Notes de cours 04

Chaines deMARKOV : une initiation

Table des matières

1 Probabilités conditionnelles

On se place dans le cadre d'un espace probabilisé $(\Omega, \mathscr{T}, \mathbb{P})$. On rappelle que $\mathscr{T}$ est l'ensemble (on dit la « tribu ») des événements. On rappelle que l'ensemble des événements contient l'événements certain, l'événement impossible et est stable par conjonction, disjonction et négation (ce qui, en tant que parties de Ω dit que $\mathscr{T}$ contient Ω, $\emptyset$ et est stable par intersection dénombrable, réunion dénombrable et passage au complémentaire).

L'application $\mathbb{P} : \mathscr{T} \to [0,1] \subset \mathbb{R}$ est une (mesure de) probabilité. Elle vérifie par définition les deux prpriétés suivantes :

(i) $\mathbb{P}(\Omega) = 0$, $\mathbb{P}(\emptyset) = 0$ et

(ii) pour toute famille dénombrable d'événements $(A_n)_{n\in\mathbb{N}}$ deux à deux incompatibles,

$$\mathbb{P}\left(\bigcup_{n=0}^{+\infty} A_n\right) = \sum_{n=0}^{+\infty} \mathbb{P}(A_n)$$

Une variable aléatoire X, à valeurs dans un ensemble $\mathbb{X}$ fini ou numérique (par exemple $\mathbb{N}$, $\mathbb{R}$ ou $\mathbb{C}$) est une fonction $X : \Omega \to \mathbb{X}$.

Les événements– pour l'essentiel– se décrivent à l'aide de propositions portant sur des variables aléatoires. Par exemple : $\{X = 2\}$ ou $\{X \in [0,1]$ et $Y \leq -1\}$. On peut évaluer la probabilité de tels événements.

1.1 Définition

Evaluer la probabilité conditionnelle d'un événement A relativement à un événement B, c'est évaluer la proportion de configurations satisfaisant à la fois les événements A et B parmi celles satisfaisant l'événement B. En formule, si $\mathbb{P}(B) > 0$,

$$\underbrace{\mathbb{P}(A|B)}_{\text{probabilité de } A \textit{ sachant } B} := \frac{\mathbb{P}(A \text{ et } B)}{\mathbb{P}(B)}$$

Une remarque de notation importante : « A sachant B » ou « $(A|B)$ » **n'est pas** un événément, ça n'est pas défini. La notation $\mathbb{P}(A|B)$ ne **signifie pas** que l'on applique la probabilité $\mathbb{P}$ à $A|B$, qui n'a pas de sens ! Pour être plus clair et syntaxiquement correct, on devrait toujours dire « la probabilité, sachant B, de A ».

Proposition-Définition 1. *Si $(\Omega, \mathscr{T}, \mathbb{P})$ est un espace probabilisé et $B \in \mathscr{T}$ vérifie $\mathbb{P}(B) > 0$ alors l'application $\mathbb{P}_B$ définie par*

$$\forall A \in \mathscr{T}, \mathbb{P}_B(A) = \mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$$

est une probabilité et $(\Omega, \mathscr{T}, \mathbb{P}_B)$ est un espace probabilisé.

Cette mesure de probabilité (notée aussi $\mathbb{P}(.|B)$), s'appelle la probabilité conditionnelle sachant l'événement *B.*

Important ! ! ! Il est crucial de ne pas confondre $\mathbb{P}$, la probabilité « première » et les probabilités conditionnelles du type $\mathbb{P}(.|B)$ qui s'en déduisent. Quand on calcule une probabilité d'événement de type $\mathbb{P}(A)$, c'est « avant » de disposer de la moindre information, les probabilités conditionnelles sont par contre calculées alors que l'on dispose d'informations supplémentaires [1].

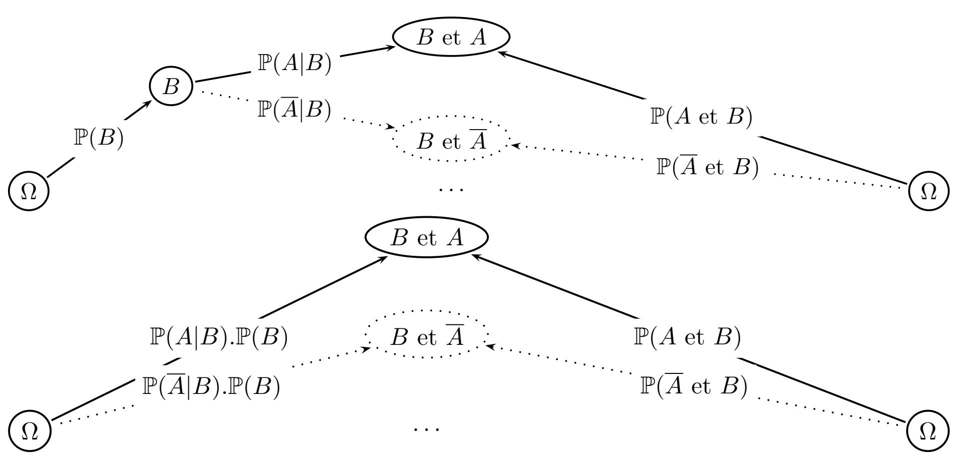


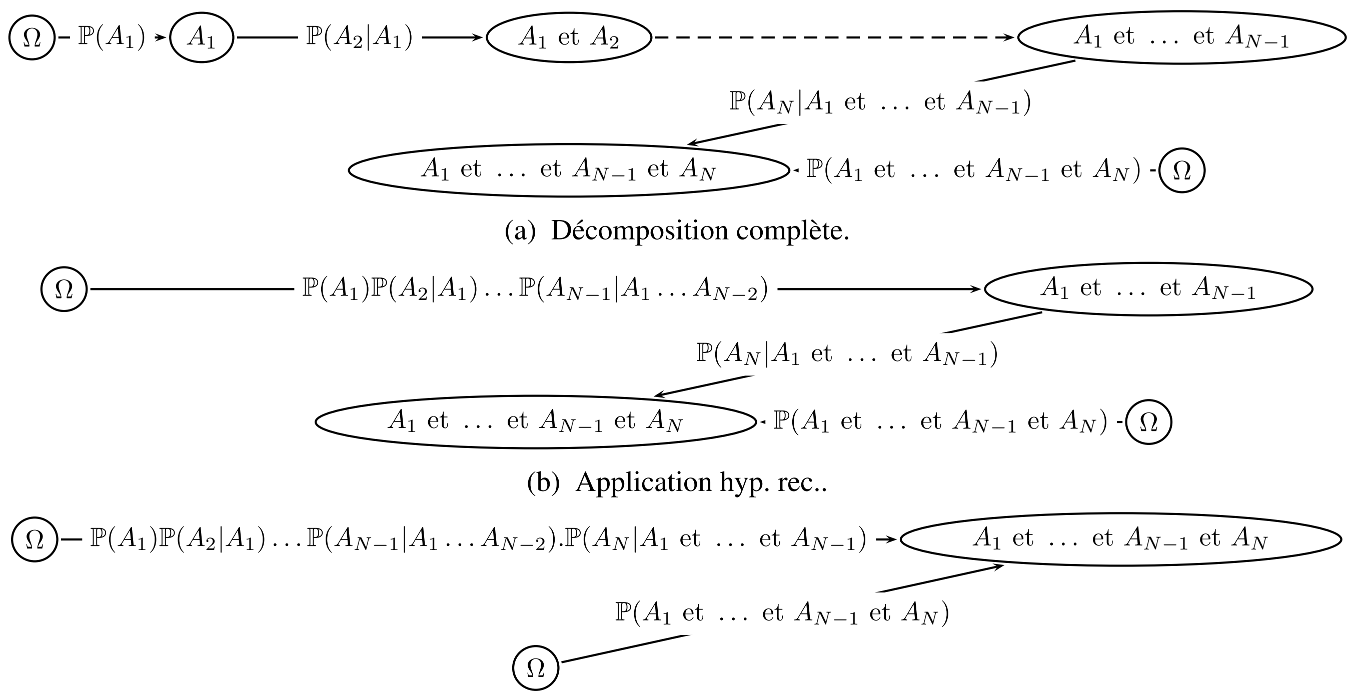
FIGURE 1 – Définition de la probabilité conditionnelle illustrée par deux arbres de choix équivalents. Chaque noeud est un événement. On indique sur la flèche la probabilité conditionnelle du but sachant la source.

1. On peut réfléchir au problème suivant : Comparer la « probabilité que je gagne au loto ce samedi sachant que j'ai joué une seule grille simple » avec la « probabilité que je gagne au loto ce samedi sachant que j'ai joué une seule grille simple et que mon petit doigt (qui ne se trompe jamais), m'a dit que la somme des nombres tirés à ce tirage est inférieure ou égale à 61 ».

1.2 La formule des probabilités composées

Si $A_1,\dots,A_N$ sont $N \geq 2$ événements et $\mathbb{P}(A_1 \text{ et } \dots \text{ et } A_{N_1}) \neq 0$ alors

$$\mathbb{P}(A_1 \text{ et } \dots A_N) = \mathbb{P}(\bigcap_{n=1}^{N} A_n) = \mathbb{P}(A_1)\mathbb{P}(A_2|A_1)\dots\mathbb{P}(A_N|A_1 \text{ et } \dots \text{ et } A_{N-1})$$



(a) Décomposition complète.

(b) Application hyp. rec..

(c) Application déf. proba. conditionnelle.

FIGURE 2 – La démonstration par récurrence de la formule des probabilités composées.

Exercice 1.— Dans une population de $2.N$ mouches comportant autant de femelles que de mâles, on extrait les individus les uns après les autres jusqu'à obtention d'un couple. On appelle T le nombre de tirages effectués. Donner la loi de T.

Indication:Poser $X_n = 1$ si une femelle est tirée au tirage n, $X_n = 0$ sinon, conditionner suivant la valeur de X_1 et utiliser la formule des proba. composées.

1.3 Un exemple simple, le BAYES du pauvre

Pour illustrer la signification de ces probabilités conditionnelles, supposons que nous disposions d'une population de 36 individus, composée de $\frac{1}{3}$ de garçons et de $\frac{2}{3}$ de filles. Parmi les garçons, 33% portent des lunettes alors que c'est le cas pour seulement 25% des filles.

Une question simple est la proportion d'individus portant des lunettes.

On peut régler cette question de deux façons.

La première, naïve, consiste à dénombrer les individus :

Il y a 12 garçons dont 4 portent des lunettes, il y a 24 filles dont 6 portent des lunettes. Il y a donc, sur les 36 individus, 10 individus porteurs de lunettes, ce qui fait une proportion de $\frac{10}{36} = 27,8\%$ de porteurs de lunettes dans cette population.

On peut faire ce raisonnement sans connaitre le nombre d'individus. Si N est le nombre total d'individus,

Il y a $\frac{1}{3}.N$ garçons et $\frac{1}{3}.\frac{1}{3}.N$ garçons porteurs de lunettes, il y a $\frac{2}{3}.N$ filles et $\frac{1}{4}.\frac{2}{3}.N$ filles porteuses de lunettes, ce qui fait un total de $(\frac{1}{9}+\frac{1}{6}).N = \frac{5}{18}.N$ de porteurs de lunettes et donc une proportion de $27,8\%$ de la population.

La seconde (en apparence plus savante, mais on fait la même chose ! !). On considère l'expérience aléatoire de tirer au sort (uniformément) un individu I et on considère les événements

$A =$« I porte des lunettes », $B =$« I est une fille » et $C =$« I est un garçon ».

La question est la valeur de $\mathbb{P}(A)$.

L'utilisation d'une phrase entre « » pour désigner un événement peut mener à des ambiguités et est une mauvaise pratique. D'un point de vue de la méthode, il est préférable d'introduire immédiatement des variables aléatoires, p.ex. des *indicatrices*, permettant l'écriture usuelle de ces événements. Soit S et M définies par

$$S = \begin{cases} 0 & \text{si } I \text{ est un garçon} \\ 1 & \text{si } I \text{ est une fille} \end{cases} \text{ et } M = \begin{cases} 0 & \text{si } I \text{ sans lunettes} \\ 1 & \text{si } I \text{ à lunettes} \end{cases}$$

On a alors

$$A = \{M = 1\}, B = \{S = 1\} \text{ et } C = \{S = 0\}$$

et les données du problème s'expriment par les faits suivants

$$\mathbb{P}(B) = \frac{2}{3}, \mathbb{P}(C) = \frac{1}{3}, \mathbb{P}(A|B) = \frac{1}{4} \text{ et } \mathbb{P}(A|C) = \frac{1}{3}$$

On a donc

$$\mathbb{P}(A \text{ et } B) = \mathbb{P}(A|B).\mathbb{P}(B) = \frac{1}{4}.\frac{2}{3}, \mathbb{P}(A \text{ et } C) = \mathbb{P}(A|C).\mathbb{P}(C) = \frac{1}{3}.\frac{1}{3}$$

et, sachant que $A = (A \text{ et } B)$ ou $(A \text{ et } C)$, avec une alternative exclusive, on a

$$\mathbb{P}(A) = \mathbb{P}(A \text{ et } B) + \mathbb{P}(A \text{ et } C) = \frac{5}{18}$$

Ces méthodes donnent évidemment le même résultat, la première a l'avantage de la naïveté et de la simplicité conceptuelle, la seconde ouvre la porte à une possibilité de généralisation dans le cas d'une population infinie.

On peut illustrer le calcul fait par une suite de schémas, *cf.* Fig. 3, à interpréter proprement.

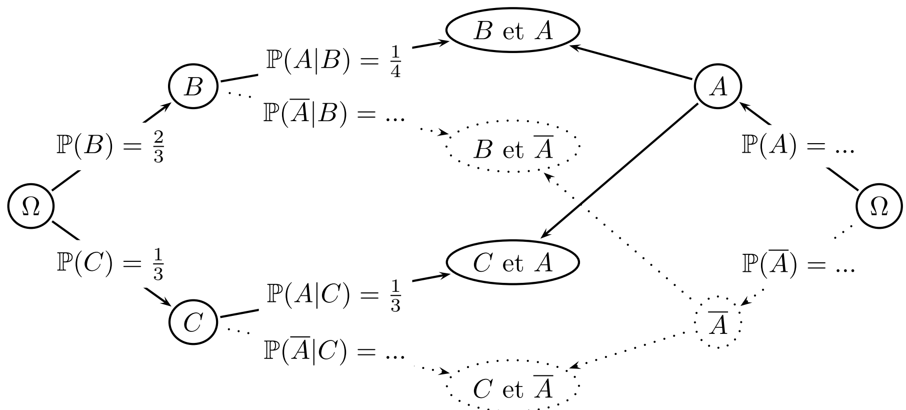
Exercice 2.— Un laboratoire a mis au point un test pour déceler des souris malades. Des essais prouvent que :

— 96 fois sur 100, le test donne un résultat positif quand la souris est effectivement malade.

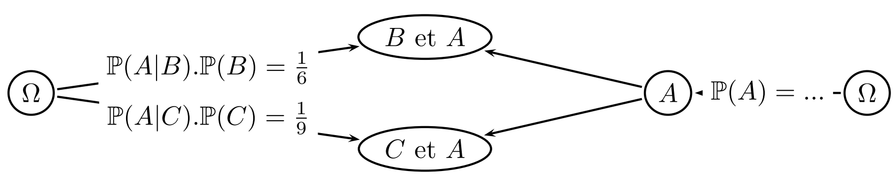
— 94 fois sur 100, le test donne un résultat négatif quand la souris n'est pas malade.

1. Dans une population de souris comprenant 3% de malades, on pratique le test sur une souris choisie au hasard et on constate que le test donne un résultat positif. Quelle est la probabilité que la souris soit malade ? Donner une formule, un ordre de grandeur numérique et commenter le résultat obtenu.

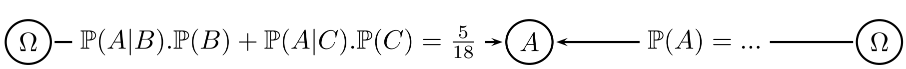
2. La formule de BAYES est la version abstraite de ces calculs. Enoncer cette formule et la redémontrer.



(a) Décomposition complète.



(b) On ne regarde que les branches menant à A, utilisation de la formule d'une probabilité conditionnelle.



(c) Utilisation de l'additivité

FIGURE 3 – Arbre de choix de l'exemple et formule des probabilités totales. Chaque noeud est un événement, que l'on décompose (flèche) en union disjointe d'événements. On indique sur la flèche la probabilité conditionnelle du but sachant la source.

1.4 La formule des probabilités totales

Dans l'exemple précédent, on a illustré la formule fondamentale des probabilités totales : Si $B_1, \ldots B_N$ est un système complet d'événements incompatibles (de probabilité > 0), on a, pour tout évenement A,

$$\mathbb{P}(A) = \sum_{n=1}^{N} \mathbb{P}(A|B_n).\mathbb{P}(B_n)$$

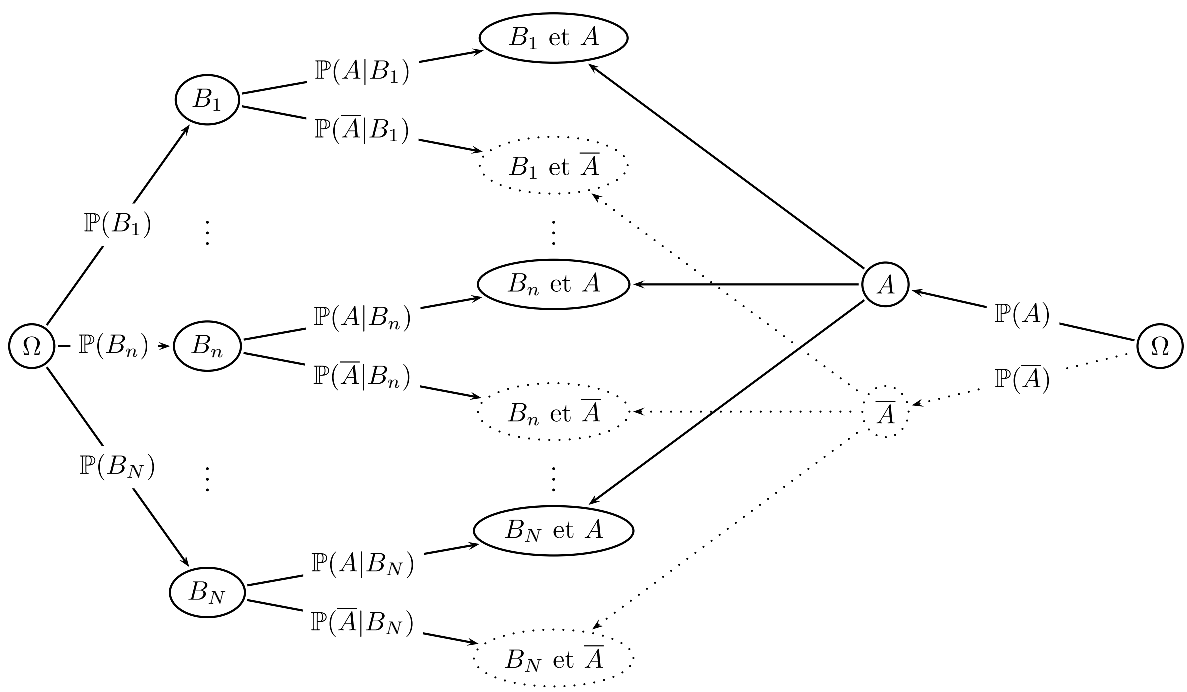


FIGURE 4 – Arbre de choix et formule des probabilités totales. Chaque noeud source est un événement, que l'on décompose (flèches) en union disjointe d'événements. On indique sur la flèche la probabilité conditionnelle du but sachant la source.

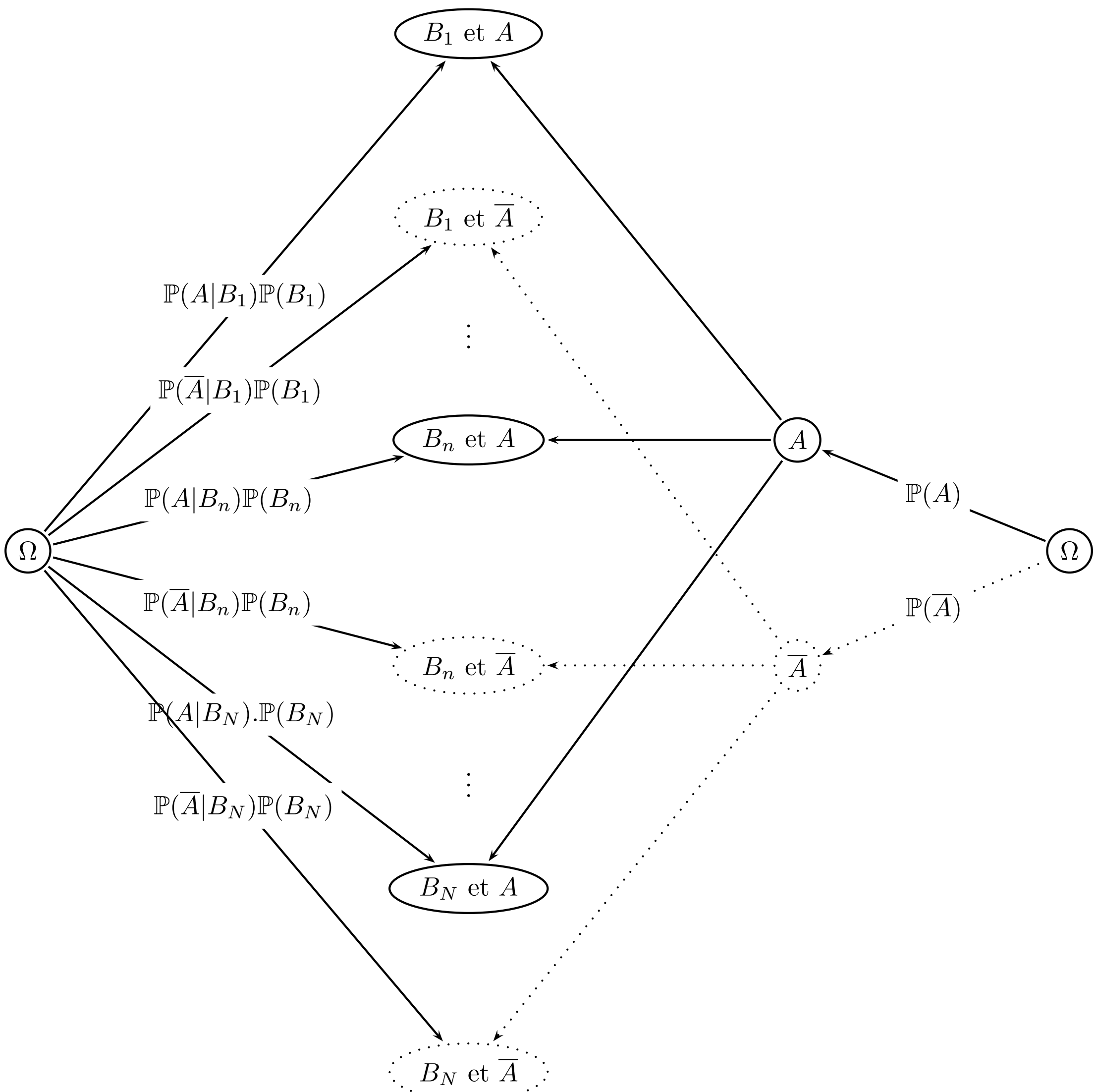


FIGURE 5 – Simplification en utilisant la déf. de proba. conditionnelle.

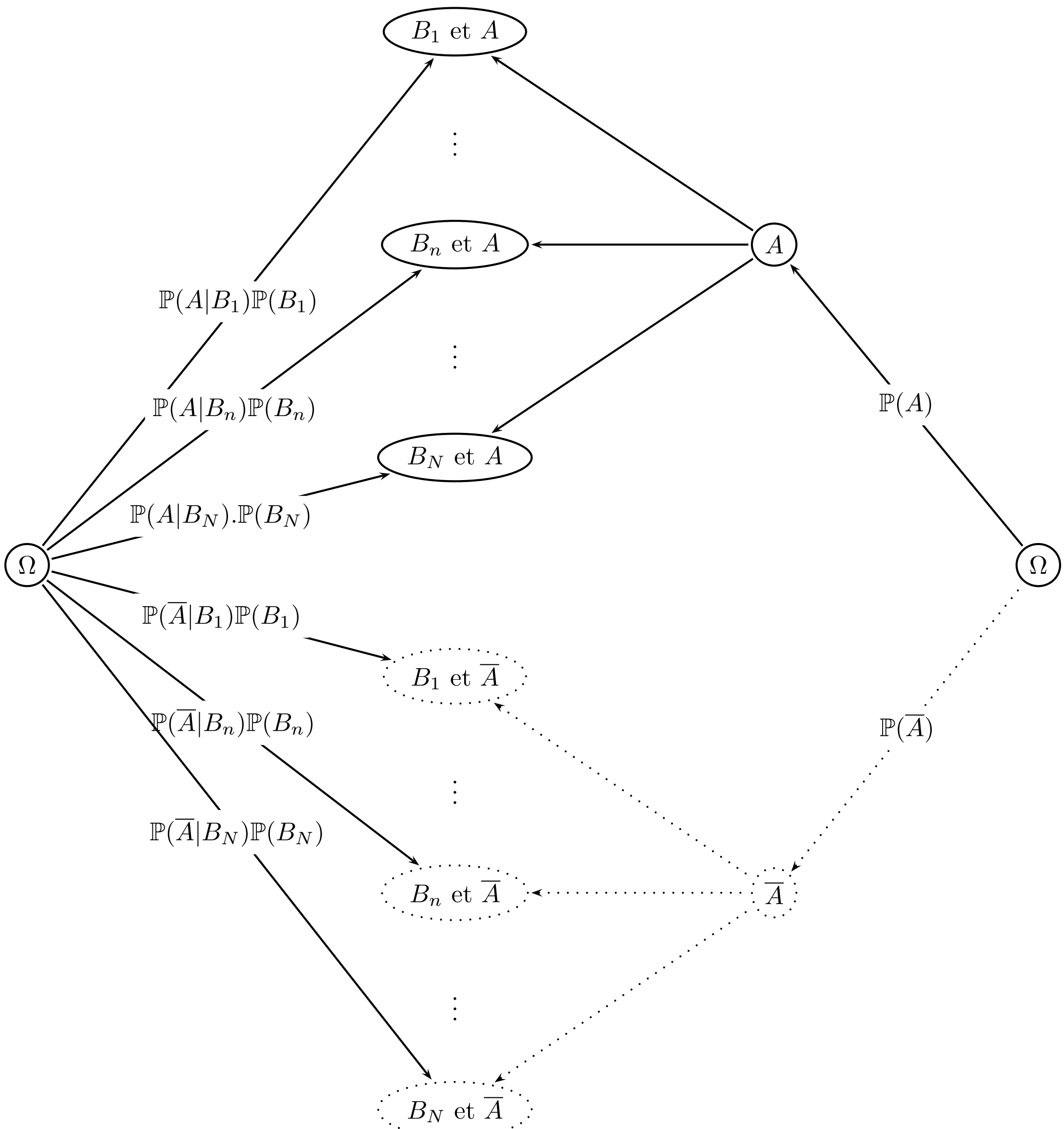


FIGURE 6 – Réordonnancement.

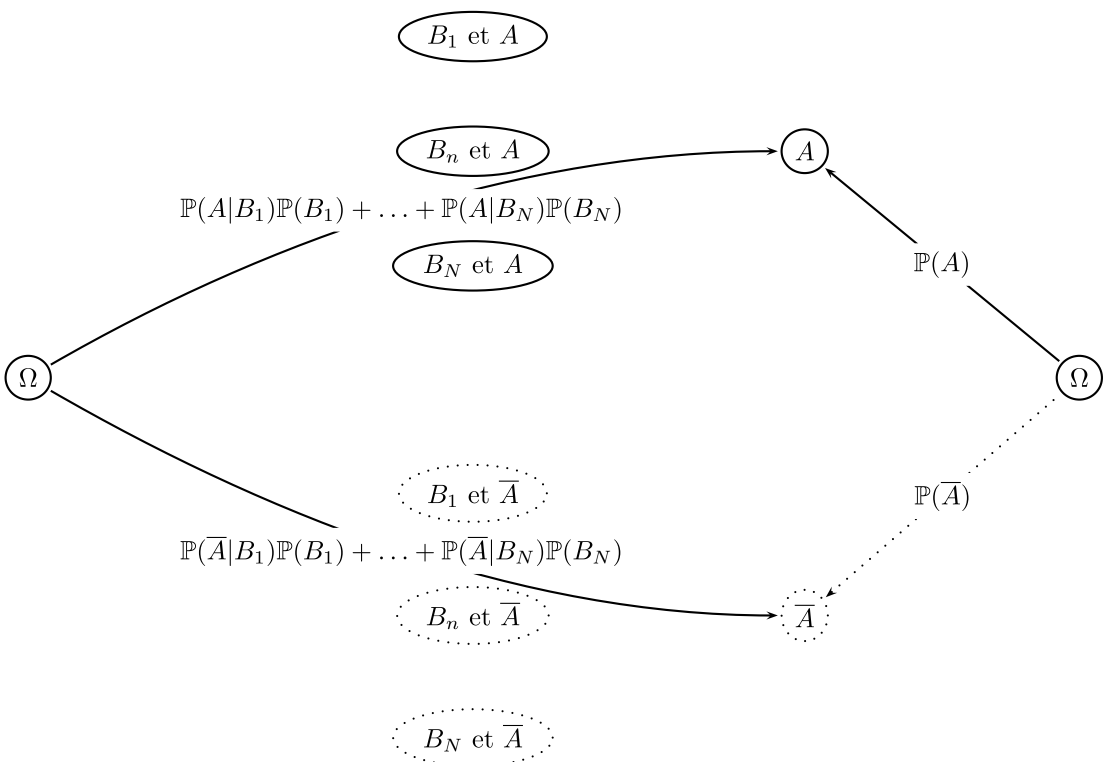


FIGURE 7 – Simplification en utilisant l'axiome d'additivité.

2 Chaînes de MARKOV

Il s'agit d'une méthode de modélisation pour des phénomènes évoluant stochastiquement dans le temps. On considère le temps discret. Chaque étape dans l'évolution se déduit de la précédente en faisant intervenir le hasard.

Le système que l'on considère peut être dans un certain nombre (fini) d'états. Par exemple, supposons que (pour une raison que j'ignore) un champignon puisse prendre trois couleurs : Rouge (R), Vert(V) ou Bleu(B).

On constate que ce champignon, s'il est d'un certaine couleur à un instant donné, il passe à une autre couleur, au top chrono suivant, avec certaines probabilités

Pour fixer les idées

1. S'il est R, il devient R avec probabilité 0.5, V avec proba 0.3 et B avec proba 0.2,
2. S'il est V, il devient R avec probabilité 0.2, V avec proba 0.4 et B avec proba 0.4,
3. S'il est B, il devient R avec probabilité 0.3, V avec proba 0.3 et B avec proba 0.4,

La question que l'on se pose est sachant qu'au départ $n = 0$ le champignon est R, quelle est la proba qu'à l'instant $n = N(= 10)$, il soit B ?

Modélisation

Si on suit notre méthode de modélisation, on voit que les variables qui importent sont $X_0, X_1, \ldots, X_n, \ldots, X_N$, X_n donnant la couleur du champignon à l'instant n. On sait que $X_0 = R$ et ce qu'on cherche c'est

$$\mathbb{P}(X_N = R), \mathbb{P}(X_N = V) \text{ et } \mathbb{P}(X_N = B)$$

Ce que donne l'énoncé, c'est la *loi conditionnelle* de X_{n+1} sachant X_n. On a

1. (R) : $\mathbb{P}(X_{n+1} = R|X_n = R) = 0.5$, $\mathbb{P}(X_{n+1} = V|X_n = R) = 0.3$ et $\mathbb{P}(X_{n+1} = B|X_n = R) = 0.2$,
2. (V) : $\mathbb{P}(X_{n+1} = R|X_n = V) = 0.2$, $\mathbb{P}(X_{n+1} = V|X_n = V) = 0.4$ et $\mathbb{P}(X_{n+1} = B|X_n = V) = 0.4$,
3. (B) : $\mathbb{P}(X_{n+1} = R|X_n = B) = 0.3$, $\mathbb{P}(X_{n+1} = V|X_n = B) = 0.3$ et $\mathbb{P}(X_{n+1} = B|X_n = B) = 0.4$.

Pour calculer la loi de X_{n+1}, appliquons la formule des probabilités totales, les définitions des probabilités conditionnelles en « décomposant les événements suivant les valeurs de X_n »,

$$\begin{aligned}
\mathbb{P}(X_{n+1} = R) &= \mathbb{P}(X_{n+1} = R|X_n = R)\mathbb{P}(X_n = R) + \mathbb{P}(X_{n+1} = R|X_n = V)\mathbb{P}(X_n = V) + \mathbb{P}(X_{n+1} = R|X_n = B)\mathbb{P}(X_n = B) \\
\mathbb{P}(X_{n+1} = V) &= \mathbb{P}(X_{n+1} = V|X_n = R)\mathbb{P}(X_n = R) + \mathbb{P}(X_{n+1} = V|X_n = V)\mathbb{P}(X_n = V) + \mathbb{P}(X_{n+1} = V|X_n = B)\mathbb{P}(X_n = B) \\
\mathbb{P}(X_{n+1} = B) &= \mathbb{P}(X_{n+1} = B|X_n = R)\mathbb{P}(X_n = R) + \mathbb{P}(X_{n+1} = B|X_n = V)\mathbb{P}(X_n = V) + \mathbb{P}(X_{n+1} = B|X_n = B)\mathbb{P}(X_n = B)
\end{aligned}$$

i.e.

$$\begin{aligned}
\mathbb{P}(X_{n+1} = R) &= 0.5\mathbb{P}(X_n = R) + 0.2\mathbb{P}(X_n = V) + 0.3\mathbb{P}(X_n = B) \\
\mathbb{P}(X_{n+1} = V) &= 0.3\mathbb{P}(X_n = R) + 0.4\mathbb{P}(X_n = V) + 0.3\mathbb{P}(X_n = B) \\
\mathbb{P}(X_{n+1} = B) &= 0.2\mathbb{P}(X_n = R) + 0.4\mathbb{P}(X_n = V) + 0.4\mathbb{P}(X_n = B)
\end{aligned}$$

Ce qui n'est pas sans rappeler le calcul matriciel.

Représentation graphique de chaîne

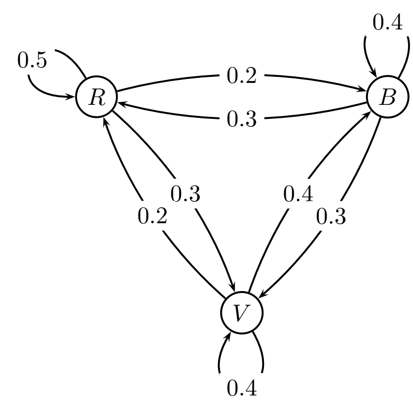


FIGURE 8 – Représentation graphique de l'exemple : on place les états et, sur des flèches, les probabilités de transition d'un état à l'autre

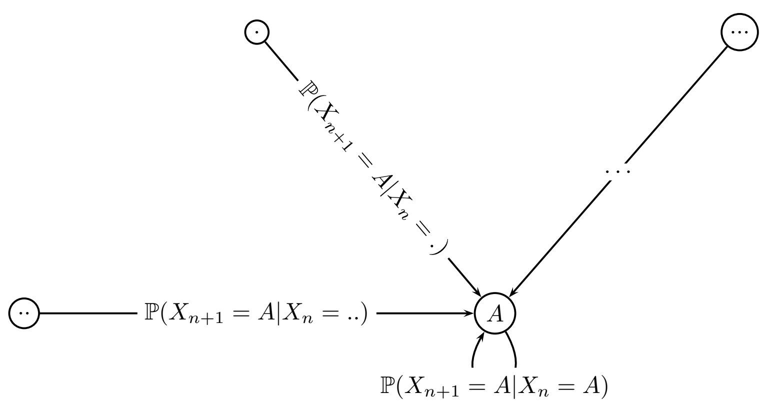


FIGURE 9 – Un noeud, probabilités de l'atteindre, la ligne *A* de la matrice de transition

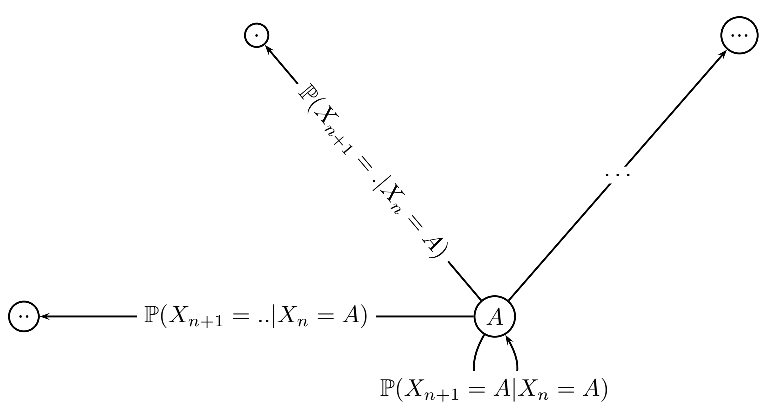


FIGURE 10 – Un noeud, probabilités des destinations, la colonne *A* de la matrice de transition, leur somme vaut 1

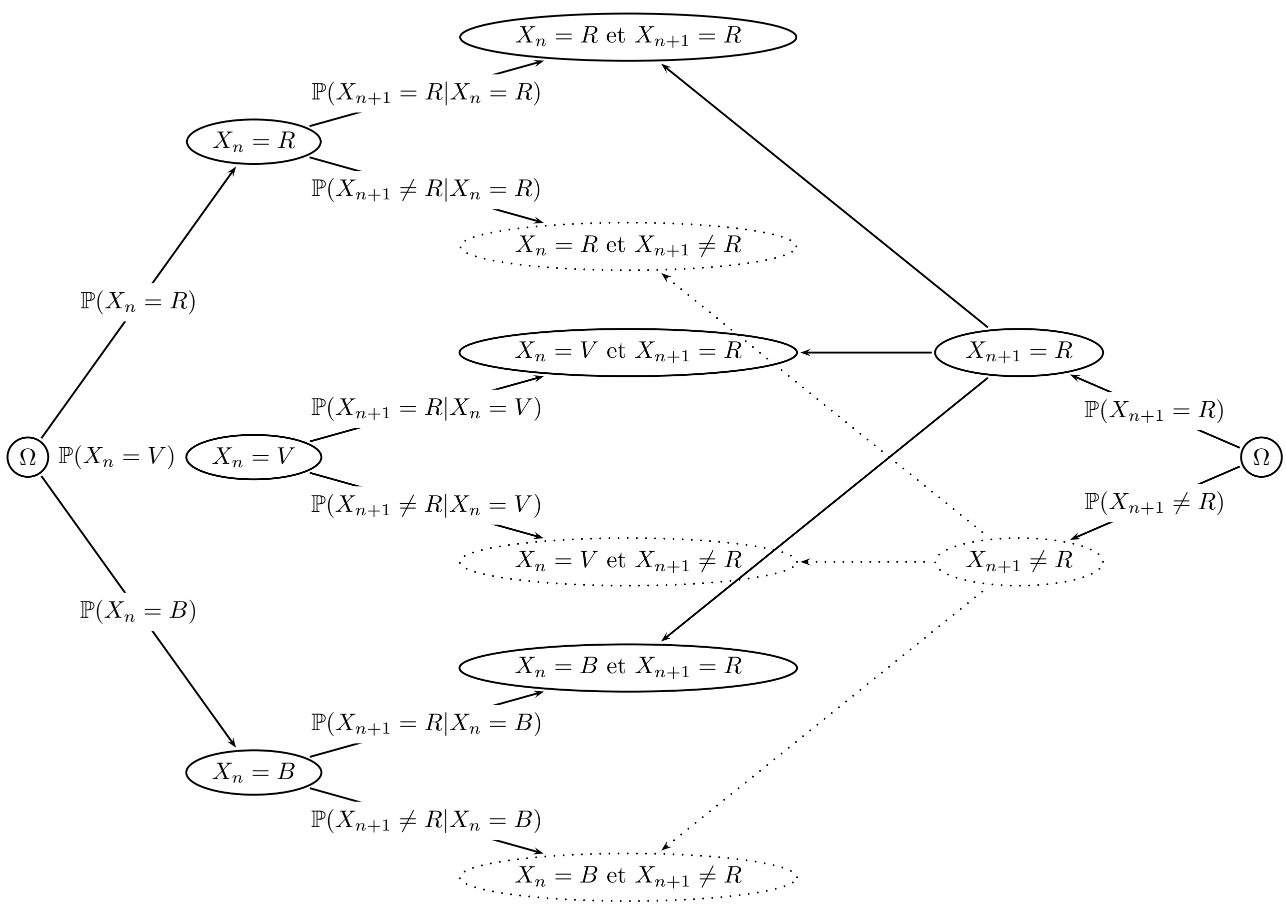


FIGURE 11 – Représentation de l'arbre des choix (initial) pour le calcul de $\mathbb{P}(X_{n+1} = R)$. Cet arbre se traite ensuite comme celui des Fig. 4,5,6 et 7

Calcul matriciel et matrice de transition

Posons

$$A = \begin{pmatrix} 0.5 & 0.2 & 0.3 \\ 0.3 & 0.4 & 0.3 \\ 0.2 & 0.4 & 0.4 \end{pmatrix}, P_n = \begin{pmatrix} \mathbb{P}(X_n = R) \\ \mathbb{P}(X_n = V) \\ \mathbb{P}(X_n = B) \end{pmatrix}$$

La relation précédente se traduit par

$$\forall n \in \{0, \dots, N-1\}, P_{n+1} = A.P_n$$

La récurrence des suites géométriques donne que $P_N = A^N.P_0$ où le vecteur P_0 traduit la position initiale.

Simulation informatique

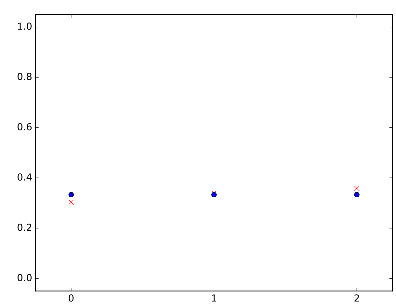


FIGURE 12 – Simulation d'une chaîne de MARKOV, $N = 10$, A et P_0 donné dans le texte. Simulation sur $NS = 100$ expériences. $\circ$=théorique, $\times$=simulé

Avec

$$A' = \begin{pmatrix} 0.5 & 0.2 & 0.3 \\ 0.1 & 0.4 & 0.3 \\ 0.4 & 0.4 & 0.4 \end{pmatrix}$$

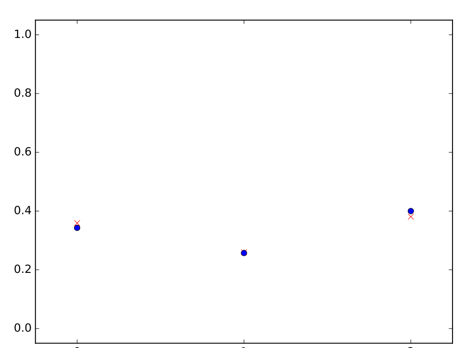


FIGURE 13 – Simulation d'une chaîne de MARKOV, $N = 10$, A', P_0 donné dans le texte. Simulation sur $NS = 1000$ expériences. $\circ$=théorique, $\times$=simulé

Utiliser le script `simulation-markov.py`.

Exercice 3.— On jette n fois une pièce, pile sortant avec probabilité p à chaque tirage. Soit P_n la probabilité pour que le nombre de piles soit pair dans n jets (comptant 0 comme nombre pair). Montrer que

$$P_n = (q-p)P_{n-1} + p$$

où $q = 1-p$.

En déduire que $P_n = \frac{1}{2}(1+(q-p)^n)$ pour $n \geq 0$.

Exercice 4.— On dispose d'une urne qui contient trois boules numérotées de 0 à 2.

On s'intéresse à une suite d'épreuves définies de la manière suivante :

— On choisit au hasard une boule dans cette urne.

— Si j est le numéro de la boule tirée, on enlève de l'urne toutes les boules dont le numéro est strictement supérieur à j. Pour le tirage suivant, l'urne ne contient plus que les boules numérotées de 0 à j.

On considère alors la variable aléatoire réelle X_k égale au numéro de la boule obtenue à la $k^{ème}$ épreuve ($k \geq 0$)

On note alors U_k la matrice unicolonne définie par :

$$U_k = \begin{pmatrix} \mathbb{P}[X_k = 0] \\ \mathbb{P}[X_k = 1] \\ \mathbb{P}[X_k = 2] \end{pmatrix}$$

où $\mathbb{P}[X_k = j]$ est la probabilité de tirer la boule numéro j à la $k^{ème}$ épreuve.

On convient de définir la matrice U_0 par :

$$U_0 = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

1. Préliminaire algèbrique. Soient

$$A = \begin{pmatrix} 1 & \frac{1}{2} & \frac{1}{3} \\ 0 & \frac{1}{2} & \frac{1}{3} \\ 0 & 0 & \frac{1}{3} \end{pmatrix} \text{ et } P = \begin{pmatrix} 1 & 1 & 1 \\ 0 & -1 & -2 \\ 0 & 0 & 1 \end{pmatrix}$$

1.a. Calculer P^2 et déterminer l'inverse P^{-1} de la matrice P.

1.b. Montrer que $A = P.D.P^{-1}$ où

$$D = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \frac{1}{2} & 0 \\ 0 & 0 & \frac{1}{3} \end{pmatrix}$$

1.c. En déduire, pour $k \in \mathbb{N}^*$, une formule explicite donnant A^k.

2. **2.a.** Vérifier que $U_1 = A.U_0$.

2.b. Déterminer la loi de X_2. (On pourra montrer que $U_2 = A.U_1$.) Calculer l'espérance et la variance de X_2

3. Par utilisation de la formule des probabilités totales, prouver que pour tout entier naturel $k \geq 1$:

$$U_{k+1} = A.U_k$$

4. Écrire, pour un entier naturel k, U_k en fonction de A et U_0

5. Pour tout k de $\mathbb{N}^*$, donner la loi de X_k et vérifier que l'on a :

$$\lim_{k\to+\infty} \mathbb{P}[X_k = 0] = 1, \quad \lim_{k\to+\infty} \mathbb{P}[X_k = 1] = 0, \quad \lim_{k\to+\infty} \mathbb{P}[X_k = 2] = 0$$