Introduction aux chaînes de Markov

S. Lemaire

Polycopié pour l'U.E. "Chaînes de Markov"
L3 Biologie-Santé et L3 Biodiversité des Organismes et Ecologie.

Table des matières

I Rappels et compléments sur les variables aléatoires discrètes 3

1 Espace de probabilité 3
1.1 Exercices 4

2 Probabilité conditionnelle 5
2.1 Définitions et propriétés 5
2.2 Indépendance de deux événements 6
2.3 Indépendance conditionnelle 6
2.4 Indépendance d'une famille finie d'événements 7
2.5 Exemples d'application 7
2.6 Exercices 8

3 Variable aléatoire discrète 9
3.1 Loi d'une variable aléatoire discrète 9
3.2 Fonction de répartition d'une variable aléatoire discrète 11
3.3 Moments d'une variable aléatoire discrète 12
3.4 Exercices 13

4 Couple de variables aléatoires discrètes 14
4.1 Loi du couple (X, Y) 14
4.2 Lois conditionnelles 16
4.3 Indépendance de deux variables aléatoires 16
4.4 Indépendance conditionnellement à une variable aléatoire 17
4.5 Espérance et covariance d'un couple de variables aléatoires numériques discrètes 17
4.6 Espérance conditionnelle d'une variable aléatoire réelle discrète 18
4.7 Vecteurs aléatoires discrets 19
4.8 Exercices 20

II Simulations et analyse des résultats 24

5 Simulation d'une expérience aléatoire 24
5.1 Loi des grands nombres 24
5.2 Loi empirique 25
5.3 Fonction de répartition empirique 26
5.4 Exercices 26

6 Simuler une expérience aléatoire sur ordinateur 27
6.1 Nombres pseudo-aléatoires 27
6.2 Simulation d'une variable aléatoire discrète 28

III Introduction aux chaînes de Markov 30

7 Généralités 30
7.1 Définitions et exemples . . . 30
7.2 Graphe associé à une matrice de transition . . . 31
7.3 Caractérisations d'une chaîne de Markov homogène . . . 32
7.4 Simulation des premiers états d'une chaîne de Markov homogène . . . 33
7.5 Exemples de chaînes de Markov . . . 34

8 Loi de la chaîne à un instant donné 35

9 Loi invariante et comportement asymptotique de la loi de X_n 38
9.1 Loi de probabilité invariante . . . 38
9.2 Convergence en loi . . . 40
9.3 Exemples de comportements en loi . . . 41
9.4 Résultats théoriques . . . 43

10 Temps d'atteinte d'un état 46
10.1 Probabilité d'atteinte d'un état . . . 47
10.2 Temps moyen d'atteinte d'un état . . . 48
10.3 Temps d'atteinte d'un sous-ensemble d'états . . . 48
10.4 Les chaînes absorbantes . . . 49

11 Fréquence de passages dans un état et estimation 50
11.1 Exemples d'évolutions de la fréquence des passages dans un état . . . 50
11.2 Chaînes de Markov irréductibles . . . 50
11.3 Comportements asymptotiques . . . 51
11.4 Estimation des paramètres . . . 51

12 Le processus de Galton-Watson 53
12.1 Définition et exemples . . . 53
12.2 Propriétés . . . 53
12.3 Evolution en moyenne de la taille de la population . . . 55
12.4 Probabilité d'extinction de la population issue d'un seul ancêtre . . . 55
12.5 Probabilité d'extinction d'un processus de Galton-Watson issu de k ancêtres . . . 57
12.6 Complément : comportement asymptotique du processus de Galton-Watson dans le cas sur-critique . . . 57
12.7 Exercices . . . 58

13 Références bibliographiques 58

Première partie

Rappels et compléments sur les variables aléatoires discrètes

1 Espace de probabilité

Certaines expériences effectuées dans des conditions déterminées ont un résultat qui comporte un élément d'incertitude ou de hasard, dépendant de facteurs non contrôlés. On les appelle des expériences aléatoires : le lancer d'un dé, le rendement d'un champ de blé, le résultat de l'autofécondation d'une plante hétérozygote... sont des expériences aléatoires.
Mathématiquement, une telle expérience est représentée par le *tirage* d'un élément ω dans un ensemble Ω représentant toutes les issues possibles.
Certains faits associés à cette expérience aléatoire peuvent se produire ou non, on les appelle des événements.
Mathématiquement, un événement sera représenté par une partie de Ω.

événements	représentation ensembliste
événement certain	Ω
événement impossible	$\emptyset$
l'événement A est réalisé	$\omega \in A$
l'événement A n'est pas réalisé	$\omega \in A^c = \Omega \setminus A$
les événements A et B sont réalisés	$\omega \in A \cap B$
A et B sont des événements incompatibles	$A \cap B = \emptyset$
l'événement A ou l'événement B est réalisé	$\omega \in A \cup B$
si l'événement A a lieu alors l'événement B a lieu	$A \subset B$

Tab. 1 – Quelques événements associés à une expérience aléatoire dont le résultat est ω et l'ensemble des résultats possibles Ω

► *Exemple 1.* On considère une famille ayant $n \geq 2$ enfants. On note E_i l'événement "le i-ème enfant est un garçon" pour $i \in \{1, \ldots, n\}$.
Décrire à l'aide des ensembles E_i, les événements suivants

1. F "la famille a au moins un garçon"
2. G "seul l'aîné est un garçon"

Solution : $F = \cup_{i=1}^n E_i$ et $G = E_1 \cap (\cap_{i=2}^n E_i^c)$.

Si l'ensemble des issues possibles de l'expérience aléatoire peut être décrit par un ensemble Ω fini ou dénombrable, chaque point ω de Ω est affecté d'une *probabilité* $P(\omega)$ qui représente la chance qu'a l'expérience d'avoir l'issue représentée par ω. C'est un nombre $P(\omega)$ entre 0 et 1 telle que $\sum_{\omega\in\Omega} P(\omega) = 1$.
La probabilité $P(A)$ qu'un événement A se produise est alors la somme des probabilités $P(\omega)$ sur l'ensemble des éléments ω de A : $P(A) = \sum_{\omega\in A} P(\omega)$.
On définit ainsi une application P définie sur l'ensemble des parties de Ω, noté $\mathcal{P}(\Omega)$, à valeurs dans $[0, 1]$ qui vérifie les propriétés suivantes :

(i) $P(\Omega) = 1$;

(ii) Si $A_1, A_2, \ldots, A_m, \ldots,$ est une suite d'événements deux à deux incompatibles, alors

$$P(\overset{+\infty}{\underset{i=1}{\cup}} A_i) = \sum_{i=1}^{+\infty} P(A_i).$$

► *Exemple 2.* Considérons l'expérience aléatoire consistant à lancer deux dés à six faces.
On peut représenter le résultat d'un lancer des deux dés comme un couple (k, ℓ) de deux entiers compris entre 1 et 6 (on peut considérer par exemple qu'on lance les dés l'un après l'autre et que l'on note dans l'ordre les résultats des deux lancers). L'ensemble des résultats possibles est $\Omega = \{1, \ldots, 6\}^2$. Chaque élément de Ω a autant de chance d'être le résultat d'un lancer. On munit donc Ω de l'équiprobabilité P : $P(\{\omega\}) = \frac{1}{36}$ pour tout $\omega \in \Omega$.
La probabilité d'un événement A est alors $P(A) = \sum_{\omega\in A} \frac{1}{36} = \frac{\mathrm{Card}(A)}{36}$.
Par exemple l'événement "obtenir un double" est décrit par l'ensemble $A = \{(i, i),\ i \in \{1, \ldots, 6\}\}$. Donc, $P(A) = \frac{1}{6}$.
Si on s'intéresse uniquement à la somme des chiffres obtenus avec les deux dés, on pourra décrire l'ensemble des résultats possibles par $\hat{\Omega} = \{2, 3, ..., 12\}$. Mais attention, la probabilité $\hat{P}$ sur cet ensemble $\hat{\Omega}$ n'est pas

l'équiprobabilité ; pour déterminer $\hat{P}$, revenons à la description d'un résultat du lancer de deux dés par un couple d'entiers : l'événement "obtenir une somme égale à k" est décrit par le sous-ensemble :

$$A_k = \{(i,j) \in \{1,...,6\}^2,\ i+j=k\}$$

de Ω. Cet événement a donc pour probabilité $\hat{P}(k) = P(A_k) = \frac{k-1}{36}$ si $k \in \{2,...,7\}$ et $\hat{P}(k) = P(A_k) = \frac{13-k}{36}$ si $k \in \{8,...,12\}$.

► *Exemple 3.* On interroge 5 personnes au hasard dans un groupe de 100 personnes. Si on numérote de 1 à 100 les individus, on peut décrire le résultat d'un tel sondage par une succession de 0 et de 1 : $(x_1,\ldots,x_{100})$ avec $x_k = 1$ si le k-ième individu du groupe a été interrogé et $x_k = 0$ si le k-ième individu du groupe n'a pas été interrogé. Avec ce codage, l'ensemble des résultats possibles pour un tel sondage est $\Omega = \{(x_1,\ldots,x_{100}) \in \{0,1\}^{100},\ \sum_{i=1}^{100} x_i = 5\}$. Comme le choix des 5 personnes se fait au hasard, chacun des éléments $\omega \in \Omega$ a autant de chance d'être le résultat du sondage : la probabilité P sur Ω associée à cette expérience est donc l'équiprobabilité : $P(\{\omega\}) = \frac{1}{\binom{100}{5}}$ pour tout $\omega \in \Omega$.

Dans le cas général d'un ensemble Ω non dénombrable (par exemple $[0,1]$, $\mathbb{R}$, $\{0,1\}^{\mathbb{N}}$ ...), on est amené à restreindre l'ensemble des événements à un ensemble $\mathcal{A}$ plus petit que $\mathcal{P}(\Omega)$; on définit $\mathcal{A}$ comme un sous-ensemble de $\mathcal{P}(\Omega)$ vérifiant les propriétés suivantes :

– si $A \in \mathcal{A}$ alors le complémentaire de A dans Ω appartient à $\mathcal{A}$;
– si $(A_i)_{i\in\mathbb{N}^*}$ est une suite d'éléments de $\mathcal{A}$ alors la réunion de ces ensembles $\cup_{i\in\mathbb{N}^*} A_i$ appartient à $\mathcal{A}$.

Un tel ensemble est appelé une tribu (ou σ-algèbre) sur Ω.
Une probabilité sur $(\Omega,\mathcal{A})$ est alors, comme précédemment, une application $P : \mathcal{A} \to [0,1]$ vérifiant les propriétés (i) et (ii). Le triplet $(\Omega,\mathcal{A},P)$ est appelé *espace de probabilité.*

1.1 Exercices

Vous trouverez d'autres exercices ainsi des documents de cours sur les notations ensemblistes et la manipulation du signe somme dans la classe WIMS.

▷ *Exercice 1.* On effectue une expérience aléatoire dont l'ensemble des résultats possibles est noté Ω. On désigne par A, B et C trois événements qui peuvent se réaliser au cours de cette expérience. Par abus de notations, A, B et C désignent aussi les parties de Ω qui décrivent ces événements : par exemple, si ω est le résultat de l'expérience, il est équivalent de dire que A est réalisé et que $\omega \in A$.

1. Donner l'écriture ensembliste de l'événement F suivant :
   parmi les 3 événements, seul l'événement A est réalisé
2. Décrire par une phrase, l'événement F^c, puis donner l'écriture ensembliste de F^c.

▷ *Exercice 2.* La fonction indicatrice d'un événènement A est la fonction $\mathbb{1}_A$ définie sur Ω par $\mathbb{1}_A(x) = \begin{cases} 1 & \text{si } x \in A \\ 0 & \text{si } x \notin A \end{cases}$
Soit A et B deux événements. Vérifier qu'on a les inégalités suivantes :

$$\mathbb{1}_{A^c} = 1 - \mathbb{1}_A, \quad \mathbb{1}_{A\cap B} = \mathbb{1}_A\mathbb{1}_B, \quad \mathbb{1}_{A\cup B} = \mathbb{1}_A + \mathbb{1}_B - \mathbb{1}_A\mathbb{1}_B$$

$\mathbb{1}_{A\times B}(x,y) = \mathbb{1}_A(x)\mathbb{1}_B(y)$ pour tout $(x,y) \in \Omega^2$,

▷ *Exercice 3.* Un sac contient n pions numérotés de 1 à n. Pour les trois expériences aléatoires suivantes :

E1 : *choisir successivement k pions en remettant chaque pion dans le sac avant d'en choisir un autre*

E2 : *choisir successivement k pions sans remettre le pion dans le sac avant d'en choisir un autre*

E3 : *choisir au hasard une poignée de k pions*

trouver parmi les ensembles suivants un ensemble qui a les deux propriétés suivantes :
– chaque résultat possible pour cette expérience aléatoire est décrit par un élément de cet ensemble,
– tous les éléments décrivent des résultats qui ont la même chance d'arriver.

A1 l'ensemble des n-uplets $(x_1,\ldots,x_n) \in \{1,...,k\}^n$
A2 l'ensemble des k-uplets $(x_1,\ldots,x_k) \in \{1,...,n\}^k$
A3 l'ensemble des n-uplets $(x_1,\ldots,x_n) \in \{0,1\}^n$
A4 l'ensemble des k-uplets $(x_1,\ldots,x_k) \in \{1,...,n\}^k$ dont les coefficients sont tous différents
A5 l'ensemble des k-uplets $(x_1,\ldots,x_k) \in \{0,...,n\}^k$ tels que $\sum_{i=1}^k x_i = n$
A6 l'ensemble des n-uplets $(x_1,...,x_n) \in \{0,1\}^n$ tels que $\sum_{i=1}^n x_i = k$

On explicitera ce que représentent les coordonnées x_i dans les ensembles choisis.

▷ *Exercice 4.* On effectue une expérience aléatoire dont l'ensemble des résultats possibles est $\Omega = \{1, ..., 9\} \times \{1, ..., 11\}$. Pour chaque $(i, j) \in \Omega$, on note $p(i, j)$ la probabilité que (i, j) soit le résultat de l'expérience et pour tout événement $E \subset \Omega$, on note $P(E)$ la probabilité que l'événement E se réalise.
Compléter la formule ci-dessous afin d'exprimer la probabilité de l'événement

$$E = \{(i, j) \in \Omega,\ 3 \leq i + j \leq 12\},$$

en fonction uniquement des valeurs de p :

$$P(E) = \sum_{i=\boxed{}}^{\boxed{}} \sum_{j=\boxed{}}^{\boxed{}} p(i, j)$$

NB : on écrira $min(a, b)$ pour désigner le plus petit des deux nombres a et b et $max(a, b)$ pour désigner le plus grand des deux nombres a et b. On n'utilisera pas de sommes de la forme $\sum_{i=k}^{l}$ avec $k > l$.

2 Probabilité conditionnelle

On désigne par $(\Omega, \mathcal{A}, P)$ un espace de probabilité associé à une expérience aléatoire.

2.1 Définitions et propriétés

Définition. Soit A et B deux événements. On suppose que $P(A) > 0$.
On définit la *probabilité conditionnelle que l'événement B soit réalisé sachant que l'événement A s'est réalisé* par : $P(B \mid A) = \frac{P(B \cap A)}{P(A)}$.

Cela définit une nouvelle probabilité sur Ω qui tient compte de l'information "l'événement A est réalisé".

Pour comprendre d'où vient cette formule, plaçons-nous dans le cas d'une expérience aléatoire dont l'ensemble des résultats possibles Ω est au plus dénombrable. L'ensemble Ω est muni de la probabilité P : pour tout $\omega \in \Omega$, $P(\{\omega\})$ désigne la probabilité que le résultat de l'expérience soit ω. Supposons que cette expérience aléatoire ait été réalisée mais que l'on ignore son issue ; on sait seulement que l'événement A (supposé de probabilité non nulle) a eu lieu. On doit définir une nouvelle probabilité P_A qui tienne compte de cette information. D'une part, sachant que l'événement A est réalisé, on posera $P_A(\{\omega\}) = 0$ si $\omega \in \Omega \setminus A$. D'autre part, il n'y a pas de raison de changer le rapport entre les probabilités de deux élements de A, ce qui amène à poser $P_A(\{\omega\}) = cP(\{\omega\})$ pour tout $\omega \in A$, c étant une constante indépendante de ω. Pour que P_A soit une probabilité, c doit valoir $1/P(A)$. La probabilité P_A est maintenant entièrement définie : $P_A(\{\omega\}) = \frac{P(\{\omega\})}{P(A)} \mathbb{1}_A(\omega)$ pour tout $\omega \in \Omega$. Par conséquent, pour tout événement B, $P_A(B) = \frac{P(B \cap A)}{P(A)}$.

► *Exemple 4.* Pierre apprend qu'une famille avec deux enfants a acheté l'appartement voisin du sien. Il se dit qu'il y a une chance sur deux la famille ait une fille et un garçon (il a supposé que chaque enfant a une probabilité $1/2$ d'être un garçon et $1/2$ d'être une fille).
Dans la conversation avec des voisins, il apprend ensuite qu'au moins un des deux enfants de cette famille est une fille. Avec cette nouvelle information, il y a maintenant 2 chances sur 3 que cette famille ait un garçon et une fille.

L'utilisation des probabilités conditionnelles est un moyen de calculer la probabilité d'une intersection d'événements puisque

$$P(A \cap B) = P(A|B)P(B)$$

et donc de décomposer le calcul de la probabilité d'un événement A en introduisant un événement B dont on connait la probabilité. En effet, on a :

$$P(A) = P(A \cap B) + P(A \cap B^c) = P(A|B)P(B) + P(A|B^c)(1 - P(B)).$$

Ces deux relations se généralisent de la façon suivante :

Proposition 1 *1. Soit $A_1, \ldots, A_m$ m événements. On suppose que $P(A_1 \cap \ldots \cap A_m) > 0$. La probabilité de leur intersection est :*

$$P(A_1 \cap \ldots \cap A_m) = P(A_1)P(A_2|A_1) \ldots P(A_m|A_1 \cap \ldots \cap A_{m-1})$$

2. *Soit $B_1, \ldots, B_m$ une partition*[1] *de l'ensemble Ω et A un événement. Alors,*

$$P(A) = \sum_{i=1}^{m} P(A|B_i)P(B_i).$$

Preuve.

1. La preuve se fait pas récurrence sur m : par définition, on a $P(A_1 \cap A_2) = P(A_2|A_1)P(A_1)$.
   Soit $m \geq 3$. Supposons le résultat vrai pour les $m-1$ événements $A_1, \ldots, A_{m-1}$.
   Comme $A_1 \cap \ldots \cap A_m \subset A_1 \cap \ldots \cap A_{m-1}$, $P(A_1 \cap \ldots \cap A_{m-1}) > 0$. On peut donc conditionner par rapport à $A_1 \cap \ldots \cap A_{m-1}$:
   $$P(A_1 \cap \ldots \cap A_m) = P(A_m|A_{m-1} \cap \ldots \cap A_1)P(A_{m-1} \cap \ldots \cap A_1).$$
   On obtient le résultat en appliquant l'hypothèse de récurrence pour exprimer $P(A_{m-1} \cap \ldots \cap A_1)$
2. $P(A) = \sum_{i=1}^{m} P(A \cap B_i)$ car les ensembles $A \cap B_i$ sont deux à deux disjoints. On conclut en utilisant que $P(A \cap B_i) = P(A|B_i)P(B_i)$.

□

2.2 Indépendance de deux événements

Définition. Deux événements A et B sont *indépendants* si $P(B \cap A) = P(B)P(A)$.

N.B. Comme $P(B \cap A) = P(B \mid A)P(A)$ si $P(A) > 0$, A et B sont indépendants si et seulement si $P(B \mid A) = P(B)$ c'est-à-dire si le fait de savoir que A est réalisé ne change pas la probabilité de B.

N.B. Un événement de probabilité nulle est indépendant de n'importe quel autre événement.

La notion d'indépendance n'est pas toujours intuitive comme le montre l'exemple suivant :

► *Exemple 5.* Soit n un entier supérieur ou égal à deux. On suppose que toutes les répartitions des sexes des enfants d'une famille à n enfants sont équiprobables. Les deux événements :
– M_n : "la famille de n enfants a des enfants des deux sexes"
– F_n : "la famille de n enfants a au plus une fille"
sont indépendants si et seulement si $n = 3$.

2.3 Indépendance conditionnelle

Définition. Soit C un événement tel que $P(C) > 0$. Deux événements A et B sont dit *indépendants conditionnellement à l'événement C* si $P(A \cap B \mid C) = P(A \mid C)P(B \mid C)$.

N.B. Deux événements A et B peuvent être indépendants et ne pas être indépendants conditionnellement à un événement C. Considérer par exemple les événements suivants associés au lancer de deux dés :

– A : "le résultat du premier dé est impair"
– B : "le résultat du second dé est pair"
– C : "les résultats des deux dés sont de même parité"

$P(A \cap B) = \frac{1}{4} = P(A)P(B)$, $P(A \cap B|C) = 0$ alors que $P(A|C) = P(B|C) = \frac{1}{2}$.

De même, deux événements peuvent ne pas être indépendants mais être indépendants conditionnellement à un autre événement. Par exemple, considérons deux pièces A_1 et A_2. Supposons que A_1 a une probabilité $0,9$ de tomber sur face et que A_2 a une probabilité $0,1$ de tomber sur face. On choisit au hasard une des deux pièces et on lance la pièce choisie deux fois. Notons F_i l'événement "on a obtenu face au i-ème lancer" pour $i = 1$ ou 2. Les événements F_1 et F_2 sont indépendants conditionnellement à l'événement "la pièce lancée est A_1". Par contre, les événements F_1 et F_2 ne sont pas indépendants.

Propriété 2 *Soit A, B et C des événements tels que $P(C) > 0$ et $P(B \cap C) > 0$.*
Deux événements A et B sont indépendants conditionnellement à l'événement C si et seulement si

$$P(A|B \cap C) = P(A|C).$$

[1]On dit que $B_1, \ldots, B_m$ définissent une partition de Ω si ce sont des ensembles deux à deux disjoints dont la réunion $\cup_{i=1}^{m} B_i$ est égale à Ω

2.4 Indépendance d'une famille finie d'événements

Définition. m événements $A_1, A_2, \ldots, A_m$ sont dits *indépendants* si et seulement si la probabilité de l'intersection d'un nombre quelconque d'entre eux est le produit de leurs probabilités : pour tout $k \in \{2, \ldots, m\}$ et $1 \leq i_1 < \ldots < i_k \leq m$,

$$P(A_{i_1} \cap \ldots \cap A_{i_k}) = P(A_{i_1}) \ldots P(A_{i_k}).$$

N.B. Trois événements peuvent être indépendants deux à deux mais ne pas être indépendants dans leur ensemble. C'est le cas par exemple des événements A, B et C définis à la remarque précédente.

2.5 Exemples d'application

▷ *Exercice 5. Tests diagnostiques.*
Des tests diagnostiques aident au dépistage de nombreuses pathologies. Lorsqu'un fabricant livre un test servant à détecter une maladie m, il fournit en général deux caractéristiques :
- la *sensibilité* (Se) du test : c'est la probabilité que le test soit positif pour une personne malade. On la souhaite proche de 1 (un test toujours positif aurait une sensibilité égale à 1, et n'aurait pourtant aucun intérêt).
- la *spécificité* (Sp) du test : c'est la probabilité que le test soit négatif pour une personne saine (une spécificité de 0.9 signifie qu'environ 10% de personnes non malades auront un test positif (appelé "faux positif").

On note p la proportion d'individus ayant la maladie m dans la population ciblée par le test (p est appelée la *prévalence* de la maladie m). On choisit au hasard un individu dans cette population.

1. Donner une expression de Se et Sp à l'aide des événements M "l'individu a la maladie m" et T "le test est positif pour cet individu".

On note LR (*Likelihood Ratio*) le rapport : $LR = \frac{Se}{1-Sp}$. Le test étant utilisé pour un diagnostic, une exigence minimale est que LR soit supérieur à 1 (on le supposera dans toute la suite).

2. Expliquer pourquoi et montrer que $P(T) < Se$.

Un usager ou un medecin qui utilise le test pour faire un diagnostic s'intéresse lui aux quantités suivantes :
- la *valeur prédictive positive* du test (VPP) : c'est la probabilité d'être atteint de la maladie m si le test est positif ;
- la *valeur prédictive négative* du test (VPN) : c'est la probabilité de n'être pas atteint par la maladie m lorsque le test est négatif ;

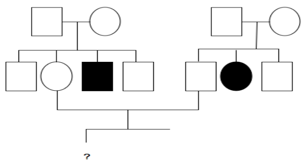
3. Donner une expression de VPP et VPN à l'aide des événements M et T.
4. Donner une expression de VPP en fonction de p et LR.
5. Que vaut VPP et VPN pour un test dont la sensibilité est de 0.8 et la spécificité est de 0.9 lorsque $p = 0.01$? lorsque $p = 0.3$?
6. Montrer que VPP $> p$. Que peut-on dire de VPP si p augmente ? si LR augmente ?

Combinaisons de tests : il arrive que deux tests T_1 et T_2 soient utilisés pour le diagnostic d'une même maladie m. On peut souvent supposer que les résultats des tests sont indépendants conditionnellement au fait d'avoir ou non la maladie m. On supposera dans la suite que l'on a deux tests T_1 et T_2 dont les résultats sont indépendants conditionnellement au fait d'avoir la maladie m.

7. On construit un test TT à partir des tests T_1 et T_2 en disant que le résultat du test TT est positif si et seulement si les résultats aux deux tests T_1 et T_2 sont positifs. Donner l'expression du rapport LR du test TT en fonction des rapports LR_1 et LR_2 des tests T_1 et T_2.
8. On suppose que les deux tests T_1 et T_2 ont les mêmes caractéristiques : $Se = 0.8$ et $Sp = 0.9$. On fait passer les tests T_1 et T_2 a une personne choisie au hasard dans une population qui a une prévalence de 0.01. Déterminer :
   (a) la probabilité que le résultat du test T_1 soit positif ;
   (b) la probabilité que le résultat du test T_2 soit positif sachant que le résultat du test T_1 est positif ;
   (c) la probabilité que la personne soit atteinte de la maladie m sachant que les résultats des tests T_1 et T_2 sont positifs.
9. Commenter les résultats obtenus.

▷ *Exercice 6. Le risque d'albiniste pour un enfant à naitre* L'albinisme oculo-cutané de type 1 est à "transmission autosomale récessive". Il a pour origine la mutation d'un gène T qui code une enzyme appelée tyrosinase dont le dysfonctionnement empêche la synthèse de la mélanine. L'allèle fonctionnel de ce gène qui code pour la tyrosinase fonctionnelle est noté T^+. L'allèle muté de ce gène, qui code une tyrosinase déficiente est noté T^a. L'expression de T^+ est dominante, celle de T^a est récessive.

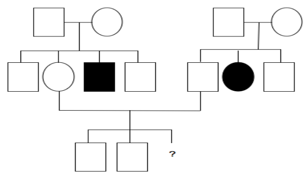
1. Quelle est la probabilité pour qu'un individu non albinos dont les deux parents sont hétérozygotes pour ce gène soit lui aussi hétérozygote ?



Le schéma ci-contre représente l'arbre généalogique d'une famille touchée par l'albinisme. On peut déduire de cet arbre que les parents de l'enfant à naitre ne sont pas albinos et que les grands-parents sont hétérozygotes pour le gène T.

2. Quelle est la probabilité pour que les parents de l'enfant à naître soient hétérozygotes ?
3. Quelle est la probabilité pour que l'enfant à naitre ne soit pas albinos ?

Les individus atteints d'albinisme sont représentés en noir.



On considère le même couple quelques années plus tard. Il a déjà deux enfants non albinos et attend un troisième enfant (voir figure ci-contre).

4. Quelle est la probabilité pour que l'enfant à naitre ne soit pas albinos ?

2.6 Exercices

▷ *Exercice 7.*

1. Donner une formulation mathématique de l'affirmation suivante : *"un fumeur a plus de chance de développer un cancer du poumon qu'un non fumeur"*.
2. Soit A et B deux événements. On suppose que $P(A) > 0$ et $0 < P(B) < 1$.
   Montrer que les quatre inégalités : $P(A|B) \geq P(A)$, $P(B|A) \geq P(B)$, $P(A|B^c) \leq P(A)$ et $P(A|B) \geq P(A|B^c)$ sont équivalentes.
3. Donner des formulations équivalentes à l'affirmation *"un fumeur a plus de chance de développer un cancer du poumon qu'un non fumeur"*.

▷ *Exercice 8.* On considère une expérience aléatoire dont l'ensemble des résultats est noté Ω. On note P une probabilité sur Ω associée à cette expérience aléatoire. On s'intéresse à deux événements A et B qui ont une probabilité strictement comprise entre 0 et 1 d'être réalisés. On suppose connu

$$p = P(A),\ q = P(B) \text{ et } r = P(A|B).$$

Donner l'expression de $P(B|A^c)$ en fonction de p, q et r.

▷ *Exercice 9.* Il manque une carte dans un jeu de 52 cartes et on ignore laquelle. On tire une carte au hasard dans ce jeu incomplet, c'est un coeur. Quelle est la probabilité que la carte manquante soit un coeur ?

▷ *Exercice 10.* Un laboratoire a mis au point un alcootest dont les propriétés sont les suivantes :
– il se révèle positif pour quelqu'un qui n'est pas en état d'ébriété dans 2% des cas,
– il se révèle positif pour quelqu'un qui est en état d'ébriété dans 96% des cas.
Dans un département donné, on estime que 3% des conducteurs sont en état d'ébriété.

1. Quelle est la probabilité que lors d'un controle, l'alcootest se révèle positif ?
2. Un controle dans ce département avec cet alcootest s'est révélé positif. Quelle est la probabilité que le conducteur ne soit pas malgré tout en état d'ébriété ?
3. Si un controle se révèle négatif, quelle est la probabilité que le conducteur controlé ne soit effectivement pas en état d'ébriété ?

▷ *Exercice 11.* Dans une population donnée, 73% des victimes d'une infection virale présente un symptôme qui n'atteint que 8% de la population non infectée. On sait de plus que 30% de la population présente ce symptôme.

1. Quelle est la probabilité qu'un individu choisi au hasard dans cette population soit infecté ?
2. Quelle est la probabilité qu'un individu présentant le symptôme ne soit pas infecté ?
3. Quelle est la probabilité qu'un individu ne présentant pas le symptôme, soit infecté ?

▷ *Exercice 12.*

Supposons que dans une population, on observe trois phénotypes différents notés A, B et C avec la répartition suivante parmi les mères de famille : 21% de phénotypes A, 55% de phénotypes B, 24% de phénotypes C.
Le tableau suivant donne la proportion d'enfants de phénotype donné en fonction du phénotype de la mère.

	enfant A	enfant B	enfant C
mère A	35%	25%	40%
mère B	40%	19%	41%
mère C	40%	42%	18%

Par exemple, le tableau indique que 35% des enfants dont la mère est de phénotype A ont le phénotype A.
On choisit un enfant au hasard dans cette population.

1. Quelle est la probabilité qu'il n'ait pas le phénotype B et que sa mère n'ait pas le phénotype B ?
2. L'enfant choisi n'a pas le phénotype B. Quelle est la probabilité que sa mère n'ait pas le phénotype B ?

$\triangleright$ *Exercice 13.* Pour préserver l'anonymat des personnes interrogées pour un sondage sur la consommation de drogues, on demande à chacun de lancer une pièce bien équilibrée puis de répondre
– "oui" si la pièce tombe sur "face" ;
– "oui" si la pièce tombe sur "pile" et qu'ils ont consommé de la drogue ;
– "non" si la pièce tombe sur "pile" et qu'ils n'ont pas consommé de drogue.
Le résultat du lancer de la pièce reste inconnu du sondeur.
Donner une expression de la probabilité qu'une personne tirée au hasard dans cette population réponde "oui" à ce sondage, en fonction de p. Déterminer la probabilité qu'une personne qui a répondu "oui" au sondage ait consommé de la drogue.

$\triangleright$ *Exercice 14.* On choisit au hasard 3 enfants différents dans un groupe constitué de $n > 3$ garçons et de $r > 3$ filles.

1. Quelle est la probabilité que l'on ait choisi dans l'ordre : un garçon, une fille, un fille ?
2. Quelle est la probabilité que l'on ait choisi en tout : un garçon et deux filles ?

$\triangleright$ *Exercice 15.* On suppose que A, B et C sont des événements indépendants de probabilité respectivement a, b et c. Exprimer en fonction de a, b et c la probabilité de l'événement $F = A \cup (B^c \cap C^c)$.

$\triangleright$ *Exercice 16.* Soit B un événement de probabilité $0 < P(B) < 1$. Montrer qu'un événement A est indépendant de B si et seulement si $P(A|B^c) = P(A|B)$.

$\triangleright$ *Exercice 17.* On dispose de sachets de graines de mufliers provenant de cultures différentes. On sème les graines d'un paquet. Les événements "le premier muflier a des fleurs jaunes", "le deuxième muflier a des fleurs jaunes" peuvent-ils être considérés comme indépendants ?

$\triangleright$ *Exercice 18.* Soit $\alpha, \beta \in]0, 100[$. On dispose d'un test dont la fonction est de détecter la présence d'une substance M dans le sang. Le test est positif pour α % des échantillons sanguins contenant la substance M et est négatif pour β % des échantillons sanguins ne contenant pas la substance M. On effectue le test sur deux échantillons de sang d'une même personne pour savoir si la substance M est présente dans son sang ou non. Les événements "le résultat du test avec le premier échantillon est positif" et "le résultat du test avec le deuxième échantillon est positif" sont-ils indépendants ?

$\triangleright$ *Exercice 19.* On s'intéresse à un gène d'une plante qui peut s'exprimer sous deux formes A et a. Par autofécondation, on obtient à partir d'une plante hétérozygote pour ce gène, une plante appartenant à la génération $F1$. On effectue une autofécondation sur cette plante pour obtenir une plante de génération $F2$,...
Notons E_i l'événement "la plante de la génération Fi est hétérozygote". Montrer que les événements E_1 et E_3 sont indépendants conditionnellement à l'événement E_2. Les événements E_1 et E_3 sont-ils indépendants ?

$\triangleright$ *Exercice 20.* Lorsque la rivière qui longe le village XXX déborde, la probabilité que la source du village soit polluée est de 0.3 le lendemain. Dans le cas où cette rivière déborde, un employé municipal est chargé de prélever 3 échantillons d'eau de la source le lendemain et d'analyser séparément chaque échantillon. L'analyse d'un échantillon n'est pas totalement fiable :
– dans seulement 80 % des cas, l'analyse d'un échantillon contenant de l'eau polluée indiquera que l'eau est polluée.
– dans seulement 90 % des cas, l'analyse d'un échantillon contenant de l'eau saine indiquera que l'eau est non polluée.
Si l'eau contenue dans 2 échantillons sur les 3 échantillons prélevés est déclarée polluée par l'analyse effectuée, quelle est la probabilité que le maire se trompe en déclarant que l'eau est polluée ?

3 Variable aléatoire discrète

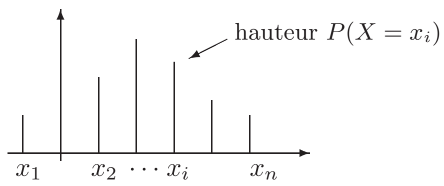
Dans une expérience aléatoire, une *variable aléatoire* (en abrégé v.a.) X est une quantité dont la valeur, a priori incertaine, est déterminée à l'issue de l'expérience. Elle est donc représentée comme une application X définie sur l'ensemble Ω des résultats possibles de l'expérience. La valeur prise par l'application X à la suite d'une expérience aléatoire, est appelée une *réalisation* de la variable aléatoire X. Par exemple, lors du tirage d'un individu dans une population, la taille, le poids ou le revenu annuel de l'individu tiré sont des réalisations de variables aléatoires. Le nombre obtenu en lançant un dé est une réalisation d'une variable aléatoire à valeurs dans $\{1, \ldots, 6\}$.

3.1 Loi d'une variable aléatoire discrète

Lorsqu'une variable aléatoire X ne peut prendre qu'un nombre fini ou dénombrable de valeurs $x_1, \ldots, x_n$ (avec n fini ou $n = +\infty$), on dit que la variable est discrète. On peut affecter à chaque valeur x_i une probabilité, celle de l'événement "X prend la valeur x_i". Cet événement se note en abrégé "$X = x_i$", il est défini par l'ensemble $\{\omega \in \Omega,\ X(\omega) = x_i\}$. Les nombres $P(X = x_i)$ pour $i \in \{1, \dots, n\}$ sont positifs ou nuls et leur somme est égale à 1. Ils définissent donc une probabilité sur l'ensemble $\mathcal{X} = \{x_1, \dots, x_n\}$ que l'on appelle la *loi* de X.

Définition. Soit X une v.a. à valeurs dans un ensemble fini ou dénombrable $\mathcal{X} = \{x_1, \dots, x_n\}$.
La loi de X est la probabilité μ sur $\mathcal{X}$ définie par $\mu(\{x_i\}) = P(X = x_i)$ pour tout $i \in \{1, \dots, n\}$.
On appellera *support de la loi de* X l'ensemble des valeurs x_i prises par X avec une probabilité strictement positive.

Lorsque X est une variable aléatoire à valeurs réelles, on peut représenter graphiquement la loi de X à l'aide d'un diagramme en bâtons : pour chaque valeur x_i du support de la loi de X, on dessine un bâton de hauteur $P(X = x_i)$.



► *Exemple 6.* Le résultat du lancer d'un dé est une variable aléatoire qui a une probabilité $\frac{1}{6}$ de prendre chacun des entiers de 1 à 6. On dit que X suit la loi uniforme sur l'ensemble $\{1, \dots, 6\}$. Si au cours d'un lancer, on a obtenu le nombre 5, 5 est la *réalisation* de la variable aléatoire X pour cette expérience aléatoire.
Le carré du nombre obtenu en lançant un dé est la variable aléatoire X^2. Sa loi est la loi uniforme sur l'ensemble $\{1, 4, 9, 16, 25, 36\}$.

► *Exemple 7.* L'indicatrice d'un événement A est une variable aléatoire qui peut prendre deux valeurs 0 ou 1 :

$$\mathbb{1}_A(\omega) = \begin{cases} 1 & \text{si } \omega \in A \\ 0 & \text{si } \omega \in \Omega \setminus A \end{cases}$$

Sa loi est donnée par les relations $P(\mathbb{1}_A = 1) = p$, $P(\mathbb{1}_A = 0) = 1 - p$. Cette loi portée par $\{0, 1\}$ est appelée la *loi de Bernoulli de paramètre* p et notée en abrégé $\mathcal{B}(p)$.

► *Exemple 8.* Si on lance n fois une pièce qui a une probabilité $p \in]0, 1[$ de tomber sur "pile", alors le nombre de fois où cette pièce tombe sur "pile" au cours des n lancers est une variable aléatoire X qui prend ses valeurs dans $\{0, \dots, n\}$ et dont la loi est définie par :

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k} \text{ pour tout } k \in \{0, \dots, n\}.$$

On dit que X suit la loi binomiale $\mathcal{B}(n, p)$.

Preuve. On décrit le résultat des n lancers par un n-uplet $(x_1, \dots, x_n)$ avec

$$x_i = \begin{cases} 1 & \text{si la pièce tombe sur pile au } i-\text{ième lancer} \\ 0 & \text{si la pièce tombe sur face au } i-\text{ième lancer} \end{cases}$$

Le nombre de fois où l'événement A est réalisé au cours des n expériences est une réalisation de la variable X définie par l'application

$$X : \begin{array}{lcl} \Omega & \to & \{0, \dots, n\} \\ (x_1, \dots, x_n) & \mapsto & \sum_{i=1}^n x_i \end{array}$$

Par exemple, si on lance 4 fois la pièce ($n = 4$) et que la pièce tombe sur "face" seulement au 3-ième lancer, le résultat de ces 4 lancers sera décrit par $\omega = (1, 1, 0, 1)$ et la réalisation de la v.a. X est $X(\omega) = 3$.
L'ensemble des résultats possibles pour ces n expériences est alors décrit par l'ensemble $\Omega = \{0, 1\}^n$. La probabilité que le résultat des n lancers soit décrit par $\omega = (x_1, \dots, x_n)$ est $P(\{\omega\}) = p(x_1)...p(x_n)$ avec $p(1) = p = 1 - p(0)$.
Donc, $P(\{\omega\}) = p^{X(\omega)}(1-p)^{n-X(\omega)}$ pour tout $\omega \in \Omega$.
Déterminons la loi de X : soit $k \in \{0, \dots, n\}$. Comme pour tout $\omega \in \{X = k\}$, $P(\{\omega\}) = p^k(1-p)^{n-k}$, on en déduit que $P(X = k) = \text{Card}(\{X = k\})p^k(1-p)^{n-k}$. On conclut utilisant que

$$\{X = k\} = \{(x_1, \dots, x_n) \in \{0, 1\}^n,\ \sum_{i=1}^n x_i = k\} \text{ a } \binom{n}{k} \text{ éléments.}$$

□

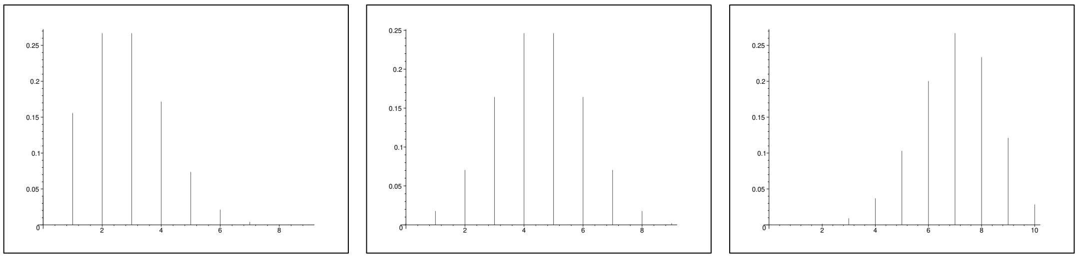


Fig. 1 – De gauche à droite, diagrammes des lois $\mathcal{B}(9, 0.3)$, $\mathcal{B}(9, 0.5)$ et $\mathcal{B}(10, 0.7)$

La loi d'une variable aléatoire permet de calculer la probabilité de n'importe quel événement dépendant uniquement de cette variable aléatoire : pour tout $A \subset \mathcal{X}$,

$$P(X \in A) = \sum_{x \in A} P(X = x) = \sum_{i=1}^{n} P(X = x_i) \mathbb{1}_{\{x_i \in A\}}.$$

► *Exemple 9.* Si X désigne le nombre obtenu en lançant un dé, la probabilité d'obtenir un nombre strictement supérieur à 4 s'écrit : $P(X > 4) = P(X = 5) + P(X = 6) = \frac{1}{3}$.

3.2 Fonction de répartition d'une variable aléatoire discrète

Définition. Pour une variable aléatoire réelle X, on appelle *fonction de répartition de X* la fonction F_X définie par

$$F_X(t) = P(X \leq t) \text{ pour tout } t \in \mathbb{R}.$$

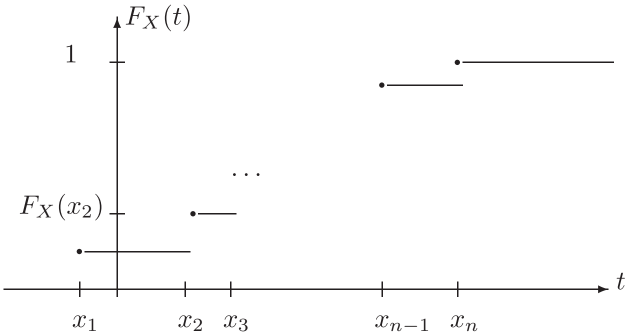


Fig. 2 – Fonction de répartition d'une variable aléatoire X dont les éléments du support de la loi sont $x_1 < x_2 < \ldots < x_n$.

La fonction de répartition est donc une fonction croissante qui a une limite égale à 0 en $-\infty$ et 1 en $+\infty$, et qui fait des sauts en chaque valeur x_i dont la hauteur est $P(X = x_i)$. Remarquons que la connaissance de la fonction de répartition d'une v.a. est équivalente à la connaissance de sa loi.
La probabilité que X prenne une valeur dans un intervalle $]a, b]$ c'est-à-dire la probabilité de l'événement "$X \in]a, b]$" peut se calculer facilement en utilisant la fonction de répartition de X :

$$P(a < X \leq b) = F_X(b) - F_X(a).$$

► *Exemple 10.* (*Loi géométrique sur* $\mathbb{N}$) Soit $n \geq 1$ un entier. On lance n fois une pièce qui a une probabilité $p \in]0, 1[$ de tomber sur pile. On note X_n le nombre de fois où la pièce tombe sur "face" avant de tomber sur "pile" (on pose $X_n = n$ si la pièce n'est pas tombée sur pile au cours des n lancers).
X_n est une variable aléatoire à valeurs dans $\{0, \ldots, n\}$ et on a $P(X_n = k) = p(1-p)^k$ pour tout $k \in \{0, \ldots, n-1\}$ et $P(X_n = n) = (1-p)^n$.
La probabilité que la pièce ne tombe pas sur "pile" si on la lance n fois est $(1-p)^n$. Cette probabilité tend vers 0 lorsque n tend vers $+\infty$. On peut donc définir une variable aléatoire X à valeurs dans $\mathbb{N}$ comme la limite de X_n lorsque n tend vers $+\infty$, correspondant au nombre de fois où la pièce tombe sur "face" avant de tomber sur "pile" si on la lance autant de fois qu'il faut pour que la pièce tombe au moins une fois sur "pile". La loi de X est appelée la loi géométrique sur $\mathbb{N}$: $P(X = k) = p(1-p)^k$ pour tout $k \in \mathbb{N}$. La fonction de répartition de X au

point $k \in \mathbb{N}$ est $P(X \le k) = \sum_{i=0}^{k} p(1-p)^i = 1-(1-p)^{k+1}$ (car $\sum_{i=0}^{k} a^i = \frac{1-a^{k+1}}{1-a}$ pour $a \neq 1$) et la fonction de répartition en tout point $t \in [k, k+1[$ est $P(X \le t) = P(X \le k) = 1-(1-p)^{k+1}$.
Remarquons que l'on peut faire le calcul de $P(X \le k)$ différemment en utilisant l'événement contraire "$X > k$" : c'est l'événement "*la pièce est tombée au moins $k+1$ fois sur face avant de tomber sur "pile"*". Donc $P(X \le k) = 1 - P(X > k) = 1-(1-p)^{k+1}$.

3.3 Moments d'une variable aléatoire discrète

Soit X une variable aléatoire réelle à valeurs dans un ensemble $\mathcal{V}$ au plus dénombrable.

Définition. Si X ne prend que des valeurs positives, on définit son espérance par : $E(X) = \sum_{x \in \mathcal{V}} xP(X = x)$ (cette quantité peut être égale à $+\infty$).
Dans le cas où X prend des valeurs positives et négatives, on dit que X admet une espérance seulement si $E(|X|) = \sum_{x \in \mathcal{V}} |x|P(X = x)$ est finie[2]. On définit l'espérance de X par : $E(X) = \sum_{x \in \mathcal{V}} xP(X = x)$.

▶ *Exemple 11.* Si X est la variable aléatoire désignant le nombre obtenu en lançant un dé, alors $E(X) = \sum_{i=1}^{6} \frac{i}{6} = 3,5$.

▶ *Exemple 12.* Si $\mathcal{V} = \mathbb{N}$ alors $E(X) = \sum_{i=0}^{+\infty} iP(X = i)$.

▶ *Exemple 13.* On dit qu'une variable aléatoire X à valeurs dans $\mathbb{N}$ suit la loi de Poisson de paramètre $a > 0$ si : $P(X = k) = e^{-a}\frac{a^k}{k!}$ pour tout $k \in \mathbb{N}$. Son espérance est $E(X) = a$.
En effet, $E(X) = \sum_{k=0}^{+\infty} ke^{-a}\frac{a^k}{k!} = a\sum_{k=1}^{+\infty} e^{-a}\frac{a^{k-1}}{(k-1)!}$. On conclut en faisant le changement de variable, $j = k-1$ dans la dernière somme.

▶ *Exemple 14.* *(suite de l'exemple 10)*
On lance un pièce jusqu'à ce qu'elle tombe sur "pile". On note p la probabilité que cette pièce tombe sur "pile" à chaque lancer. On note X le nombre de fois où la pièce tombe sur "face" avant de tomber sur "pile". Pour calculer l'espérance de X on utilise par exemple que $\sum_{k=0}^{n} ka^{k-1} = \frac{1}{(1-a)^2}(1-a^n(n(1-a)+1)$ pour[3] $a \neq 1$. On obtient :

$$E(X) = \sum_{k=0}^{+\infty} kp(1-p)^k = \lim_{n\to+\infty} \sum_{k=0}^{n} kp(1-p)^k = \frac{1-p}{p}.$$

Application : si on répète une expérience aléatoire dans les mêmes conditions et si à chaque expérience la chance de succès est de 0.3 alors l'espérance du nombre d'échecs avant un succès est $\frac{7}{3} \simeq 2.33$

Proposition 3 *Pour toute fonction f, la variable aléatoire $f(X)$ admet une espérance si*

$$E(|f(X)|) = \sum_{x \in \mathcal{V}} |f(x)|P(X = x) \text{ est finie}$$

et on a alors, $E(f(X)) = \sum_{x \in \mathcal{V}} f(x)P(X = x)$.

Preuve. Notons $\mathcal{Y} = \{y_1, \ldots, y_n, ...\}$ l'ensemble des valeurs prises par la variable aléatoire $Y = |f(X)|$. Comme les ensembles $A_y = \{x \in \mathcal{V},\ y = |f(x)|\}$ pour $y \in \mathcal{Y}$ définissent une partition de $\mathcal{V}$, pour énumérer les éléments de $\mathcal{V}$, il suffit de commencer par énumérer les éléments de A_{y_1}, puis ceux de A_{y_2}, puis ceux de A_{y_3} ... Donc,

$$\sum_{x \in \mathcal{V}} |f(x)|P(X = x) = \sum_{y \in \mathcal{Y}} y(\sum_{x \in A_y} P(X = x)).$$

En utilisant que $P(Y = y) = \sum_{x \in A_y} P(X = x)$, on obtient : $\sum_{x \in \mathcal{V}} |f(x)|P(X = x) = E(Y)$. Comme $E(Y)$ est finie, $\sum_{x \in \mathcal{V}} f(x)P(X = x)$ est bien définie et peut être obtenue en sommant les termes de la famille $\{f(x)P(X = x),\ x \in \mathcal{V}\}$ dans n'importe quel ordre, ce qui permet de conclure. □

▶ *Exemple 15.*
- Soit a et b deux réels. Si $E(|\ X\ |)$ est finie, $E(aX + b) = aE(X) + b$.
- $E(X^2) = \sum_{x \in \mathcal{V}} x^2 P(X = x)$.
- Si $s \in [-1, 1]$ et X est une variable aléatoire à valeurs dans $\mathbb{N}$ alors $E(|s^X|) = \sum_{k=0}^{+\infty} |s|^k P(X = k) \le 1$ car $|s| \le 1$. Donc, la variable aléatoire s^X admet une espérance qui s'écrit $E(s^X) = \sum_{k=0}^{+\infty} s^k P(X = k)$; cette fonction du paramètre s est appelée la *fonction génératrice de la variable aléatoire* X. Par exemple, si X suit la loi de Bernoulli $\mathcal{B}(p)$, alors $E(s^X) = s^0(1-p) + s^1 p = (1-p) + sp$.

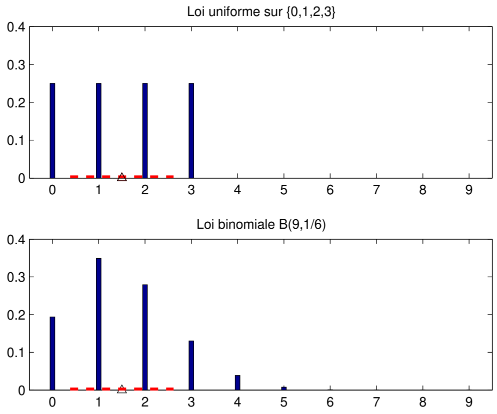
[2]Si $\{a_i,\ i \in I\}$ est une famille dénombrable de réels telle que $\sum_{i\in I} |\ a_i\ | < +\infty$, alors on peut sommer les termes de la famille $\{a_i,\ i \in I\}$ dans n'importe quel ordre, on obtiendra toujours le même résultat que l'on note $\sum_{i\in I} a_i$. Ce n'est pas le cas par exemple de la famille de nombres $\{a_k = \frac{(-1)^k}{k}, k \in \mathbb{N}^*\}$

[3]On peut montrer cette formule par récurrence sur n par exemple.

Définition. On appelle *moment d'ordre k de X*, l'espérance de X^k lorsqu'elle existe.
Lorsque $E(X^2)$ est finie, on définit la *variance de X* par $\text{Var}(X) = E((X - E(X))^2) = E(X^2) - E(X)^2$ et *l'écart-type de X* par $\sigma(X) = \sqrt{\text{Var}(X)}$.

N.B. La variance mesure l'amplitude moyenne des fluctuations de la variable autour de sa moyenne.

▶ *Exemple 16.* Les deux diagrammes ci-contre montrent deux lois différentes ayant même espérance et même variance. Il s'agit de la loi uniforme sur $\{0, 1, 2, 3\}$ (diagramme du haut) et de la loi binomiale $\mathcal{B}(3, 9/16)$ (diagramme du bas) : leur espérance est 1.5 et leur variance est 1.25. Le triangle indique la valeur de l'espérance. Le segment en pointillés délimite l'intervalle $[E(X) - \sigma(X), E(X) - \sigma(X)]$.



▶ *Exemple 17.* Si X est une variable aléatoire de loi de Bernoulli $\mathcal{B}(p)$, alors $E(X) = p$ et plus généralement $E(X^k) = p$ pour tout $k \in \mathbb{N}^*$. La variance de X est $\text{Var}(X) = p(1-p)$.

▶ *Exemple 18.* Considérons l'expérience aléatoire consistant à interroger une personne au hasard parmi un groupe de n individus pour connaitre son score, un entier entre 0 et 100, à une compétition. On désignera par $\Omega = \{\omega_1, \ldots, \omega_n\}$ l'ensemble des individus du groupe et par $X(\omega_i)$ le score obtenu par la i-ème personne du groupe à cette compétition et on munira Ω de l'équiprobabilité P. Pour tout $i \in \{0, 1, \ldots, 100\}$, $P(X = i)$ est la proportion de personnes dans le groupe ayant obtenu le score i. La moyenne des scores obtenus est $\frac{1}{n}\sum_{i=1}^n X(\omega_i)$. Si on regroupe les individus qui ont obtenu le même score dans l'expression de la moyenne, on obtient :

$$\frac{1}{n}\sum_{i=1}^{n} X(\omega_i) = \sum_{i=0}^{100} iP(X = i) = E(X).$$

Donc dans cet exemple, l'espérance de X correspond à la moyenne des scores du groupe.
On montre de la même façon que le moment d'ordre k de X s'écrit $E(X^k) = \frac{1}{n}\sum_{i=1}^n (X(\omega_i))^k$.

Proposition 4 *L'espérance de X est la constante qui approche le mieux la variable aléatoire X au sens suivant : la fonction $a \mapsto \mathrm{E}((X-a)^2)$ atteint son minimum en $\mathrm{E}(X)$. De plus, pour tout $a \in \mathbb{R}$, $\mathrm{E}((X-a)^2) = \text{Var}(X) + (\mathrm{E}(X) - a)^2$.*

Preuve. En introduisant $E(X)$ et en utilisant la linéarité de l'espérance, on obtient :

$$\begin{aligned}
\mathrm{E}((X-a)^2) &= \mathrm{E}(((X - \mathrm{E}(X)) + \mathrm{E}(X) - a)^2) \\
&= \mathrm{E}((X - \mathrm{E}(X))^2) + (\mathrm{E}(X) - a)^2 + 2(\mathrm{E}(X) - a)\mathrm{E}(X - \mathrm{E}(X)) \\
&= \text{Var}(X) + (\mathrm{E}(X) - a)^2.
\end{aligned}$$

□

3.4 Exercices

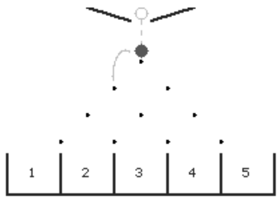
▷ *Exercice 21.* On considère une variable aléatoire X qui prend ses valeurs dans l'ensemble $\{-1, 0, 1, 2\}$.

i	-1	0	1	2
$P(X = i)$	1/10	1/10	2/5	2/5

1. Calculer $P(|X| \leq 1)$
2. Calculer la valeur de la fonction de répartition de X au point 0.6
3. Déterminer l'espérance et la variance de X.
4. Déterminer la loi de X^2.

▷ *Exercice 22.*

Une bille tombe sur une pyramide de clous comme cela est représenté sur le schéma. A chaque fois qu'elle arrive sur un clou, elle a une probabilité 0.2 de rebondir à droite du clou et une probabilité 0.8 de rebondir à gauche du clou. On note X le numéro de la boite dans laquelle tombe une telle bille. Déterminer la loi de X.



▷ *Exercice 23.* Montrer que si X est une variable aléatoire discrète telle que $E(X^2)$ est finie alors pour tous réels a et b, $\mathrm{Var}(aX+b) = a^2\mathrm{Var}(X)$.

▷ *Exercice 24.* Une source émet une suite de lettres choisies indépendamment les unes des autres parmi les lettres **a**, **b** et **c** suivant la loi de probabilité décrite par le tableau :

Lettre	a	b	c
Probabilité	0.3	0.2	0.5

1. On note X le nombre de fois où la lettre **a** apparaît parmi les sept premières lettres émises.
   (a) Calculer $P(X \geq 3)$.
   (b) Si vous deviez parier sur le nombre de fois où la lettre **a** apparaitra sur les 7 prochaines lettres émises, quel serait votre pari ? Quelle est la probabilité que vous gagniez votre pari ?
2. On note Y la variable aléatoire donnant le nombre de lettres émises avant que la lettre **b** apparaisse. Calculer $P(Y > 5)$.

▷ *Exercice 25.* Une analyse en laboratoire est délicate : elle réussit 7 fois sur 10.

1. Si on effectue cette analyse 3 fois, quelle est la probabilité de la réussir une seule fois ?
2. Soit $k \in \mathbb{N}^*$. Calculer la probabilité qu'il faille recommencer k fois cette analyse pour la réussir.
3. Si cette analyse coûte 100 euros. Quel budget doit-on prévoir si on veut avoir au moins 99% de chance de réussir cette analyse une fois ? deux fois ?

4 Couple de variables aléatoires discrètes

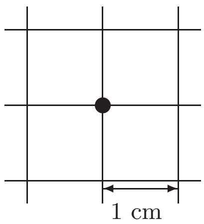
Soit X et Y deux variables aléatoires définies sur un même espace de probabilité $(\Omega, \mathcal{A}, P)$, à valeurs dans des ensembles au plus dénombrables notés $\mathcal{X}$ et $\mathcal{Y}$ respectivement.

4.1 Loi du couple (X, Y)

Définition. La loi du couple (X, Y) est une probabilité sur $\mathcal{X} \times \mathcal{Y}$ notée $P_{(X,Y)}$ et définie par $P_{(X,Y)}((x, y)) = P((X, Y) = (x, y))$ pour tout couple $(x, y) \in \mathcal{X} \times \mathcal{Y}$.

N.B.L'événement $\{(X, Y) = (x, y)\}$ est l'événement "X prend la valeur x <u>et</u> Y prend la valeur y". On l'écrit aussi $\{X = x \text{ et } Y = y\}$ ou encore $\{X = x,\ Y = y\}$.

▷ *Exercice 26.*



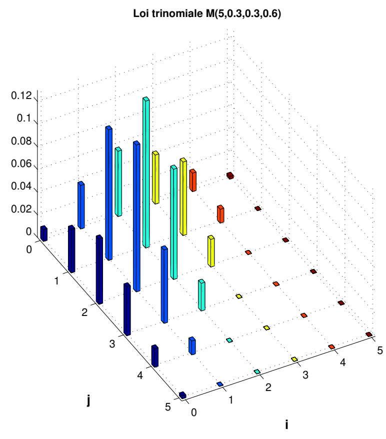
Une fourmi se déplace sur un grillage régulier : à l'instant 0 elle se trouve en $(0, 0)$. A chaque instant, elle choisit au hasard de se déplacer d'1cm vers l'une des 4 directions possibles : vers la droite, vers la gauche, vers le haut ou vers le bas. La position (en cm) de la fourmi après un 1er déplacement est décrite par un couple de variables aléatoires $Z = (X, Y)$. Quelle est la loi de Z ?

► *Exemple 19.* Au cours de la réplication de l'ADN, une base peut être substituée par une autre : on dit qu'il y a *transition* si une purine est remplacée par une purine (par exemple une adénine a par une guanine g) ou si une pyrimidine est remplacée par une autre pyrimidine. On dit qu'il y a *transvection* si une purine est remplacée par une pyrimidine ou inversement. On note p la probabilité qu'il y ait une transition et q la probabilité qu'il y ait une transvection sur une base donnée $(p + q < 1)$.
On suppose que le résultat de la réplication au niveau d'une base ne dépend pas de ce qui se passe sur les autres bases de la séquence d'ADN.
Le nombre de transitions et le nombre de transvections sur une séquence d'ADN ayant n bases, observée lors d'une réplication, sont décrits par des variables aléatoires notées respectivement X et Y. Le couple (X, Y) prend ses

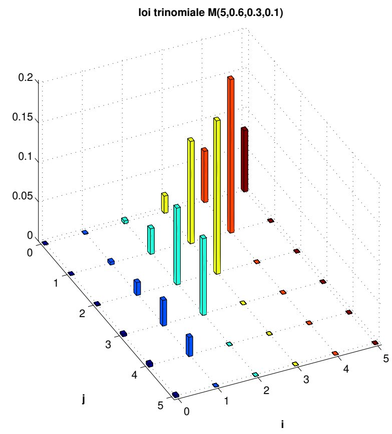
valeurs dans l'ensemble $V = \{0, \dots, n\} \times \{0, \dots, n\}$. Soit $(k,l) \in V$. Si $k+l > n$, alors $P((X,Y) = (k,l)) = 0$. Si $k+l \leq n$, alors

$$P((X,Y) = (k,l)) = \frac{n!}{k!l!(n-k-l)!} p^k q^l (1-p-q)^{n-k-l}.$$

On dit que $(X, Y, n-X-Y)$ suit la loi trinomiale $\mathcal{M}(n, (p, q, 1-p-q))$.



Représentation de la loi trinomiale $\mathcal{M}(5, (0.3, 0.3, 0.6))$: $P(X = i,\ Y = j)$ est la hauteur de la barre positionnée au point (i, j).



Représentation de la loi trinomiale $\mathcal{M}(5, (0.6, 0.3, 0.1))$.

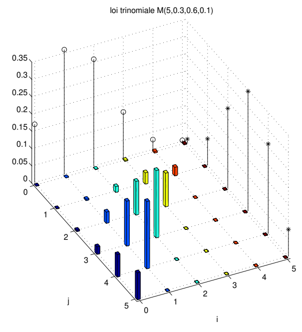
N.B. On appellera *support de la loi d'un couple de variables aléatoires discrètes*, l'ensemble des valeurs que ce couple de variables aléatoires prend avec probabilité strictement positive. Dans l'exemple 19, le support de la loi de (X, Y) est $\{(k, l) \in \{0, \dots, n\}^2,\ k+l \leq n\}$.

Si on connait la loi du couple (X, Y), on peut retrouver la loi de X et la loi de Y qui sont appelées les *lois marginales* du couple (X, Y).

Proposition 5 *Pour tout $x \in \mathcal{X}$, $P(X = x) = \sum_{y \in \mathcal{Y}} P(X = x,\ Y = y)$.*
Pour tout $y \in \mathcal{Y}$, $P(Y = y) = \sum_{x \in \mathcal{X}} P(X = x,\ Y = y)$.

Par contre, si on connait les deux lois marginales, on ne peut pas en général retrouver la loi du couple.

► *Exemple 20.* Dans l'exemple 19, page 14, la loi de X est la loi binomiale $\mathcal{B}(n, p)$ et la loi de Y est la loi binomiale $\mathcal{B}(n, q)$.



Représentation de la loi d'un couple (X, Y) et des lois des marginales c'est-à-dire de la loi de X (au fond) et de la loi de Y (à droite). Ici $(X, Y, 5-X-Y)$ suit la loi trinomiale $\mathcal{M}(5, (0.3, 0.6, 0.1))$ et donc les lois de X et de Y sont respectivement les lois binomiales $\mathcal{B}(5, 0.3)$ et $\mathcal{B}(5, 0.6)$.

Plus généralement, à partir de la loi du couple (X, Y) on peut déterminer la loi de n'importe quelle fonction de X et de Y :

Proposition 6 *Soit f une fonction définie sur $\mathcal{X} \times \mathcal{Y}$ et soit $\mathcal{Z}$ l'ensemble des valeurs prises par la variable aléatoire $Z = f(X, Y)$.*
Pour tout $z \in \mathcal{Z}$, si on note $A_z = \{(x, y) \in \mathcal{X} \times \mathcal{Y},\ f(x, y) = z\}$ alors $P(Z = z) = \sum_{(x,y) \in A_z} P(X = x,\ Y = y)$

► *Exemple 21.* Si X et Y sont deux variables aléatoires à valeurs dans $\{1, \dots, n\}$ et $\{1, \dots, m\}$ respectivement alors $X+Y$ est une variable aléatoire à valeurs dans $\{2, \dots, n+m\}$ et pour tout $k \in \{2, \dots, n+m\}$, l'ensemble $A_k = \{(x, y) \in \{1, \dots, n\} \times \{1, \dots, m\},\ x+y = k\}$ s'écrit $A_k = \{(x, k-x),\ x$ est entier et $\max(1, k-m) \leq$

$x \leq \min(k-1,n)\}$ et donc $P(X+Y=k) = \sum_{i=1}^{n} P(X=i \text{ et } Y=k-i) = \sum_{i=\max(1,k-m)}^{\min(k-1,n)} P(X=i \text{ et } \; Y=k-i)$ car $P(X=i, \text{ et } Y=k-i)=0$ si $k-i \notin \{1,\ldots,m\}$. On peut aussi faire le calcul en décomposant en fonction des valeurs de Y :

$$P(X+Y=k) = \sum_{j=1}^{m} P(X=k-j \text{ et } Y=j) = \sum_{j=\max(1,k-n)}^{\min(k-1,m)} P(X=k-j \text{ et } Y=j).$$

▷ *Exercice 27.* Dans l'exercice 26 page 14, quelle est la loi de X ? la loi de Y ? la loi de $D=\sqrt{X^2+Y^2}$?

4.2 Lois conditionnelles

En utilisant les probabilités conditionnelles, on obtient des décompositions de la probabilité de l'événement $\{(X,Y)=(x,y)\}$ lorsque cette probabilité est non nulle :

$$P((X,Y)=(x,y)) = P(Y=y|X=x)P(X=x) = P(X=x|Y=y)P(Y=y).$$

Définition. Etant donné, un événement A de probabilité non nulle, on appelle *la loi conditionnelle de Y sachant l'événement A*, la loi de probabilité sur $\mathcal{Y}$ définie par la famille de probabilités

$$\{P(Y=y|A), \; y \in \mathcal{Y}\}.$$

Pour tout $x \in \mathcal{X}$ tel que $P(X=x)>0$, on peut définir la loi conditionnelle de Y sachant $\{X=x\}$:

Définition. On appelle *lois conditionnelles de Y sachant X*, la famille des lois conditionnelles de Y sachant $\{X=x\}$ pour x parcourant le support de la loi de X.

L'ensemble de ces lois conditionnelles décrit l'influence de X sur la valeur de Y. La donnée de la loi de X et des lois conditionnelles de Y sachant X détermine évidemment la loi conjointe de X et de Y (donc la loi de Y), et réciproquement.

► *Exemple 22.* On dispose d'un sac contenant n jetons numérotés de 1 à n. On choisit au hasard un jeton du sac et on le retire, on note X_1 son numéro. On choisit alors un deuxième jeton dans le sac ; on note X_2 son numéro. D'après l'énoncé, la loi de X_1 est la loi uniforme sur l'ensemble $\{1,\ldots,n\}$ et pour tout $k \in \{1,\ldots,n\}$, la loi conditionnelle de X_2 sachant que $\{X_1=k\}$ est la loi uniforme sur $\{1,\ldots,k-1,k+1,\ldots,n\}$. On peut alors en déduire que la loi de (X_1,X_2) est la loi uniforme sur $\{(i,j)\in\{1,\ldots,n\}^2, \; i\neq j\}$ et que la loi de X_2 est aussi la loi uniforme sur $\{1,\ldots,n\}$.

► *Exemple 23.* Dans l'exemple 19, page 14, un calcul montre que pour tout $k\in\{0,\ldots,n\}$, la loi conditionnelle de Y sachant $\{X=k\}$ est la loi $\mathcal{B}(n-k,\frac{q}{1-p})$

▷ *Exercice 28.* Dans l'exercice 26 page 14, quelles sont les lois conditionnelles de Y sachant X ?

4.3 Indépendance de deux variables aléatoires

Définition. Les deux variables aléatoires X et Y sont dites *indépendantes* si

$$\text{pour tout } (x,y)\in\mathcal{X}\times\mathcal{Y}, \; P(X=x, \; Y=y) = P(X=x)P(Y=y)$$

En utilisant les lois conditionnelles de X sachant Y, on obtient une autre caractérisation de l'indépendance entre X et Y :

Proposition 7 *Les deux variables aléatoires X et Y sont indépendantes si et seulement si pour tout y appartenant au support de la loi de Y, la loi de X conditionnellement à l'événement $\{Y=y\}$ coïncide avec la loi de X.*

► *Exemple 24.* Dans l'exemple 22, la loi conditionnelle de X_2 sachant $\{X_1=k\}$ dépend de k donc, X_1 et X_2 ne sont pas indépendantes.

► *Exemple 25.* Dans l'exemple 19, page 14, on a vu que la loi conditionnelle de Y sachant $\{X=k\}$ dépend de k. Donc, les variables aléatoires X et Y ne sont pas indépendantes.

4.4 Indépendance conditionnellement à une variable aléatoire

Définition. Soit Z une autre variable aléatoire discrète définie sur Ω à valeurs dans un ensemble $\mathcal{Z}$ au plus dénombrable.
Les variables aléatoires X et Y sont dites *indépendantes conditionnellement à la variable aléatoire* Z si X et Y sont indépendantes conditionnellement à tout événement de la forme $\{Z = z\}$, c'est-à-dire si pour tout $(x, y) \in \mathcal{X} \times \mathcal{Y})$ et pour tout z appartenant au support de la loi de Z,

$$P(X = x,\ Y = y|Z = z) = P(X = x|Z = z)P(Y = y|Z = z).$$

Propriété 8 *Soit X, Y et Z des variables aléatoires définies sur un même espace de probabilité $(\Omega, \mathcal{A}, P)$ et à valeurs dans des ensembles au plus dénombrables $\mathcal{X}$, $\mathcal{Y}$ et $\mathcal{Z}$ respectivement.*
Les variables aléatoires X et Y sont indépendantes conditionnellement à la variable aléatoire Z si et seulement si pour tout (y, z) dans le support de la loi de (Y, Z), la loi conditionnelle de X sachant $\{Y = y$ et $Z = z\}$ ne dépend pas de y.
Dans ce cas, la loi conditionnelle de X sachant $\{Y = y$ et $Z = z\}$ est égale à la loi conditionnelle de X sachant $\{Z = z\}$.

Preuve. Si $P(X = x|Y = y, Z = z) = T(z, x)$ pour tout $x \in \mathcal{X}$ et $(y, z) \in supp((X, Y))$,

$$\begin{aligned} P(X = x|Z = z) &= \sum_{y \in \mathcal{Y}} P(X = x, Y = y|Z = z) = \sum_{y \in \mathcal{Y}} P(X = x|Y = y, Z = z)P(Y = y|Z = z) \\ &= T(x, z) \sum_{y \in \mathcal{Y}} P(Y = y|Z = z) = T(x, z). \end{aligned}$$

D'où l'egalité des lois conditionnelles de X sachant $\{Y = y$ et $Z = z\}$ et de X sachant $\{Z = z\}$. On a aussi obtenu que

$$P(X = x, Y = y|Z = z) = P(X = x|Y = y, Z = z)P(Y = y|Z = z) = P(X = x|Z = z)P(Y = y|Z = z) \quad (1)$$

pour tout (y, z) dans le support de la loi de (Y, Z). Si (y, z) n'est pas dans le support de la loi de (Y, Z) alors $P(X = x, Y = y|Z = z) = 0$ et $P(Y = y|Z = z) = 0$ donc l'égalité (1) est encore satisfaite. On a montré l'indépendance conditionnelle de (X, Y) sachant Z.
Inversement, supposons l'indépendance conditionnelle de (X, Y) sachant Z. Alors, $P(X = x|Y = y, Z = z) = P(X = x, Y = y|Z = z)/P(Y = y|Z = z) = P(X = x|Z = z)$. □

► *Exemple 26.* (*La fortune d'un joueur*) On considère un joueur qui dispose initialement de s euros ($s \in \mathbb{N}$). Il joue à un jeu de hasard, dont l'enjeu à chaque partie est de 1 euro, jusqu'à ce qu'il soit en possession de m euros ($m \in \mathbb{N}$ tel que $m \geq s$) ou jusqu'à ce qu'il ait dépensé tout son argent.
La somme dont dispose le joueur après la 3-ième partie et la somme dont il dispose après la 1-ière partie ne sont pas des variables aléatoires indépendantes. Par contre, ce sont des variables aléatoires indépendantes conditionnellement à la variable aléatoire donnant la fortune du joueur après la deuxième partie.

4.5 Espérance et covariance d'un couple de variables aléatoires numériques discrètes

On suppose ici que les variables aléatoires discrètes X et Y sont à valeurs réelles (i.e. $\mathcal{X}$ et $\mathcal{Y}$ sont supposés inclus dans $\mathbb{R}$)

Définition. On appelle *espérance du couple* (X, Y), le couple des espérances $(E(X), E(Y))$ lorsque $E(X)$ et $E(Y)$ sont bien définies.

Proposition 9 *Soit f une fonction définie sur $\mathcal{X} \times \mathcal{Y}$. Si $E(|f(X, Y)|)$ est finie alors*

$$E(f(X, Y)) = \sum_{x \in \mathcal{X}} \sum_{y \in \mathcal{Y}} f(x, y)P(X = x,\ Y = y).$$

Preuve. La preuve est analogue à la preuve de la proposition 3. □

► *Exemple 27.*

1. Si $E(|X|)$ et $E(|Y|)$ sont finies alors pour tout $a, b \in \mathbb{R}$, $E(|aX+bY|)$ est finie et $E(aX+bY) = aE(X)+bE(Y)$.

2. Si $E(X^2)$ et $E(Y^2)$ sont finies alors $E(|XY|)$ est finie et

$$E(XY) = \sum_{x\in\mathcal{X}}\sum_{y\in\mathcal{Y}} xyP(X=x,\ Y=y) = \sum_{y\in\mathcal{Y}}\sum_{x\in\mathcal{X}} xyP(X=x,\ Y=y)$$

Définition. Supposons que $E(X^2) < +\infty$ et que $E(Y^2) < +\infty$.
La *covariance entre X et Y*, notée $\mathrm{Cov}(X,Y)$, est définie par $\mathrm{Cov}(X,Y) = E((X-E(X))(Y-E(Y)))$.
En particulier, $\mathrm{Cov}(X,X) = \mathrm{Var}(X)$.

Propriété 10 *Supposons que $E(X^2) < +\infty$ et que $E(Y^2) < +\infty$. Alors,*

(i) $\mathrm{Cov}(X,Y) = E(XY) - E(X)E(Y)$.

(ii) si $E(Z^2) < +\infty$ et a est un réel, alors $\mathrm{Cov}(X+aY,Z) = \mathrm{Cov}(Z,X+aY) = \mathrm{Cov}(X,Z)+a\mathrm{Cov}(Y,Z)$. *En particulier* $\mathrm{Var}(X+Y) = Var(X) + Var(Y) + 2\mathrm{Cov}(X,Y)$.

(iii) $|\mathrm{Cov}(X,Y)| \le \sigma(X)\sigma(Y)$ *et l'égalité est réalisée si et seulement si il existe $\lambda \in \mathbb{R}$ et $c \in \mathbb{R}$ tels que $P(Y = \lambda X + c) = 1$.*

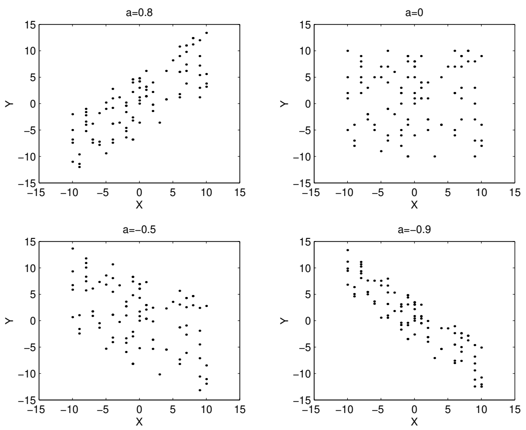
Preuve. (i) et (ii) s'obtiennent en utilisant la propriété de linéarité de l'espérance décrite dans l'exemple 27.
Pour montrer (iii), il suffit de considérer $\mathrm{Var}(X+tY) = \mathrm{Var}(X)+t^2\mathrm{Var}(Y)+2t\mathrm{Cov}(X,Y)$. C'est un polynome d'ordre 2 en t, positif ou nul. Donc, le discriminant $\mathrm{Cov}(X,Y)^2 - \mathrm{Var}(X)\mathrm{Var}(Y)$ doit être négatif ou nul.
D'autre part, si $\mathrm{Cov}(X,Y)^2 = \mathrm{Var}(X)\mathrm{Var}(Y)$, alors $\mathrm{Var}(X+tY) = 0$. Cela ne peut se produire que si $X+tY$ est une variable aléatoire constante avec probabilité 1. (Rappelons que si $X \ge 0$, $E(X) = 0$ implique que $xP(X = x) = 0$ pour tout $x \in \mathcal{X}$, c'est-à-dire $P(X = x) = 0$ pour tout $x \neq 0$ et donc $P(X = 0) = 1$). □

Définition. Considérons deux variables aléatoires X et Y qui admettent des moments d'ordre 2 finis et des variances non nulles.
On définit la *coefficient de corrélation entre X et Y* comme le rapport

$$\mathrm{Corr}(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sigma(X)\sigma(Y)}.$$

Propriété 11 *Supposons que X et Y admettent des moments d'ordre 2 finis et des variances non nulles.* $\mathrm{Corr}(X,Y)$ *est un réel dont la valeur absolue est majorée par* 1 *avec égalité si et seulement si il existe $\lambda \in \mathbb{R}^*$ et $c \in \mathbb{R}$ tels que $Y = \lambda X + c$ (dans ce cas, le signe de* $\mathrm{Corr}(X,Y)$ *est donné par celui de λ).*

▶ *Exemple 28.* Soit X et U deux variables aléatoires indépendantes centrées d'écart-type σ non nul. Notons Y la variable aléatoire définie par $Y = aX + \sqrt{1-a^2}U$. Alors, le vecteur $Z = (X,Y)$ est d'espérance nulle, de covariance $a\sigma^2$ et de coefficient de corrélation a.
Sur les figures ci-contre sont représentées 100 réalisations du vecteur aléatoire Z pour différentes valeurs de a, lorsque X et U suivent la loi uniforme sur $\{-10,-9,\dots,10\}$.



▶ *Exemple 29.* Dans l'exemple 19, page 14, pour $i \in \{1,\dots,n\}$, on note :
– ϵ_i la variable aléatoire qui vaut 1 s'il y a une transition à la i-ème base et 0 sinon.
– η_i la variable aléatoire qui vaut 1 s'il y a une transvection à la i-ème base et 0 sinon.
Alors, pour tout tout $i,j \in \{1,\dots,n\}$, $\mathrm{Cov}(\epsilon_i,\eta_j) = -pq\mathbb{1}_{\{i=j\}}$. En remarquant que $X = \sum_{i=1}^n \epsilon_i$ et $Y = \sum_{j=1}^n \eta_j$ on obtient que $E(X) = np$, $E(Y) = nq$, $\mathrm{Var}(X) = np(1-p)$, $\mathrm{Var}(Y) = nq(1-q)$ et que la covariance entre X et Y vaut $-npq$.

4.6 Espérance conditionnelle d'une variable aléatoire réelle discrète

On suppose ici que les variables aléatoires discrètes X et Y sont à valeurs réelles et que $E(|X|)$ et $E(|Y|)$ sont finies.

Définition. Si A est un événement de probabilité strictement positive, *l'espérance conditionnelle de Y sachant l'événement A* notée $E(Y|A)$ est définie comme l'espérance de la loi conditionnelle de Y sachant A : $E(Y|A) = \sum_{y\in\mathcal{Y}} yP(Y = y \mid A)$.

N.B.L'expression de $E(Y|A)$ a bien un sens car $\sum_{y\in\mathcal{Y}} |y|P(Y = y \mid A) \le \sum_{y\in\mathcal{Y}} |y|P(Y = y) = E(|Y|)$.

Propriété 12 *Si A est un événement de probabilité strictement positive, $E(Y|A) = \frac{E(Y\mathbf{1}_A)}{P(A)}$.*

Preuve. $E(Y|A) = \sum_{y\in\mathcal{Y}} \frac{y}{P(A)} P(\{Y = y\} \cap A)$. Comme $P(\{Y = y\} \cap A) = E(\mathbb{1}_{\{Y=y\}}\mathbb{1}_A)$, et $Y = \sum_{y\in\mathcal{Y}} y\mathbb{1}_{\{Y=y\}}$, on obtient le résultat en intervertissant la somme et l'espérance. □

Propriétés 13 *Notons $\mathcal{S}$ le support de la loi de X*

1. *$\sum_{x\in\mathcal{S}} E(Y|X = x)P(X = x) = E(Y)$.*
2. *Si X et Y sont indépendantes, alors $E(Y|X = x) = E(Y)$ pour tout $x \in \mathcal{S}$.*
3. *Pour toutes fonctions Φ et Ψ telles que $E(|\Phi(X)\Psi(Y)|) < +\infty$ et pour tout $x \in \mathcal{S}$,*

$$E(\Phi(X)\Psi(Y)|X = x) = \Phi(x)E(\Psi(Y)|X = x).$$

En particulier, $E(XY|X = x) = xE(Y|X = x)$.

Preuve.

1. Pour montrer que la première égalité a un sens, il faut vérifier que la somme

$$I = \sum_{x\in\mathcal{S}} |E(Y|X = x)|P(X = x) \text{ est finie.}$$

$I = \sum_{x\in\mathcal{S}} |\sum_{y\in\mathcal{Y}} yP(Y = y|X = x)|P(X = x)$. Donc, $I \le \sum_{x\in\mathcal{S}} \sum_{y\in\mathcal{Y}} |y|P(Y = y, X = x)$. En intervertissant les deux sommes, on obtient que cette double somme est égale à $E(|Y|)$ qui est finie par hypothèse. En refaisant le calcul sans les valeurs absolues, on obtient que

$$\sum_{x\in\mathcal{S}} E(Y|X = x)P(X = x) = E(Y).$$

2. Si X et Y sont indépendants alors la loi conditionnelle de Y sachant $\{X = x\}$ est égale à la loi de Y. Donc, $E(Y|X = x) = E(Y)$.
3. Sur l'évènement $\{X = x\}$, la variable aléatoire $\phi(X)$ est constante et égale à $\phi(x)$. Donc, par linéarité de l'espérance,

$$E(\phi(X)\psi(Y)|X = x) = E(\phi(x)\psi(Y)|X = x) = \phi(x)E(\psi(Y)|X = x).$$

□

► *Exemple 30.* Dans l'exemple 28, page 18, $E(Y|X = k) = ak$ pour tout $k \in \{-10, \ldots, 10\}$.

► *Exemple 31.* Dans l'exemple 22 page 16, $E(X_2|X_1 = k) = \frac{1}{2(n-1)}(n(n + 1) - 2k)$ pour tout $k \in \{1, \ldots, n\}$.

► *Exemple 32.* Dans l'exemple 19, page 14, pour tout $k \in \{0, \ldots, n\}$, l'espérance conditionnelle de Y sachant $\{X = k\}$ est $(n - k)\frac{q}{1-p}$.

N.B. Supposons que $E(X^2)$ et $E(Y^2)$ sont finies et considérons la fonction $\phi : x \mapsto E(Y|X = x)$ définie sur le support de la loi de X. Si on connait la valeur de X, mais pas celle de Y, on peut montrer que $\phi(X)$ est la meilleur façon d'approcher Y par une fonction de X au sens des moindres carrés :

$$E((Y - \phi(X))^2) \le E((Y - \psi(X))^2) \text{ pour toute fonction } \psi \text{ telle que } E(\psi(X)^2) < +\infty.$$

La variable aléatoire $\phi(X)$ est notée $E(Y|X)$.

4.7 Vecteurs aléatoires discrets

Les définitions données pour un couple de variables aléatoires discrètes s'étendent au cas d'un nombre fini de variables aléatoires discrètes : soit $X_1, \ldots, X_r$, $r \ge 2$ variables aléatoires définies sur un même espace de probabilité $(\Omega, \mathcal{A}, P)$, X_i étant à valeurs dans un ensemble au plus dénombrable $\mathcal{X}_i$.

Loi d'un vecteur aléatoire. La loi de $(X_1, \ldots, X_r)$ est une probabilité $P_{(X_1,\ldots,X_r)}$ définie sur $\mathcal{X}_1 \times \ldots \times \mathcal{X}_r$ par :

$$P_{(X_1,\ldots,X_r)}((x_1, \ldots, x_r)) = P((X_1, \ldots, X_r) = (x_1, \ldots, x_r)) \text{ pour tout } (x_1, \ldots, x_r) \in \mathcal{X}_1 \times \ldots \times \mathcal{X}_r.$$

Indépendance. Les variables aléatoires $X_1, \dots, X_r$ sont dites *indépendantes* si

$$P(X_1 = x_1, \dots, X_r = x_r) = \prod_{i=1}^{r} P(X_i = x_i) \text{ pour tout } (x_1, \dots, x_r) \in \mathcal{X}_1 \times \dots \times \mathcal{X}_r.$$

Lois conditionnellement à un vecteur. Soit $s \in \{1, \dots, r-1\}$.
Les lois conditionnelles de $(X_1, \dots, X_s)$ conditionnellement au vecteur $(X_{s+1}, \dots, X_r)$ sont formées par l'ensemble des lois conditionnelles de $(X_1, \dots, X_s)$ sachant l'événement $\{(X_{s+1}, \dots, X_r) = (x_{s+1}, \dots, x_r)\}$ pour $(x_{s+1}, \dots, x_r)$ parcourant le support de la loi de $(X_{s+1}, \dots, X_r)$.

Loi multinomiale On répète n fois dans les mêmes conditions une expérience aléatoire $\mathcal{E}$ qui a r issues possibles notées $I_1, \dots, I_r$. Pour $k \in \{1, \dots, r\}$, notons $p(k)$ la probabilité que le résultat de l'expérience soit I_k et désignons par X_k le nombre d'expériences dont le résultat est I_k ; $(p(1), \dots, p(r))$ définit une probabilité sur $\{1, \dots, r\}$ et $X_1, \dots, X_r$ sont des variables aléatoires à valeurs dans $\{0, 1, \dots, n\}$ telles que $\sum_{i=1}^{r} X_i = n$. Pour tout $(j_1, \dots, j_r) \in \{0, \dots, n\}^r$ tel que $\sum_{k=1}^{r} j_i = n$, l'événement "*sur les n expériences, le résultat I_1 est observé j_1 fois, le résultat I_2 est observé j_2 fois, ..., le résultat I_r est observé j_r fois*" s'écrit : $\{(X_1, \dots, X_r) = (j_1, \dots, j_r)\}$. Sa probabilité est

$$P((X_1, \dots, X_r) = (j_1, \dots, j_r)) = \frac{n!}{j_1! \dots j_r!} (p(1))^{j_1} \dots (p(r))^{j_r} \tag{2}$$

On dit que le vecteur aléatoire $(X_1, \dots, X_r)$ suit la loi multinomiale $\mathcal{M}(n, (p(1), \dots, p(r)))$.

Preuve de l'égalité (2) : On peut coder le résultat des n expériences par une suite $(x_1, \dots, x_n)$ de n chiffres : $x_i = k$ signifiant que le résultat de la i-ième expérience est I_k. L'ensemble des résultats possibles est alors $\Omega = \{1, \dots, r\}^n$. La probabilité d'observer une telle suite $\omega = (x_1, \dots, x_n)$ est $P(\omega) = p(x_1) \dots p(x_r)$. En regroupant les termes identiques dans le produit, on obtient $P(\omega) = p(1)^{X_1(\omega)} \dots p(r)^{X_r(\omega)}$. Par conséquent, si $j_1, \dots, j_r$ sont des entiers positifs dont la somme est égale à n, l'événement

$$A_{j_1, \dots, j_r} = \{X_1 = j_1 \text{ et } X_2 = j_2 \text{ et } \dots \text{ et } X_r = j_r\}$$

a pour probabilité : $P(A_{j_1,\dots,j_r}) = \sum_{\omega \in A_{j_1,\dots,j_r}} p(1)^{j_1} \dots p(r)^{j_r} = \text{Card}(A_{j_1,\dots,j_r}) p(1)^{j_1} \dots p(r)^{j_r}$. □

N.B. En particulier, pour $j \in \{1, \dots, r\}$, X_j suit la loi binomiale $\mathcal{B}(n, p(j))$ et si $k \in \{1, \dots, r\}$ est différent de j, $(X_j, X_k, n - X_j - X_k)$ suit la loi trinomiale $\mathcal{M}(n, (p(j), p(k), 1 - p(j) - p(k)))$.

► *Exemple 33.* si on note X_i le nombre de fois où le dé tombe sur i en 10 lancers pour $i = 1, \dots, 6$ alors $(X_1, \dots, X_6)$ suit la loi $\mathcal{M}(10, (1/6, 1/6, 1/6, 1/6, 1/6, 1/6))$.

4.8 Exercices

▷ *Exercice 29.* On dispose d'un sac contenant 7 boules blanches, 7 boules noires et 8 boules rouges indiscernables au toucher. On note X le nombre de boules blanches et Y le nombre de boules noires que l'on obtient en tirant au hasard deux boules différentes.

Compléter le tableau en y mettant les valeurs de $P((X, Y) = (i, j))$ pour $0 \leq i \leq 2$ et $0 \leq j \leq 2$, afin que le tableau décrive la loi du couple (X, Y).

$i \setminus j$	0	1	2
0			
1			
2			

▷ *Exercice 30.*

Soit $\alpha \in [0, 1]$. Soit (X, Y) un couple de variables aléatoires à valeurs dans $\{0, 1, 2\} \times \{0, 1\}$. Le tableau suivant donne la probabilité $P\{(X, Y) = (x, y)\}$, pour tout $(x, y) \in \{0, 1, 2\} \times \{0, 1\}$,

$y \setminus x$	0	1	2
0	$\frac{1}{4}$	0	$\frac{1}{4}$
1	$\frac{\alpha}{2}$	$\frac{1-\alpha}{2}$	0

1. Déterminer la loi des marginales X et Y, puis les lois conditionnelles de X sachant Y.
2. Les variables aléatoires X et Y sont-elles indépendantes ?
3. Déterminer l'espérance de X et de Y.

4. Déterminer la covariance entre X et Y.
5. Déterminer l'espérance conditionnelle de X sachant que $\{Y = y\}$ pour $y \in \{0, 1\}$.
6. En utilisant la question 5, retrouver l'expression de l'espérance de X d'une autre façon.
7. Déterminer la loi de $Z = X + Y$.

▷ *Exercice 31.* Soit X et Y deux variables aléatoires discrètes vérifiant $\mathrm{Var}(X) = 4$, $\mathrm{Var}(Y) = 1$ et $\mathrm{Corr}(X, Y) = -0.5$. On pose $Z = 2X - 3Y + 4$. Déterminer $\mathrm{Var}(Z)$ et $\mathrm{Cov}(X, Z)$.

▷ *Exercice 32.* On considère un couple de variables aléatoires (X, Y) qui prend ses valeurs dans l'ensemble $\{-3, 2, -1\} \times \{-1, 0, 1, 2\}$. Le tableau décrit les lois conditionnelles de X sachant Y : La case de la i-ième ligne de et la j-ième colonne du tableau suivant contient la probabilité que X prenne la valeur i sachant que l'événement $\{Y = j\}$ est réalisé :

i	$P(X = i \mid Y = -1)$	$P(X = i \mid Y = 0)$	$P(X = i \mid Y = 0)$	$P(X = i \mid Y = 2)$
-3	1/3	1/6	1/2	1/6
-2	1/3	2/3	0	1/3
-1	1/3	1/6	1/2	1/2

1- Déterminer l'espérance conditionnelle de X sachant que $\{Y = k\}$ pour $k \in \{-1, 0, 1, 2\}$:

k	-1	0	1	2
$E(X \mid Y = k)$				

2- Le tableau ci-dessous décrit la loi de Y.

k	-1	0	1	2
P(Y = k)	4/9	1/3	1/9	1/9

Déterminer l'espérance de X.
Déterminer la probabilité que le couple (X, Y) prenne la valeur $(-2, 0)$.
Déterminer la loi de X.

▷ *Exercice 33.* Soient X et Y des variables aléatoires à valeurs dans $\mathbb{N}$. Compléter l'expression ci-dessous de la probabilité de l'événement $\{(X, Y) \in E\}$ où $E = \{(i, j) \in \mathbb{N}^2$ tels que $4 \leq i \leq 7$ et $3 \leq i + j < 8\}$

$$P((X,Y) \in E) = \sum_{i=???}^{???} \Big(\sum_{j=???}^{???} P(X = i \text{ et } Y = j) \Big) = \sum_{j=???}^{???} \Big(\sum_{i=???}^{???} P(X = i \text{ et } Y = j) \Big).$$

▷ *Exercice 34.* Quelle est la probabilité pour que le lancer de quatre dés donne deux six et deux trois ?

▷ *Exercice 35.* Paul choisit au hasard un numéro entier entre 1 et 10. Jean choisit au hasard un numéro entier entre 1 et 10. Notons X la variable aléatoire désignant le numéro tiré par Paul et Y la variable aléatoire désignant le numéro tiré par Jean.
Ecrire l'événement A : "*Paul et Jean choisissent le même chiffre*" en utilisant les variables aléatoires X et Y, puis calculer la probabilité qu'un tel événement A se réalise.
Faire de même avec l'événement B : "*le plus grand des deux numéros est supérieur ou égal à 9*" et l'événement C : "*la somme des deux numéros est supérieure ou égale à 15*".

▷ *Exercice 36.* Une source émet une suite de lettres choisies indépendamment les unes des autres parmi les lettres **a**, **b** et **c** suivant la loi de probabilité décrite par le tableau :

Lettre	a	b	c
Probabilité	0.3	0.2	0.5

1. Quelle est la probabilité que sur les 10 premières lettres émises, la lettre **a** apparaisse 3 fois et la lettre **b** apparaisse 2 fois ?
2. On vous dit que dans les 10 premières lettres émises, la lettre **b** est apparue 2 fois, quelle est la probabilité que la lettre **a** soit apparue 3 fois ?
3. Le nombre de lettres émises avant d'observer pour la première fois la lettre **b** est une réalisation d'une variable aléatoire X. Déterminer la loi de X.
4. On note Y le nombre de lettres **c** déjà émises au moment où une lettre **b** apparaît pour la première fois. Déterminer la loi de conditionnelle de Y sachant que $X = 10$.

▷ *Exercice 37.*

1. On considère un couple (X, Y) de variables aléatoires à valeurs dans les entiers positifs ou nuls. Exprimer la probabilité de l'événement $\{X - Y \leq 8\}$ uniquement à l'aide des probabilités $P(X = i$ et $Y = j)$ pour $i \in \mathbb{N}$ et $j \in \mathbb{N}$.

Application : on a mis deux appareils photos à détecteur à infrarouge sur des lieux de passage d'animaux. On note a et b les probabilités qu'un animal passe devant les détecteurs 1 et 2 respectivement au cours d'une journée et X et Y le nombre de jours passés au moment du 1er déclenchement des appareils photos 1 et 2 respectivement.

2. Que représente l'événement $\{X - Y \leq 8\}$? Calculer $P(X - Y \leq 8)$ lorsque $a = b = 1/4$.

3. On introduit une variable aléatoire ϵ qui vaut 1 si l'appareil photo 1 s'est déclenché le premier jour et qui vaut 0 dans le cas contraire.
   (a) Déterminer la loi conditionnelle de X sachant que $\epsilon = 1$ puis la loi conditionnelle de X sachant que $\epsilon = 0$.
   (b) Montrer que la loi conditionnelle de X sachant que $\epsilon = 0$ est la même que la loi de $X + c$ pour un entier c à déterminer.
   (c) On pose $m_0 = E(X|\epsilon = 0)$ et $m_1 = E(X|\epsilon = 1)$. Utiliser la relation existant entre $E(X)$, m_0 et m_1 et la question précédente pour trouver la valeur de $E(X)$.
   (d) Utiliser la même technique que dans la question précédente pour trouver $E(X^2)$.
   (e) En déduire $Var(X - Y)$.

▷ *Exercice 38.* On dispose d'une pièce qui a une probabilité $p \in]0, 1[$ de tomber sur "pile" et $1 - p$ de tomber sur "face".

1. Calculer la probabilité que la pièce tombe n fois sur "face" si on la lance n fois. On note u_n cette probabilité.
2. Montrer que la suite $(u_n)_n$ tend vers 0 lorsque n tend vers $+\infty$.
3. On définit τ comme étant le nombre de lancers que l'on doit faire avant d'obtenir pour la première fois "pile" ou encore le nombre de fois où la pièce tombe sur "face" avant de tomber sur "pile"(on pose $\tau = +\infty$ si on n'obtient jamais "pile"). Déduire de la question précédente que $P(\tau = +\infty) = 0$.
4. Déterminer la loi de τ.
5. On définit la variable aléatoire X par $X = 1$ si la pièce tombe sur "pile" au premier lancer et $X = 0$ si la pièce tombe sur "face" au premier lancer. Déterminer les lois conditionnelles de τ sachant X.
6. Exprimer l'espérance conditionnelle de τ sachant $\{X = 0\}$ en fonction de $E(\tau)$. En déduire la valeur de l'espérance de τ.

▷ *Exercice 39.* On modélise le nombre de graines d'une plante par une variable aléatoire X de loi de Poisson de paramètre $a > 0$ ($P(X = k) = e^{-a}\frac{a^k}{k!}$ pour tout $k \in \mathbb{N}$). On suppose que chacune de ces graines a une probabilité p de germer et ceci indépendamment des autres graines.

1. Soit $k \in \{0, \ldots, n\}$. Déterminer la probabilité que sur n graines, k germent.
2. On note Y le nombre des graines d'une plante qui ont germé. Pour tout $n \in \mathbb{N}$, déterminer la loi conditionnelle de Y sachant que $X = n$. En déduire la loi de Y.
3. Dans cette question, on ne fait aucune hypothèse sur la loi de X, on suppose seulement que $E(X) = a$. Pour $k \in \text{supp}(X)$, déterminer $E(Y|X = k)$ et en déduire $E(Y)$.

▷ *Exercice 40.* On sait que le croisement de deux phénotypes de pois donne naissance à quatre types de pois :
- Type 1 - pois ronds et jaunes
- Type 2 - pois ronds et verts
- Type 3 - pois ridés et jaunes
- Type 4 - pois ridés et verts

Chaque caractère couleur et forme des pois est géré par un gène ayant un allèle dominant (jaune pour la couleur et rond pour la forme) et un allèle récessif. On croise un grand nombre de pois de races purs jaunes et ronds avec des pois verts et ridés de sorte d'obtenir que des pois de génération 1 hétérozygotes pour les deux gènes. On croise des pois de génération 1 pour obtenir des pois de génération 2. On fait l'hypothèse que les deux caractères sont transmis de manière indépendante.

1. Déterminer la probabilité qu'un pois de génération 2 soit de type i pour $i \in \{1, \ldots, 4\}$.
2. On note X_i le nombre de pois de type i sur 20 pois de génération 2 pour $i \in \{1, \ldots, 4\}$.
   (a) Quelle est la loi de X_i ?
   (b) Quelle est la loi de (X_1, X_2, X_3, X_4) ?
   (c) Sur les 20 pois, on sait que 10 sont de type 1, quelle est la probabilité d'avoir 4 pois de type 2 et 4 pois de type 3 ?

▷ *Exercice 41.* Pour obtenir des informations sur le nombre N de grenouilles vivant dans un étang, on en capture en premier lieu un nombre m, on les marque puis on les relache. Après un temps suffisamment long pour que les grenouilles relachées se soient dispersées, on capture de nouveau n grenouilles de l'étang. On supposera que le nombre de grenouilles entre les deux captures n'a pas changé et qu' à chaque capture les grenouilles de l'étang ont toute la même probabilité d'être attrapées.

1. On note A_i l'événement "la i-ème grenouille capturée est une grenouille marquée et on pose $Y_i = \mathbb{1}_{A_i}$. Déterminer la loi de Y_1, puis la loi du couple (Y_1, Y_2)
2. En déduire la loi de Y_2.
3. Soit $i \in \{1, \ldots, n-1\}$. Déterminer les lois conditionnelles de Y_{i+1} sachant $(Y_1, \ldots, Y_i)$.
4. Montrer que pour tout $(y_1, \ldots, y_i) \in \{0, 1\}^i$ tel que $s(y, i) = \sum_{k=1}^{i} y_k \in [m + i - N, m]$,

$$P(Y_i = y_i, Y_{i-1} = y_{i-1}, \ldots, Y_1 = y_1) = \frac{m!(N-m)!}{(m - s(y, i))!(N - m - (i - s(y, i)))!} \frac{(N-i)!}{N!}.$$

5. On pose $S_i = \sum_{j=1}^{i} Y_j$ pour tout $i \in \{1, \ldots, n\}$. Que représente S_i ? Déterminer sa loi.

On suppose que les grenouilles marquées sont numérotées de 1 à m. Pour $i \in \{1, \ldots, m\}$, on note B_i l'événement "la grenouille numéro i a été capturée une deuxième fois" et on pose $Z_i = \mathbb{1}_{B_i}$.

6. Déterminer la loi de Z_i.
7. Calculer $\mathrm{Cov}(Z_i, Z_j)$ pour tout $i, j \in \{1, \ldots, m\}$.
8. Que représente $W = \sum_{j=1}^{m} Z_j$?
9. Déterminer $E(W)$ et $\mathrm{Var}(W)$.

Deuxième partie
Simulations et analyse des résultats

5 Simulation d'une expérience aléatoire

La simulation, c'est l'expérimentation sur un modèle ; le modèle étant une représentation simplifiée d'un phénomène que l'on cherche à étudier. Le modèle cherche à donner une description abstraite de certains aspects de la réalité en terme de variables et de relations entre ces variables. Expérimenter sur le modèle consiste à faire varier la valeur des variables d'entrée afin d'observer la réaction du modèle.
Lorsqu'on reproduit un phénomène faisant intervenir le hasard, les résultats que l'on obtient sont alors des réalisations de variables aléatoires dont la loi est souvent inconnue.

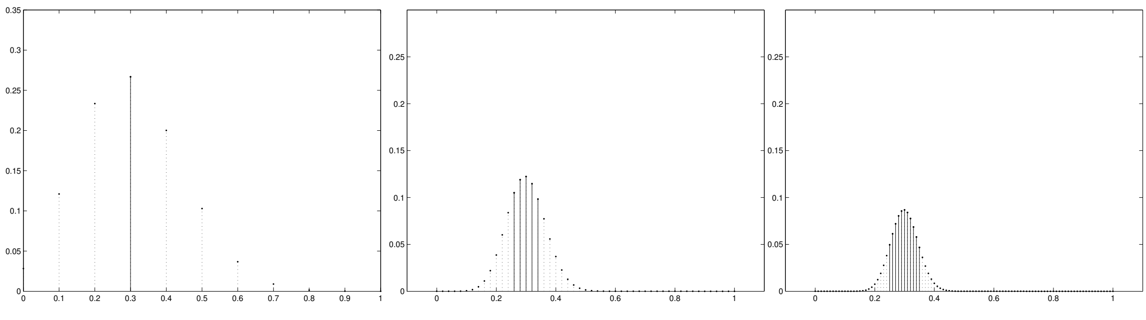
5.1 Loi des grands nombres

Soit A un événement qui peut être réalisé au cours d'une expérience aléatoire $\mathcal{E}$. On aimerait connaitre la probabilité p que l'événement A se réalise. Lorsqu'on effectue n fois cette expérience aléatoire, *dans des conditions identiques*, et de telle manière que les facteurs aléatoires présents lors des différentes expérimentations soient *indépendantes* (c'est-à-dire sans influence mutuelle), le nombre de fois où l'événement A a été réalisé au cours de ces n expériences, est une réalisation d'une variable aléatoire S_n de loi binomiale $B(n, p)$. La fréquence, $T_n = \frac{S_n}{n}$, avec laquelle cet événement se réalise est une variable aléatoire d'espérance p et de variance $\frac{p(1-p)}{n}$. Comme la variance de T_n tend vers 0, lorsque n est très grand, il y a de fortes chances que la valeur observée de T_n soit proche de p. Plus précisément, on a le résultat suivant :

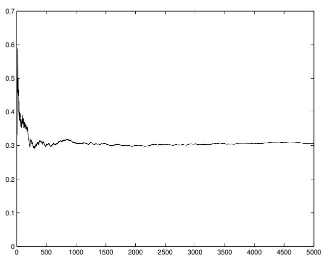
$$\text{pour tout } \epsilon > 0,\ P(|T_n - p| > \epsilon) \leq \frac{1}{4n\epsilon^2} \underset{n\to+\infty}{\to} 0.$$

Plus généralement, on a les deux résultats suivants :

Théorème 14 *Soit $X_1, \ldots, X_n, \ldots$ une suite de variables aléatoires définies sur un même espace de probabilité $(\Omega, \mathcal{A}, P)$, indépendantes et de même loi. On suppose que ces variables aléatoires admettent une espérance m finie.*
- Loi faible des grands nombres : *pour tout $\epsilon > 0$, $P(|\frac{1}{n}\sum_{i=1}^n X_i - m| > \epsilon) \underset{n\to+\infty}{\to} 0$.*
- Loi forte des grands nombres : *il existe un événement Λ de probabilité 1 tel que pour tout $\omega \in \Lambda$, $\frac{1}{n}\sum_{i=1}^n X_i(\omega) \underset{n\to+\infty}{\to} m$.*



De gauche à droite, représentation des lois de T_{10}, T_{50} et T_{100} lorsque $p = 0.3$. On a $P(|T_{10} - p| > 0.05) \simeq 0.77$, $P(|T_{50} - p| > 0.05) \simeq 0.44$ et $P(|T_{100} - p| > 0.05) \simeq 0.23$



Une réalisation de $n \mapsto T_n$ lorsque $p = 0.3$

N.B. Lorsqu'une propriété est vraie pour un ensemble de tirages ω de probabilité 1, on dira que la propriété est vraie *avec probabilité 1* ou pour *presque tout* ω ou encore *presque sûrement.*

Lorsque les variables aléatoires ont un moment d'ordre 2 fini, la loi faible des grands nombres est une conséquence directe de l'inégalité de Bienayme-Tchebychev :

Si X est une variable aléatoire ayant un moment d'ordre 2 fini, alors pour tout $\epsilon > 0$,

$$P(|X - E(X)| > \epsilon) \leq \frac{\operatorname{Var}(X)}{\epsilon^2}.$$

5.2 Loi empirique

Soit X est une variable aléatoire discrète à valeurs dans un ensemble $\mathcal{X} = \{e_1, \ldots, e_k\}$ (k fini ou $k = +\infty$) de loi μ.

5.2.1 Définition et propriétés

Soit $(X_1, \ldots, X_n)$ n variables aléatoires indépendantes de même loi μ. On dit $(X_1, \ldots, X_n)$ est un *n-échantillon de loi μ.*

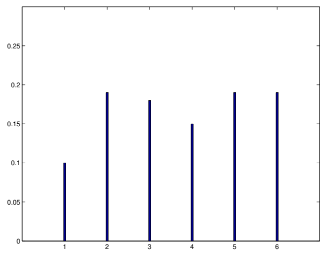
A partir d'une réalisation $(X_1(\omega), \ldots, X_n(\omega))$ de cet échantillon, on peut construire une nouvelle probabilité $\bar{\mu}_n(\omega)$ en définissant $\bar{\mu}_n(\omega, e_i)$ comme la fréquence d'apparition de la valeur e_i dans l'échantillon :

$$\bar{\mu}_n(\omega, e_i) = \frac{1}{n} \sum_{k=1}^{n} \mathbb{1}_{\{X_k(\omega) = e_i\}} \text{ pour tout } i \in \{1, \ldots, k\}.$$

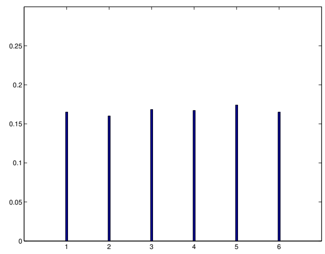
On appelle $\bar{\mu}_n(\omega)$ *la loi empirique de la réalisation* $(X_1(\omega), \ldots, X_n(\omega))$. Cela permet d'approcher la loi μ à l'aide des observations $(X_1(\omega), \ldots, X_n(\omega))$:

Propriété 15 *Pour presque tout $\omega \in \Omega$,* $\sup_{i \in \{1,\ldots,k\}} |\bar{\mu}_n(\omega, e_i) - \mu(e_i)| \underset{n \to +\infty}{\to} 0.$

On peut représenter une réalisation d'une telle fonction à l'aide d'un diagramme en bâtons, le bâton d'abscisse k ayant comme hauteur la proportion de valeurs de l'échantillon valant k.



Loi empirique des résultats de 100 lancers d'un dé



Loi empirique des résultats de 5000 lancers d'un dé

5.2.2 Exemples d'estimateurs empiriques

L'espérance de $\bar{\mu}_n(\omega)$ est la moyenne des valeurs $(X_1(\omega), \ldots, X_n(\omega))$ (appelée *la moyenne empirique* et notée $\bar{X}_n(\omega)$).

La variance de $\bar{\mu}_n(\omega)$ est $\bar{\sigma}_n^2(\omega) = \frac{1}{n}\sum_{i=1}^n (X_i(\omega) - \bar{X}_n(\omega))^2$. Elle est appelée *la variance empirique* de l'échantillon.

Par application de la loi forte des grands nombres, on obtient les propriétés suivantes :

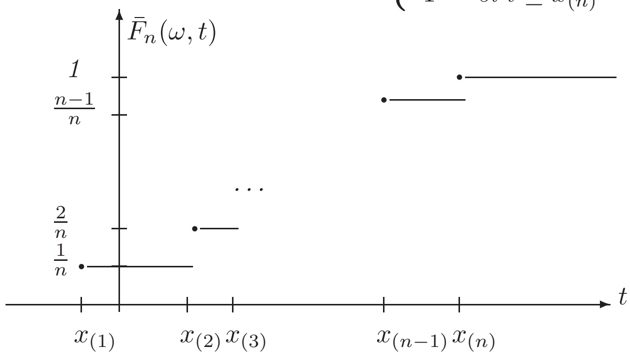
Propriétés 16
1. *Si $E(X) < +\infty$ alors pour presque tout $\omega \in \Omega$, $\bar{X}_n(\omega)$ tend vers $E(X)$ lorsque n tend vers $+\infty$.*
2. *Si $E(X^2) < +\infty$ alors pour presque tout $\omega \in \Omega$, $\bar{\sigma}_n^2(\omega)$ tend vers* $\operatorname{Var}(X)$ *lorsque n tend vers $+\infty$.*

5.3 Fonction de répartition empirique

La fonction de répartition de $\bar{\mu}_n(\omega)$ est la fonction sur $\mathbb{R}$ définie par : $\bar{F}_n(\omega) : t \mapsto \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{X_i(\omega)\leq t\}}$. Elle est appelée *la fonction de répartition empirique* de l'échantillon : $\bar{F}_n(\omega)(t)$ correspond simplement à la proportion de valeurs de l'échantillon qui sont inférieures ou égales à t.

Propriétés 17 – *Notons $x_{(1)}, \ldots, x_{(n)}$ la suite obtenue en ordonnant par ordre croissant les n observations $X_1(\omega), \ldots, X_n(\omega)$.*
La fonction de répartition empirique de l'échantillon s'écrit aussi :

$$\bar{F}_n(\omega)(t) = \begin{cases} 0 & \text{si } t < x_{(1)} \\ \frac{i}{n} & \text{si } x_{(i)} \leq t < x_{(i+1)} \qquad (1 \leq i \leq n-1) \\ 1 & \text{si } t \geq x_{(n)} \end{cases}$$



– *Soit F la fonction de répartition de X. Pour presque tout $\omega \in \Omega$,* $\sup_{t\in\mathbb{R}} |\bar{F}_n(\omega)(t) - F(t)| \underset{n\to+\infty}{\to} 0$ *(théorème de Glivenko-Cantelli)*

5.4 Exercices

▷ *Exercice 42.* On dispose d'un modèle d'infection de cellules par un virus. On effectue 100 simulations d'infection de cellules à l'aide de ce modèle avec les mêmes paramètres. Le tableau ci-dessous donne le nombre de simulations où k cellules sont infectées après une journée d'exposition :

Nb cellules infectées	10	13	14	15	16	17	18	19	20	21	25
Nb de simulations	1	2	2	5	10	10	25	23	12	8	2

Décrire la loi empirique du nombre de cellules infectées sur ces 100 simulations. Donner la valeur de l'estimateur empirique de l'espérance et de la variance de la loi du nombre de cellules infectées après une journée d'exposition.

Chez des moutons d'un même troupeau, on mesure l'importance d'une infection par des nématodes (vers parasites intestinaux) en comptant le nombre d'oeufs présents dans les fèces de l'animal au mois de juin. On considère que l'importance de l'infection résulte de deux effets additifs et indépendants que l'on ne peut pas distinguer à la mesure : la résistance lié aux caractéristiques génétiques de chaque animal et la variation naturelle dues aux conditions climatiques. Pour estimer la part de la variance expliquée par les caractéristiques génétiques d'un troupeau, on fait une mesure du nombres d'oeufs pour chaque animal du troupeau deux années successives. L'écart-type des mesures pour le troupeau est de 250 oeufs/gr, le coefficient de corrélation entre les deux années est de 0.4. On en déduit que 40 % de la variance observée est expliquée par les caractéristiques génétiques d'un troupeau.
Proposer une modélisation qui permet d'arriver au résultat annoncé en explicitant bien les hypothèses émises et les approximations faites.

▷ *Exercice 43.* On a répété une expérience aléatoire dans les mêmes conditions 100 fois. Le résultat d'une expérience est décrite par une variable aléatoire X à valeurs dans $\{-2, -1, 0, 1\}$. On dispose donc de 100 réalisations de la variable aléatoire X.
Le tableau suivant donne le nombre de fois où chaque valeur a été observée au cours de ces 100 expériences :

Valeurs	-2	-1	0	1
Effectifs observés	19	46	14	21

Déterminer la valeur observée de

1. l'estimateur empirique de $P(X \leq -0.5)$.
2. l'estimateur empirique de l'espérance de X.

▷ *Exercice 44.* Soit X une variable aléatoire dont la loi est définie par le tableau suivant :

k	-2	-1	0	1	2
$P(X = k)$	0.18	0.18	0.17	0.22	0.25

Soit $(X_n)_n$ une suite de variables aléatoires indépendantes et de même loi que X.

1. Lorsque n tend vers $+\infty$, la variable aléatoire $\frac{1}{n}\sum_{i=1}^{n}(X_i)^2$ converge avec probabilité 1 vers une constante. Quelle est la valeur de cette constante ?
2. Lorsque n tend vers $+\infty$, la variable aléatoire $\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}_{\{X_i\geq 1\}}$ converge avec probabilité 1 vers une constante. Quelle est la valeur de cette constante ?

$\triangleright$ *Exercice 45.* Soit X et Y deux variables aléatoires à valeurs dans $\{0,...,100\}$. On répète une expérience n fois dans les mêmes conditions de sorte d'obtenir n réalisations $(X_1(\omega),Y_1(\omega)),\dots,(X_n(\omega),Y_n(\omega))$ du vecteur aléatoire $Z=(X,Y)$. Que peut-on dire du comportement des suites suivantes lorsque n tend vers $+\infty$?

$$(\frac{1}{n}\sum_{i=1}^{n}\mathbb{1}_{\{X_i(\omega)+Y_i(\omega)\leq 2\}})_n \text{ et } (\frac{1}{n}\sum_{i=1}^{n}X_i(\omega)Y_i(\omega))_n$$

$\triangleright$ *Exercice 46.* Soit X et Y deux variables aléatoires à valeurs dans $\{0,...,50\}$. On répète une expérience n fois dans les mêmes conditions de sorte d'obtenir n réalisations $(X_1(\omega),Y_1(\omega)),\dots,(X_n(\omega),Y_n(\omega))$ du vecteur aléatoire $Z=(X,Y)$ ($Z_1=(X_1,Y_1)$, ..., $Z_n=(X_n,Y_n)$ définit un n-échantillon de Z). On pose $\bar{X}_n=\frac{1}{n}\sum_{i=1}^{n}X_i$ et $\bar{Y}_n=\frac{1}{n}\sum_{i=1}^{n}Y_i$ et

$$c_n=\frac{1}{n}\sum_{i=1}^{n}(X_i-\bar{X}_n)(Y_i-\bar{Y}_n).$$

Montrer que $c_n=\frac{1}{n}\sum_{i=1}^{n}X_iY_i-\bar{X}_n\bar{Y}_n$. En déduire que c_n converge avec probabilité 1 vers $\mathrm{Cov}(X,Y)$.
Application Le tableau suivant donne la largeur et la longueur (en cm) des pétales de 20 fleurs d'une espèce d'iris :

Largeur	1.4	1.5	1.5	1.3	1.5	1.3	1.6	1.0	1.3	1.4
Longueur	4.7	4.5	4.9	4.0	4.6	4.5	4.7	3.3	4.6	3.9
Largeur	1.0	1.5	1.0	1.4	1.3	1.4	1.5	1.0	1.5	1.1
Longueur	3.5	4.2	4.0	4.7	3.6	4.4	4.5	4.1	4.5	3.9

Donner une estimation de la covariance entre la longueur et la largeur des pétales de cet espèce d'iris basée sur ces données.

6 Simuler une expérience aléatoire sur ordinateur

6.1 Nombres pseudo-aléatoires

Les logiciels de calcul dispose d'un algorithme déterministe que nous appellerons `rand` qui fournit une liste de nombres $(u_n)_n$ dans $]0,1[$ tel que $(u_1,\dots,u_n)$ ressemble à n nombres « tirés aléatoirement entre 0 et 1 » pour tout n. Cette liste est obtenue par un procédé le plus souvent récursif de la forme $u_{n+1}=f(u_n,\dots,u_{n-k})$ pour un certain k fixé. Un tel procédé est appelé « générateur de nombres pseudo-aléatoires ». Ce générateur a besoin d'un vecteur $(u_0,\dots,u_k)$ pour l'initialisation de la récurrence. On appelle ce vecteur la graine (« seed » en anglais) du générateur.

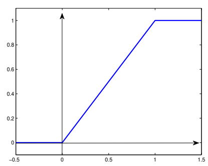
6.1.1 Loi d'un nombre tiré au hasard entre 0 et 1

Le résultat du tirage au hasard d'un nombre entre 0 et 1 est une variable aléatoire U qui doit avoir les propriétés suivantes :

- la valeur de U peut être n'importe quel nombre entre 0 et 1 (ce n'est donc plus une variable aléatoire discrète, sa loi ne peut plus être définie par les coefficients $P(U=x)$ pour $x\in[0,1]$),
- la chance que U soit compris entre deux nombres $0\leq a<b\leq 1$ doit être égale à $b-a$: pour tout $0\leq a<b\leq 1$, $P(a\leq U\leq b)=b-a$.

En particulier,
- $P(U=x)=0$ pour tout $x\in[0,1]$;
- La fonction de répartition de U, $F:t\mapsto P(U\leq t)$, est définie par :

$$F(t)=\begin{cases}0 & \text{si } t<0\\ t & \text{si } t\in[0,1]\\ 1 & \text{si } t>1\end{cases}$$



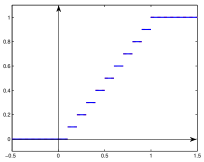
Graphe de F sur l'intervalle $[-0.5,1.5]$

On dit que U suit la *loi uniforme sur l'intervalle* $[0,1]$.
Comme $P(a\leq U\leq b)$ peut aussi s'écrire comme $\int_a^b f(x)dx$ avec f l'indicatrice de l'intervalle $[0,1]$ c'est-a-dire : $f(x)=\begin{cases}1 & \text{si } x\in[0,1]\\ 0 & \text{si } x\notin[0,1]\end{cases}$.
On dit que la loi de U admet pour densité la fonction $f=\mathbb{1}_{[0,1]}$.

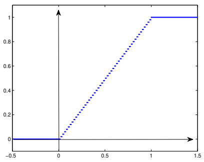
Proposition 18 *La loi uniforme sur $[0,1]$ est "la limite" de la loi uniforme sur l'ensemble $I_n = \{\frac{1}{n}, \frac{2}{n}, \ldots, \frac{n-1}{n}, 1\}$ lorsque n tend vers $+\infty$ au sens où la fonction de répartition de la loi uniforme sur I_n converge en tout point vers F.*

Preuve. Pour tout entier n, considérons la loi uniforme sur l'ensemble I_n. Sa fonction de répartition est la fonction F_n définie par :

$$F_n(t) = \begin{cases} 0 & \text{si } t < \frac{1}{n} \\ \frac{i}{n} & \text{si } \frac{i}{n} \leq t < \frac{i+1}{n} \text{ pour } i \in \{1, \ldots, n-1\} \\ 1 & \text{si } t \geq 1 \end{cases}$$



Graphe de F_{10} sur l'intervalle $[-0.5, 1.5]$



Graphe de F_{100} sur l'intervalle $[-0.5, 1.5]$

Lorsque n tend vers l'infini, $F_n(t)$ tend vers $F(t)$ en tout point t : en effet, les deux fonctions coïncident déjà en tout point $t \notin]0,1[$. Si $t \in]0,1[$ alors $t - \frac{1}{n} \leq F_n(t) \leq t$ et donc $|F_n(t) - F(t)| \leq \frac{1}{n}$ tend vers 0 lorsque n tend vers $+\infty$. □

6.1.2 Hypothèses

Nous supposerons que nous disposons d'une fonction `rand` idéale c'est-à-dire qui vérifie les deux propriétés suivantes :

(i) un appel à la fonction `rand` donne une réalisation d'une variable aléatoire de loi uniforme sur $[0,1]$.

(ii) les appels successifs à la fonction `rand` fournissent une réalisation d'une suite de variables aléatoires indépendantes de loi uniforme sur $[0,1]$.

6.2 Simulation d'une variable aléatoire discrète

Pour simuler une expérience aléatoire sur ordinateur, il faut être capable de générer des réalisations d'une variable aléatoire. Par exemple, simuler le lancer d'un dé, revient à générer une réalisation d'une variable aléatoire de loi uniforme sur $\{1, \ldots, 6\}$.

On peut, à partir d'une variable aléatoire U de loi uniforme sur $[0,1]$, construire des variables aléatoires discrètes de n'importe quelle loi (et même n'importe quelle variable aléatoire réelle). Il suffit pour cela d'appliquer à U une transformation f bien choisie.

▶ *Exemple 34.* Soit $p \in]0,1[$. Si U suit la loi uniforme sur $[0,1]$ alors la variable aléatoire $Y = \mathbb{1}_{\{X \leq p\}}$ suit la loi de Bernoulli $\mathcal{B}(p)$.

▶ *Exemple 35.* Si U suit la loi uniforme sur $[0,1]$ alors Y =Ent$(6U)$+1 est une variable aléatoire de loi uniforme sur $\{1, \ldots, 6\}$ ($Ent(a)$ désigne le plus grand entier inférieur ou égal à a).

De façon générale, on peut construire une variable aléatoire X qui prend un nombre fini ou dénombrable de valeurs $x_1, \ldots, x_n$ avec probabilité respectivement $p_1, \ldots, p_n$ en posant :

$$X = \begin{cases} x_1 & \text{si } 0 \leq U \leq p_1 \\ x_2 & \text{si } p_1 < U \leq p_1 + p_2 \\ \ldots & \\ x_i & \text{si } p_1 + \ldots + p_{i-1} < U \leq p_1 + \ldots + p_i \\ \ldots & \\ x_n & \text{si } p_1 + \ldots + p_{n-1} < U \leq 1 \end{cases}$$

Algorithme de simulation d'une réalisation de la variable aléatoire X :

```
u ← rand
i ← 1
q ← p1
Tant que u > q
   i ← i + 1
   q ← q + pi
FinTantque
X ← xi
```

En effet, $P(X_1 = x_1) = P(U \leq p_1) = p_1$ et $P(X = x_i) = P(p_1 + \ldots + p_{i-1} < U \leq p_1 + \ldots + p_i) = p_i$ pour tout $i \in \{2, \ldots, n\}$.

N.B. Pour une variable aléatoire réelle quelconque X, on a le résultat suivant :

Soit F la fonction de répartition de X, notons F^{-1} la *pseudo-inverse* de F (appelée aussi *fonction quantile*) : $F^{-1}(x) = \inf\{t,\ F(t) \geq x\}$ pour tout $x \in]0,1[$.

Si U est une variable aléatoire de loi uniforme sur $[0,1]$ alors $F^{-1}(U)$ a même loi que X.

▷ *Exercice 47.* On considère une variable aléatoire U de loi uniforme sur l'intervalle $[0, 1]$. Déterminer la loi de la variable aléatoire $X = -5 + 2\mathbb{1}_{\{U \leq 0.25\}} - 2\mathbb{1}_{\{U > 0.72\}}$.

Troisième partie

Introduction aux chaînes de Markov

Lorsque deux systèmes identiques à un instant donné peuvent avoir des comportements différents dans le futur, on est amené à introduire une suite de variables aléatoires $(X_t)_t$ pour décrire leurs évolutions : X_t servant à définir l'état du système étudié à l'instant t. Si le système étudié est une population, l'état du système à un instant donné peut être décrit simplement par un nombre lorsqu'on s'intéresse uniquement à la taille de cette population, ou par un ensemble de nombres tels que l'ensemble des positions de chaque individu lorsqu'on s'intéresse à la répartition spatiale de la population.

On va ici se limiter à des systèmes dont l'état peut être décrit par une variable aléatoire ou un vecteur aléatoire discret.

En général l'évolution futur d'un système dépendant au moins de son état présent, les variables aléatoires décrivant l'état du système à chaque instant ne pourront pas être considérées comme indépendantes. On va s'intéresser aux situations où l'évolution future d'un système ne dépend du passé qu'au travers de son état présent et pour simplifier on n'étudiera pas l'évolution du système en temps continu, mais son évolution en une suite infinie d'instants $0 = t_0 < t_1 < \ldots < t_n < ....$ On travaillera donc avec une suite de variables aléatoires discrètes $(X_n)_{n\in\mathbb{N}}$, chaque variable aléatoire étant à valeurs dans un ensemble fini ou infini dénombrable noté $\mathcal{X}$ que l'on identifiera à $\{1, \ldots, N\}$ si $\mathcal{X}$ est composé de N éléments et à $\mathbb{N}^*$ si $\mathcal{X}$ est infini.

7 Généralités

Soit $X_0, X_1, \ldots, X_n, \ldots$ une suite de variables aléatoires définies sur un même espace de probabilité $(\Omega, \mathcal{A}, P)$ et à valeurs dans $\mathcal{X}$.

7.1 Définitions et exemples

Définition. La suite $(X_n)_{n\in\mathbb{N}}$ est une *chaîne de Markov* si pour tout $n \in \mathbb{N}$ et pour tout $(i_1, \ldots, i_{n+1}) \in \mathcal{X}^{n+1}$ tel que $P(X_n = i_n, X_{n-1} = i_{n-1}, \ldots, X_0 = i_0) > 0$,

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

ne dépend que des valeurs de n, i_n et i_{n+1}. L'ensemble $\mathcal{X}$ est appelé *l'espace des états* de la chaîne de Markov $(X_n)_n$.

Proposition 19 *Soit* $(X_n)_{n\in\mathbb{N}}$ *une* chaîne de Markov *d'espace d'états* $\mathcal{X}$.
Pour tout $n \in \mathbb{N}$ *et pour tout* $(i_0, \ldots, i_{n+1}) \in \mathcal{X}^{n+1}$ *tel que* $P(X_n = i_n, X_{n-1} = i_{n-1}, \ldots, X_0 = i_0) > 0$, *on a*

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

Preuve. Soit $n \in \mathbb{N}$, i et j deux états possibles. Posons $Z = (X_0, \ldots, X_{n-1})$ le vecteur décrivant les positions successives du système jusqu'à l'instant $n - 1$ (Z est à valeurs dans $\mathcal{X}^n$). Par définition, pour tout état $h \in \mathcal{X}^n$, $P(X_{n+1} = j|X_n = i, Z = h)$ ne dépend que de n, i et j. Notons cette probabilité conditionnelle $Q_n(i, j)$. Montrons que $P(X_{n+1} = j|X_n = i) = Q_n(i, j)$.

$$\begin{aligned} P(X_{n+1} = j|X_n = i) &= \frac{P(X_{n+1} = j, X_n = i)}{P(X_n = i)} = \sum_{h\in\mathcal{X}^n} \frac{P(X_{n+1} = j, X_n = i, Z = h)}{P(X_n = i)} \\ &= \sum_{h\in\mathcal{X}^n} \frac{Q_n(i, j)P(X_n = i, Z = h)}{P(X_n = i)} \\ &= Q_n(i, j). \end{aligned}$$

□

N.B. Dire qu'une suite $(X_n)_n$ est une chaîne de Markov signifie que pour tout $n \in \mathbb{N}$, X_{n+1} est indépendant du vecteur aléatoire $(X_0, \ldots, X_{n-1})$ conditionnellement à la variable aléatoire X_n.
Les états successifs d'un système peuvent donc être décrits par une chaîne de Markov si la connaissance de

l'état du système à l'instant présent apporte autant d'information sur le futur que la connaissance de tout le passé. Les chaînes de Markov sont des processus sans mémoire.

► *Exemple 36.* (*Fortune d'un joueur*) On considère un joueur qui dispose initialement de s euros ($s \in \mathbb{N}$). Il joue à un jeu de hasard, dont l'enjeu à chaque partie est de 1 euro, jusqu'à ce qu'il soit en possession de m euros ($m \in \mathbb{N}$ tel que $m \geq s$) ou jusqu'à ce qu'il ait dépensé tout son argent. On suppose qu'à chaque partie, il a une probabilité p de gagner et $(1-p)$ de perdre et ceci indépendamment des résultats des autres parties ($p \in]0,1[$). La fortune du joueur après la n-ième partie est décrite par une variable aléatoire notée X_n. La suite $(X_n)_n$ est une chaîne de Markov.

► *Exemple 37.* On considère une population constituée de cellules de type A et de cellules de type B. Entre deux instants successifs t_n et $t_{n+1} = t_n + \Delta t$, on suppose qu'exactement une cellule choisie au hasard dans la population se divise donnant deux cellules-filles identiques à la cellule-mère. Le nombre de cellules de type A dans la population à l'instant t_n est une variable aléatoire X_n pour tout n et la suite (X_n) ainsi définie est une chaîne de Markov.

► *Exemple 38.* *(Autofécondation)* 0n dispose d'une plante que l'on croise avec elle-même (génération 0). On choisit une plante au hasard parmi les plantes obtenues par cette autofécondation (génération 1). On répète le processus. On s'intéresse à un gène qui a deux allèles notées A et a. On note X_n le génotype pour ce gène de la plante choisie parmi les plantes de la génération n.
Alors $(X_n)_n$ définit une chaîne de Markov à trois états AA, Aa et aa.

Définition. Soit $(X_n)_{n\in\mathbb{N}}$ une chaîne de Markov à valeurs dans $\mathcal{X}$.
La chaîne de Markov est dite *homogène* si pour tout $(i,j) \in \mathcal{X}^2$ et tout $n \in \mathbb{N}$, la probabilité de passer à l'état j à l'instant $n+1$ sachant qu'on était à l'état i à l'instant n ne dépend pas de n :

$$P(X_{n+1} = j | X_n = i) = Q(i,j) \text{ pour tout } n \in \mathbb{N}.$$

Les nombres $Q(i,j),\ (i,j) \in \mathcal{X}^2$ s'appellent les *probabilités de transition* de la chaîne de Markov.

N.B. Lorsque $\mathcal{X} = \{1,\ldots,N\}$, la matrice de transition Q est représentée par un tableau à N lignes et N colonnes appelée *matrice de transition* de la chaîne de Markov $(X_n)_n$:

$$Q = \begin{pmatrix} Q(1,1) & \cdots & Q(1,N) \\ \vdots & & \vdots \\ Q(N,1) & \cdots & Q(N,N) \end{pmatrix}$$

Que $\mathcal{X}$ soit un ensemble fini ou dénombrable, on appellera $Q = (Q(i,j))_{(i,j)\in\mathcal{X}^2}$ la *matrice de transition* de la chaîne de Markov homogène $(X_n)_n$.

N.B. La loi de X_0 est appelée *la loi initiale de la chaîne de Markov.* On l'écrira

$$\pi_0 = (P(X_0 = 1), P(X_0 = 2), \ldots, P(X_0 = N-1), P(X_0 = N)) \text{ si } \mathcal{X} = \{1,\ldots,N\}.$$

Comme la i-ième ligne de la matrice de transition d'une chaîne de Markov homogène $(X_n)_n$ définit la loi conditionnelle de X_n sachant que $\{X_{n-1} = i\}$, on a la propriété suivante :

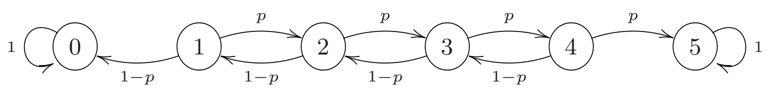
Propriété 20 *Soit $(X_n)_{n\in\mathbb{N}}$ une chaîne de Markov homogène à valeurs dans $\mathcal{X}$ de matrice de transition Q. La matrice Q est une* matrice stochastique*, c'est-à-dire : pour tout $i,j \in \mathcal{X}$, $Q(i,j) \geq 0$ et pour tout $i \in \mathcal{X}$, $\sum_{j\in\mathcal{X}} Q(i,j) = 1$.*

► *Exemple 39.* Dans les exemples 36 et 38, les chaînes de Markov $(X_n)_n$ sont homogènes. Par contre, dans l'exemple 37, la suite $(X_n)_n$ est une chaîne de Markov non homogène.

7.2 Graphe associé à une matrice de transition

Pour visualiser l'évolution d'une chaîne de Markov homogène, il est souvent utile de représenter la matrice de transition Q de la chaîne de Markov par un graphe orienté : les noeuds du graphe sont les états possibles pour la chaîne de Markov, une flèche allant de l'état i à l'état j indique qu'il y a une probabilité strictement positive que le prochain état de la chaîne soit l'état j si elle est actuellement dans l'état i. On met le poids $Q(i,j)$ à la flêche allant de l'état i à l'état j.

▶ *Exemple 40.* Lorsque $m = 5$, le graphe décrivant l'évolution de la fortune du joueur dans l'exemple 36 est :



▷ *Exercice 48.* Dans l'exemple 38, déterminer la matrice de transition et dessiner le graphe associé à cette matrice de transition.

Définition. Une suite d'états $(x_1, x_2, \ldots, x_m)$ définit un *chemin de longueur m allant de x_1 à x_m* dans le graphe associé à la chaîne de Markov homogène si et seulement si

$$Q(x_1, x_2)Q(x_3, x_4)\ldots Q(x_{m-1}, x_m) > 0.$$

Grâce au graphe associé à Q, on peut voir la suite $(X_n)_n$ comme marquant les positions successives d'un pion que l'on déplace sur les noeuds du graphe : si le pion est sur le noeud i, on choisit de le déplacer au noeud j avec probabilité $Q(i, j)$ et ceci indépendamment de la trajectoire passée.
On peut directement lire sur le graphe la probabilité que le pion emprunte un chemin fixé :

Proposition 21 *Soit $i_0, \ldots, i_n \in \mathcal{X}$. Pour tout $k \in \mathbb{N}$.*

$$P(X_{n+k} = i_n, \ldots, X_{k+1} = i_1 | X_k = i_0) = Q(i_0, i_1)Q(i_1, i_2)Q(i_2, i_3)\cdots Q(i_{n-1}, i_n)$$

Preuve. Posons $A_j = \{X_{k+j} = i_j\}$ pour tout $j \in \{0, \ldots, n\}$. On a :

$$P(A_n \cap \ldots \cap A_1 | A_0) = P(A_n | A_{n-1} \cap \ldots \cap A_0)P(A_{n-1} | A_{n-2} \cap \ldots \cap A_0)\ldots P(A_1 | A_0)$$

Comme $(X_n)_n$ est une chaîne de Markov homogène de matrice de transition Q, $P(A_j | A_{j-1} \cap \ldots \cap A_0) = Q(i_{j-1}, i_j)$ pour tout $j \in \{1, \ldots, n\}$, ce qui permet de conclure. □

7.3 Caractérisations d'une chaîne de Markov homogène

Proposition 22 *La suite $(X_n)_{n\in\mathbb{N}}$ est une chaîne de Markov homogène si et seulement si il existe une matrice Q ayant la propriété suivante : pour tout $n \in \mathbb{N}$ et pour tout $i_0, \ldots, i_n \in \mathcal{X}$,*

$$P(X_n = i_n,\ X_{n-1} = i_{n-1}, \ldots, X_0 = i_0) = P(X_0 = i_0)Q(i_0,\ i_1)\ldots Q(i_{n-1}, i_n).$$

Dans ce cas, Q est la matrice de transition de la chaîne de Markov homogène $(X_n)_{n\in\mathbb{N}}$.

Preuve.

- Supposons que $(X_n)_n$ soit une chaîne de Markov homogène. Notons Q sa matrice de transition.
  D'après la proposition 21 avec $k = 0$,

$$\begin{aligned} P(X_n = i_n,\ X_{n-1} = i_{n-1}, \ldots, X_0 = i_0) &= P(X_n = i_n,\ X_{n-1} = i_{n-1}, \ldots, X_1 = i_1 | X_0 = i_0)P(X_0 = i_0) \\ &= P(X_0 = i_0)Q(i_0,\ i_1)\ldots Q(i_{n-1}, i_n). \end{aligned}$$

- Réciproquement, supposons que pour tout $n \in \mathbb{N}$ et pour tout $i_0, \ldots, i_n \in \mathcal{X}$,

$$P(X_n = i_n,\ X_{n-1} = i_{n-1}, \ldots, X_0 = i_0) = P(X_0 = i_0)Q(i_0,\ i_1)\ldots Q(i_{n-1}, i_n).$$

  Pour tout $n \in \mathbb{N}$ et pour tout $i_0, \ldots, i_{n+1} \in \mathcal{X}$ tels que $P(X_n = i_n,\ X_{n-1} = i_{n-1}, \ldots, X_0 = i_0) > 0$,

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

  ce qui montre que $(X_n)_n$ est une chaîne de Markov homogène de matrice de transition Q.

□

Proposition 23 *Soit $\mathcal{X}$ et $\mathcal{Y}$ deux ensembles finis ou dénombrables. Soit $(Y_n)_n$ une suite de variables aléatoires indépendantes et de même loi, à valeurs dans l'ensemble $\mathcal{Y}$.*
Soit f une fonction définie sur $\mathcal{X} \times \mathcal{Y}$ à valeurs dans $\mathcal{X}$.
On définit une suite de variables aléatoires $(X_n)_{n\in\mathbb{N}}$ à valeurs dans $\mathcal{X}$ par :

- *X_0 est une variable aléatoire indépendante de la suite $(Y_n)_{n\in\mathbb{N}^*}$ (c'est-à-dire que pour tout n, $X_0, Y_1, \ldots, Y_n$ sont des variables aléatoires indépendantes).*

– $X_{n+1} = f(X_n, Y_{n+1})$ *pour tout* $n \in \mathbb{N}$

Alors, $(X_n)_{n\in\mathbb{N}}$ *est une chaîne de Markov homogène.*

Preuve. Soit $n \in \mathbb{N}$ et $i_0, \dots, i_{n+1} \in \mathcal{X}$.

$$P(X_n = i_n, \dots, X_0 = i_0) = P(i_n = f(i_{n-1}, Y_n), \dots, i_1 = f(i_0, Y_1), X_0 = i_0).$$

Comme $X_0, Y_1, \dots, Y_n$ sont des variables aléatoires indépendantes et comme les variables aléatoires $Y_1, \dots, Y_n$ ont même loi, $P(X_n = i_n, \dots, X_0 = i_0) = P(i_n = f(i_{n-1}, Y_1)) \cdots P(i_1 = f(i_0, Y_1))P(X_0 = i_0)$. Donc, d'après la proposition précédente, $(X_n)_n$ est une chaîne de Markov homogène de matrice de transition Q définie par $Q(i,j) = P(j = f(i, Y_1))$ pour tout $i, j \in \mathcal{X}$. □

► *Exemple 41.* Dans l'exemple 36, la fortune du joueur X_n après la n-ième partie, vérifie la relation de récurrence : $X_n = X_{n-1} + U_n \mathbb{1}_{\{X_n \in \{1,\dots,m-1\}\}}$ où U_n désigne la variable aléatoire qui vaut 1 si le joueur gagne la n-ième partie et qui vaut -1 si le joueur perd la n-ième partie. Comme le joueur joue à un jeu de hasard, la suite (U_n) est une suite de variables aléatoires indépendantes et de même loi : $P(U_n = 1) = p = 1 - P(U_n = -1)$ et cette suite est indépendante de X_0.

7.4 Simulation des premiers états d'une chaîne de Markov homogène

Souvent la description de l'évolution d'un système qui pourra être modélisée par une chaine de Markov homogène permet de trouver une relation de récurrence comme celle donnée dans la proposition 23 : $X_{n+1} = f(X_n, Y_{n+1})$ avec $(Y_n)_n$ une suite de variables aléatoires indépendantes et de même loi qui est indépendante de X_0.

On dispose alors d'une méthode simple pour simuler les états successifs de ce système.

Algorithme de simulation d'une réalisation de $(X_0, \dots, X_n)$ **:**

`Simuler une réalisation` x_0 `de` X_0 `suivant la loi initiale de la chaine de Markov.`

`Pour` i `allant de` 1 `à` n `:`

  `Simuler une réalisation` y_i `d'une v.a. de même loi que` Y_i `;`

  $x_i \leftarrow f(x_{i-1}, y_i)$ `;`

`FinDeLaBoucle`

`Retourner le vecteur` $(x_0, \dots, x_n)$`.`

Lorsque l'on cherche à simuler les premiers états d'une chaîne de Markov homogène (X_n) d'espace d'états finis $\mathcal{X} = \{1, \dots, N\}$ décrite uniquement par sa loi initiale et sa matrice de transition Q on peut utiliser l'algorithme suivant qui repose sur la proposition 22 :

Algorithme de simulation d'une réalisation de $(X_0, \dots, X_n)$ **si** $(X_n)_n$ **est une chaîne de Markov homogène de matrice de transition** Q **:**

`Simuler une réalisation` x_0 `de` X_0 `suivant la loi initiale de la chaîne de Markov ;`

`Pour` k `allant de` 1 `à` n `:`

  `Simuler une réalisation` x_k `d'une v.a. dont la loi est égale à la loi conditionnelle de` X_k `sachant que` $X_{k-1} = x_{k-1}$ `(cette loi` $\mathcal{X}$ `décrite est décrite par le vecteur ligne` $(Q(x_{k-1}, 1), \dots, Q(x_{k-1}, N))$`.`

`FinDeLaBoucle`

`Retourner le vecteur` $(x_0, \dots, x_n)$`.`

N.B. Rappelons que pour tout $k \in \mathcal{X}$, la loi conditionnelle de X_i sachant que $X_{i-1} = k$ est une loi sur $\mathcal{X}$ dont les coefficients sont décrits par la k-ième ligne de la matrice de transition Q :

$$Q(k, \ell) = P(X_i = \ell | X_{i-1} = k).$$

Avec le logiciel **R**, une réalisation d'une variable aléatoire Z à valeurs dans $\mathcal{X} = \{1, \dots, N\}$ peut être simulée en utilisant la fonction `sample( x = 1 :N, size=1,replace = TRUE, prob = p )` où `p` désigne un vecteur de taille N tel que `p[i]`$= P(Z = i)$ pour tout $i \in \mathcal{X}$.

7.5 Exemples de chaînes de Markov

▷ *Exercice 49.* On constitue une séquence de lettres prises dans l'alphabet fini $\mathcal{A} = \{a, c, g, t\}$ en tirant au hasard chaque lettre parmi $\mathcal{A}$. On définit une suite de variables aléatoires $(X_n)_n$ en posant : $X_0 = 0$, et pour $n \in \mathbb{N}^*$,
- $X_n = 0$ si la n-ième lettre de la séquence n'est pas la lettre a.
- $X_n = k$ si la n-ième lettre de la séquence est a et si elle constitue la k-ième occurence d'une suite successive de a, pour $k \in \mathbb{N}^*$.

1. Quelles sont les valeurs de X_1, X_2 et X_3 si les 3 premières lettres tirées sont dans l'ordre a, a, c ?
2. Pour tout entier n, exprimer X_{n+1} en fonction de X_n et de la $(n+1)$-ième lettre tirée.
3. Expliquer comment faire pour simuler $X_1, \ldots, X_n$.
4. Montrer que $(X_n)_n$ constitue une chaîne de Markov homogène sur $\mathbb{N}$. Déterminer sa matrice de transition et dessiner le graphe associé.

▷ *Exercice 50.* Un patient qui arrive dans un cabinet médical est dirigé vers une salle d'attente. S'il y a déjà 3 personnes qui attendent, le patient découragé repart. Un médecin est présent en permanence dans ce cabinet. Il vient toutes les 20 mn dans la salle d'attente pour voir s'il y a des patients en attente. Si c'est le cas, il prend en consultation l'un des patients, sinon il revient 20 mn plus tard. On supposera qu'une consultation ne dure pas plus de 20 mn. On discrétise le temps en intervalles de temps (t_n, t_{n+1}) de durée 20 mn. On modélise le nombre de personnes qui arrivent à ce cabinet médical pendant les intervalles de temps successifs (t_0, t_1), (t_1, t_2), (t_2, t_3),... par des variables aléatoires indépendantes et de même loi A_1, A_2, A_3,.... La loi de ces variables aléatoires est décrite par le tableau ci-dessous :

i	0	1	2	3
$P(A_1 = i)$	0.1	0.4	0.3	0.2

On note X_0 le nombre de personnes dans la salle d'attente à l'instant t_0 et pour $n \in \mathbb{N}^*$, on note X_n le nombre de personnes qui sont dans la salle d'attente lorsque le médecin arrive à l'instant t_n dans la salle d'attente.

1. Si au moment où le medecin arrive dans la salle d'attente il y a deux personnes dans la salle d'attente, quelle est la probabilité pour qu'il y en ait trois lorsqu'il revient 20 mn plus tard ?
2. Exprimer X_{n+1} en fonction de X_n et de A_{n+1}.
3. Montrer que $(X_n)_n$ est une chaîne de Markov homogène et déterminer sa matrice de transition.

▷ *Exercice 51.* On croise deux souris, puis on choisit au hasard deux souris de sexes opposés dans leurs descendants directs. On croise alors les souris sélectionnées, puis on recommence. On s'intéresse à un gène qui peut s'exprimer sous deux formes A et a. Il y a alors 6 génotypes pour le couple de souris sélectionné à la n-ième génération $E_1 = AA \times AA$, $E_2 = AA \times Aa$, $E_3 = Aa \times Aa$, $E_4 = Aa \times aa$, $E_5 = aa \times aa$ et $E_6 = AA \times aa$.

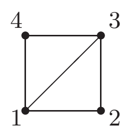
1. Expliquer comment simuler la suite des génotypes des couples de souris sélectionnés.
2. Montrer que la suite des génotypes des couples de souris sélectionnés constitue une chaîne de Markov.
3. Déterminer sa matrice de transition.

▷ *Exercice 52. Modèle simple pour une séquence d'ADN dont le premier nucléotide est a* : pour tout $n \geq 1$, on note X_n la n-ième base composant une séquence d'ADN en partant d'une de ses extrêmités. On suppose que (X_n) est une chaîne de Markov homogène d'espace d'états $\mathcal{A} = \{a, c, g, t\}$ et d'état initial a (on identifiera a à l'état 1, c à l'état 2, g à l'état 3 et t à l'état 4). On note Q sa matrice de transition.

1. Exprimer à l'aide de la matrice Q, la probabilité que la séquence commence par le motif $aacg$.
2. Montrer que la 4-ième base est indépendante de la seconde sachant que la 3-ième base est a.
3. Expliquer comment simuler les n premières bases d'une séquence pour ce modèle.

▷ *Exercice 53.*

Une fourmi se déplace le long des arêtes du dessin ci-contre de la façon suivante : arrivée à un sommet, elle choisit au hasard une arête partant de ce sommet et la parcourt jusqu'à atteindre un autre sommet.



Montrer que la suite des sommets visités par la fourmi est une chaîne de Markov homogène et donner la matrice de transition de cette chaîne.
On observe à un instant donné, que la fourmi se trouve au sommet 4. Quelle est la probabilité pour que la fourmi se retrouve au sommet 4 après avoir parcouru 3 arêtes ?

▷ *Exercice 54.*

1	2	3	4	5	6	7	8	9

Le tableau d'un jeu de l'oie simple est constitué de 9 cases numérotées de 1 à 9. La case 1 est la case de départ et la case 9 est la case d'arrivée. Pour faire avancer le pion, on lance un dé à 6 faces numérotés de 1 à 6 et on avance le pion d'un nombre de cases égal au nombre obtenu avec le dé. Le jeu s'arrête lorsque le pion tombe exactement sur la case 9. Sinon le pion recule.

Par exemple, si le pion se trouve sur la case 8 et si le dé tombe sur 3, le pion va à la case 7. Si au coup suivant, le dé tombe sur 1, le pion retourne sur la case 8. On supposera que lorsque le jeu s'arrête, les positions suivantes du pion sont toujours 9.

1. Montrer que les positions successives du pion définissent une chaîne de Markov homogène sur les entiers de 1 à 9. Donner la matrice de transition de cette chaîne de Markov.
   Indication : pour chaque position i possible du pion et pour chaque nombre k obtenu en lançant le dé, déterminer la prochaine position $f(i,k)$ du pion. Les résultats obtenus pourront être écrits sous forme d'un tableau.
2. Dans le véritable jeu de l'oie, toutes les cases ne sont pas identiques. On modifie le plateau du jeu en supposant que la case 7 est une case "oubliette", ce qui signifie que si le pion tombe sur cette case, il y reste indéfiniment. Déterminer la matrice de transition de la chaîne de Markov qui décrit les positions successsives du pion sur ce nouveau plateau.
3. *Une autre variante* : on modifie le plateau initial en supposant que sur la case 4, il est écrit "attendre deux tours". Les positions successives du pion sur le plateau ne constitue plus une chaîne de Markov. Expliquez pourquoi.
   Montrer que l'on peut tout de même décrire le mouvement du pion dans ce nouveau jeu par une chaîne de Markov homogène en ajoutant deux cases fictives au plateau.

▷ *Exercice 55.* (*Chaîne d'Ehrenfest*) On modélise le mouvement de $m \geq 2$ molécules entre deux compartiments notés A et B de la façon suivante : on discrétise le temps en intervalles de même durée notée Δt. On suppose que pendant un intervalle de temps Δt, une molécule tirée au hasard change de compartiment. On note X_n le nombre de molécules dans le premier compartiment à l'instant $n\Delta t$.

Montrer que (X_n) est une chaîne de Markov homogène dont on donnera la matrice de transition Q et le graphe associé.

▷ *Exercice 56.* On considère un système capable de reconnaitre les lettres d'une séquence formée à partir des lettres `a, c, g, t` et capable de donner les positions du motif `aac` dans la séquence.
Description du fonctionnement du système : le système peut être dans quatre états différents numérotés de 0 à 3 :
– le système est dans l'état 3 si les trois dernières lettres qu'il a lues forment le motif `aac`,
– le système est dans l'état 2 si les deux dernières lettres qu'il a lues forment le motif `aa`,
– le système est dans l'état 1 si la dernière lettre qu'il a lue est `a` et si l'avant-dernière lettre lue n'est pas `a`,
– dans les situations autres que les trois situations précédentes, le système est dans l'état 0.
On fait lire au système une séquence de lettres que l'on constitue en tirant les lettres indépendamment les unes des autres, de sorte que :
– la lettre `a` est tirée avec probabilité $\frac{1}{3}$,
– les lettres `c` et `t` sont tirées avec probabilité $\frac{1}{4}$ chacune,
– la lettre `g` est tirée avec probabilité $\frac{1}{6}$.
L'état du système après la lecture de la n-ième lettre est une variable aléatoire que l'on note X_n (par hypothèse $X_0 = 0$).

1. Montrer que la suite $(X_n)_n$ définit une chaîne de Markov homogène. Déterminer sa matrice de transition.
2. Représenter le graphe associé à la matrice de transition de $(X_n)_n$.

▷ *Exercice 57.* Le modèle simple de Wright-Fisher décrit l'évolution de la fréquence des allèles A et a d'un même gène dans une population de taille fixe. Ce modèle suppose que les croisements se forment au hasard relativement à ce gène (hypothèse de panmixie) qu'il n'y a ni mutation, ni effet de la sélection.
Hypothèses :
– on suppose que la taille de la population est constante égale à N à chaque génération (les individus de la génération k sont par définition, les descendants directs d'un croisement de deux individus de la génération $k-1$). Les gènes des individus d'une génération constituent un ensemble de $2N$ gènes.
– Pour constituer l'ensemble des gènes des individus de la génération $k+1$, on tire au hasard chaque gène parmi les gènes des individus de la génération k (tirage avec remise).
Pour tout $k \in \mathbb{N}$, on note X_k le nombre d'allèles A dans l'ensemble des gènes des individus de la k-ième génération. Montrer que la suite $(X_k)_k$ est une chaîne de Markov homogène et déterminer sa matrice de transition Q.

8 Loi de la chaîne à un instant donné

Dans toute la suite du cours, $(X_n)_{n\in\mathbb{N}}$ désigne une chaîne de Markov homogène de matrice de transition Q et d'espace d'états $\mathcal{X}$, fini ou dénombrable. D'après la proposition 22, la loi d'une chaîne de Markov homogène est entièrement déterminée par la donnée de la loi de X_0 et de sa matrice de transition Q. On peut déduire de cette proposition plusieurs formules simples permettant le calcul des probabilités de transition en m étapes et le calcul de la loi de X_n.

Proposition 24 (Probabilités de transition en m étapes)

Soit $i, j \in \mathcal{X}$ deux états et $n, m \in \mathbb{N}$ avec $m \geq 1$. Notons H_{n-1} un événement ne dépendant que des valeurs de $X_0, \ldots, X_{n-1}$ et tel que $P(\{X_n = i\} \cap H_{n-1}) > 0$. La probabilité $P(X_{n+m} = j | \{X_n = i\} \cap H_{n-1})$

ne dépend que de m, i et j. Notons-la $p^{(m)}(i,j)$. Elle peut s'obtenir par récurrence :

$$p^{(1)}(i,j) = Q(i,j) \text{ et pour tout } m \in \mathbb{N}^*,\ p^{(m)}(i,j) = \sum_{\ell \in \mathcal{X}} Q(i,\ell) p^{(m-1)}(\ell,j). \tag{3}$$

N.B. La probabilité que la chaîne de Markov passe de l'état i à j en m coups est égale à $p^{(m)}(i,j)$. La proposition dit simplement que cette probabilité est égale à la probabilité de passer de i à un certain état ℓ en un coup, puis d'aller de l'état ℓ à l'état j en $m-1$ coups.

Si l'ensemble des états $\mathcal{X}$ est fini, la formule (3) dit que $p^{(m)}(i,j)$ est le coefficient (i,j) du produit matricielle de Q par la matrice $p^{(m-1)} = (p^{(m-1)}(i,j))_{(i,j)\in\mathcal{X}^2}$. La formule (3) a toujours un sens lorsque $\mathcal{X}$ est dénombrable ce qui permet d'étendre la multiplication à des matrices à coefficients positifs ayant un nombre infini de lignes et de colonnes. Donc, en utilisant les notations matricielles, on en déduit que $p^{(m)}(i,j)$ est le coefficient (i,j) de la matrice Q^m et en particulier :

Corollaire 25 *Soit $m \in \mathbb{N}^*$ et $n \in \mathbb{N}$.*
Pour tout $i, j \in \mathcal{X}$, $P(X_{n+m} = j \mid X_n = i)$ est le coefficient (i,j) de la matrice Q^m.

Preuve de la proposition 24. Le fait que $P(X_{n+m} = j | \{X_n = i\} \cap H_{n-1})$ ne dépend que de i, j et m peut se montrer par récurrence sur m.

- Pour $m = 1$, la propriété est une conséquence directe de la définition d'une chaîne de Markov homogène ; en effet, H_{n-1} est un événement de la forme "$(X_{n-1}, \ldots, X_0) \in A$" et

$$\begin{aligned} &P(X_{n+1} = j | X_n = i, (X_{n-1}, \ldots, X_0) \in A) = \\ &\quad \sum_{(\ell_{n-1},\ldots,\ell_0)\in A} P(X_{n+1} = j | X_n = i, X_{n-1} = \ell_{n-1}, \ldots, X_0 = \ell_0) \frac{P(X_n = i, X_{n-1} = \ell_{n-1}, \ldots, X_0 = \ell_0)}{P(X_n = i, (X_{n-1}, \ldots, X_0) \in A)} \\ &\quad = Q(i,j) \end{aligned}$$

- Soit $m \geq 2$. Supposons que le résultat soit vrai pour $m-1$, c'est-à-dire que pour tout n, $i \in \mathcal{X}$, $j \in \mathcal{X}$ et pour tout événement H_{n-1} ne dépendant que de $(X_0, \ldots, X_{n-1})$, $P(X_{n+m-1} = j | \{X_n = i\} \cap H_{n-1})$ ne dépend ni de n, ni de H_{n-1}. On note cette probabilité $p^{(m-1)}(i,j)$. Soit $n \in \mathbb{N}^*$. En décomposant par rapport à la valeur de X_{n+1} on obtient :

$$P(X_{n+m} = j | \{X_n = i\} \cap H_{n-1}) = \sum_{k \in \mathcal{X}} P(X_{n+m} = j | \{X_{n+1} = k\} \cap G_n) P(X_{n+1} = k | \{X_n = i\} \cap H_{n-1})$$

  avec $G_n = \{X_n = i\} \cap H_{n-1}$. Comme G_n est un événement ne dépendant que de $X_0, \ldots, X_n$, l'hypothèse de récurrence implique que $P(X_{n+m} = j | \{X_{n+1} = k\} \cap G_n) = p^{(m-1)}(k,j)$.
  D'autre part, $P(X_{n+1} = k | \{X_n = i\} \cap H_{n-1}) = Q(i,k)$. Donc, $P(X_{n+m} = j | \{X_n = i\} \cap H_{n-1})$ ne dépend que de i, de j et de m, ce qui termine la preuve par récurrence.

Remarquons enfin que si on note $p^{(m)}(k,j)$ la valeur de $P(X_{n+m} = j | \{X_{n+1} = k\} \cap G_n)$ on a aussi obtenu dans la preuve l'égalité suivante : $p^{(m)}(i,j) = \sum_{k\in\mathcal{X}} Q(i,k) p^{(m-1)}(k,j)$. □

On peut déduire de la proposition 24 une façon simple de calculer la loi de X_n puisque

$$P(X_n = j) = \sum_{i \in \mathcal{X}} P(X_n = j \mid X_0 = i) P(X_0 = i) \ \ :$$

Corollaire 26 (Loi de X_n) *On identifie ici l'espace des états $\mathcal{X}$ à l'ensemble $\{1, 2, \ldots, N\}$. Pour $n \in \mathbb{N}$, notons π_n le vecteur ligne donnant la loi de X_n :*

$$\pi_n = (P(X_n = 1), P(X_n = 2), \ldots, P(X_n = N)).$$

- *Pour tout $n \in \mathbb{N}$, $\pi_n = \pi_0 Q^n$.*
- *Les lois de X_1, X_2, …, X_n, … peuvent s'obtenir de façon récursive :*

$$\text{pour tout } k \in \mathbb{N}, \text{ et } j \in \{1, \ldots, N\},\ \pi_{k+1}(j) = \sum_{i=1}^{N} \pi_k(i) Q(i,j),$$

  ce qui s'écrit matriciellement, $\pi_{k+1} = \pi_k Q$.

▷ *Exercice 58.* Considérons un système qui peut se trouver dans 3 états différents notés $1, 2, 3$ à chaque unité de temps. Supposons que l'on puisse décrire l'évolution de l'état du système par une chaîne de Markov homogène $(X_n)_{n\geq 0}$ de matrice de transition $Q = \begin{pmatrix} 0.5 & 0.2 & 0.3 \\ 0.1 & 0.8 & 0.1 \\ 0.6 & 0 & 0.4 \end{pmatrix}$, X_n décrivant l'état du système à l'instant n. On ne connait pas l'état du système à l'instant 0 mais on sait qu'il a une chance sur 2 de s'être trouvé dans l'état 1 à l'instant 0 et une chance sur 2 de s'être trouvé dans l'état 2.
Quelle est la loi de X_3 ? A l'instant 1, on a observé que le système se trouvait dans l'état 2. Quelle est la probabilité qu'il se trouve dans l'état 3 à l'instant 4 ?

Solution. Notons π_k le vecteur contenant les coefficients de la loi de X_k : $\pi_k = (P(X_k = 1), P(X_k = 2), P(X_k = 3))$. D'après l'énoncé, la loi initiale est $\pi_0 = (1/2, 1/2, 0)$.

1. D'après le corollaire 26, $\pi_3 = \pi_0 Q^3 = (0.338, 0.432, 0.23)$.
2. On demande de déterminer $P(X_4 = 3|X_1 = 2)$. D'après le corollaire 25, $P(X_4 = 3|X_1 = 2) = [Q^3](2, 3) = 0.183$.

▷ *Exercice 59.* (suite de l'exercice 36, page 31) Donner une formule permettant de calculer la loi de X_n en fonction des coefficients de la loi de X_{n-1} lorsque la fortune qu'il cherche à atteindre est $m \geq 4$. On suppose que la fortune intiale du joueur est de 1 euro. Utiliser cette formule pour déterminer la loi de la fortune du joueur après la première partie, la deuxième partie, puis la troisième partie. Quelle est la probabilité qu'il joue au plus 4 parties ?
Solution. D'après l'énoncé,
- $X_n = X_{n-1}$ si $X_{n-1} = 0$ ou $X_{n-1} = m$;
- $X_n = X_{n-1} - 1$ si $1 \leq X_{n-1} \leq m - 1$ et si le joueur perd la n-ème partie.
- $X_n = X_{n-1} + 1$ si $1 \leq X_{n-1} \leq m$ et si le joueur gagne la n-ème partie.

Soit $k \in \{1, \ldots, m - 1\}$.

$$\begin{aligned} P(X_n = k) &= P(X_n = k \text{ et } X_{n-1} = k-1) + P(X_n = k \text{ et } X_{n-1} = k+1) \\ &= P(X_n = k|X_{n-1} = k-1)P(X_{n-1} = k-1) + P(X_n = k|X_{n-1} = k+1)P(X_{n-1} = k+1) \text{ avec} \end{aligned}$$

- $P(X_n = k|X_{n-1} = k - 1) = p$ si $k \geq 2$ et $P(X_n = k|X_{n-1} = k - 1) = 0$ si $k = 1$;
- $P(X_n = k|X_{n-1} = k + 1) = 1 - p$ si $k \leq m - 2$ et $P(X_n = k|X_{n-1} = k + 1) = 0$ si $k = m - 1$.

De même,
$P(X_n = 0) = P(X_n = 0 \text{ et } X_{n-1} = 0) + P(X_n = 0 \text{ et } X_{n-1} = 1) = P(X_{n-1} = 0) + (1 - p)P(X_{n-1} = 1)$.
$P(X_n = m) = P(X_n = m \text{ et } X_{n-1} = m) + P(X_n = m \text{ et } X_{n-1} = m - 1) = P(X_{n-1} = m) + pP(X_{n-1} = m - 1)$.
Notons $u_n(k) = P(X_n = k)$ pour tout $k \in \{0, \ldots, m\}$. La loi de X_n est décrite par le vecteur $u_n = (u_n(0), \ldots, u_n(m))$.
On a obtenu :

$$\begin{cases} u_n(0) & = u_{n-1}(0) + (1-p)u_{n-1}(1) \\ u_n(1) & = (1-p)u_{n-1}(2) \\ u_n(k) & = pu_{n-1}(k-1) + (1-p)u_{n-1}(k+1) \text{ si } k \in \{2, \ldots, m-2\} \\ u_n(m-1) & = pu_{n-1}(m-2) \\ u_n(m) & = pu_{n-1}(m-1) + u_{n-1}(m) \end{cases}$$

Pour $s = 1$, on obtient : $u_0 = (0, 1, 0, \ldots, 0)$, $u_1 = (1 - p, 0, p, 0, \ldots, 0)$, $u_2 = (1 - p, p(1 - p), 0, p^2, 0, \ldots, 0)$ et $u_3 = (1 - p + p(1 - p)^2, 0, 2p^2(1 - p), 0, p^3, 0, \ldots, 0)$.
Le joueur s'arrête de jouer s'il n'a plus d'euro ou s'il a m euros. Comme par convention, $X_n = X_{n-1}$ si $X_{n-1} = 0$ ou m, la probabilité qu'il joue au plus 4 parties est égale à $a = P(X_4 = 0) + P(X_4 = m)$. Si $m > 4$ alors $a = 1 - p + p(1 - p)^2$ et si $m = 4$ alors $a = 1 - p + p(1 - p)^2 + p^3$.

▷ *Exercice 60.* Une source émet une suite de bits 0 et 1. On suppose que la loi du premier bit émis est donnée par le tableau suivant :

0	1
11/20	9/20

On suppose que les chiffres sucessivement émis constituent une chaîne de Markov homogène dont la matrice de transition est : $\begin{pmatrix} \frac{3}{5} & \frac{2}{5} \\ \frac{11}{20} & \frac{9}{20} \end{pmatrix}$. Quelle est la probabilité que le premier bit 1 émis par la source soit le troisième bit de la suite ?

▷ *Exercice 61.* (*suite de l'exercice 38, page 31 sur l'autofécodation d'une plante*)
Lorsqu'on ne connait pas le génotype de la plante initiale, on peut par exemple supposer que l'on ait autant de chance que la plante soit de type AA, Aa ou aa pour le gène étudié : cela signifie que l'on prend comme loi initiale de la chaîne de Markov $(X_n)_n$ la loi uniforme sur les 3 états de la chaîne.

1. Calculer sous cette hypothèse la loi de X_1, puis la loi de X_2.
2. Montrer que pour tout $n \in \mathbb{N}$, $Q^n = \begin{pmatrix} 1 & 0 & 0 \\ 1/2 - 1/2^{n+1} & 1/2^n & 1/2 - 1/2^{n+1} \\ 0 & 0 & 1 \end{pmatrix}$
3. En déduire une expression de la loi de X_n pour tout $n \in \mathbb{N}$. Par quelle loi peut-on approcher la loi de X_n lorsque n est grand ?

▷ *Exercice 62. Modèle de substitution.* On s'intéresse uniquement aux mutations correspondant à des substitutions d'un nucléotide par un autre sur une base lors de la réplication d'une séquence d'ADN. On supposera qu'à chaque réplication et sur chaque base de la séquence, une purine a une probabilité $p \in [0, 1[$ d'être substituée en une pyrimidine et une pyrimidine a une probabilité $q \in [0, 1[$ d'être substituée en une purine, ceci indépendamment des autres bases et de ce qui s'est passé dans les réplications précédentes. On considère la suite des séquences obtenues à partir d'une séquence donnée par réplications successives : la n-ième séquence désigne la séquence obtenue par réplication de la $(n-1)$-ième séquence. A la k-ième base de la séquence, on associe la variable aléatoire $X_n^{(k)}$ qui vaut 1 si elle est occupée par une purine et qui vaut 2 si elle est occupée par une pyrimidine.

1. Ecrire la matrice de transition de la chaîne de Markov $X_n^{(k)}$. On la notera T dans la suite.
2. Montrer que pour tout $n \in \mathbb{N}$, $(p+q)T^n = (1-p-q)^n \begin{pmatrix} p & -p \\ -q & q \end{pmatrix} + \begin{pmatrix} q & p \\ q & p \end{pmatrix}$.
3. Que peut-on dire de T^n lorsque n tend vers $+\infty$?
4. On pose $a = P(X_0 = 1) = 1 - P(X_0 = 2)$.
   (a) Que peut-on dire de la loi de X_n lorsque n tend vers $+\infty$?
   (b) Que peut-on dire de $P(X_n = X_0)$ lorsque n tend vers $+\infty$?
5. Montrer que si $k \neq l$ alors $(X_n^{(k)}, X_n^{(l)})_n$ définit une chaîne de Markov homogène et déterminer sa matrice de transition.

▷ *Exercice 63.* On considère une population de cellules toutes identiques (génération 0). Si on ne tient pas compte de l'apparition possible de mutations, chaque cellule de cette population donne naissance à deux cellules identiques avec probabilité p et meurt sans se diviser avec probabilité $1-p$ ($p \in [0, 1]$). Les cellules issues de la division des cellules de la génération $n-1$ constituent les cellules de la génération n. On note X_n le nombre de cellules à la génération n.

1. Montrer que la suite $(X_n)_{n\geq 0}$ définit une chaîne de Markov homogène dont on déterminera l'espace des états.
2. On suppose que la population à la géneration 0 est constituée d'une cellule. Déterminer la loi du nombre de cellules à la génération 2.
3. Si on suppose maintenant que la population de cellules à la génération 0 est constituée de deux cellules, quelle est la loi du nombre de cellules à la génération 2 ?

▷ *Exercice 64.* Une séquence d'ADN peut être vue comme une suite de lettres `a, c, g, t`. On choisit de la modéliser comme une réalisation d'une chaîne de Markov homogène. Plus précisément, pour tout $n \geq 1$, on note X_n la n-ième base composant une séquence d'ADN en partant d'une de ses extrêmités et on suppose que (X_n) est une chaîne de Markov homogène d'espace d'états $\mathcal{A} = \{$ `a,c,g,t`$\}$ (on identifiera `a` à l'état 1, `c` à l'état 2, `g` à l'état 3 et `t` à l'état 4) de matrice de transition Q.

$$Q = \begin{pmatrix} 0.2 & 0.1 & 0.6 & 0.1 \\ 0.3 & 0.1 & 0.1 & 0.5 \\ 0.3 & 0.2 & 0.2 & 0.3 \\ 0.6 & 0.1 & 0.2 & 0.1 \end{pmatrix}.$$

La loi de la première base est décrite par le tableau suivant :

a	c	g	t
0	0.6	0	0.4

1. Déterminer la probabilité que la séquence commence par le codon `tag`.
2. Déterminer la probabilité que la troisième base soit `c`.
3. Déterminer la probabilité que le 4-ième nucléotide soit un `a` sachant que le 2-ième nucléotide est un `a`.
4. Déterminer la probabilité que les bases 1 et 2 forment le motif `ta`, sachant que les bases 3 et 4 forment le motif `ca`.

9 Loi invariante et comportement asymptotique de la loi de X_n

Dans cette partie $\mathcal{X}$ désigne un ensemble ayant un nombre fini d'éléments que l'on peut identifier à $\{1, \dots, r\}$ avec r un entier strictement positif, ou bien un ensemble dénombrable comme $\mathbb{N}^*$ et $\mathbb{Z}$ et $(X_n)_n$ désigne une chaîne de Markov homogène d'espace d'états $\mathcal{X}$.

9.1 Loi de probabilité invariante

Définition. Soit Q une matrice stochastique sur l'ensemble $\mathcal{X} = \{1, \dots, r\}$ avec $r \in \mathbb{N}^*$ ou $r = +\infty$.
On dit qu'un vecteur $v = (v(1), \dots, v(r))$ est *invariant par* Q si $vQ = v$ c'est-à-dire si pour tout $j \in \mathcal{X}$,

$$v(j) = \sum_{i\in\mathcal{X}} v(i)Q(i,j).$$

On dit qu'un vecteur $\mu = (\mu(1), \dots, \mu(r))$ définit une *loi de probabilité invariante par* Q si les coefficients de μ sont positifs, leur somme est égale à 1 et si μ est invariant par Q.
Si $(X_n)_{n\in\mathbb{N}}$ est une chaîne de Markov homogène de matrice de transition Q, une loi de probabilité invariante par Q est aussi appelée *une loi stationnaire de la chaîne* $(X_n)_{n\in\mathbb{N}}$.

N.B. Lorsque r est fini, trouver les lois de probabilité invariantes par Q revient donc à trouver les vecteurs $(x_1, \dots, x_r)$ à coefficients positifs ou nuls solution du système suivant formé de $r+1$ équations :

$$\begin{cases} x_1 & + & x_2 & + & \dots & + & x_r & = & 1 \\ x_1(Q(1,1)-1) & + & x_2 Q(2,1) & + & \dots & + & x_r Q(r,1) & = & 0 \\ x_1 Q(1,2) & + & x_2(Q(2,2)-1) & + & \dots & + & x_r Q(r,2) & = & 0 \\ \dots \\ x_1 Q(1,r) & + & x_2 Q(2,r) & + & \dots & + & x_r(Q(r,r)-1) & = & 0 \end{cases}$$

N.B. Si μ_1 et μ_2 sont deux lois de probabilité invariantes par Q alors pour tout $\alpha \in [0,1]$, $\alpha\mu_1 + (1-\alpha)\mu_2$ est encore une loi de probabilité invariante par Q.

▶ *Exemple 42.* (*Chaîne à deux états*) Soit $(X_n)_n$ une chaîne de Markov homogène d'espace d'états $\{1,2\}$ et de matrice de transition $Q = \begin{pmatrix} 1-p & p \\ q & 1-q \end{pmatrix}$ avec $p, q \in [0,1]$. Si $p+q > 0$, la loi de probabilité $\mu = (\frac{q}{p+q}, \frac{p}{p+q})$ sur $\{1,2\}$ est l'unique loi de probabilité invariante par Q. Si $p = q = 0$ alors Q est la matrice identité, toutes les lois de probabilité sur $\{1,2\}$ sont invariantes par Q.

Propriété 27 *Soit $(X_n)_n$ est une chaîne de Markov de matrice de transition Q. Supposons que la loi de X_0 soit invariante par Q.*
Alors, pour tout $n \in \mathbb{N}$, la loi de X_n est égale à la loi de X_0.
Plus généralement, pour tout $n \in \mathbb{N}$ et $m \in \mathbb{N}$, $(X_m, \dots, X_{m+n})$ a même loi que $(X_0, \dots, X_n)$ (on dit alors que la suite $(X_n)_n$ est stationnaire).

N.B. Comme $P(X_{n+1} = x) - P(X_n = x) = P(X_{n+1} = x, X_n \neq x) - P(X_{n+1} \neq x, X_n = x)$, dire que la loi de la chaîne de Markov ne varie pas, signifie donc que si on considère un grand nombre de systèmes dont l'évolution est décrite par cette chaîne de Markov, on s'attend à ce que la proportion de systèmes qui arrivent dans l'état x entre deux instants successifs soit très proche de la proportion de ceux qui étaient dans l'état x et qui le quittent.

▶ *Exemple 43.* Si le déplacement quotidien d'un vélo en location entre les différents points de prêts/restitutions de la société de location se modélise fidèlement par une chaîne de Markov homogène, choisir le nombre de vélos disponibles dans les différents points de prêts/restitutions proportionnellement aux coefficients d'une probabilité invariante par la chaîne de Markov assure une stabilité de la répartition des vélos au cours du temps.

▶ *Exemple 44.* *Modélisation markovienne d'une séquence d'ADN de longueur K*
Une séquence d'ADN peut être vue comme une suite de lettres `a`, `c`, `g` et `t`. Modéliser une séquence d'ADN de longueur K par une chaîne de Markov homogène, consiste à considérer les K nucléotides de la séquence en partant de l'extrêmité 5' comme une réalisation des K premiers états d'une chaîne de Markov homogène $(X_n)_{n\geq 1}$. Dans les applications, on choisit en général la matrice de transition Q de $(X_n)_{n\geq 1}$ comme une matrice dont tous les coefficients sont strictement positifs et on choisit X_1, le premier nucléotide de la séquence d'ADN, suivant la loi de probabilité invariante[4] notée μ dans la suite. Dans ce cas, on aura la même probabilité de trouver le nucléotide `a` (resp. `c`, `g` et `t`) sur la base 1 que sur n'importe quelle autre base. Cela est vrai aussi pour n'importe quelle suite de nucléotides.
Regardons par exemple le motif `tag`. Comme on a la même probabilité de trouver le motif `tag` en première position[5] qu'à n'importe quelle autre position de la séquence, alors le nombre moyen de motifs `tag` dans la séquence $(X_1, \dots, X_K)$ est $(K-2)\mu(\mathtt{t})Q(\mathtt{t},\mathtt{a})Q(\mathtt{a},\mathtt{g})$. Comme la probabilité que le motif `tag` occupe en même temps les bases $(i, i+1, i+2)$ et les bases $(j, j+1, j+2)$ est nulle pour $|i-j| \leq 2$ et vaut $\mu(t)Q(t,a)Q(a,g)Q^\ell(g,t)Q(t,a)Q(a,g)$ si $|i-j| = 2+\ell$ pour $\ell \geq 1$ alors la variance du nombre de motifs `tag` est

$$(K-2)p_{tag} - (K-2)^2 p_{tag}^2 + 2\frac{p_{tag}^2}{\mu(t)} \sum_{\ell=1}^{K-5} (K-\ell-4) Q^\ell(g,t) \text{ où } p_{tag} = \mu(t)Q(t,a)Q(a,g).$$

▷ *Exercice 65.* *(suite de l'exemple 38, page 31 sur l'autofécondation d'une plante)* Déterminer les lois stationnaires de la chaîne de Markov $(X_n)_n$ décrivant le génotype de la plante choisie à la génération n.

▷ *Exercice 66.* Pour la chaîne de Markov décrivant l'évolution de la fortune du joueur (exemple 36), montrer que toutes les lois de probabilité de la forme suivante sont des lois de probabilité invariantes par la chaîne : $\mu = (a, \underbrace{0, \dots, 0}_{m-1 \text{ fois}}, 1-a)$ avec $a \in [0,1]$.

[4] On verra dans la suite du cours qu'une telle matrice de taille finie et dont tous les coefficients sont strictement positifs admet une unique loi de probabilité invariante

[5] On dira qu'il y a un motif `tag` en positif ℓ si la ℓ-ième base est un `t`, la $(\ell+1)$-ième base est un `a` et si la $(\ell+2)$-ième base est un `g`.

▷ *Exercice 67.* Deux joueurs A et B s'affrontent à un jeu de hasard. On définit le score du joueur A de la façon suivant : initialement le score de A est 0. A chaque partie gagnée, son score augmente de 1. A chaque partie perdue son score diminue de 1. On suppose que le jeu est équitable. On note X_n le score du joueur A après la n-ième partie.

1. Montrer que la suite $(X_n)_n$ est une chaîne de Markov homogène sur $\mathbb{Z}$ et donner sa matrice de transition Q.
2. Soit $n \in \mathbb{N}^*$ et $\ell \in \{0, \dots, n\}$. Calculer la probabilité que sur n parties le joueur A en gagne ℓ.
3. Déduire de la question précédente que, pour tout $n, k \in \mathbb{N}$ et $x, y \in \mathbb{Z}$, $P(X_{n+k} = y | X_n = x) = C_k^{(y-x+k)/2} 1/2^k$ si $(y - x + k)/2$ est un entier compris entre 0 et k.
4. Montrer que si la famille de réels positifs $(z_i)_{i\in\mathbb{Z}}$ est invariante par la matrice de transition Q, c'est-à-dire si $z_j = \sum_{i\in\mathbb{Z}} z_i Q(i,j)$ pour tout $i \in \mathbb{Z}$ alors pour tout $i \in \mathbb{Z}$, $z_i = z_0 + i(z_1 - z_0)$.
5. Déduire de la question précédente que la chaîne de Markov n'a pas de loi stationnaire.
6. En utilisant la formule de Stirling suivante :
$$\frac{n!e^n}{\sqrt{2\pi n}n^n} \textit{ tend vers } 1 \textit{ lorsque } n \textit{ tend vers } +\infty,$$
montrer que pour tout $x, y \in \mathbb{Z}$ et $n \in \mathbb{N}$, $P(X_{n+k} = y | X_n = x)$ tend vers 0 lorsque k tend vers $+\infty$.

▷ *Exercice 68.* (*suite de l'exercice 49*) Montrer que la chaîne de Markov décrite dans l'exercice 49 admet, comme unique probabilité invariante, la loi géométrique[6] sur $\mathbb{N}$ de paramètre $\frac{3}{4}$.

▷ *Exercice 69.* On construit une suite de 100 bits 0 et 1 de la façon suivante : on tire le premier bit X_1 suivant la loi $\mu_1 = (1/3, 2/3)$. Une fois les n premiers bits choisis, on tire le $(n+1)$-ième bit X_{n+1} de sorte que $P(X_{n+1} = 1 | X_n = 1) = 1/2$ et $P(X_{n+1} = 1 | X_n = 0) = 1/4$.

1. En moyenne combien de bits 1 y aura-t-il dans une telle suite ?
2. En moyenne combien de motifs 000111 y aura-t-il dans une telle suite ?

▷ *Exercice 70.*
Dans le modèle de substitution en temps discret proposé par Kimura, la substitution d'un nucléotide par un autre à une base donnée est décrite par une chaîne de Markov homogène dont les états sont $e_1 = a$, $e_2 = g$, $e_3 = c$ et $e_4 = t$ et dont la matrice transition est de la forme
$$Q = \begin{pmatrix} 1-\alpha-2\beta & \alpha & \beta & \beta \\ \alpha & 1-\alpha-2\beta & \beta & \beta \\ \beta & \beta & 1-\alpha-2\beta & \alpha \\ \beta & \beta & \alpha & 1-\alpha-2\beta \end{pmatrix}$$
avec $\alpha, \beta > 0$ tels que $\alpha + 2\beta < 1$.

1. Montrer que la chaîne de Markov admet la loi uniforme comme loi de probabilité invariante.
2. On dispose d'une séquence d'ADN dont le premier nucléotide est un `a`. En supposant que chaque base a subi des mutations suivant le modèle de Kimura au cours des réplications sucessives et que les nucléotides sur la séquence d'origine sont distribués suivant la loi uniforme, quelle est la probabilité que la séquence avant la dernière réplication commence par le nucléotide `t` ?

▷ *Exercice 71.* Soit $(X_n)_{n\in\mathbb{N}}$ une chaîne de Markov homogène de matrice de transition Q et de loi initiale μ. On suppose que μ est une loi de probabilité invariante par Q. On note $\mathcal{A}$ le support de μ (c'est-à-dire l'ensemble des états i pour lesquels $\mu(i) > 0$). Soit $n \in \mathbb{N}^*$. Pour tout $k \in \{0, \dots, n\}$, on pose $Y_k = X_{n-k}$.
Déterminer la loi conditionnelle de Y_k sachant que $\{Y_0 = i_0, \dots, Y_{k-1} = i_{k-1}\}$ pour tout $i_0, \dots, i_{k-1} \in \mathcal{A}$.

▷ *Exercice 72.* (*suite de l'exercice 55*) On suppose qu'au départ, chaque molécule est placée au hasard dans l'un des deux compartiments. Quelle est la loi de X_0. Déterminer la loi de X_1 ? Que peut-on dire de la loi de X_n pour tout $n \geq 2$?

9.2 Convergence en loi

Définition. Soit $(Z_n)_n$ une suite de variables aléatoires réelles à valeurs dans $\mathcal{X}$.

1. Soit μ une loi de probabilité sur $\mathcal{X}$. On dit que *la loi de Z_n converge vers μ lorsque n tend vers $+\infty$* si pour tout $x \in \mathcal{X}$, $P(Z_n = x)$ converge vers $\mu(x)$ lorsque n tend vers $+\infty$.
2. On dit que la loi de Z_n converge lorsque n tend vers $+\infty$ s'il existe une loi de probabilité μ sur $\mathcal{X}$ telle que $P(Z_n = x)$ tende vers $\mu(x)$ pour tout $x \in \mathcal{X}$.

► *Exemple 45.* Notons Z_n une variable aléatoire de loi de Bernoulli de paramètre $p_n = \frac{1}{2} - \frac{1}{4^n}$ pour tout $n \in \mathbb{N}^*$. Alors la loi de Z_n converge vers la loi de Bernoulli de paramètre 1/2.

► *Exemple 46.* Soit a un réel strictement positif. Pour tout entier $n \geq a$, notons Z_n une variable aléatoire de loi de binomiale $\mathcal{B}(n, a/n)$. Alors, pour tout $k \in \mathbb{N}$, $P(Z_n = k)$ tend vers $q_k = e^{-a}\frac{a^k}{k!}$ lorsque n tend vers $+\infty$. L'ensemble des coefficients q_k, $k \in \mathbb{N}$ forme une loi de probabilité sur $\mathbb{N}$ puisque $\sum_{k=0}^{+\infty} q_k = 1$. Cette loi est appelée la loi de Poisson de paramètre a. On en déduit donc que la loi de Z_n converge vers la loi de Poisson de paramètre a lorsque n tend vers $+\infty$.

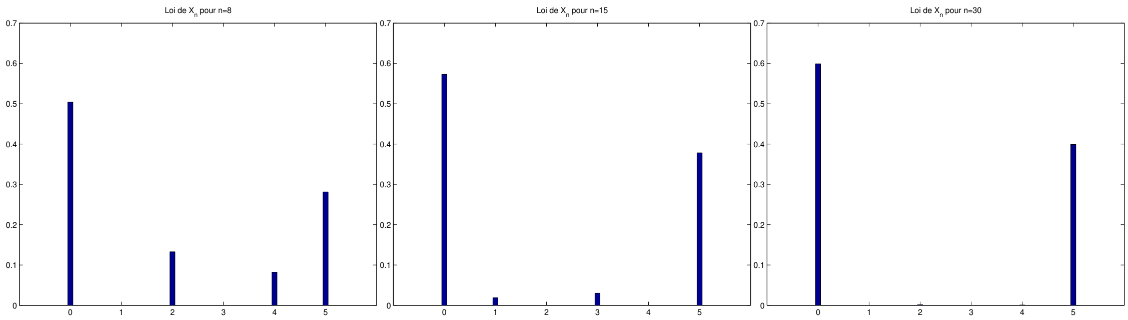
[6] La loi géométrique sur $\mathbb{N}$ de paramètre $a \in [0, 1]$ est définie comme la probabilité sur $\mathbb{N}$ qui affecte, à chaque entier positif k, le « poids » $(1-a)^k a$.

9.3 Exemples de comportements en loi

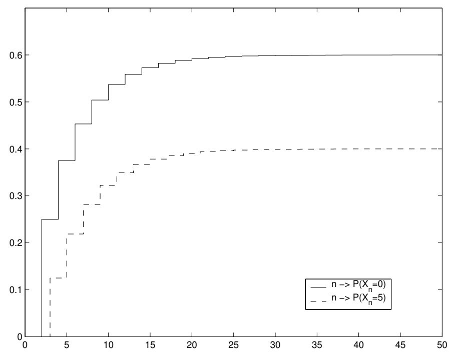
Nous donnons dans ce paragraphe des exemples illustrant différents types d'évolutions de la loi à un instant donné d'une chaîne de Markov homogène.

9.3.1 Exemple d'une chaîne de Markov dont la loi converge vers une loi qui dépend de la loi initiale de la chaîne.

Dans l'exemple de la chaîne de Markov (X_n) décrivant l'évolution de la fortune d'un joueur jouant à un jeu de hasard équitable (voir exemple 36), la loi de la fortune du joueur converge vers la loi $\mu = (1-\frac{s}{m}, 0, \dots, 0, \frac{s}{m})$ sur $\{0, \dots, m\}$: pour tout $i \in \{1, \dots, m-1\}$, $P(X_n = i)$ tend vers 0, $P(X_n = 0)$ tend vers $1-\frac{s}{m}$ et $P(X_n = m)$ tend vers $\frac{s}{m}$. Dans cet exemple, toutes les lois dont le support est $\{0, m\}$ sont invariantes par la matrice de transition de la chaîne.



Les diagrammes en bâtons ci-dessus décrivent les lois de la fortune du joueur après la 8-ième partie, la 15-ième partie et la 30-ième partie dans le cas où sa fortune initiale est de 2 euros et où il cherche à atteindre la somme de $m = 5$ euros.
Les courbes ci-contre montrent l'évolution de la probabilité que le joueur n'ait plus d'argent (respectivement dispose de 5 euros) après la n-ième partie en fonction de n.

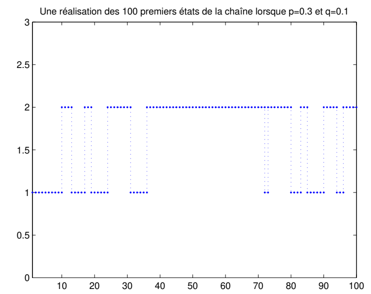


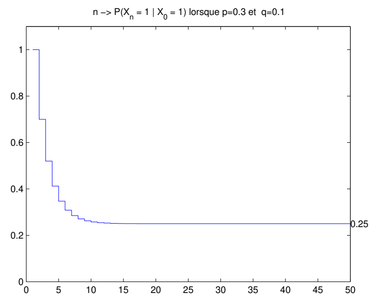
9.3.2 Exemple d'une chaîne de Markov à deux états

Soit $(X_n)_n$ une chaîne de Markov homogène d'espace d'états $\{1, 2\}$ et de matrice de transition $Q = \begin{pmatrix} 1-p & p \\ q & 1-q \end{pmatrix}$ avec $p, q \in [0, 1]$. Pour tout $n \in \mathbb{N}$,

$$(p+q)Q^n = (1-p-q)^n \begin{pmatrix} p & -p \\ -q & q \end{pmatrix} + \begin{pmatrix} q & p \\ q & p \end{pmatrix}.$$

Si $0 < p+q < 2$ alors Q^n tend vers la matrice $L = \begin{pmatrix} \frac{q}{p+q} & \frac{p}{p+q} \\ \frac{q}{p+q} & \frac{p}{p+q} \end{pmatrix}$ lorsque n tend vers $+\infty$. Cela montre que quelle que soit la loi de X_0, la loi de X_n converge vers la loi $\mu = (\frac{q}{p+q}, \frac{p}{p+q})$ qui est l'unique loi invariante par Q.
Par contre, si $p = q = 1$ c'est-à-dire si la matrice de transition s'écrit $Q = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$, alors pour tout $n \in \mathbb{N}$, $P(X_{2n} = 1 \mid X_0 = 1) = 1$ et $P(X_{2n+1} = 1 \mid X_0 = 1) = 0$. Dans ce cas, la loi de la chaîne de Markov partant de l'état 1 ne converge pas.
Enfin si $p = q = 0$, $X_n = X_0$ pour tout $n \in \mathbb{N}$ et toutes les probabilités sur $\{1, 2\}$ sont invariantes par Q.





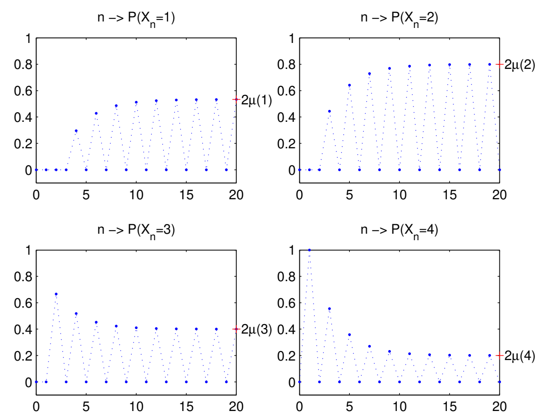
Les deux figures concernent l'exemple de la chaîne de Markov homogène à deux états avec $p = 0.3$ et $q = 0.1$. Sa loi invariante est $\mu = (1/4, 3/4)$. Sur la figure de gauche, on peut voir une réalisation de $(X_0, \dots, X_{100})$ et sur la figure de droite, la courbe $n \mapsto P(X_n = 1|X_0 = 1)$.

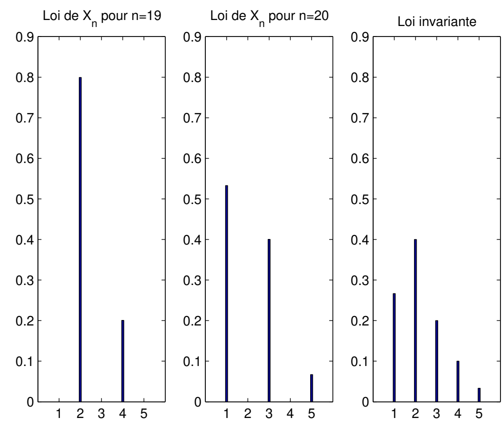
9.3.3 Exemple d'une chaîne dont les lois aux instants pairs et impairs se stabilisent vers des lois différentes

Considérons la chaîne de Markov homogène $(X_n)_n$ partant de l'état 5 et dont le graphe associé à la matrice de transition est :

$$\begin{tikzcd} \mathbf{1} \arrow[r, bend left, "1"] & \mathbf{2} \arrow[l, bend left, "2/3"] \arrow[r, bend left, "1/3"] & \mathbf{3} \arrow[l, bend left, "2/3"] \arrow[r, bend left, "1/3"] & \mathbf{4} \arrow[l, bend left, "2/3"] \arrow[r, bend left, "1/3"] & \mathbf{5} \arrow[l, bend left, "1"] \end{tikzcd}$$

Aux instants pairs, la chaîne de Markov se trouvera dans les états 1, 3 ou 5 et aux instants impairs, elle se trouvera dans les états 2 ou 4. Les figures ci-dessous montrent que lorsque n augmente, les lois de X_{2n} et de X_{2n+1} se stabilisent rapidement vers deux lois différentes que l'on peut comparer à la loi invariante $\mu = (4/15, 2/5, 1/5, 1/10, 1/30)$ de la chaîne : la loi de X_{2n} se stabilise vers la loi $\mu_0 = (2\mu(1), 0, 2\mu(3), 0, 2\mu(5))$. La loi de X_{2n+1} se stabilise vers la loi $\mu_1 = (0, 2\mu(2), 0, 2\mu(4), 0)$.



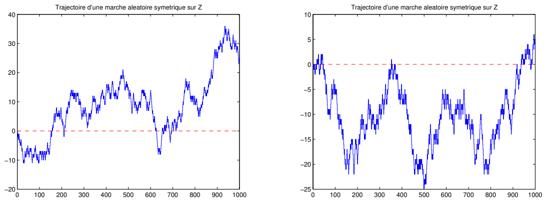


9.3.4 Exemple d'une chaîne de Markov sur $\mathbb{Z}$ dont la loi ne converge pas

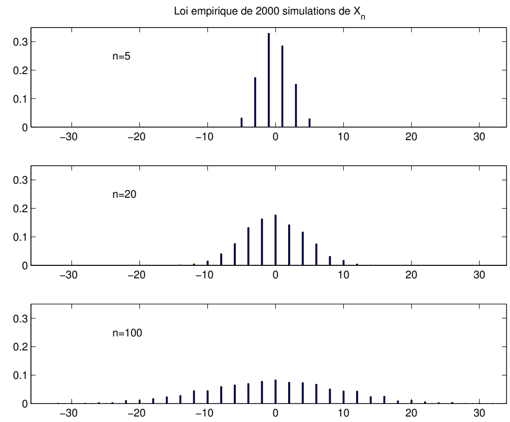
Considérons la marche aléatoire symétrique sur $\mathbb{Z}$ définie par $X_{n+1} = X_n + \epsilon_n$ avec (ϵ_n) une suite de variables aléatoires indépendantes, de loi uniforme sur $\{-1, 1\}$ et indépendante de X_0. C'est une chaîne de Markov homogène sur $\mathbb{Z}$ dont le graphe associé à la matrice de transition est :

$$\begin{tikzcd} \cdots \arrow[r, bend left, "0.5"] & \mathbf{-2} \arrow[l, bend left, "0.5"] \arrow[r, bend left, "0.5"] & \mathbf{-1} \arrow[l, bend left, "0.5"] \arrow[r, bend left, "0.5"] & \mathbf{0} \arrow[l, bend left, "0.5"] \arrow[r, bend left, "0.5"] & \mathbf{1} \arrow[l, bend left, "0.5"] \arrow[r, bend left, "0.5"] & \mathbf{2} \arrow[l, bend left, "0.5"] \arrow[r, bend left, "0.5"] & \cdots \arrow[l, bend left, "0.5"] \end{tikzcd}$$

Dans cet exemple, pour tout $k \in \mathbb{N}$, $P(X_n = k)$ tend vers 0 lorsque n tend vers $+\infty$. Donc, la loi de X_n ne converge pas lorsque n tend vers $+\infty$. On peut montrer qu'il n'existe aucune loi de probabilité invariante par la matrice de transition de cette chaîne de Markov (voir exercice 67, page 40).



Deux simulations des 1000 premiers pas d'une marche aléatoire symétrique partie de 0.



Evolution de la loi empirique de 2000 simulations de la marche aléatoire symétrique $(X_n)_n$ partie de 0.

9.4 Résultats théoriques

Sur tous les exemples vus précédemment, on observe que si la loi d'une chaîne de Markov homogène $(X_n)_n$ converge, alors sa loi limite est une loi invariante par la matrice de transition de $(X_n)_n$. On a le résultat général suivant :

Proposition 28

1. Cas d'une chaîne de Markov homogène $(X_n)_n$ d'espace d'états fini $\mathcal{X} = \{1, \dots, r\}$ et de matrice de transition Q :
   (a) *Supposons que pour tout $k \in \mathcal{X}$, $P(X_n = k)$ tend vers un nombre $L(k)$ lorsque n tend vers $+\infty$, alors $L = (L(1), \dots, L(r))$ est une probabilité invariante par Q.*
   (b) *Si pour tout $k, \ell \in \mathcal{X}$, $Q^n(k, \ell)$ tend vers un nombre $\mu(\ell)$ ne dépendant pas de k, alors μ est l'unique probabilité invariante par Q.*
2. Cas d'une chaîne de Markov homogène $(X_n)_n$ d'espace d'états $\mathcal{X} = \mathbb{N}^*$ et de matrice de transition Q :
   (a) *Supposons que pour tout $k \in \mathcal{X}$, $P(X_n = k)$ tend vers un nombre $L(k)$ lorsque n tend vers $+\infty$, alors le vecteur $L = (L(1), L(2), \dots)$ est invariant par Q et $\sum_{k=1}^{+\infty} L(k) \leq 1$.*
   (b) *Si pour tout $k, \ell \in \mathcal{X}$, $Q^n(k, \ell)$ tend vers un nombre $\mu(\ell)$ ne dépendant pas de k, alors :*
   – *soit $\mu(\ell) = 0$ pour tout $\ell \in \mathcal{X}$ et dans ce cas il n'existe pas de probabilité invariante par Q ;*
   – *soit $\mu = (\mu(\ell),\ \ell \in \mathcal{X})$ définit une probabilité sur $\mathcal{X}$ qui est l'unique probabilité invariante par Q.*

Preuve dans le cas d'une chaîne de Markov d'espace d'états fini.

1. Supposons que pour tout $k \in \mathcal{X}$, $P(X_n = k)$ tend vers un nombre $L(k)$. Comme limite d'une suite de termes compris entre 0 et 1, $L(k) \in [0, 1]$. Comme pour tout $n \in \mathbb{N}$, $\sum_{k \in \mathcal{X}} P(X_n = k) = 1$ et que $\mathcal{X}$ est fini, on a $\sum_{k \in \mathcal{X}} L(k) = 1$.
   Enfin, comme pour tout $i, j \in \mathcal{X}$ et $n \in \mathbb{N}$, $P(X_{n+1} = j) = \sum_{i \in \mathcal{X}} P(X_n = i)Q(i, j)$, en faisant tendre n vers $+\infty$ dans les deux membres de l'égalité, on obtient : $L(j) = \sum_{i \in \mathcal{X}} L(i)Q(i, j)$. Donc L définit bien une loi de probabilité invariante par Q.
2. La même preuve que précédemment montre que si pour tout $k, \ell \in \mathcal{X}$, $Q^n(k, \ell)$ tend vers un nombre $\mu(\ell)$ ne dépendant pas de k, alors μ est une probabilité invariante par Q. Montrons que c'est la seule : soit ν une loi de probabilité invariante par Q. Par définition, pour tout $n \in \mathbb{N}$ et pour tout $\ell \in \mathcal{X}$, $\nu(\ell) = \sum_{k \in \mathcal{X}} \nu(k)Q^n(k, \ell)$. En faisant tendre n vers $+\infty$, on obtient que pour tout $\ell \in \mathcal{X}$, $\nu(\ell) = \sum_{k \in \mathcal{X}} \nu(k)\mu(\ell) = \mu(\ell)$. Cela montre que μ est l'unique probabilité invariante.

□

Les exemples précédents montrent par contre qu'il faut ajouter des hypothèses si on veut donner un résultat de convergence de la loi de X_n lorsque n tend vers $+\infty$. Nous allons donner ici des conditions suffisantes pour que la loi de X_n converge dans le cas où l'espace des états de la chaîne est fini ; le théorème suivant affirme que, pour une chaîne de Markov dont l'espace d'états est fini, si on peut trouver un entier $k \geq 1$ tel que l'on puisse passer en exactement k étapes de n'importe quel état i à n'importe quel état j avec probabilité strictement positive alors, la loi de X_n converge à une vitesse exponentielle vers son unique loi stationnaire :

Théorème 29 *Soit $(X_n)_n$ est une chaîne de Markov homogène à espace d'états $\mathcal{X}$* ***fini*** *et de matrice de transition Q vérifiant : il existe un entier $k \in \mathbb{N}^*$ telle que tous les coefficients de Q^k soient* ***strictement positifs****. Alors, la loi de X_n converge vers l'unique loi de probabilité μ invariante par Q à une vitesse exponentielle. Plus précisément, il existe $A > 0$ et $\rho \in [0, 1[$ tels que pour tout $n \in \mathbb{N}^*$ et pour tout $x, y \in \mathcal{X}$,*

$$|Q^n(y, x) - \mu(x)| \leq A\rho^n.$$

Preuve.

1. On suppose dans un premier temps que tous les coefficients de la matrice de transition Q sont strictement positifs. Notons N le nombre d'états de la chaîne et $\delta = \min(Q(i, j),\ i, j \in \mathcal{X})$. On a $\delta > 0$ et $N\delta \leq 1$ puisque $1 = \sum_j Q(i, j) \geq N\delta$. Pour $j \in \mathcal{X}$, on considère les suites $m_{j,n} = \min_i Q^n(i, j)$ et $M_{j,n} = \max_i Q^n(i, j)$. On montre qu'elle ont les propriétés suivantes :
   (i) $(m_{j,n})_n$ est une suite croissante et $(M_{j,n})_n$ est une suite décroissante.
   (ii) $M_{j,n} - m_{j,n} \leq (1 - N\delta)(M_{j,n} - m_{j,n})$ pour tout $n \in \mathbb{N}$.
   Donc, pour tout $j \in \mathcal{X}$, $(M_{j,n})_n$ et $(m_{j,n})_n$ ont une limite commune que l'on note $\pi(j)$ et pour tout n $m_{n,j} \leq \pi(j) \leq M_{n,j}$. On en déduit que pour tout $n \in \mathbb{N}$ et $i, j \in \mathcal{X}$, $|Q^n(i, j) - \pi(j)| \leq (1 - N\delta)^n$ (puisque $Q^n(i, j) - M_{n,j} \leq Q^n(i, j) - \pi(j) \leq Q^n(i, j) - m_{n,j}$). D'après la proposition 28, π définit une loi de probabilité invariante par Q et c'est la seule probabilité invariante par Q.
2. Cas général : il existe $k \in \mathbb{N}$ tel que tous les coefficients de Q^k soient strictement positifs.
   On déduit de la première partie que $M_{j,nk} - m_{j,nk} \leq (1 - N\eta)^n$ pour tout $n \in \mathbb{N}$ et $j \in \mathcal{X}$, avec $\eta = \min(Q^k(i, j),\ i, j \in \mathcal{X})$. Comme $(M_{n,j} - m_{n,j})_n$ est une suite décroissante, $M_{j,n} - m_{j,n} \leq (1 - N\eta)^{[n/k]} \leq (1 - N\eta)^{n/k-1}$. On en déduit que les deux suites $(M_{j,n})_n$ et $(m_{j,n})_n$ ont la même limite et que pour tout $i, j \in \mathcal{X}$ et $n \in \mathbb{N}$, $|Q^n(i, j) - \pi(j)| \leq A\rho^n$ avec $\rho = (1 - N\delta)^{1/k}$ et $A = \rho^{-1}$.

Montrons les propriétés (i) et (ii) :

- $m_{j,n+1} = \min_i \sum_k Q(i, k)Q^n(k, j) \geq \min_i \sum_k Q(i, k)m_{j,n} = m_{j,n}$.
  De même, $M_{j,n+1} \leq \max_i \sum_k Q(i, k)M_{j,n} = M_{j,n}$.
- Soit x et y deux états, posons $\Delta_n(x, y, z) = Q^n(x, z) - Q^n(y, z)$. On a

$$\Delta_{n+1}(x, y, z) = Q^{n+1}(x, z) - Q^{n+1}(y, z) = \sum_{i\in\mathcal{X}} (Q(x, i) - Q(y, i))Q^n(i, z).$$

  Notons $I = \{i \in \mathcal{X},\ Q(x, i) \geq Q(y, i)\}$. Alors,

$$\begin{aligned}\Delta_{n+1}(x, y, z) &\leq \sum_{i\in I} (Q(x, i) - Q(y, i))M_{z,n} + \sum_{i\notin I} (Q(x, i) - Q(y, i))m_{z,n} \\ &= \sum_{i\in I} (Q(x, i) - Q(y, i))(M_{z,n} - m_{z,n})\end{aligned}$$

  car $\sum_{i\in I}(Q(x, i) - Q(y, i)) + \sum_{i\notin I}(Q(x, i) - Q(y, i)) = 0$.
  Enfin, $\sum_{i\in I}(Q(x, i)-Q(y, i)) = 1-\sum_{i\notin I} Q(x, i)-\sum_{i\in I} Q(y, i) \leq 1-N\delta$, puisque $Q(k, l) \geq \delta$ pour tout $k, l \in \mathcal{X}$.
  On en déduit que $\Delta_n(x, y, z) \leq (1 - N\delta)(M_{z,n} - m_{z,n})$ pour tout $x, y, z \in \mathcal{X}$ et donc que $M_{z,n+1} - m_{z,n+1} \leq (1 - N\delta)(M_{z,n} - m_{z,n})$.

□

► *Exemple 47.* Si $(X_n)_n$ est une chaîne de Markov homogène sur $\{1, 2\}$ dont la matrice de transition est $Q = \begin{pmatrix} 1-p & p \\ q & 1-q \end{pmatrix}$ avec $0 < p + q < 2$, elle est récurrente positive. L'expression de de Q^n (voir l'exemple traité au paragraphe 9.3.2, page 41) montre que pour tout $x, y \in \{1, 2\}$, $\mid P(X_n = x \mid X_0 = y) - \mu(x) \mid\leq A\rho^n$ avec $\rho =\mid 1 - (p - q) \mid$ et $A = \frac{1}{p+q} \max(p, q)$ et μ sa loi de probabilité invariante.

▷ *Exercice 73.* Décrire le comportement asymptotique de la chaîne de Markov donnée dans l'exercice 70, page 40. Lorsqu'il y a eu un grand nombre de réplications, quelle est la probabilité d'observer sur une base donnée le même nucléotide que sur la séquence d'origine ?

▷ *Exercice 74.* Soit $(X_n)_n$ une chaîne de Markov homogène de matrice de transition Q. Posons $Y_n = X_{2n}$ pour tout $n \in \mathbb{N}$.
Montrer que $(Y_n)_n$ est une chaîne de Markov homogène de matrice de transition Q^2.
Application : on considère la chaîne de Markov décrite dans le paragraphe 9.3.3, page 42. Déterminer la matrice de transition de la chaîne de Markov $Y_n = X_{2n}$ et montrer que la loi de Y_n converge vers $\mu = (8/15, 0, 2/5, 0, 1/15)$.

▷ *Exercice 75.* (*suite de l'exercice 49, page 34*)

1. Que vaut $P(X_n = k)$ pour $k > n$? Pour tout $k \in \{0, \dots, n\}$, déterminer $P(X_n = k)$.

2. En déduire que la loi de X_n converge vers la loi géométrique sur $\mathbb{N}$ de paramètre 3/4.
3. montrer que $|P(X_n = k) - 3/4(1/4)^k| \leq C/4^n$ pour $k \in \mathbb{N}$ et $n \in \mathbb{N}$ où C est une constante. Cela montre que la vitesse de convergence de la loi de X_n vers la loi géométrique sur $\mathbb{N}$ de paramètre 3/4 est exponentielle.

▷ *Exercice 76.*

Considérons la dynamique suivante qui modélise l'évolution d'un écosystème méditerranéen[7]. A l'origine la forêt méditerranéenne, sur roche calcaire à faible altitude, était très certainement dominée par des chênes (chênes pubescents). Mais l'action de l'homme a éradiqué ces forêts primitives pour leur substituer des parcours pastoraux, des vergers ... Puis l'abandon de toute activité agricole, au lieu de conduire à la restauration naturelle de ces chênaies, a bien souvent favorisé l'implantation d'une autre espèce, le pin d'Alep, après un passage par un état de garrigue. Or ces forêts de substitution, hautement inflammables, subissent de manière récurrente le passage du feu (incendies volontaires ou non), le sol mis à nu se recouvre pendant un temps de pelouses ; ces forêts sont donc condamnées à une perpétuelle reconstitution. Pour étudier l'évolution à long terme de l'écosystème, on propose de modéliser cette évolution par une chaîne de Markov homogène $(X_n)_{n\in\mathbb{N}}$ à cinq états $e_1, \ldots, e_5$:

e_1 : chênaie, e_2 : zone de production agricole (vigne, vergers, ...), e_3 : pelouse, e_4 : garrigue, e_5 : pinède

Le temps est discrétisé en intervalles de temps de durée Δt (exprimée en années). Pour tout $n \in \mathbb{N}$, X_n désignera l'état de cet écosystème à l'instant $t_n = n\Delta t$. La matrice de transition de cette chaîne de Markov est la matrice Q définie par :

$$Q = \begin{pmatrix} 0.8 & 0.2 & 0 & 0 & 0 \\ 0 & 0.7 & 0.3 & 0 & 0 \\ 0 & 0 & 0.4 & 0.6 & 0 \\ 0 & 0 & 0 & 0.2 & 0.8 \\ 0.1 & 0 & 0.25 & 0 & 0.65 \end{pmatrix}$$

1. Représenter le graphe associé à la matrice de transition de $(X_n)_n$.
2. On note τ le plus petit entier $n \geq 1$ tel que $X_n \neq e_1$.
   Déterminer la loi de τ. Si à l'instant t_0, l'écosystème est une chênaie, combien de temps en moyenne restera-t-il dans cet état d'après le modèle ?
3. Suite à un incendie, l'écosystème est constitué d'une pelouse à l'instant t_n. Quelle est la probabilité qu'alors l'écosystème soit une pinède à l'instant t_{n+3} ?
4. Expliquer ce que retourne et trace la fonction `calcul` suivante :
```
calcul <- function(A,n){
  # n est un entier strictement plus grand que 1, A est une matrice carré
  R <- A; m <- numeric(n-1)
  for(i in 2:n){
    B <-R%*%A
    m[i-1] <- max(abs(B-R))
    R <- B
  }
  par(lty=2,pch='+')
  plot(x=1:(n-1), y=log(m), type='b', xlab='n', ylab='')
  return(list(mat=R,val=m))
}
```
5. Utiliser la fonction `calcul(A,n)` avec pour `A` la matrice de transition de $(X_k)_k$ et `n`=50. Qu'observe-t-on ? Quelle information cela donne-t-il sur l'évolution en loi de la chaîne de Markov ?
6. Les hypothèses du théorème sont-elles satisfaites par $(X_n)_n$?
7. Si on suppose que l'écosystème méditerranéen évolue suivant ce modèle, les proportions de terrains occupés par des chênes, des pinèdes et de la garrigue vont-elles se stabiliser au bout d'un certain nombre d'années ? Si oui vers quelles valeurs ? (la réponse doit être justifiée en détail)
8. Ecrire une fonction `recosyst`(x0,n) qui simule une réalisation de $X_1, \ldots, X_n$ en supposant que X_0=x0. Cette fonction devra retourner un vecteur x de taille $n + 1$ de sorte que le coefficient i de x soit la réalisation obtenue de X_{i-1} pour tout $1 \leq i \leq n$.
   Utiliser la fonction suivante pour simuler et tracer une réalisation de $X_0, X_1, \ldots, X_n$ avec $n = 50$ et $X_0 = 1$:
```
dessintraj <- function(x0,n){
  X <- recosyst(x0,n)
  par(lty=2,pch='+')
  plot(x=0:n, y=X, type='b', xlab='n', ylab='Xn')
}
```
   Si on interprète la courbe comme décrivant l'évolution sur $50\Delta t$ années de l'état d'une parcelle recouverte de chênes dans une zone méditérranéenne, combien de temps cette parcelle restera-t-elle une chênaie ? Combien de fois la parcelle sera-t-elle recouverte d'une pinède et ensuite se retrouvera au stade pelouse ?
9. On considère la fonction `trace` suivante. Expliquer ce que contient le coefficient (i, j) de la matrice `R` à la fin de l'exécution de la fonction

[7]Cet exercice est inspiré d'un exemple tiré du livre *Modélisation et simulation d'écosystèmes*, P. Coquillard et D. Hill, Masson 1997.

```
 trace <- function(x0,n){
 R <- matrix(data=0, ncol=5, nrow=n+1)
 X <- recosyst(x0,n)
 for (j in 1:5){
   for(i in 1:(n+1)){
      R[i,j] <- mean(X[1:i]==j)
   }
 }
 matplot(x=0:n, y=R,type='s',xlab='n',ylab=' ')
 return(R[n+1,])
}
```

Exécuter la commande `trace1(x0=1,n=5000)`. Recommencer en changeant la valeur de `x0`. Qu'observe-t-on ? Commenter les figures et les valeurs affichées.

10 Temps d'atteinte d'un état

Dans toute la section, $(X_n)_n$ désignera une chaîne de Markov homogène d'espace d'états $\mathcal{X}$ fini ou dénombrable de matrice de transition Q : on pourra interpréter X_n comme modélisant l'état d'un système à l'instant n.

La loi initiale de la chaîne (X_n) ayant été fixée, seuls interviennent dans l'étude de l'évolution de la chaîne les états susceptibles d'être atteints. On pose

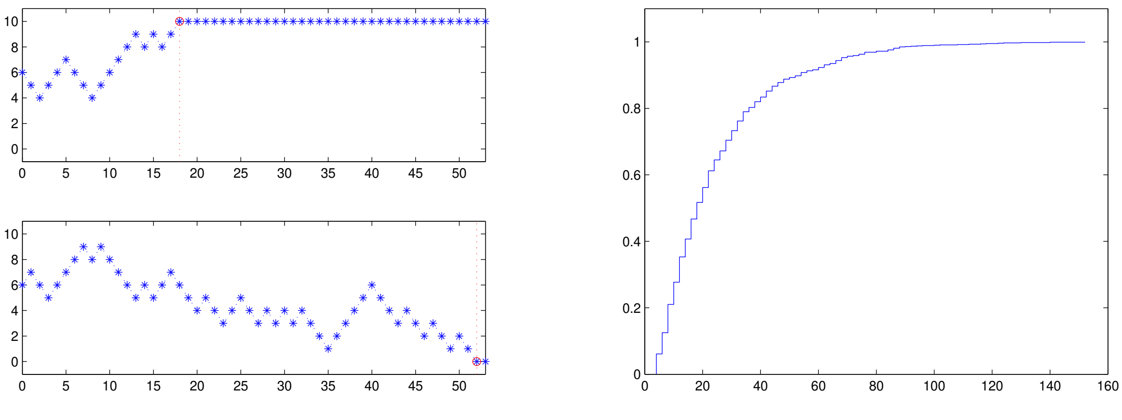
$$\mathcal{X}_a = \{x \in \mathcal{X}, \text{ il existe } n \geq 0, \; P(X_n = x) > 0\}.$$

Pour un état $e \in \mathcal{X}_a$, on notera $T_e^{(n)}$ le temps qu'il faut à la chaîne pour atteindre l'état e strictement après l'instant n :

- $T_e^{(n)}$ désigne donc le plus petit entier $k > 0$ tel que $X_{n+k} = e$ si la chaîne passe par e après l'instant n
- $T_e^{(n)} = +\infty$ si la chaîne ne passe pas par l'état e après l'instant n.

Pour simplifier les notations $T_e^{(0)}$ sera aussi noté simplement T_e.

► *Exemple 48.* (*suite de l'exemple 36*)
Dans le cas de la chaîne de Markov décrivant l'évolution de la fortune du joueur, $\{T_0 < +\infty\}$ désigne l'événement "le joueur se ruine". Dans le cas où le jeu s'arrête à cause de la ruine du joueur, T_0 représente le nombre de parties qu'il a jouées. De façon générale, $\min(T_0, T_m)$ est le nombre de parties d'un jeu qui se termine si le joueur n'a plus d'argent ou si il a atteint la somme m.



Les deux figures à gauche montrent l'évolution de la fortune de deux joueurs qui jouent indépendamment l'un et l'autre à un même jeu équitable ; ils commencent tous les deux avec 6 euros et cherchent à atteindre la somme de 10 euros.
Le premier joueur atteint effectivement la somme de 10 euros après avoir joué 18 fois : la valeur de T_{10} est 18 et la valeur de T_0 est $+\infty$.
Le second joueur n'atteint pas la somme de 10 euros et perd son dernier euro à la 52-ème partie : la valeur de T_{10} est $+\infty$ et celle de T_0 est 52.
Pour étudier la loi du nombre de parties $T = \min(T_0, T_m)$, on a fait 1000 simulations de l'évolution de la fortune au cours d'un tel jeu. Sur les 1000 jeux simulés, 428 jeux ont abouti à la ruine du joueur, le nombre moyenne de parties pour un jeu a été de 24,8 parties. Le jeu qui a duré le plus longtemps a compté 152 parties. La figure de droite montre la fonction de répartition empirique des 1000 réalisations obtenues de T.

▶ *Exemple 49.* (*suite de l'exemple 38*)
Dans le cas de la chaîne de Markov décrivant l'évolution des génotypes de plantes obtenues par autofécondations successives à partir d'une plante hétérozygote, $\{T_{AA} < +\infty\}$ est l'événement "on a obtenu après un nombre fini d'autofécondations une plante de génotype AA". Si cet événement est réalisé, T_{AA} représente le numéro de la première génération où on obtient une plante de génotype AA.

10.1 Probabilité d'atteinte d'un état

Proposition 30 *Soit $i, e \in \mathcal{X}_a$.*

– *Pour tout $n \in \mathbb{N}$ et $k \in \mathbb{N}^*$, la probabilité d'atteindre l'état e à l'instant $n+k$ pour la première fois après l'instant n sachant que $X_n = i$ ne dépend pas de n, on la notera $f_{i,e}(k) = P(T_e^{(n)} = k|X_n = i)$. Elle vérifie l'équation suivante :*

$$f_{i,e}(1) = Q(i,e) \text{ et } f_{i,e}(k) = \sum_{j \in \mathcal{X}\setminus\{e\}} Q(i,j) f_{j,e}(k-1) \text{ pour tout } k \geq 2. \tag{4}$$

– *La probabilité d'atteindre l'état e après l'instant n, sachant que $X_n = i$ ne dépend pas de n, on la notera $f_{i,e}$: $f_{i,e} = P(T_e^{(n)} < +\infty|X_n = i)$. Elle vérifie :*

$$f_{i,e} = Q(i,e) + \sum_{j \in \mathcal{X}\setminus\{e\}} Q(i,j) f_{j,e}. \tag{5}$$

N.B.

- L'équation pour $f_{i,e}(k)$ traduit simplement le fait que pour arriver pour la première fois dans l'état e en k pas, en partant de l'état i, il faut aller de l'état i à un état $j \neq e$, puis partant de j, arriver pour la première fois en e en $k-1$ pas.
- De même, l'équation pour $f_{i,e}$ traduit le fait que la chaîne atteint e à partir d'un état i soit directement, soit passe de l'état i à un état $j \neq e$ puis atteint e à partir de l'état j.
- $f_{i,e} = 1$ si et seulement si $f_{j,e} = 1$ pour tout état $j \neq e$ tel que $Q(i,j) > 0$.

▶ *Exemple 50.* (*suite de l'exemple 36*)
Supposons que le joueur dispose de 1 euro et joue à un jeu de hasard où la mise est de 1 euro jusqu'à ce qu'il ait atteint la somme de 3 euros ou jusqu'à ce qu'il soit ruiné. Rappelons que p désigne la probabilité qu'il gagne une partie et donc gagne 1 euro et $1-p$ est la probabilité qu'il perde la partie et donc perde 1 euro. On cherche la probabilité qu'il réussisse ainsi à avoir 3 euros, c'est-à-dire $f_{1,3}$ avec les notations de la proposition 30. L'équation (4) avec $i = 1$ et $e = 3$ s'écrit $f_{1,3} = pf_{2,3} + (1-p)f_{0,3}$. Déjà $f_{0,3} = 0$ puisque le joueur ne peut pas jouer s'il n'a pas d'argent initialement. Il reste à écrire l'équation (4) pour $f_{2,3}$ qui est la probabilité qu'un joueur arrive à obtenir 3 euros s'il a au départ 2 euros. On obtient : $f_{2,3} = 1 - p + (1-p)f_{1,3}$. On a donc à résoudre un système de 2 équations à 2 inconnues : $\begin{cases} f_{1,3} = pf_{2,3} \\ f_{2,3} = p + (1-p)f_{1,3} \end{cases}$.
Ce système a une seule solution qui est $f_{2,3} = \frac{p}{1-p(1-p)}$ et $f_{1,3} = \frac{p^2}{1-p(1-p)}$.
En particulier, si le jeu est équitable c'est-à-dire si $p = 1/2$, on a $f_{1,3} = 1/3$ ce qui signifie que si le joueur a initialement 1 euro, il a une chance sur trois d'arriver à obtenir 3 euros. La probabilité que le jeu s'arrête car le joueur est ruiné se calcule de façon analogue en écrivant les équations satisfaites par $f_{1,0}$, $f_{2,0}$ et $f_{3,0}$ et en résolvant le système obtenu. Ce calcul permet de voir que $f_{1,0} + f_{1,3} = 1$, ce qui signifie que le joueur s'arrête forcément au bout d'un nombre fini de parties soit parce qu'il a réussi à obtenir 3 euros, soit parce qu'il est ruiné.

Preuve de la proposition 30.

– On a $\{T_e^{(n)} = 1\} = \{X_{n+1} = e\}$ et pour $k \geq 2$, $\{T_e^{(n)} = k\} = \{X_{n+k} = e, X_{n+k-1} \neq e, \ldots, X_{n+1} \neq e\}$. En décomposant cet événement en fonction de la valeur de X_n, et en utilisant que $(X_n)_n$ est une chaîne de Markov homogène, on obtient :

$$\begin{aligned} P(T_e^{(n)} = k|X_n = i) &= \sum_{j \in \mathcal{X}\setminus\{e\}} P(X_{n+k} = e, X_{n+k-1} \neq e, \ldots, X_{n+2} \neq e,\ X_{n+1} = j|X_n = i) \\ &= \sum_{j \in \mathcal{X}\setminus\{e\}} P(X_{n+k} = e, X_{n+k-1} \neq e, \ldots, X_{n+2} \neq e|X_{n+1} = j,\ X_n = i) P(X_{n+1} = j|X_n = i) \\ &= \sum_{j \in \mathcal{X}\setminus\{e\}} P(X_{n+k-1} = e, X_{n+k-2} \neq e, \ldots, X_{n+1} \neq e|X_n = j) Q(i,j) \\ &= \sum_{j \in \mathcal{X}\setminus\{e\}} Q(i,j) P(T_e^{(n)} = k-1|X_n = j) \end{aligned}$$

Cette égalité et le fait que $P(T_e^{(n)} = 1|X_n = i) = Q(i,e)$ permet de montrer que pour tout $k \in \mathbb{N}^*$, $P(T_e^{(n)} = k|X_n = i)$ ne dépend pas de n.

– Comme $P(T_e^{(n)} < +\infty | X_n = i) = \sum_{k=1}^{+\infty} P(T_e^{(n)} = k | X_n = i)$, la probabilité d'atteindre e après l'instant n sachant que $X_n = i$ ne dépend pas non plus de n. En utilisant l'expression récursive pour $f_{i,j}(k)$, on a :

$$f_{i,e} = Q(i,e) + \sum_{k=2}^{+\infty} \sum_{j \in \mathcal{X} \setminus \{e\}} Q(i,j) f_{j,e}(k-1).$$

Comme tous les termes de la double somme sont positifs, on peut intervertir les deux sommes, ce qui donne l'équation (5).

□

10.2 Temps moyen d'atteinte d'un état

Il est intéressant d'avoir des renseignements sur le temps moyen que met la chaîne de Markov pour atteindre un état accessible e à partir d'un état i, c'est-à-dire sur $E(T_e^{(n)} \mid X_n = i)$. Remarquons déjà que si il y a une probabilité strictement positive à partir de i de ne jamais atteindre l'état e alors $E(T_e^{(n)} \mid X_n = i) = +\infty$. Dans le cas contraire, on pourra déterminer le temps moyen que met la chaîne de Markov pour atteindre un état accessible e à partir d'un état i en résolvant un système d'équation linéaires comme l'indique le résultat suivant :

Corollaire 31 *Soit $i, e \in \mathcal{X}_a$. $E(T_e^{(n)} \mid X_n = i)$ ne dépend pas de n, on le notera $m_{i,e}$. On a*

$$m_{i,e} = 1 + \sum_{j \in \mathcal{X} \setminus \{e\}} Q(i,j) m_{j,e} \tag{6}$$

(les deux termes de l'égalité pouvant être infinis)

Preuve. Supposons d'abord que $f_{i,e} = 1$. Comme la loi conditionnelle de $T_e^{(n)}$ sachant que $X_n = i$ ne dépend pas de n, il en est de même de $E(T_e^{(n)} \mid X_n = i)$ que l'on notera $m_{i,e}$. Comme $m_{i,e} = \sum_{k=1}^{+\infty} k f_{i,e}^{(k)}$, on obtient grâce à l'expression récursive de $f_{i,e}(k)$:

$$\begin{aligned}
m_{i,e} &= Q(i,e) + \sum_{k=2}^{+\infty} \sum_{j \in \mathcal{X} \setminus \{e\}} Q(i,j) k f_{j,e}(k-1) \\
&= Q(i,e) + \sum_{j \in \mathcal{X} \setminus \{e\}} Q(i,j) \sum_{k=2}^{+\infty} f_{j,e}(k-1) + \sum_{j \in \mathcal{X} \setminus \{e\}} Q(i,j) \sum_{k=2}^{+\infty} (k-1) f_{j,e}(k-1) \\
&= Q(i,e) + \sum_{j \in \mathcal{X} \setminus \{e\}} Q(i,j) \sum_{\ell=1}^{+\infty} f_{j,e}(\ell) + \sum_{j \in \mathcal{X} \setminus \{e\}} Q(i,j) \sum_{\ell=1}^{+\infty} \ell f_{j,e}(\ell) \\
&= Q(i,e) + \sum_{j \in \mathcal{X} \setminus \{e\}} Q(i,j) f_{j,e} + \sum_{j \in \mathcal{X} \setminus \{e\}} Q(i,j) m_{j,e}
\end{aligned}$$

Il suffit alors d'utiliser l'équation (5) pour établir l'égalité annoncée : $m_{i,e} = 1 + \sum_{j \in \mathcal{X} \setminus \{e\}} Q(i,j) m_{j,e}$.
Il reste à traiter le cas où $f_{i,e} < 1$. Dans ce cas $m_{i,e} = +\infty$ et il existe un état $j \neq e$ tel que $Q(i,j) > 0$ et $f_{j,e} < 1$. Pour cet état j, on a $Q(i,j) m_{j,e} = +\infty$. Donc, l'égalité est encore vraie si $f_{i,e} < 1$. □

▶ *Exemple 51.* Considérons une chaîne de Markov homogène $(X_n)_n$ à deux états $\mathcal{X} = \{1, 2\}$ dont la matrice de transition est $Q = \begin{pmatrix} 1-p & p \\ q & 1-q \end{pmatrix}$ avec $0 < p, q < 1$.
On a $f_{1,1} = f_{1,2} = f_{2,1} = f_{2,2} = 1$, $m_{2,1} = \frac{1}{q}$, $m_{1,1} = 1 + \frac{p}{q}$, $m_{1,2} = \frac{1}{p}$ et $m_{2,2} = 1 + \frac{q}{p}$.

10.3 Temps d'atteinte d'un sous-ensemble d'états

L'étude faite pour $T_e^{(n)}$ peut s'étendre au temps $T_A^{(n)}$ que met la chaîne de Markov à atteindre un sous-ensemble A d'états strictement après l'instant n :

$$T_A^{(n)} = \inf(\{k > 0,\ X_{k+n} \in A\}) = \min(T_e^{(n)}, e \in A).$$

Proposition 32 *Soit i un état accessible.*

1. *Pour tout $n \in \mathbb{N}$ et $k \in \mathbb{N}^*$, la probabilité d'atteindre le sous-ensemble A à l'instant $n+k$ pour la première fois après l'instant n sachant que $X_n = i$ ne dépend pas de n, on la notera $f_{i,A}(k) = P(T_A^{(n)} = k | X_n = i)$. Elle vérifie l'équation suivante :*

$$f_{i,A}(1) = \sum_{a \in A} Q(i,a) \text{ et } f_{i,A}(k) = \sum_{j \in \mathcal{X} \setminus A} Q(i,j) f_{j,A}(k-1) \text{ pour tout } k \geq 2. \tag{7}$$

2. *La probabilité d'atteindre le sous-ensemble A après l'instant n, sachant que $X_n = i$ ne dépend pas de n, on la notera $f_{i,A}$: $f_{i,A} = P(T_A^{(n)} < +\infty | X_n = i)$. Elle vérifie :*

$$f_{i,A} = \sum_{a \in A} Q(i,a) + \sum_{j \in \mathcal{X} \setminus A} Q(i,j) f_{j,A}. \tag{8}$$

3. *$E(T_A^{(n)} \mid X_n = i)$ ne dépend pas de n, on le notera $m_{i,A}$. On a*

$$m_{i,A} = 1 + \sum_{j \in \mathcal{X} \setminus A} Q(i,j) m_{j,A} \tag{9}$$

(les deux termes de l'égalité pouvant être infinis)

▶ *Exemple 52.* (suite de l'exemple 50) Déterminons la durée moyenne du jeu (exprimée en parties) pour un joueur possédant 1 euro initialement et voulant obtenir 3 euros. Cela revient à chercher $m_{1,A} = E(T_A | X_0 = 1)$ avec $A = \{0, 3\}$. On écrit les équations (9) pour $m_{1,A}$ et $m_{2,A}$: $\begin{cases} m_{1,A} = 1 + p m_{2,A} \\ m_{2,A} = 1 + (1-p) m_{1,A}. \end{cases}$ En résolvant ce système, on obtient que $m_{1,A} = \frac{1+p}{1-p(1-p)}$ et $m_{1,A} = \frac{2-p}{1-p(1-p)}$. En particulier, lorsque le jeu est équitable, l'espérance du nombre de parties est égale à 2, que le joueur ait initialement 1 ou 2 euros.

10.4 Les chaînes absorbantes

Définition. On dit qu'un état e est *absorbant* si $Q(e,e) = 1$ (si la chaîne entre dans cet état, elle y reste avec probabilité 1).

▶ *Exemple 53.* (suite de l'exemple 38). Pour la chaîne de Markov décrivant les génotypes des plantes obtenues par autofécondations successives, les états correspondants aux génotypes AA et aa sont des états absorbants.

Nous allons décrire quelques propriétés simples des chaînes de Markov ayant des états absorbants. Mais avant cela, introduisons une définition :

Définition. On dit qu'un état i conduit à un état j (ou encore que j est accessible à partir de i) s'il existe un chemin dans le graphe associé à la matrice de transition qui permet de passer de i à j. Autrement dit, i conduit à j s'il existe un entier n tel que $(Q^n)(i,j) > 0$.

Proposition 33 *Considérons une chaîne de Markov homogène $(X_n)_n$ d'espace d'états $\mathcal{X}$ qui a au moins un état absorbant. Notons A l'ensemble des états absorbants de la chaîne.*

(i) Si e est un état qui conduit à un état absorbant, alors la chaîne de Markov ne passera qu'un nombre fini de fois dans l'état e et donc $Q^n(i,e)$ tend vers 0 lorsque n tend vers $+\infty$ quel que soit l'état i,

(ii) Si le nombre d'états est <u>fini</u> et si tous les états conduisent à un état absorbant, alors
- *$P(X_n \in A) \underset{n \to +\infty}{\to} 1$;*
- *avec probabilité 1, la chaîne atteint un état absorbant en temps fini (i.e. $P(T_A < +\infty) = 1$) ;*
- *il existe des constantes $b > 0$ et $0 < c < 1$ tel que pour tout n et pour tout états $i, j \notin A$, $(Q^n)(i,j) \leq bc^n$.*

N.B. $P(T_A < +\infty) = \lim_{n \to +\infty} P(T_A \leq n) = \lim_{n \to +\infty} P(X_n \in A)$.

▶ *Exemple 54.* (suite de l'exemple 36) Quelle que soit la fortune m que le joueur cherche à atteindre, la proposition 33 dit que le jeu s'arrêtera après un nombre fini de parties (en effet, (ii) s'applique puisque tous les états $0 < i < m$ conduisent aux deux absorbants 0 et m).

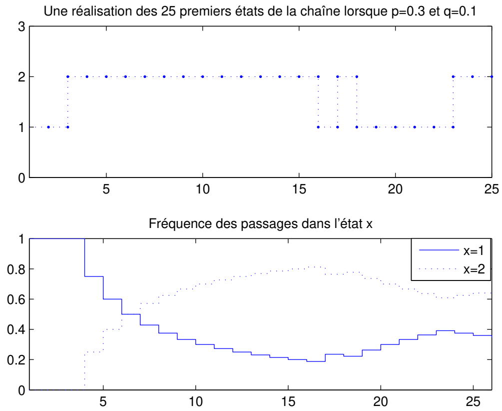
▶ *Exemple 55.* (suite de l'exemple 38). La proposition 33 dit seulement que la probabilité d'avoir une plante hétérozygote à la n-ième génération tend vers 0 lorsque n tend vers $+\infty$.

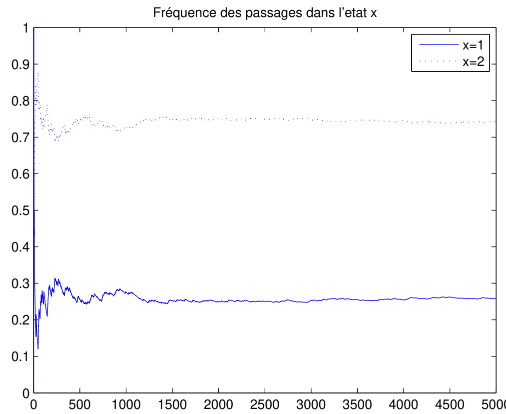
11 Fréquence de passages dans un état et estimation

Dans toute cette section, $(X_n)_{n\geq 0}$ désigne une chaîne de Markov homogène d'espace d'états $\mathcal{X}$ fini ou dénombrable et de matrice de transition Q.
Pour un état x et un entier $n \in \mathbb{N}$, notons $V_n(x)$ le nombre de fois où la chaîne de Markov visite l'état x au cours des n premières étapes : $V_n(x) = \sum_{k=0}^{n-1} \mathbb{1}_{\{X_k=x\}}$. Si $X_0, X_1, \ldots, X_n$ décrivent les états successifs d'un système aux instants $0, 1, \ldots, n$, la variable aléatoire $\frac{V_n(x)}{n}$ correspond à la proportion de temps que le système passe dans l'état x avant l'instant n. C'est une variable aléatoire qui est facilement observable. Nous allons décrire le comportement de $\frac{1}{n}V_n(x)$ lorsque n tend vers $+\infty$.

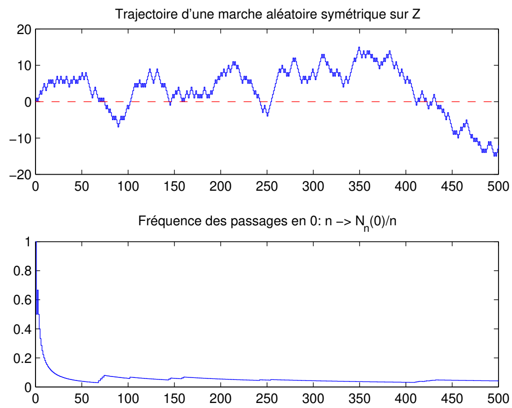
11.1 Exemples d'évolutions de la fréquence des passages dans un état

► *Exemple 56.* Les figures ci-dessous concernent la chaîne de Markov homogène à deux états (voir exemple 42, page 39) avec $p = 0.3$ et $q = 0.1$: les figures de gauche montrent une réalisation de $(X_0, X_1, \ldots, X_{25})$ et en dessous les valeurs de $(V_1(x), \frac{V_2(x)}{2}, \ldots, \frac{V_{26}(x)}{26})$ associées pour les deux états $x = 1$ et $x = 2$. La figure de droite a été obtenue en simulant les 5000 premiers pas de la chaîne de Markov et en représentant les réalisations de $n \mapsto \frac{V_n(x)}{n}$ obtenues pour $x = 1$ et $x = 2$.





► *Exemple 57.* Les figures ci-contre montrent une réalisation des 500 premiers pas d'une marche aléatoire symétrique sur $\mathbb{Z}$ (voir exemple 67) et en dessous la réalisation correspondante de la fréquence des passages en 0 i.e. $(\frac{V_1(0)}{1}, \ldots, \frac{1}{500}V_{500}(0))$.



► *Exemple 58.* (suite de l'exemple 44 portant sur la modélisation markovienne d'une séquence d'ADN de longueur K)
Si X_ℓ décrit le ℓ-ème nucléotide en partant de l'extrêmité 5' d'une séquence d'ADN, $\frac{V_n(a)}{n}$ correspond à la proportion d'adénines parmi les n premiers nucléotides.

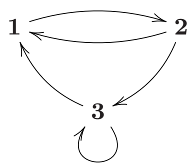
11.2 Chaînes de Markov irréductibles

Pour simplifier la description du comportement de $\frac{V_n(x)}{n}$ lorsque n tend vers $+\infty$, nous allons considérer les chaînes de Markov homogènes irréductibles.

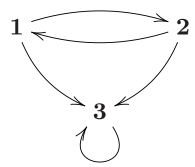
Définition. Une chaîne de Markov est dite *irréductible* si pour tout couple d'états différents i et j, i conduit à j c'est-à-dire s'il existe un chemin de longueur finie dans le graphe associé à la matrice de transition de la chaîne permettant de passer de i à j. Autrement dit, une chaîne de Markov est irréductible si pour tout couple d'états i et j avec $i \neq j$, il existe un entier $k_{i,j}$ tel que la chaîne ait une probabilité positive de passer de i à j en $k_{i,j}$ étapes ($Q^{k_{i,j}}(i,j) > 0$).

▶ *Exemple 59.*

Graphe (a)



Graphe (b)



Le graphe (a) donne un exemple de graphe associé à la matrice de transition d'une chaîne de Markov homogène irréductible.
Le graphe (b) donne un exemple de graphe associé à la matrice de transition d'une chaîne de Markov homogène qui n'est pas irréductible : il n'existe pas de chemin partant de l'état 3 et aboutissant à l'état 1 par exemple.

▶ *Exemple 60.* La marche aléatoire symétrique sur $\mathbb{Z}$ est irréductible (voir exemples 67 et 57).

▶ *Exemple 61.*

1. Si tous les coefficients de Q sont strictement positifs alors la chaîne est irréductible.
2. Une chaîne de Markov qui a au moins deux états et dont l'un des états est absorbant n'est pas irréductible.

11.3 Comportements asymptotiques

Soit x un état. Notons $(\tilde{X}_n)_n$ une chaîne de Markov homogène de même matrice de transition Q que $(X_n)_n$ et partant de x à l'instant 0. Grâce à la propriété de Markov et l'homogénéité en temps, si la chaîne de Markov $(X_n)_n$ passe par un état x pour la première fois à un instant aléatoire alors le temps noté R_x qu'elle mettra pour revenir dans l'état x a même loi que le temps noté $\tilde{T}_x$ que la chaîne $(\tilde{X}_n)_n$ met pour retourner en x en étant partie de x. En particulier, $E(R_x|(X_n)_n$ passe par $x) = E(\tilde{T}_x)$; cette espérance conditionnelle correspond au temps moyen que met la chaîne $(X_n)_n$ une fois passée en x pour y revenir ; elle est aussi égale à $m_{x,x}$ avec les notations du corollaire 31. De plus, si la chaîne est déjà passée deux fois (resp. trois fois, ..., 10 fois ...) dans l'état x, le temps qu'elle mettra pour y revenir une nouvelle fois aura la même loi que $\tilde{T}_x$.
En regardant l'évolution de la chaîne sur une très longue durée, on a les propriétés suivantes que l'on admettra :

Théorème 34 *Supposons que $(X_n)_n$ soit une chaîne de Markov homogène* ***irréductible*** *d'espace d'états $\mathcal{X}$ fini ou dénombrable.*

1. *Pour tout $x \in \mathcal{X}$, $\frac{1}{n}V_n(x)$ tend vers $\frac{1}{m_{x,x}}$ avec probabilité 1 lorsque n tend vers $+\infty$ (avec la convention $\frac{1}{m_{x,x}} = 0$ si $m_{x,x} = +\infty$).*
2. *Si de plus la chaîne admet une probabilité invariante μ, alors elle est unique et vérifie $\frac{1}{m_{x,x}} = \mu(x)$ pour tout $x \in \mathcal{X}$.*
3. *Si de plus la chaîne n'admet pas de probabilité invariante, alors $m_{x,x} = +\infty$ pour tout $x \in \mathcal{X}$.*

11.4 Estimation des paramètres

On considère dans cette section des chaînes de Markov homogènes ayant un nombre fini d'états. Les chaînes de Markov à espace d'états fini admettent au moins une loi de probabilité invariante. D'après le théorème 34, si cette chaîne de Markov est en plus irréductible alors elle admet une unique probabilité invariante et on peut donner une estimation des coefficients de cette loi invariante à partir de l'observation des états successifs pris par la chaîne de Markov pendant un temps long. Plus généralement, on a le résultat suivant :

Théorème 35 *Soit $(X_n)_n$ une chaîne de Markov homogène* ***irréductible*** *d'espace d'états* ***fini*** *noté $\mathcal{X}$. Pour $x \in \mathcal{X}$ et $n \in \mathbb{N}^*$, posons $V_n(x)$ le nombre de fois où la chaîne passe en x au cours des n premiers instants : $V_n(x) = \sum_{i=0}^{n-1} \mathbb{1}_{\{X_i = x\}}$.*

1. *Elle admet une unique probabilité invariante que l'on note* μ *;*
2. *Pour tout* $x \in \mathcal{X}$*, avec probabilité 1,* $\frac{1}{n}V_n(x)$ *tend vers* $\mu(x)$ *lorsque* n *tend vers* $+\infty$ *;*
3. *Plus généralement, pour toute fonction* $f : \mathcal{X} \to \mathbb{R}$ *,* $\frac{1}{n}\sum_{k=0}^{n-1} f(X_k)$ *converge vers* $\sum_{x\in\mathcal{X}} f(x)\mu(x)$ *avec probabilité* 1 *lorsque* n *tend vers* $+\infty$*.*

Pour obtenir des estimateurs des coefficients de la matrice de transition, on applique le théorème 35 à la suite $(Y_n)_n$ définie par $Y_n = (X_n, X_{n+1})$ pour tout $n \geq 0$:

Lemme 36 *Soit* $(X_n)_n$ *une chaîne de Markov homogène irréductible d'espace d'états* $\mathcal{X} = \{e_1, \ldots, e_r\}$ *et de matrice de transition* Q*. Pour tout* $n \in \mathbb{N}$*, posons* $Y_n = (X_n, X_{n+1})$*.*

1. *La suite* $(Y_n)_n$ *est chaîne de Markov homogène de matrice de transition* T *définie par :*
$$T((x_1, x_2); (y_1, y_2)) = Q(y_1, y_2)\mathbb{1}_{\{x_2=y_1\}} \text{ pour tout } x_1, x_2, y_1, y_2 \in \mathcal{X}$$
2. *La suite* $(Y_n)_n$ *restreinte à l'ensemble* $\mathcal{E} = \{(i, j) \in \mathcal{X},\ Q(i, j) > 0\}$ *est une chaîne de Markov homogène irréductible et admet comme loi de probabilité invariante la probabilité* ν *sur* $\mathcal{E}$ *définie par :* $\nu(y) = \mu(y_1)Q(y_1, y_2)$ *pour tout* $y = (y_1, y_2) \in \mathcal{E}$*.*

Corollaire 37 *Soit* $(X_n)_n$ *une chaîne de Markov homogène et irréductible à valeurs dans un ensemble fini* $\mathcal{X} = \{1, \ldots, N\}$*. Notons* Q *sa matrice de transition et* μ *sa loi de probabilité invariante.*
Pour $x, y \in \mathcal{X}$ *et* $n \in \mathbb{N}^*$*, posons* $V_n(x) = \sum_{k=0}^{n-1} \mathbb{1}_{\{X_k=x\}}$ *et* $V_n(x, y) = \sum_{k=0}^{n-1} \mathbb{1}_{\{(X_k,X_{k+1})=(x,y)\}}$*.*
- *Pour tout* $x \in \mathcal{X}$*,* $\frac{1}{n}V_n(x) \underset{n\to+\infty}{\longrightarrow} \mu(x)$ *avec probabilité 1.*
- *Pour tout* $x, y \in \mathcal{X}$*,* $\frac{V_n(x,y)}{V_n(x)} \underset{n\to+\infty}{\longrightarrow} Q(x, y)$ *avec probabilité 1.*

Application à l'estimation Pour tout $x, y \in \mathcal{X}$, $\hat{Q}_n(x, y) = \frac{V_n(x,y)}{V_n(x)}$ définit un estimateur du coefficient $Q(x, y)$ construit à partir des observations des n premiers états $X_0, \ldots, X_{n-1}$ de la chaîne de Markov. On dit que cet estimateur[8] est *fortement consistant* car $\frac{V_n(x,y)}{V_n(x)}$ converge vers $Q(x, y)$ avec probabilité 1 lorsque n tend vers $+\infty$. La matrice $\hat{Q}_n$ définie ainsi est une matrice stochastique. Le vecteur $\hat{\mu}_n = (\frac{V_n(e_1)}{n}, \ldots, \frac{V_n(e_r)}{n})$ définit un estimateur fortement consistant de la loi invariante μ par Q.
Remarquons que $\hat{\nu}_n$ définit une probabilité qui n'est pas invariante par la matrice $\hat{Q}_n$: $\hat{\nu}_n\hat{Q}_n = (\frac{\tilde{V}_n(e_1)}{n}, \ldots, \frac{\tilde{V}_n(e_r)}{n})$ où $\tilde{V}_n(x) = \sum_{i=1}^{n} \mathbb{1}_{\{X_i=x\}}$ correspond au nombre de fois où la chaîne de Markov passe par x entre les instants 1 à n. Comme $|\tilde{V}_n(x) - V_n(x)| \leq 2$ pour tout x, $\max(|\hat{\nu}_n\hat{Q}_n(x) - \hat{\nu}_n(x)|,\ x \in \mathcal{X}) \leq \frac{2}{n}$.

Démonstration du corollaire 37. Pour démontrer la seconde assertion, on applique le théorème 35 à la chaîne de Markov (Y_n) définie par $Y_n = (X_n, X_{n+1})$ pour tout n. On obtient que pour tout $i, j \in \mathcal{X}$ tel que $Q(i, j) > 0$, $\frac{V_n(i,j)}{n}$ tend vers $\mu(i)Q(i, j)$ avec probabilité 1. Comme $\frac{V_n(i)}{n}$ tend vers $\mu(i)$ avec probabilité 1, on en déduit que $\frac{V_n(i,j)}{V_n(i)}$ tend vers $Q(i, j)$ avec probabilité 1. □

Démonstration du lemme 36. Soit $n \in \mathbb{N}$. Pour tout $y_1 = (y_1^{(1)}, y_1^{(2)}) \in \mathcal{X}$, $\ldots$, $y_{n-1} = (y_{n-1}^{(1)}, y_{n-1}^{(2)}) \in \mathcal{X}$ $i = (i_1, i_2) \in \mathcal{X}$ et $j = (j_1, j_2) \in \mathcal{X}$, on a

$$\begin{aligned} P(Y_{n+1} = j \mid Y_n = i, Y_{n-1} = y_{n-1}, \ldots, Y_0 = y_0) &= P(X_{n+2} = j_2 \mid X_{n+1} = i_2, X_n = i_1, \ldots, X_0 = y_0^{(1)})\mathbb{1}_{\{j_1=i_2\}}. \\ &= Q(i_2, j_2)\mathbb{1}_{\{j_1=i_2\}}. \end{aligned}$$

Donc, $(Y_n)_n$ est une chaîne de Markov homogène de matrice de transition T. On restreint l'ensemble des états de la chaîne à l'ensemble $\mathcal{E} = \{(i, j) \in \mathcal{X},\ Q(i, j) > 0\}$ puisque $P(Y_n = y) = P(X_n = y^{(1)})Q(y^{(1)}, y^{(2)}) = 0$ si $y \notin \mathcal{E}$.
Comme la chaîne $(X_n)_n$ est irréductible, il en est de même pour la chaîne $(Y_n)_n$ sur $\mathcal{E}$: en effet, soit $y, z \in \mathcal{E}$, on a $Q(y^{(1)}, y^{(2)}) > 0$ et $Q(z^{(1)}, z^{(2)}) > 0$. De plus, il existe un chemin $y^{(2)} = u_0, \ldots, u_n = z^{(1)}$ dans le graphe associé à Q reliant $y^{(2)}$ et $z^{(1)}$. Donc, $(y^{(1)}, y^{(2)}), (u_0, u_1), \ldots, (u_{n-1}, u_n), (u_n, z^{(2)})$ définit un chemin dans le graphe associé à T allant de y à z.
Comme μ est invariante par Q, $\sum_{y_1,y_2\in\mathcal{X}} \nu(y_1, y_2) = \sum_{y_1,y_2\in\mathcal{X}} \mu(y_1)Q(y_1, y_2) = \sum_{y_2\in\mathcal{X}} \mu(y_2) = 1$. Donc ν est une probabilité. Enfin, pour tout $y = (y_1, y_2) \in \mathcal{X}^2$,

$$\nu T(y) = \sum_{x_1,x_2\in\mathcal{X}} \nu(x_1, x_2)T((x_1, x_2); (y_1, y_2)) = \sum_{x_1\in\mathcal{X}} \mu(x_1)Q(x_1, y_1)Q(y_1, y_2) = \mu(y_1)Q(y_1, y_2) = \nu(y).$$

Donc ν est bien invariante par T. □

[8] Cet estimateur a d'autres propriétés intéressantes, c'est l'estimateur que l'on obtient par la méthode dite du maximum de vraisemblance.

► *Exemple 62.* Un brin de l'ADN du virus HIV1 compte 9718 nucléotides. Le tableau ci-dessous donne le nombre d'occurrences des 16 dinucléotides sur ce brin :

aa	ac	ag	at	ca	cc	cg	ct	ga	gc	gg	gt	ta	tc	tg	tt
1112	561	1024	713	795	413	95	470	820	457	661	432	684	342	590	548

Proposer une modélisation d'une telle séquence d'ADN par une chaîne de Markov homogène stationnaire qui rende compte du comptage des dinucléotides.

Solution
On modélise la suite de nucléotides comme les 9718 premiers états d'une chaîne de Markov homogène $(X_n)_n$ d'espace d'états $\mathcal{X} = \{e_1 = a, e_2 = c, e_3 = g, e_4 = t\}$. Notons $V(x, y)$ le nombre de fois où le nucléotide x est suivi par le nucléotide y et $V(x) = \sum_{y \in \{a,c,g,t\}} V(x, y)$ le nombre de nucléotides x parmi les 9717 premiers nucléotides. On prend comme matrice de transition, la matrice Q définie par $Q(x, y) = \frac{V(x,y)}{V(x)}$ pour tout $x, y \in \{a, c, g, t\}$, ce qui donne :

$$Q = \begin{pmatrix} 1112/3410 & 561/3410 & 1024/3410 & 713/3410 \\ 795/1773 & 413/1773 & 95/1773 & 470/1773 \\ 820/2370 & 457/2370 & 661/2370 & 432/2370 \\ 684/2164 & 342/2164 & 590/2164 & 548/2164 \end{pmatrix}$$

Pour avoir une chaîne stationnaire on choisit comme loi initiale, la loi invariante par Q qui est décrite par le vecteur $\mu = (\frac{2492462740900}{7100315238309}, \frac{431851952133}{2366771746103}, \frac{577267826325}{2366771746103}, \frac{1580493162035}{7100315238309}) \simeq (0.3510, \ 0.1825, \ 0.2439, \ 0.2226)$.
On peut remarquer que les coordonnées du vecteur μ ne diffèrent de celles du vecteur $(\frac{V(a)}{9717}, \frac{V(c)}{9717}, \frac{V(g)}{9717}, \frac{V(t)}{9717})$ que de 10^{-4}.

12 Le processus de Galton-Watson

Le processus de branchement discret ou processus de Galton-Watson est une chaîne de Markov homogène sur $\mathbb{N}$ qui décrit l'évolution de la taille des générations successives d'une population où chaque individu donne naissance à un nombre aléatoire d'individus. Dans ce modèle, on suppose que les variables aléatoires décrivant le nombre de descendants[9] de chaque individu sont indépendantes et de même loi.

12.1 Définition et exemples

Définition. Soit μ une loi de probabilité sur $\mathbb{N}$ et $\{X_{n,i}, i \in \mathbb{N}^*, \ n \in \mathbb{N}\}$ une famille de variables aléatoires indépendantes de loi μ. Soit U une variable aléatoire sur $\mathbb{N}^*$ indépendante de la famille $\{X_{n,i}, i \in \mathbb{N}, \ n \in \mathbb{N}\}$. La suite de variables aléatoires $(Z_n)_n$ définie par :

$$Z_0 = U \text{ et pour } n \in \mathbb{N}, \quad Z_{n+1} = \begin{cases} \sum_{i=1}^{Z_n} X_{n,i} & \text{si } Z_n \geq 1, \\ 0 & \text{si } Z_n = 0 \end{cases} \tag{10}$$

est appelée un *processus de Galton-Watson de loi de reproduction* μ.

► *Exemple 63.* *Division cellulaire* : on considère une population de cellules toutes identiques (génération 0). Si on ne tient pas compte de l'apparition possible de mutations, chaque cellule de cette population donne naissance à deux cellules identiques avec probabilité p et meurt sans se diviser avec probabilité $1 - p$ ($p \in [0, 1]$). Les cellules issues de la division des cellules de la génération $n - 1$ constituent les cellules de la génération n. La loi de reproduction μ vérifie $\mu(0) = 1 - p$ et $\mu(2) = p$.

► *Exemple 64.* *Survivance des noms de familles :* le processus de branchement a été introduit pour la première fois par Francis Galton en 1873 pour étudier la survivance des noms de famille en Grande-Bretagne. Comme les noms de famille sont transmis par le père, $\mu(k)$ représente dans ce modèle, la probabilité qu'un homme ait k fils.

12.2 Propriétés

Dans la suite, $(Z_n)_n$ désigne un processus de Galton-Watson de loi de reproduction μ.

Proposition 38 *La suite $(Z_n)_n$ est une chaîne de Markov homogène sur $\mathbb{N}$ de matrice de transition Q définie par $Q(i, j) = P(\sum_{k=1}^{i} X_{0,k} = j)$ pour tout $i, j \in \mathbb{N}$.*
- *Si $\mu(1) \neq 1$, l'état 0 est un état absorbant et tous les autres états sont transitoires.*
- *Si $\mu(1) = 1$, tous les états sont absorbants.*

[9] Par descendants d'un individu I, on désignera uniquement les individus engendrés par I.

Preuve. Soit $n \in \mathbb{N}^*$ et $k_0, \dots, k_n \in \mathbb{N}$. D'après la définition de Z_n et en utilisant la convention que $\sum_{i=1}^{0} a_i = 0$:

$$P(Z_0 = k_0, \dots, Z_n = k_n) = P(U = k_0, \sum_{i=1}^{k_0} X_{0,i} = k_1, \dots, \sum_{i=1}^{k_{n-1}} X_{n-1,i} = k_n)$$

Par hypothèse, U, $\{X_{0,i},\ i \in \mathbb{N}\}, \dots, \{X_{n-1,i},\ i \in \mathbb{N}\}$ sont toutes indépendantes. Donc,

$$P(U = k_0, \dots, Z_n = k_n) = P(U = k_0) P(\sum_{i=1}^{k_0} X_{0,i} = k_1) \dots P(\sum_{i=1}^{k_{n-1}} X_{n-1,i} = k_n)$$

Enfin, comme les variables aléatoires $X_{n,i}$ sont toutes de même loi μ,

$$P(U = k_0, \dots, Z_n = k_n) = P(U = k_0) Q(k_0, k_1) \dots Q(k_{n-1}, k_n)$$

en posant $Q(k, \ell) = P(\sum_{i=1}^{k} X_{0,i} = \ell)$ pour tout $k, \ell \in \mathbb{N}$. Donc,

$$Q(0, j) = \mathbb{1}_{j=0} \text{ et } Q(i, j) = P(X_{0,1} + \ldots + X_{0,i} = j) \text{ si } i \in \mathbb{N}^*.$$

Nature des états : $\mu(1) = 1$ signifie que chaque individu donne naissance à un unique individu, la taille de la population reste fixe : tous les états sont absorbants.
On suppose maintenant que $\mu(1) < 1$.

- Si $\mu(0) > 0$, alors pour tout $i \in \mathbb{N}^*$, $Q(i, 0) = \mu(0)^i > 0$ (c'est la probabilité que les i individus de la population n'aient pas de descendants). Comme 0 est un état absorbant, $P(\text{pour tout } n \in \mathbb{N}^*,\ Z_n \neq i | Z_0 = i) \geq Q(i, 0) > 0$. Donc, pour tout $i \in \mathbb{N}^*$, i est transitoire.
- Supposons que $\mu(0) = 0$. Cela signifie que tous les individus donnent naissance à au moins un individu avec probabilité 1. Comme $\mu(1) < 1$, chaque individu a une probabilité strictement positive d'avoir au moins 2 descendants. Donc, partant d'une population avec $i > 0$ individus, il y a une probabilité strictement positive que les générations suivantes aient toujours au moins $i + 1$ individus : précisément, il existe $l \geq 2$ tel que $\mu(l) > 0$. Donc, $Q(i, il) > 0$ et comme $Q(i, j) = 0$ si $j < i$,

  $$P(Z_n \neq i, \text{ pour tout } n > 0 | Z_0 = i) \geq P(X_{0,1} = l, \dots, X_{0,i} = l) = \mu(l)^i > 0.$$

  Cela montre que i est transitoire.

□

Proposition 39 (Comportement asymptotique) *Notons Ext l'événement "la population s'éteint au bout d'un nombre fini de générations" :*

$$Ext = \{\text{Il existe } n \in \mathbb{N},\ Z_n = 0\} = \{T < +\infty\} \text{ où } T = \inf(n > 0,\ Z_n = 0) \quad .$$

Posons $q = P(Ext)$ la probabilité que la population s'éteigne.

(i) La suite $(P(Z_n = 0))_n$ est croissante et converge vers q.

(ii) Si $\mu(1) \neq 1$ alors avec probabilité 1, la population s'éteint ou sa taille tend vers $+\infty$ et donc

$$q + P(\{Z_k \underset{k \to +\infty}{\to} +\infty\}) = 1.$$

Preuve.

(i) Par construction $Z_n = 0$ implique que $Z_{n+1} = 0$. Donc $(P(Z_n = 0))_n$ est une suite croissante.
D'autre part, $q = P(T < +\infty) = \lim_{n \to +\infty} P(T \leq n)$. Or, pour tout $n \in \mathbb{N}^*$, $\{T \leq n\} = \{Z_n = 0\}$. Donc, $(P(Z_n = 0))_n$ converge vers q.

(ii) Soit $N \in \mathbb{N}^*$. Comme tous les états $1, \dots, N - 1$ sont transitoires,

$$P(Z_n \in]0, N[\text{ pour une infinité de } n) \leq \sum_{k=1}^{N-1} P(Z_n \text{ passe une infinité de fois en } k) = 0.$$

Donc, l'événement "il n'y a pas extinction de la population et il existe un entier $N > 0$ tel que la taille d'une infinité de générations soit inférieure à N" est de probabilité nulle :

$$P(\text{Il existe } N \in \mathbb{N}^*,\ Z_n \in]0, N[\text{ pour une infinité de } n) \leq \sum_{N=1}^{+\infty} P(Z_n \in]0, N[\text{ pour une infinité de } n) = 0.$$

L'événement contraire est "au bout d'un nombre fini de générations, la population s'éteint ou pour tout entier N, on peut trouver une génération à partir de laquelle la taille de la population dépasse toujours N". Il arrive avec probabilité 1. Donc, $P(\text{Ext} \cup \{Z_n \underset{n \to +\infty}{\to} +\infty\}) = 1$.
Comme les deux événements "extinction de la population" et "explosion de la taille de la population" sont incompatibles, on peut écrire le résultat différemment

$$q + P(Z_n \underset{n \to +\infty}{\to} +\infty) = 1.$$

□

12.3 Evolution en moyenne de la taille de la population

On suppose dans ce paragraphe que la loi de reproduction μ a une espérance finie notée m ($m = \sum_{k=0}^{+\infty} k\mu(k)$) et que la taille initiale de la population Z_0 est d'espérance finie.

Proposition 40 *Pour tout $n \in \mathbb{N}^*$, la taille moyenne d'une population à la n-ième génération est*

$$E(Z_n) = m^n E(Z_0) \qquad .$$

Preuve. Soit $k \in \mathbb{N}$. Pour calculer $E(Z_n)$, on va d'abord calculer l'espérance conditionnelle $E(Z_n|Z_{n-1} = k)$ pour tout $k \in \mathbb{N}$.
S'il n'y a plus d'individu dans la population à la $(n-1)$-ième génération alors c'est encore le cas à la n-ième génération. Donc, $E(Z_n|Z_{n-1} = 0) = 0$.
Si $k > 0$, $E(Z_n|Z_{n-1} = k) = E(\sum_{l=1}^{k} X_{n-1,l}|Z_{n-1} = k)$. D'après la relation (10), Z_{n-1} dépend uniquement des variables aléatoires $X_{k,i}$ pour $k < n-1$ et $i \in \mathbb{N}^*$. Donc, les variables aléatoires $X_{n-1,l}$ sont indépendantes de Z_{n-1}. D'autre part, elles sont de loi μ. Donc, $E(Z_n|Z_{n-1} = k) = mk$. Calculons $E(Z_n)$:

$$E(Z_n) = \sum_{k=0}^{+\infty} E(Z_n|Z_{n-1} = k)P(Z_{n-1} = k) = \sum_{k=1}^{+\infty} mkP(Z_{n-1} = k) = mE(Z_{n-1}).$$

Par conséquent, $E(Z_n) = m^n E(Z_0)$ pour tout $n \in \mathbb{N}$. □

Définition. Le processus de Galton-Watson $(Z_n)_n$ est dit *sous-critique* si $m < 1$, *critique* si $m = 1$ et *sur-critique* si $m > 1$.

N.B. Si $m < 1$ et si $E(Z_0) < +\infty$, $(E(Z_n))_n$ tend vers 0 et donc la population s'éteint en un nombre fini de générations avec probabilité 1. En effet, pour tout n, $P(Z_n \geq 1) \leq E(Z_n)$, d'après l'inégalité de Markov[10]. Donc, $(P(Z_n = 0))_n$ tend vers 1. Comme la probabilité que la population s'éteigne en un nombre fini de générations est minorée par $P(Z_n = 0)$ pour tout n, elle vaut 1.

12.4 Probabilité d'extinction de la population issue d'un seul ancêtre

On suppose dans ce paragraphe qu'initialement la population est formée d'une seule personne : $\mathbf{Z_0 = 1}$. On cherche à calculer la probabilité que la population s'éteigne : $q = P(\text{Ext})$.

Théorème 41 *Notons f l'application définie sur $[-1, 1]$ par $f(x) = \sum_{k=0}^{+\infty} x^k \mu(k)$.*
- *Pour tout $n \in \mathbb{N}^*$, $P(Z_n = 0) = f(P(Z_{n-1} = 0))$.*
- *La probabilité d'extinction q est la plus petite solution positive de l'équation $x = f(x)$.*

Preuve.
- Notons $\mathcal{S}$ le support de la loi de Z_1 c'est-à-dire l'ensemble des entiers k tels que $P(Z_1 = k) > 0$. Alors,
$$P(Z_n = 0) = \sum_{k \in \mathcal{S}} P(Z_n = 0 \mid Z_1 = k)P(Z_1 = k).$$
Comme $Z_0 = 1$, la loi de Z_1 est μ. Supposons que $Z_1 = k$ avec $k \in \mathbb{N}^*$, c'est-à-dire que la population à la génération 1 est constituée de k individus $I_1, \ldots, I_k$. Notons $V_n^{(i)}$ le nombre d'individus de la génération n qui sont des descendants de I_i. Alors, la loi conditionnelle de Z_n sachant que $Z_1 = k$ est la même que la loi de $\sum_{i=1}^{k} V_n^{(i)}$. D'autre part, les variables aléatoires $V_n^{(1)}, \ldots, V_n^{(k)}$ sont indépendantes et de même loi que Z_{n-1}. Donc,
$$P(Z_n = 0 \mid Z_1 = k) = P(\sum_{i=1}^{k} V_n^{(i)} = 0) = P(\{V_n^{(1)} = 0\} \cap \ldots \cap \{V_n^{(k)} = 0\}) = P(Z_{n-1} = 0)^k$$
Cela montre que $P(Z_n = 0) = f(P(Z_{n-1} = 0))$.
- On fait tendre n vers $+\infty$ dans les deux membres de la relation de récurrence. En utilisant que la suite $(P(Z_n = 0))_n$ tend vers q et le théorème de convergence dominée, on obtient que $\sum_{k=0}^{+\infty} P(Z_{n-1} = 0)^k \mu(k)$ converge vers $\sum_{k=0}^{+\infty} q^k \mu(k)$. Cela montre que $q = f(q)$.
Il reste à montrer que q est la plus petite solution positive de cette équation. Soit e un réel positif vérifiant $e = f(e)$, montrons que pour tout $n \in \mathbb{N}^*$, $P(Z_n = 0) \leq e$.
On a $P(Z_1 = 0) = \mu(0) = f(0) \leq f(e) = e$ car f est croissante sur $[0, 1]$. Supposons que $P(Z_n = 0) \leq e$. Alors, $P(Z_{n+1} = 0) = f(P(Z_n = 0)) \leq f(e) = e$. Donc, $P(Z_n = 0) \leq e$ pour tout $n \in \mathbb{N}^*$.
Comme $(P(Z_n = 0))_n$ tend vers q, on en déduit que $q \leq e$.

[10]Inégalité de Markov : si Y est une variable aléatoire positive ou nulle, pour tout $\epsilon > 0$, $P(Y \geq \epsilon) \leq \frac{E(Y)}{\epsilon}$

□

Définition. La fonction $f : s \mapsto \sum_{k=0}^{+\infty} s^k \mu(k)$ est appelée la *fonction génératrice de la loi de reproduction* μ (si X est une variable aléatoire de loi μ alors $f(s) = E(s^X)$ pour $s \in [-1, 1]$, f est aussi appelée la fonction génératrice de X)

Remarquons que 1 est toujours solution de l'équation $x = f(x)$. Une étude de la fonction f permet d'obtenir le résultat suivant :

Corollaire 42 *Notons m l'espérance de la loi de reproduction ($m \in \mathbb{R}_+ \cup \{+\infty\}$).*
- *Si $\mu(0) = 0$, alors $q = 0$.*
- *Si $\mu(0) > 0$ et $m \leq 1$, alors $q = 1$.*
- *Si $\mu(0) > 0$ et $m > 1$, alors $q \in]0, 1[$ et c'est l'unique solution dans l'intervalle $[0, 1[$ de l'équation $f(s) = s$.*

Preuve.
- Si $\mu(0) = 0$, chaque individu de la population donne naissance à au moins un individu avec probabilité 1. Donc, pour tout $n \in \mathbb{N}^*$, $P(Z_n = 0 | Z_0 = 1) = 0$, cela implique que la population a une probabilité nulle de s'éteindre.
- Si $\mu(0) > 0$ et si $\mu(0) + \mu(1) = 1$, alors $f(s) = \mu(0) + (1 - \mu(0))s$ et l'unique solution de $f(s) = s$ est 1. Remarquons que dans ce cas $m < 1$.
- Supposons maintenant que $\mu(0) > 0$ et que $\mu(0) + \mu(1) < 1$. Posons $h(s) = f(s) - s$. On a $h(0) = \mu(0) \in]0, 1[$ et $h(1) = 0$. On va utiliser les propriétés suivantes de f (communes à toutes les fonctions génératrices de loi de probabilité sur $\mathbb{N}$) :
  - la fonction f est au moins dérivable deux fois sur $] - 1, 1[$: pour tout $s \in] - 1, 1[$, $f'(s) = \sum_{k=1}^{+\infty} k s^{k-1} \mu(k)$, $f''(s) = \sum_{k=2}^{+\infty} k(k-1) s^{k-1} \mu(k)$.
  - f' est donc croissante sur $[0, 1[$
  - $\lim_{s \uparrow 1} f'(s) = m$.

  Pour $s \in [0, 1[$, h est donc deux fois dérivable, $h'(s) = f'(s) - 1$ et $h''(s) = f''(s)$. Comme il existe $k \geq 2$ tel que $\mu(k) > 0$, h'' est strictement positive sur $[0, 1[$ et donc la fonction h' est strictement croissante sur l'intervalle $[0, 1[$. On a $h'(0) = \mu(1) - 1 < 0$ et $h'(s)$ tend vers $m - 1$ lorsque s tend vers 1. Le comportement de h est différent suivant que $m \leq 1$ ou $m > 1$ (faire un tableau de variations pour h).
  - Si $m \leq 1$ alors h' est strictement négative sur $[0, 1[$ et donc h est strictement décroissante sur l'intervalle $]0, 1[$. Par conséquent, la seule solution de $f(s) = s$ dans l'intervalle $[0, 1]$ est 1.
  - Si $m > 1$ alors il existe un unique $a \in]0, 1[$ tel que $h'(a) = 0$. Cela montre que $h(s) = 0$ a une unique solution dans l'intervalle $[0, 1[$.

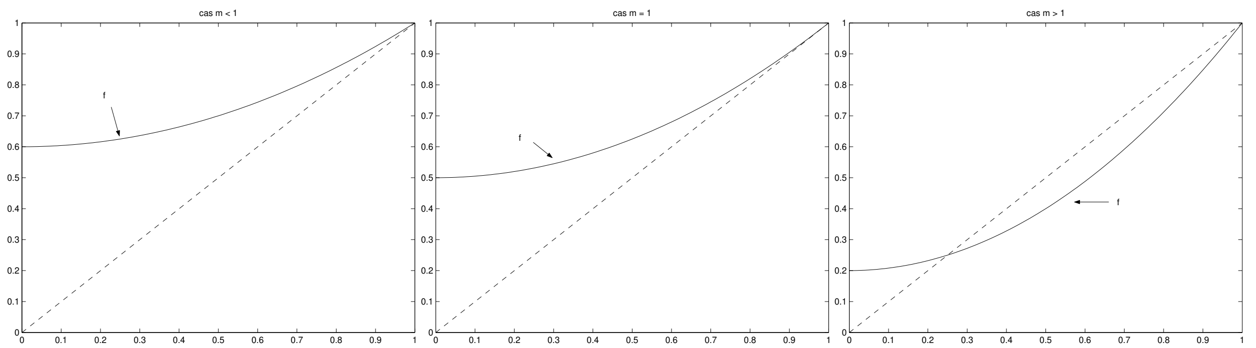


FIG. 3 – Position de la courbe de f par rapport à la diagonale suivant que $m < 1$, $m = 1$ ou $m > 1$ lorsque $\mu(0) > 0$ et $\mu(0) + \mu(1) < 1$.

□

► *Exemple 65. Division cellulaire :* $\mu(0) = 1 - p$ et $\mu(2) = p$. La fonction génératrice de μ est la fonction f définie par $f(s) = 1 - p + ps^2$.
La probabilité d'extinction est égale à $(1 - p)/p$ si $p > 1/2$ et à 1 si $p \leq 1/2$.

► *Exemple 66. Loi de reproduction géométrique :* $\mu(k) = p(1 - p)^k$ pour $k \in \mathbb{N}$. La fonction génératrice de μ est la fonction f définie sur $[-1, 1]$ par

$$f(s) = \sum_{k=0}^{+\infty} s^k \mu(k) = p \sum_{k=0}^{+\infty} (s(1-p))^k = \frac{p}{1 - s(1-p)}.$$

La probabilité d'extinction est égale à $p/(1 - p)$ si $p < 1/2$ et à 1 si $p \geq 1/2$.

12.5 Probabilité d'extinction d'un processus de Galton-Watson issu de k ancêtres

Soit $(Z_n)_n$ un processus de Galton-Watson de loi de reproduction μ issu de $k \geq 2$ ancêtres ($Z_0 = k$). Notons $I_1, \dots, I_k$ les individus de la population initiale. Pour $i \in \{1, \dots, k\}$, posons $Z_0^{(i)} = 1$ et pour $n \geq 1$, posons $Z_n^{(i)}$ le nombre de descendants de la génération n issus de l'individu I_i.
Les suites $(Z_n^{(i)})_{n\geq 1}$ sont des processus de Galton-Watson indépendants issus d'un ancêtre et de loi de reproduction μ et $Z_n = \sum_{i=1}^k Z_n^{(i)}$ pour tout $n \in \mathbb{N}$.
En particulier, le processus $(Z_n)_n$ s'éteint signifie que les populations issues des individus $I_1, \dots, I_k$ s'éteignent toutes. En utilisant l'indépendance des suites $(Z_n^{(i)})_n$, on obtient :

Corollaire 43 *La probabilité que le processus de Galton-Watson $(Z_n)_n$ issu de k ancêtres s'éteigne est q^k où q est la plus petite solution positive de l'équation $f(s) = s$, f étant la fonction génératrice de la loi de reproduction.*

Preuve.

$$\begin{aligned} P(\text{il existe } n \in \mathbb{N}^*,\ Z_n = 0) &= P(\text{pour tout } i \in \{1, \dots, k\},\ \text{ il existe } n_i \in \mathbb{N},\ Z_n^{(i)} = 0) \\ &= \prod_{i=1}^k P(\text{il existe } n \in \mathbb{N}^*,\ Z_n^{(i)} = 0) = q^k \end{aligned}$$

□

12.6 Complément : comportement asymptotique du processus de Galton-Watson dans le cas sur-critique

Soit $(Z_n)_n$ un processus de Galton-Watson, issu d'un ancêtre ($Z_0 = 1$), de loi de reproduction μ. On suppose que la loi de reproduction μ a une espérance finie $m > 1$ (cas sur-critique) et qu'elle admet une variance σ^2 finie. Le théorème suivant décrit le comportement asymptotique de la suite $(W_n)_n$ définie par $W_n = \frac{Z_n}{m^n}$ pour tout $n \in \mathbb{N}$.

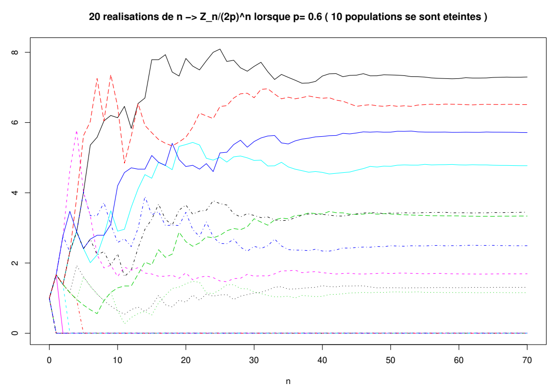
Théorème 44 *La suite $(W_n)_n$ converge avec probabilité 1 vers une variable aléatoire W. De plus,*

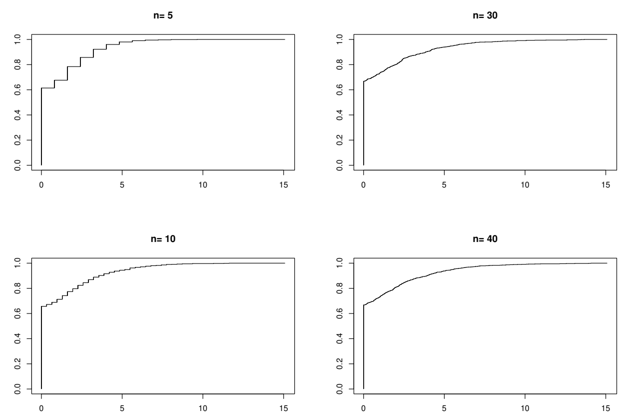
(i) $E((W_n - W)^2) \underset{n\to+\infty}{\longrightarrow} 0$,

(ii) $E(W) = 1$ *et* $\mathrm{Var}(W) = \frac{\sigma^2}{m^2 - m}$,

(iii) $P(W = 0) = q$ *où q est la probabilité d'extinction du processus de Galton-Watson $(Z_n)_n$.*

► *Exemple 67.* Etude par simulations de l'évolution de la taille d'une population de bactéries obtenues par divisions cellulaires à partir d'une seule bactérie (exemple 63). On suppose que chaque bactérie a une probabilité $p = 0.6$ de se diviser. La variable aléatoire Z_n est le nombre de cellules de la génération n. Ici $m = 2p$ et la probabilité d'extinction de la population est $q = 2/3$.





Sur la figure de gauche, on observe bien que chaque réalisation de $n \to \frac{\ln(Z_n)}{(2p)^n}$ semble se stabiliser lorsque n est grand vers une valeur qui dépend de la réalisation.
Les quatre figures de droite montrent la fonction de répartition empirique de 1000 réalisations de $W_n = \frac{Z_n}{(2p)^n}$ pour quatre valeurs de n ; on observe bien que la fonction de répartition empirique se stabilise lorsque n augmente ; la proportion de réalisations nulles de W_{40} est 0.668, la moyenne empirique des 1000 réalisations de W_{40} est de 1.029 et sa variance empirique est de 4.325. La courbe “limite” doit être le graphe de la fonction de répartition de la variable aléatoire W introduite dans le théorème 44 (ici $q = 2/3$, $E(W) = 1$ et $\mathrm{Var}(W) = 4$).

12.7 Exercices

▷ *Exercice 77.* On suppose que la loi de reproduction μ et la loi de Z_0 ont des moments d'ordre 1 et 2 finis. On note m et σ^2 respectivement l'espérance et la variance de la loi de reproduction.

1. En procédant comme pour calculer $E(Z_n)$, montrer que $E(Z_n^2) = m^2 E(Z_{n-1}^2) + \sigma^2 m^{n-1} E(Z_0)$ pour tout $n \in \mathbb{N}^*$.
2. En déduire que la variance de Z_n s'écrit :

$$\mathrm{Var}(Z_n) = \begin{cases} m^2 n \mathrm{Var}(Z_0) + \sigma^2 m^{n-1} \frac{m^n - 1}{m-1} & \text{si } m \neq 1 \\ \mathrm{Var}(Z_0) + n\sigma^2 & \text{si } m = 1 \end{cases}$$

3. Etudier le comportement asymptotique de la suite $(\mathrm{Var}(Z_n))_n$.

▷ *Exercice 78.* Soit $p \in]0,1[$. On rappelle que $\mu = \{\mu(k),\ k \in \mathbb{N}\}$ est la loi géométrique de paramètre p sur $\mathbb{N}$ si pour tout $k \in \mathbb{N}$, $\mu(k) = p(1-p)^k$.
On considère une population d'individus qui a les propriétés suivantes : un individu de cette population évolue vers la maturité avec probabilité $\frac{4}{5}$. Un individu immature ne se reproduit pas. On suppose que le nombre de descendants qu'a un individu mature suit une loi géométrique sur $\mathbb{N}$ de paramètre p et est indépendant du nombre de personnes dans la population et du nombre de descendants qu'ont les autres individus de cette population. On note X le nombre de descendants d'un individu de cette population.

1. Déterminer la loi de X.
2. Déterminer la fonction génératrice G de X.
3. Donner la valeur de l'espérance de X.
4. Au départ la population est constituée d'un unique individu (génération 0). Déterminer en fonction de la valeur de p les probabilités
   – qu'il n'y ait plus d'individus dans la population à la génération 2.
   – que cette population s'éteigne au bout d'un nombre fini de générations.
   – que la taille de la population tende vers $+\infty$ lorsque le nombre de générations augmente.

13 Références bibliographiques

Cours de probabilités

– Bogaert P. *"Probabilités pour scientifiques et ingénieurs. Introduction au calcul des probabilités"*, deBoeck (2005)
– J-M et R Morvan *"Bien débuter en probabilités, Exercices avec rappels de cours de probabilités discrètes"*, Cepadues
– K. P. Baclawski, *"Introduction to Probability with R"*, Chapman & Hall.

Ouvrages généraux sur la modélisation en biologie

– Allman E., Rhodes J. *"Mathematical models in biology"*, Cambridge University Press (2004)
– Coquillard P., Hill D. *"Modélisation et simulation d'écosystèmes"*, Masson (1997)
– Delmas J-F, Jourdain B., *"Modèles aléatoires, Applications aux sciences de l'ingénieur et du vivant"*, Springer (2006)
– Istas J. *"Mathematical modeling for the life sciences"*, Springer (2005)

Sur les chaînes de Markov et leurs applications

– Engel A. *"Processus aléatoires pour les débutants"*, Cassini (2011)
– Istas J. *"Mathematical modeling for the life sciences"*, Springer (2005)
– Tan Wai-Yuan, *"Stochastic models with applications to genetics, cancer, AIDS and other Biomedical systems"*, World Scientific (2002)
– Ycart B. *"Chaînes de Markov"*, `http://www-lmc.imag.fr/lmc-sms/Bernard.Ycart/polys/cha_mark.ps` (fichier postscript)

Sur l'utilisation des chaînes de Markov en bioinformatique

– Deonier R., Tavaré S., Waterman M., *"Computational Genome Analysis"*, Springer (2005)
– Robin S., F. Rodolphe, S. Schbath, *"ADN, mots et modèles"*, Belin (2003)