CHAÎNES DE MARKOV
ANALYSE DE SÉQUENCES BIOLOGIQUES

L3 Génie Biologique et Informatique – Second semestre 2012-2013

MIKAEL FALCONNET
mikael.falconnet@genopole.cnrs.fr

Table des matières

1 Probabilités sur un ensemble fini ou dénombrable

1.1 Univers, évènements et mesures de probabilité

On convient de représenter une *expérience aléatoire* $\mathscr{E}$, c'est à dire une expérience dont l'issue est soumise au hasard, par l'ensemble Ω de tous les résultats possibles à cette expérience. Cet ensemble est appelé *univers*, *espace des possibles* ou encore *espace d'états*. Un résultat possible de l'expérience $\mathscr{E}$ est classiquement noté ω.

Exemple I (Quelques exemples d'univers)**.**

- *L'univers pour le jeu de pile ou face est* $\Omega = \{\mathtt{P}, \mathtt{F}\}$.
- *L'univers pour le jeu de pile ou face répété deux fois est*

$$\Omega = \{(\mathtt{P},\mathtt{P}), (\mathtt{P},\mathtt{F}), (\mathtt{F},\mathtt{P}), (\mathtt{F},\mathtt{F})\}.$$

- *L'univers pour la durée de vie d'une ampoule électrique est* $\Omega = [0, +\infty[$.

Un *évènement aléatoire* A lié à l'expérience $\mathscr{E}$ est un sous-ensemble de Ω dont on peut dire au vu de l'expérience s'il est réalisé ou non.

Exemple II (Quelques exemples d'évènements)**.**

- *« Obtenir pile » lors d'un jeu de pile ou face.*
- *« Obtenir au moins un pile » lors de plusieurs lancers de pièce.*
- *« L'ampoule survit plus de 200 heures » lors de la durée de vie d'une ampoule électrique.*

Les évènements aléatoires étant des ensembles, nous allons rapidement rappeler les opérations élémentaires sur les ensembles afin de décrire la réalisation d'évènements.

Considérons un ensemble Ω, c'est à dire une collection d'objets appelés éléments de Ω. Une partie A de Ω est aussi un ensemble, appelé sous-ensemble de Ω. L'appartenance d'un élément ω au sous-ensemble A est notée $\omega \in A$, et $\omega \notin A$ signifie que le point ω n'appartient pas à A.

Rappelons les opérations élémentaires sur les parties d'un ensemble.

Intersection : l'intersection des ensembles A et B notée $A \cap B$ est l'ensemble des points appartenant à la fois à A et à B.

Réunion : la réunion de deux ensembles A et B notée $A \cup B$ est l'ensemble des points appartenant à au moins l'un des deux ensembles.

Ensemble vide : l'ensemble vide, noté $\emptyset$, est l'ensemble ne contenant aucun élément.

Ensembles disjoints : les ensembles A et B sont dits disjoints si $A \cap B = \emptyset$.

Complémentaire : le complémentaire de l'ensemble $A \subset \Omega$ dans Ω, noté A^c ou $\Omega \setminus A$, est l'ensemble des éléments n'appartenant pas à A. Les ensembles A et A^c sont disjoints.

Les opérations ensemblistes précédemment décrites s'interprètent de la manière suivante avec les évènements. Soient A et B deux évènements.

Non : la réalisation de l'évènement contraire à A est représenté par A^c : le résultat de l'expérience n'appartient pas à A.

Et : l'évènement « A et B sont réalisés » est représenté par $A \cap B$: le résultat de l'expérience se trouve à la fois dans A et dans B.

Ou : l'évènement « A ou B sont réalisés » est représenté par $A \cup B$: le résultat de l'expérience se trouve soit dans A soit dans B soit dans les deux.

Implication : le fait que la réalisation de l'évènement A entraîne la réalisation de B se traduit par $A \subset B$.

Incompatibilité : si $A \cap B = \emptyset$, A et B sont dits incompatibles. Un résultat de l'expérience ne peut être à la fois dans A et dans B.

Nous cherchons à définir, pour un ensemble possible de réalisations de l'expérience A, la vraisemblance accordée a priori à A (avant le résultat de l'expérience). Nous voulons donc associer à chaque évènement un nombre $\mathbb{P}(A)$ compris entre 0 et 1, qui représente la chance que cet évènement soit réalisé à la suite de l'expérience.

Définition 1.1. *Soit $\mathscr{E}$ une expérience aléatoire d'univers Ω. On appelle* mesure de probabilité *sur Ω (ou plus simplement* probabilité*) une application $\mathbb{P}$ qui associe à tout évènement aléatoire A un nombre réel $\mathbb{P}(A)$ telle que*
(i) Pour tout A tel que $\mathbb{P}(A)$ existe, on a $0 \leqslant \mathbb{P}(A) \leqslant 1$.
(ii) $\mathbb{P}(\emptyset) = 0$ et $\mathbb{P}(\Omega) = 1$.
(iii) $A \cap B = \emptyset$ implique que $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$.

Si Ω est fini ou dénombrable, toute probabilité $\mathbb{P}$ sur Ω est parfaitement déterminée par la donnée d'un ensemble de nombres $\{p(\omega) : \omega \in \Omega\}$ vérifiant
- $\forall \omega \in \Omega,\ p(\omega) \geqslant 0$.
- $\sum\limits_{\omega \in \Omega} p(\omega) = 1$.

On a alors $\mathbb{P}(\{\omega\}) = p(\omega)$ et pour tout A, $\mathbb{P}(A) = \sum_{\omega \in A} p(\omega)$.

On note $|E|$ le cardinal d'un ensemble E. Si Ω est fini, la probabilité $\mathbb{P}$ définie sur Ω par

$$\mathbb{P}(\{\omega\}) = \frac{1}{|\Omega|},$$

est appelée *probabilité uniforme* sur Ω. C'est la probabilité qui rend toutes les issues équiprobables. On a alors, pour tout évènement A

$$\mathbb{P}(A) = \frac{|A|}{|\Omega|},$$

formule que l'on paraphrase souvent ainsi

$$\mathbb{P}(A) = \frac{\text{nombre de cas favorables}}{\text{nombre de cas possibles}}.$$

Exemple III (Lancers de deux pièces). *L'univers Ω associé au lancers de deux pièces est défini par*

$$\Omega = \{(\mathrm{P},\mathrm{P}),(\mathrm{P},\mathrm{F}),(\mathrm{F},\mathrm{P}),(\mathrm{F},\mathrm{F})\}.$$

Le cardinal de Ω est donc égal à quatre. Si l'on met la probabilité uniforme sur Ω, alors la probabilité de l'évènement A défini comme « obtenir au moins un pile » est donné par

$$\mathbb{P}(A) = \frac{|A|}{|\Omega|} = \frac{|\{(\mathrm{P},\mathrm{P}),(\mathrm{P},\mathrm{F}),(\mathrm{F},\mathrm{P})\}|}{4} = \frac{3}{4}.$$

1.2 Variables aléatoires, espérance, variance

Une *variable aléatoire* est une fonction dont la valeur dépend de l'issue d'une expérience aléatoire $\mathscr{E}$ d'univers Ω. On dit qu'une variable aléatoire X est *discrète* si elle prend un nombre de valeurs fini ou dénombrables. L'ensemble des issues ω sur lesquelles X prend une valeur fixée x forme l'évènement $\{\omega : X(\omega) = x\}$ que l'on note $[X = x]$. La probabilité de cet évènement est notée $\mathbb{P}(X = x)$.

La fonction $p_X : x \mapsto \mathbb{P}(X = x)$ est appelée la *loi* de la variable aléatoire X. Si $\{x_1, x_2, \ldots\}$ est l'ensemble des valeurs possibles pour X, on a

$$p_X(x_i) \geqslant 0, \quad \sum_i p_X(x_i) = 1.$$

Exemple IV. *Soit S_2 le nombre de piles obtenus lors du lancer de deux pièces. L'ensemble des valeurs possibles pour S_2 est $\{0,1,2\}$. Si l'on munit l'univers Ω associé à cette expérience aléatoire de la probabilité uniforme $\mathbb{P}$, il vient*

$$\mathbb{P}(S_2 = 0) = \mathbb{P}(\{(\mathrm{F},\mathrm{F})\}) = \frac{1}{4},$$

puis

$$\mathbb{P}(S_2 = 1) = \mathbb{P}(\{(\mathrm{P},\mathrm{F}),(\mathrm{F},\mathrm{P})\}) = \frac{1}{2},$$

et enfin

$$\mathbb{P}(S_2 = 2) = \mathbb{P}(\{(\mathrm{P},\mathrm{P})\}) = \frac{1}{4}.$$

On remarque qu'il est possible de réécrire l'évènement A défini par « obtenir au moins un pile » comme l'évènement $[S_2 \geqslant 1]$. On retrouve alors

$$\mathbb{P}(A) = \mathbb{P}(S_2 \geqslant 1) = \mathbb{P}(S_2 = 1) + \mathbb{P}(S_2 = 2) = \frac{3}{4}.$$

Exercice I. *On étudie le lancer de deux dés à six faces équilibrés*

1. *Quel est l'espace d'états Ω associé à cette expérience ? Donner son cardinal.*
2. *On munit Ω de la probabilité uniforme $\mathbb{P}$. On définit la variable aléatoire X comme la somme des résultats de chaque dé. Déterminer la loi de X.*

Lorsqu'elle existe, on appelle *espérance* ou *moyenne* d'une variable aléatoire discrète X la quantité notée $\mathbb{E}(X)$ définie par

$$\mathbb{E}(X) = \sum_i x_i \cdot \mathbb{P}(X = x_i) = \sum_i x_i \cdot p_X(x_i).$$

L'espérance est toujours définie si X prend un nombre fini de valeurs, ou bien si X est à valeurs positives.

Exemple V. *Soit S_2 définie comme ci-dessus. Il vient*

$$\mathbb{E}(S_2) = 0 \cdot \mathbb{P}(S_2 = 0) + 1 \cdot \mathbb{P}(S_2 = 1) + 2 \cdot \mathbb{P}(S_2 = 2) = 1.$$

Lorsqu'elle existe, on appelle *variance* d'une variable aléatoire discrète X la quantité notée $\mathrm{Var}(X)$ définie par

$$\mathrm{Var}(X) = \mathbb{E}([X - \mathbb{E}(X)]^2) = \sum_i [x_i - \mathbb{E}(X)]^2 \cdot \mathbb{P}(X = x_i).$$

Proposition 1.2. *Soit X une variable aléatoire discrète telle que $\mathbb{E}(X)$ et $\mathrm{Var}(X)$ existent. On a alors*

$$\mathrm{Var}(X) = \mathbb{E}(X^2) - [\mathbb{E}(X)]^2.$$

Exercice II. *Démontrer la proposition 1.2.*

Exemple VI. *Soit S_2 définie comme ci-dessus. Il vient*

$$\mathbb{E}(S_2^2) = 0^2 \cdot \mathbb{P}(S_2 = 0) + 1^2 \cdot \mathbb{P}(S_2 = 1) + 2^2 \cdot \mathbb{P}(S_2 = 2) = \frac{3}{2},$$

et

$$\mathrm{Var}(S_2) = \frac{3}{2} - 1^2 = \frac{1}{2}.$$

Exercice III. *On dit qu'une variable aléatoire X suit une loi de Bernoulli de paramètre p si elle prend les valeurs 0 et 1 uniquement, et si*

$$\mathbb{P}(X = 1) = p \quad \textit{et} \quad \mathbb{P}(X = 0) = 1 - p.$$

Une telle variable aléatoire peut être interprétée comme l'obtention d'un « succès » lors d'une expérience aléatoire, le succès étant codé comme 1 et l'échec comme 0.

1. *Lors du lancer d'une pièce de monnaie équilibrée, on considère X la variable aléatoire valant 1 si on obtient pile et 0 si on obtient face. La variable aléatoire X suit-elle une loi de Bernoulli ? Si oui, de quel paramètre ? Si non, pourquoi ?*
2. *Lors du lancer d'un dé non pipé, on considère la variable aléatoire Y valant 1 si l'on obtient un nombre supérieur ou égal à cinq, et 0 sinon. La variable aléatoire X suit-elle une loi de Bernoulli ? Si oui, de quel paramètre ? Si non, pourquoi ?*

3. *Lors d'un questionnaire à choix multiple comportant trois questions avec deux réponses possibles par question (vrai/faux), un étudiant répond au hasard à chaque question. On admet que l'étudiant a autant de chance de se tromper que de répondre juste et que les questions sont indépendantes. On note Z la variable aléatoire comptant le nombre de réponses justes obtenues par l'étudiant. La variable aléatoire Z suit-elle une loi de Bernoulli ? Si oui, de quel paramètre ? Si non, pourquoi ?*
4. *Calculer l'espérance et la variance d'une variable aléatoire suivant une loi de Bernoulli de paramètre p.*

Exercice IV. *Dans beaucoup de modèles probabilistes d'évolution de séquences d'ADN, le nombre de substitutions (changements d'un nucléotide par un autre) que subit un site pendant un temps t est donné par une loi de Poisson de paramètre λt où λ est le taux moyen de substitution par unité de temps.*

On dit que la variable aléatoire N suit une loi de Poisson *de paramètre α si elle est à valeurs entières et si*

$$\mathbb{P}(N = n) = \mathrm{e}^{-\alpha}\frac{\alpha^n}{n!}, \quad \textit{pour tout entier } n.$$

1. *Vérifier que $\sum_n \mathbb{P}(N = n) = 1$. On pourra utiliser le fait que*

$$\mathrm{e}^{\alpha} = 1 + \alpha + \frac{\alpha^2}{2} + \cdots + \frac{\alpha^n}{n!} + \ldots$$

2. *Calculer l'espérance de N.*
3. *Ceci valide-t-il l'appellation « taux moyen de substitution par unité de temps » pour λ dans les modèles d'évolution de séquences d'ADN ?*

1.3 Probabilités conditionnelles, indépendance

L'idée de base du conditionnement est la suivante : une information supplémentaire concernant l'expérience modifie la vraisemblance que l'on accorde à l'évènement étudié.

Par exemple, pour un lancer de deux dés, la probabilité de l'évènement « la somme est supérieure ou égale à 10 » vaut $\frac{1}{6}$ sans information supplémentaire. En revanche si l'on sait que le résultat d'un des dés est 6, elle est égale à $\frac{1}{2}$ tandis qu'elle est égale à 0 si le résultat d'un des dés est 2.

Exercice V. *Comment obtient on ces résultats ?*

Soient $\mathbb{P}$ une mesure de probabilité sur Ω et B un évènement tel que $\mathbb{P}(B) > 0$. La *probabilité conditionnelle* de A sachant B est le réel $\mathbb{P}(A|B)$ défini par

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

Remarque 1.3. *L'application* $A \mapsto \mathbb{P}(A|B)$ *est une mesure de probabilité sur* Ω.

Les évènements A et B sont dits *indépendants* si

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B).$$

On remarque que si $\mathbb{P}(A) > 0$ et $\mathbb{P}(B) > 0$, on a les équivalences

$$A \text{ et } B \text{ indépendants } \Leftrightarrow \mathbb{P}(A|B) = \mathbb{P}(A) \Leftrightarrow \mathbb{P}(B|A) = \mathbb{P}(B).$$

Soient A_1, $A_2, \dots, A_n$ des évènements. Ils sont dits *indépendants* (on précise parfois : dans leur ensemble) si pour tout $k \in \{1, \dots, n\}$ et pour tout ensemble d'entiers distincts $\{i_1, \dots, i_k\} \subset \{1, \dots n\}$, on a

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

Des variables aléatoires $X_1, X_2, \dots, X_n$ sont dites *indépendantes* si les évènements $[X_1 = x_1], [X_2 = x_2], \dots, [X_n = x_n]$ sont indépendants pour tout n-uplet $(x_1, \dots, x_n)$.

Remarque 1.4. *Des variables aléatoires peuvent être indépendantes deux à deux, c'est à dire*

$$\mathbb{P}(X_i = x, X_j = y) = \mathbb{P}(X_i = x)\mathbb{P}(X_j = y), \quad \forall i \neq j, \quad \forall x, y$$

sans être indépendantes dans leur ensemble.

Exercice VI. *L'exercice suivant est un peu artificiel et contre-intuitif, mais il permet de vérifier que vous comprenez les notions suivantes : variables aléatoires, indépendance deux à deux, indépendance dans leur ensemble. Cet exercice illustre en particulier que l'indépendance deux à deux de variables aléatoires n'entraîne pas l'indépendance dans leur ensemble.*

Soient X *et* Y *deux variables aléatoires indépendantes équidistribuées telles que*

$$\mathbb{P}(X = 1) = \frac{1}{2} \quad \textit{et} \quad \mathbb{P}(X = -1) = \frac{1}{2}.$$

On définit une troisième variable aléatoire Z *par* $Z = XY$.

1. *Quelles sont les valeurs possibles de* Z *?*
2. *Donner la loi de* Z.
3. *Montrer que* X *et* Z *sont indépendantes ainsi que* Y *et* Z.
4. *Montrer que* X, Y *et* Z *ne sont pas indépendantes.*

Proposition 1.5. *Soient* X *et* Y *deux variables aléatoires discrètes définies sur le même univers* Ω. *On a alors*

$$\mathbb{E}(aX + bY) = a\mathbb{E}(X) + b\mathbb{E}(Y),$$

pour tous nombres réels a *et* b. *Si de plus* X *et* Y *sont indépendantes*

$$\mathrm{Var}(aX + bY) = a^2\mathrm{Var}(X) + b^2\mathrm{Var}(Y).$$

Exercice VII. *Démontrer la proposition 1.5.*

Remarque 1.6. *Il peut arriver que* $\mathrm{Var}(X+Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$ *bien que* X *et* Y *ne soient pas indépendantes, mais ceci est faux dans le cas général*

Exercice VIII. *Soient* X_1 *et* X_2 *deux variables aléatoires indépendantes suivant une loi de Bernoulli de paramètre* $p \in]0,1[$. *On pose* $S_2 = X_1 + X_2$.

1. *Donner la loi de* S_2.
2. *Calculer l'espérance et la variance de* S_2.
3. *Donner la loi de* S_2 *sachant que* $X_1 = 1$.

Exercice IX. *Au jeu des petits chevaux, il faut obtenir un 6 lors d'un lancer de dé pour pouvoir sortir un cheval de son enclos. On s'interroge sur le nombre moyen de lancers nécessaires afin de sortir un cheval de son enclos. Soit* T *le nombre de lancers nécessaires.*

1. *Quelles sont les valeurs possibles de* T *?*
2. *Écrire l'évènement* $[T = n]$ *avec* $n \geqslant 1$, *à l'aide d'évènements décrits par des variables aléatoires de Bernoulli* $X_1, \dots, X_n$.
3. *En déduire la loi de* T.
4. *Calculer l'espérance de* T. *On admet que*

$$1 + 2x + 3x^2 + \cdots + (n+1)x^n + \cdots = \frac{1}{(1-x)^2}, \quad \textit{pour tout } x \in [0,1[.$$

La loi que l'on vient d'exhiber est appelée loi géométrique, *elle joue un rôle important dans les chaînes de Markov à temps discret.*

Exercice X. *On considère un casino pas tout à fait honnête avec les joueurs qui s'y rendent et qui met à disposition de ses clients deux sortes de dés. Parmi les dés,* 99% *sont équilibrés mais* 1% *d'entre eux sont pipés de telle sorte que l'on obtienne un* 6 *une fois sur deux.*

On choisit au hasard un dé D *sur une table et on note* X *le résultat que l'on obtient en le lançant.*

1. *Donnez* $\mathbb{P}(X = 6 \mid D = d_{\mathrm{equ}})$ *et* $\mathbb{P}(X = 6 \mid D = d_{\mathrm{pip}})$ *?*
2. *Donnez* $\mathbb{P}(X = 6, D = d_{\mathrm{equ}})$ *et* $\mathbb{P}(X = 6, D = d_{\mathrm{pip}})$.
3. *Quelle est la probabilité d'obtenir un six avec le dé ?*
4. *On lance le dé trois fois de suite et on obtient un* 6 *à chaque fois. Naturellement, on se demande si on a choisi un dé pipé. Quelle est la probabilité que ce soit le cas ?*

Exercice XI. *Un étudiant doit répondre à un questionnaire à choix multiples comportant cinq questions. Chaque question comporte quatre réponses possibles dont une et une seule est correcte. On suppose que l'étudiant n'a pas révisé et qu'il répond au hasard à chaque question de manière indépendante. On note* X *le nombre de bonnes réponses obtenus par le candidat.*

Donner la loi de X. *On pourra représenter l'expérience à l'aide d'un arbre.*

2 Premiers pas en modélisation

Quand on lance une pièce de monnaie, on se trouve dans l'incapacité de prédire le résultat que l'on va obtenir. En conséquence, il paraît raisonnable d'utiliser un modèle aléatoire pour décrire ce type d'expérience. Il reste donc à comprendre comment construire le modèle.

Tout d'abord, il faut s'interroger sur les résultats possibles de l'expérience. Dans le cas d'un lancer de pièce, ces valeurs sont `pile` ou `face` [1]. Ensuite, il convient de s'interroger sur la loi de probabilité que l'on va utiliser. Dans le cas d'un lancer de pièce, on aurait envie de dire que les évènements « obtenir pile » ou « obtenir face » sont équiprobables, c'est à dire qu'ils arrivent avec la même probabilité, mais pour l'instant rien ne justifie un tel a priori. Ceci sera donné par la loi des grands nombres que l'on introduira dans cette section.

Dans le cadre des séquences d'ADN, les mécanismes qui gèrent leur évolution au cours du temps ou leur composition spatiale sont loin d'être compris et maîtrisés. En revanche, il paraît raisonnable d'utiliser des modèles aléatoires pour comprendre et explorer des données génomiques. Cependant, il n'est pas aisé de construire des modèles pour ce problème même si le principe de la modélisation est assez simple à comprendre. Il s'agit d'avoir un équilibre entre la pertinence biologique, l'exploitation mathématique et l'efficacité algorithmique. Il ne sert à rien d'avoir un modèle prenant en compte trop d'effets biologiques ou biochimiques en même temps. Un tel modèle serait inexploitable mathématiquement et n'aurait d'utilité pour personne. Au contraire, un modèle trop simple donnerait des résultats mathématiques aussi jolis qu'inutiles puisque les données dont on disposerait ne validerait jamais un tel modèle.

Loi des grands nombres pour un pile ou face. Dans le cadre d'un grand nombre de lancers de la même pièce de monnaie, on s'interroge sur la proportion de piles que l'on devrait obtenir. Intuitivement, on voudrait penser que le nombre de pile et le nombre de faces obtenus ont tendance à s'équilibrer, c'est à dire que la proportion de piles devrait converger vers $1/2$. Mais cette proportion dépend d'une expérience aléatoire, il va donc falloir choisir un modèle pour illustrer notre expérience, puis donner un sens à cette convergence.

On convient que le résultat du lancer d'une pièce de monnaie ne dépend pas de tous les lancers précédents et n'influence pas les lancers suivants. On peut par exemple supposer qu'on change de lanceur à chaque fois et qu'il n'y a pas d'apprentissage possible pour le lanceur. Comme on n'utilise toujours la même pièce, on peut également supposer que la probabilité d'obtenir pile est la même à chaque lancer et vaut $p \in]0,1[$. Avec une pièce parfaitement équilibrée, p serait égal à $1/2$.

1. On exclut le cas où la pièce atterrit sur la tranche même si cela se produit dans "Blitzgiving." How I met your mother. Nov. 22, 2010 (USA).

Pour notre modélisation, on va se donner une suite de variables aléatoires $(X_n)_{n\geqslant 1}$ indépendantes et équidistribuées, c'est à dire ayant toutes la même loi de probabilité donnée par

$$\mathbb{P}(X_i = 1) = p, \quad \mathbb{P}(X_i = 0) = 1 - p,$$

où 1 symbolise `pile` et 0 symbolise `face`.

On note S_n et $\bar{X}_n$ les variables aléatoires définies par

$$S_n = \sum_{i=1}^{n} X_i, \quad \bar{X}_n = \frac{S_n}{n}.$$

Elles représentent respectivement le nombre et la proportion de piles que l'on a obtenus au cours des n premiers lancers. Ce sont bien des variables aléatoires car leur valeur dépend de l'expérience aléatoire à laquelle on s'intéresse.

Proposition 2.1. *La variable aléatoire S_n est à valeurs dans $\{0, 1, \ldots, n\}$. Sa loi est donnée par*

$$\mathbb{P}(S_n = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad \text{pour tout } k \in \{0, 1, \ldots, n\},$$

où $\binom{n}{k}$ donne le nombre de sous-ensembles différents à k éléments que l'on peut former à partir d'un ensemble contenant n éléments.

La loi de S_n est appelée *loi binomiale* de paramètre n et p. Dans le cas particulier où n vaut 1, on retrouve la loi de Bernoulli de paramètre p. L'espérance et la variance de S_n sont donnés par

$$\mathbb{E}(S_n) = np, \quad \mathrm{Var}(S_n) = np(1-p).$$

Démonstration. L'évènement $[S_n = k]$ est égal à l'évènement « obtenir k piles lors de n lancers ». Cet évènement peut s'écrire comme la réunion sur tous les k-uplets $(i_1, \ldots, i_k)$ de $\{1, \ldots, n\}$ des évènements

$$A_{i_1,\ldots,i_k} = \left(\bigcap_{j=1}^{k} [X_{i_j} = 1] \right) \cap \left(\bigcap_{i \notin \{i_1,\ldots,i_k\}} [X_i = 0] \right).$$

Les évènements $A_{i_1,\ldots,i_k}$ sont deux à deux disjoints et chacun d'entre eux a pour probabilité $p^k(1-p)^{n-k}$ grâce à l'indépendance des variables aléatoires $X_1, \ldots, X_n$. Le nombre de k-uplets différents que l'on peut former à partir de $\{1, \ldots, n\}$ est donné par $\binom{n}{k}$ ce qui achève le calcul de $\mathbb{P}(S_n = k)$.

Pour le calcul de l'espérance et de la variance, on utilise la proposition 1.5. □

Puisqu'à présent on connaît la loi de S_n, on peut en déduire la loi exacte de $\bar{X}_n$, mais ce n'est pas ce qui nous intéresse. Par contre, on remarque que l'espérance de $\bar{X}_n$ est donnée par p. En moyenne, on s'attend donc à obtenir une proportion p de piles au cours de l'expérience. En fait, on a un résultat plus fort qui est que lorsque n tend vers l'infini, la proportion de piles X_n tend vers p au sens suivant.

Théorème 2.2. *Quand n devient grand, $\bar{X}_n$ converge en probabilité vers p, c'est à dire que pour tout $\varepsilon > 0$*

$$\mathbb{P}(|\bar{X}_n - p| \geqslant \varepsilon) \to 0, \quad \textit{quand} \quad n \to \infty.$$

Pour démontrer ce théorème, on va utiliser l'inégalité de Bienaymé-Tchebychev.

Proposition 2.3 (Inégalité de Bienaymé-Tchebychev)**.** *Soit Y une variable aléatoire telles que $\mathbb{E}(Y)$ et $\mathrm{Var}(Y)$ existent. On a alors*

$$\mathbb{P}(|Y - \mathbb{E}(Y)| \geqslant \alpha) \leqslant \frac{\mathrm{Var}(Y)}{\alpha^2}, \quad \forall \alpha > 0.$$

Exercice XII. *Soit m un nombre réel fixé et α un nombre réel strictement positif. On veut comparer les fonctions*

$$x \mapsto \alpha^2 \mathbf{1}\{|x - m| \geqslant \alpha\} = \begin{cases} 1 & si\ |x-m| \geqslant \alpha, \\ 0 & si\ |x-m| < \alpha, \end{cases} \quad et \quad x \mapsto (x-m)^2.$$

1. *On considère le cas $m = 0$ et $\alpha = 1$. Représenter sur un même graphique les fonctions*
$$x \mapsto \mathbf{1}\{|x| \geqslant 1\} = \begin{cases} 1 & si\ |x| \geqslant 1, \\ 0 & si\ |x| < 1, \end{cases} \quad et \quad x \mapsto x^2.$$
2. *Représenter sur un même graphique les fonctions*
$$x \mapsto \alpha^2 \mathbf{1}\{|x - m| \geqslant \alpha\} \quad et \quad x \mapsto (x-m)^2.$$
3. *Montrer que pour tout nombre réel x,*
$$\mathbf{1}\{|x - m| \geqslant \alpha\} \leqslant \frac{(x-m)^2}{\alpha^2}.$$

Démonstration de la proposition 2.3. Soit Z la variable aléatoire définie par

$$Z = \mathbf{1}\{|Y - \mathbb{E}(Y)| \geqslant \alpha\},$$

c'est à dire valant 1 si $|Y - \mathbb{E}(Y)| \geqslant \alpha$ et 0 sinon. C'est une variable aléatoire suivant une loi de Bernoulli de paramètre $\mathbb{P}(|Y - \mathbb{E}(Y)| \geqslant \alpha)$. En particulier, on a

$$\mathbb{E}(Z) = \mathbb{P}(|Y - \mathbb{E}(Y)| \geqslant \alpha).$$

Grâce à l'exercice XII, on sait que

$$\alpha^2 Z \leqslant [Y - \mathbb{E}(Y)]^2.$$

En prenant l'espérance de part et d'autre de l'inégalité (qui est conservée), on obtient

$$\alpha^2 \mathbb{P}(|Y - \mathbb{E}(Y)| \geqslant \alpha) = \mathbb{E}(\alpha^2 Z) \leqslant \mathbb{E}([Y - \mathbb{E}(Y)]^2) = \mathrm{Var}(Y),$$

ce qui achève la preuve. □

Le fait que l'inégalité entre deux variables aléatoires $Y_1 \leqslant Y_2$ soit conservée par passage à l'espérance vient du fait que l'espérance d'une variable aléatoire positive est positive. En appliquant ceci à la variable aléatoire $Y_2 - Y_1$, on obtient $\mathbb{E}(Y_2 - Y_1) \geqslant 0$, ce qui entraîne $\mathbb{E}(Y_2) - \mathbb{E}(Y_1) \geqslant 0$.

Démonstration du théorème 2.2. On applique l'inégalité de Bienaymé-Tchebychev à la variable aléatoire $\bar{X}_n$. Comme

$$\mathbb{E}(\bar{X}_n) = p, \quad \mathrm{Var}(\bar{X}_n) = \mathrm{Var}\left(\frac{S_n}{n}\right) = \frac{1}{n^2}\mathrm{Var}(S_n) = \frac{p(1-p)}{n},$$

il vient alors

$$\mathbb{P}(|\bar{X}_n - p| \geqslant \varepsilon) \leqslant \frac{p(1-p)}{n\varepsilon^2} \leqslant \frac{1}{4n\varepsilon^2} \to 0, \quad \text{quand} \quad n \to \infty. \qquad \square$$

Exercice XIII. *Soit n un entier strictement positif. On considère n boîtes numérotées de 1 à n dans lesquelles on va placer aléatoirement n bonbons. Pour tout $i \in \{1,\dots,n\}$, on note X_i la variable aléatoire qui désigne le numéro de la boîte dans laquelle est placée le i-ème bonbon. On suppose que les variables aléatoires $(X_i)_{i=1}^n$ sont indépendantes et équidistribuées suivant la loi uniforme sur $\{1,\dots,n\}$.*

On définit N_n comme le nombre de boîtes vides et on s'intéresse à la convergence en probabilité de la proportion N_n/n de boîtes vides.

1. *Soit $k \in \{1,\dots,n\}$. Écrire en fonction des variables aléatoires $(X_i)_{i=1}^n$ l'évènement A_k défini comme « la boîte k est vide ».*
2. *Donner la probabilité de l'évènement A_k pour $k \in \{1,\dots,n\}$.*
3. *Les évènements $(A_k)_{k=1}^n$ sont-ils indépendants ? Justifier.*
4. *Pour tout $k \in \{1,\dots,n\}$, on définit la variable aléatoire Y_k par $Y_k = \mathbf{1}_{A_k}$. Donner la loi de Y_k.*
5. *Les variables aléatoires $(Y_k)_{k=1}^n$ sont-elles indépendantes ?*
6. *Écrire le nombre N_n de boîtes vides en fonction des variables aléatoires $(Y_k)_{k=1}^n$.*
7. *Calculer l'espérance de N_n.*
8. *On souhaite calculer la variance de N_n et pour cela on doit calculer $\mathbb{E}(N_n^2)$.*
   (a) *Écrire N_2^2 et N_3^2 en fonction de Y_1, Y_2 et Y_3.*
   (b) *Écrire N_n^2 en fonction de $(Y_k)_{k=1}^n$.*
   (c) *Soient k et ℓ deux entiers distincts dans $\{1,\dots,n\}$ Que représente la variable aléatoire $Y_k Y_\ell$? Quelle est la loi de cette variable aléatoire et quelle est son espérance ?*
   (d) *Combien y a-t-il de couples d'entiers (k,ℓ) distincts dans $\{1,\dots,n\}$?*
   (e) *Déduire de ce qui précède $\mathbb{E}(N_n^2)$ puis $\mathrm{Var}(N_n)$.*

9. *On s'intéresse à la limite de* (u_n) *quand* n *tend vers l'infini avec* $u_n = \left(1 - \frac{t}{n}\right)^n$ *avec* t *un nombre réel fixé. On définit* $v_n = \ln\left(1 - \frac{t}{n}\right)$.
   (a) *Donner un développement limité à l'ordre* 1 *de* (v_n).
   (b) *En déduire que* $v_n/n \to -t$ *quand* $n \to +\infty$.
   (c) *Donner la limite de* (u_n).
10. *Montrer que* $\mathbb{E}(N_n/n) \to \frac{1}{\mathrm{e}}$ *quand* $n \to +\infty$.
11. *Montrer que* $\mathrm{Var}(N_n/n) \to 0$ *quand* $n \to +\infty$.
12. *Montrer que si* $\left|\mathbb{E}(N_n/n) - \frac{1}{\mathrm{e}}\right|$ *alors*

$$\mathbb{P}\left(\left|\frac{N_n}{n} - \frac{1}{\mathrm{e}}\right| > \varepsilon\right) \leqslant \mathbb{P}\left(\left|\frac{N_n}{n} - \mathbb{E}\left(\frac{N_n}{n}\right)\right| > \varepsilon/2\right).$$

13. *Déduire de tout ce qui précède que* $\frac{N_n}{n} \to \frac{1}{\mathrm{e}}$ *quand* $n \to +\infty$.

3 Chaînes de Markov à temps discret et espace d'états fini ou dénombrable

Dans la section précédente, nous avons conclu que le modèle indépendant n'était pas satisfaisant pour comprendre la composition spatiale des séquences d'ADN. Nous allons donc introduire un nouveau modèle qui sera dit markovien. Avant de comprendre comment calquer un tel modèle sur notre problématique, il convient de présenter rigoureusement ce type de processus et d'en étudier les propriétés mathématiques.

3.1 Courrier des lectrices

Charlotte la marmotte est une grosse dormeuse, trop peut-être. En effet, son énorme besoin de sommeil la condamne au célibat faute de trouver un compagnon capable de s'adapter à un tel train de vie. Et puis un jour, Charlotte la marmotte rencontre Firmin le lémurien. C'est le coup de foudre. Toutefois, Charlotte la marmotte est légèrement méfiante. Elle a déjà eu des crush pour Hector le castor, Tonio le blaireau, Nicolas le koala, Éric le porc-épic, etc., mais à chaque fois leur rythme de vie n'était pas compatible avec le sien et cela a donné lieu à une rupture qui était malheureusement prévisible.

Cette fois, Charlotte va faire confiance à la science et essayer de prédire

"Combien de temps par jour en moyenne dort Firmin ?"

Comprendre "Est-ce que Charlotte prend rendez-vous chez l'esthéticienne ?".

Nous allons utiliser une chaîne de Markov pour répondre à la question de Charlotte la marmotte et l'aider à prendre une décision pour laquelle *Marmopolitan* n'aurait pas suffi.

3.2 Définition et propriété de Markov

Pourquoi une chaîne de Markov ? Une chaîne de Markov est un processus aléatoire à temps discret dont la principale caractéristique est *l'absence de mémoire* et l'existence de probabilités de transition entre les états de la chaîne. Cela signifie que seul l'état actuel du processus a une influence sur l'état qui va suivre.

Exemple VII (Disponibilité de deux machines, Ruegg (1989))**.** *Une unité de production comprend deux machines automatiques qui fonctionnent indépendamment l'une de l'autre. Chaque machine a la fiabilité p au cours d'une journée, ce qui signifie que sa probabilité de tomber en panne pendant cette période est égale à $1-p$. Dans le cas où elle est tombée en panne, elle sera réparée pendant la nuit* *et se retrouvera et état de marche le lendemain, sauf si les deux machines sont tombées en panne le même jour, auquel cas une seule machine peut être réparée.*

On souhaite comprendre le comportement du processus $(X_n)_{n\geqslant 1}$, où X_n représente le nombre de machines en panne au matin du n-ième jour. Les probabilités de transition sont données par

$$\begin{aligned}
\mathrm{p}(0,0) &= p(2-p),\\
\mathrm{p}(0,1) &= (1-p)^2,\\
\mathrm{p}(1,0) &= p,\\
\mathrm{p}(1,1) &= (1-p),
\end{aligned}$$

que l'on peut résumer à l'aide de la matrice de transition suivante

$$\begin{array}{c c}
 & \begin{array}{cc} \quad 0 & \qquad 1 \end{array} \\
\begin{array}{c} 0 \\ 1 \end{array} & \begin{pmatrix} p(2-p) & (1-p)^2 \\ p & 1-p \end{pmatrix}.
\end{array}$$

Définition 3.1. *Soient S un ensemble fini ou dénombrable et $(X_n)_{n\in\mathbb{N}}$ une suite de variables aléatoires définies sur le même univers Ω muni de la mesure de probabilité $\mathbb{P}$ et à valeurs dans S. On dit que (X_n) est une chaîne de Markov homogène si*

- *i)* (Propriété de Markov) *pour tout entier n et tous $x_0,\ldots,x_{n+1}\in S$*

$$\mathbb{P}(X_{n+1}=x_{n+1}\,|\,X_{0:n}=x_{0:n})=\mathbb{P}(X_{n+1}=x_{n+1}\,|\,X_n=x_n);$$

- *ii)* (Homogénéité) *la probabilité de transition $\mathbb{P}(X_{n+1}=y\,|\,X_n=x)$ ne dépend pas de n, pour tous $(x,y)\in S^2$. On note $\mathrm{p}(x,y)$ cette probabilité de transition.*

La loi de X_0 est appelée loi initiale *de la chaîne de Markov et on appelle* matrice de transition *(de dimensions possiblement infinies) la matrice définie par*

$$P=(\mathrm{p}(x,y))_{x,y\in S}.$$

Remarque 3.2. *La matrice P vérifie*

$$\mathrm{p}(x,y)\geqslant 0 \quad \textit{et} \quad \sum_y \mathrm{p}(x,y)=1.$$

Une telle matrice est dite stochastique *et on remarque que chaque ligne d'une telle matrice est une mesure de probabilité sur S.*

Comportement de Firmin. Charlotte la marmotte observe Firmin le lémurien et constate qu'il a trois activités dans sa vie (auxquelles elle souhaite secrètement en ajouter une quatrième [2]) : dormir, manger et se brosser la fourrure. Dans cet exemple, un état est une activité pour Firmin et on a alors

$$S=\{\text{Dormir, Manger, se Brosser la fourrure}\}=\{\mathrm{D},\mathrm{M},\mathrm{B}\}.$$

2. Le scrabble

Pour satisfaire la propriété de Markov i) de la définition 3.1, nous allons supposer que chaque minute, Firmin le lémurien peut changer d'activité ou la poursuivre, et que cela dépend uniquement de l'activité en cours.

Remarque 3.3. *L'absence de mémoire dans le cas de Firmin est possiblement discutable, mais on peut toujours supposer qu'il est tombé d'une branche en étant petit et que cela lui a laissé quelques séquelles.*

On a vu qu'une chaîne de Markov est formellement définie par une matrice de transition et une loi initiale. Pour étudier le comportement de Firmin dans un cadre Markovien, il convient donc de définir les différents paramètres. Nous reviendrons plus tard sur la possibilité de les estimer.

– Quand Firmin dort, il ne se réveille pas la minute suivante avec probabilité 0.9.
– Si Firmin se réveille, il mange ou bien il se brosse la fourrure avec équiprobabilité.
– Lorsqu'il mange, généralement un fruit, cela dure une minute puis il change d'activité.
– Après le repas, il se brosse la fourrure avec probabilité 0.3 ou bien il fait entame une sieste avec probabilité 0.7.
– Se brosser la fourrure, c'est agréable, mais c'est fatigant. Après une minute, Firmin s'écroule avec probabilité 0.8 ou bien il continue avec probabilité 0.2. Comme il vient juste de se brosser la fourrure, il ne va pas prendre le risque de se tâcher avec de la pulpe, donc il ne se nourrit jamais après le brossage.

La matrice P correspondant à la description précédente est la suivante

$$P = \begin{array}{c} \\ \mathrm{D} \\ \mathrm{M} \\ \mathrm{B} \end{array} \begin{array}{c} \begin{array}{ccc} \mathrm{D} & \mathrm{M} & \mathrm{B} \end{array} \\ \begin{pmatrix} 0.9 & 0.05 & 0.05 \\ 0.7 & 0 & 0.3 \\ 0.8 & 0 & 0.2 \end{pmatrix} \end{array}.$$

Les chaînes de Markov sont toujours mieux décrites à l'aide de diagrammes. Le diagramme correspondant à la matrice stochastique P est représentée sur la figure 1.

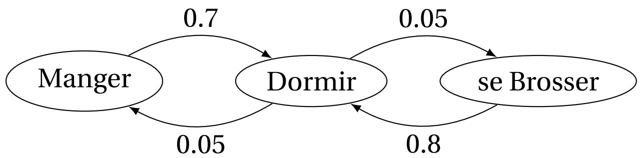


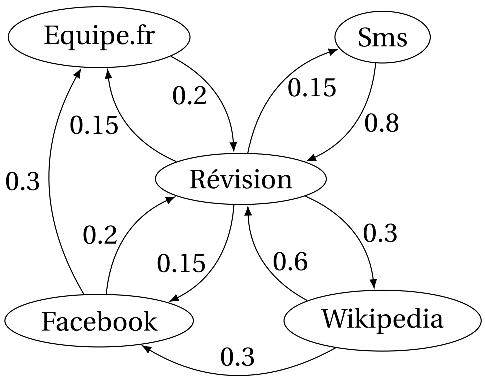
FIGURE 1 – Firmin's French way of life

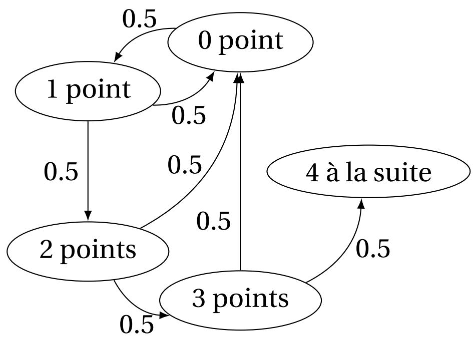
Remarque 3.4. *Nous aurions pu dessiner une boucle de* Dormir *vers lui-même avec pour valeur* 0.9*, mais elle est implicitement présente puisque la somme des* *valeurs des arcs orientés quittant un état est égale à* 1*. cette boucle n'apportant pas d'information supplémentaire, nous choisissons de ne pas la représenter afin de ne pas surcharger le diagramme.*

Exercice XIV. *Dessiner les diagrammes correspondant aux matrices de transition suivantes :*

$$P = \begin{pmatrix} 1-\alpha & \alpha \\ \beta & 1-\beta \end{pmatrix}, \quad P = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 1/2 & 1/2 \\ 1/2 & 0 & 1/2 \end{pmatrix}.$$

Exercice XV. *Donner les matrices stochastiques correspondant aux diagrammes suivants.*





Dans toute la suite, nous abrégeons l'expression « une chaîne de Markov homogène de loi initiale μ et de probabilité de transition P » par « Markov(μ, P) ».

Voici le premier résultat sur la loi d'une trajectoire pour une chaîne de Markov.

Théorème 3.5. *Si un processus discret* (X_n) *est Markov*(μ, P)*, alors pour tout entier n et tout* $(n+1)$*-uplet* $x_{0:n} \in S^{n+1}$

$$\mathbb{P}(X_{0:n} = x_{0:n}) = \mu(x_0)\mathrm{p}(x_0, x_1)\mathrm{p}(x_1, x_2)\ldots\mathrm{p}(x_{n-1}, x_n). \tag{3.1}$$

Démonstration. Nous allons procéder par récurrence. Soit H_n le prédicat de récurrence « pour tout $(n+1)$-uplet $x_{0:n} \in S^{n+1}$, (3.1) est vérifiée ».

Par définition de la loi initiale μ, nous avons $\mathbb{P}(X_0 = x_0) = \mu(x_0)$ pour tout $x_0 \in S$. Ainsi H_0 est vrai.

À présent, fixons $n \geqslant 0$ et supposons que H_n est vrai. Nous montrons que H_{n+1} est vraie. Soit $x_{0:n+1} \in S^{n+2}$. En remarquant que $[X_{0:n+1} = x_{0:n+1}] \subset [X_{0:n} = x_{0:n}]$ et en conditionnant par rapport à la trajectoire de la chaîne jusqu'au temps n, nous obtenons

$$\mathbb{P}(X_{0:n+1} = x_{0:n+1}) = \mathbb{P}(X_{n+1} = x_{n+1} \,|\, X_{0:n} = x_{0:n})\mathbb{P}(X_{0:n} = x_{0:n}).$$

En utilisant la propriété de Markov, il vient

$$\mathbb{P}(X_{0:n+1} = x_{0:n+1}) = \mathbb{P}(X_{n+1} = x_{n+1} \,|\, X_n = x_n)\mathbb{P}(X_{0:n} = x_{0:n}).$$

Comme H_n est vrai, on a

$$\mathbb{P}(X_{0:n} = x_{0:n}) = \mu(x_0)\mathrm{p}(x_0, x_1)p(x_1, x_2)\dots\mathrm{p}(x_{n-1}, x_n),$$

ce qui, combiné à $\mathbb{P}(X_{n+1} = x_{n+1} \,|\, X_n = x_n) = \mathrm{p}(x_n, x_{n+1})$, permet de conclure que H_{n+1} est vraie. □

3.3 Loi de X_n

Le théorème 3.5 nous donne la loi de la trajectoire d'une chaîne de Markov mais ne permet pas de répondre à la question « Quelle est la probabilité qu'après n pas, la chaîne de Markov soit dans un état donné ? ». Dans notre exemple, nous sommes particulièrement intéressé par la probabilité qu'après n minutes Firmin soit en train de dormir. Nous allons voir que ce problème se réduit au calcul de la puissance n-ième de la matrice P.

Proposition 3.6. *Soit (X_n) Markov(μ, P). Alors*

$$\mathbb{P}(X_n = x) = (\mu P^n)(x), \quad \forall x \in S.$$

Démonstration. Une fois encore, nous allons procéder par récurrence. Pour tout entier n, on note $\mu^{(n)}$ la loi de probabilité de X_n, c'est à dire que pour tout x, on a $\mu^{(n)}(x) = \mathbb{P}(X_n = x)$. Soit H_n le prédicat de récurrence « pour tout $x \in S$, $\mu^{(n)}(x) = (\mu P^n)(x)$ ».

Comme P^0 est égale à l'identité, nous avons

$$(\mu P^0)(x) = \mu(x) = \mu^{(0)}(x), \quad \forall x \in S.$$

Ainsi, H_0 est vrai. À présent, fixons $n \geqslant 0$, supposons que H_n est vrai et montrons que H_{n+1} est vrai. Il vient

$$\mu^{(n+1)}(x) = \mathbb{P}(X_{n+1} = x) = \sum_y \mathbb{P}(X_{n+1} = x \,|\, X_n = y)\mathbb{P}(X_n = y) = \sum_y \mathrm{p}(y, x)\mu^{(n)}(y).$$

Comme H_n est vrai, nous avons $\mu^{(n)}(x) = (\mu P^n)(x)$ pour tout $y \in S$ et ainsi

$$\mu^{(n+1)}(x) = \sum_y \mathrm{p}(y, x)(\mu P^n)(x) = \sum_y (\mu P^n)(x)\mathrm{p}(y, x) = (\mu P^{n+1})(x).$$

Ceci permet de conclure que H_{n+1} est vraie. □

Regardons ce que donne la proposition 3.6 appliquée au cas de Firmin. Supposons que Firmin dort la première minute. Alors,

$$\mu^{(0)} = \begin{pmatrix} 1 & 0 & 0 \end{pmatrix}.$$

Après une minute, on a

$$\mu^{(1)} = \mu^{(0)} P = \begin{pmatrix} 0.9 & 0.05 & 0.05 \end{pmatrix},$$

et après deux puis trois minutes

$$\mu^{(2)} = \begin{pmatrix} 0.885 & 0.045 & 0.07 \end{pmatrix} \quad \text{et} \quad \mu^{(3)} = \begin{pmatrix} 0.884 & 0.04425 & 0.07175 \end{pmatrix}.$$

On peut remarquer que $\mu^{(2)}$ et $\mu^{(3)}$ sont proches. On aimerait affirmer que la loi de X_n converge quand n est grand. La suite du cours est consacrée à déterminer la convergence de la loi de X_n et sa loi limite.

Beaucoup de propriétés en temps long des chaînes de Markov sont liées à la notion de mesure de probabilité invariante.

Définition 3.7. *On dit qu'une mesure de probabilité λ est* invariante *(ou* stationnaire*) pour P si*

$$\lambda P = \lambda.$$

Dans notre exemple, la mesure définie par

$$\pi = \begin{pmatrix} \frac{160}{181} & \frac{8}{181} & \frac{13}{181} \end{pmatrix}$$

est invariante pour P. Un calcul approché nous donne

$$\frac{160}{181} \approx 0.884, \quad \frac{8}{181} \approx 0.044, \quad \text{et} \quad \frac{13}{181} \approx 0.072,$$

et l'on peut constater que $\mu^{(3)} = \begin{pmatrix} 0.884 & 0.04425 & 0.07175 \end{pmatrix}$ est proche de π. Effectivement, nous allons montrer que sous certaines hypothèses, la loi de X_n converge vers la mesure invariante de la matrice de transition P.

D'autres calculs nous donne

$$P^2 = \begin{pmatrix} 0.885 & 0.045 & 0.07 \\ 0.87 & 0.035 & 0.095 \\ 0.88 & 0.04 & 0.08 \end{pmatrix},$$

et

$$P^3 = \begin{pmatrix} 0.884 & 0.04425 & 0.07175 \\ 0.8835 & 0.0435 & 0.073 \\ 0.884 & 0.044 & 0.072 \end{pmatrix}.$$

On a donc l'impression que

$$\mathrm{p}^{(n)}(x, y) \to \pi(y) \quad \text{quand } n \to \infty.$$

Cette conjecture est vraie sous certaines hypothèses que nous détaillons dans la suite et qui sont vérifiées par notre exemple.

Exercice XVI. *Paul le poulpe subit un entraînement pour apprendre à choisir un objet A parmi deux objets A et B. Pour cela, on lui fait subir des essais consécutifs dans lesquels il reçoit une récompense si il choisit le bon objet. Paul peut être dans un des trois états suivants :*

(1) *Paul ne sait pas quel objet est récompensé et choisit A ou B avec équiprobabilité.*

(2) *Paul se rappelle que A est récompensé et choisit A, mais peut oublier par la suite.*

(3) *Paul choisit toujours A.*

On suppose qu'après chaque essai il peut changer d'état suivant la matrice de transition

$$P = \begin{pmatrix} 1/2 & 1/2 & 0 \\ 1/2 & 1/12 & 5/12 \\ 0 & 0 & 1 \end{pmatrix}.$$

Soit (X_n) *le processus de Markov associé à l'état de Paul où* X_n *représente l'état de Paul avant le* $(n+1)$*-ième essai. On suppose que Paul est dans l'état 1 avant le premier essai, c'est à dire* $X_0 = 1$.

1. *Dessiner le diagramme correspondant à la matrice de transition* P.
2. *Calculer* $\mathbb{P}(X_1 = 1)$, $\mathbb{P}(X_1 = 2)$ *et* $\mathbb{P}(X_1 = 3)$.
3. *Calculer* $\mathbb{P}(X_2 = 3)$.
4. *Exprimer* $\mathbb{P}(X_{n+1} = 1)$ *en fonction de* $\mathbb{P}(X_n = 1)$ *et* $\mathbb{P}(X_n = 2)$.
5. *Exprimer* $\mathbb{P}(X_{n+1} = 2)$ *en fonction de* $\mathbb{P}(X_n = 1)$ *et* $\mathbb{P}(X_n = 2)$.
6. *Montrer par récurrence que la suite de matrices en ligne* (λ_n) *définie par*

$$\lambda_n = \begin{pmatrix} \mathbb{P}(X_n = 1) & \mathbb{P}(X_n = 2) \end{pmatrix},$$

vérifie

$$\lambda_n = \lambda_0 Q^n, \quad \textit{avec} \quad Q = \begin{pmatrix} 1/2 & 1/2 \\ 1/2 & 1/12 \end{pmatrix}.$$

7. *On admet que les valeurs propres de Q sont* $\frac{5}{6}$ *et* $-\frac{1}{4}$. *Déterminer les constantes a et b telles que*

$$\mathbb{P}(X_n = 1) = a\left(\frac{5}{6}\right)^n + b\left(-\frac{1}{4}\right)^n, \quad \forall n \in \mathbb{N}.$$

8. *Quelle est la probabilité que Paul choisisse l'objet A au n-ième essai ?*

Exercice XVII. *Considérons la matrice de transition suivante :*

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

Vérifier que $\lambda = \begin{pmatrix} 1/3 & 2/3 \end{pmatrix}$ *est une mesure de probabilité invariante pour* P.

Exercice XVIII. *Soit $\alpha \in]0,1]$. Considérons la matrice de transition suivante :*

$$P = \begin{pmatrix} 1-\alpha & \alpha \\ \alpha & 1-\alpha \end{pmatrix}.$$

Déterminer l'unique mesure de probabilité invariante de P.

3.4 Décomposition en classes de communication

Il est parfois possible de « casser » une chaîne de Markov en composantes plus petites, chacune d'entre elle étant relativement facile à comprendre, et qui toutes ensembles permette de comprendre le comportement de la chaîne globale. Ceci est possible grâce aux classes de communication de la chaîne.

Rappelons que $\mathrm{p}^{(n)}(x,y)$ est défini comme $\mathbb{P}(X_n = y \mid X_0 = x)$ que l'on réécrit

$$\mathbb{P}(X_n = y \mid X_0 = x) = \mathbb{P}_x(X_n = y).$$

Définition 3.8. *Soient x et y deux états de S. On dit que x* mène *à y, noté $x \to y$ si*

$$\mathbb{P}_x(X_n = y \text{ pour au moins un } n \geqslant 0) > 0,$$

ce qui est équivalent au fait qu'il existe un entier $n \geqslant 0$ tel que $\mathrm{p}^{(n)}(x,y) > 0$. On dit que x communique *avec y, noté $x \leftrightarrow y$ si x mène à y et y mène à x.*

Théorème 3.9. *Pour des états distincts x et y, les propositions suivantes sont équivalentes.*

(i) x mène à y ;

(ii) il existe un entier n et des états $x_0, x_1, \ldots, x_n$ tels que $x_0 = x$, $x_n = y$ et $\mathrm{p}(x_0,x_1)\mathrm{p}(x_1,x_2)\ldots\mathrm{p}(x_{n-1},x_n) > 0$.

Démonstration. On remarque que

$$\mathrm{p}^{(n)}(x,y) \leqslant \mathbb{P}_x(X_n = y \text{ pour au moins un } n \geqslant 0) \leqslant \sum_{n\geqslant 0} \mathrm{p}^{(n)}(x,y)$$

et que

$$\mathrm{p}^{(n)}(x,y) = \sum_{x_1,\ldots,x_{n-1}} \mathrm{p}(x,x_1)\mathrm{p}(x_1,x_2)\ldots\mathrm{p}(x_{n-1},y). \qquad \square$$

Le théorème fait simplement la correspondance entre l'existence d'au moins un chemin de longueur n débutant en x et finissant en y dans le diagramme de la chaîne de Markov et les probabilités de transition. Ainsi le fait que x mène à y se traduit par le fait qu'il existe un chemin reliant x à y.

La relation $\leftrightarrow$ est une relation d'équivalence sur S. En effet,

1. si x communique avec y et y communique avec z, alors x communique avec z ;

2. l'état x communique avec lui-même ;
3. si x communique avec y, alors y communique avec x.

Ainsi, on peut décomposer S en *classes de communication.*

Définition 3.10. *On dit qu'une classe C est fermée si*

$$x \in C,\ x \to y \quad \textit{implique}\ y \in C.$$

Ainsi, une classe fermée est une classe dont on ne peut pas s'échapper. Un état x est dit absorbant si $\{x\}$ est une classe fermée.

Exercice XIX. *Décomposer en classes de communication les chaînes de Markov des exercices XIV et XV et déterminer dans chaque cas lesquelles sont fermées.*

Définition 3.11. *Une chaîne de Markov est dite irréductible si elle ne possède qu'une seule classe de communication.*

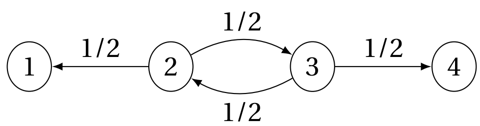
Autrement dit, une chaîne de Markov est irréductible si on ne peut pas casser la chaîne en plusieurs sous-chaînes.

Exercice XX. *Trouver les classes de communication de la matrice de transition*

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

Expliquer pourquoi l'irréductibilité sera l'une des conditions nécessaires pour la convergence vers l'équilibre.

Exercice XXI. *On considère la chaîne de Markov correspondant au diagramme suivant.*



1. *Décomposer la chaîne en classes de communication. Que peut on dire des état* 1 *et* 4 *?*
2. *Écrire la matrice de transition P de la chaîne.*
3. *Vérifier que toute mesure de probabilité λ invariante pour P s'écrit*

$$\lambda = \begin{pmatrix} \alpha & 0 & 0 & 1-\alpha \end{pmatrix},$$

avec $\alpha \in [0,1]$.

3.5 Apériodicité

Nous cherchons le comportement limite des probabilités de transition $\mathrm{p}^{(n)}(x,y)$ quand n devient grand. Regardons un exemple pour lequel un comportement uniforme n'existe pas.

Exemple VIII. *Soit P la matrice de transition donnée par*

$$P = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}.$$

Nous constatons que $P^2 = Id$, ce qui implique que $P^{2n} = Id$ et $P^{2n+1} = P$ pour tout entier naturel n. En particulier, nous avons

$$\mathrm{p}^{2n}(1,1) = 1 \quad \textit{et} \quad \mathrm{p}^{2n}(1,1) = 0,$$

pour tout entier naturel n. Une telle suite ne peut pas converger. C'est le fait que nous sommes face à une chaîne périodique de période 2 *que la convergence n'est pas possible.*

Définition 3.12. *Un état x est dit* apériodique *si* $\mathrm{p}^{(n)}(x,x) > 0$ *pour tout n assez grand.*

Lemme 3.13. *Supposons que P est une matrice de transition irréductible et possède au moins un état apériodique x. Alors, pour tous états y et z, nous avons*

$$\mathrm{p}^{(n)}(y,z) > 0,$$

pour tout n assez grand. En particulier, tous les états sont apériodiques.

Démonstration. Comme la chaîne est irréductible, y mène à x et x mène à z. Il existe donc des entiers r et s tels que

$$\mathrm{p}^{(r)}(y,x) > 0 \quad \text{et} \quad \mathrm{p}^{(s)}(x,z) > 0.$$

En d'autres termes il existe un chemin menant de y à x et un chemin menant de x à z. Comme l'état x est apériodique, quand n est assez grand, on a

$$\mathrm{p}^{(n)}(x,x) > 0.$$

En d'autres termes, il existe une boucle de longueur n partant de x. En mettant bout à bout le chemin menant de y à x, la boucle partant de x et le chemin de x à z, on construit un chemin de longueur $r+n+s$ menant de y à z. C'est à dire

$$\mathrm{p}^{(r+n+s)}(y,z) \geqslant \mathrm{p}^{(r)}(y,x)\mathrm{p}^{(n)}(x,x)\mathrm{p}^{(s)}(x,z) > 0. \qquad \square$$

Exercice XXII. *Soit P une matrice de transition possédant un état x tel que* $\mathrm{p}(x,x) > 0$. *Montrer que l'état x est apériodique.*

3.6 Récurrence et transience

Nous avons vu qu'il était possible de décomposer une chaîne de Markov en différentes classes de communication pour en comprendre le comportement. Nous allons voir que certaines classes sont transitoires tandis que l'on passe infiniment souvent dans d'autres.

Définition 3.14. *Soit (X_n) Markov(μ, P). On dit qu'un état x est* récurrent *si*

$$\mathbb{P}_x(X_n = x, \text{ pour une infinité de } n) = 1.$$

On dit qu'un état x est transient *si*

$$\mathbb{P}_x(X_n = x, \text{ pour une infinité de } n) = 0.$$

Un état récurrent est donc un état par lequel on repasse une infinité de fois tandis qu'un état transient est un état qu'on finira par quitter pour toujours.

Soit T_x le temps de passage en x (ou de retour si l'on part de x) défini par

$$T_x = \inf\{n \geqslant 1 : X_n = x\}.$$

On peut définir par récurrence les temps de passage successifs en x par

$$T_x^{(0)} = 0, \quad T_x^{(1)} = T_x,$$

et pour tout $r \geqslant 1$

$$T_x^{(r+1)} = \inf\{n \geqslant T_x^{(r)} + 1 : X_n = x\}.$$

On peut alors définir la longueur des excursions en dehors de x par

$$S_x^{(r)} = \begin{cases} T_x^{(r)} - T_x^{(r-1)} & \text{si } T_x^{(r-1)} < \infty, \\ 0 & \text{sinon.} \end{cases}$$

Lemme 3.15. *Pour tout $r \geqslant 2$, conditionnellement à $T_x^{(r-1)} < \infty$, $S_x^{(r)}$ est indépendant de $\{X_m : m \leqslant T_x^{(r-1)}\}$, et*

$$\mathbb{P}(S_x^{(r)} = n \,|\, T_x^{(r-1)} < \infty) = \mathbb{P}_x(T_x = n).$$

Démonstration. Admis. □

Soit V_x le nombre de visites en x défini par

$$V_x = \sum_{n=0}^{\infty} \mathbf{1}\{X_n = x\}.$$

On peut remarquer que l'évènement $[X_n = x, \text{ pour une infinité de } n]$ est égal à l'évènement $[V_x = \infty]$. Ainsi un état x est récurrent si et seulement si $\mathbb{P}_x(V_x =$

$\infty) = 1$. Or l'évènement $[V_x = \infty]$ s'écrit comme la limite de la suite d'évènements décroissants $\big([V_x > r]\big)_{r\in\mathbb{N}}$. En effet, si r est un nombre entier, on a alors $[V_x > r+1] \subset [V_x > r]$ et

$$\mathbb{P}_x(V_x = \infty) = \lim_{r\to\infty} \mathbb{P}_x(V_x > r).$$

La suite de cette section consiste donc à étudier la limite ci-dessus, à savoir est elle différente ou égale à 1.

Lemme 3.16. *Pour tout entier r, nous avons*

$$\mathbb{P}_x(V_x > r) = \mathbb{P}_x(T_x < \infty)^r.$$

Démonstration. Remarquons si l'on part de x, le nombre de visites en x est strictement plus grand que r si et seulement si le temps du r-ième passage en x est fini. Ainsi

$$\mathbb{P}_x(V_x > r) = \mathbb{P}_x(T_x^{(r)} < \infty).$$

On a donc en remarquant que $T_x^{(r+1)} = S_x^{(r+1)} + T_x^{(r)}$, si $T_x^{(r)} < \infty$,

$$\mathbb{P}_x(V_x > r+1) = \mathbb{P}_x(T_x^{(r+1)} < \infty) = \mathbb{P}_x(T_x^{(r)} < \infty, S_x^{(r+1)} < \infty),$$

et ainsi en utilisant le lemme 3.15

$$\mathbb{P}_x(V_x > r+1) = \mathbb{P}_x(S_x^{(r+1)} < \infty \mid T_x^{(r)} < \infty)\mathbb{P}_x(T_x^{(r)} < \infty) = \mathbb{P}_x(T_x < \infty)\mathbb{P}_x(T_x^{(r)} < \infty).$$

Une récurrence achève la preuve. □

On utilise dans la suite le lemme suivant

Lemme 3.17. *On a*

$$\mathbb{E}_x(V_x) = \sum_{n=0}^{\infty} \mathrm{p}^{(n)}(x,x) \quad \textit{ou bien} \quad \mathbb{E}_x(V_x) = \sum_{r\in\mathbb{N}} \mathbb{P}_x(V_x > r). \tag{3.2}$$

Démonstration. On a immédiatement

$$\mathbb{E}_x(V_x) = \sum_{n=0}^{\infty} \mathbb{E}_x(\mathbf{1}\{X_n = x\}) = \sum_{n=0}^{\infty} \mathbb{P}_x(X_n = x) = \sum_{n=0}^{\infty} \mathrm{p}^{(n)}(x,x).$$

L'autre formule s'obtient par un jeu de réécriture de la manière suivante

$$\begin{aligned}
\mathbb{E}_x(V_x) =& \mathbb{P}_x(V_x = 1) + 2\mathbb{P}_x(V_x = 2) + 3\mathbb{P}_x(V_x = 3) + 4\mathbb{P}_x(V_x = 4) + \cdots \\
=& \quad \mathbb{P}_x(V_x = 1) \\
& + \mathbb{P}_x(V_x = 2) + \mathbb{P}_x(V_x = 2) \\
& + \mathbb{P}_x(V_x = 3) + \mathbb{P}_x(V_x = 3) + \mathbb{P}_x(V_x = 3) \\
& + \mathbb{P}_x(V_x = 4) + \mathbb{P}_x(V_x = 4) + \mathbb{P}_x(V_x = 4) + \mathbb{P}_x(V_x = 4) \\
& + \cdots
\end{aligned}$$

Si on somme les termes de la première colonne, on obtient

$$\mathbb{P}_x(V_x = 1) + \mathbb{P}_x(V_x = 2) + \mathbb{P}_x(V_x = 3) + \mathbb{P}_x(V_x = 4) + \cdots = \mathbb{P}_x(V_x > 0).$$

En sommant sur la deuxième colonne, on obtient

$$\mathbb{P}_x(V_x = 2) + \mathbb{P}_x(V_x = 3) + \mathbb{P}_x(V_x = 4) + \cdots = \mathbb{P}_x(V_x > 1).$$

On continue ainsi sur chaque colonne et on obtient bien la formule désirée. □

Théorème 3.18. *On a la dichotomie suivante*
(i) Si $\mathbb{P}_x(T_x < \infty) = 1$, alors l'état x est récurrent et $\sum_{n=0}^{\infty} \mathrm{p}^{(n)}(x,x) = \infty$.
(i) Si $\mathbb{P}_x(T_x < \infty) < 1$, alors l'état x est transient et $\sum_{n=0}^{\infty} \mathrm{p}^{(n)}(x,x) < \infty$.
En particulier, chaque état x est soit récurrent, soit transient.

Démonstration. Supposons que $\mathbb{P}_x(T_x < \infty) = 1$. On a alors

$$\mathbb{P}_x(V_x > r) = \mathbb{P}_x(T_x < \infty)^r = 1,$$

pour tout $r \geqslant 0$. Ceci entraîne $\mathbb{P}_x(V_x = \infty) = 1$ et donc que l'état x est récurrent.

Supposons à présent que $\mathbb{P}_x(T_x < \infty) < 1$. Grâce à (3.2) et au lemme 3.16, on obtient

$$\mathbb{E}(V_x) = \sum_{r=0}^{\infty} \mathbb{P}_x(V_x > r) = \sum_{r=0}^{\infty} \mathbb{P}_x(T_x < \infty)^r = \frac{1}{1 - \mathbb{P}_x(T_x < \infty)} < \infty.$$

En particulier, on a $\mathbb{P}_x(V_x = \infty) = 0$. □

Théorème 3.19. *Soit C une classe de communication. Alors soit tous les état de C sont transients, soit ils sont tous récurrents.*

Démonstration. Supposons que C possède un état transient x. Soit y un autre état de C. Montrons que y est aussi transient. Comme x et y communiquent, il existe n et m tels que

$$\mathrm{p}^{(n)}(x,y) > 0 \quad \text{et} \quad \mathrm{p}^{(m)}(y,x) > 0.$$

Comme pour tout $r \geqslant 0$, on a

$$\mathrm{p}^{(n+r+m)}(x,x) \geqslant \mathrm{p}^{(n)}(x,y)\mathrm{p}^{(r)}(y,y)\mathrm{p}^{(m)}(y,x),$$

ce qui se réécrit

$$\mathrm{p}^{(r)}(y,y) \leqslant \frac{\mathrm{p}^{(n+r+m)}(x,x)}{\mathrm{p}^{(n)}(x,y)\mathrm{p}^{(m)}(y,x)},$$

il vient

$$\sum_{r=0}^{\infty} \mathrm{p}^{(r)}(y,y) \leqslant \frac{1}{\mathrm{p}^{(n)}(x,y)\mathrm{p}^{(m)}(y,x)} \sum_{r=0}^{\infty} \mathrm{p}^{(n+r+m)}(x,x) < \infty.$$ □

Théorème 3.20. *Toute classe de communication C finie et fermée est récurrente.*

Démonstration. Supposons que la chaîne de Markov commence en C. Comme la classe C est fermée, la chaîne reste à l'intérieur de C au cours du temps. Comme C est finie, il existe un x tel que

$$\mathbb{P}(X_n = x, \text{ pour une infinité de } n) > 0.$$

En effet, dans le cas contraire, on aurait pour chaque $x \in C$ un entier (aléatoire) n_x tel que $\mathbb{P}(X_n \neq x, \text{ pour tout } n \geqslant n_x) = 1$. Mais on aurait alors, comme C est finie, $\mathbb{P}(X_n \notin C, \text{ pour tout } n \geqslant \max_x n_x) = 1$, ce qui contredit le fait que la classe est fermée.

Comme

$$\mathbb{P}(X_n = x, \text{ pour une infinité de } n) = \mathbb{P}(T_x < \infty) \times \\ \times \mathbb{P}_x(X_n = x, \text{ pour une infinité de } n),$$

on en déduit que

$$\mathbb{P}_x(X_n = x \text{ pour une infinité de } n) > 0,$$

ce qui montre que x n'est pas transient et qu'en conséquence il est récurrent. On termine en utilisant le théorème 3.19. □

Théorème 3.21. *Supposons que P est irréductible et récurrente (automatique si S fini). Alors pour tout $y \in S$,*

$$\mathbb{P}(T_y < \infty) = 1.$$

Démonstration. On écrit tout d'abord

$$\mathbb{P}(T_y < \infty) = \sum_{n \geqslant 1} \mathbb{P}(T_y = n) = \sum_{n \geqslant 1} \sum_{x \in S} \mathbb{P}(T_y = n, X_0 = x).$$

On utilise la propriété de Markov en conditionnant par le premier pas. :

$$\mathbb{P}(T_y = n, X_0 = x) = \mathbb{P}(X_0 = x)\mathbb{P}_x(T_y = n).$$

Ainsi,

$$\mathbb{P}(T_y < \infty) = \sum_{n \geqslant 1} \sum_{x \in S} \mathbb{P}(X_0 = x)\mathbb{P}_x(T_y = n) = \sum_{x \in S} \mathbb{P}(X_0 = x)\mathbb{P}_x(T_y < \infty).$$

Montrons que $\mathbb{P}_x(T_y < \infty) = 1$ pour tout $x \in S$. Comme la chaîne est irréductible, il existe un entier m tel que $\mathrm{p}^{(m)}(x, y) > 0$. Nous avons d'une part

$$\begin{aligned} &\mathbb{P}_y(X_n = y \text{ pour une infinité de } n) \\ =&\mathbb{P}_y(X_n = y \text{ pour une infinité de } n \geqslant m+1) \\ =&\sum_{z \in S} \mathbb{P}_y(X_n = y \text{ pour une infinité de } n \geqslant m+1 \mid X_m = z)\mathbb{P}_y(X_m = z) \\ =&\sum_{z \in S} \mathbb{P}_z(T_y < \infty)\mathrm{p}^{(m)}(y, z). \end{aligned}$$

et d'autre part

$$\mathbb{P}_y(X_n = y \text{ pour une infinité de } n) = 1,$$

car y est un état récurrent. Ainsi

$$\sum_{z\in S} \mathbb{P}_z(T_y < \infty)\mathrm{p}^{(m)}(y,z) = 1.$$

Mais on a $\sum_{z\in S} \mathrm{p}^{(m)}(y,z) = 1$, donc pour tout z tel que $\mathrm{p}^{(m)}(y,z) > 0$, on a $\mathbb{P}_z(T_y < \infty) = 1$ sinon l'égalité n'est pas vérifiée. En particulier pour $z = x$, on obtient $\mathbb{P}_x(T_y < \infty) = 1$. □

Voici enfin un des résultats importants pour les chaînes de Markov. Non seulement il est important, mais en plus sa preuve est très intuitive bien qu'un peu technique à mettre en place.

Théorème 3.22 (Convergence vers l'équilibre)**.** *Soit (X_n) Markov(μ, P) avec P une matrice de transition irréductible et apériodique possédant une mesure invariante π. Supposons de plus que S est fini. Alors*

$$\mathbb{P}(X_n = x) \to \pi(x) \quad \textit{quand } n \to \infty \textit{ pour tout } x,$$

et

$$\mathrm{p}^{(n)}(x,y) \to \pi(y) \quad \textit{quand } n \to \infty \textit{ pour tous } x, y.$$

Démonstration. Nous allons utiliser une technique de couplage pour la démonstration. Soit (Y_n) Markov(π, P) indépendante de (X_n). Fixons un état b dans S et considérons le temps de couplage T défini par

$$T = \inf\{n \geqslant 1 : X_n = Y_n = b\}.$$

Le temps T correspond au premier temps de rencontre en b des deux chaînes. Montrons que $\mathbb{P}(T < \infty) = 1$.

Considérons la nouvelle chaîne de Markov (W_n) à valeurs dans $S \times S$ et définie par $W_n = (X_n, Y_n)$. On peut exprimer la matrice de transition $\tilde{P}$ de (W_n) en fonction de P de la manière suivante

$$\tilde{\mathrm{p}}\big((x,u),(y,v)\big) = \mathrm{p}(x,y)\cdot \mathrm{p}(u,v).$$

La loi initiale μ de (W_n) est quant à elle donnée par

$$\mu(x,u) = \mathbb{P}(W_0 = (x,u)) = \mathbb{P}(X_0 = x, Y_0 = u) = \lambda(x)\cdot\pi(u).$$

Montrons que $\tilde{P}$ est irréductible. Soient (x,u) et (y,v) deux états de $S\times S$. On peut voir que

$$\tilde{\mathrm{p}}^{(n)}\big((x,u),(y,v)\big) = \mathrm{p}^{(n)}(x,y)\cdot \mathrm{p}^{n}(u,v).$$

Or comme P est apériodique et irréductible, on sait d'après le lemme 3.13 qu'il existe un entier n_0 tel que pour tout entier $n \geqslant n_0$

$$\mathrm{p}^{(n)}(x,y) > 0 \quad \text{et} \quad \mathrm{p}^{n}(u,v) > 0.$$

Ainsi, $\tilde{P}$ est également irréductible (et apériodique) sur l'espace d'états $S \times S$ qui est fini. D'après le théorème 3.21, tous les états sont récurrents pour la chaîne (W_n), en particulier l'état (b,b). On a donc $\mathbb{P}(T_{(b,b)} < \infty) = 1 = \mathbb{P}(T < \infty)$.

Définissons le processus (Z_n) par

$$Z_n = \begin{cases} X_n & \text{si } n < T, \\ Y_n & \text{si } n \geqslant T. \end{cases}$$

Nous admettons que le processus (Z_n) est aussi Markov(λ, P). Il vient alors

$$\mathbb{P}(Z_n = x) = \mathbb{P}(X_n = x, n < T) + \mathbb{P}(Y_n = x, n \geqslant T),$$

et donc en remarquand que Z_n a la même loi que X_n et que Y_n est distribuée suivant la mesure invariante π

$$\begin{aligned}
|\mathbb{P}(X_n = x) - \pi(x)| &= |\mathbb{P}(Z_n = x) - \mathbb{P}(Y_n = x)| \\
&= |\mathbb{P}(X_n = x, n < T) + \mathbb{P}(Y_n = x, n \geqslant T) - \mathbb{P}(Y_n = x)| \\
&= |\mathbb{P}(X_n = x, n < T) - \mathbb{P}(Y_n = x, n < T)| \\
&\leqslant \max\{\mathbb{P}(X_n = x, n < T), \mathbb{P}(Y_n = x, n < T)\} \\
&\leqslant \mathbb{P}(n < T).
\end{aligned}$$

Comme $\mathbb{P}(T < \infty) = 1$, il vient $\mathbb{P}(n < T) \to 0$ quand n tend vers l'infini, ce qui achève la preuve. □

Remarque 3.23. *La vitesse de convergence vers la loi stationnaire est de l'ordre de* $|\zeta|^n$ *où* ζ *est la valeur propre de* P *différente de* 1 *et de plus grand module (qui est strictement plus petit que* 1*).*

D'après le théorème 3.22 et la modélisation que nous adoptée pour le comportement de Firmin, nous pouvons déduire à ce stade de réflexion que

"Rapidement (quelques minutes), la probabilité que Firmin soit en train de dormir chaque minute est de l'ordre de 88.4%.*"*

Ceci ne signifie pas que Firmin dorme en moyenne 88.4% du temps. Pour une telle affirmation, il est nécessaire de disposer d'une loi des grands nombres comme dans un modèle de pile ou face.

3.7 Théorèmes ergodiques

Le théorème 2.2 est une application de la loi des grands nombres pour des variables aléatoires indépendantes.

Dans le cadre des chaînes de Markov, il existe aussi une loi des grands nombres sous réserve que la chaîne satisfasse de bonnes hypothèses. Nous omettrons leur démonstration dans ce cours.

Théorème 3.24. *Supposons que S est fini. Soit P une matrice de transition irréductible et apériodique. Alors,*

– *P possède une unique mesure de probabilité invariante λ.*
– *Pour toute fonction bornée $f : S \to \mathbb{R}$, pour tout $\varepsilon > 0$, nous avons*

$$\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n} f(X_i) - \bar{f}\right| \geqslant \varepsilon\right) \to 0 \quad \textit{quand} \quad n \to \infty,$$

avec $\bar{f} = \sum_{x\in S} \pi(x) f(x)$.

Ainsi, en prenant la fonction f définie par

$$f(x) = \mathbf{1}\{x = \mathrm{D}\},$$

et en appliquant le théorème 3.24, on obtient pour tout $\varepsilon > 0$

$$\mathbb{P}(|R_n(\mathrm{D}) - \pi(\mathrm{D})| \geqslant \varepsilon) \to 0 \quad \text{quand} \quad n \to \infty,$$

où

$$R_n(\mathrm{D}) = \frac{1}{n}\sum_{i=1}^{n} \mathbf{1}\{X_i = \mathrm{D}\}.$$

C'est à dire que la proportion de temps que Firmin passe à dormir converge en probabilité vers $\pi(\mathrm{D}) \approx 88.4\%$.

On pourrait opposer à notre raisonnement le fait que le choix de la matrice de transition P en amont rendait une telle conclusion inévitable. Ceci est juste mais si l'on fait uniquement l'hypothèse que le comportement de Firmin est Markovien d'ordre 1, alors il est possible d'estimer les paramètres de la matrice P grâce au théorème suivant.

Théorème 3.25. *Supposons S fini. Soit P une matrice de transition irréductible et apériodique. Supposons que (X_n) est Markov (π, P) avec π l'unique mesure de probabilité invariante de P. Alors,*

$$\frac{1}{n}\sum_{i=1}^{n} f(X_{i-1}, X_i) \xrightarrow[n\to\infty]{} \sum_{x,y\in S} f(x,y)\pi(x)p(x,y) \quad \textit{en probabilité},$$

pour toute fonction bornée $f : S \times S \to \mathbb{R}$.

Ainsi,

$$\frac{1}{n}\sum_{i=1}^{n} \mathbf{1}\{X_{i-1} = x, X_i = y\} \xrightarrow[n\to\infty]{} \pi(x)p(x,y) \quad \text{en probabilité}.$$

4 Utilisations des chaînes de Markov en biologie

4.1 Les modèles de Wright-Fisher

Le modèle de Wright-Fisher est le modèle le plus célèbre (et le plus simple) de l'évolution de la fréquence d'un gène à deux allèles dans une population finie. On suppose que la population est de taille constante, que les générations ne se recouvrent pas et que les unions sont indépendantes du caractère étudié.

Le modèle simple. On néglige dans un premier temps les phénomènes de mutation et sélection. On note $2N$ la taille de la population et $n = 0, 1, 2, \dots$ les générations successives. Sur le locus étudié, on peut trouver deux allèles différents notés A et B. La variable aléatoire X_n compte le nombre d'allèles A à la génération n.

La population à la génération $n+1$ est déduite de celle de la génération n de la manière suivante : chaque individu de la nouvelle génération pioche un allèle au hasard parmi ceux de la génération $2N$, s'attribue son type puis le remet dans le pool d'allèles.

1. Quelle est la loi de X_{n+1} sachant que $X_n = i$? Expliciter les probabilités de transition $\mathrm{p}(i, j) := \mathbb{P}(X_{n+1} = j | X_n = i)$.
2. En déduire les classes de communication de la chaîne.
3. Écrire un script qui prend en entrée une taille de population $2N$, un nombre initial $x \in \{0, \dots, 2N\}$ d'allèles A et un temps t et qui fait évoluer le nombre d'allèles d'allèles A jusqu'à la génération t.
4. Représenter plusieurs trajectoires de cette chaîne de Markov pour $N = 5, 10, 30$.
5. Donner une procédure pour estimer la probabilité que la chaîne soit absorbée en $2N$ en fonction du point de départ x.
6. Une conjecture pour l'expression théorique ?

Prise en compte des mutations Supposons en plus des hypothèses du modèle basique précédent que l'allèle A mute en allèle B avec probabilité $u \in (0, 1)$ et que l'allèle B mute en A avec probabilité $v \in (0, 1)$.

1. Si le nombre d'allèles A à la génération n est i, combien d'allèles de type A sont disponibles après mutations ?
2. Expliciter les nouvelles probabilités de transition de la chaîne.
3. Quel changement s'est il produit par rapport à la chaîne précédente ?
4. À l'aide du logiciel `R`, calculer la mesure de probabilité invariante π de la chaîne pour $N = 10$ et différents paramètres u et v.

5. À partir de quelle valeur de N, le logiciel ne parvient plus à calculer π ?
6. À l'aide du théorème ergodique, proposez une procédure pour estimer π.

4.2 Modélisation de séquences d'ADN

Essayons de comprendre à l'aide d'un exemple pourquoi les modèles markoviens sont importants pour l'analyse de séquences d'ADN.

Exemple : dinucléotides CpG. La notation « CpG » représente ici et dans toute la suite « $5' - CG - 3'$ ». Les cytosines impliquées dans un dinucléotide CpG sont fréquemment méthylées dans le génome des mammifères Bird (1980), et cette méthylation entraîne une hausse des fréquences de mutation de CpG vers TpG, et en conséquence, vers CpA sur le brin complémentaire. Ces dinucléotides intéressent particulièrement les biologistes, car il existe des portions dans le génome, appelés îlots CpG, qui possèdent une concentration en CpG plus élevée que dans le reste de la séquence (voir Bird (1986), Aïssani and Bernardi (1991), Aïssani and Bernardi (1991)). Il s'avère que ces régions sont souvent des zones fonctionnelles de l'ADN Antequera and Bird (1999), et l'existence de telles zones suggère que le phénomène de méthylation est réprimé dans les îlots CpG. La compréhension de ce mécanisme apparaît comme un challenge important en évolution moléculaire.

Maintenant que nous savons que les dinucléotides sont important, nous sommes intéressés par des modèles probabilistes dans lesquels la probabilité d'apparition d'une lettre dépend de la lettre précédente. C'est ainsi que les modèles markoviens trouvent leur place dans l'analyse de séquences biologiques.

Estimation de l'ordre du modèle. La définition 3.1 est la description d'une chaîne de Markov homogène d'ordre 1 au sens où la nature d'un site dépend d'un seul état dans le passé. Cette définition peut être étendue, par exemple si l'on souhaite que la nature d'un site ne dépende non plus uniquement du site qui le précède mais aussi de l'état encore avant. Un tel processus est encore une chaîne de Markov mais d'ordre 2. Dans la suite, on explique un critère pour choisir le « bon » ordre d'un modèle pour la séquence d'ADN que l'on observe.

Définition 4.1. *On dit qu'un processus* (X_n) *est une chaîne de Markov homogène d'ordre m si*

– (Propriété de Markov) *pour tout* $n \geqslant m$ *et* $x_1, \dots, x_{n+1} \in S$

$$\mathbb{P}(X_{n+1} = x_{n+1} \mid X_{1:n} = x_{1:n}) = \mathbb{P}(X_{n+1} = x_{n+1} \mid X_{n-m+1:n} = x_{n-m+1:n});$$

– (Homogénéité) *la probabilité de transition* $\mathbb{P}(X_{n+1} = y \mid X_{n-m+1:n} = x)$ *ne dépend pas de n, pour tous* $(x, y) \in S^m \times S$. *On note* $p(x, y)$ *cette probabilité de transition.*

La loi de $X_{0:m-1}$ *est appelée* loi initiale *de la chaîne de Markov et on appelle* matrice de transition *(de dimensions possiblement infinies) la matrice définie par*

$$P = (p(x, y))_{(x,y)\in S^m \times S}.$$

La matrice P vérifie

$$p(x,y) \geqslant 0 \quad et \quad \sum_y p(x,y) = 1.$$

On remarque que chaque ligne d'une telle matrice est une mesure de probabilité sur S.

Remarque 4.2. *Quand $S = \mathscr{A}$, le nombre de paramètres libres pour une chaîne de Markov $(\mu, P, 1)$ est* 15*,* 3 *pour la mesure invariante et* 4×3 *pour la matrice de transition. Pour une chaîne de Markov (μ, P, m), ce nombre devient* $(4^m - 1) + 4^m \times 3$. *On voit ainsi apparaître le coût de l'extension de la mémoire d'un modèle markovien.*

D'après le théorème 3.5, on sait que la probabilité d'observer une séquence d'ADN $x_{1:N}$ sous le modème $(\mu, P, 1)$ est

$$\mathbb{P}(X_{1:N} = x_{1:N}) = \mu(x_0)\, p(x_0, x_1)\, p(x_1, x_2) \dots p(x_{N-1}, x_N).$$

Une telle formule peut être généralisée pour les modèles d'ordre m.

Théorème 4.3. *Si un processus $(X_n)_n$ est Markov (μ, P, m), alors pour tout $N \geqslant m$ et tout $x_{1:N} \in S^N$*

$$\mathbb{P}(X_{1:N} = x_{1:N}) = \mu(x_{1:m})\, p(x_{1:m}, x_{m+1})\, p(x_{2:m+1}, x_{m+2}) \dots p(x_{N-m+1}, x_N).$$

Exemple IX. *Considérons la séquence de longueur* 8

$$\text{seq} = \texttt{aatatagc}.$$

Si l'on suppose que cette séquence est générée par un modèle Markov$(\mu, P, 1)$, alors on a

$$\mathbb{P}(X_{1:8} = \text{seq}) = \mu(\texttt{a})\, p(\texttt{a},\texttt{a})\, p(\texttt{a},\texttt{t})\, p(\texttt{t},\texttt{a})\, p(\texttt{a},\texttt{t})\, p(\texttt{t},\texttt{a})\, p(\texttt{a},\texttt{g})\, p(\texttt{g},\texttt{c}). \tag{4.1}$$

Si l'on suppose que cette séquence est générée par un modèle Markov$(\mu, P, 2)$, alors on a

$$\mathbb{P}(X_{1:8} = \text{seq}) = \mu(\texttt{aa})\, p(\texttt{aa},\texttt{t})\, p(\texttt{at},\texttt{a})\, p(\texttt{ta},\texttt{t})\, p(\texttt{at},\texttt{a})\, p(\texttt{ta},\texttt{g})\, p(\texttt{ag},\texttt{c}). \tag{4.2}$$

Si l'on suppose que cette séquence est générée par un modèle Markov$(\mu, P, 3)$, alors on a

$$\mathbb{P}(X_{1:8} = \text{seq}) = \mu(\texttt{aat})\, p(\texttt{aat},\texttt{a})\, p(\texttt{ata},\texttt{t})\, p(\texttt{tat},\texttt{a})\, p(\texttt{ata},\texttt{g})\, p(\texttt{tag},\texttt{c}). \tag{4.3}$$

L'équation (4.1) peut être réécrite

$$\mathbb{P}(X_{1:8} = \text{seq}) = \mu(\texttt{a})\, p(\texttt{a},\texttt{a})^1\, p(\texttt{a},\texttt{t})^2\, p(\texttt{a},\texttt{g})^1\, p(\texttt{g},\texttt{c})^1\, p(\texttt{t},\texttt{a})^2.$$

On peut voir que l'exposant correspondant à $p(\texttt{a},\texttt{a})$ est égal à 1 qui est exactement le nombre de fois que le dinucléotides aa apparaît dans la séquence seq. L'exposant correspondant à $p(\texttt{a},\texttt{t})$ est égal à 2, le nombre de fois où le dinucléotide at apparaît dans la séquence seq. en général, pour une chaîne de Markov d'ordre 1,

$$\mathbb{P}(X_{1:N} = s) = \mu(s_1) \prod_{(x,y)\in\mathscr{A}\times\mathscr{A}} p(x,y)^{N_s(xy)},$$

où $N_s(xy)$ est le nombre d'apparitions du dinucléotide xy dans la séquence s.

De même, l'équation (4.2) peut être réécrite

$$\mathbb{P}(X_{1:8} = \text{seq}) = \mu(\texttt{aa})p(\texttt{aa},\texttt{t})^1 p(\texttt{at},\texttt{a})^2 p(\texttt{ta},\texttt{t})^1 p(\texttt{ta},\texttt{g})^1 p(\texttt{ag},\texttt{c})^1.$$

Proposition 4.4. *Si un processus $(X_n)_n$ est Markov(μ, P, m) alors pour tout $N \geqslant m$ et tout $s \in S^N$*

$$\mathbb{P}(X_{1:N} = s) = \mu(s_{1:m}) \prod_{(x,y)\in\mathscr{A}^m\times\mathscr{A}} p(x,y)^{N_s(xy)}, \tag{4.4}$$

où $N_s(xy)$ est le nombre d'apparitions du motif xy dans la séquence s.

Étant donnée une séquence s, on se demande quel model markovien correspond le mieux à s, c'est à dire lequel maximise $\mathbb{P}(X_{1:N} = s)$. Tout d'abord, on peut fixer m et chercher l'ensemble de paramètres $\theta = (\mu, P)$ tel que

$$\ell_m(\theta|s) = \mu(s_{1:m}) \prod_{(x,y)\in\mathscr{A}^m\times\mathscr{A}} p(x,y)^{N_s(xy)},$$

est maximum. La fonction $\theta \mapsto \ell_m(\theta|s)$ est appelée la *vraisemblance*. La théorie du maximum de vraisemblance nous dit que l'ensemble de paramètres $\hat{\theta}_m$ qui maximise ℓ_m étant donnée s est

$$\hat{\mu}(x) = \mathbf{1}\{x = s_{1:m}\}, \quad \forall x \in \mathscr{A}^m \quad \text{et} \quad \hat{p}(x,y) = \frac{N_s(xy)}{N_s(x\bullet)}, \quad \forall (x,y) \in \mathscr{A}^m \times \mathscr{A}.$$

De plus, nous avons

$$\ell_{m+1}(\hat{\theta}_{m+1}|s) \geqslant \ell_{m+1}(\hat{\theta}_m|s),$$

ainsi plus large est l'ensemble des paramètres, mieux le modèle s'accorde à la séquence. Mais le nombre de paramètres croît exponentiellement vite avec m, on veut donc pénaliser ce coût. On définit ainsi un critère BIC (Bayesian information criterion) pour choisir l'ordre du modèle :

$$BIC = -L_m + 3 \times 4^m \log(N),$$

où L_m est la log-vraisemblance du modèle définie par

$$L_m = \log\left(\prod_{(x,y)\in\mathscr{A}^m\times\mathscr{A}} \hat{p}(x,y)^{N_s(xy)}\right) = \sum_{(x,y)\in\mathscr{A}^m\times\mathscr{A}} N_s(xy) \log\left[\frac{N_s(xy)}{N_s(x\bullet)}\right].$$

On choisira ainsi l'ordre m pour lequel le critère BIC est le plus faible.

A Questionnaires à choix multiple

Université d'Évry Val d'Essonne | *Lundi 4 février*
L3 GBI, 2nd semestre 2012-2013 | *Durée : 10 min*

Entraînement aux interrogations de mathématiques

(Documents et calculatrices interdits)

Nom, Prénom :

- Vous recevez 1 point pour chaque réponse correcte et vous perdez 0.5 point pour chaque réponse incorrecte.
- Si vous ne répondez pas à une question, vous ne perdez ni ne gagnez aucun point.
- Vous ne pouvez pas avoir de note négative. Vous aurez donc au moins 0 et au plus 5.

On lance un dé [a] à six faces non pipé et une pièce [b] équilibrée sur laquelle il est inscrit `pair` [c] d'un côté et `impair` [d] de l'autre. L'univers Ω associé à cette expérience aléatoire est donné par

$$\{1,2,3,4,5,6\} \times \{\texttt{pair}, \texttt{impair}\}.$$

On munit Ω de la probabilité uniforme. Soient A l'évènement « le chiffre du dé est pair », B l'évènement « la pièce indique `pair` » et C l'évènement « la parité du dé et l'indication de la pièce correspondent ».
On a alors :

	Vrai	Faux
• $(1, \texttt{pair}) \in A$	☐	☐
• $(2, \texttt{impair}) \in B^c$	☐	☐
• $C = A \cap B$	☐	☐
• $\mathbb{P}(A) = \frac{1}{2}$	☐	☐
• $\mathbb{P}(C) = \frac{1}{2}$	☐	☐

a. *Kość do gry* en polonais.
b. *Moneta* en polonais.
c. Un nombre est dit *pair* si il est divisible par deux.
d. Un nombre est dit *impair* si il n'est pas divisible par deux.

Université d'Évry Val d'Essonne — *Lundi 11 février*
L3 GBI, 2^{nd} semestre 2012-2013 — *Durée : 10 min*

Interrogation de mathématiques n° 1

(Documents et calculatrices interdits)

Nom, Prénom :

– Vous recevez 1 point pour chaque réponse correcte et vous perdez 0.5 point pour chaque réponse incorrecte.
– Si vous ne répondez pas à une question, vous ne perdez ni ne gagnez aucun point.
– Vous ne pouvez pas avoir de note négative. Vous aurez donc au moins 0 et au plus 5.

Soient X et Y deux variables aléatoires indépendantes suivant une loi de Bernoulli de paramètres respectifs p et q, c'est à dire

$$\mathbb{P}(X=1) = 1-\mathbb{P}(X=0) = p, \quad \mathbb{P}(Y=1) = 1-\mathbb{P}(Y=0) = q.$$

On définit par ailleurs Z comme $Z = XY$. On a alors :

	Vrai	Faux
• Z suit une loi de Bernoulli de paramètre pq.	☐	☐
• X^2 suit une loi de Bernoulli de paramètre p^2.	☐	☐
• X et Z sont indépendantes.	☐	☐
• $\mathrm{Var}(X-Y) = p(1-p)+q(1-q)$.	☐	☐
• $1-X$ et Y^2 sont indépendantes.	☐	☐

Université d'Évry Val d'Essonne | *Lundi 18 février*
L3 GBI, 2^{nd} semestre 2012-2013 | *Durée : 10 min*

Interrogation de mathématiques n° 2

(Documents et calculatrices interdits)

Nom, Prénom :

- Vous recevez 1 point pour chaque réponse correcte et vous perdez 0.5 point pour chaque réponse incorrecte.
- Si vous ne répondez pas à une question, vous ne perdez ni ne gagnez aucun point.
- Vous ne pouvez pas avoir de note négative. Vous aurez donc au moins 0 et au plus 5.

On lance un dé à six faces non pipé et on note D la variable aléatoire représentant le score obtenu. Si $D \geqslant 5$, on pioche une boule au hasard dans une urne u_1 contenant une boule noire et trois boules blanches. Si $D < 5$, on pioche une boule au hasard dans une urne différente u_2 contenant une boule blanche et trois boules noires. On note B la variable aléatoire représentant la couleur obtenue que l'on abrègera par `n` pour noire et `b` pour blanche.
On a alors :

	Vrai	Faux
• $\mathbb{P}(B = \mathtt{n} \mid D \geqslant 5) = \dfrac{1}{4}$	☐	☐
• $\mathbb{P}(B = \mathtt{n}, D \leqslant 4) = \dfrac{1}{2}$	☐	☐
• $\mathbb{P}(B = \mathtt{b}) = \dfrac{5}{12}$	☐	☐
• Les variables aléatoires B et D sont indépendantes.	☐	☐
• Si l'on ajoute des boules noires dans l'urne u_1, on peut parvenir à $\mathbb{P}(B = \mathtt{n}) = \dfrac{1}{2}$.	☐	☐

Université d'Évry Val d'Essonne — *Lundi 8 avril*

L3 GBI, 2^nd^ semestre 2012 – 2013 — *Durée : 20 min*

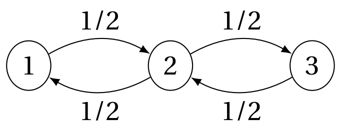
Interrogation de mathématiques n° 3

(Documents et calculatrices interdits)

Nom, Prénom :

– Vous recevez 1 point pour chaque réponse correcte et vous perdez 0.5 point pour chaque réponse incorrecte.
– Si vous ne répondez pas à une question, vous ne perdez ni ne gagnez aucun point.
– Vous ne pouvez pas avoir de note négative. Vous aurez donc au moins 0 et au plus 8.

Soit $(X_n)_{n\geqslant 0}$ Markov(μ, P) où le diagramme correspond à P est le suivant.



On a alors :

	Vrai	Faux
• $\mathrm{p}(1,1) = \frac{1}{2}$	☐	☐
• $\mathbb{P}(X_2 = 3 \mid X_0 = 1) = \frac{1}{4}$	☐	☐
• Si $\mu = \begin{pmatrix}\frac{1}{3} & \frac{1}{3} & \frac{1}{3}\end{pmatrix}$, alors $\mathbb{P}(X_0 = 1, X_1 = 1, X_2 = 2) = \frac{1}{4}$.	☐	☐
• Si $\mu = \begin{pmatrix}1 & 0 & 0\end{pmatrix}$, alors $\mathbb{P}(X_0 = 1, X_1 = 1, X_2 = 2) = \frac{1}{4}$.	☐	☐
• Il existe une loi initiale telle que $\mathbb{P}(X_2 = 3) = 1$.	☐	☐
• Une mesure invariante de la chaîne est $\begin{pmatrix}1/4 & 1/4 & 1/2\end{pmatrix}$	☐	☐
• L'état 1 mène à l'état 3.	☐	☐
• L'état 3 communique avec l'état 3.	☐	☐

Université d'Évry Val d'Essonne | *Lundi 15 avril*
L3 GBI, 2^{nd} semestre 2012 – 2013 | *Durée : 15 min*

Interrogation de mathématiques n° 4

(Documents et calculatrices interdits)

Nom, Prénom :

- Vous recevez 1 point pour chaque réponse correcte et vous perdez 0.5 point pour chaque réponse incorrecte.
- Si vous ne répondez pas à une question, vous ne perdez ni ne gagnez aucun point.
- Vous ne pouvez pas avoir de note négative. Vous aurez donc au moins 0 et au plus 7.

Soit P la matrice de transition suivante

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

On a alors :

	Vrai	Faux
• L'état 3 mène à l'état 1.	☐	☐
• L'état 2 communique avec l'état 3	☐	☐
• L'ensemble $\{1,2\}$ est une classe de communication.	☐	☐
• L'ensemble $\{3\}$ est une classe de communication fermée.	☐	☐
• La loi $\begin{pmatrix}1/4 & 1/2 & 1/4\end{pmatrix}$ est invariante pour P.	☐	☐
• L'état 1 est apériodique.	☐	☐
• $\mathbb{P}_1(\text{toucher } 3) = \dfrac{1}{3}$.	☐	☐

Université d'Évry Val d'Essonne | *Lundi 6 mai*
L3 GBI, 2^nd semestre 2012 – 2013 | *Durée : 10 min*

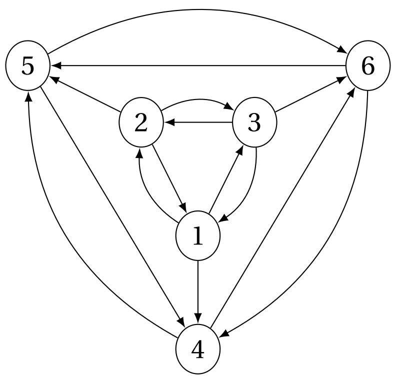
Interrogation de mathématiques n° 5

(Documents et calculatrices interdits)

Nom, Prénom :

– Vous recevez 1 point pour chaque réponse correcte et vous perdez 0.5 point pour chaque réponse incorrecte.
– Si vous ne répondez pas à une question, vous ne perdez ni ne gagnez aucun point.
– Vous ne pouvez pas avoir de note négative. Vous aurez donc au moins 0 et au plus 5.

On considère la chaîne de Markov associée au diagramme suivant. Chaque arc orienté correspond à une probabilité de transition égale à 1/3.



On a alors :

	Vrai	Faux
• L'état 5 est apériodique.	☐	☐
• L'ensemble $\{4,5,6\}$ est une classe de communication fermée.	☐	☐
• L'état 2 est récurrent.	☐	☐
• Partant de 5, on est sûr de toucher l'état 6.	☐	☐
• $\mathbb{P}_3(\text{toucher } 1) = \frac{1}{2}$.	☐	☐

B Contrôles continus

Université d'Évry Val d'Essonne — *Lundi 25 février*
L3 GBI, 2^{nd} semestre 2012-2013 — *Durée : 1h30*

Contrôle continu de MB61 n° 1

(Documents et calculatrices interdits)

– Chaque question est notée sur 1,25 point et vous aurez au moins 0 et au plus 20.
– Une bonne rédaction est un élément d'appréciation important pour les copies.
– Il est permis d'admettre un résultat et de répondre aux questions suivantes en l'utilisant.
– Les deux parties sont indépendantes.

Une entreprise située sur deux sites distincts utilise le transporteur Jason S. afin de convoyer les pièces d'un site à l'autre, une fois par jour.

Partie 1

Dans cette partie, on s'intéresse à la probabilité que Jason S. effectue la livraison en retard.

1. Le trajet habituel de Jason S. comporte trois feux de circulation affichant soit la couleur rouge, soit la couleur verte. On suppose que la couleur des trois feux est aléatoire et on utilise la notation `r` pour rouge et `v` pour verte.
   (a) Donner l'ensemble Ω des configurations de feux possiblement rencontrées par Jason S.. Par exemple, s'il rencontre un feu vert puis deux feux rouges, la configuration est $(\mathtt{v},\mathtt{r},\mathtt{r})$.
   (b) Donner le sous-ensemble de Ω correspondant à l'évènement A défini par « Jason S. rencontre au moins trois feux rouges. ».
   (c) On munit Ω de la probabilité uniforme $\mathbb{P}$. Montrer que $\mathbb{P}(A) = \frac{1}{8}$.
2. Il arrive qu'au début du trajet habituel de Jason S. un camion de livraison bloque le passage. Il emprunte alors un trajet alternatif qui comporte quatre feux de circulation.
   (a) Quel est le cardinal de l'ensemble Ω' des configurations de feux possiblement rencontrées par Jason S. sur le trajet alternatif ?
   (b) On munit Ω' de la mesure de probabilité uniforme $\mathbb{P}$. Montrer que $\mathbb{P}(A) = \frac{5}{16}$.
3. On suppose que la probabilité de présence d'un camion de livraison est de $\frac{1}{5}$. Montrer que la probabilité que Jason S. rencontre au moins deux feux rouges sur son trajet est de $\frac{13}{80}$.

4. Pour effectuer son trajet, Jason S. emprunte sa voiture avec probabilité $\frac{8}{13}$ et son vélo avec probabilité $\frac{5}{13}$. Lorsque Jason S. roule à vélo, il ne tient pas compte des feux rouges rencontrés. En voiture, il sera en retard si il rencontre au moins deux feux rouges. Quelle est la probabilité que Jason S. livre le second site en retard ?

Partie 2

Dans cette partie, on suppose que la probabilité pour que Jason S. livre le second site avec retard est de $\frac{1}{10}$ et que les retards successifs sont indépendants. On note X_n la variable aléatoire égale à 1 si le transporteur est en retard le n-ième jour et 0 sinon.

5. Soit S la variable aléatoire égale au nombre de retards ayant eu lieu pendant les 100 premiers jours.
   (a) Exprimer S en fonction de $(X_n)_{n=1}^{100}$.
   (b) Quelle est la loi de S ?
   (c) Donner l'espérance et la variance de S.
6. Le coût d'une livraison est de 10 euros. En cas de retard du transporteur, le coût de la livraison est divisé par deux. On note C la variable aléatoire égale au coût total des livraisons sur les 100 premiers jours.
   (a) Exprimer C en fonction de S.
   (b) En déduire le prix moyen payé par l'entreprise sur 100 jours.
7. Dorénavant et pour une période indéterminée, l'entreprise décide de changer de société de transport au premier retard de celui-ci. À partir du lendemain du premier retard, elle fera appel à Ryan G. On note T la variable aléatoire représentant le numéro du jour où pour la première fois l'entreprise utilise Ryan G..
   (a) Justifier que l'ensemble des valeurs prises par T est égal à $\mathbb{N} \setminus \{0,1\}$.
   (b) Écrire l' évènement $[T \geqslant 3]$ en fonction de X_1 et l'évènement $[T \geqslant 4]$ en fonction de X_1 et X_2.
   (c) Montrer que $\mathbb{P}(T \geqslant k) = \left(\frac{9}{10}\right)^{k-2}$ pour tout $k \geqslant 2$.
   (d) À partir de quel entier k a-t-on $\mathbb{P}(T \geqslant k) \leqslant \frac{1}{2}$? On donne $\log 2 \approx 0.69$ et $\log(9/10) \approx -0.1$.
   (e) Soit $\ell \geqslant 2$. Montrer que $\mathbb{P}(T \geqslant k+\ell \mid T \geqslant \ell)$ ne dépend pas de k.

Fin

Université d'Évry Val d'Essonne — *Lundi 22 avril*
L3 GBI, 2^{nd} semestre 2012-2013 — *Durée : 1h30*

Contrôle continu de MB61 nº 1

(Documents et calculatrices interdits)

– Une bonne rédaction est un élément d'appréciation important pour les copies.
– Il est permis d'admettre un résultat et de répondre aux questions suivantes en l'utilisant.
– Les trois exercices sont indépendants.
– **Il est recommandé de lire l'intégralité du sujet et de commencer par ce que vous jugez le plus facile**.

Dans tout le sujet, la probabilité d'un évènement A sera notée $\mathbb{P}(A)$. L'espérance et la variance d'une variable aléatoire X seront respectivement notées $\mathbb{E}(X)$ et $\mathrm{Var}(X)$.

Exercice 1

On considère la suite $(X_n)_{n\geqslant 1}$ de variables aléatoires *indépendantes* et telles que pour tout entier $n \geqslant 1$

$$\mathbb{P}(X_n = 1) = \frac{1}{2} \quad \text{et} \quad \mathbb{P}(X_n = -1) = \frac{1}{2}.$$

Pour tout entier $n \geqslant 1$, on définit la variable aléatoire S_n par

$$S_n = \sum_{k=1}^{n} X_k.$$

1. Soit un entier $n \geqslant 1$. Calculer $\mathbb{E}(S_n)$ et $\mathrm{Var}(S_n)$ en fonction de n.
2. Montrer que $\left(\mathrm{Var}\left(\frac{S_n}{n}\right)\right)_{n\geqslant 1}$ converge vers 0 quand n tend vers $+\infty$.
3. En déduire que $\left(\frac{S_n}{n}\right)_{n\geqslant 1}$ converge en probabilité vers 0.

Exercice 2

On dispose d'une urne contenant quatre boules numérotées de 1 à 4. On effectue dans cette urne une succession de tirages d'une boule avec remise et on suppose qu'à chaque tirage, chaque boule a la même probabilité d'être tirée.

Pour tout entier $n \geqslant 1$, on note X_n la variable aléatoire égale au nombre de numéros distincts obtenus après le n-ième tirage. On a donc $X_1 = 1$ et par exemple si les huit premiers tirages sont 2, 2, 3, 1, 3, 2, 4, 3, on a $X_1 = 1$, $X_2 = 1$, $X_3 = 2$, $X_4 = 3$, $X_5 = 3$, $X_6 = 3$, $X_7 = 4$, $X_8 = 4$.

1. Déterminer la loi de X_2. Calculer $\mathbb{E}(X_2)$ et $\mathrm{Var}(X_2)$.
2. Soit M la matrice carré d'ordre 4 définie par

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

   et U_n le vecteur colonne défini pour tout entier $n \geqslant 1$ par

$$U_n = \begin{pmatrix} \mathbb{P}(X_n = 1) \\ \mathbb{P}(X_n = 2) \\ \mathbb{P}(X_n = 3) \\ \mathbb{P}(X_n = 4) \end{pmatrix}.$$

   (a) Établir pour tout entier $n \geqslant 1$ la relation $U_{n+1} = MU_n$.

   (b) Montrer que $U_n = M^{n-1}U_1$ pour tout entier $n \geqslant 1$.
3. Après calculs, on constate que U_n s'écrit pour tout entier $n \geqslant 1$

$$U_n = \left(\frac{1}{4}\right)^n V_1 + 3\left(\frac{1}{2}\right)^n V_2 + 3\left(\frac{3}{4}\right)^n V_3 + V_4,$$

   avec

$$V_1 = \begin{pmatrix} 1 \\ -3 \\ 3 \\ -1 \end{pmatrix}, \quad V_2 = \begin{pmatrix} 0 \\ 1 \\ -2 \\ 1 \end{pmatrix}, \quad V_3 = \begin{pmatrix} 0 \\ 0 \\ 1 \\ -1 \end{pmatrix}, \quad V_4 = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}.$$

   (a) Donner la loi de X_n et calculer $\mathbb{E}(X_n)$.

   (b) Déterminer la limite de $(\mathbb{E}(X_n))$.

Exercice 3

On étudie l'évolution aléatoire d'un étudiant dans un ascenseur (mal conçu) au sein d'un bâtiment décomposé en 7 niveaux de la manière suivante :

- un rez-de-chaussée correspondant au niveau 0 ;
- 4 étages correspondants aux niveaux 1, 2, 3 et 4 ;
- 2 niveaux souterrains correspondants aux niveaux -1 et -2.

L'ascenseur évolue de la manière suivante :

- quand il se trouve à un niveau intermédiaire x avec $x \in \{-1, \ldots, 3\}$, il monte d'un niveau avec probabilité $\frac{1}{2}$ ou descend d'un niveau avec probabilité $\frac{1}{2}$;
- quand il se trouve au niveau 4, il descend au niveau 3 ;
- quand il se trouve au niveau -2, il monte au niveau -1.

Pour tout entier $n \geqslant 1$, on désigne par X_n le niveau aléatoire auquel se trouve l'étudiant après n étapes, et on note X_0 le niveau auquel l'étudiant entre dans l'ascenseur. On admet que $(X_n)_{n \geqslant 0}$ est une chaîne de Markov.

1. Donner la matrice de transition P de la chaîne (X_n) et dessiner son diagramme de transition.
2. La matrice P est elle irréductible ? Justifier.
3. (a) Calculer P^2.
   (b) La chaîne (X_n) est elle apériodique ? Justifier.
4. On considère à présent la chaîne (Y_n) des étapes paires, c'est-à-dire que pour tout entier n, on pose $Y_n = X_{2n}$. La matrice de transition P' des étapes paires est donnée par

$$P' = \begin{array}{c} \\ -2 \\ 0 \\ 2 \\ 4 \end{array} \begin{array}{c} \begin{array}{cccc} -2 & \ \ 0 & \ \ 2 & \ \ 4 \end{array} \\ \begin{pmatrix} 1/2 & 1/2 & 0 & 0 \\ 1/4 & 1/2 & 1/4 & 0 \\ 0 & 1/4 & 1/2 & 1/4 \\ 0 & 0 & 1/2 & 1/2 \end{pmatrix} \end{array}.$$

Pour tout $x \in \{-2, 0, 2, 4\}$, on note $h(x)$ la probabilité que partant du niveau x, on atteigne le niveau 4 avant le niveau -2, c'est à dire

$$h(x) = \mathbb{P}_x\big(\text{la chaîne } (Y_n) \text{ atteint 4 avant } -2\big).$$

(a) Justifier que $h(-2) = 0$ et $h(4) = 1$.

(b) En conditionnant par rapport au premier pas de la chaîne (Y_n), montrer que

$$h(0) = \frac{1}{2}h(-2) + \frac{1}{2}h(2) \quad \text{et} \quad h(2) = \frac{1}{2}h(0) + \frac{1}{2}h(4).$$

(c) Montrer que $h(0) = \frac{1}{3}$.

Fin

Références

Aïssani, B. and G. Bernardi (1991). CpG islands, genes and isochores in the genomes of vertebrates. *Gene 106*, 185–195.

Antequera, F. and A. P. Bird (1999). CpG islands as genomic footprints of promoters that are associated with replication origins. *Current Biology 9*, R661–R667.

Aïssani, B. and G. Bernardi (1991). CpG islands : features and distribution in the genome of vertebrates. *Gene 106*, 173–183.

Bird, A. P. (1980). DNA methylation and the frequency of CpG in animal DNA. *Nucleic Acids Research 8*, 1499–1504.

Bird, A. P. (1986). CpG-rich islands and the function of DNA methylation. *Nature 321*, 209–213.

Ruegg, A. (1989). *Processus stochastiques*, Volume 6 of *Méthodes Mathématiques pour l'Ingénieur [Mathematical Methods for the Engineer]*. Lausanne : Presses Polytechniques Romandes. Avec applications aux phénomènes d'attente et de fiabilité. [With applications to queueing and reliability].