbb{R}$ ,  $|\psi_X(t)| \leq 1$ .
2.  $\psi_X(0) = 1$ .
3. Pour tout  $t \in \mathbb{R}$ ,  $\psi_X(-t) = \overline{\psi_X(t)}$ .
4. Pour tout  $t \in \mathbb{R}$ ,  $\psi_{aX+b}(t) = e^{itb} \psi_X(at)$ .
5.  $\psi_X(t)$  est continue sur  $\mathbb{R}$ .

**Proposition 1.2.** Si le moment d'ordre  $n$  d'une variable aléatoire  $X$  existe, alors la fonction caractéristique de  $X$  est  $n$  fois dérivable et :

$$\mathbb{E}(X^n) = \frac{1}{i^n} \psi_X^{(n)}(0)$$

**Exemple 1.8** (Loi normal standard). La fonction caractéristique de la loi normale standard est :

$$\psi(t) = e^{-t^2/2}$$

**Exemple 1.9** (Loi exponentielle). La fonction caractéristique de la loi exponentielle de paramètre  $\lambda$  est :

$$\psi(t) = \frac{\lambda}{it - \lambda}$$

Si  $X_1, X_2, \dots, X_n$  sont des variables indépendantes, alors

$$\psi_{\sum_i X_i}(t) = \prod_i \psi_{X_i}(t)$$

## 1.6 Espérance conditionnelle - Variance conditionnelle

L'un des concepts les plus utiles de la théorie des probabilités est celui de la probabilité conditionnelle et de l'espérance conditionnelle. La raison est double. Premièrement, dans la pratique, nous nous intéressons souvent au calcul des probabilités et des espérances lorsqu'une information partielle est disponible ; par conséquent, les probabilités et les espérances souhaitées sont conditionnelles. Deuxièmement, pour calculer une probabilité, il est souvent extrêmement utile d'abord *conditionner* une variable aléatoire appropriée.

### 1.6.1 Densité conditionnelle

Soient  $X$  et  $Y$  deux variables aléatoires de densité jointe  $f(x, y)$ .

**Définition 1.8.** On définit la densité **conditionnelle** de  $X$  sachant  $Y = y$  par :

$$f_{X|Y}(x, y) = \begin{cases} \frac{P(X = x, Y = y)}{P_Y(Y = y)} & \text{si } X \text{ et } Y \text{ sont discrètes} \\ \frac{f(x, y)}{f_Y(y)} & \text{si } X \text{ et } Y \text{ sont continues} \end{cases}$$

**Exemple 1.10.** Soit un couple aléatoire discrète  $(X, Y)$  définie par la densité jointe suivante :

$$P(X = 1, Y = 1) = 0.5; P(X = 1, Y = 2) = 0.1; P(X = 2, Y = 1) = 0.1; P(X = 2, Y = 2) = 0.3$$

Déterminer la loi conditionnelle  $f_{X|Y=1}(x, y)$ .

La densité conditionnelle est par définition

$$f_{X|Y=1}(x, y) = \frac{P(X = x, Y = 1)}{P_Y(Y = 1)}$$

Or,  $P_Y(Y = 1) = \sum_x P(X = x, Y = 1) = p(1, 1) + p(2, 1) = 0.6$ . D'où

$$f_{X|Y=1}(1, 1) = \frac{P(X = 1, Y = 1)}{P_Y(Y = 1)} = \frac{p(1, 1)}{P_Y(1)} = \frac{5}{6}$$

$$f_{X|Y=1}(2, 1) = \frac{P(X = 2, Y = 1)}{P_Y(Y = 1)} = \frac{p(2, 1)}{P_Y(1)} = \frac{1}{6}$$

### 1.6.2 Espérance conditionnelle

Une espérance conditionnelle est une expression calculée depuis une distribution conditionnelle. On écrit  $\mathbb{E}(Y|X = x)$  pour l'espérance de  $Y$  sachant  $X = x$ .

**Définition 1.9.** L'espérance conditionnelle de  $Y$  sachant  $X = x$  est

$$\mathbb{E}(Y|X = x) = \begin{cases} \sum_y y \mathbb{P}(Y = y|X = x) & \text{cas discret} \\ \int_{\mathbb{R}} y f_{Y|X}(x, y) dy & \text{cas continu} \end{cases}$$

**Exemple 1.11.** On donne  $Y = \begin{cases} 1 & \text{avec une probabilité } \frac{1}{8} \\ 2 & \text{avec une probabilité } \frac{7}{8} \end{cases}$

et  $X|Y = \begin{cases} 2Y & \text{avec une probabilité } \frac{3}{4} \\ 3Y & \text{avec une probabilité } \frac{1}{4} \end{cases}$

Déterminer  $\mathbb{E}(X|Y = y)$ .

Si  $Y = 1$ , alors  $(X|Y = 1) = \begin{cases} 2 & \text{avec une probabilité } \frac{3}{4} \\ 3 & \text{avec une probabilité } \frac{1}{4} \end{cases}$

D'où  $\mathbb{E}(X|Y = 1) = \sum_x x \mathbb{P}(X|Y = 1) = 2 \times \frac{3}{4} + 3 \times \frac{1}{4} = \frac{9}{4}$ .

Si  $Y = 2$ , alors  $(X|Y = 2) = \begin{cases} 4 & \text{avec une probabilité } \frac{3}{4} \\ 6 & \text{avec une probabilité } \frac{1}{4} \end{cases}$

D'où  $\mathbb{E}(X|Y = 2) = \sum_x x \mathbb{P}(X|Y = 2) = 4 \times \frac{3}{4} + 6 \times \frac{1}{4} = \frac{18}{4}$ .

Donc  $\mathbb{E}(X|Y = y) = \begin{cases} \frac{9}{4} & \text{si } Y = 1 \text{ avec une prob} = \frac{1}{8} \\ \frac{18}{4} & \text{si } Y = 2 \text{ avec une prob} = \frac{7}{8} \end{cases}$ .

**Exemple 1.12.** Soit la densité jointe d'un couple aléatoire définie par :

$$f(x, y) = \frac{2}{xy}, \quad \text{pour } 1 < y < x < e$$

Déterminer  $\mathbb{E}(Y|X = x)$

Par définition  $\mathbb{E}(Y|X=x) = \int_{\mathbb{R}} y f_{X|Y}(x, y) dy$ .

Or  $f_{X|Y}(x, y) = \frac{f(x, y)}{f_X(x)}$  avec

$$\begin{aligned} f_X(x) &= \int_1^x \frac{2}{xy} dy = \frac{2}{x} \left[ \ln(y) \right]_1^x \\ &= \frac{2 \ln(x)}{x}, \text{ pour } 1 < x < e. \end{aligned}$$

D'où  $f_{X|Y}(x, y) = \frac{2}{xy \ln(x)} = \frac{1}{y \ln(x)}$  pour  $1 < y < x$ , et par conséquent,

$$\mathbb{E}(Y|X=x) = \int_1^x \frac{y}{y \ln(x)} dy = \frac{1}{\ln(x)} \left[ y \right]_1^x = \frac{x-1}{\ln(x)}$$

### Propriétés de l'espérance conditionnelle

**1. Linéarité :** Pour toutes constantes  $a, b$  et  $X, Y$  et  $Z$  des variables aléatoires,

$$\mathbb{E}(aY + bZ|X=x) = a\mathbb{E}(Y|X=x) + b\mathbb{E}(Z|X=x)$$

**2. Indépendance :** Si  $X$  et  $Y$  sont deux variables aléatoires indépendantes, alors  $\mathbb{E}(Y|X=x) = \mathbb{E}(Y)$

**3.**  $\mathbb{E}(g(X)|X=x) = \mathbb{E}(X)$  où  $g$  est une transformation.

**4. Espérance totale :**  $\mathbb{E}(Y) = \mathbb{E}(\mathbb{E}(Y|X=x))$  : La moyenne totale est la moyenne des moyennes.

**Exemple 1.13.** Reprenons l'exemple 1.11.

$$\mathbb{E}(X|Y=y) = \begin{cases} \frac{9}{4} \text{ avec une prob} = \frac{1}{8} \\ \frac{18}{4} \text{ avec une prob} = \frac{7}{8} \end{cases}.$$

Donc, l'espérance totale est :

$$\mathbb{E}(X) = \frac{9}{4} \times \frac{1}{8} + \frac{18}{4} \times \frac{7}{8} = \frac{135}{32}$$

**Exercice 1.1.** On jette une pièce de monnaie équilibrée. Soit  $Y$  la variable aléatoire désignant le nombre des lancers avant d'obtenir Face pour la première fois.

1. Déterminer la loi de  $Y$  ainsi que son espérance et sa variance.
2. Simuler 10 000 réalisations de  $Y$ .
3. Soit  $(X|Y=y) \sim \mathcal{P}(\lambda Y)$ .
  - a) Calculer  $\mathbb{E}(X|Y=y)$  et  $\mathbb{E}(X)$ .
  - b) En prenant  $\lambda = 2$  et à l'aide des simulations, donner la valeur de  $\mathbb{E}(X|Y=y)$ .

*Solution .* 1. Soit  $P(F) = p$  et  $P(P) = 1 - p$ . On a  $(n-1)$  premières épreuves donnant  $P$  et la  $n$ ème épreuve donne  $F$ , et puisque les épreuves sont indépendantes, alors

$$\mathbb{P}(Y=n) = (1-p)^{n-1} p, \quad n=1, 2, \dots$$

Donc  $Y \sim G(p)$ .

2. Soit  $Z = (X|Y=y) \sim \mathcal{P}(\lambda Y)$ ,  $\mathbb{E}(Z) = \mathbb{E}(X|Y=y) = \lambda Y$ .

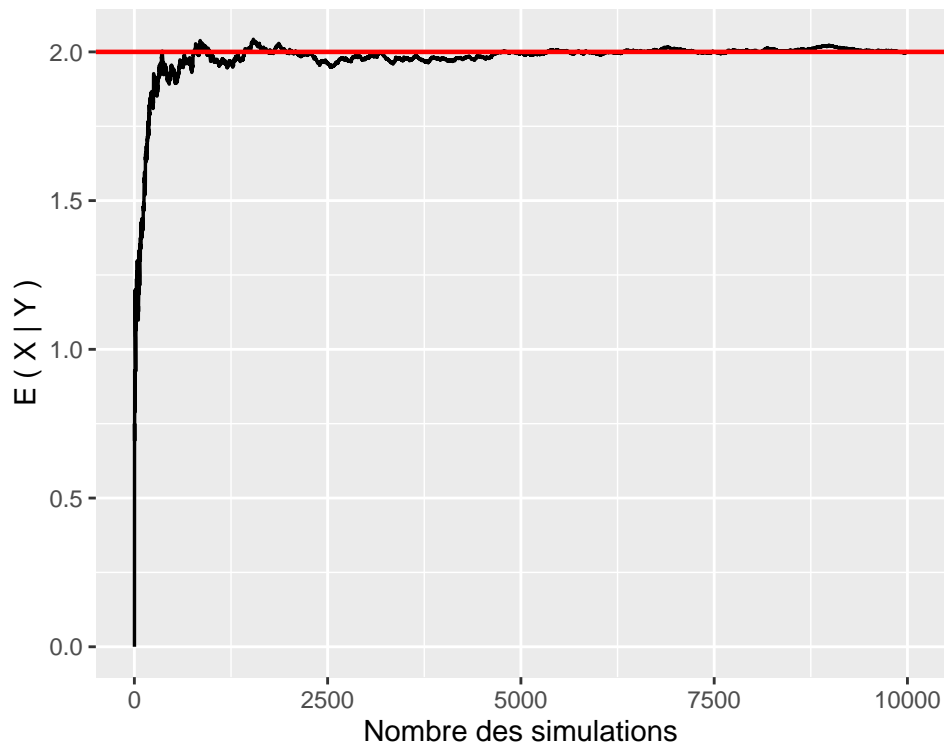
$$\begin{aligned} \mathbb{E}(X) &= \mathbb{E}(\mathbb{E}(X|Y=y)) = \mathbb{E}(\lambda Y) \\ &= \lambda \mathbb{E}(Y) = \lambda \frac{1-p}{p} \end{aligned}$$

```
set.seed(1)
y=rgeom(10000, prob=0.5)
```

```
set.seed(1)
x=rpois(10000, lambda = 2*y)
mean(x)
```

```
## [1] 2.0029
```

La figure suivante donne l'évolution de  $\mathbb{E}(X|Y = y)$  en fonction du nombre des simulations.



### 1.6.3 Variance conditionnelle

Similairement à l'espérance conditionnelle, la variance conditionnelle est une variance prise par rapport à une distribution conditionnelle. Étant donné les variables aléatoires  $X$  et  $Y$ , soit  $m_X = \mathbb{E}(Y|X = x)$ . La variance conditionnelle  $\mathbb{V}(Y|X = x)$  est définie comme suit :

**Définition 1.10.** La variance conditionnelle de  $Y$  sachant  $X = x$  est

$$\mathbb{V}(Y|X = x) = \begin{cases} \sum (y - m_X) \mathbb{P}(Y = y|X = x) & \text{cas discret} \\ \int_{\mathbb{R}} (y - m_X)^2 f_{Y|X}(x, y) dy & \text{cas continu} \end{cases}$$

La variance conditionnelle possède les propriétés suivantes :

#### Propriétés de la variance conditionnelle

1.  $\mathbb{V}(Y|X = x) = \mathbb{E}(Y^2|X = x) - \mathbb{E}(Y|X = x)^2$ .

2.  $\mathbb{V}(aY + b|X = x) = a^2\mathbb{V}(Y|X = x).$

3. **Variance totale :**  $\mathbb{V}(Y) = \mathbb{E}(\mathbb{V}(Y|X)) + \mathbb{V}(\mathbb{E}(Y|X)).$

**Exemple 1.14.** Soit  $X$  une variable aléatoire uniforme sur  $]0, 1[$ . Si  $X = x$ , alors  $Y$  suit une loi uniforme sur  $]0, x[$ . Déterminer la variance de  $Y$ .

La distribution conditionnelle de  $Y|X = x$  est uniforme sur  $]0, x[$ . On déduit alors,

$$\mathbb{E}(Y|X = x) = \frac{x}{2} \text{ et } \mathbb{V}(Y|X = x) = \frac{x^2}{12}$$

La propriété de la variance totale donne,

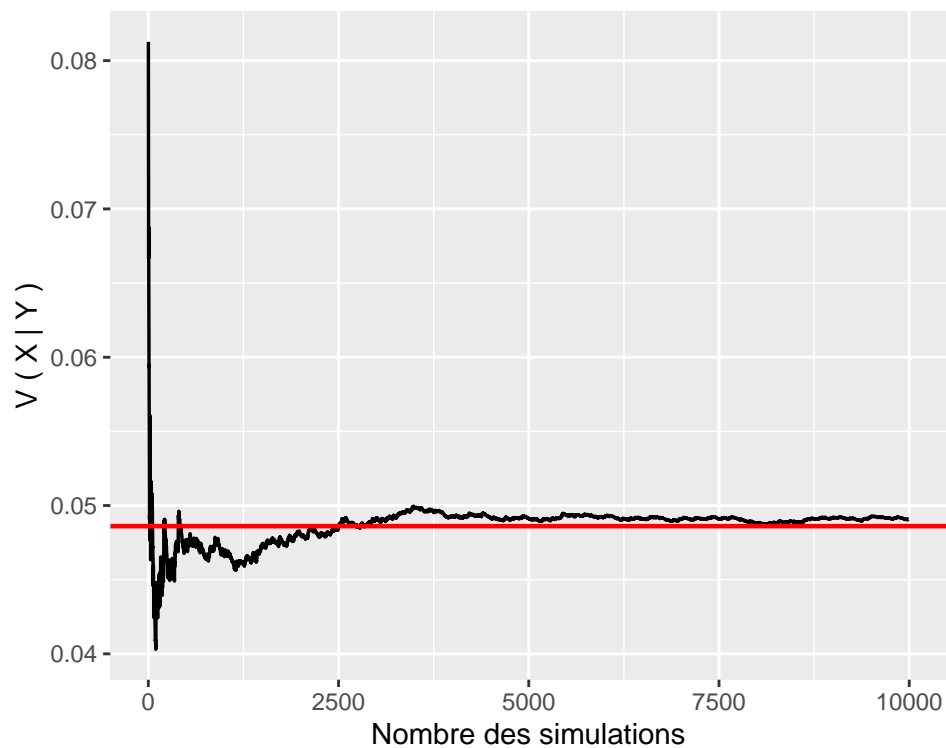
$$\begin{aligned} \mathbb{V}(Y) &= \mathbb{E}(\mathbb{V}(Y|X)) + \mathbb{V}(\mathbb{E}(Y|X)) = \mathbb{E}\left(\frac{X^2}{12}\right) + \mathbb{V}\left(\frac{X}{2}\right) \\ &= \frac{1}{12}\mathbb{E}(X^2) + \frac{1}{4}\mathbb{V}(X) = \frac{1}{12} \times \frac{1}{3} + \frac{1}{4} \times \frac{1}{12} = \frac{7}{144} = 0.04861. \end{aligned}$$

Sous R, on peut procéder comme suit :

```
set.seed(1)
x=runif(10000,0,1) # réalisations uniformes sur ]0,1[
y=runif(10000,0,x) # réalisations uniformes sur ]0,x[
var(y)
```

[1] 0.04906725

L'évolution de la variance conditionnelle suivant le nombre des simulations est donnée par le graphique suivant :





### 1.6.4 Théorèmes limites

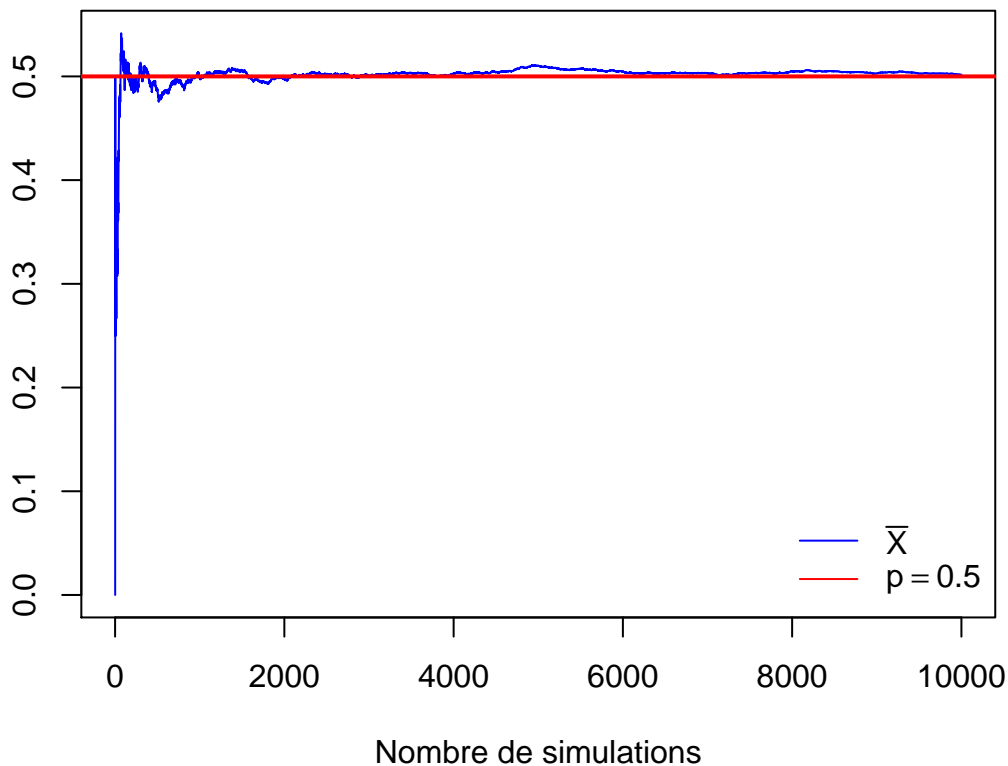
**Théorème 1.1** (Loi forte des grands nombres). Si  $X_1, X_2, \dots, X_n$  sont des variables aléatoires iid de moyenne  $m$ , alors

$$\mathbb{P} \left[ \lim_{n \rightarrow \infty} \bar{X}_n = m \right] = 1$$

La moyenne des  $n$  premiers termes d'une suite de variables aléatoires iid converge presque sûrement vers l'espérance mathématique  $\mathbb{E}(X_i) = m$ , lorsque  $n$  tend vers l'infini.

Vérifions ce théorème à l'aide des simulations. On considère l'expérience suivante : On lance une pièce de monnaie équilibrée  $n$  fois. Soit  $X_i = 1$  si Face et  $X_i = 0$  si Pile. D'après la loi des grands nombres,  $\bar{X}_n$  converge vers  $\mathbb{E}(X_i) = p = 0.5$ .

```
set.seed(1)
s=sample(c(0,1),10000, replace = T)
x_bar=cumsum(s)/(1:10000)
plot(x_bar, type="l", col="blue", xlab="Nombre de simulations",
      ylab="")
abline(h=0.5, col="red", lwd=2)
legend("bottomright", legend=c(expression(bar(X)), expression(p==0.5)),
      lty=1, col=c("blue", "red"), bty="n")
```



**Théorème 1.2** (Loi faible des grands nombres). Si  $X_1, X_2, \dots, X_n$  sont des variables aléatoires iid de moyenne  $m$ , alors

$$\lim_{n \rightarrow \infty} \mathbb{P} \left[ |\bar{X}_n - m| < \epsilon \right] = 1, \quad \forall \epsilon > 0$$

**Théoreme 1.3** (Théorème central limite). Si  $X_1, X_2, \dots, X_n$  sont des variables aléatoires iid( $m, \sigma^2$ ), alors

$$\lim_{n \rightarrow \infty} \mathbb{P} \left[ \frac{\bar{X}_n - m}{\sigma/\sqrt{n}} < t \right] = \mathbb{P}(Z < t), \quad \text{avec } Z \sim N(0, 1)$$

### Illustration sous R

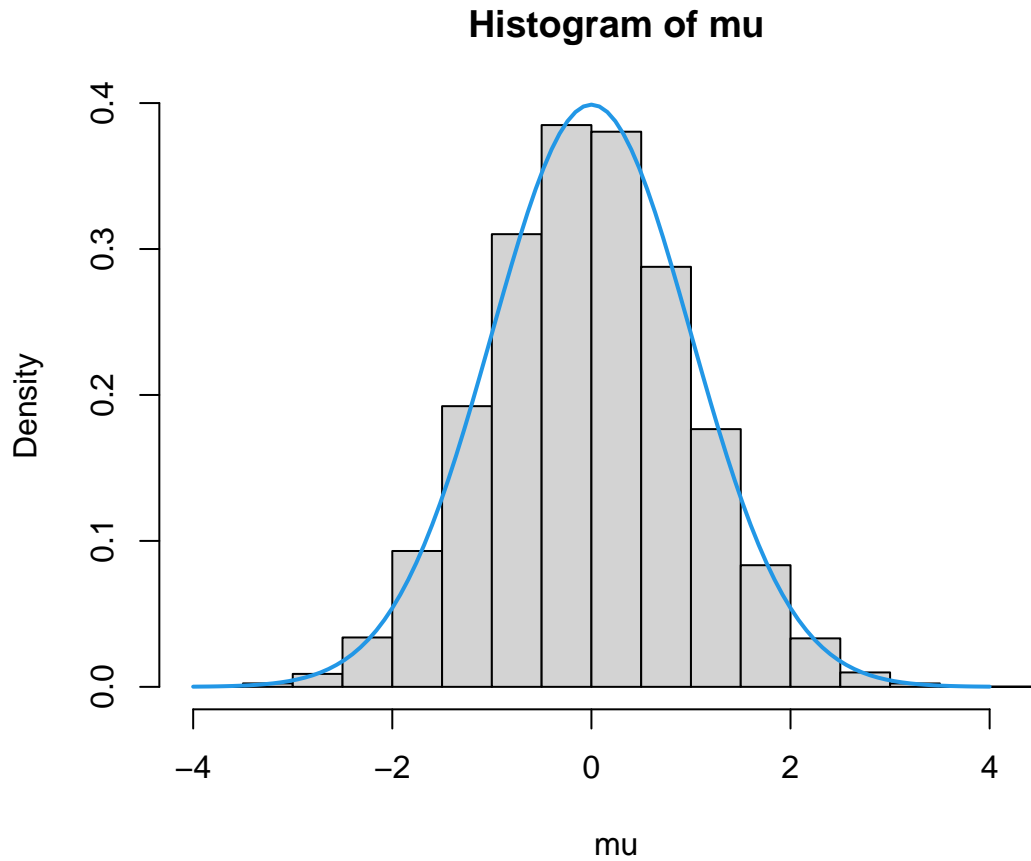
Soit  $X_1, X_2, \dots, X_{100}$  une suite de variables aléatoires telle que  $X_i \stackrel{iid}{\sim} \mathcal{P}(\lambda = 4)$ . On sait que  $\mathbb{E}(X_i) = \mathbb{V}(X_i) = \lambda = 4$ .

1. Simuler 100 réalisations du Poisson de paramètre  $\lambda = 4$  puis déduire la variable de  $z$

$$z = \frac{\bar{X} - m}{\sigma/\sqrt{n}}$$

2. Répéter 100000 les instructions précédentes et stocker le résultat dans un objet que l'on appelle `mu`.
3. Représenter l'histogramme de `mu`. Ajouter sur le même graphique la courbe de la densité de la loi normale standard.

```
# 1
set.seed(1)
simPois=rpois(n=100, lambda=4)
z=(mean(simPois)-4)/(2/sqrt(100))
# 2
Z=function(){
  simPois=rpois(n=100, lambda=4)
  (mean(simPois)-4)/(2/sqrt(100))
}
set.seed(1)
mu=replicate(100000, Z())
# 3
hist(mu,freq = F)
curve(dnorm,-4,4,add=T,col=4,lwd=2)
```



## 1.7 Processus stochastique

### Exercices

#### Exercice 1

Les citoyens de Sousse retirent de l'argent d'un distributeur de billets selon la fonction de probabilité suivante :

$x_i$ en DT	100	200	500
$\mathbb{P}(X = x_i)$	0.25	0.55	0.2



# CHAÎNES DE MARKOV

---

La notion de ce qu'on appelle aujourd'hui une chaîne de Markov a été conçue par le mathématicien russe A.A. Markov.

Une chaîne de Markov est un système mathématique qui subit des transitions d'un état à un autre selon un ensemble donné de règles probabilistes. Les chaînes de Markov sont des processus stochastiques, mais ils diffèrent en ce sens qu'ils doivent manquer de "mémoire". Autrement dit, la probabilité de l'état suivant du système ne dépend que **seulement** de l'état actuel du système.

Les chaînes de Markov sont largement utilisées dans de nombreux domaines tels que la finance, la théorie des jeux et la génétique.

## 2.1 Définitions et exemples

Soit  $X_0, X_1, \dots, X_n, \dots$  une suite de variables aléatoires sur un espace de probabilité  $(\Omega, \mathcal{A}, \mathbb{P})$  à valeurs dans  $E$  (fini ou infini dénombrable).

**Définition 2.1.** La suite  $(X_n)_{n \in \mathbb{N}}$  est une chaîne de Markov si pour tout  $n \in \mathbb{N}$  et pour tout  $(x_0, x_1, \dots, x_{n+1}) \in E^{n+2}$  tel que  $\mathbb{P}(X_n = x_n, X_{n-1} = x_{n-1}, \dots, X_0 = x_0) > 0$ ,  $\mathbb{P}(X_{n+1} = x_{n+1} | X_n = x_n, X_{n-1} = x_{n-1}, \dots, X_0 = x_0)$  ne dépend que des valeurs  $n, x_n$  et  $x_{n+1}$ .

L'ensemble  $E$  est appelé l'espace des états de la chaîne de Markov et les  $x_i$  sont appelés états.

**Proposition 2.1.** Soit  $(X_n)_{n \in \mathbb{N}}$  une chaîne de Markov d'espace d'états  $E$ . Pour tout  $n \in \mathbb{N}$  et pour tout  $(x_0, x_1, \dots, x_{n+1}) \in E^{n+2}$ , on a

$$\mathbb{P}(X_{n+1} = x_{n+1} | X_n = x_n, X_{n-1} = x_{n-1}, \dots, X_0 = x_0) = \mathbb{P}(X_{n+1} = x_{n+1} | X_n = x_n)$$

**Définition 2.2.** Une chaîne de Markov  $(X_n)_{n \in \mathbb{N}}$  est dite homogène (dans le temps) si  $\mathbb{P}(X_{n+1} = x_{n+1} | X_n = x_n)$  ne dépend pas de  $n$ .

La dynamique du processus est alors entièrement caractérisé par les  $p_{ij} = \mathbb{P}(X_{n+1} = j | X_n = i)$ , appelés probabilités de transitions de l'état  $i$  à l'état  $j$  si la chaîne est homogène, ou plus généralement

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

**Définition 2.3.** Une matrice  $P = (P_{ij}, i, j \in E)$  est une matrice stochastique, ou matrice de transition si chaque ligne  $P_i = (P_{ij}, j \in E)$  est une distribution de probabilité.

Pour tout  $i, j \in E$ ,  $p_{ij}$  peut alors s'interpréter comme la probabilité d'aller à l'état  $j$  sachant qu'on se trouve à l'état  $i$  à l'instant précédent.

Par exemple, la matrice suivante est une matrice stochastique

$$P = \begin{pmatrix} 0.2 & 0.8 \\ 0.4 & 0.6 \end{pmatrix}$$

```
(P<-matrix(c(0.2,0.4,0.8,0.6),nc=2))
```

```
      [,1] [,2]
[1,]  0.2  0.8
[2,]  0.4  0.6
```

```
# vérifions que P est une matrice stochastique
P>=0 & P<=1 # 0<= p <=1
```

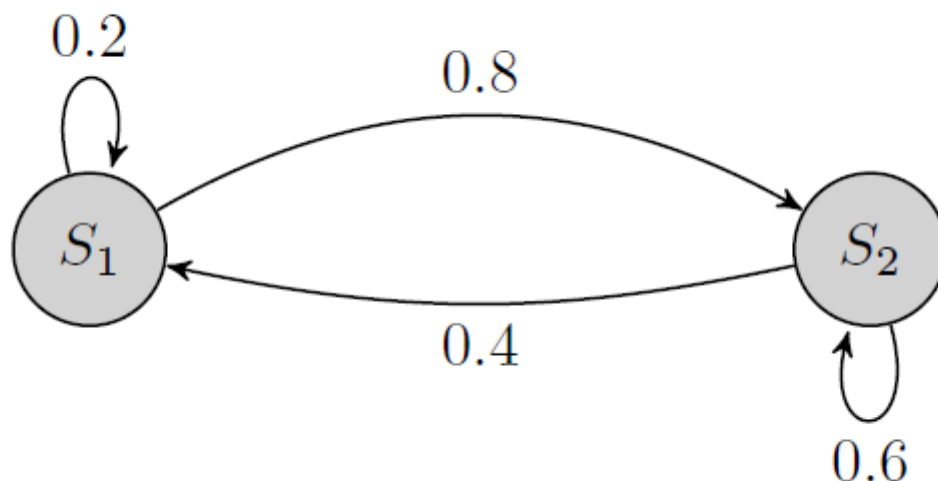
```
      [,1] [,2]
[1,] TRUE TRUE
[2,] TRUE TRUE
```

```
rowSums(P)==1 # somme pj = 1
```

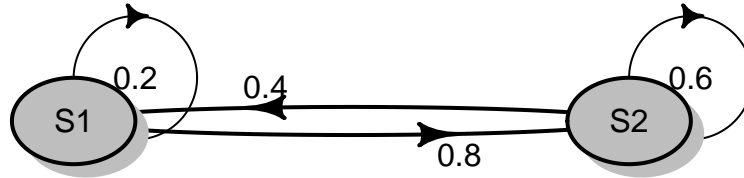
```
[1] TRUE TRUE
```

A toute matrice de transition, on peut associer un [graphe orienté](#). Les sommets sont les états de la chaîne, et l'orientation est donnée par la probabilité  $p_{ij} > 0$ .

Le graphe de la chaîne associé à la matrice de transition définie ci-dessus est donné par :



```
# représenter un graphe sous R
# install.packages(diagram)
library(diagram) # charger l'extension 'diagram'
P<-matrix(c(0.2,0.4,0.8,0.6),nc=2)
colnames(P) <- c("S1", "S2")
rownames(P) <- c("S1", "S2")
plotmat(t(P), pos=c(2),box.size = 0.05, box.prop = 0.7,
        box.col = "gray", arr.pos = 0.6, relsize = 1.1)
```



**Exemple 2.1.** Lors d'une ronde, une méthode simple pour tromper l'ennemi et d'avance ou reculer d'une manière aléatoire, en tirant pile ou face. On désigne par  $X_n$  la position de garde à l'instant  $n$ . L'évolution peut se modéliser à l'aide d'une matrice de transition

$$P = \begin{matrix} & \begin{matrix} N & E & S & O \end{matrix} \\ \begin{matrix} N \\ E \\ S \\ O \end{matrix} & \begin{pmatrix} 0 & 0.5 & 0 & 0.5 \\ 0.5 & 0 & 0.5 & 0 \\ 0 & 0.5 & 0 & 0.5 \\ 0.5 & 0 & 0.5 & 0 \end{pmatrix} \end{matrix}$$

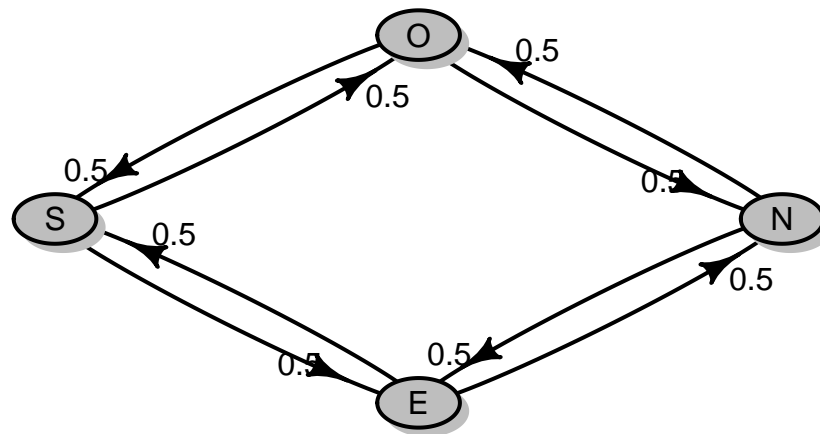
L'ensemble des états,  $E$ , est formé de 4 états  $\{N, E, S, O\}$ . La probabilité de se déplacer à l'état  $E$  sachant qu'on se trouve à l'état  $N$  est :

$$\mathbb{P}(X_{n+1} = E | X_n = N) = p_{12} = 0.5$$

Le graphe orienté correspondant à cette matrice de transition est :

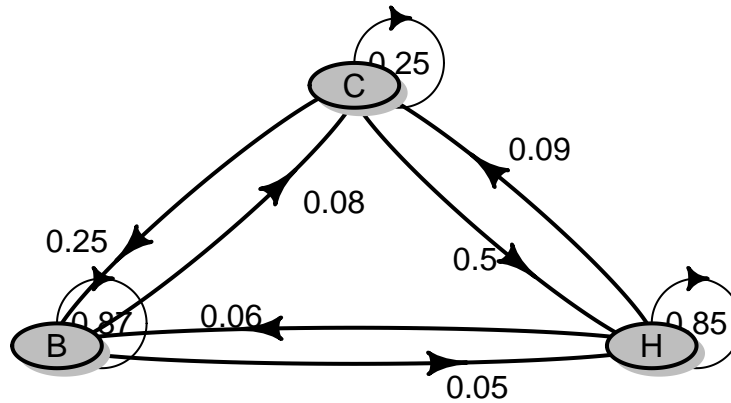
```

nn=c("N", "E", "S", "O")
P1=matrix(rep(c(0,0.5,0,0.5,0.5,0,0.5,0),2), nc=4,
          dimnames = list(nn,nn))
plotmat(t(P1), box.size = 0.05, box.prop = 0.6, box.col = "gray",
        arr.pos = 0.7, relsize = 0.9)
  
```



**Exemple 2.2.** La tendance d'un marché boursier peut être haussière (H), ou baissière (B) ou en consolidation (marché en range) (C). Le graphe associé est donné comme suit :





On déduit la matrice de transition

$$P = \begin{pmatrix} 0.85 & 0.06 & 0.09 \\ 0.05 & 0.87 & 0.08 \\ 0.5 & 0.25 & 0.25 \end{pmatrix}$$

On note qu'un jour haussier est suivi par un autre jour haussier dans 85% des temps, par un jour baissier dans 6% des temps et par un jour de consolidation dans 9% des temps.

**Exercice 2.1.** Dans un certain pays, il ne fait jamais beau deux jours de suite. Si un jour il fait beau, le lendemain il peut neiger ou pleuvoir avec autant de chances. Si un jour il pleut ou il neige, il y a une chance sur deux qu'il ait changement de temps le lendemain, et s'il y a changement, il y a une chance sur deux que ce soit pour du beau temps.

1. Former une chaîne de Markov et en déterminer sa matrice de transition. Représenter le graphe orienté associé.
2. Si on suppose l'on a que deux états (beau temps et mauvais temps), déterminer la matrice de la nouvelle chaîne ainsi obtenue.

### Solution 1.1

1. On a l'ensemble des états suivants  $E = \{B, P, N\}$ . Le temps pour un jour ne dépend que seulement de temps du jour précédent. On a donc un processus de Markov.

Il ne fait jamais beau deux jours de suite  $\Rightarrow \mathbb{P}(X_{n+1} = B | X_n = B) = 0$ .

Si un jour il fait beau, le lendemain il peut neiger ou pleuvoir avec autant de chances  $\Rightarrow \mathbb{P}(X_{n+1} = P | X_n = B) = \mathbb{P}(X_{n+1} = N | X_n = B) = 0.5$ .

Si un jour il pleut ou il neige, il y a une chance sur deux qu'il ait changement de temps le lendemain  $\Rightarrow \mathbb{P}(X_{n+1} = P | X_n = P) = \mathbb{P}(X_{n+1} = N | X_n = P) = 0.5$ .

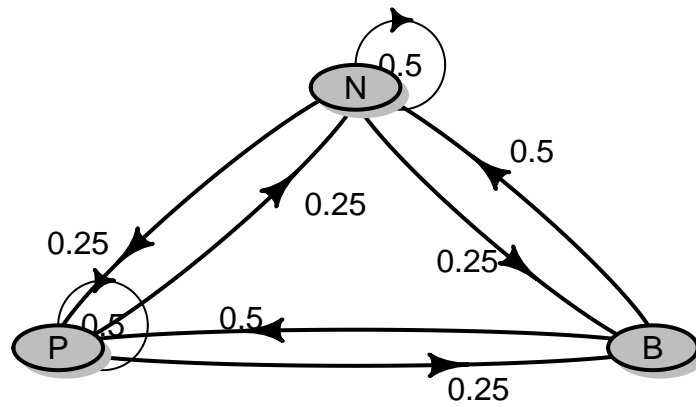
et s'il y a changement, il y a une chance sur deux que ce soit pour du beau temps  $\mathbb{P}(X_{n+1} = B|X_n = P) = \mathbb{P}(X_{n+1} = B|X_n = N) = 0.5 \times 0.5 = 0.25$ .

Ainsi, la matrice stochastique est donnée par :

$$P = \begin{pmatrix} 0 & 0.5 & 0.5 \\ 0.25 & 0.5 & 0.25 \\ 0.25 & 0.25 & 0.5 \end{pmatrix}$$

et le graphe orienté associé est :

```
noms=c("B","P","N")
PB=matrix(c(0,0.25,0.25,0.5,0.5,0.25,0.5,0.25,0.5), nc=3,
          dimnames = list(noms,noms))
plotmat(t(PB), box.size = 0.06, box.prop = 0.5, box.col = "gray",
        arr.pos = 0.6, relsize = 0.85, dtext = 0.6)
```



2. Le nouvel ensemble des états est  $E_1 = \{B, M\}$ ,  $M$  pour désigner mauvais temps (soit il pleut, soit il neige).

$$\mathbb{P}(X_{n+1} = B|X_n = M) = \mathbb{P}(X_{n+1} = B|X_n = P) + \mathbb{P}(X_{n+1} = B|X_n = N) = 0.5 + 0.5 = 1$$

$$\mathbb{P}(X_{n+1} = B|X_n = B) = 0.$$

$$\begin{aligned} \mathbb{P}(X_{n+1} = M|X_n = M) &= \mathbb{P}(X_{n+1} = P|X_n = P) + \mathbb{P}(X_{n+1} = P|X_n = N) \\ &= \mathbb{P}(X_{n+1} = N|X_n = N) + \mathbb{P}(X_{n+1} = N|X_n = N) \\ &= \frac{1}{2} + \frac{1}{4} = \frac{3}{4} \end{aligned}$$

$$\begin{aligned}\mathbb{P}(X_{n+1} = M | X_n = B) &= \mathbb{P}(X_{n+1} = P | X_n = B) = \mathbb{P}(X_{n+1} = N | X_n = B) \\ &= \frac{1}{4}\end{aligned}$$

Ainsi, la matrice de transition d'espace d'états  $\{B, M\}$  est :

$$P_1 = \begin{pmatrix} 0 & 1 \\ \frac{1}{4} & \frac{3}{4} \end{pmatrix}$$

## 2.2 Loi des $X_n$

Le comportement d'une chaîne de Markov  $X$  dépend entièrement de sa matrice de transition  $P$ , et de la position initiale  $X_0$ . On appelle  $\mu_0$  la **loi initiale** de  $X$ , c'est une mesure définie par

$$\mu_0(x) = \mathbb{P}(X_0 = x)$$

Connaissant  $\mu_0$  et  $P$ , on peut calculer directement la loi de  $X_n$ .

La distribution de  $X_1$  est :

$$\begin{aligned}\mathbb{P}(X_1 = j) &= \sum_{i=1}^N \mathbb{P}(X_1 = j | X_0 = i) \mathbb{P}(X_0 = i) \\ &= \sum_{i=1}^N \mu_0(i) p_{ij} = (\mu_0 P)_j \text{ pour tout } j\end{aligned}$$

La distribution de  $X_2$  est :

$$\begin{aligned}\mathbb{P}(X_2 = j) &= \sum_{i=1}^N \mathbb{P}(X_2 = j | X_0 = i) \mathbb{P}(X_0 = i) \\ &= \sum_{i=1}^N \mu_0(i) (P^2)_{ij} = (\mu_0 P^2)_j \text{ pour tout } j\end{aligned}$$

**Théoreme 2.1.** Soit  $\{X_0, X_1, X_2, \dots\}$  une CM et  $P$  sa matrice de transition carré d'ordre  $N$ . Si la distribution initiale,  $\mu_0$  est donnée, alors la distribution de probabilité de  $X_n$  est donnée par  $\mu_0 P^n$ .

$$X_0 \sim \mu_0 \implies X_n \sim \mu_0 P^n$$

### Probabilité d'une trajectoire

Rappelons qu'une trajectoire est une séquence de valeurs pour  $X_0, X_1, \dots, X_n$ . La propriété de Markov nous permet de trouver la probabilité de n'importe quelle trajectoire en multipliant la probabilité de départ et toutes les probabilités à un pas ultérieures.

**Proposition 2.2.** Pour toute suite  $\{x_0, x_1, \dots, x_n\}$  dans  $E$ , on a  $P(X_0 = x_0, X_1 = x_1, X_2 = x_2, \dots, X_n = x_n) = \mu_0(x_0) P(x_0, x_1) P(x_1, x_2) \dots P(x_{n-1}, x_n)$ .

**Théoreme 2.2** (Equations de Chapman-Kolmogorov). Pour tout  $(i, j) \in E^2$  et tout couple  $(m, n)$  d'entiers positifs, on a l'identité :

$$\begin{aligned}\mathbb{P}(X_{m+n} = j | X_0 = i) &= \sum_{k \in E} \mathbb{P}(X_m = k | X_0 = i) \mathbb{P}(X_n = j | X_0 = k) \\ \text{ou encore } p_{ij}^{(m+n)} &= \sum_{k \in E} p_{ik}^{(m)} p_{kj}^{(n)} \text{ ou } (P^{m+n})_{ij} = (P^m)_{ik} (P^n)_{kj}\end{aligned}$$

**Exemple 2.3.** Soit une CM définie par la matrice de transition

$$P = \begin{pmatrix} 0.2 & 0.4 & 0.1 & 0.3 \\ 0 & 0.5 & 0.4 & 0.1 \\ 0.1 & 0 & 0.6 & 0.3 \\ 0.4 & 0.15 & 0.25 & 0.2 \end{pmatrix}$$

et de distribution initiale  $\mu_0 \sim (0.3, 0.2, 0.5, 0)$ .

1. Déterminer  $\mathbb{P}(X_2 = 2)$ .
2. Déterminer la probabilité des trajectoires : 1,2,3,4 et 2,3,1,1.

1.  $\mathbb{P}(X_2 = 2) = (\mu_0 P^2)_2$ .

$$\begin{aligned} \text{Or } \mu_0 P^2 &= (0.3, 0.2, 0.5, 0) \begin{pmatrix} 0.2 & 0.4 & 0.1 & 0.3 \\ 0 & 0.5 & 0.4 & 0.1 \\ 0.1 & 0 & 0.6 & 0.3 \\ 0.4 & 0.15 & 0.25 & 0.2 \end{pmatrix} \begin{pmatrix} 0.2 & 0.4 & 0.1 & 0.3 \\ 0 & 0.5 & 0.4 & 0.1 \\ 0.1 & 0 & 0.6 & 0.3 \\ 0.4 & 0.15 & 0.25 & 0.2 \end{pmatrix} \\ &= (0.167, 0.193, 0.41, 0.23) \end{aligned}$$

D'où  $\mathbb{P}(X_2 = 2) = 0.193$ .

Sous R, le produit matriciel s'effectue à l'aide de la fonction `%*%`. Pour la puissance d'ordre  $n \geq 2$  d'une matrice carré, on peut utiliser la fonction `%^%` définie ci-dessous.

```
P4=matrix(c(0.2,0,0.1,.4,.4,.5,0,.15,.1,.4,.6,.25,.3,.1,.3,.2), nc=4)
mu0=c(.3,.2,.5,0)
mu0%*%P4%*%P4 # %*% : produit matriciel
```

```
      [,1] [,2] [,3] [,4]
[1,] 0.167 0.193 0.41 0.23
```

```
# puissance nième d'une matrice carré
power_matrix=function(A,n){
  if(n==0) return(diag(rep(1,nrow(A))))
  if(n==1) return(A)
  if(n>1) return(A%*%power_matrix(A,(n-1)))
}
"%^%"<-power_matrix
P4%^%3 # P4^3
```

```
      [,1] [,2] [,3] [,4]
[1,] 0.1415 0.259 0.3835 0.216
[2,] 0.1385 0.193 0.4405 0.228
[3,] 0.1925 0.163 0.3885 0.256
[4,] 0.1670 0.244 0.3670 0.222
```

2.  $\mathbb{P}(1, 2, 3, 4) = \mathbb{P}(X_0 = 1) \times P_{12} \times P_{23} \times P_{34} = 0.3 \times 0.4 \times 0.4 \times 0.3 = 0.0144$ .

$$\mathbb{P}(2, 3, 1, 1) = \mathbb{P}(X_0 = 2) \times P_{23} \times P_{31} \times P_{11} = 0.2 \times 0.4 \times 0.1 \times 0.2 = 0.0016.$$

Pour le calcul de la probabilité d'une trajectoire donnée sous R, on peut utiliser la fonction suivante :

```

ProbTraj<-function(traj,mu0,X0, P){
  n<-length(traj)-1
  pi<-NULL
  for(k in 1:n) pi[k]<-P[traj[k],traj[k+1]]
  mu0[X0]*prod(pi)
}
ProbTraj(c(1,2,3,4),mu0,X0=1,P4)

```

```
[1] 0.0144
```

```
ProbTraj(c(2,3,1,1),mu0,X0=2,P4)
```

```
[1] 0.0016
```

**Exercice 2.2.** On considère une CM à trois états dont la matrice de transition est la suivante :

$$P = \begin{pmatrix} 0.6 & 0.2 & 0.2 \\ 0.4 & 0 & 0.6 \\ 0 & 0.8 & 0.2 \end{pmatrix}$$

Calculer  $\mathbb{P}(X_2 = 3 | X_0 = 1)$ .

### Solution 1.2

$\mathbb{P}(X_2 = 3 | X_0 = 1)$  est la probabilité de se déplacer de l'état 1 à l'état 2 en **deux** étapes. Donc les trajectoires possibles sont : (1,2,3); (1,1,3) ou (1,3,3) avec  $\mu_0 = (1, 0, 0)$ .

$$\begin{aligned} \mathbb{P}(X_2 = 3 | X_0 = 1) &= (P^2)_{13} = \begin{pmatrix} 0.6 & 0.2 & 0.2 \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \end{pmatrix} \begin{pmatrix} \cdot & \cdot & 0.2 \\ \cdot & \cdot & 0.4 \\ \cdot & \cdot & 0.2 \end{pmatrix} \\ &= 0.6 \times 0.2 + 0.2 \times 0.6 + 0.2 \times 0.2 = 0.28. \end{aligned}$$

```

P=matrix(c(0.6,0.4,0,0.2,0,0.8,0.2,0.6,0.2),nc=3)
P[1,]%*%P[,3]

```

```

[,1]
[1,] 0.28

```

```

# ou encore
# (1,2,3); (1,1,3); (1,3,3)
ProbTraj(c(1,2,3),c(1,0,0),X0=1,P)+ProbTraj(c(1,1,3),c(1,0,0),X0=1,P)+
  ProbTraj(c(1,3,3),c(1,0,0),X0=1,P)

```

```
[1] 0.28
```

**Exercice 2.3.** On considère une chaîne de Markov à deux états dont la matrice de transition est la suivante :

$$P = \begin{pmatrix} 0.4 & 0.6 \\ 0.6 & 0.4 \end{pmatrix}$$

Calculer la probabilité de l'événement  $A_3$  de passer pour  $n \leq 5$  trois fois par l'état 2 si la loi initiale est  $\mu_0 = (0.3, 0.7)$ .

**Solution 1.3**

$A_3 = \{n \leq 5, \text{ on passe 3 fois par l'état 2} \}$ .

Tout d'abord, déterminons les différentes possibilités.

Si l'état initial est l'état 1, on aura les possibilités suivantes :

$$x_1^1 = (1, 1, 2, 2, 2), x_1^2 = (1, 2, 1, 2, 2), x_1^3 = (1, 2, 2, 1, 2) \text{ et } x_1^4 = (1, 2, 2, 2, 1)$$

$$\text{D'où } \mathbb{P}_1(A_3) = \mathbb{P}(X = x_1^1) + \mathbb{P}(X = x_1^2) + \mathbb{P}(X = x_1^3) + \mathbb{P}(X = x_1^4)$$

$$\text{Avec } \mathbb{P}(X = x_1^1) = \mathbb{P}(X_0 = 1) \times p_{11} \times p_{12} \times p_{22} \times p_{22} = 0.3 \times 0.4 \times 0.6 \times 0.4 \times 0.4 = 0.01152;$$

$$\mathbb{P}(X = x_1^2) = \mathbb{P}(X_0 = 1) \times p_{12} \times p_{21} \times p_{12} \times p_{22} = 0.3 \times 0.6^3 \times 0.4 = 0.02592;$$

$$\mathbb{P}(X = x_1^3) = \mathbb{P}(X_0 = 1) \times p_{12} \times p_{22} \times p_{21} \times p_{12} = 0.3 \times 0.6 \times 0.4 \times 0.6^2 = 0.02592;$$

$$\mathbb{P}(X = x_1^4) = \mathbb{P}(X_0 = 1) \times p_{12} \times p_{22} \times p_{22} \times p_{21} = 0.3 \times 0.6^2 \times 0.4^2 = 0.01728;$$

$$\text{Donc } \mathbb{P}_1(A_3) = 0.01152 + 0.02592 + 0.02592 + 0.01728 = 0.08064.$$

Si l'état initial est l'état 2, on aura :

$$x_2^1 = (2, 1, 1, 2, 2), x_2^2 = (2, 1, 2, 1, 2), x_2^3 = (2, 1, 2, 2, 1), x_2^4 = (2, 2, 1, 2, 1), x_2^5 = (2, 2, 1, 1, 2) \text{ et } x_2^6 = (2, 2, 2, 1, 1).$$

$$\text{D'où } \mathbb{P}_2(A_3) = \mathbb{P}(X = x_2^1) + \mathbb{P}(X = x_2^2) + \mathbb{P}(X = x_2^3) + \mathbb{P}(X = x_2^4) + \mathbb{P}(X = x_2^5) + \mathbb{P}(X = x_2^6).$$

$$\mathbb{P}_2(A_3) = 0.3192.$$

$$\text{Ainsi } \mathbb{P}(A_3) = \mu_0(1) \times \mathbb{P}_1(A_3) + \mu_0(2) \times \mathbb{P}_2(A_3) = 0.247632.$$

```
tab1=matrix(c(1,1,2,2,2,1,2,1,2,2,1,2,2,1,2,1,2,2,1,1),nr=5)
mu0=c(0.3,0.7)
P2=matrix(c(0.4,0.6,0.6,0.4),nc=2)
(probs1=apply(tab1,2,ProbTraj,mu0,1,P2))
```

```
[1] 0.01152 0.02592 0.02592 0.01728
```

```
sum(probs1)      # P1(A3)
```

```
[1] 0.08064
```

```
tab2=matrix(c(2,1,1,2,2,2,1,2,1,2,2,1,2,2,1,2,2,1,1,2,2,2,1,1),nr=5)
(probs2=apply(tab2,2,ProbTraj,mu0,2,P2))
```

```
[1] 0.04032 0.09072 0.06048 0.06048 0.04032 0.02688
```

```
sum(probs2)      # P2(A3)
```

```
[1] 0.3192
```

```
PA3=mu0%*%c(sum(probs1),sum(probs2)) ; as.vector(PA3)      #P(A3)
```

```
[1] 0.247632
```

## 2.3 Classification des états

L'espace des états d'une CM peut être partitionné en un ensemble de classes communicantes disjointes.

**Définition 2.4.** On dit que l'état  $j$  est **accessible** à partir de l'état  $i$ , s'il existe un entier  $n \geq 0$  tel que  $P_{ij}^{(n)} > 0$ . On note  $i \rightarrow j$ .

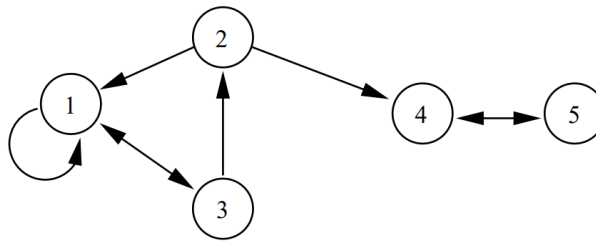
Sur le graphe, si  $i \neq j$ ,  $i \rightarrow j$  s'il existe un chemin (orienté) du sommet  $i$  vers le sommet  $j$ .

Si  $j$  est accessible à partir de  $i$  et  $i$  est accessible depuis  $j$ , on dit que  $i$  et  $j$  **communiquent** et on note  $i \leftrightarrow j$ .

On note que si  $i \rightarrow k$  et  $k \rightarrow j$ , alors  $i \rightarrow j$  (la relation d'accessibilité est transitive).

**Définition 2.5.** Les états  $i$  et  $j$  sont dans la même classe communicante si  $i \leftrightarrow j$ .

**Exemple 2.4.** Trouvez les classes communicantes associées au diagramme de transition suivant.



D'après le diagramme, on remarque que  $1 \leftrightarrow 2$ ,  $2 \leftrightarrow 3$  et  $4 \leftrightarrow 5$ . Donc on a deux classes communicantes  $\{1, 2, 3\}$  et  $\{4, 5\}$ .

**Définition 2.6.** Une classe d'états communicante est fermée s'il est impossible de quitter cette classe. Autrement dit, la classe  $C$  communicante est fermée si  $p_{ij} = 0$  pour tout  $i \in C$  et  $j \notin C$ .

Dans l'exemple précédent :

- La classe  $\{1, 2, 3\}$  n'est pas fermée car on peut la quitter ( $2 \rightarrow 4$ ).
- La classe  $\{4, 5\}$  est fermée car on ne peut pas la quitter.

**Définition 2.7.** Si la chaîne de Markov ne possède qu'une unique classe, c'est à dire que tous ses éléments communiquent, la chaîne sera dite **irréductible**.

Un état  $i$  est dite **absorbant** si  $\{i\}$  est une classe fermée.

**Exemple 2.5.** On considère la CMD à trois états 1, 2, 3 et de matrice stochastique

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

Comme  $1 \leftrightarrow 2$  et  $2 \leftrightarrow 3$ , alors cette CM est **irréductible**.

Pour étudier la classification d'une CMD sous R, on peut procéder comme suit :

```
# install.packages("markovchain") # Installation de l'extension markovchain
library(markovchain)
pp=matrix(c(0.5,0.5,0,0.5,0.25,1/3,0,0.25,2/3), nc=3,
          dimnames = list(c("1","2","3"),c("1","2","3")))
cm1=new("markovchain", transitionMatrix=pp)
is.irreducible(cm1) # tester si la CM est irreductible
```

```
[1] TRUE
```

**Exemple 2.6.** Soit la CMD à quatre états définie par la matrice stochastique

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

Cette CM possède trois classes :  $\{1, 2\}$ ,  $\{3\}$  et  $\{4\}$ , donc elle est **non irréductible**.

L'état 4 est **absorbant** car  $\{4\}$  est une classe fermée.

```
p15=matrix(c(0.5,0.5,0,0,0.5,0.5,0,0,0,0,0.25,0,0,0,0.75,1), nc=4,
           dimnames = list(c("1","2","3","4"),c("1","2","3","4")))
cm2=new("markovchain", transitionMatrix= p15)
is.irreducible(cm2) # tester si la CM est irréductible
```

```
[1] FALSE
```

```
absorbingStates(cm2) # déterminer les états absorbants
```

```
[1] "4"
```

## 2.4 Temps d'absorption

Dans cette section on se pose la question suivante : Etant donné une chaîne  $X$  et  $x$  dans l'espace d'états  $E$ , quel est le temps moyen (éventuellement infini) que met  $X$  à arriver au temps  $x$ .

### 2.4.1 Temps d'arrêt

Pour  $x \in E$  on définit le temps aléatoire

$$T_x = \min\{n \geq 0 : X_n = x\},$$

premier moment où la chaîne atteint l'état  $x$ .

**Définition 2.8.** Soit  $A$  un sous ensemble de l'espace d'états  $E$ . Le **temps d'atteinte** de  $A$  est la variable aléatoire  $T_A$  définie par :

$$T_A = \min\{n \geq 0 : X_n \in A\},$$

Notons que la probabilité d'atteindre  $A$  depuis  $i$  est  $h_{A|i} = P(T_A < \infty | X_0 = i)$ .

$T_A$  est le temps nécessaire pour atteindre l'ensemble  $A$  pour la première fois.  $T_A$  peut prendre les valeurs  $0, 1, 2, \dots, \infty$ .

Si la chaîne ne peut jamais atteindre  $A$ , alors  $T_A = \infty$ .

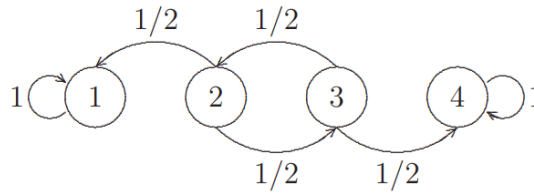
Le temps d'atteinte est appelé aussi **temps d'arrêt**. Si  $A$  est une classe fermée,  $T_A$  sera appelé le **temps d'absorption** et  $h_{A|i}$  la **probabilité d'absorption**.

**Définition 2.9.** Le **temps moyen d'atteinte** de  $A$ , à partir de  $i$  est :

$$m_{iA} = \mathbb{E}(T_A | X_0 = i)$$

**Exemple 2.7.** On considère la chaîne suivante





- 1) Déterminer le vecteur des probabilités d'atteinte de l'état 4.
- 2) Déterminer le temps moyen d'atteinte de  $A = \{1, 4\}$ , à partir de 2.

## 2.5 Simulation d'une chaîne de Markov discrète

### 2.5.1 Exemple illustratif

L'idée générale pour simuler une chaînes de Markov discrète peut être illustrée par un exemple simple à 2 états. Supposons que notre espace d'états est  $\{1, 2\}$  et que la matrice de transition est :

$$P = \begin{pmatrix} 0.2 & 0.8 \\ 0.4 & 0.6 \end{pmatrix}$$

Supposons maintenant que notre chaîne de Markov commence à l'état 1 de sorte que  $X_0 = 1$ . Puisque nous commençons à l'état 1, nos probabilités de transition sont définies par la première ligne de  $P$ . Notre chaîne peut soit rester à l'état 1 avec une probabilité  $P_{11} = 0.2$ , soit passer à l'état 2 avec une probabilité  $P_{12} = 0.8$ . Par conséquent, pour simuler  $X_1$ , il faut générer une variable aléatoire selon les probabilités  $P_{11} = P(X_1 = 1 | X_0 = 1) = 0.2$  et  $P_{12} = P(X_1 = 2 | X_0 = 1) = 0.8$ .

En général, nous pouvons générer n'importe quelle variable aléatoire discrète selon un ensemble de probabilités  $p = \{p_1, \dots, p_K\}$  avec la méthode de transformation inverse. Notez également que cela équivaut à prendre un seul tirage à partir d'une distribution multinomiale avec un vecteur de probabilité  $p$  - nous utilisons cette méthode dans l'algorithme ci-dessous.

#### Algorithme

1. Obtenir la matrice de transition de probabilité  $S \times S$ ,  $P$ .
2. Définir  $t = 0$ .
3. Choisissez un état initial  $X_0 = x_0$ .
4. pour  $t = 1, \dots, T$  :
  - a) Obtenir la ligne de  $P$  correspondante à l'état actuel  $X_t$
  - b) Génère  $X_{t+1}$  à partir d'une distribution multinomiale avec un vecteur de probabilité égal à la ligne que nous avons obtenue ci-dessus.

Nous implémentons cet algorithme dans la fonction suivante :

```

# simuler des CMD selon la matrice de transition P
sim_CMD <- function( P, N = 50, x0=1 ) {
  # nombre des états
  n_etats <- nrow(P)
  etats <- numeric(N)
  # valeur de l'état initial
  etats[1] <- x0
  for(t in 2:N) {
    # vecteur de prob pour simuler l'état suivant X_{t+1}
    p <- P[etats[t-1], ]
    ## tirer depuis la distribution multinomiale et déterminer l'état
    etats[t] <- which(rmultinom(1, 1, p) == 1)
  }
}

```

```

}
return(etats)
}

```

### 2.5.1.1 Simulation de l'exemple illustratif

```
(P <- matrix(c(0.2,0.4,0.8,0.6), nc=2))
```

```

      [,1] [,2]
[1,]  0.2  0.8
[2,]  0.4  0.6

```

```

sim1<-sim_CMD(P,N=100, x0=1)
head(sim1)

```

```
[1] 1 2 1 2 2 1
```

```

sim11<-replicate(sim_CMD(P,N=50, x0=1),n=100)
mean(sim11[2,sim11[1,]==1]==1) #  $p_{11}=p(X_1=1|X_0=1)=0.2$ 

```

```
[1] 0.16
```

```

mean(sim11[1,sim11[1,]==1]==1) #  $p_{12}=p(X_1=1|X_0=1)=0.2$ 

```

```
[1] 1
```

# PROCESSUS DE POISSON

---



# PROCESSUS GAUSSIEN

---



# APPLICATIONS

---

Some *significant* applications are demonstrated in this chapter.

## 5.1 Example one

## 5.2 Example two





---

# BIBLIOGRAPHIE

- Dobrow, Robert P. 2016. *Introduction to Stochastic Processes with R*. Hoboken, New Jersey : John Wiley & Sons, Inc.
- Fewster, Rachel. 2014. *Stochastic Processes*. Department of Statistics, University of Auckland,NZ. <https://www.stat.auckland.ac.nz/~fewster/325/notes.php>.
- Foata, Dominique, and Aimé Fuchs. 2002. *Processus Stochastiques*. 2ième ed. Sciences Sup. Dunod.
- Lessard, Sabin. 2014. *Processus Stochastiques : Cours et Exercices Corrigés*. ellipses.
- Pierre, Brémaud. 2009. *Initiation Aux Probabilités et Aux Chaînes de Markov*. 2e édition entièrement révisée. Berlin Heidelberg : Springer.
- R Core Team. 2020. *R : A Language and Environment for Statistical Computing*. Vienna, Austria : R Foundation for Statistical Computing. <https://www.R-project.org/>.
- Ross, Sheldon M. 2019. *Introduction to Probability Models*. Twelfth. San Diego, CA, USA : Academic Press.
- Ross, S. M. 1996. *Stochastic Processes*. Wiley Series in Probability and Statistics : Probability and Statistics. Wiley.
- Tijms, H. C. 2003. *A First Course in Stochastic Models*. Wiley. <https://books.google.tn/books?id=WibF8iVHaiMC>.