Un cours sur les probabilités

Licence Informatique UHP

Blaise Potard

2005

Table des matières

Chapitre 1

Introduction

1.1 Pipo introductif

Les probabilités en France ont longtemps constituées (avec la logique) le parent pauvre des mathématiques, essentiellement sous l'influence de la «nouvelle école mathématique» de N. Bourbaki, qui disait justement que «l'espace des probabilités n'est pas localement paracompacte»[1] et que du coup, ce n'était pas vraiment la peine de s'y intéresser. C'est l'une des raisons pour lesquelles les probabilités ne sont toujours, assez souvent, pas étudiées en cours de mathématique, mais en cours de physique, de biologie, voire d'informatique ou d'économie. Même si le formalisme de la construction peut, il est vrai, laisser à désirer, cela n'empêche pas les gens qui cherchent à faire des choses utiles de s'en servir ; en réalité, les probabilités sont quelque chose d'extrêmement utile pour toute la recherche appliquée : de la modélisation des processus stochastiques (mécanique des fluides,...) à la mécanique quantique, l'importance pour le physicien n'est pas à démontrer. L'informaticien s'en sert également beaucoup, pour la conception d'heuristiques probabilistes, et dans tous les travaux sur l'intelligence artificielle. En particulier, l'un des modèles les plus utilisés pour cela sont les chaînes (ou modèles) de Markov : par exemple, pour l'OCR (reconnaissance de caractère) ou la reconnaissance vocale. Certains s'en servent également (avec un succès relatif) pour faire de la traduction automatique de texte, etc.

1.2 Plan du cours

Ce cours n'a pour ambition que d'être une introduction à ce domaine (passionnant) que constitue les probabilités. Après avoir vu les notions de probabilité et de variable aléatoire discrète, nous étudierons en détail une heuristique pour résoudre le problème de la coupe minimale et nous présenterons les chaînes de Markov.

1.3 Remerciements

Ce cours s'inspire largement des cours de probabilités de M. Cichon (que vous devez connaître) et de M. Koiran (de l'ÉNS Lyon). Le cours de M. Koiran est de niveau maîtrise et est assez complet (mais il est d'un niveau assez relevé). Une transcription pour les gens interessés est disponible à cette adresse : http ://laure.gonnord.org/site-ens/mim/probas/proba.ps

1.4 Exploitation des masses

Tous les commentaires (questions, remarques, et surtout repérage d'erreurs et imprécisions) sont bien entendu les bienvenus, et peuvent être adressés à l'auteur à cette adresse : potard@loria.fr

[1] ou quelque chose d'approchant...

Chapitre 2

Probabilités sur un espace fini

2.1 Définitions

On s'intéresse à une expérience aléatoire qui conduit à la réalisation d'un seul résultat parmi un nombre fini de résultats possibles ω_1, ..., ω_n. On note $\Omega = \{\omega_1, \ldots, \omega_n\}$, l'ensemble de ces résultats.

Exemple 2.1.1

1. *Jet d'une pièce à pile ou face :* $\Omega = \{P, F\}$.
2. *Jet d'un dé :* $\Omega = \{1, 2, 3, 4, 5, 6\}$.

Si on mesure la fréquence d'apparition du résultat ω_k au cours d'un grand nombre de répétitions de l'expérience i.e. on calcule le rapport $F_k = \frac{N_k}{N}$ du nombre N_k d'expériences dont le résultat est ω_k sur le nombre total d'expériences N, on constate qu'elle fluctue de moins en moins. La limite $p_k \geq 0$ de F_k lorsque $N \to +\infty$ correspond à la notion intuitive de *probabilité*.

On appelle *événement* une partie A de Ω. La fréquence de A, c'est-à-dire la proportion d'expériences dont le résultat est dans A, est égale à $\sum_{\omega_k \in A} F_k$. On est donc amené à associer la probabilité $\sum_{\omega_k \in A} p_k$ à l'événement A. Comme la fréquence de Ω vaut 1, on obtient $\sum_{k=1}^{n} p_k = 1$.

Définition 2.1.2 *Une probabilité $\mathbb{P}$ sur un ensemble fini $\Omega = \{\omega_1, \ldots, \omega_n\}$ est une pondération $p_1, \ldots, p_n$ des éléments de cet ensemble telle que*

$$\forall 1 \leq k \leq n, \; p_k \geq 0 \; \text{et} \; \sum_{k=1}^{n} p_k = 1$$

On attribue à tout événement $A \subseteq \Omega$ le nombre

$$\mathbb{P}(A) = \sum_{\omega_k \in A} p_k$$

qui est appelé probabilité de l'événement *A.*

Exemple 2.1.3 *Jet de deux dés à six faces : $\Omega = \{(i, j) : 1 \leq i, j \leq 6\}$ où i désigne la valeur de la face supérieure du premier dé et j celle du second. Pour des raisons de symétrie (si les dés ne sont pas pipés), on munit Ω de la pondération suivante :*

$$\forall 1 \leq i, j \leq 6, \; p_{(i,j)} = \frac{1}{36}$$

Soit A l'événement : "les valeurs des deux dés sont identiques".

$$A = \{(1, 1), \ldots, (6, 6)\} \; \text{et} \; \mathbb{P}(A) = \sum_{i-1}^{6} p_{(i,i)} = \frac{6}{36} = \frac{1}{6}$$

On note S la somme des deux dés et $\{S = k\}$ l'événement $\{(i,j) : S(i,j) = k\}$. On a $S(i,j) = i + j$. Donc

$$
\begin{array}{rclrcl}
\{S=2\} & = & \{(1,1)\} & \mathbb{P}(S=2) & = & \frac{1}{36} \\
\{S=3\} & = & \{(1,2),\ (2,1)\} & \mathbb{P}(S=3) & = & \frac{1}{18} \\
\{S=4\} & = & \{(1,2),\ (2,2),\ (3,1)\} & \mathbb{P}(S=4) & = & \frac{1}{12} \\
\{S=5\} & = & \{(1,4),\ (2,3),\ (3,2),\ (4,1)\} & \mathbb{P}(S=5) & = & \frac{1}{9} \\
\{S=6\} & = & \{(1,5),\ (2,4),\ (3,3),\ (4,2),\ (5,1)\} & \mathbb{P}(S=6) & = & \frac{5}{36} \\
\{S=7\} & = & \{(1,6),\ (2,5),\ (3,4),\ (4,3),\ (5,2),\ (6,1)\} & \mathbb{P}(S=7) & = & \frac{1}{6} \\
\{S=8\} & = & \{(2,6),\ (3,5),\ (4,4),\ (5,3),\ (6,2)\} & \mathbb{P}(S=8) & = & \frac{5}{36} \\
\{S=9\} & = & \{(3,6),\ (4,5),\ (5,4),\ (6,3)\} & \mathbb{P}(S=9) & = & \frac{1}{9} \\
\{S=10\} & = & \{(4,6),\ (5,5),\ (6,4)\} & \mathbb{P}(S=10) & = & \frac{1}{12} \\
\{S=11\} & = & \{(5,6),\ (6,5)\} & \mathbb{P}(S=11) & = & \frac{1}{18} \\
\{S=12\} & = & \{(6,6)\} & \mathbb{P}(S=12) & = & \frac{1}{36}
\end{array}
$$

2.2 Terminologie concernant les événements

- Si $A, B \subseteq \Omega$, l'événement 'A ou B' (réalisé lorsque A ou B le sont) est noté $A \cup B$.
- Si $A, B \subseteq \Omega$, l'événement 'A et B' (réalisé lorsque A et B le sont) est noté $A \cap B$.
- On appelle événement contraire de A et on note A^c l'événement $\Omega \setminus A$.

2.3 Propriétés élémentaires

1. $\mathbb{P}(\Omega) = 1$.
2. Si $A \cap B = \emptyset$, alors $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B)$.
3. $\mathbb{P}(A^c) = 1 - \mathbb{P}(A)$.
4. $\mathbb{P}(A \cap B) = \mathbb{P}(A) - \mathbb{P}(A \cap B^c)$.

1. est une application direct de la définition. 2. est évident en écrivant $\sum_{\omega_k \in A \cup B} p_k = \sum_{\omega_k \in A} p_k + \sum_{\omega_k \in B} p_k$ (car A et B sont disjoints). 3. $A^c \cap A = \emptyset$ et $A^c \cup A = \Omega$, et on applique 2. 4. $A = (A \cap B) \cup (A \cap B^c)$ et $(A \cap B) \cap (A \cap B^c) = \emptyset$, puis on applique 2.

2.4 Probabilité de l'événement $A \cup B$

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$$

Notons que $A \cup B$ est égal à l'union disjointe $(A \cap B^c) \cup (A \cap B) \cup (A^c \cap B)$, d'où le résultat est immédiat.

2.5 Probabilités uniformes

Dans le cas où les symétries font que tous les résultats possibles $\omega_1, \ldots, \omega_n$ jouent le même rôle, ces résultats doivent avoir la même pondération $\frac{1}{\mathrm{card}(\Omega)}$. On dit alors qu'il sont *équiprobables*. On a alors, pour tout événement $A \subseteq \Omega$,

$$\mathbb{P}(A) = \sum_{\omega_k \in \Omega} \frac{1}{\mathrm{card}(\Omega)} = \frac{\mathrm{card}(A)}{\mathrm{card}(\Omega)}$$

Cette probabilité $\mathbb{P}$ s'appelle *probabilité uniforme* sur Ω.

Exemple 2.5.1 *Dans le cas du jet de deux dés non pipés, $\Omega = \{(i,j) : 1 \leq i, j \leq 6\}$ est muni de la probabilité uniforme.*

Remarque 2.5.2 *Si on s'intéresse à la somme des deux dés, on peut choisir $\Omega = \{2, 3, \ldots, 12\}$, ensemble des valeurs prises par cette somme. Mais faute de propriétés de symétrie, on ne sait pas munir cet espace d'une probabilité naturelle. Cette pondération n'est pas uniforme.*

Dans l'exemple 2.1.3, en travaillant sur l'espace $\{(i,j) : 1 \leq i,j \leq 6\}$ *des couples des valeurs des deux dés muni de la probabilité uniforme, nous avons pu construire la pondération naturelle sur les valeurs de la somme des deux dés.*

Cet exemple permet de bien comprendre l'importance du choix de l'espace de probabilité sur lequel on travaille.

Exemple 2.5.3 *Dans une classe de* $n \leq 365$ *élèves, quelle est la probabilité de l'événement : "2 élèves au moins ont le même anniversaire" que l'on note* A *?*

On choisit comme espace de probabilité $\Omega = \{f : [1,n] \mapsto [1,365]\}$ *où, pour* $1 \leq i \leq n$, $f(i)$ *représente le jour d'anniversaire du* $i^{\text{ème}}$ *élève dans l'ordre alphabétique.*

Même si les naissances ne sont pas vraiment équiréparties au long de l'année, on munit Ω *de la probabilité uniforme. On a* $\operatorname{card}(\Omega) = 365^n$. *Pour calculer la probabilité de* A, *on peut calculer la probabilité de l'événement contraire* A^c *: "tous les élèves ont des dates d'anniversaire différentes".*

On a $A^c = \{f : [1,n] \mapsto [1,365] \text{ injective}\}$.

Donc $\mathbb{P}(A^c) = \dfrac{365}{365} \times \dfrac{364}{365} \times \cdots \times \dfrac{365-n+1}{365}$,

et $\mathbb{P}(A) = 1 - \dfrac{365}{365} \times \dfrac{364}{365} \times \cdots \times \dfrac{365-n+1}{365}$.

On peut vérifier que, dès que $n \geq 23$, *cette probabilité est supérieure à* $\frac{1}{2}$.

2.6 Probabilité conditionnelle

La notion de probabilité conditionnelle permet de prendre en compte l'information dont on dispose (à savoir qu'un événement B est réalisé) pour actualiser la probabilité que l'on donne à un événement A.

Définition 2.6.1 *Soit* Ω *muni d'une probabilité* $\mathbb{P}$ *et* $A, B \subseteq \Omega$. La probabilité conditionnelle de l'événement A sachant l'événement B *est notée* $\mathbb{P}(A \mid B)$ *et définie par*

$$\mathbb{P}(A \mid B) = \begin{cases} \frac{\mathbb{P}(A\cap B)}{\mathbb{P}(B)} & \text{si } \mathbb{P}(B) > 0 \\ \mathbb{P}(A) & \text{sinon.} \end{cases}$$

Exemple 2.6.2

1. *Dans une famille qui comporte deux enfants, l'un est une fille. On cherche la probabilité que l'autre soit un garçon.*
   On choisit $\Omega = \{FF, FG, GF, GG\}$ *où, par exemple,* FG *signifie que l'aîné des enfants est une fille et le second un garçon. Cet espace est muni de la probabilité uniforme.*
   On note
   $$\begin{aligned} A &= \{\text{un des enfants est un garçon}\} = \{FG, GF, GG\}, \\ B &= \{\text{un des enfants est une fille}\} = \{FF, FG, GF\}. \end{aligned}$$
   On a $\mathbb{P}(B) = \frac{\operatorname{card}(B)}{\operatorname{card}(\Omega)} = \frac{3}{4}$. *Comme* $A \cap B = \{FG, GF\}$, $\mathbb{P}(A \cap B) = \frac{\operatorname{card}(A\cap B)}{\operatorname{card}(\Omega)} = \frac{1}{2}$. *Donc la probabilité recherchée est* $\mathbb{P}(A \mid B) = \frac{\mathbb{P}(A\cap B)}{\mathbb{P}(B)} = \frac{1/2}{3/4} = \frac{2}{3}$.
2. *On suppose maintenant que l'aîné des enfants est une fille. On veut alors connaître la probabilité pour que l'autre soit un garçon.*
   En reprenant la démarche ci-dessus, on obtient que cette probabilité vaut $\frac{1}{2}$.

Dans certains problèmes, ce sont les probabilités conditionnelles que l'on connaît naturellement et on est amené à utiliser la définition sous la forme $\mathbb{P}(A \cap B) = \mathbb{P}(A \mid B)\mathbb{P}(B)$.

Exemple 2.6.3 *Parmi* 10 *pièces mécaniques,* 4 *sont défectueuses. On prend successivement deux pièces au hasard dans le lot (sans remise). Quelle est la probabilité pour que les deux pièces soient correctes ?*

On note A_1, *l'événement "la première pièce est bonne" et* A_2 *l'événement "la seconde pièce est bonne". Comme, au départ, il y a* 6 *pièces bonnes sur* 10, $\mathbb{P}(A_1) = \frac{6}{10} = \frac{3}{5}$. *Lorsque l'on a retiré une pièce bonne, il reste* 5 *pièces bonnes sur* 9. *D'où* $\mathbb{P}(A_2 \mid A_1) = \frac{5}{9}$. *On conclut que la probabilité cherchée est* $\mathbb{P}(A_1 \cap A_2) = \mathbb{P}(A_2 \mid A_1)\mathbb{P}(A_1) = \frac{5}{9} \times \frac{3}{5} = \frac{1}{3}$.

Le résultat suivant porte le nom de *formule de Bayes.*

Proposition 2.6.4 *Soit $B_1, \ldots, B_m$, une partition de Ω (c'est-à-dire, des sous-ensembles disjoints de Ω dont la réunion est Ω) et $A \subseteq \Omega$ tel que $\mathbb{P}(A) > 0$. Alors, pour tout $1 \leq i \leq n$,*

$$\mathbb{P}(B_i \mid A) = \frac{\mathbb{P}(A \mid B_i)\mathbb{P}(Bi)}{\sum_{j=1}^m \mathbb{P}(A \mid B_j)\mathbb{P}(B_j)}.$$

Démonstration
Le numérateur du second membre est égal à $\mathbb{P}(A \cap B_i)$. Le dénominateur vaut $\sum_{j=1}^m \mathbb{P}(A \cap B_j)$ et, comme les B_j forment une partition de Ω, il est égal à $\mathbb{P}(A)$. Donc le second membre est bien égal à $\frac{\mathbb{P}(A \cap B_i)}{\mathbb{P}(A)}$. $\square$.

Exemple 2.6.5 *Pour dépister une maladie, on applique un test sanguin. Si le patient est atteint, le test donne un résultat positif dans 99% des cas. Mais il se peut aussi que le résultat du test soit positif alors que le patient est en bonne santé et cela se produit dans 2% des cas. La proportion de personnes malades dans la population soumise au test est de 10^{-3}. Calculer la probabilité pour qu'un patient soit en bonne santé sachant que le résultat de son test est positif.*

Soit P l'événement "le test donne un résultat positif" et M l'événement "le patient est malade". On a $\mathbb{P}(P \mid M) = 0,99$, $\mathbb{P}(P \mid M^c) = 0,02$ et $\mathbb{P}(M) = 0,001$, d'où $\mathbb{P}(M^c \mid P) = \frac{\mathbb{P}(P|M^c)\mathbb{P}(M^c)}{\mathbb{P}(P|M^c)\mathbb{P}(M^c)+\mathbb{P}(P|M)\mathbb{P}(M)} = \frac{222}{233}$.

2.7 Indépendance

Définition 2.7.1 *Soit Ω muni d'une probabilité $\mathbb{P}$. Deux événements A et B sont dits* indépendants *si $\mathbb{P}(A \cap B) = \mathbb{P}(A) \times \mathbb{P}(B)$.*

Remarque 2.7.2 *L'indépendance de A et B se caractérise aussi par les relations $\mathbb{P}(A \mid B) = \mathbb{P}(A)$ ou $\mathbb{P}(B \mid A) = \mathbb{P}(B)$, c'est-à-dire que la probabilité donnée à l'événement A (resp. B) n'est pas modifiée par l'information que l'événement B (resp. A) est réalisé.*

Définition 2.7.3 *m évènements $A_1, \ldots, A_m$, sont dits* indépendants *si $\forall I \subseteq [1, m], \mathbb{P}\left(\bigcup_{i \in I} A_i\right) = \prod_{i \in I} \mathbb{P}(A_i)$.*

Remarque 2.7.4 *Il ne suffit pas que $\mathbb{P}(A_1 \cap \cdots \cap A_m) = \prod_{i=1}^m \mathbb{P}(A_i)$ pour que les événements soient indépendants. Pour que 3 événements soient indépendants, il ne suffit pas qu'il soient 2 à 2 indépendants.*

Contre-exemple :
Jet de deux pièces à Pile ou Face : $\Omega = \{PP, PF, FP, FF\}$ où, par exemple, PF signifie que la première pièce donne Pile et la seconde Face. Cet espace est muni de la probabilité uniforme. On note A l'événement "la première pièce donne Pile", B l'événement "la seconde pièce donne Face" et C l'événement "les deux pièces donnent le même résultat".

$$\begin{array}{rclrcll}
A & = & \{PP, PF\} & \mathbb{P}(A) & = & \frac{1}{2} & \\
B & = & \{PF, FF\} & \mathbb{P}(B) & = & \frac{1}{2} & \\
C & = & \{PP, FF\} & \mathbb{P}(C) & = & \frac{1}{2} & \\
A \cap B & = & \{PF\} & \mathbb{P}(A \cap B) & = & \frac{1}{4} & = \quad P(A)P(B) \\
A \cap C & = & \{PP\} & \mathbb{P}(A \cap C) & = & \frac{1}{4} & = \quad P(A)P(C) \\
B \cap C & = & \{FF\} & \mathbb{P}(B \cap C) & = & \frac{1}{4} & = \quad P(B)P(C) \\
A \cap B \cap C & = & \emptyset & \mathbb{P}(A \cap B \cap C) & = & 0 & \neq \quad P(A)P(B)P(C).
\end{array}$$

Ainsi les événements A, B et C sont 2 à 2 indépendants mais pas indépendants.

Chapitre 3

Variables aléatoires discrètes

3.1 Espace de probabilité

Dans le cas d'un espace Ω fini, nous avons défini un événement comme une partie quelconque de Ω. Mais si on souhaite modéliser le temps de première obtention de Pile dans une suite de jets d'une pièce à Pile ou Face, on choisit naturellement $\Omega = \{P, F\}^*$, ensemble qui est infini. Dès que Ω est infini, pour pouvoir construire une probabilité qui satisfasse des propriétés intuitives, il est souvent utile de restreindre les événements que l'on considère à une sous-classe de $P(\Omega)$ appelée *tribu*.

Définition 3.1.1 *Une tribu $\mathcal{A}$ sur Ω est une famille de parties de Ω qui vérifie les trois propriétés suivantes :*

1. $\emptyset, \Omega \in \mathcal{A}$.
2. $A \in \mathcal{A} \to A^c \in \mathcal{A}$.
3. *Si $(A_i)_{i\in I}$ est une famille dénombrable d'éléments de $\mathcal{A}$ alors $\bigcup_{i\in I} A_i$ est dans $\mathcal{A}$ (en d'autres termes, $\mathcal{A}$ est close par union dénombrable).*

On appelle événements *les éléments de $\mathcal{A}$.*

Remarque 3.1.2 *Si $\mathcal{A}$ est une tribu :*

1. *$\mathcal{A}$ est aussi close par intersection dénombrable, i.e. si $(A_i)_{i\in I}$ est une famille dénombrable d'éléments de $\mathcal{A}$ alors $\bigcap_{i\in I} A_i$ est dans $\mathcal{A}$ (car l'intersection est le complémentaire de l'union des complémentaire).*
2. *$\mathcal{A}$ est aussi appelée σ-algèbre.*

Exemple 3.1.3 *$\{\emptyset, \Omega\}$ est la plus petite tribu sur Ω. $P(\Omega)$ est la plus grosse tribu sur Ω. Si $A \subseteq \Omega$, $\{\emptyset, A, A^c, \Omega\}$ est une tribu sur Ω.*

Définition 3.1.4 *Soit Ω muni d'une tribu $\mathcal{A}$. On appelle* probabilité sur $(\Omega, \mathcal{A})$ *une application $\mathbb{P} : \mathcal{A} \mapsto [0, 1]$ qui vérifie*

1. $\mathbb{P}(\Omega) = 1$.
2. *Si $(A_i)_{i\in I}$ est une famille dénombrable d'éléments de $\mathcal{A}$ deux à deux disjoints, ($\forall i \neq j \in I, A_i \cap A_j = \emptyset$), alors $\mathbb{P}\left(\bigcup_{i\in I} A_i\right) = \sum_{i\in I} \mathbb{P}(A_i)$.*

Le triplet $(\Omega, \mathcal{A}, \mathbb{P})$ s'appelle un espace de probabilité*, la propriété 2 la* σ-additivité.

Dans toute la suite, nous travaillerons sans le spécifier nécessairement systématiquement sur un espace de probabilité $(\Omega, \mathcal{A}, \mathbb{P})$.

En utilisant la propriété de σ-additivité, on peut facilement vérifier que l'on a toujours pour tout couple d'événements A et B : $\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B) - \mathbb{P}(A \cap B)$. Les définitions de la probabilité conditionnelle de A sachant B et celle de l'indépendance des événements données au chapitre précédent restent également valables. Comme nous pouvons maintenant être amenés à considérer des familles infinies d'événements, nous préciserons simplement qu'une telle famille est indépendante si toute sous-famille finie l'est.

Définition 3.1.5 *On appelle* variable aléatoire discrète *une application $X : \Omega \mapsto F$ où F est un ensemble dénombrable. Pour $x \in F$, on note de façon concise $\{X = x\}$ l'événement $\{\omega \in \Omega : X(\omega) = x\}$. La famille des nombres $(\mathbb{P}(X = x))_{x\in F}$ s'appelle* la loi de X.

Notons que $(\{X = x\})_{x\in F}$ est une famille dénombrable d'événements deux à deux disjoints telle que $\bigcup_{x\in F}\{X = x\} = \Omega$. Donc, par la propriété de σ-additivité, $\sum_{x\in F} \mathbb{P}(X = x) = \mathbb{P}\left(\bigcup_{x\in F}\{X = x\}\right) = \mathbb{P}(\Omega) = 1$.

Exemple 3.1.6 *Dans le cas du jet de deux dés, la somme S des deux dés est une variable aléatoire discrète à valeurs dans $F = [2, 12]$ dont nous avons calculé la loi dans l'exemple 2.1.3 :*

$$\begin{aligned}
&\mathbb{P}(S = 2) = \mathbb{P}(S = 12) = \tfrac{1}{36}\\
&\mathbb{P}(S = 3) = \mathbb{P}(S = 11) = \tfrac{1}{18}\\
&\mathbb{P}(S = 4) = \mathbb{P}(S = 10) = \tfrac{1}{12}\\
&\mathbb{P}(S = 5) = \mathbb{P}(S = 9) = \tfrac{1}{9}\\
&\mathbb{P}(S = 6) = \mathbb{P}(S = 8) = \tfrac{5}{36}\\
&\mathbb{P}(S = 7) = \tfrac{1}{6}.
\end{aligned}$$

Il faut noter que la loi de S est la probabilité naturelle dont on doit munir l'ensemble $[2, 12]$ lorsque l'on s'intéresse à la somme de deux dés.

3.2 Indépendance

Définition 3.2.1 *n variables aléatoires discrètes $X_1, X_2, \ldots, X_n$ à valeurs respectivement dans $F_1, \ldots, F_n$ sont dites* indépendantes *si $\forall x_1 \in F_1, \ldots, \forall x_n \in F_n, \mathbb{P}(X_1 = x_1, \ldots, X_n = x_n) = \prod_{i=1}^{n} \mathbb{P}(X_i = x_i)$.*

Une famille quelconque de variables aléatoires discrètes est dite indépendante si toute sous-famille finie l'est.

3.3 Lois discrètes usuelles

Ce sont des lois qui portent sur $F \subseteq \mathbb{N}$.

Définition 3.3.1 (Loi de Bernoulli de paramètre p où $0 \leq p \leq 1$)
On dit que X suit la loi de Bernoulli de paramètre p et on note $X \sim \mathcal{B}(p)$ si : $\mathbb{P}(X = 1) = p$ et $\mathbb{P}(X = 0) = 1 - p$. On a alors $\forall x \in \{0, 1\}, \mathbb{P}(X = x) = p^x(1 - p)^{1-x}$ (convention : $0^0 = 1$).

Définition 3.3.2 (Loi binômiale de paramètres $n \in \mathbb{N} \setminus \{0\}$ et p où $0 \leq p \leq 1$)
C'est la loi de la somme $S = X_1 + \cdots + X_n$ de n variables de Bernoulli de paramètre p indépendantes $X_1, \cdots, X_n$. On a alors, pour $k \in F = \{0, \ldots, n\}$,

$$\begin{aligned}
\mathbb{P}(S = k) &= \mathbb{P}(X_1 + \cdots + X_n = k)\\
&= \mathbb{P}\left(\bigcup_{\substack{x_i \in \{0,1\}\\ x_1 + \cdots + x_n = k}} \{X_1 = x_1, \ldots, X_n = x_n\}\right)\\
&= \sum_{\substack{x_i \in \{0,1\}\\ x_1 + \cdots + x_n = k}} \mathbb{P}(X_1 = x_1, \ldots, X_n = x_n) \ (\sigma-\ \textit{additivité})\\
&= \sum_{\substack{x_i \in \{0,1\}\\ x_1 + \cdots + x_n = k}} \mathbb{P}(X_1 = x_1) \ldots \mathbb{P}(X_n = x_n) \ (\textit{indépendence})\\
&= \sum_{\substack{x_i \in \{0,1\}\\ x_1 + \cdots + x_n = k}} p^{x_1+\cdots+x_n}.(1-p)^{n-x_1-\cdots-x_n}\\
&= p^k.(1-p)^{n-k}.\mathrm{card}(\{(x_1, \ldots, x_n) \in \{0, 1\}^n : x_1 + \cdots + x_n = k\})\\
&= C_n^k p^k (1-p)^{n-k}
\end{aligned}$$

Si $\forall 0 \leq k \leq n$, $\mathbb{P}(S = k) = C_n^k p^k.(1 - p)^{n-k}$, on note $S \sim \mathcal{B}(n, p)$.

Définition 3.3.3 (Loi de Poisson de paramètre $\lambda > 0$)
On dit que N suit la loi de Poisson de paramètre $\lambda > 0$ et on note $N \sim P(\lambda)$ si $\forall n \in \mathbb{N}$, $\mathbb{P}(N = n) = e^{-\lambda}\frac{\lambda^n}{n!}$.

Définition 3.3.4 (Loi géométrique de paramètre $p \in]0,1]$) *C'est la loi du temps de premier succès dans une suite d'expériences aléatoires indépendantes où la probabilité de succès est p. Une telle suite se modélise à l'aide d'une suite $(X_i)_{i\geq 1}$ de variables indépendantes et identiquement distribuées suivant la loi de Bernoulli de paramètre p. L'événement "la i-ième expérience est un succès" s'écrit alors $\{X_i = 1\}$ et le temps T de premier succès est donné par $T = \inf\{i \geq 1 : X_i = 1\}$. Pour $k \geq 1$, en utilisant l'indépendance des X_i on obtient*

$$\begin{aligned} \mathbb{P}(T = k) &= \mathbb{P}(X_1 = 0, \dots, X_{k-1} = 0, X_k = 1) \\ &= \mathbb{P}(X_1 = 0) \cdots \mathbb{P}(X_{k-1} = 0)\mathbb{P}(X_k = 1) \\ &= (1-p)^{k-1} p. \end{aligned}$$

Si $\forall k \in \mathbb{N} \setminus \{0\}$, $\mathbb{P}(T = k) = p.(1-p)^{k-1}$, on note $T \sim Geo(p)$.

Exemple 3.3.5 *Soit $S \sim Geo(p)$ et $T \sim Geo(q)$ deux variables indépendantes. On cherche la loi de $Z = \min(S, T)$.*

1. *Pour $k \in \mathbb{N} \setminus \{0\}$, calculer $\mathbb{P}(S \geq k)$.*

$$\begin{aligned} \text{On a } \mathbb{P}(S \geq k) &= \sum_{i=k}^{\infty} \mathbb{P}(S = i) \\ &= \sum_{i=k}^{\infty} p.(1-p)^{i-1} \\ &= p.(1-p)^{k-1}.\sum_{i=0}^{\infty}(1-p)^i \\ &= p.(1-p)^{k-1}.\frac{1}{p} \\ &= (1-p)^{k-1} \end{aligned}$$

2. *En déduire $\mathbb{P}(Z \geq k)$.*

$$\begin{aligned} \mathbb{P}(Z \geq k) &= \mathbb{P}(S \geq k).\mathbb{P}(T \geq k) \\ &= (1-p)^{k-1}.(1-q)^{k-1} \end{aligned}$$

3. *Quelle est la loi de Z ?*

$$\begin{aligned} \mathbb{P}(Z = k) &= \mathbb{P}(Z \geq k) - \mathbb{P}(Z \geq k+1) \\ &= (1-p)^{k-1}.(1-q)^{k-1} - (1-p)^k.(1-q)^k \\ &= (1-p)^{k-1}.(1-q)^{k-1}[1 - (1-p).(1-q)] \\ &= (1-p)^{k-1}.(1-q)^{k-1}[p + q - pq] \end{aligned}$$

3.4 Espérance

Définition 3.4.1 (Espérance d'une variable aléatoire) *Soit $X : \Omega \mapsto \mathbb{R}$ une variable aléatoire discrète. Si $X(\Omega)$ est dénombrable, on pose $A = X(\Omega)$. Si la série $\sum_{a\in A} |a|\mathbb{P}(X = a)$ converge, on dit que l'espérance de X existe et on la note :*

$$E(X) = \sum_{a\in A} a\mathbb{P}(X = a)$$

Intuitivement, l'espérance correspond à la "valeur moyenne" de la variable aléatoire.

Exemple 3.4.2 *Jet d'un dé non-pipé : $\Omega = 1, 2, 3, 4, 5, 6, X(w) = w, \mathbb{P}$ distribution uniforme. $E(X) = \frac{1}{6}\sum_{i=1}^{6} i = \frac{7}{2}$.*

Proposition 3.4.3 (Espérance d'une somme) *Soient X et Y deux variables aléatoires. Si $E(X)$ et $E(Y)$ existent, alors $E(X + Y) = E(X) + E(Y)$.*

Démonstration
$X(\Omega) = A = \{a_1, a_2, ..., a_n, ...\}, Y(\Omega) = B = \{b_1, ..., b_n, ...\}$.

$$\begin{aligned}
E(X+Y) &= \sum_{i,j}(a_i + b_j)\mathbb{P}(X = a_i \text{ et } Y = b_j) \\
&= \sum_{i,j} a_i\mathbb{P}(X = a_i \text{ et } Y = b_j) + \sum_{i,j} b_j\mathbb{P}(X = a_i \text{ et } Y = b_j) \\
&= \sum_{i} a_i\mathbb{P}(X = a_i) + \sum_{j} b_j\mathbb{P}(Y = b_j) \\
&= E(X) + E(Y).
\end{aligned}$$

$\square$.

Exemple 3.4.4 *On tire deux dés, l'espérance de la somme est la somme des espérances, donc* 7.

Proposition 3.4.5 (Espérance d'un produit de variables indépendantes) *Si $X, Y : \Omega \mapsto \mathbb{R}$ sont des variables aléatoires indépendantes, alors si $E(X)$ et $E(Y)$ existent, $E(XY) = E(X) \times E(Y)$.*

Démonstration

$$\begin{aligned}
E(XY) &= \sum_{i,j} a_i b_j \mathbb{P}(X = a_i\ textrmet\ Y = b_j) \\
&= \sum_{i,j} a_i b_j \mathbb{P}(X = a_i)\mathbb{P}(Y = b_j)(\text{indépendance}) \\
&= \left(\sum_{i} a_i\mathbb{P}(X = a_i)\right)\left(\sum_{j} b_j\mathbb{P}(Y = b_j)\right)
\end{aligned}$$

$\square$.

Remarque 3.4.6 *En général, l'espérance du produit n'est pas égal au produit des espérances.*

3.5 Variance

Tout comme l'espérance, la variance (et sa racine carré, l'écart-type) est indispensable lorsque l'on fait des statistiques. Intuitivement, la variance mesure l'écart d'un ensemble de donnée par rapport à sa moyenne.

Définition 3.5.1 *Soit X une variable aléatoire telle que $E(X)$ existe. Alors on définit la* variance *de X de la manière suivante :*

$$Var(X) = E\left((X - m)^2\right) \text{ avec } m = E(X)$$

On note aussi $Var(X) = \sigma^2$, où σ s'appelle l'écart-type.

Remarque 3.5.2 $Var(X) = E(X^2 + m^2 - 2mX) = E(X^2) + m^2 - 2mE(X) = E(X^2) - E(X)^2$.

Proposition 3.5.3
- *Pour tout $\lambda \in \mathbb{R}, Var(\lambda X) = \lambda^2 Var(X)$*
- *Si X et Y sont deux variables indépendantes, $Var(X+Y) = Var(X) + Var(Y)$*

Demonstration laissée en exercice.

Chapitre 4

Application au problème de la coupe minimale

4.1 Préliminaires

Soit G un multigraphe (non-orienté) : si $x \neq y$ sont deux sommets, il peut y avoir plusieurs arêtes entre x et y, mais pas de boucle. On suppose G connexe (i.e. il existe toujours un chemin reliant deux sommets de G). Une coupe de $G = (V, E)$ (où V est l'ensemble des sommets et E est l'ensemble des arêtes) est un sous-ensemble $C \subseteq E$ tel que $(V, E - C)$ est non-connexe. On cherche une coupe de cardinal minimal.

Contraction d'arête : soit $e \in E$ une arête reliant x à y. À partir de G on construit un nouveau graphe G' tel que l'ensemble des sommets de G' soit $V' = V - \{x, y\} \cup \{v(x, y)\}$, où $v(x, y)$ est un nouveau sommet né de la fusion de x et y. Pour deux sommets u et v, $w(u, v)$ désigne le nombre d'arêtes reliant u à v dans G et $w'(u, v)$ désigne le nombre d'arêtes reliant u à v dans G'. Si $v(x, y) \notin \{u, v\}, w'(u, v) = w(u, v)$. Pour $u \neq v(x, y), w'(u, v(x, y)) = w(u, x) + w(u, y)$. En d'autres termes, les arêtes qui étaient reliées à x ou y dans G sont à présent reliées à $v(x, y)$, sauf les arêtes reliant x à y qui sont supprimées.

Remarque 4.1.1 *$\forall v \in V, degre(v) \geq$ cardinal de la coupe minimal (rappel : le degré d'un sommet est le nombre d'arêtes reliées à ce sommet).*

4.2 Algorithme

Tant que $|V| > 2$ **{**
    Choisir une arête et la contracter ; }
Retourner le nombre d'arêtes restantes.

Proposition 4.2.1 *L'algorithme retourne toujours une surestimation du cardinal de la coupe minimal.*

Proposition 4.2.2 *Toute coupe de G est une coupe de G'.*

Théorème 1
Soit C une coupe minimale de G. La probabilité d'obtenir C par l'algorithme est supérieure ou égale à $\frac{2}{n^2}$, où $n = |V|$.

Démonstration
Soit E_i l'évènement : "à la contraction i, l'algorithme ne contracte pas une arête de C". On cherche $\mathbb{P}(E_1 \cap E_2 ... \cap E_{n-2})$.

Soit $k = |C|$. Tous les sommets sont de degré supérieur ou égal à k donc $|E| \geq \frac{kn}{2}$ (car $|E| = \frac{1}{2} \sum_{v \in V} \text{degre}(v)$).

À l'étape 1 on choisit une arête de E au hasard :

$$\mathbb{P}(E1) = \frac{|E| - k}{|E|} \geq \frac{\frac{kn}{2} - k}{\frac{kn}{2}} = \frac{n-2}{n} = 1 - \frac{2}{n}.$$

Par les probabilités conditionnelles, on a :

$$\mathbb{P}(E_1 \cap E_2 \cap ... \cap E_i) = \mathbb{P}(E_1 \cap E_2 \cap ... \cap E_{i-1})\mathbb{P}(E_i | E_1 \cap E_2 \cap ... \cap E_{i-1})$$

On suppose que pendant les $i-1$ premières étapes on n'a contracté aucune arête de C. Dans le graphe courant, les sommets sont de degré supérieur ou égal à k (car toute coupe de G est une coupe de G') et le graphe courant compte $n-(i-1)$ sommets.

Donc $\mathbb{P}(E_i|E_1 \cap E_2 \cap ... \cap E_{i-1}) \geq \frac{k\frac{n-i+1}{2}-k}{k\frac{n-i+1}{2}} = 1 - \frac{2}{n-i+1}$

Finalement, par récurrence on arrive à :

$$\begin{aligned}
\mathbb{P}(E_1 \cap E_2 \cap ... \cap E_{n-2}) &\geq \prod_{i=1}^{n-2} \frac{n-i-1}{n-i+1} \\
&\geq \frac{2}{n(n-1)} \text{(élimination diagonale)} \\
&\geq \frac{2}{n^2}
\end{aligned}$$

$\square$.

4.3 Algorithme amélioré

Faire tourner l'agorithme de contraction r fois
Retourner le minimum des résultats obtenus

La probabilité d'erreur est majorée par $(1-\frac{2}{n^2})^r$. Pour obtenir une probabilité d'erreur inférieure à un ϵ donné, il suffit de prendre (les calculs sont très simples) :

$$r \geq \frac{n^2}{2} \ln\left(\frac{1}{\epsilon}\right)$$

C'est un algorithme de "Monte-carlo" (la probabilité d'erreur est supérieure à 0), par opposition aux algorithmes de "Las Vegas", où on obtient toujours le bon résultat (mais le temps de calcul est une variable aléatoire).

Chapitre 5

Chaînes de Markov

Les chaînes de Markov constituent l'exemple le plus simple des processus stochastiques, lorsque dans l'étude d'une suite de variables aléatoires, on abandonne l'hypothèse d'indépendance. Il s'agit d'un processus à temps discret - d'où le nom de “chaîne”.

5.1 Présentation

Définition 5.1.1 *Soit $(X_n)_{n\geq 0}$ une suite de variables aléatoires à valeurs dans un ensemble S, fini ou dénombrable. Dans la suite, pour simplifier les notations, on remplacera souvent S par la numérotation de ses éléments. Soit un ensemble $(p_{i,j})_{(i\in S, j\in S)}$ de probabilités de transition. On dit que la suite (X_n) est une* chaîne de Markov, *si, pour tout $n > 1$ et toute suite $(i_0,\ldots,i_{n-1},i_n)$ d'éléments de S, on a la relation suivante entre probabilités conditionnelles :*

$$\mathbb{P}(X_n = i_n | X_0 = i_0,\ldots,X_{n-1} = i_{n-1}) = \mathbb{P}(X_n = i_n | X_{n-1} = i_{n-1}) = p_{i_{n-1},i_n}. \tag{5.1}$$

Autrement dit, dans l'évolution au cours du temps, l'état du processus à l'instant n ne dépend que de celui à l'instant $n-1$ précédent, mais non de ses états antérieurs. Le processus est *sans mémoire* ou *non héréditaire.*

La probabilité $p_{i,j}$ est appelée *la probabilité de passage de l'état i à l'état j*, en une étape, ou en une opération, ou encore, en une transition.

Exemple 5.1.2 (Modèle de diffusion) *On dispose de deux boîtes contenant chacune r boules, et il y a au total $2r$ boules : r noires et r blanches. L'état du sytème est déterminé par le nombre de boules blanches dans la première boîte ($S = \{0,1,...,r\}$). Une transition consiste à tirer une boule de la première boîte et une de la deuxième et à les échanger.*

- $p_{i,j} = 0$ *si* $|j-i| \geq 2$
- $p_{i,i-1} = \frac{i^2}{r^2}$ *si* $i \geq 1$
- $p_{i,i+1} = \frac{(r-i)^2}{r^2}(i < r)$
- $p_{i,i} = \frac{2i(r-i)}{r^2}$.

Exemple 5.1.3 (Marche au hasard dans $\mathbb{Z}^k$ ($S = \mathbb{Z}^k$)) *À chaque étape, on se déplace d'une unité dans une direction : $(x_1,x_2,...,x_k)$ a 2^k voisins de la forme $(x_1,...,x_i \pm 1,...,x_k)$. Chaque probabilité de transition vaut $\frac{1}{2^k}$.*

Si $k \leq 2$, on revient toujours au point de départ. Si $k \geq 3$, on revient au point de départ avec une probabilité strictement inférieure à 1.

Définition 5.1.4 (Distribution) *On appelle* distribution *à l'instant n la famille Π^n : $(\Pi_i^n)_{i\,\in\,S}$ telle que $\forall i \in S, \Pi_i^n = \mathbb{P}(X_n = i)$.*

Remarque 5.1.5 *L'évolution d'une chaîne de Markov est complètement déterminée par les $p_{i,j}$ et par la distribution initiale Π^0.*

5.2 Représentation

En principe, on considère dans cette partie des chaînes de Markov à ensemble d'états fini. Cela ne nous empêchera pas par la suite d'utiliser ces représentations pour des chaînes avec ensembles d'états infinis !

Définition 5.2.1 *La matrice*

$$\mathcal{P} = \begin{pmatrix} p_{0,0} & p_{0,1} & p_{0,2} & \cdots \\ p_{1,0} & p_{1,1} & p_{1,2} & \cdots \\ \vdots & \vdots & \vdots & \vdots\vdots\vdots \end{pmatrix}$$

dont les coefficients sont les probabilités de transition $p_{i,j}$ *est appelée matrice* de passage *(ou* de transition*) de la chaîne.*

Toute matrice de transition $\mathcal{P} = (p_{i,j})_{i,j\in S}$ vérifie les propriétés suivantes :

1. pour tout couple (i, j), on a $p_{i,j} \geq 0$;
2. pour tout $i \in S$, on a $\sum_{j\in S} p_{i,j} = 1$.

Une matrice $\mathcal{P}$, qui vérifie ces deux conditions, est appelée *matrice stochastique.* Si $\mathcal{P}$ et $\mathcal{P}^T$ sont stochastiques alors $\mathcal{P}$ est *bistochastique.*

Définition 5.2.2 (Le graphe associé à une matrice de transition)
À toute matrice de transition, on peut associer son graphe. *Les* sommets *du graphe sont les différents états de la chaîne. Il y a une* flèche, *étiquetée* $p_{i,j}$, *entre le sommet étiqueté* s_i *et le sommet étiqueté* s_j *si, et seulement si, la probabilité de transition de l'état* s_i *à l'état* s_j *est strictement positive :* $p_{i,j} > 0$.

5.3 Persistance

Définition 5.3.1 *Un état i est dit* persistant *si, partant de l'état i, on retourne à l'état i avec une probabilité 1. Sinon i est dit* transitoire.

Pour un évènement A, on pose $p_i(A) = \mathbb{P}(A|X_0 = i)$. *La probabilité de se trouver un jour dans l'état* j *sachant que* $X_0 = i$ *est* $p_i(\cup_{n=1}^{+\infty} X_n = j)$, *que l'on note* $f_{i,j}$. *Soit* $f_{i,j}^{(n)}$ *la probabilité de passer pour la première fois en* j *à l'instant* n *sachant que* $X_0 = i$: $f_{i,j}^{(n)} = p_i(X_1 \neq j, X_2 \neq j, ..., X_{n-1} \neq j, X_n = j)$. *On a alors* $f_{i,j} = \sum_{n=1}^{+\infty} f_{i,j}^{(n)}$. *Par définition,*

$$\begin{aligned} i \text{ transitoire } &\Leftrightarrow f_{i,i} < 1 \\ i \text{ persistant } &\Leftrightarrow f_{i,i} = 1 \end{aligned}$$

Proposition 5.3.2
$$p_i(X_n = i \text{ pour une infinité de valeur de } n) = \begin{cases} 1 & \text{si } i \text{ est persistant} \\ 0 & \text{sinon} \end{cases}$$

En effet, $p_i(|\{n \geq 1; X_n = i\}| \geq k) = f_{i,i}^k = \begin{cases} & 1 \quad \text{si } i \text{ est persistant} \\ \longrightarrow_{k\to+\infty} & 0 \quad \text{si } i \text{ transitoire} \end{cases}$

Remarque 5.3.3
$$p_i(X_n = j \text{ pour une infinité de valeur de } n) = \begin{cases} 0 & \text{si } j \text{ est transitoire} \\ f_{i,j} & \text{si } j \text{ est persistant} \end{cases}$$

Théorème 2 *On note* $p_{i,j}^{(n)}$ *la probabilité d'être dans l'état* j *à l'instant* n *en partant de l'état* i. *On a le théorème suivant :*

$$i \text{ est transitoire } \Leftrightarrow \sum_{n=1}^{+\infty} p_{i,i}^{(n)} < +\infty$$

Démonstration

On suppose que $\sum_{n=1}^{+\infty} p_{i,i}^{(n)} < +\infty$. D'après le lemme de Borel-Cantelli (on ne l'a pas vu en cours, mais si ça vous intéresse, google pourra vous en dire plus), $p_i(X_n = i$ pour une infinité de valeurs de $n) = 0$. Donc i est transitoire.

Réciproquement, soit i transitoire (donc $f_{i,i} < 1$). On va montrer que $\sum_{n=1}^{+\infty} p_{i,i}^{(n)} \leq \frac{f_{i,i}}{1-f_{i,i}}$.

$$p_{i,i}^{(n)} = \sum_{s=1}^{n} f_{i,i}^{(s)} p_{i,i}^{(n-s)}$$

$$\begin{aligned}
\sum_{n=1}^{+\infty} p_{i,i}^{(n)} &= \sum_{n=1}^{+\infty}\sum_{s=1}^{n} f_{i,i}^{(s)} p_{i,i}^{(n-s)} \\
&= \sum_{s=1}^{+\infty}\sum_{n=s}^{+\infty} f_{i,i}^{(s)} p_{i,i}^{(n-s)} \\
&= \sum_{s=1}^{+\infty} f_{i,i}^{(s)} \sum_{n=0}^{+\infty} p_{i,i}^{(n)} \\
&\leq \left(\sum_{s=1}^{+\infty} f_{i,i}^{(s)}\right)\left(1+\sum_{n=1}^{+\infty} p_{i,i}^{(n)}\right).
\end{aligned}$$

Or, $\sum_{s=1}^{+\infty} f_{i,i}^{(s)} = f_{i,i}$. D'où le résultat. $\square$.

Exemple 5.3.4 (Marche au hasard sur $\mathbb{Z}$) *Est-ce que* 0 *est transitoire ?*

$$p_{0,0}^{(m)} = \begin{cases} 0 & \text{si } m \text{ impair} \\ \frac{C_{2n}^n}{2^{2n}} & \text{si } m = 2n \end{cases}$$

$\frac{C_{2n}^n}{2^{2n}} \sim \frac{1}{\sqrt{n\pi}}$ *(en utilisant l'équivalent de la factorielle).*

On en déduit que la série $\sum_{n=1}^{+\infty} p_{0,0}^{(n)}$ *diverge, donc l'état* 0 *(comme tout état* i*) est persistant.*

5.4 Chaînes irréductibles

Définition 5.4.1 *Une chaîne est dite irréductible si* $\forall\ (i,j)\ f_{i,j} > 0$. *En d'autres termes, de tout état il est possible atteindre n'importe quel autre état. Cela peut aussi s'écrire :* $\forall\ (i,j), \exists n \geq 1, p_{i,j}^{(n)} > 0$.

Théorème 3 *Pour une chaîne de Markov irréductible, l'une des deux propriétés suivante est vérifiée :*
(i) tous les états sont transitoires ; et dans ce cas on a :

$$\forall\ (i,j)\ p_i(X_n = j \text{ infiniment souvent}) = 0 \quad \sum_{n\geq 1} p_{i,j}^{(n)} < +\infty$$

(ii) tous les états sont persistants :

$$\forall\ (i,j)\ p_i(X_n = j \text{ infiniment souvent}) = 1 \quad \sum_{n\geq 1} p_{i,j}^{(n)} = +\infty$$

Démonstration cf. dernier cours $\square$.

Exemple 5.4.2 (Chaînes irréductibles finies) *Elles sont toutes persistantes. Soit* $S = \{1, ..., n\}$.

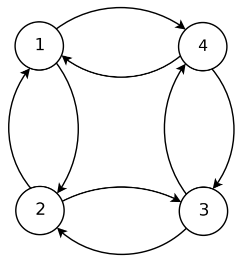
$$\forall t \geq 1, \sum_{j=1}^{n} p_{i,j}^{(t)} = \sum_{j=1}^{n} p_i(X_t = j) = 1$$

Supposons que la chaîne est transitoire, i.e. $\forall i,j, \sum_{t\geq 1} p_{i,j}^{(t)} < +\infty$.
Alors $\sum_{j=1}^{n}\sum_{t\geq 1} p_{i,j}^{(t)} < +\infty$ *(comme somme finie d'éléments finis).*
Or $\sum_{j=1}^{n}\sum_{t\geq 1} p_{i,j}^{(t)} = \sum_{t\geq 1}\sum_{j=1}^{n} p_{i,j}^{(t)} = +\infty$ *ce qui est absurde.*

5.5 Distributions stationnaires

Définition 5.5.1 *Une* distribution stationnaire *est une ditribution de probabilités* Π *sur l'ensemble des états telle que* $\mathcal{P}\Pi = \Pi$ *(*$\mathcal{P}$ *est la matrice de transition).*

Exemple 5.5.2



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

Chaque transition a une probabilité de $1/2$. *On cherche une distribution stationnaire* Π.

$$\begin{cases} \Pi_2 + \Pi_4 & = & 2\Pi_1 \\ \Pi_1 + \Pi_3 & = & 2\Pi_2 \\ \Pi_2 + \Pi_4 & = & 2\Pi_3 \\ \Pi_1 + \Pi_3 & = & 2\Pi_4 \end{cases}$$

D'où $\Pi_1 = \Pi_2 = \Pi_3 = \Pi_4 = 1/4$.
Évolution d'une chaîne :

$$\Pi^0 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \end{pmatrix} \quad \Pi^1 = \mathcal{P}\Pi^0 = \begin{pmatrix} 0 \\ 1/2 \\ 0 \\ 1/2 \end{pmatrix} \quad \Pi^2 = \begin{pmatrix} 1/2 \\ 0 \\ 1/2 \\ 0 \end{pmatrix} \quad \Pi^3 = \Pi^1$$

Exemple 5.5.3 (Modélisation d'une file d'attente) *Le temps d'attente de chaque client est 2 minutes. Pendant ce tmps il peut arriver 0, 1 ou 2 clients avec des probabilités respectives* t_0, t_1, t_2, *avec* $t_0 + t_1 + t_2 = 1$.

Un état est le nombre de clients en train d'attendre. L'espace des états S *est alors* $\mathbb{N}$. *Pour* $i \geq 1$, *on a* $p_{i,i-1} = t_0,\ p_{i,i} = t_1,\ p_{i,i+1} = t_2$. *On a de plus :* $p_{0,0} = t_0, p_{0,1} = t_1, p_{0,2} = t_2$. *La matrice de transition est donc :*

$$\mathcal{P} = \begin{pmatrix} t_0 & t_0 & 0 & 0 & 0 & \dots \\ t_1 & t_1 & t_0 & 0 & 0 & \dots \\ t_2 & t_2 & t_1 & t_0 & 0 & \dots \\ 0 & 0 & t_2 & t_1 & t_0 & \dots \\ 0 & 0 & 0 & t_2 & t_1 & \dots \\ \vdots & \vdots & \vdots & \ddots & \ddots & \ddots \end{pmatrix}$$

$$\mathcal{P}\Pi = \Pi \Longrightarrow \begin{cases} \Pi_0 & = & t_0\Pi_0 + t_0\Pi_1 \\ \Pi_1 & = & t_1\Pi_0 + t_1\Pi_1 + t_0\Pi_2 \\ \Pi_2 & = & t_2\Pi_0 + t_2\Pi_1 + t_1\Pi_2 + t_0\Pi_3 \\ \Pi_k & = & t_2\Pi_{k-1} + t_1\Pi_k + t_0\Pi_{k+1} \ \textit{pour } k \geq 3 \end{cases}$$

On en déduit :

$$t_0\Pi_{k+1} + (t_1 - 1)\Pi_k + t_2\Pi_{k-1} = 0$$

$$t_0\Pi_{k+1} - (t_0 + t_2)\Pi_k + t_2\Pi_{k-1} = 0$$

$$\begin{array}{rcl} \textit{Soit } Q(X) & = & t_0X^2 - (t_0 + t_2)X + t_2 \\ & = & (X - 1)(t_0X - t_2) \end{array}$$

Q a pour racines 1 et $\frac{t_2}{t_0}$. *Si* $t_2 = t_0$, *la solution générale est de la forme* $ak + b$. *Si* $t_2 \neq t_0$, *la solution générale est* $a + b\left(\frac{t_2}{t_0}\right)^k$.

On rappelle que l'on a les conditions supplémentaires suivantes sur Π *: tous les* Π_k *sont positifs, et la somme des* Π_k *est égale à* 1.

- *Si* $t_2 = t_0$, *la seule solution générale qui pourrait convenir pour que* $\sum \Pi_k$ *ne diverge pas est* $a = b = 0$. *Mais dans ce cas tous les* Π_k *sont nuls, on n'a donc pas de solution.*
- *Si* $t_2 > t_0$, *pour la même raison il n'y a pas de solution.*
- *Si* $t_2 < t_0$, *il faut* $a = 0$ *et la série des* Π_k *est une série géométrique de raison* < 1, *donc converge. On trouve* $\Pi_0 = bt_2, \Pi_1 = \frac{bt_2(1-t_0)}{t_0}$ *et* $b = \frac{t_0 - t_2}{t_2}$.