

Introduction aux Processus Stochastiques

Cours ESIEA 4A

Arnaud LIONNET

04/09/2015, 11:50

Table des matières

1	Processus stochastiques	4
1.1	Exemples de processus stochastiques	4
1.1.1	Marches aléatoires	4
1.1.2	Files d'attente	5
1.1.3	Processus de Galton–Watson	5
1.1.4	Ruine du joueur	6
1.2	Rappels de probabilité	7
1.2.1	Espace de probabilité, variables aléatoires, loi	7
1.2.2	Espérance	14
1.2.3	Exemples de lois “classiques”	17
1.2.4	Caractérisation d’une loi sur \mathbb{R}	19
1.2.5	Conditionnement	20
1.2.6	Indépendance	22
1.3	Généralités sur les processus stochastiques	23
1.3.1	Définition	23
1.3.2	Loi d’un processus et marginales	24
1.3.3	Temps d’arrêt	25
1.4	Exercices	26
2	Chaînes de Markov	27
2.1	Définition et descriptions d’une chaîne de Markov	27
2.1.1	Définition, exemples	27
2.1.2	Noyau, représentation matricielle	28
2.1.3	Graphe de transition	29
2.2	Distribution : marginales et lois invariantes	30
2.2.1	Marginales	30
2.2.2	Lois invariantes	32
2.3	Classification des états	34
2.3.1	Communication	34
2.3.2	Réurrence et transience	35
2.3.3	Réurrence positive et récurrence nulle	37

2.3.4	Périodicité	38
2.4	Calcul de la probabilité et de l'espérance du temps d'absorption	38
2.5	Comportement en temps long	40
3	Processus de Poisson et files d'attentes	43
3.1	Loi exponentielle de loi de Poisson	43
3.1.1	La loi exponentielle	43
3.1.2	La loi de Poisson	44
3.2	Indépendance et stationnarité des accroissements	45
3.2.1	Indépendance des accroissements	45
3.2.2	Stationnarité des accroissements	45
3.3	Processus de comptage et processus de Poisson	46
3.3.1	Processus de comptage	46
3.3.2	Processus de Poisson	47
3.4	Processus de Poisson généralisés	48
3.4.1	Processus de Poisson marqué	48
3.4.2	Processus de Poisson composé	49
3.5	Processus Markovien de saut	49
3.5.1	Description chaîne de Markov induite	50
3.5.2	Description "générateur"	50
3.5.3	Lien entre les deux descriptions	51
3.5.4	Lois invariantes	51
3.6	Exemples de modèles de files d'attente	52
3.6.1	$M/M/0$	52
3.6.2	$M/M/1$	52
3.6.3	$M/M/s$	53
3.6.4	$M/M/\infty$	54
3.6.5	$M/M/1/k$	54
3.6.6	$M/M/s/0$	54
4	Processus Gaussiens et mouvement Brownien	55
4.1	Loi normale et variables aléatoires Gaussiennes réelles	55
4.1.1	Loi standard	55
4.1.2	Loi normale, cas général	57
4.1.3	Le théorème central de la limite	57
4.2	Vecteurs Gaussiens	58
4.3	Processus Gaussiens	60
4.4	Différente caractérisation du mouvement brownien	61
4.5	Quelques exercices autour du mouvement brownien	63

Ces notes s'appuient sur le travail des précédents enseignants du cours de Processus Stochastiques, et notamment sur les notes développées par Charles-Edouard Bréhier pour sa version du cours. La version actuelle des notes contient localement un peu plus, un peu moins ou des choses légèrement différentes que ce qui sera traité en cours. Il est donc fortement recommandé de s'y reporter pour compléter le cours, mais cela n'est pas suffisant.

Quelques références :

Markov Chains, Brémaud.

Probability and Random Processes, Grimmett & Ztirzaker.

Probabilités, Ouvrard (volume 1 : probabilités élémentaires, volume 2 : plus avancé).

Chapitre 1

Processus stochastiques

1.1 Exemples de processus stochastiques

Un *processus stochastique*, ou *processus aléatoire* (c'est synonyme), est un phénomène aléatoire dépendant du temps. Un exemple est la météo : on connaît celle d'aujourd'hui ou d'hier, mais celle de demain, d'après demain ou de la semaine prochaine sont inconnues.

Passons en revue quelques modèles aléatoires classiques.

1.1.1 Marches aléatoires

La marche aléatoire simple sur \mathbb{Z} est un des exemples les plus élémentaires et dont les généralisations sont les plus variées.

Soit $(X_n)_{n=1,2,\dots}$ une suite de variables aléatoires indépendantes pouvant prendre uniformément (c'est-à-dire avec la même probabilité $1/2$) les valeurs 1 et -1 . On part de l'origine, $S_0 = 0$, et on pose ensuite $S_{n+1} = S_n + X_{n+1}$. S_n représente la position à l'instant n , tandis que le mouvement entre deux instants n et $n + 1$ est donné par X_{n+1} . On peut facilement généraliser le processus à la marche aléatoire dans \mathbb{Z}^2 si les X_n prennent uniformément les valeurs $(1, 0)$, $(-1, 0)$, $(0, 1)$ et $(0, -1)$, donc avec probabilité $1/4$ pour chaque cas. Ou plus généralement dans \mathbb{Z}^d pour tout entier d . Que peut-on dire du comportement de la position S_n ?

Ce processus est parfois surnommé la marche de l'étudiant ivre (ou tout autre variation autour de l'excès d'alcool), dans une ville quadrillée comme Manhattan. Une question que l'on peut se poser est : notre personnage finira-t'il par arriver chez lui ? Il se trouve que la réponse est oui, cela arrivera avec probabilité 1. Malheureusement pour un poisson ou un astronaute ivre, ce

résultat est vrai quand $d = 1$ ou 2 mais cesse de l'être dès que $d \geq 3$. Par contre, même pour $d \leq 2$, il se trouve que le temps moyen qu'il faut au marcheur ivre pour regagner sa maison est infini ...

De nombreuses généralisations de la marche aléatoire sont possibles. Le saut $X_n = S_n - S_{n-1}$ peut suivre d'autres lois. La marche peut avoir lieu sur tout autre graphe que \mathbb{Z}^d , qu'il soit fini ou infini. On peut aussi faire des marches renforcées. Par exemple, sur \mathbb{Z} , on met initialement un poids $w_i = 1$ sur chaque sommet i et l'on commence avec $S_0 = 0$. A chaque instant n , on est en $S_n = k$ et l'on saute vers $k - 1$ ou $k + 1$ avec des probabilités proportionnelles au nombre de fois w_{k-1} et w_{k+1} où les sommets $k - 1$ et $k + 1$ ont été utilisées, puis on rajoute un poids 1 sur le sommet où l'on vient d'atterrir. On peut montrer qu'avec probabilité 1, la marche aléatoire finit par se stabiliser sur 5 sommets.

On verra dans le chapitre 2 comment étudier ces processus quand ils n'ont "pas de mémoire", comme la marche aléatoire simple.

1.1.2 Files d'attente

Dans le 3^e chapitre du cours, on s'intéressera à la modélisation de files d'attente, par des processus comptant le nombre d'individus par file à chaque instant, avec différents choix du nombre de files, et différentes modélisations de l'arrivée de clients, et de leur traitement.

Le modèle le plus simple (et là aussi assez universel) est donné par le processus de Poisson (par référence au mathématicien français Poisson, du début du 19^{ème} siècle) : on ne considère pas le service des clients, seulement le processus de leur arrivée.

Les arrivées ont lieu à des instants aléatoires, une à la fois, avec une propriété d'absence de mémoire : la probabilité d'une arrivée après un instant t sachant qu'elle a lieu après l'instant s , ne dépend que de $t - s$. La durée entre deux arrivées suit alors une loi exponentielle (comme dans la désintégration radioactive), et le nombre de clients à un instant donné dans la file suit une loi particulière, la loi de Poisson.

1.1.3 Processus de Galton–Watson

Historiquement, ce modèle, ainsi que des variantes, a été introduit pour étudier la probabilité d'extinction d'un nom de famille dans un arbre généalogique. Il a donné lieu à de nombreuses généralisations en biologie des

populations. Il s'agit de compléter les approches déterministes (par des équations de récurrence, ou des équations différentielles), comme le modèle logistique, au comportement trop "rigide" ; ici on peut prendre en compte plus de "diversité".

On part d'un seul individu à l'instant initial : on pose $X_0 = 1$.

A chaque instant n , on a une population de X_n individus. Pour déterminer la taille de la population à l'instant suivant, on décide que chacun des individus, juste avant de mourir, donne naissance à un nombre aléatoire Z_{n+1}^k , pour $1 \leq k \leq X_n$, de façon indépendante des autres, et du passé, selon une loi de probabilité dite de reproduction, fixée initialement.

A chaque génération, on a une identité du type $X_{n+1} = \sum_{k=1}^{X_n} Z_{n+1}^k$.

La question est la suivante : quelle est la probabilité que la population s'éteigne (ce qui se produit quand les individus d'une génération ont tous aucune descendance) ?

Essentiellement, le résultat ne dépend que de la moyenne de la loi de reproduction : lorsqu'elle est inférieure ou égale à 1, il y a extinction presque sûre, tandis que si elle est strictement supérieure à 1, il y a une probabilité strictement positive que la population ne s'éteigne jamais.

Ce processus ne prend en compte que le nombre d'individus. Des modèles plus sophistiqués peuvent prendre en compte différentes caractéristiques des individus, comme les gènes qu'ils portent. On peut aussi considérer une version spatiale du processus de population où chaque individu a une position, et peut se déplacer entre les temps n et $n + 1$.

1.1.4 Ruine du joueur

On considère la situation d'un joueur contre la banque au casino : il dispose d'une fortune initiale notée a , et à chaque instant n sa fortune peut soit augmenter d'une unité, avec probabilité p , soit diminuer d'une unité, avec probabilité q (avec $p + q = 1$) ; les résultats (augmenter ou diminuer) à des instants distincts sont supposés indépendants.

Le jeu est stoppé soit lorsque la fortune du joueur devient nulle, soit quand c'est la banque qui ne peut plus payer, c'est-à-dire lorsque la fortune du joueur atteint une valeur fixée au départ, notée N .

Mathématiquement, la dynamique est la même qu'une marche aléatoire (biaisée lorsque $p \neq q$, en moyenne on a plus de chance d'augmenter ou de diminuer sa fortune, si $p > q$, ou l'inverse si $p < q$), sauf lorsqu'on atteint l'un des deux états 0 ou N : ensuite il ne se passe plus rien.

Le jeu se termine-t-il ou continue-t-il indéfiniment ? Combien de temps dure t-il en moyenne ? Une question naturelle dans ce contexte est de déter-

miner, en fonction de sa fortune initiale, la fortune du joueur au moment où le jeu s'arrête : plus précisément, quelle est la probabilité que l'arrêt du jeu soit dû à la ruine du joueur, ou à celle de la banque.

1.2 Rappels de probabilité

Dans cette section, on rappelle certains concepts de base des probabilités, notamment le formalisme moderne des probabilités, quelques lois de probabilité classiques, le conditionnement et l'indépendance.

1.2.1 Espace de probabilité, variables aléatoires, loi

Quand on parle d'une *probabilité* dans le langage courant, on ne parle pas toujours de la même chose. Par exemple, quand on dit que la probabilité de sortir un 6 avec un dé équilibré est de $\frac{1}{6}$ on comprend que si l'expérience du lancer de dé était répétée un grand nombre de fois, la fréquence d'apparition du 6 serait approximativement $\frac{1}{6}$. Mais lorsqu'on dit que les Cleveland Cavaliers ont 25% de chances de remporter les playoffs l'année prochaine, ou que les marchés estiment que la Grèce a 15% de chances de faire faillite d'ici la fin de l'année, il ne s'agit pas de fréquence ou d'expérience répétable ; ce qu'on exprime est un degré de certitude, ou une quantification d'un avis. Indépendamment de ces considérations philosophiques, les probabilités se sont développées depuis le 16^e–17^e siècle, mais sans qu'il ait jamais été clair ce qu'est vraiment une *variable aléatoire* ou un *évènement*, au vu du niveau clarté et de rigueur que l'on exige de nos jours.

Le formalisme introduit par Andreï Kolmogorov en 1933 donne un cadre formel aux probabilités. Il permet de décrire rigoureusement les évènements, variables aléatoires et probabilités en termes d'*éventualités*.

Éventualités, évènements, mesures de probabilités

Vocabulaire et intuition. On considère une situation aléatoire (météo de la semaine prochaine, résultat d'un lancer de pièce ou de dé, prix du pétrole dans 6 mois). On modélise cette situation à l'aide d'un ensemble Ω . Les éléments $\omega \in \Omega$ sont les *éventualités* (ou possibilités, ou scénarios, ou parfois échantillons ou réalisations). L'ensemble Ω est parfois appelé espace des éventualités, ou des possibilités.

Les *évènements* sont des sous-ensembles de Ω , c'est à dire des collections d'éventualités ω . Il faut y penser comme *ce que l'on peut observer*.

Exemple 1.1. Pour une situation qui consiste en un lancer de dé, on peut prendre $\Omega = \{1, 2, 3, 4, 5, 6\}$. Dans cette modélisation, les éventualités sont les résultats possibles du lancer. 2, 3 et 6 sont des éventualités. L'évènement $E = \{2, 4, 6\}$ n'est autre que l'évènement "le résultat est pair". L'évènement $\{6\}$ est l'évènement "le résultat est un 6".

Exemple 1.2. Imaginons maintenant un jeu basé sur la somme du résultat de deux dés, un bleu et un rouge. On peut prendre $\Omega = \{2, 3, 4, \dots, 11, 12\}$, chaque éventualité étant le résultat possible pour la somme. $G = \{2, 3, 4\}$ est l'évènement "la somme est inférieure ou égale à 4".

Mais on peut aussi modéliser la situation avec $\Omega = \{1, 2, 3, 4, 5, 6\}^2 = \{(i, j) : 1 \leq i, j \leq 6\}$ (l'ensemble des paires de résultats possibles pour chaque dé), où l'éventualité $(3, 6)$ signifie que le résultat du dé bleu est 3 et celui du dé rouge est 6. L'évènement $G =$ "la somme est inférieure ou égale à 4" est alors $\{(1, 1); (1, 2); (2, 1); (1, 3); (2, 2); (3, 1)\}$.

On dénote par \mathcal{F} l'ensemble de tous les évènements, et on l'appelle *tribu* (ou σ -*algèbre*) des évènements. Une *probabilité* \mathbb{P} assigne aux évènements un nombre dans $[0, 1]$, une masse.

Exemple 1.3. Dans le premier exemple ci-dessus, si le dé est équilibré, on assigne à chaque évènement élémentaire $\{i\}$ la probabilité $\frac{1}{6}$. Alors on a, par exemple, $\mathbb{P}(\{1\}) = \mathbb{P}(\{4\}) = \frac{1}{6}$. Pour l'évènement E ci-dessus, "le résultat est pair", on peut le décomposer en évènements élémentaires dont on connaît les probabilités : $E = \{2, 4, 6\} = \{2\} \cup \{4\} \cup \{6\}$, et ces évènements plus simples sont incompatibles (d'intersection nulle, deux à deux). Donc $\mathbb{P}(E) = \mathbb{P}(\{2\}) + \mathbb{P}(\{4\}) + \mathbb{P}(\{6\}) = \frac{1}{2}$. Pour l'évènement $D =$ "le résultat est 1 ou 6", on a

$$\mathbb{P}(D) = \mathbb{P}(\{1, 6\}) = \mathbb{P}(\{1\} \cup \{6\}) = \mathbb{P}(\{1\}) + \mathbb{P}(\{6\}) = \frac{1}{6} + \frac{1}{6} = \frac{1}{3}.$$

Cependant on peut considérer le cas d'un dé truqué tel que $\mathbb{P}(\{1\}) = \mathbb{P}(\{2\}) = \mathbb{P}(\{3\}) = \frac{6}{30}$ et $\mathbb{P}(\{4\}) = \mathbb{P}(\{5\}) = \mathbb{P}(\{6\}) = \frac{4}{30}$. Alors $\mathbb{P}(E) = \mathbb{P}(\{2\}) + \mathbb{P}(\{4\}) + \mathbb{P}(\{6\}) = \frac{14}{30}$ (strictement inférieur à $\frac{1}{2}$). Pour D on a

$$\mathbb{P}(D) = \mathbb{P}(\{1, 6\}) = \mathbb{P}(\{1\} \cup \{6\}) = \mathbb{P}(\{1\}) + \mathbb{P}(\{6\}) = \frac{6}{30} + \frac{4}{30} = \frac{1}{3}.$$

La donnée d'un triplé $(\Omega, \mathcal{F}, \mathbb{P})$ —l'espace Ω des éventualités, l'ensemble \mathcal{F} de tous les évènements et la mesure de probabilité \mathbb{P} — s'appelle un *espace de probabilité*.

Définitions rigoureuses. Pour qu’une fonction \mathbb{P} , qui donne une masse (≥ 0) aux évènements, soit une probabilité, il faut qu’elle satisfasse un minimum de propriétés intuitives.

Définition 1.4. $\mathbb{P} : \mathcal{F} \rightarrow [0, +\infty[$ est une probabilité, ou mesure de probabilité (ou loi de probabilité), si et seulement si elle satisfait les conditions suivantes :

1. $\mathbb{P}(\Omega) = 1$
2. Pour toute famille dénombrable $(A_i)_{i=1,2,\dots}$ d’évènements deux à deux disjoints (i.e. $A_i \cap A_j = \emptyset$ pour $i \neq j$), on a

$$\mathbb{P}\left(\bigcup_{i=1,2,\dots} A_i\right) = \sum_{i=1,2,\dots} \mathbb{P}(A_i).$$

La première propriété est simplement une condition de normalization, qui fixe la probabilité de l’évènement $\Omega =$ “il se passe quelque chose” à 1. La deuxième propriété s’appelle l’additivité. Elle a déjà été utilisée implicitement dans l’exemple précédent. Elle est extrêmement analogue à la propriété de la fonction cardinal qui à un ensemble renvoie son nombre d’éléments : si deux ensembles n’ont aucun élément en commun, le nombre d’éléments dans l’union est la somme des nombres d’éléments dans chacun. En fait ... la fonction cardinal est une *mesure*, mais pas une *mesure de probabilité*. Elle satisfait le point 2 dans la définition mais pas le 1. Si l’on regarde les sous-ensembles de $\{1, \dots, 36\} = \Omega$, avec $\text{card}(A) =$ nombre d’éléments dans A , alors la fonction $\mathbb{P} = \frac{1}{36}\text{card}$ qui à un sous-ensemble A associe $\frac{1}{36}\text{card}(A)$ est une mesure de probabilité. Elle donne la taille relative de A dans Ω .

Remarque 1.5. En principe, puisqu’une mesure de probabilité est une fonction sur \mathcal{F} , il faudrait donner sa valeur pour tous les évènements de \mathcal{F} . Mais par l’additivité, on peut facilement trouver la probabilité d’évènement plus complexes à partir de la probabilité d’évènements simples, comme on l’a fait pour les évènements D et E ci-dessus. Donc en pratique, la probabilité \mathbb{P} est *caractérisée* si on connaît sa valeur pour les évènements les plus simples.

Une autre propriété intuitive des probabilités, qui n’a pas été incluse dans la définition est que si un évènement en contient un autre, alors sa probabilité est plus grande. Pour reprendre le premier exemple du tirage d’un dé, il est clair que si le résultat est pair (évènement E), alors le résultat est supérieur ou égal à 2. Notons F ce dernier évènement, $F = \{2, 3, 4, 5, 6\}$. Alors que le dé soit truqué ou pas, peu importe les probabilités des différents évènements élémentaires, on doit avoir $\mathbb{P}(E) \leq \mathbb{P}(F)$ puisque $E \subseteq F$. On dit que \mathbb{P}

est une fonction *croissante*. Cette propriété découle en fait de la condition d'additivité, c'est pourquoi elle n'est pas dans la définition. De même, la probabilité de $\emptyset = \text{“impossible”}$ vaut toujours 0. (Montrez-le!)

Encore une propriété intuitive : si A est un évènement, la probabilité de son contraire A^c devrait être $1 - \mathbb{P}(A)$. Cette propriété découle de l'additivité et de la normalization $\mathbb{P}(\Omega) = 1$. (Montrez-le!)

[*Ce qui suit sur les tribus est hors-programme.*] Il y a un certain nombres de subtilités au sujet de la tribu \mathcal{F} . J'ai écrit plus haut que les évènements sont *des* sous-ensembles de Ω . Ce qui ne veut pas dire que *tous* les sous-ensembles de Ω sont toujours des évènements.

Le problème. Quand on considère $\Omega = [0, 1]$, quel pourrait être l'équivalent de la fonction “taille relative” vue ci-dessus ? Sous l'axiome du choix, on peut montrer qu'il est *impossible* d'avoir une probabilité \mathbb{P} sur l'ensemble $\mathcal{F} = \mathcal{P}([0, 1])$ de toutes les parties de $[0, 1]$, telle que pour tout intervalle $[a, b]$ de $[0, 1]$ on ait $\mathbb{P}([a, b]) = b - a$. On doit donc être un peu moins ambitieux en général.

Ce qu'on voudrait : pouvoir déclarer tous les intervalles comme des évènements (car ce sont des ensembles assez élémentaires) ; leur donner une masse $\mathbb{P}([a, b]) = b - a$; et que ce soit suffisant.

La solution. La théorie de la mesure assure que c'est possible. On peut déclarer certains sous-ensembles qui nous intéressent comme étant des évènements, et leur donner une masse. Puis on combine les évènements élémentaires et on attribue une masse au résultat de façon consistante. En poussant ce procédé jusqu'au bout, on obtient une tribu au sens de la définition ci-dessus, sur laquelle on a une fonction de masse qui est une mesure de probabilité au sens de la définition ci-dessus.

Définition 1.6. On dit qu'un ensemble \mathcal{F} de sous-ensembles des Ω est une tribu si et seulement si :

1. $\Omega \in \mathcal{F}$ (et $\emptyset \in \mathcal{F}$) ;
2. si A est un évènement (i.e. $A \in \mathcal{F}$), son complémentaire est aussi un évènement : $A^c \in \mathcal{F}$;
3. si $(A_i)_{i=1,2,\dots}$ est une famille dénombrable d'évènements, alors son union aussi est un évènement : $\cup_{i \geq 1} A_i \in \mathcal{F}$.

La première propriété dit que “on n'a rien observé” et “on a observé quelque chose” sont toujours des évènements. La deuxième dit que si A est un évènement, c'est à dire si on peut observer A se réaliser, alors on peut aussi l'observer ne pas se réaliser, c'est à dire observer son complémentaire

se réaliser. La troisième dit que si les A_i sont des évènements, alors l'évènement "l'un d'entre eux au moins se réalise" est encore un évènement. Les lois de De Morgan impliquent qu'il en va de même pour l'intersection, "tous les évènements A_i se réalisent". (Montrez-le!)

Ce qu'il faut retenir. Il y a deux cadres dans ce cours.

1. Le cas discret : l'espace Ω est dénombrable. Dans ce cas il n'y a aucun soucis. On peut (si l'on veut, mais on n'est pas obligé) considérer la tribu $\mathcal{F} = \mathcal{P}(\Omega)$ faite de *tous* les sous-ensembles de Ω , et se donner la valeur de \mathbb{P} sur les évènements élémentaires : les singletons.
2. Le cas indénombrable réel : Ω est \mathbb{R} , $[0, +\infty[$, $[0, 1]$, ou autre. Dans ce cas on se donne comme évènements élémentaires les intervalles. La tribu engendrée en combinant tous les intervalles se note $\mathcal{B}(\mathbb{R})$ (respectivement $\mathcal{B}([0, +\infty[)$, etc) et s'appelle la tribu Borélienne. Elle ne contient pas tous les sous-ensembles de \mathbb{R} . On se donne la valeur de \mathbb{P} pour les intervalles. La masse de tout autre évènement dans $\mathcal{B}(\mathbb{R})$ s'en déduit.

Variable aléatoire

On a vu jusqu'à présent de quoi parler des probabilités pour des évènements. On peut décrire rigoureusement le fait qu'il pleuvra ou non demain, mais pas encore la température qu'il fera demain (enfin ... a priori).

Une *variable aléatoire* X est une fonction de l'espace Ω des éventualités à valeurs dans un ensemble S , généralement appelé *espace des états*. Les valeurs $X(\omega)$ pour diverses éventualités $\omega \in \Omega$ sont appelées *réalisations* de X .

Il y a deux cadres importants. Lorsque S est un ensemble de valeurs réelles, comme dans les cas $S = \mathbb{Z}$, $S =]a, b]$, $S = [0, +\infty[$ ou $S = \mathbb{R}$, on parle de *variable aléatoire réelle*. Lorsque S est un ensemble dénombrable, comme dans les cas $S = \mathbb{Z}$, $S = \{0, 1\}$ ou $S = \{\times, \square, \triangle, \circ\}$, on parle de *variable aléatoire discrète*. Ces deux cas ne sont bien sûr pas mutuellement exclusifs.

[*Ce qui suit sur la mesurabilité est hors-programme.*] En fait, la définition complètement rigoureuse d'une variable aléatoire rajoute une condition, qui est la *mesurabilité*.

D'abord, il faut mentionner que S doit être équipé d'une tribu de sous-ensemble, notée \mathcal{S} .

1. Dans le cas où S est dénombrable, on prend (toujours) comme tribu $\mathcal{S} = \mathcal{P}(S)$, engendrée par les singletons.

2. Dans le cas où S est une partie indénombrable de \mathbb{R} , on prend $\mathcal{S} = \mathcal{B}(S)$, engendrée par les intervalles.

Maintenant, pour que la fonction X soit une variable aléatoire, il faut que pour tout $A \in \mathcal{S}$, l'ensemble $\{\omega \in \Omega \mid X(\omega) \in A\}$ des éventualités pour lesquelles X est dans A est bien un évènement, c'est-à-dire est dans \mathcal{F} . On note cet évènement de façon abrégée $\{X \in A\}$. Il est difficile d'exhiber des fonctions non-mesurables. Dans ce cours, toutes les fonctions considérées seront mesurables.

Exemple 1.7. Considérons la somme X des résultats R_1 et R_2 du lancer de deux dés. On peut prendre comme espace $\Omega = \{1, \dots, 6\}^2$. L'évènement "la somme est inférieure ou égale à 3" est alors

$$\{X \leq 3\} = \{\omega \in \Omega \mid X(\omega) \leq 3\} = \{(1, 1); (1, 2); (2, 1)\}.$$

L'évènement "le résultat du premier dé est un 6" est

$$\{R_1 = 6\} = \{\omega \in \Omega \mid R_1(\omega) = 6\} = \{(6, 1); (6, 2); (6, 3); (6, 4); (6, 5); (6, 6)\}.$$

R_1 , R_2 et X sont des variables aléatoires. Si $\omega = (\omega_1, \omega_2)$ est une éventualité, on a ici $R_2(\omega) = \omega_2$ et $X(\omega) = \omega_1 + \omega_2$.

Exercice 1. On considère dans le tirage ci-dessus qu'on gagne la différence entre le résultat du premier dé et le résultat du deuxième dé, si cette différence est positive. Ecrire ce gain comme une variable aléatoire G , en fonction de R_1 et R_2 , puis expliciter les évènements $\{G \geq 4\}$ et $\{G = 0\}$.

Exemple 1.8. On considère trois lancers de pièce successifs et la variable aléatoire τ donnant le premier instant où pile apparaît (τ vaut $+\infty$ si on n'a que des faces). Ici on peut prendre $\Omega = \{P, F\}^3$, où $\omega = (\omega_1, \omega_2, \omega_3) = (F, P, P)$ est l'évènement où la première pièce donne face et les 2 suivantes pile. Alors $T(\omega) = \inf\{i \in \{1, 2, 3\} \mid \omega_i = P\}$.

Exercice 2. Décrire la variable aléatoire N donnant le nombre de piles apparus, et décrire l'évènement $\{N = 2\}$.

Loi d'une variable aléatoire

On a défini une variable aléatoire comme une fonction des éventualités. Bien souvent, on s'intéresse surtout à sa *loi*.

Définition 1.9. Soit Ω un espace d'éventualités, avec une tribu \mathcal{F} d'évènements et une mesure de probabilité \mathbb{P} sur les évènements, $\mathbb{P} : \mathcal{F} \rightarrow [0, 1]$. Soit $X : \Omega \rightarrow S$ une variable aléatoire discrète ou réelle. Sa loi, ou distribution, est la fonction μ_X définie comme suit.

1. Si X est discrète, pour tout $x \in S$ on définit $\mu_X(\{x\}) = \mathbb{P}(\{X = x\})$.
2. Si X est à valeurs réelles (indénombrables), pour tout intervalle $[a, b]$, on définit $\mu_X([a, b]) = \mathbb{P}(\{X \in [a, b]\})$.

En fait ... μ_X est une mesure de probabilité sur la tribu \mathcal{S} de S .

Remarque 1.10. On fait un certain nombre de simplifications d'écriture.

- Dans le cas où l'ensemble S est dénombrable, il y a une correspondance entre une mesure μ , c'est-à-dire une fonction donnant une masse aux sous-ensembles de S , et une fonction m donnant une masse à chaque élément de S . La correspondance se fait via $\mu(\{x\}) = m(x)$ pour les singletons, et ensuite pour tout sous-ensemble A de S on a $\mu(A) = \sum_{x \in A} m(x)$. On se permet donc parfois l'abus de notation d'écrire $\mu(x)$ au lieu de $\mu(\{x\})$.
- De même, pour les probabilités d'évènements, on écrit $\mathbb{P}(X = x)$ pour $\mathbb{P}(\{X = x\})$ et $\mathbb{P}(X \in [a, b])$ pour $\mathbb{P}(\{X \in [a, b]\})$.

Exercice 3. On lance deux dé équilibrés et on considère la somme $X = R_1 + R_2$ des deux résultats. Quelle est la loi de X ?

On suppose maintenant les dés non-équilibrés, avec $\mathbb{P}(R_1 = j) = \mathbb{P}(R_2 = j) = \frac{6}{30}$ si $j = 1, 2$ ou 3 et $\frac{4}{30}$ si $j = 4, 5$ ou 6 . Quelle est la loi de X ?

Exemple 1.11. Considérons le résultat R du lancer d'un dé équilibré. Il y a plusieurs façons de modéliser/formaliser cette situation.

- a) On prend $\Omega = \{1, \dots, 6\}$, et $\mathbb{P}(\{i\}) = \frac{1}{6}$ pour tout $i \in \Omega$. On prend $R(\omega) = \omega$, et $S = \Omega$. Alors la loi de R est telle que $\mu_R(\{i\}) = \frac{1}{6}$ pour tout $i \in S$.
- b) On prend $\Omega = \{i, +, \triangle, \square, \text{cinq}, \circ, \text{Batman}\}$. La probabilité \mathbb{P} est représentée par le vecteur ligne $(\frac{1}{6} \quad \frac{1}{6} \quad \frac{1}{6} \quad \frac{1}{6} \quad \frac{1}{6} \quad \frac{1}{12} \quad \frac{1}{12})$ ce qui signifie que

$$\mathbb{P}(\{i\}) = \frac{1}{6}, \dots, \mathbb{P}(\{\text{cinq}\}) = \frac{1}{6}, \mathbb{P}(\{\circ\}) = \frac{1}{12}, \mathbb{P}(\{\text{Batman}\}) = \frac{1}{12}.$$

On prend R définie par

$$R(i) = 1, \dots, R(\text{cinq}) = 5, R(\circ) = R(\text{Batman}) = 6,$$

et $S = \{1, \dots, 6\}$. Alors la loi de R est telle que $\mu_R(\{i\}) = \frac{1}{6}$ pour tout $i \in S$. En effet on a (par exemple et le reste peut se vérifier de même)

$$\mu_R(6) = \mathbb{P}(R = 6) = \mathbb{P}(\{\circ, \text{Batman}\}) = \dots = \frac{1}{12} + \frac{1}{12} = \frac{1}{6}.$$

c) On prend $\Omega = [0, 1[$. On prend pour \mathbb{P} la mesure qui à chaque intervalle associe sa longueur. On définit R par

$$R(\omega) = 1 \text{ si } \omega \in \left[0, \frac{1}{6}\right[, \quad 2 \text{ si } \omega \in \left[\frac{1}{6}, \frac{2}{6}\right[, \quad \dots$$

Alors la loi de R est telle que $\mu_R(\{i\}) = \frac{1}{6}$ pour tout $i \in S$. En effet on a (par exemple et le reste peut se vérifier de même)

$$\mu_R(6) = \mathbb{P}(\{R = 6\}) = \mathbb{P}\left(\left[\frac{5}{6}, 1\right[\right) = 1 - \frac{5}{6} = \frac{1}{6}.$$

On voit donc qu'il y a en général plus d'une façon de décrire une situation aléatoire. Le "détail de l'implémentation", le choix d'un ensemble de scénarios Ω , peut varier. En revanche, la loi μ_X de la variable aléatoire reste la même. En un sens, la variable $X = \text{"résultat du dé"}$ reste aussi la même si on s'abstient de vouloir savoir ce que c'est au juste (une fonction sur Ω). Et l'évènement $\{X = 6\}$ reste aussi le même si on s'abstient de vouloir savoir ce que c'est (un sous-ensemble de Ω) et de l'écrire comme étant $\{\omega \in \Omega | X(\omega) = 6\}$.

1.2.2 Espérance

L'*espérance* d'une variable aléatoire est un résumé de sa distribution sous la forme d'un simple nombre. Elle donne vague une idée de la position générale (des réalisations) de la variable aléatoire. Plus précisément il s'agit de la moyenne des différentes réalisations pondérées par leurs probabilités.

Fixons un espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$ sur lequel est défini une variable aléatoire X , de loi μ_X . Pour pouvoir parler d'espérance (ou moyenne), il faut au minimum que X soit une variable aléatoire réelle (ou plus généralement à valeurs dans un espace vectoriel), ce que l'on suppose maintenant, et que l'on abrège en *v.a.r.* parfois.

Cas d'une variable aléatoire (réelle) discrète

On rappelle que c'est une variable aléatoire à valeurs dans un ensemble S au plus dénombrable. L'espérance de X est définie par

$$\mathbb{E}[X] = \sum_{x \in S} x \mathbb{P}(X = x) = \sum_{x \in S} x \mu_X(x).$$

Plus généralement, pour toute fonction $\phi = \mathbb{R} \rightarrow \mathbb{R}$ continue (disons),

$$\mathbb{E}[\phi(X)] = \sum_{x \in S} \phi(x) \mathbb{P}(X = x) = \sum_{x \in S} \phi(x) \mu_X(x).$$

Cas d'une variable aléatoire (réelle) continue

La variable aléatoire réelle X est dite *continue* si il existe une fonction positive $f_X : \mathbb{R} \rightarrow [0, +\infty[$ satisfaisant $\int_{-\infty}^{+\infty} f_X(x) dx = 1$ et telle que pour tout intervalle $[a, b] \subseteq \mathbb{R}$ on ait

$$\mu_X([a, b]) = \mathbb{P}(X \in [a, b]) = \int_a^b f_X(x) dx.$$

En prenant un petit intervalle de taille $\delta x > 0$ autour du point $x \in \mathbb{R}$, par exemple $I = [x - \frac{\delta x}{2}, x + \frac{\delta x}{2}]$, on voit que

$$\mu_X(I) = \mathbb{P}\left(x - \frac{\delta x}{2} \leq X \leq x + \frac{\delta x}{2}\right) = \int_{x - \frac{\delta x}{2}}^{x + \frac{\delta x}{2}} f_X(x) dx \approx f_X(x) \delta x.$$

Donc $f_X(x) dx$ donne la probabilité que X soit dans un intervalle de taille infinitésimale dx autour de x : $f_X(x) dx = \mathbb{P}(X = x \pm dx) =: \mu_X(dx)$.

Attention : $f(x) \neq \mathbb{P}(X = x)$! On voit en faisant $\delta x \rightarrow 0$ dans l'intégrale ci-dessous que pour une v.a. continue on a nécessairement $\mathbb{P}(X = x) = 0$.

Par contre, $f(x) = \lim_{\delta x \rightarrow 0} \frac{1}{\delta x} \mathbb{P}\left(x - \frac{\delta x}{2} \leq X \leq x + \frac{\delta x}{2}\right)$.

La fonction f s'appelle la *densité de probabilité de X* .

L'espérance de X est définie par

$$\mathbb{E}[X] = \int_{-\infty}^{+\infty} x f_X(x) dx \quad \text{que l'on écrit aussi} \quad \int_{-\infty}^{+\infty} x \mu_X(dx),$$

Noter comment définition de l'espérance pour une variable aléatoire réelle continue généralise la définition précédente pour le cas discret : les sommes discrètes $\sum_{x \in S}$ sont remplacées par des sommes continues $\int_{x \in \mathbb{R}}$ et les $\mathbb{P}(X = x) = \mu_X(x)$ sont remplacés par des $f_X(x) dx = \mu_X(dx)$.

Plus généralement, pour toute fonction $\phi = \mathbb{R} \rightarrow \mathbb{R}$ continue (disons),

$$\mathbb{E}[\phi(X)] = \int_{-\infty}^{+\infty} \phi(x) f_X(x) dx \quad \text{que l'on écrit aussi} \quad \int_{-\infty}^{+\infty} \phi(x) \mu_X(dx).$$

Exercice 4. On considère une v.a.r. continue X , ayant pour densité de probabilité f_X . On se donne $a < b$ dans \mathbb{R} , et $c \in \mathbb{R}$. Calculer $\mathbb{P}(X = c)$, $\mathbb{P}(X \in [a, b])$, $\mathbb{P}(X \in]a, b])$, $\mathbb{P}(X \in [a, b[)$, $\mathbb{P}(X \in]a, b[)$.

La question de l'intégrabilité

Les sommes infinies (si S est infini) et les intégrales écrites ci-dessus supposent que le terme général et l'intégrande soient *intégrables*.

\implies Revoir vos cours sur les séries et les intégrales.

Pour rappel rapide, tant que le terme général ou l'intégrande est positive, on peut toujours écrire des sommes infinies ou des intégrales, le résultat à valeur dans $[0, +\infty]$ est bien défini ($+\infty$ est inclus). Quand le terme général ou l'intégrande n'est pas de signe constant, il faut d'abord s'assurer qu'en prenant la valeur absolue de l'intégrande, le résultat est dans $[0, +\infty[$: valeur finie.

Propriétés de l'espérance

Propriété 1.12. Soient X et Y des v.a.r. et λ un réel.

- L'espérance est linéaire :

$$\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y] \quad \text{et} \quad \mathbb{E}[\lambda X] = \lambda \mathbb{E}[X].$$

- L'espérance préserve le signe : si $X \geq 0$, alors $\mathbb{E}[X] \geq 0$.
De façon équivalente, $X \geq Y \implies \mathbb{E}[X] \geq \mathbb{E}[Y]$.
- Si $X \geq 0$ et $\mathbb{E}[X] = 0$, alors $X = 0$.

Variance

La *variance* d'une variable aléatoire est elle aussi un résumé de sa distribution sous la forme d'un nombre. Elle donne une idée de l'étalement général de la variable aléatoire autour de son espérance. Plus précisément il s'agit de la moyenne (pondérée) des carrés des distances entre les réalisations et l'espérance. Autrement dit,

$$\text{Var}[X] = E\left[(X - E[X])^2\right].$$

Un simple calcul montre que

$$\text{Var}[X] = E[X^2] - E[X]^2.$$

Comme le montre clairement la première définition, la variance est toujours positive et l'on définit l'*écart-type* (ou *déviatation standard*) comme étant

$$\sigma_X = \text{Var}[X]^{\frac{1}{2}}.$$

L'écart-type a l'avantage de s'exprimer dans la même unité que X (si la v.a. X est une distance en m, la variance est en m^2 tandis que l'écart-type et en m aussi).

On peut déjà mentionner que si X et Y sont deux v.a.r. indépendantes (cf le rappel de la définition plus bas), $\text{Var}[X + Y] = \text{Var}[X] + \text{Var}[Y]$.

1.2.3 Exemples de lois “classiques”

Lois discrètes

La loi uniforme sur $S = \{1, \dots, N\}$. Elle est caractérisée par le fait que $\mu(i) = \frac{1}{N}$, pour tout $i \in S$. Il en résulte pour tout sous-ensemble $A \subseteq S$ on a $\mu(A) = \frac{\text{card}(A)}{N}$.

Noter qu'on peut aisément généraliser au cas d'une loi uniforme sur tout autre partie finie S de \mathbb{R} .

Exercice 5. Très souvent, dans le langage courant, quand on parle de “tirer quelque chose au hasard”, on sous-entend “suivant la loi uniforme”. Qu'en est-il de la loi uniforme sur \mathbb{N} ? Que veut dire “tirer un nombre entier au hasard”?

Notation : si une variable aléatoire X suit la loi uniforme sur S , c'est-à-dire quand la loi de X est la loi uniforme, on note $X \sim \mathcal{U}(S)$.

La loi de Bernoulli p . C'est une loi sur $S = \{0, 1\}$, paramétrée par un paramètre $p \in [0, 1]$. Elle est définie par $\mu(1) = p$ et $\mu(0) = 1 - p$.

Notation : si une variable aléatoire X suit la loi de Bernoulli de paramètre p , on note $X \sim \mathcal{B}(p)$.

Elle est souvent associée à l'idée de succès/échec dans une expérience aléatoire. Par exemple, si on tire à pile ou face avec une pièce équilibrée et que l'on regarde la v.a. X qui vaut 1 si pile sort, 0 si face sort, alors X suit la loi de Bernoulli de paramètre $p = \frac{1}{2}$. Si on tire un dé équilibré et que l'on regarde la v.a. X qui vaut 1 si le résultat est 6, 0 dans tous les autres cas, alors X suit la loi de Bernoulli de paramètre $p = \frac{1}{6}$.

La loi binomiale (n, p) . C'est une loi à deux paramètres, $n \in \mathbb{N}$ et $p \in [0, 1]$, sur les sous-ensembles de $S = \{0, \dots, n\}$. Elle est caractérisée par $\mu(k) = \binom{n}{k} p^k (1 - p)^{n-k}$ pour tout $k \in \{0, \dots, n\}$.

Notation : si une variable aléatoire X suit la loi de Bernoulli de paramètre p , on note $X \sim \mathcal{B}(n, p)$.

On peut interpréter une telle v.a. X comme la somme $Y_1 + \dots + Y_n$ de v.a. indépendante suivant la loi de Bernoulli $\mathcal{B}(p)$, c'est-à-dire du nombre de succès sur n tirages.

La loi géométrique p . C'est une loi à un paramètre, $p \in [0, 1]$, sur les sous-ensembles de \mathbb{N}^* . Elle est caractérisée par $\mu(k) = (1 - p)^{k-1}p$ pour tout $k \in \mathbb{N}^*$.

Notation : si une variable aléatoire X suit la loi géométrique de paramètre p , on note $X \sim \mathcal{G}(p)$.

On peut interpréter une telle v.a. X comme le numéro du premier succès lors de tirages indépendants suivant une loi $\mathcal{B}(p)$.

La loi de Poisson λ . C'est une loi à un paramètre, $\lambda \in [0, +\infty[$, sur les sous-ensembles de \mathbb{N} . Elle est caractérisée par $\mu(\{k\}) = \frac{\lambda^k}{k!} e^{-\lambda}$ pour tout $k \in \mathbb{N}$.

Notation : si une variable aléatoire X suit la loi géométrique de paramètre p , on note $X \sim \mathcal{P}(p)$. Voir chapitre 3 pour plus de détails.

Lois continues

Loi uniforme sur $S = [a, b]$. Elle est caractérisée par la densité de probabilités $f(x) = \frac{1}{b-a} 1_{[a,b]}(x) = \frac{1}{b-a}$ si $x \in [a, b]$, $= 0$ sinon.

Notation : $X \sim \mathcal{U}([a, b])$.

Exercice 6. Déterminer $\mu([x, y])$ pour $x, y \in \mathbb{R}$ (pas forcément seulement dans S). Déterminer $\mu(]-\infty, x])$ pour tout $x \in \mathbb{R}$.

Loi normale standard Elle est caractérisée par la densité sur \mathbb{R} définie par $f(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right)$.

Notation : $X \sim \mathcal{N}(0, 1)$. On verra au chapitre 4 le cas plus général $\mathcal{N}(m, \sigma^2)$.

Loi exponentielle λ Elle est caractérisée par la densité sur \mathbb{R} définie par $f(x) = \lambda \exp(-\lambda x)$ si $x \geq 0$, $= 0$ sinon.

Notation : $X \sim \mathcal{E}(\lambda)$. Voir chapitre 3 pour plus de détails.

Exercice 7. Déterminer l'espérance et la variance pour chacune des lois présentées.

1.2.4 Caractérisation d'une loi sur \mathbb{R}

Cette sous-section a pour but d'évoquer brièvement les différentes façons de caractériser une loi sur \mathbb{R} (c'est-à-dire pour les variables aléatoires avec $S \subseteq \mathbb{R}$). La question est de savoir, étant donné deux lois μ_1 et μ_2 , si elles sont égales. Plus souvent, dans la pratique, on cherche à identifier la loi μ_1 d'une variable aléatoire X , c'est à dire à montrer que c'est bien la loi connue μ_2 .

Bien sûr, comme μ_1 et μ_2 sont deux mesures de probabilité sur S , c'est-à-dire des fonctions $\mathcal{S} \rightarrow \mathbb{R}$, dire qu'elles sont égales c'est par définition dire que $\mu_1(A) = \mu_2(A)$ pour tout $A \in \mathcal{S}$. Mais cela fait souvent trop de sous-ensembles A à vérifier. D'après la remarque 1.5, il suffit de vérifier que pour tous les intervalles I (les ensembles élémentaires) on a $\mu_1(I) = \mu_2(I)$. (Pour des lois discrètes, même quand S n'est pas une partie de \mathbb{R} , la situation est souvent celle-là : on vérifie que $\mu_1(x) = \mu_2(x)$ pour tout $x \in S$.) Notons que pour des lois de variables aléatoires continues, si f_1 et f_2 sont les densités de probabilité, on peut aussi vérifier que $f_1 = f_2$.

On recense ci-dessous quelques alternatives.

Fonction de répartition

C'est la fonction

$$\begin{aligned} F_\mu : \mathbb{R} &\rightarrow \mathbb{R} \\ x &\mapsto \mu([-\infty, x]), = \mathbb{P}(X \leq x) \end{aligned}$$

quand μ est la loi d'une variable aléatoire X , auquel cas on note F_X sa fonction de répartition.

Il s'agit d'une fonction croissante, tendant vers 0 en $-\infty$, vers 1 en $+\infty$; dans le cas général, elle est continue à droite (et peut contenir des sauts, en nombre au plus dénombrable).

Dans le cas d'une variable aléatoire continue, F_X est toujours continue, et vérifie $F_X(x) = \int_{-\infty}^x f_X(t)dt$.

Dans le cas des variables discrètes, au contraire on a des plateaux, avec des sauts aux points n , dont la hauteur est donnée par la probabilité de cette valeur.

Si $F_{\mu_1} = F_{\mu_2}$, cela veut dire que $\mu_1([-\infty, x]) = \mu_2([-\infty, x])$ pour tout interval $I =]-\infty, x]$. Et ces intervalles particuliers génèrent tous les autres intervalles.

Espérance pour des fonctions-test

Supposons que μ_1 et μ_2 sont les lois de variables aléatoires X_1 et X_2 , pour fixer les idées. Si pour toute fonction $\phi : \mathbb{R} \rightarrow \mathbb{R}$ ou \mathbb{C} bornée on a

$$\mathbb{E}[\phi(X_1)] = \mathbb{E}[\phi(X_2)],$$

alors on a bien $\mu_1 = \mu_2$.

Fonction caractéristique

Si X est une variable aléatoire de loi μ , sa fonction caractéristique est

$$\begin{aligned}\varphi_\mu &= \varphi_X : \mathbb{R} \rightarrow \mathbb{C} \\ t &\mapsto \mathbb{E}[e^{itX}].\end{aligned}$$

L'idée est que si $\varphi_{X_1} = \varphi_{X_2}$, alors on a une grande famille de fonctions bornées $\phi_t : x \mapsto e^{itx}$ pour lesquelles $\mathbb{E}[\phi_t(X_1)] = \mathbb{E}[\phi_t(X_2)]$. Et cela suffit à identifier les lois.

Exercice 8. Déterminer la fonction de répartition et la fonction caractéristique pour les lois classiques ci-dessus.

1.2.5 Conditionnement

On rappelle ici quelques éléments du conditionnement de probabilités ou d'espérances.

Probabilité conditionnellement à un évènement

Définition 1.13. Soit A et B deux évènements. On suppose $\mathbb{P}(B) > 0$. La probabilité conditionnelle de A sachant B , notée $\mathbb{P}(A|B)$, est définie par

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

La possibilité de diviser par $\mathbb{P}(B)$ est assurée par l'hypothèse. La prise de l'intersection d'un ensemble avec B correspond à la connaissance a priori de la réalisation de l'évènement B . La division correspond à une normalisation : l'application

$$\begin{aligned}\mathbb{P}(\cdot|B) &: \mathcal{F} \rightarrow \mathbb{R} \\ A &\mapsto \mathbb{P}(A|B)\end{aligned}$$

est alors une loi de probabilités sur l'espace (Ω, \mathcal{F}) .

Quelques propriétés générales :

- si $B \subset A$, alors $\mathbb{P}(A|B) = 1$;
- A et B sont indépendants si et seulement si $\mathbb{P}(A|B) = \mathbb{P}(A)$ (cf sous-section suivante) ;
- $\mathbb{P}(A \cap B) = \mathbb{P}(A|B)\mathbb{P}(B) = \mathbb{P}(B|A)\mathbb{P}(A)$; d'où

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

- plus généralement, si $\Omega = \bigcup_{i \in I} B_i$, union au plus dénombrable ($I \subset \mathbb{N}$) et disjointe ($i \neq j \Rightarrow B_i \cap B_j = \emptyset$), telle que pour tout $i \in I$ on a $\mathbb{P}(B_i) > 0$, alors (formule dite des probabilités totales)

$$\mathbb{P}(A) = \sum_{i \in I} \mathbb{P}(A|B_i)\mathbb{P}(B_i);$$

- avec les notations précédentes, on a la formule de Bayes : si $\mathbb{P}(A) > 0$, pour tout $j \in I$,

$$\mathbb{P}(B_j|A) = \frac{\mathbb{P}(A|B_j)\mathbb{P}(B_j)}{\sum_{i \in I} \mathbb{P}(A|B_i)\mathbb{P}(B_i)}.$$

Espérance conditionnellement à un évènement

On ne traite ici que du cas d'une variable aléatoire discrète. La construction générale est plus complexe, et on n'en aura pas besoin.

Définition 1.14. *Considérons une variable aléatoire réelle discrète X à valeurs dans S , et un évènement B tel que $\mathbb{P}(B) > 0$. On définit l'espérance de X sachant B , notée $\mathbb{E}[X|B]$, comme étant la quantité*

$$\mathbb{E}[X|B] = \sum_{x \in S} x\mathbb{P}(X = x|B).$$

De même, si ϕ est une fonction de S dans \mathbb{R} , on a

$$\mathbb{E}[\phi(X)|B] = \sum_{x \in S} \phi(x)\mathbb{P}(X = x|B).$$

Espérance conditionnellement à une variable aléatoire

On considère maintenant une deuxième variable aléatoire Y à valeurs dans S .

Définition 1.15. On définit l'espérance de X sachant Y , notée $\mathbb{E}[X|Y]$, comme étant la variable aléatoire

$$\mathbb{E}[X|Y] = \sum_{y \in S: \mathbb{P}(Y=y) \neq 0} \mathbb{E}[X|Y = y] 1_{\{Y=y\}}.$$

C'est une fonction $g(Y)$ de la valeur prise par Y . Si l'écrit $S = \{y_1, y_2, y_3, \dots\}$, on peut alors écrire

$$\mathbb{E}[X|Y] = g(Y) = \begin{cases} \mathbb{E}[X|Y = y_1] & \text{quand } Y = y_1 \\ \vdots \\ \mathbb{E}[X|Y = y_j] & \text{quand } Y = y_j \\ \vdots \end{cases}.$$

On note deux propriétés simples donnant des possibilités de calcul :

- si $X = f(Y)$ pour une fonction f , alors $\mathbb{E}[X|Y] = X = f(Y)$;
- si X et Y sont indépendantes (voir ci-dessous pour le rappel de la définition), alors $\mathbb{E}[X|Y] = \mathbb{E}[X]$.

1.2.6 Indépendance

Définition 1.16. Soit $n \in \mathbb{N}$, et X_0, \dots, X_n des variables aléatoires à valeurs dans l'espace d'états $S \subseteq \mathbb{R}$. On dit qu'elles sont indépendantes si pour tout $A_0, \dots, A_n \in \mathcal{S}$ on a

$$\mathbb{P}(X_0 \in A_0, \dots, X_n \in A_n) = \mathbb{P}(X_0 \in A_0) \dots \mathbb{P}(X_n \in A_n).$$

Si I est un ensemble infini, une famille de variables aléatoires $(X_i)_{i \in I}$ est une famille de variables aléatoires indépendantes si toute sous-famille finie est constituée de variables aléatoires indépendantes : pour tout $n \in \mathbb{N}$, et tout $i_0, \dots, i_n \in I$, distincts deux à deux, les variables aléatoires X_{i_0}, \dots, X_{i_n} sont indépendantes.

Dans la définition ci-dessus, si S est discret, alors il suffit de prendre pour les A_i des singletons $\{x_i\}$ de S . Si les X_i sont réelles, on prend en toute généralité des intervalles $[a_i, b_i] \subseteq \mathbb{R}$.

Rappelons la notion d'indépendance 2 à 2 (pour une famille finie ou infinie indexée par I) : pour tous i, j distincts, X_i et X_j sont indépendantes. Attention : l'indépendance 2 à 2 n'entraîne pas l'indépendance, comme le démontre le contre-exemple suivant !

Exemple 1.17. On considère le lancer de deux dés à 6 faces, successivement. On définit les événements :

$$\begin{aligned} A &= \{\text{le lancer du dé 1 est impair}\} \\ B &= \{\text{le lancer du dé 2 est impair}\} \\ C &= \{\text{la somme des deux lancers est impaire}\}. \end{aligned}$$

A, B, C sont indépendants 2 à 2, mais pas "dans leur ensemble".

On a en fait la caractérisation suivante :

Proposition 1.18. Soit $n \in \mathbb{N}$. Les variables aléatoires X_0, \dots, X_n sont indépendantes si et seulement si pour toute famille ϕ_0, \dots, ϕ_n de fonctions $S \rightarrow \mathbb{R}$ bornées, on a

$$\mathbb{E}[\phi_0(X_0) \dots \phi_n(X_n)] = \mathbb{E}[\phi_0(X_0)] \dots \mathbb{E}[\phi_n(X_n)].$$

Si des variables aléatoires sont indépendantes et suivent toutes la même loi, on parle de *variables aléatoires indépendantes et identiquement distribuées*, abrégé en *va iid*.

Corollaire 1.19. Si X et Y sont deux v.a.r. indépendantes,

$$\text{Var}[X + Y] = \text{Var}[X] + \text{Var}[Y].$$

1.3 Généralités sur les processus stochastiques

D’abord une remarque sur le vocabulaire. Stochastique vient du grec (stokhastikos) tandis qu’aléatoire vient du latin (aleatorius) et veulent tous les deux dire la même chose. On peut donc parler de façon équivalente de *processus stochastique* ou de *processus aléatoire*. Mais pour quelque raison, l’expression processus stochastique semble un petit peu plus employée.

Un processus stochastique représente l’évolution au cours du temps d’un système dont la dynamique est aléatoire. Donc en plus de la dimension “aléatoire” des variables de la section précédente, on a maintenant aussi une dimension “temps”.

1.3.1 Définition

On se donne un espace de probabilité (Ω, \mathcal{F}, P) . On peut vouloir étudier un processus stochastique en temps discret ou en temps continu. Dans le cas du temps discret, on prend comme ensemble des temps $\mathbb{T} = \{0, 1, 2, \dots, T\}$

pour $T \in \mathbb{N}^*$, si l'on considère un horizon fini T , ou $\mathbb{T} = \mathbb{N}$, si l'on considère un horizon infini ($T = +\infty$). Dans le cas du temps continu, on prend comme ensemble des temps $\mathbb{T} = [0, T]$ pour $T \in \mathbb{R}_+^*$, si l'on considère un horizon fini T , ou $\mathbb{T} = \mathbb{R}_+^* = [0, +\infty[$, si l'on considère un horizon infini ($T = +\infty$).

Définition 1.20. *Un processus stochastique, sur Ω et à valeurs dans l'espace d'états S , est une famille $(X_t)_{t \in \mathbb{T}}$ de variables aléatoires, sur Ω et à valeur dans S .*

Autrement dit, pour chaque temps $t \in \mathbb{T}$, X_t est une variable aléatoire.

On peut s'intéresser, pour chaque temps t fixé de \mathbb{T} , à la variable aléatoire X_t , sa distribution μ_{X_t} , ses réalisations $X_t(\omega)$, etc. Mais on peut vouloir s'intéresser à ce qu'il se passe au cours du temps pour une réalisation donnée. Pour une éventualité $\omega \in \Omega$ fixée, la fonction $X.(\omega) : \mathbb{T} \rightarrow S$, $t \mapsto X_t(\omega)$ est appelée *trajectoire* associée à l'éventualité ω .

Une remarque (abstraite) : on peut considérer un processus stochastique $X = (X_t)_{t \in \mathbb{T}}$ comme étant en fait une seule variable aléatoire X sur Ω , mais à valeur dans l'espace des trajectoires $\mathcal{S} = S^{\mathbb{T}} = \text{Fonctions}(\mathbb{T} \rightarrow S)$. X est donc une *fonction aléatoire*, et ses réalisations $X.(\omega)$ sont précisément les trajectoires $t \mapsto X_t(\omega)$.

Exemple 1.21. Tous les exemples considérés dans la première section.

1.3.2 Loi d'un processus et marginales

Etant donné un processus stochastique $X = (X_t)_{t \in \mathbb{T}}$ sur l'espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$, la *loi du processus*, μ_X , est sa loi en tant que variable aléatoire à valeur dans l'espace des trajectoires. Dans ce cours, on n'en aura pas besoin. Pour chaque temps $t \in \mathbb{T}$, la loi μ_{X_t} de la variable aléatoire X_t est appelée *loi marginale* ou tout simplement *marginale* du processus au temps t .

Il y a deux fonctions associées au processus $X = (X_t)_{t \in \mathbb{T}}$ et qui permettent d'avoir une idée partielle de sa loi μ_X , de la même façon que l'espérance et la variance d'une v.a. réelle donnent une idée partielle de sa distribution. La première est la fonction "espérance au cours du temps :

$$m : \begin{array}{l} \mathbb{T} \rightarrow \mathbb{R} \\ t \mapsto m(t) = \mathbb{E}[X_t]. \end{array}$$

La seconde est la fonction de covariance :

$$c : \begin{array}{l} \mathbb{T} \times \mathbb{T} \rightarrow \mathbb{R} \\ (s, t) \mapsto c(s, t) = \text{cov}(X_s, X_t). \end{array}$$

On rappelle que pour deux variables aléatoires réelles X et Y , la covariance est définie par

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

On a naturellement $\text{cov}(X, X) = \text{Var}[X]$. Proche de la covariance, on a la corrélation qui en est une version renormalisée pour être toujours dans $[-1, 1]$:

$$\text{corr}(X, Y) = \frac{\text{cov}(X, Y)}{\text{Var}[X]^{\frac{1}{2}} \text{Var}[Y]^{\frac{1}{2}}} = \frac{\text{cov}(X, Y)}{\sigma_X \sigma_Y}.$$

Exercice 9. Prouver que la covariance est toujours dans $[-1, 1]$.

Exercice 10. On considère une suite de v.a. $(X_n)_{n \in \mathbb{N}^*}$ i.i.d. suivant la loi $\mathcal{B}(p)$, $p \in [0, 1]$, ainsi qu'une suite de v.a. $(Z_n)_{n \in \mathbb{N}^*}$ indépendantes entre elles et indépendante des X_n , telle que chaque $Z_n \sim \mathcal{B}(n, p)$. On définit $Y_0 = 0$ et pour tout $n \geq 1$, $Y_n = Y_{n-1} + X_n$.

On s'intéresse au processus $Y = (Y_n)_{n \geq 1}$ et $Z = (Z_n)_{n \geq 1}$. Quelles sont les marginales de chaque processus ? Les deux processus ont-ils la même loi ?

1.3.3 Temps d'arrêt

[Ce qui suit sur les temps d'arrêt est hors-programme.]

On veut fréquemment parler de temps qui sont aléatoires, car dépendant d'un processus aléatoire. Par exemple, dans un exemple de la section 2 on considèrerait le tirage répété d'une pièce (ce qui donne un processus stochastique) et on regardait le premier instant τ_0 où le lancer donnait pile. C'est une variable aléatoire.

Regardons deux autres exemples dans le même modèle de tirage de pièce. On considère le processus de richesse du joueur de pile ou face commençant avec 10 EUR, gagnant 1 EUR à chaque pile mais perdant 1 EUR à chaque face, et jouant 20 tours. Le joueur peut décider de s'arrêter au premier temps τ_1 où sa richesse s'est éloigné de 5 EUR de sa richesse initiale. Ou alors il peut décider de s'arrêter au temps τ_2 où sa richesse est maximale.

Comme τ_0 , ces temps τ_1 et τ_2 sont des variables aléatoires. Mais on sent qu'il y a un problème avec τ_2 : pour savoir s'il faut s'arrêter au temps $t = 9$, il faut connaître le futur !

Pour écarter ce genre de temps aléatoire, on utilise une notion de "bon temps aléatoire", qu'on appelle *temps d'arrêt*. Intuitivement, un temps d'arrêt est tel que, à tout temps fixe t dans \mathbb{T} , on peut savoir si le temps aléatoire

choisi pour s'arrêter est arrivé ou pas, c'est à dire on sait si l'évènement $\{\tau \leq t\}$ s'est réalisé ou pas, avec l'information disponible au temps t .

On ne donnera pas ici la définition propre d'un temps d'arrêt. On se contentera de savoir que tout temps aléatoire τ défini comme étant "le premier instant où quelque chose se passe" est bien un temps d'arrêt. A l'inverse, tout temps aléatoire défini comme étant "le dernier instant où quelque chose se passe", ou bien défini utilisant le maximum d'un processus aléatoire (comme τ_2) n'est pas un temps d'arrêt.

1.4 Exercices

Exercice 11. Soit $(X_n)_{n \in \mathbb{N}^*}$ une suite de variables aléatoires réelles iid, et soit A un sous-ensemble de \mathbb{R} (mesurable), par exemple A est un intervalle.

On pose $T_1 = \inf \{n \geq 1; X_n \in A\}$, et $T_2 = \inf \{n \geq T_1 + 1; X_n \in A\}$.

Montrer que T_1 et T_2 sont deux temps d'arrêt.

Déterminer la loi de T_1 , la loi du couple (T_1, T_2) . En particulier, voir que T_1 et T_2 sont finis presque sûrement.

En déduire la loi de T_2 . T_1 et T_2 sont-elles des variables aléatoires indépendantes ?

Déterminer la loi de $T_2 - T_1$; montrer que $T_2 - T_1$ et T_1 sont indépendantes et de même loi.

Exercice 12. Un jeu oppose deux adversaires A et B .

A chaque instant $n \in \mathbb{N}^*$, une partie a lieu ; A a la probabilité p de l'emporter, B la probabilité $q = 1 - p$ (pas de match nul). Les parties successives sont indépendantes.

Le jeu s'arrête dès que l'un des deux joueurs a gagné deux parties de plus que l'autre.

Déterminer la probabilité que A remporte le jeu.

En déduire celle que B le remporte, puis la probabilité que le jeu se poursuive indéfiniment, sans vainqueur.

Indications : remarquer que le jeu ne peut s'arrêter qu'après un nombre pair de parties. Puis introduire les événements suivants :

$$\mathcal{E}_k = \{\text{à l'instant } 2k \text{ aucun des joueurs n'a gagné le jeu}\}$$

$$\mathcal{F}_k = \{A \text{ gagne le jeu à l'instant } 2k\}.$$

Chapitre 2

Chaînes de Markov

Dans ce chapitre, on se fixe un espace d'états S dénombrable, fini ou infini. Comme vu dans le chapitre 1, on prend alors $\mathcal{S} = \mathcal{P}(S)$. Dans la théorie : sans perte de généralité, dans le cas fini on prendra $S = \{1, \dots, N\}$, et dans le cas dénombrable $S = \mathbb{N}$. Dans la pratique : il faut faire attention et savoir s'adapter !

2.1 Définition et descriptions d'une chaîne de Markov

2.1.1 Définition, exemples

On se place dans un espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$. On considère des processus stochastique à temps discret à valeurs dans S , c'est-à-dire qu'on travaille dans ce chapitre avec $\mathbb{T} = \mathbb{N}$.

Définition 2.1. *Un processus stochastique $(X_n)_{n \in \mathbb{N}}$ à valeurs dans S est une chaîne de Markov si : pour tout $n \in \mathbb{N}$, pour tout x_0, \dots, x_n dans S , vérifiant $\mathbb{P}(X_0 = x_0, \dots, X_{n-1} = x_{n-1}) \neq 0$, on a*

$$\mathbb{P}(X_n = x_n | X_{n-1} = x_{n-1}, \dots, X_0 = x_0) = \mathbb{P}(X_n = x_n | X_{n-1} = x_{n-1}). \quad (2.1)$$

Interprétation : si l'on a observé toute la trajectoire jusqu'au temps $n-1$, on n'a pas plus d'information au sujet de X_n que si on avait observé que la position actuelle $X_{n-1} = x_{n-1}$. Autrement dit, la position au prochain temps n ne dépend que de la position au temps présent $n-1$ et d'aléatoire supplémentaire. Le processus n'a pas de mémoire. La propriété exprimée par l'équation (2.1) s'appelle la propriété de Markov.

Convention importante : les états sont notés avec des lettres minuscules, les variables aléatoires avec des lettres majuscules.

On dispose d'une formulation équivalente de la propriété de Markov.

Proposition 2.2. *Pour tout $n \in \mathbb{N}$, pour tout x_0, \dots, x_n dans S , vérifiant $\mathbb{P}(X_0 = x_0, \dots, X_{n-1} = x_{n-1}) \neq 0$, on a*

$$\begin{aligned} \mathbb{P}(X_0 = x_0, \dots, X_{n-1} = x_{n-1}, X_n = x_n) \\ = \mathbb{P}(X_n = x_n | X_{n-1} = x_{n-1}) \dots \mathbb{P}(X_1 = x_1 | X_0 = x_0) \mathbb{P}(X_0 = x_0). \end{aligned} \quad (2.2)$$

Remarque : pour des raisons techniques liées à la définition des probabilités conditionnelles, il faut vérifier que tous les conditionnements ont lieu par rapport à des événements de probabilité non nulle ; l'hypothèse faite ici l'assure. En effet, pour tout $i = 0, \dots, n-1$, on a $\{X_0 = x_0, \dots, X_{n-1} = x_{n-1}\} \subseteq \{X_i = x_i\}$ donc $\mathbb{P}(X_i = x_i) > 0$.

Exercice 13 (Théorique). Prouver l'équivalence annoncée dans la proposition ci-dessus.

Exemple 2.3 (stupide). Une suite de variables aléatoires iid.

Exemple 2.4 (fondamental). Une marche aléatoire simple.

Exercice 14. Vérifier que les deux exemples précédents conviennent.

2.1.2 Noyau, représentation matricielle

Définition 2.5. *On appelle noyau de transition p une fonction*

$$\begin{aligned} P : S \times S &\rightarrow [0, 1] \\ (x, y) &\mapsto P(x, y) \end{aligned}$$

telle que, pour tout $x \in S$, l'application

$$\begin{aligned} P(x, \cdot) : S &\rightarrow [0, 1] \\ y &\mapsto P(x, y) \end{aligned}$$

est une mesure de probabilité, c'est-à-dire $\sum_{y \in S} P(x, y) = 1$.

Lien avec les chaînes de Markov : la fonction donnant les probabilités conditionnelles $\mathbb{P}(X_k = x_{k-1} | X_{k-1} = x_{k-1})$ vérifie cette condition pour $y = x_k$, et $x = x_{k-1}$ tel que $\mathbb{P}(X_{k-1} = x_{k-1}) = 0$. Pour se soustraire à cette condition technique, en fait on considère des processus issus de toutes les différentes conditions initiales possibles.

Définition 2.6. On dit que $(X_n)_{n \in \mathbb{N}}$ est une chaîne de Markov de noyau P et de loi initiale μ_0 si :

- c'est une chaîne de Markov au sens de la définition précédente ;
- la variable aléatoire X_0 suit la loi μ_0 , c'est-à-dire $\mathbb{P}(X_0 = x) = \mu_0(x)$, pour tout $x \in S$;
- les probabilités conditionnelles sont données par la fonction P :

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

Exemple de condition initiale : $X_0 = \tilde{x}_0$ presque sûrement, avec $\mu(x_0) = \delta(x_0, \tilde{x}_0)$ (symbole de Kronecker). En voyant que le cas général s'obtient par combinaison linéaire de ce cas particulier (quand on considère toutes les valeurs possibles de \tilde{x}_0), il suffit donc de comprendre le comportement lorsque la condition initiale est déterministe.

DANS LA PRATIQUE : on se donnera le noyau P en premier, ainsi que la loi initiale.

Cas d'un espace d'états fini $S = \{1, \dots, N\}$: le noyau de transition est une matrice carrée $N \times N$, qu'on appelle matrice de transition. On notera aussi $P_{xy} = P(x, y)$. Les coefficients d'une telle matrice vérifient deux propriétés :

- ils sont positifs ou nuls ;
- la somme des coefficients sur une ligne donne 1.

Exercice 15. Montrer que la deuxième condition ci-dessus est équivalente au fait que le vecteur $v = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$ est vecteur propre de P , associé à la valeur propre 1, c'est-à-dire $Pv = v$.

Lorsque S est infini : des matrices infinies... mais les formules restent les mêmes (en remplaçant sommes finies par des sommes de séries, convergentes).

2.1.3 Graphe de transition

On représente très souvent une chaîne de Markov à l'aide d'un graphe $G = (S, A, E)$ sur l'espace d'états S , appelé *graphe de transition*. Les sommets du graphe sont tous les états $x \in S$. Les arêtes orientées sont les couples $a = (x, y)$ tels que $P(x, y) > 0$. De plus, sur les arêtes orientées, on indique en étiquette la probabilité de transition correspondante (non nulle).

Remarque : on ne représente donc que les transitions qui peuvent se produire. D'ailleurs, pour étudier certaines propriétés des états que l'on verra plus bas, il n'est pas vraiment nécessaire de savoir quelle est la probabilité

$P(x, y)$ de passer de x à y , il suffit de savoir qu'elle est strictement positive. On se contente donc parfois du graphe réduit $G' = (S, A)$, sans les étiquettes.

Exercice 16. Représenter le graphe dans le cas de la marche aléatoire sur \mathbb{Z} et dans le cas de la ruine du joueur.

2.2 Distribution : marginales et lois invariantes

2.2.1 Marginales

Soit $X = (X_n)_{n \in \mathbb{N}}$ une chaîne de Markov sur l'espace d'états S , de noyau de transition P et de loi initiale μ_0 : pour tout état $x \in S$ on a $\mathbb{P}(X_0 = x) = \mu_0(x)$. Comme il a été dit dans le chapitre 1, les marginales du processus sont les lois μ_n des variables X_n . On va voir dans cette section comment les déterminer.

Par la formule des probabilités totales, puis en utilisant la propriété de Markov (2.2), pour tout entier $n \geq 1$ et tout état $x_n \in S$, on a

$$\begin{aligned} \mathbb{P}(X_n = x_n) &= \sum_{(x_0, \dots, x_{n-1}) \in S^n} \mathbb{P}(X_n = x_n, X_{n-1} = x_{n-1}, \dots, X_1 = x_1, X_0 = x_0) \\ &= \sum_{(x_0, \dots, x_{n-1}) \in S^n} \mathbb{P}(X_n = x_n | X_{n-1} = x_{n-1}) \dots \mathbb{P}(X_1 = x_1 | X_0 = x_0) \mathbb{P}(X_0 = x_0) \\ &= \sum_{(x_0, \dots, x_{n-1}) \in S^n} \mu(x_0) P(x_0, x_1) \dots P(x_{n-1}, x_n). \end{aligned}$$

Les sommes ci-dessus sont indexées par toutes les trajectoires possibles jusqu'à l'instant $n - 1$; le conditionnement de la position présente à l'instant n par son passé prend la forme du produit des probabilités de transition d'un instant à l'autre, avec une pondération supplémentaire par rapport à la position initiale à l'instant 0 : c'est la propriété de Markov !

On dispose également de la formule de récurrence suivante, en conditionnant uniquement par rapport à la position à l'instant précédent $n - 1$:

$$\begin{aligned} \mathbb{P}(X_n = x_n) &= \sum_{x_{n-1} \in S} \mathbb{P}(X_n = x_n, X_{n-1} = x_{n-1}) \\ &= \sum_{x_{n-1} \in S} \mathbb{P}(X_{n-1} = x_{n-1}) \mathbb{P}(X_n = x_n | X_{n-1} = x_{n-1}) \\ &= \sum_{x_{n-1} \in S} \mathbb{P}(X_{n-1} = x_{n-1}) P(x_{n-1}, x_n). \end{aligned}$$

Grâce à cette formule, on voit que connaissant la loi initiale μ et le noyau de transition P , on peut calculer de façon simple et itérative, la loi de la variable aléatoire X_n à n'importe quel instant n .

En fait, les équations précédentes ont une interprétation simple en terme de produit matriciel. De plus, de nombreuses quantités importantes concernant la dynamique des chaînes de Markov peuvent s'obtenir en résolvant des systèmes d'équations linéaires associés au noyau de transition.

Exercice 17. Revoir ses cours d'algèbre linéaire !

Vous connaissez le cas des matrices finies. Lorsque l'espace d'états est infini dénombrable, formellement rien n'est modifié, à condition de manipuler des séries convergentes...

La définition suivante n'en est pas vraiment une, elle précise juste une convention très utile.

Définition 2.7. On représente une loi de probabilité p sur S à l'aide d'un vecteur ligne : si $S = \{1, \dots, N\}$, il s'agit du vecteur $p = (p(1), \dots, p(N))$; si $S = \mathbb{N}$, on a un vecteur infini $(p(1), \dots, p(N), \dots)$.

On représente une fonction à valeurs réelles f définie sur S , à l'aide d'un vecteur colonne : $\begin{pmatrix} f(1) \\ \vdots \\ f(N) \end{pmatrix}$ dans le cas d'un espace d'états fini de cardinal N .

Les lois de probabilités sont donc identifiées avec les vecteurs lignes vérifiant deux conditions : (i) tous les coefficients sont positifs ou nuls ; (ii) leur somme est égale à 1.

Théorème 2.8. Pour tout $n \in \mathbb{N}$, si on note μ_n la loi de X_n , c'est-à-dire le vecteur ligne tel que $\mu_n(x_n) = \mathbb{P}(X_n = x_n)$, on a

$$\mu_n = \mu_0 P^n,$$

le produit matriciel du vecteur ligne μ_0 et de la matrice carrée $P^n = P \dots P$ (n fois).

Exercice 18. Démontrer le théorème à l'aide des formules données plus haut.

Bonus :

Théorème 2.9. Soit $f : S \rightarrow \mathbb{R}$ une fonction (supposée bornée si S est infinie).

On a $\mathbb{E}[f(X_n)] = \mu_0 P^n f$, où f est vu comme vecteur colonne.

Exercice 19. Démonstration ?

Un dernier résultat :

Théorème 2.10. *Pour tout $n \in \mathbb{N}^*$, P^n est une matrice de transition.*

De plus, pour tout $x \in S$, la ligne de P^n est une loi de probabilité ; précisément, c'est la loi de X_n lorsque la condition initiale est $X_0 = x$:

$$P^n(x, y) = \mathbb{P}(X_n = y | X_0 = x).$$

Une conséquence des théorèmes précédents :

Théorème 2.11 (Equation de Chapman-Kolmogorov). *Pour tous entiers $n, m \in \mathbb{N}$ et tous états $x, y \in S$,*

$$\begin{aligned} \mathbb{P}(X_{n+m} = y | X_0 = x) &= \sum_{z \in S} \mathbb{P}(X_{n+m} = y | X_n = z) \mathbb{P}(X_n = z | X_0 = x) \\ &= \sum_{z \in S} P^n(x, z) P^m(z, y). \end{aligned}$$

Précisément, cela découle de l'associativité du produit matriciel : $P^{n+m} = P^n P^m$.

2.2.2 Lois invariantes

On commence par une définition générale.

Définition 2.12. *Une chaîne de Markov d'espace d'états S est dite stationnaire si pour tout entier $m \in \mathbb{N}$, la loi du vecteur aléatoire de longueur $m+1$ (X_n, \dots, X_{n+m}) ne dépend pas de n .*

Le résultat est le suivant.

Théorème 2.13. *Une chaîne de Markov, de noyau de transition P , est stationnaire si et seulement si sa loi initiale vérifie (avec les conventions matricielles de la section précédente) :*

$$\mu_0 P = \mu_0.$$

Dans ce cas, pour tout $n \in \mathbb{N}$ on a $\mu_n = \mu_0$.

Cela motive la définition suivante, qui est le concept à retenir ici.

Définition 2.14. Une loi μ sur S est une loi invariante (pour le noyau de transition P) si et seulement si $\mu P = \mu$.

L'existence, l'unicité, le calcul des lois invariantes en fonction du noyau de transition, ou de façon équivalente en fonction de la structure du graphe de transition associé, sont des questions sur lesquelles on reviendra plus loin.

Le calcul d'une loi invariante μ revient à la résolution d'un système d'équations linéaires, sous la condition d'être une loi de probabilité (coefficients positifs ou nuls, et somme des coefficients égale à 1) :

$$\mu \text{ invariante} \Leftrightarrow \left(\forall j \in S, \mu(j) = \sum_{i \in S} \mu(i) P(i, j) \right).$$

Attention ! Ne pas résoudre le système $P\mu = \mu$, pour un vecteur colonne μ , comme il est courant en algèbre linéaire.

Exercice 20. Supposons que P est une matrice de transition, de taille $N \times N$, symétrique. Montrer que $(\frac{1}{N}, \dots, \frac{1}{N})$ définit une loi invariante.

Remarquer que dans le cas infini, formellement si $\mu = (1, \dots, 1, \dots)$, on a bien l'égalité $\mu P = \mu$, mais que μ n'est pas normalisable en une loi de probabilités.

[Ce qui suit sur les lois réversibles est hors-programme.]

La situation de l'exercice ci-dessus se généralise en la notion de loi réversible ; elle donne un moyen simple qui permet parfois de trouver une loi invariante.

Une loi μ est dite réversible pour le noyau de transition P si

$$\forall x, y \in S, \mu(x) P(x, y) = \mu(y) P(y, x).$$

Cette condition est appelée "condition d'équilibre détaillé", par rapport à définition de loi invariante, appelée "condition d'équilibre".

La formule ci-dessus est symétrique en x et y . L'interprétation de cette propriété se fait en termes de renversement du temps. Noter que l'existence ou non de lois réversibles dépend fortement du noyau de transition ; un noyau peut ne pas avoir de lois réversibles, mais des lois invariantes !

Le résultat est le suivant : une loi réversible est nécessairement invariante.

Exercice 21. Montrer cette propriété.

En quoi le cas d'une matrice de transition symétrique entre-t-il dans le cadre des lois réversibles ?

Définition 2.15. *Un vecteur ligne, non identiquement nul, μ tel que $\mu P = \mu$, à coefficients positifs, ne vérifiant pas nécessairement $\sum_{x \in S} \mu(x) = 1$, est appelé mesure invariante.*

Étant donnée une mesure invariante, on a deux possibilités :

- soit $\sum_{x \in S} \mu(x) < +\infty$: on peut alors définir une loi invariante $\bar{\mu}$ en normalisant μ : pour tout $x \in S$, $\bar{\mu}(x) = \frac{\mu(x)}{\sum_{y \in S} \mu(y)}$;
- soit $\sum_{x \in S} \mu(x) = +\infty$: on ne peut alors pas normaliser μ en une loi de probabilité.

2.3 Classification des états

2.3.1 Communication

Soit P un noyau de transition, sur l'ensemble S fini ou dénombrable. P est fixé, toutes les notions définies dans cette section dépendent de P .

Définition 2.16. *Si x et y sont deux états dans S , on dit que x conduit à y s'il existe un entier $n \geq 0$ tel que $P^n(x, y) > 0$. On note alors $x \rightarrow y$.*

Interprétation : $P^n(x, y)$ = probabilité de transition en n étapes de x à y . x conduit à y si cette probabilité est positive pour un certain entier n .

Autre écriture équivalente (cf interprétation au niveau du graphe) : il existe un entier $n \geq 0$, et une suite d'états, $x_0 = x, x_1, \dots, x_{n-1}, x_n = y$, telle que pour tout $0 \leq k \leq n-1$ on a $P(x_k, x_{k+1}) > 0$, c'est-à-dire que l'arête orientée reliant x_k à x_{k+1} est présente dans le graphe (réduit).

Lorsque $n = 0$, on a $P^n = Id$, donc $P^n(x, x) = 1$ pour tout $x \in S$. Conséquence : pour tout $x \in S$ on a $x \rightarrow x$, la relation est réflexive.

Transitivité de la relation (conséquence de l'équation de Chapman-Kolmogorov) :

$$\forall x, y, z \in S, (x \rightarrow y \text{ et } y \rightarrow z) \Rightarrow x \rightarrow z.$$

La symétrie n'est pas automatique : $x \rightarrow y$ n'entraîne pas $y \rightarrow x$. On introduit une notion supplémentaire.

Définition 2.17. *Si x et y sont deux états dans S , on dit que x et y communiquent si x conduit à y et y conduit à x . On note alors $x \leftrightarrow y$:*

$$x \leftrightarrow y \Leftrightarrow (x \rightarrow y \text{ et } y \rightarrow x).$$

La relation \leftrightarrow ainsi définie est réflexive, symétrique et transitive : c'est une relation d'équivalence.

Proposition 2.18. *L'espace d'états peut se partitionner en classes de communication : S s'écrit de façon unique comme une union d'ensembles C_i deux à deux disjoints, de telle sorte que pour tout $i \in I$ et tous $x, y \in S$,*

$$(x \in C_i \text{ et } y \in C_i) \Leftrightarrow x \leftrightarrow y.$$

Classe associée à l'état x : c'est l'unique élément C_i de la partition tel que $x \in C_i$. Alors, pour tout $y \in S$, $x \leftrightarrow y$ si et seulement si $y \in C_i$.

Définition 2.19. *Un noyau P est dit irréductible si tous les états communiquent.*

De façon équivalente : P est irréductible si et seulement si la partition de la Proposition précédente est triviale, au sens où il n'y a qu'un seul ensemble, $C_1 = S$.

Exemple 2.20. La marche aléatoire simple sur \mathbb{Z}^d est irréductible.

Exercice 22. Déterminer les classes de communication du noyau de transition associé au problème de la ruine du joueur.

Deux notions pour terminer :

Définition 2.21. *Un sous-ensemble C de S est dit fermé si pour tout $x \in C$ et tout $y \notin C$ on a $P(x, y) = 0$.*

Un état x tel que $C = \{x\}$ est fermé est appelé absorbant.

Interprétation : une fois entré dans un ensemble fermé par une dynamique Markovienne associée au noyau P , on n'en sort plus. Mieux : si la chaîne atteint un état absorbant, elle y reste.

2.3.2 Récurrence et transience

Définition 2.22. *Un état $x \in S$ est dit récurrent (pour le noyau P) si*

$$\mathbb{P}(X_n = x \text{ pour une infinité d'entiers } n | X_0 = x) = 1.$$

Dans le cas contraire, x est dit transient.

On définit deux quantités importantes :

$$T(x) = \inf \{n \in \mathbb{N}^*; X_n = x\} \quad \text{et} \quad N(x) = \text{Card} \{n \in \mathbb{N}; X_n = x\}.$$

$T(x)$ et $N(x)$ sont deux variables aléatoires à valeurs dans $\mathbb{N} \cup \{+\infty\}$.

On a la réécriture immédiate de la définition de la récurrence et de la transience :

$$\begin{aligned} x \text{ récurrent} &\Leftrightarrow \mathbb{P}(N(x) = \infty | X_0 = x) = 1 \\ x \text{ transient} &\Leftrightarrow \mathbb{P}(N(x) = \infty | X_0 = x) < 1. \end{aligned}$$

On a le théorème de caractérisation suivant.

Théorème 2.23.

– *Caractérisation de la récurrence :*

$$\begin{aligned} x \text{ récurrent} &\Leftrightarrow N(x) = +\infty \text{ presque sûrement sachant } X_0 = x \\ &\Leftrightarrow \mathbb{E}[N(x) | X_0 = x] = +\infty \\ &\Leftrightarrow \sum_{n=0}^{+\infty} P^n(x, x) = +\infty \\ &\Leftrightarrow \mathbb{P}(T(x) < +\infty | X_0 = x) = 1. \end{aligned}$$

– *Caractérisation de la transience :*

$$\begin{aligned} x \text{ transient} &\Leftrightarrow \mathbb{P}(N(x) = +\infty | X_0 = x) < 1 \\ &\Leftrightarrow \mathbb{E}[N(x)] < +\infty \\ &\Leftrightarrow \sum_{n=0}^{+\infty} P^n(x, x) < +\infty \\ &\Leftrightarrow \mathbb{P}(T(x) < +\infty | X_0 = x) < 1. \end{aligned}$$

- *Dans le cas transient : $N(x)$ suit une loi géométrique de paramètre $\mathbb{P}(T(x) < +\infty | X_0 = x)$, et en particulier en fait $N(x) < +\infty$ presque sûrement, sachant $X_0 = x$.*
- *On a en fait la formule générale : pour tout $n \geq 1$*

$$\mathbb{P}(N(x) = n | X_0 = x) = \mathbb{P}(T(x) < +\infty | X_0 = x)^{n-1} (1 - \mathbb{P}(T(x) < +\infty | X_0 = x)).$$

Remarque 2.24. D'après le troisième point du théorème, dans la caractérisation de la transience on peut remplacer $\mathbb{P}(N(x) = +\infty) < 1$ par $\mathbb{P}(N(x) = +\infty) = 0$, c'est-à-dire $N(x) < +\infty$ presque sûrement.

Interprétation de la formule générale : passer exactement n fois en x (en comptant la valeur à l'instant 0), partant de x , c'est revenir $n - 1$ fois de façon indépendante (conséquence de la propriété de Markov forte, hors-programme), ce qui se fait avec probabilité $\mathbb{P}(T(x) < +\infty | X_0 = x)$, puis ne pas revenir.

Proposition 2.25. *La récurrence et la transience sont des propriétés de classe : pour toute paire d'états x, y qui communiquent, alors x est récurrent si et seulement si y l'est.*

Remarque 2.26 (Culture). La marche aléatoire simple sur \mathbb{Z}^d est récurrente si $d = 1$ ou $d = 2$, et est transiente si $d \geq 3$.

On a le théorème suivant concernant la fermeture des classes de communications.

Théorème 2.27. *L'espace d'états S s'écrit de façon unique comme une partition $S = T \cup C_1 \cup C_2 \cup \dots$, où :*

- *T est l'ensemble des états transients ;*
- *pour tout indice i , C_i est une classe de communication constituée d'états récurrents, et est fermée.*

Une remarque : l'ensemble des états transients T peut être constitué de plusieurs classes de communication.

En effet, il suffit de voir qu'une classe de communication C_i est fermée si ses états sont récurrents : si $x \in C_i$ et $y \notin C_i$ on ne peut pas avoir $p(x, y) > 0$. Raisonnons par l'absurde : on aurait $x \rightarrow y$ par définition, donc y ne conduit pas à x (sinon $x \leftrightarrow y$) ; alors $\mathbb{P}(N(x) = 1 | X_0 = x) \geq p(x, y) > 0$ ce qui contredit la récurrence de x .

Exercice 23. Trouver un exemple de noyau de transition avec un ensemble d'états transients qui soit fermé. Trouver également un exemple où c'est le contraire qui se produit.

2.3.3 Récurrence positive et récurrence nulle

La transience est caractérisée par le fait que $\mathbb{P}(T(x) < +\infty | X_0 = x) < 1$, donc que $\mathbb{P}(T(x) = +\infty | X_0 = x) > 0$; dans ce cas, on a $\mathbb{E}[T(x) | X_0 = x] = +\infty$.

Dans le cas récurrent, en fait les deux cas sont possibles. Par exemple, dans le cas de la marche aléatoire simple sur \mathbb{Z} ou \mathbb{Z}^2 , on a aussi $\mathbb{E}[T(x) | X_0 = x] = +\infty$. Par contre, on verra que pour une chaîne de Markov irréductible sur un espace d'états fini, alors au contraire on a $\mathbb{E}[T(x) | X_0 = x] < +\infty$.

Ceci justifie la définition suivante :

Définition 2.28. *Soit x un état. On dit que x est récurrent positif si x est récurrent et si $\mathbb{E}[T(x) | X_0 = x] < \infty$. On dit que x est récurrent nul si x est récurrent et si $\mathbb{E}[T(x) | X_0 = x] = \infty$.*

La récurrence positive et la récurrence nulle sont des propriétés de classe : pour toute paire d'états x, y qui communiquent, alors x est récurrent positif si et seulement si y l'est.

2.3.4 Périodicité

Définition 2.29. Soit $x \in S$ un état.

On définit la période de x par : $d(x) = \text{pgcd}\{n \geq 1; P^n(x, x) > 0\}$.

On dit que x est apériodique si $d(x) = 1$.

Proposition 2.30. La période et l'apériodicité sont des propriétés de classe : si $x \leftrightarrow y$, alors $d(x) = d(y)$.

Un critère simple pour prouver l'apériodicité : si $P(x, x) > 0$, alors $d(x) = 1$.

Exemple 2.31. Dans le cas de la marche aléatoire simple, la période est 2 (indépendant de l'état par irréductibilité).

Dans le cas de la ruine du joueur, les états absorbants ont pour période 1, tandis que les autres états ont la période 2.

Remarque 2.32. Supposons le noyau de transition irréductible. Lorsque $d > 1$, on peut séparer l'espace d'états S en d sous-ensembles C_0, C_1, \dots, C_{d-1} , tels que si pour des indices i, j et des états $y_i \in C_i, y_j \in C_j$, alors le passage de y_i à y_j n'est possible qu'après une durée de la forme $(j - i) + ld$, où l est un entier : si $P^k(y_1, y_2) > 0$ alors il existe l entier tel que $k = ld + j - i$.

On peut définir C_i comme suit : $C_i = \{y \in C; \exists l \in \mathbb{N}, P^{ld+i}(x, y) > 0\}$.

Intuitivement, on visite successivement, et de façon cyclique de période d , chacun des sous-ensembles.

2.4 Calcul de la probabilité et de l'espérance du temps d'absorption

Soient P un noyau de transition sur l'espace d'états S et a un état absorbant. Etant donnée une condition initiale $x_0 \in S$, on veut calculer

$$\mathbb{P}(\{\exists n \in \mathbb{N}; X_n = a\} \mid X_0 = x_0),$$

$$\mathbb{E}[\inf \{\exists n \in \mathbb{N}; X_n = a\} \mid X_0 = x_0],$$

respectivement la probabilité d'absorption par l'état a , et l'espérance du temps d'absorption par cet état, partant de l'état initial x_0 .

La méthode est la suivante.

1. Introduire deux fonctions donnant les quantités définies précédemment, **quelle que soit la condition initiale** : pour tout $x \in S$

$$p(x) = \mathbb{P}(\{\exists n \in \mathbb{N}; X_n = a\} \mid X_0 = x) \in [0, 1],$$

$$t(x) = \mathbb{E}[\inf \{\exists n \in \mathbb{N}; X_n = a\} \mid X_0 = x] \in [0, +\infty].$$

2. Remarquer que $p(a) = 1$, $t(a) = 0$.
3. Trouver un système linéaire dont p (respectivement t) est solution, en fonction du noyau de transition P . Pour cela, **on décompose selon les valeurs possibles de X_1** : pour tout $x \neq a$,

$$\begin{aligned}
p(x) &= \mathbb{P}(\{\exists n \in \mathbb{N}; X_n = a\} \mid X_0 = x) \\
&= \sum_{y \in S} \mathbb{P}(\{\exists n \geq 1; X_n = a\} \cap \{X_1 = y\} \mid X_0 = x) \\
&= \sum_{y \in S} \mathbb{P}(\{\exists n \geq 1; X_n = a\} \mid X_1 = y) \mathbb{P}(X_1 = y \mid X_0 = x) \\
&= \sum_{y \in S} P(x, y) p(y).
\end{aligned}$$

(Remarquer que cette identité est vraie aussi pour $x = a$, car a est absorbant !) Matriciellement, $p = Pp$.

De même, pour tout $x \neq a$,

$$\begin{aligned}
t(x) &= \mathbb{E}[\inf \{\exists n \geq 1; X_n = a\} \mid X_0 = x] \\
&= \sum_{y \in S} \mathbb{E}[\inf \{\exists n \geq 1; X_n = a\} \cap \{X_1 = y\} \mid X_0 = x] \\
&= \sum_{y \in S} \mathbb{E}[\inf \{\exists n \geq 1; X_n = a\} \mid X_1 = y] \mathbb{P}(X_1 = y \mid X_0 = x) \\
&= \sum_{y \in S} P(x, y) (1 + \mathbb{E}[\inf \{\exists k \geq 0; X_k = a\} \mid X_0 = y]) \\
&= \sum_{y \in S} P(x, y) (1 + t(y)).
\end{aligned}$$

(Remarquer que cette fois c'est identité est fausse pour $x = a$: le membre de gauche est 0, celui de droite est 1 ; c'est logique, il n'y a pas de premier pas !)

Remarquer aussi que la valeur $t(x) = +\infty$ est possible : par exemple si $b \neq a$ est un autre état absorbant, on a $\mathbb{P}(X_n = b \mid X_0 = b) = 1$. L'équation précédente s'écrit alors $t(b) = t(b) + 1$, ce qui entraîne bien $t(b) = +\infty$; en "propageant" ce résultat, on voit que si x conduit à b on a aussi $t(x) = +\infty$.

Interprétation : à chaque étape, on regarde le temps qu'il faut en moyenne pour atteindre a à partir de la nouvelle position, et on doit ajouter 1 correspondant à la première étape.

4. Résoudre le système !

Pour bien comprendre comment ça fonctionne, il suffit de regarder le graphe de transition et d'effectuer le premier pas !

De plus : ne pas apprendre de formule par coeur, toujours refaire ce raisonnement. D'autant plus si l'espace d'états est de cardinal petit...

Considérons un exemple détaillé : la ruine du joueur.
à compléter !

Exercice 24. Reprendre la situation de l'exercice 12 : modéliser la dynamique à l'aide d'une chaîne de Markov, représenter le graphe des états. (Indication : on pourra considérer comme variable la différence entre le nombre de victoire respectif de chaque joueur, et obtenir un espace d'états de cardinal 5).

Déterminer la nature de chacun des états.

Calculer les probabilités et les temps moyens d'absorption.

2.5 Comportement en temps long

On suppose que P est un noyau de transition *irréductible* sur l'espace d'états S . Le premier résultat fondamental est le suivant.

Théorème 2.33. *Si le noyau est irréductible, apériodique et récurrent positif (on parle alors d'une chaîne ergodique), alors :*

- *il existe une unique loi invariante μ ;*
- *pour tous $x, y \in S$, quand $n \rightarrow +\infty$, on a*

$$P^n(x, y) \rightarrow \mu(y);$$

On a une partie de la réciproque.

Théorème 2.34. *Si un noyau irréductible admet une loi de probabilité invariante μ :*

- *elle est unique : si ν est une autre loi invariante, alors $\nu = \mu$;*
- *la chaîne est récurrente positive ;*
- *on a la formule suivante : pour tout $y \in S$*

$$\mu(y) = \frac{1}{\mathbb{E}[T(y)|X_0 = y]} \in (0, +\infty).$$

Le dernier point donne un moyen de calculer $\mathbb{E}[T(y)|X_0 = y]$ en fonction de la mesure invariante μ , calculée en résolvant un système d'équations linéaires.

L'apériodicité est une hypothèse supplémentaire dans le premier théorème.

L'irréductibilité seule assure en fait l'unicité à constante multiplicative près des mesures invariantes : si μ et ν sont deux mesures invariantes non triviales (i.e. $\forall x \in S$ on a $\mu(x) \in [0, +\infty)$), alors il existe $C \in (0, +\infty)$ telle que $\forall x \in S, \mu(x) = C\nu(x)$.

Dans le cas non irréductible, il peut y avoir non-unicité.

Exercice 25. Donner un tel exemple (par exemple si l'espace d'états se décompose en deux classes irréductibles récurrentes).

Une condition suffisante d'existence d'une mesure invariante non triviale pour un noyau irréductible est la récurrence. Dans le cas transient, selon les cas il y a existence ou non.

L'existence d'une loi invariante est alors équivalente à la possibilité de normaliser les mesures invariantes.

Par contre, l'impossibilité de normaliser ne permet pas de distinguer entre récurrence nulle et transience.

Exemple 2.35. La marche aléatoire dans \mathbb{Z}^d est irréductible, et admet toujours pour mesure invariante la mesure uniforme (non normalisable) telle que $\mu(x) = 1$ pour tout $x \in \mathbb{Z}^d$.

Néanmoins, la chaîne est récurrente si $d \leq 2$ et transiente si $d \geq 3$.

Le résultat de convergence est le suivant.

Proposition 2.36. *Si P est un noyau irréductible récurrent nul ou transient, on a pour tout $x, y \in S$, quand $n \rightarrow +\infty$:*

$$P^n(x, y) \rightarrow 0.$$

Finalement, on a un cas très important.

Théorème 2.37. *Si l'espace d'états S est fini, une chaîne irréductible est nécessairement récurrente positive.*

Remarque 2.38. Dans le cas non-irréductible, pour comprendre la dynamique (en temps long), on décompose de l'espace des états.

Partant d'un état dans une classe récurrente, donc fermée, on y reste, et on peut utiliser les résultats ci-dessus en se restreignant à une chaîne irréductible, sur cette classe.

Partant d'un état dans l'ensemble transient T , la situation est plus complexe, selon qu'on entre ou non dans une des classes récurrentes.

Pour terminer, on dispose également d'un théorème de type "loi des grands nombres".

Rappelons que si $(Z_n)_{n \in \mathbb{N}}$ est une suite de variables aléatoires réelles indépendantes et identiquement distribuées, telle que $\mathbb{E}[|Z_n|] < +\infty$, alors on a presque sûrement quand $n \rightarrow +\infty$

$$\frac{1}{N+1} \sum_{n=0}^N Z_n \rightarrow \mathbb{E}[Z_0].$$

Ce résultat fournit un *estimateur* de l'espérance, à l'aide de *moyennes empiriques* (par rapport à la variable de temps).

Dans le cadre des chaînes de Markov irréductibles récurrentes, on a le résultat suivant :

Proposition 2.39. *Notons ν une mesure invariante (non triviale), éventuellement non normalisable.*

Considérons deux fonctions bornées f et g , telles que

$$\begin{aligned} \int_S |f| d\nu &:= \sum_{x \in S} |f(x)| \nu(x) < +\infty, \\ \int_S |g| d\nu &:= \sum_{x \in S} |g(x)| \nu(x) < +\infty. \end{aligned}$$

On suppose également $\int_S g d\nu \neq 0$. Soit μ_0 une loi initiale, et $(X_n)_{n \in \mathbb{N}}$ une chaîne de Markov de noyau de transition P , de loi initiale μ_0 . On a alors la convergence presque sûre suivante, quand $N \rightarrow +\infty$

$$\frac{\sum_{n=0}^N f(X_n)}{\sum_{n=0}^N g(X_n)} \rightarrow \frac{\int_S f d\nu}{\int_S g d\nu}.$$

Si le noyau est récurrent positif, d'unique loi invariante μ , on a

$$\frac{1}{N+1} \sum_{n=0}^N f(X_n) \rightarrow \int_S f d\mu.$$

Noter que le membre de droite ne dépend pas du fait de savoir si ν est normalisable ou non, et qu'on divise numérateur et dénominateur par une même constante non nulle sans changer le quotient.

Le cas récurrent positif s'obtient en considérant $g(x) = 1$ pour tout $x \in S$, et $\nu = \mu$.

Chapitre 3

Processus de Poisson et files d'attentes

3.1 Loi exponentielle de loi de Poisson

3.1.1 La loi exponentielle

Elle a été définie au chapitre 1. Une variable aléatoire réelle X suit la *loi exponentielle de paramètre* $\lambda > 0$ si pour tout $t > 0$, on a

$$\mathbb{P}(X > t) = \exp(-\lambda t).$$

On note alors $X \sim \mathcal{E}(\lambda)$. Notons qu'on a $\mathbb{P}(X \leq 0) = 0$: X ne prend que des valeurs positives.

Exercice 26. Démontrer que $\mathbb{E}[X] = \frac{1}{\lambda}$ et $\text{Var}(X) = \frac{1}{\lambda^2}$. De plus, montrer que la fonction caractéristique de la loi $\mathcal{E}(\lambda)$ est donnée par $\varphi(t) = \frac{1}{\lambda - it}$.

La propriété fondamentale est celle d'*absence de mémoire* : pour tout $t, s > 0$,

$$\mathbb{P}(X > t + s | X > s) = \mathbb{P}(X > t).$$

Dans la suite, on va utiliser des variables aléatoires de loi exponentielle pour modéliser le temps écoulé entre deux occurrences d'un événement.

Exercice 27. Soient T_1 et T_2 des v.a. indépendantes de loi exponentielles de paramètres λ_1 et λ_2 respectivement. Quelle est la loi de $T = \inf\{T_1, T_2\}$ (c'est-à-dire le temps du premier événement qui se produit) ?

3.1.2 La loi de Poisson

La loi de Poisson de paramètre $\lambda > 0$ a aussi été introduite au chapitre

1. Une variable aléatoire discrète X suit cette loi si on a pour tout $n \in \mathbb{N}$

$$\mathbb{P}(X = n) = \frac{\lambda^n}{n!} \exp(-\lambda).$$

On note alors $X \sim \mathcal{P}(\lambda)$. Rappelons que pour tout $x \in \mathbb{R}$, on a le développement de $\exp(x)$ en série : $\exp(x) = \sum_{k=0}^{+\infty} \frac{x^k}{k!}$. Donc on a bien défini une loi de probabilité sur \mathbb{N} .

L'espérance et la variance de cette loi sont toutes deux égales à λ .

De plus, la fonction caractéristique vérifie $\varphi(t) = \exp(\lambda(\exp(it) - 1))$ pour tout $t \in \mathbb{R}$.

Exercice 28. Vérifier ces résultats sur l'espérance, la variance, la fonction caractéristique.

Voici un résultat fameux, en exercice (facultatif, plus probabilités "élémentaires" que processus stochastiques) :

Exercice 29. Soit $\lambda > 0$ un paramètre fixé.

Pour tout $n \in \mathbb{N}$, donnons X_n une variable aléatoire suivant une loi binomiale $\mathcal{B}(n, \frac{\lambda}{n})$: pour tout $0 \leq k \leq n$, on a

$$\mathbb{P}(X_n = k) = \binom{n}{k} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k}.$$

Montrer, pour tout $k \in \mathbb{N}$ fixé, la convergence suivante lorsque $n \rightarrow +\infty$:

$$\mathbb{P}(X_n = k) \rightarrow \mathbb{P}(X = k),$$

où X suit une loi de Poisson de paramètre λ .

L'interprétation du résultat précédent se fait en termes d'*événements rares* : on compte le nombre d'occurrences d'événements indépendants, se réalisant chacun avec une probabilité $\frac{\lambda}{n}$, parmi un nombre de tirages égal à n , de telle sorte que *en moyenne* le nombre de réalisations soit constant égal à λ . Lorsque n tend vers $+\infty$, ce nombre de réalisations suit donc une loi de Poisson ; le paramètre λ est interprété comme une *intensité* : il détermine avec quelle proportion en moyenne l'événement doit se produire.

Au niveau des applications, on peut ainsi considérer qu'on subdivise un intervalle de temps fixé, en n sous-intervalles de taille identique. Sur chaque

sous-intervalle, de façon indépendante, peut ou ne pas se produire un événement (arrivée d'un client, panne d'un appareil...), avec une probabilité proportionnelle à la longueur des sous-intervalles.

Plus généralement, on peut aussi découper un rectangle, un disque (par exemple une carte), et considérer comme événement un incendie, une panne...

Cette interprétation est à la base de la considération du processus de Poisson, où on ajoutera une composante dynamique pour faire évoluer le comptage au cours du temps.

3.2 Indépendance et stationarité des accroissements

On introduit dans cette section des notions importantes qui sont valables pour n'importe quelle sorte de processus stochastique. Elles seront utilisées dans ce chapitre et le suivant. On se replace donc dans le cadre du chapitre 1 avec un processus $X = (X_t)_{t \in \mathbb{T}}$, pour un ensemble de temps \mathbb{T} général, et avec un espace d'états S dénombrable ou non.

3.2.1 Indépendance des accroissements

Définition 3.1. Soit $X = (X_t)_{t \in \mathbb{T}}$ un processus. On dit qu'il est à accroissements indépendants (noté P.A.I.) si pour tout $n \in \mathbb{N}$, tout n -uplet ordonné $t_1 \leq \dots \leq t_n$, les variables aléatoires $X_{t_1} - X_0, X_{t_2} - X_{t_1}, \dots, X_{t_n} - X_{t_{n-1}}$ sont indépendantes.

Exemple 3.2. La marche aléatoire simple est un P.A.I. à temps discret.

3.2.2 Stationnarité des accroissements

Définition 3.3. Soit $X = (X_t)_{t \in \mathbb{T}}$ un processus. On dit qu'il est à accroissements stationnaires (noté P.A.S.) si pour tous instants $s \leq t$ et $s' \leq t'$ dans \mathbb{T} tels que $t' - s' = t - s$, les variables aléatoires $X_t - X_s$ et $X_{t'} - X_{s'}$ ont la même loi.

Autrement dit, la loi de l'accroissement $X_t - X_s$ ne dépend que de l'écart de temps $t - s$, pas des temps eux-même. C'est en particulier la loi de l'accroissement $X_{t-s} - X_0$.

Exemple 3.4. La marche aléatoire simple est un P.A.S. à temps discret.

Définition 3.5. Un processus qui est à la fois un P.A.I. et un P.A.S. est appelé processus à accroissements indépendants et stationnaires (noté P.A.I.S.).

Dans le cas de la marche aléatoire simple notée $(S_n)_{n \in \mathbb{N}}$, la propriété d'être un P.A.I.S. entraîne l'égalité suivante :

$$\begin{aligned}\mathbb{E}[S_{n+1}|S_n] &= \mathbb{E}[S_{n+1} - S_n|S_n] + S_n \\ &= \mathbb{E}[S_{n+1} - S_n] + S_n \\ &= \mathbb{E}[S_1] + S_n = S_n.\end{aligned}$$

Un processus vérifiant une telle propriété est appelé une *martingale*. Ce sont des processus très importants dans la théorie, ils généralisent la notion de processus à accroissement indépendants. Par manque de temps, on n'a pas pu traiter la théorie en temps discret (où les temps d'arrêt, les filtrations... ont encore un grand rôle!), et aussi parce qu'on n'a pas développé la notion d'espérance conditionnelle dans le cas général.

3.3 Processus de comptage et processus de Poisson

3.3.1 Processus de comptage

Avant de spécifier le cas du processus de Poisson, on énonce quelques généralités sur la façon de modéliser le comptage d'événements. Le processus est indexé par $\mathbb{T} = [0, +\infty[$, et est à valeurs dans \mathbb{N} .

On se donne la suite des instants où se produit un événement (dans la suite, on parlera d'instant d'arrivées, par référence au cas d'une file d'attente) : il s'agit d'une suite ordonnée de variables aléatoires réelles $T_1 \leq T_2 \leq \dots \leq T_n \dots$, sur laquelle on effectue les hypothèses suivantes.

Hypothèse 3.1. Avec probabilité 1,

1. l'ordre est strict : $0 = T_0 < T_1 < \dots < T_n < \dots$;
2. on a $T_n \rightarrow +\infty$ quand $n \rightarrow +\infty$.

Cela signifie que deux (ou plus) événements ne peuvent pas se produire en même temps, et qu'en un temps fixé T il ne peut se produire qu'un nombre fini de ces événements. On notera $T_0 = 0$ l'instant initial.

On peut compter les événements produits sur l'intervalle de temps $[0, t]$, pour tout $t > 0$, en introduisant la quantité suivante :

$$N_t = \sum_{n=1}^{+\infty} \mathbb{1}_{T_n \leq t} = \sup \{n \in \mathbb{N}; T_n \leq t\}.$$

Remarquer que $N_0 = 0$.

Le processus à temps continu $N = (N_t)_{t \in [0, +\infty[}$ est à valeurs dans \mathbb{N} . Ses trajectoires sont en escalier, continues à droite, avec des sauts de hauteur 1 aux instants T_n (faire un dessin). La connaissance du processus N ou des temps d'arrivées $(T_n)_{n \in \mathbb{N}}$ sont équivalentes. On vient de voir comment N dépend des T_n ; graphiquement, les T_n sont les instants de saut des trajectoires.

On peut aussi noter les égalités d'événements suivantes :

$$\begin{aligned}\{N_t = n\} &= \{T_n \leq t < T_{n+1}\} ; \\ \{T_n \leq t\} &= \{N_t \geq n\} .\end{aligned}$$

En rajoutant des hypothèses sur les temps d'arrivées, ou de façon équivalente sur le processus, on va pouvoir caractériser la loi du processus.

3.3.2 Processus de Poisson

On introduit tout d'abord une hypothèse sur les temps d'arrivées.

Hypothèse 3.2. *Le processus des temps d'arrivées $(T_n)_{n \in \mathbb{N}}$ est un P.A.I.S., tel que $T_1, T_2 - T_1, \dots$ sont (iid) de loi exponentielle de paramètre $\lambda > 0$.*

Interprétation : les sources respectives des événements successifs sont indépendantes, et suivent toute la même loi exponentielle.

En parallèle, on peut introduire l'hypothèse suivante sur le processus de comptage :

Hypothèse 3.3. *Le processus de comptage $(N_t)_{t \in \mathbb{R}^+}$ est un P.A.I.S. .*

On a alors le résultat suivant (admis) :

Théorème 3.6. *On suppose l'hypothèse 3.1 vérifiée.*

Si l'hypothèse 3.2 est satisfaite, alors le processus de comptage associé est un P.A.I.S.

Réciproquement, si l'hypothèse 3.3 est satisfaite, alors il existe $\lambda > 0$ tel que l'hypothèse 3.2 est vérifiée.

Le cas échéant, pour tout $t \in \mathbb{R}^+$, la variable aléatoire N_t suit une loi de Poisson de paramètre λt . On dit que N est un processus de Poisson d'intensité λ .

Exercice 30. Montrer que la somme de deux v.a. indépendantes suivant des lois de Poisson $\mathcal{P}(\lambda), \mathcal{P}(\mu)$, suit une loi de Poisson dont on précisera le paramètre.

Remarque 3.7. On a deux moyens de simuler un processus de Poisson :

- par la donnée de la suite des temps d'arrivées (tirages indépendants selon la loi $\mathcal{E}(\lambda)$ des temps entre deux arrivées successives) ;
- si on a fixé un temps final t : on commence par tirer aléatoirement un nombre $N_t = k$ selon la loi de Poisson $\mathcal{P}(\lambda t)$; puis on tire des nombres u_1, \dots, u_k , indépendants, uniformément sur l'intervalle $[0, t]$ (ils sont presque sûrement tous distincts) ; on les trie en ordre croissant, ce qui donne les temps d'arrivée successifs.

Le deuxième moyen repose sur une démonstration (admise) que ce procédé fonctionne...

Remarque 3.8. Tirer un nombre aléatoirement selon une loi donnée = générer (par un certain mécanisme) une réalisation d'un variable aléatoire, de la loi en question.

Exercice 31. Soit $N = (N_t)_{t \geq 0}$ un processus de Poisson, et soit $T > 0$ un instant fixé.

Montrer que le processus $N^{(T)}$ tel que $N_t^{(T)} = N_{T+t} - N_T$, est un autre processus de Poisson, indépendant du processus $(N_s)_{0 \leq s \leq T}$ - c'est-à-dire de tout vecteur aléatoire $(N_{s_1}, \dots, N_{s_n})$ avec $0 \leq s_1 \leq \dots \leq s_n \leq T$.

Estimation de l'intensité : $\frac{N_t}{t} \rightarrow \lambda$, quand $t \rightarrow +\infty$, presque sûrement.

3.4 Processus de Poisson généralisés

3.4.1 Processus de Poisson marqué

On considère toujours une suite de variables aléatoires exponentielles indépendantes, modélisant le temps entre deux instants d'occurrence d'un événement (tel l'arrivée d'un client dans une file d'attente).

Maintenant, on va diviser en deux (il est possible de faire plus mais cela suffira à expliquer comment ça marche) les possibilités. Par exemple, les clients sont soit des hommes, soit des femmes. Et on ne va compter qu'un seul type de possibilité.

Autrement dit, à un instant T_n , on regarde la réalisation d'une variable aléatoire X_n , indépendante du reste, prenant la valeur 1 (le type qu'on compte) avec une probabilité p , et la valeur 0 (celui qu'on élimine) avec la probabilité $q = 1 - p$. Et on ajoute la valeur de X_n au compteur.

Mathématiquement, on se donne donc des variables $0 = T_0 \leq T_1 \leq \dots \leq T_n$ vérifiant l'hypothèse 3.2, tel que les variables $(T_{i+1} - T_i)_{i \in \mathbb{N}}$ sont iid de loi exponentielle de paramètre λ . On se donne aussi une suite $(X_n)_{n \in \mathbb{N}^*}$ de va iid, de loi de Bernoulli de paramètre p . Les suites $(T_n)_{n \in \mathbb{N}^*}$ et $(X_n)_{n \in \mathbb{N}^*}$ sont également supposées indépendantes.

On a ainsi 3 processus, définis par les formules :

- $N_t = \sum_{n=1}^{+\infty} \mathbb{1}_{T_n \leq t}$ quand on compte tous les types ;
- $N_t^1 = \sum_{n=1}^{+\infty} X_n \mathbb{1}_{T_n \leq t}$ quand on compte uniquement le type 1 ;
- $N_t^0 = \sum_{n=1}^{+\infty} (1 - X_n) \mathbb{1}_{T_n \leq t}$ quand on compte uniquement le type 0.

Proposition 3.9. *On a $N_t = N_t^0 + N_t^1$. De plus, les processus N^0 et N^1 sont deux processus de Poisson indépendants, d'intensités respectives $(1 - p)\lambda$ et $p\lambda$.*

3.4.2 Processus de Poisson composé

Dans la sous-section précédente, par rapport au Processus de Poisson (simple), on s'est donné la possibilité de sélectionner ou non chaque événement dans notre compteur. Pour cela, on a introduit une suite de va iid $(X_n)_{n \in \mathbb{N}^*}$ dont les valeurs déterminent la prise en compte ou non de l'événement. On observe qu'on obtient encore un Processus de Poisson simple d'intensité différente ; en particulier, les sauts sont tous de taille 1.

Maintenant, on veut que les sauts puissent prendre d'autres valeurs, aléatoires. Il suffit pour cela de considérer une suite de va iid $(Y_n)_{n \in \mathbb{N}^*}$, de loi quelconque (par exemple gaussienne), et de poser

$$N_t^Y = \sum_{n=1}^{+\infty} Y_n \mathbb{1}_{T_n \leq t}.$$

Il est important de noter que N_t ne prend plus uniquement des valeurs entières.

Les Processus de Poisson simple et marqué sont deux cas particuliers de processus de Poisson composé.

3.5 Processus Markovien de saut

Dans les deux généralisations précédents du Processus de Poisson simple, on a modifié la loi des sauts. Mais on n'a pas modifié la loi des temps de saut : la suite des durées entre deux sauts successifs est constituée de va indépendantes, de loi exponentielle, de même paramètre. On va maintenant modifier une condition (et pas les autres) : celle que le paramètre est le même pour toutes les va exponentielles.

On va aussi supposer que l'espace d'états du processus $(X_t)_{t \in \mathbb{R}^+}$ ainsi obtenu est fini ou dénombrable, il sera noté S . Plutôt que de parler des sauts de ces processus, il est commode de garder un point de vue "position", qui change (sans possibilité de rester sur place !) aux "instants de saut".

Le fait de conserver l'indépendance entre les sauts et les temps de saut, ainsi que l'hypothèse d'accroissements stationnaires et de loi exponentielle sur ces temps de saut, est lié à la propriété de Markov (en temps continu) : on veut que le futur du processus ne dépende du passé qu'à travers l'état présent.

On ne va pas détailler précisément comment exprimer cette condition sur nos processus, on va se contenter de décrire la dynamique (de deux façons équivalentes). L'introduction de ces objets est motivée par l'étude des modèles de files d'attente dans la section suivante.

3.5.1 Description chaîne de Markov induite

La suite des états successifs du processus à temps continu est en fait donnée par une chaîne de Markov (à temps discret) d'espace d'états S .

On se donne un noyau de transition R sur S , qui a la propriété assurant qu'il n'est pas possible de rester sur un état non absorbant :

$$R(x, x) = 0 \text{ ou } 1 \text{ pour tout } x \in S.$$

La durée passée sur chacun des états est donnée par une suite de variables aléatoires indépendantes, de loi exponentielle dépendant de la position x : on note $q(x) \in [0, +\infty)$ ce paramètre. La valeur $+\infty$ signifierait qu'en fait on ne reste jamais sur la position x , et qu'on peut donc l'enlever de l'espace d'états. Par contre la valeur $q(x) = 0$ est autorisée : elle signifie qu'une fois que cette position x est atteinte, alors on ne bouge plus, l'état x est absorbant.

On a la condition $R(x, x) = 1$ si et seulement si $q(x) = 0$.

La dynamique est la suivante : supposons qu'on se situe à l'instant t à la position x ; on se donne alors une variable aléatoire T_x , exponentielle de paramètre $q(x)$; alors sur l'intervalle de temps $[t, t + T_x[$, X vaut x . Ensuite un saut se produit en $t + T_x$, où X prend la valeur $y \in S$, selon la loi de probabilité se trouvant à la ligne x de Q (comme dans une chaîne de Markov à temps discret).

3.5.2 Description "générateur"

On se donne une seule matrice indicée par $S \times S$, notée Q , qui vérifie les propriétés suivantes : pour tout $x \in S$,

$$\begin{aligned} Q(x, y) &\geq 0 \text{ pour tout } y \in S, y \neq x; \\ Q(x, x) &= - \sum_{y \in S, y \neq x} Q(x, y) \leq 0. \end{aligned}$$

On remarque que $Q(x, x) = 0$ si et seulement si $Q(x, y) = 0$ pour tout $y \neq x$.

La dynamique est la suivante : supposons qu'on se situe à l'instant t à la position x . On se donne alors une famille $(T^y)_{y \in S, y \neq x}$ de va indépendantes, de telle sorte que T^y est de loi exponentielle de paramètre $Q(x, y)$. On pose $T_x = \inf_{y \in S, y \neq x} T^y$; il s'agit d'une va exponentielle, de paramètre $\sum_{y \in S, y \neq x} Q(x, y) = -Q(x, x)$. Alors sur l'intervalle de temps $[t, t + T_x[$, X vaut x . Ensuite un saut se produit en $t + T_x$, où X prend la valeur $y \in S$, avec probabilité $\frac{Q(x, y)}{-Q(x, x)}$.

3.5.3 Lien entre les deux descriptions

Les deux descriptions précédentes sont équivalentes, avec les relations :

$$\begin{aligned} q(x) &= -Q(x, x), \\ R(x, y) &= \frac{Q(x, y)}{q(x)} \text{ pour tout } y \in S, y \neq x, \text{ si } q(x) \neq 0, \\ R(x, y) &= 0 \text{ pour tout } y \in S, y \neq x, \text{ si } q(x) = 0. \end{aligned}$$

La suite des positions successives $(Z_n)_{n \in \mathbb{N}} = (X_{T_n})_{n \in \mathbb{N}}$ est une chaîne de Markov de noyau de transition Q , qui est de plus indépendante de la suite des temps de saut $(T_n)_{n \in \mathbb{N}}$.

Remarque 3.10. La valeur $T_n = +\infty$ est possible ; il faut alors considérer que $T_m = +\infty$ pour tout $m \geq n$.

3.5.4 Lois invariantes

Avec les notations précédentes, on dit que le processus $(X_t)_{t \in \mathbb{R}^+}$ est irréductible si la chaîne induite $(Z_n)_{n \in \mathbb{N}}$ de noyau de transition R l'est et si pour tout $x \in S$ on a $q(x) > 0$.

On dispose également de la notion de loi invariante μ , telle qu'on a stationnarité : si la position initiale X_0 suit cette loi μ , alors pour tout $t > 0$ c'est aussi le cas de X_t . En cas d'existence, cela fournit un moyen de voir que le processus peut être dans un état stable. S'il n'y a pas de loi invariante, cela montre qu'il peut y avoir explosion.

Selon la description adoptée pour le processus, on dispose d'une équation différentielle sur μ .

- Première description : en définissant $\pi(x) = q(x)\mu(x)$, π est solution de $\pi R = \pi$ (c'est-à-dire π est une mesure invariante pour la chaîne induite).
- Deuxième description : μ est solution de $\mu Q = 0$.

3.6 Exemples de modèles de files d'attente

Une file d'attente est constituée de clients, d'une salle d'attente, de guichets et de serveurs. On va supposer que les temps interarrivées sont indépendants et de même type de loi (dépendant d'un ou plusieurs paramètres), et indépendants des temps de services, également indépendants et de même type de loi. On suppose que dès qu'un guichet se libère il est utilisé à nouveau s'il y avait un client dans la salle d'attente.

On utilise la notation suivante. Loi d'interarrivée / Loi de service / Nombre de serveurs / Longueur maximale. Le dernier paramètre est facultatif, par défaut il vaut ∞ , la salle d'attente a une capacité non limitée. Les deux premiers paramètres peuvent en toute généralité prendre deux valeurs : M (pour Markov) ou G (pour général). On se contente ici du cadre Markovien (donc de la valeur M) : cela signifie que les lois d'interarrivées et de services sont exponentielles (de paramètre dépendant de la nature exacte de la file).

Par conséquent on peut utiliser les outils de la section précédente sur les processus de Markov de saut. Il suffit de donner les matrices correspondant à l'une ou l'autre des descriptions pour caractériser la dynamique ; c'est ce qu'on va faire pour plusieurs exemples. Attention : le processus étudié compte le nombre de clients dans le système salle d'attente + guichets.

3.6.1 $M/M/0$

Il n'y a pas de serveur, la salle d'attente a une capacité illimitée. Les clients arrivent selon des va exponentielles de paramètre λ .

En fait, il s'agit d'un processus de Poisson d'intensité λ .

Exercice 32. Vérifier que pour tout $x \in \mathbb{N}$

$$\begin{aligned} q(x) &= \lambda, \\ R(x, x+1) &= 1. \end{aligned}$$

Donner les valeurs manquantes.

3.6.2 $M/M/1$

Loi des temps interarrivées : exponentielle de paramètre λ .

Un guichet, un serveur : loi de temps de service exponentielle de paramètre μ .

Un saut peut prendre la valeur 1 (arrivée d'un nouveau client dans la salle d'attente, ou directement au guichet s'il n'y avait personne), ou -1 (départ d'un client après avoir été servi).

On a

$$\begin{aligned} Q(x, x+1) &= \lambda \text{ pour tout } x \in \mathbb{N}, \\ Q(x, x-1) &= \mu \text{ pour tout } x \in \mathbb{N}^*, \\ Q(x, y) &= 0 \text{ si } |x-y| \geq 2. \end{aligned}$$

Noter que c'est un processus de naissance et de mort.

Exercice 33. Déterminer q et R .

La chaîne induite est irréductible.

Il y a récurrence positive si et seulement si $\rho = \frac{\lambda}{\mu} < 1$ (la moyenne du temps entre deux arrivées est supérieure à la moyenne du temps d'un service). La loi invariante est donnée par $\mu(n) = \rho^n(1-\rho)$ (loi de type géométrique); on peut prouver qu'à l'équilibre la taille moyenne de la file est $\frac{\rho}{1-\rho}$.

3.6.3 $M/M/s$

$$\begin{aligned} Q(n, n+1) &= \lambda \text{ pour tout } n \in \mathbb{N}, \\ Q(n, n-1) &= \min(s, n)\mu \text{ pour tout } n \in \mathbb{N}^*, \\ Q(x, y) &= 0 \text{ si } |x-y| \geq 2. \end{aligned}$$

Exercice 34. Donner q et Q .

Pour la deuxième ligne : il y a sortie d'un client si un service se termine. Il faut regarder combien de guichets étaient occupés : n le nombre total de clients si $n \leq s$, et s si $n > s$ (il y a alors $n-s$ clients dans la salle d'attente).

Les temps de service à chaque guichet sont indépendants et de loi exponentielle de paramètre μ . On utilise le résultat suivant : le minimum de k va iid de loi $\mathcal{E}(\mu)$ est de loi $\mathcal{E}(k\mu)$ - pour le prouver, considérer la fonction de répartition F_{\min} , ou plutôt $1 - F_{\min}$.

Ici il y a récurrence positive si et seulement si $\rho := \frac{\lambda}{\mu} < s$.

La loi invariante est alors donnée par

$$\begin{aligned} \mu(n) &= \mu_0 \frac{\rho^n}{n!} \text{ si } n \leq s; \\ \mu(n) &= \mu_0 \frac{\rho^s}{s!} \left(\frac{\rho}{s}\right)^{n-s} \text{ si } n \geq s. \end{aligned}$$

On peut en déduire des quantités à l'équilibre, comme la probabilité que tous les serveurs soient occupés, la moyenne du nombre de serveurs occupés, la taille moyenne du système, le temps d'attente moyen avant d'être servi... et utiliser ces résultats pour optimiser les caractéristiques du système.

3.6.4 $M/M/\infty$

Une infinité de serveurs.

$$\begin{aligned} Q(n, n+1) &= \lambda \text{ pour tout } n \in \mathbb{N}, \\ Q(n, n-1) &= n\mu \text{ pour tout } n \in \mathbb{N}^*, \\ Q(x, y) &= 0 \text{ si } |x - y| \geq 2. \end{aligned}$$

Récurrent positif, probabilité invariante : loi de Poisson de paramètre $\rho := \frac{\lambda}{\mu}$.

3.6.5 $M/M/1/k$

Capacité finie : espace d'états $\{0, \dots, k+1\}$.

$$\begin{aligned} Q(n, n+1) &= \lambda \text{ pour tout } 0 \leq n \leq k, \\ Q(n, n-1) &= \mu \text{ pour tout } 1 \leq n \leq k+1, \\ Q(x, y) &= 0 \text{ si } |x - y| \geq 2. \end{aligned}$$

3.6.6 $M/M/s/0$

Si tous les guichets sont occupés, tout nouveau client est rejeté (pas de salle d'attente). Espace d'états $\{0, \dots, s\}$.

$$\begin{aligned} Q(n, n+1) &= \lambda \text{ pour tout } 0 \leq n \leq s-1, \\ Q(n, n-1) &= n\mu \text{ pour tout } 1 \leq n \leq s-1, \\ Q(x, y) &= 0 \text{ si } |x - y| \geq 2. \end{aligned}$$

Chapitre 4

Processus Gaussiens et mouvement Brownien

4.1 Loi normale et variables aléatoires Gaussiennes réelles

On parle indifféremment de loi normale, de loi gaussienne, de loi de Laplace-Gauss... C'est une loi fondamentale en probabilités (et statistiques), liée au fameux théorème central de la limite (TCL).

4.1.1 Loi standard

Définition 4.1. Une variable aléatoire réelle X suit la loi normale standard, notée $\mathcal{N}(0, 1)$, si elle a pour densité de probabilité la fonction suivante : pour tout $x \in \mathbb{R}$:

$$\phi(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right).$$

\Rightarrow Revoir la notion de densité de probabilité et de fonction de répartition. La fonction de répartition est notée Φ . On n'a pas d'expression de cette fonction en termes de *fonctions usuelles* : on a seulement pour tout $x \in \mathbb{R}$:

$$\Phi(x) = \int_{-\infty}^x \phi(t) dt = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x \exp\left(-\frac{t^2}{2}\right) dt.$$

Dans certains logiciels de calcul, cette fonction Φ fait partie de la bibliothèque standard, sous le nom *erf* ("fonction d'erreur").

Voici quelques propriétés élémentaires mais fondamentales.

Théorème 4.2. (i) La fonction ϕ est bien une densité de probabilité : $\phi \geq 0$ et $\int_{-\infty}^{+\infty} \phi(t)dt = 1$;
(ii) si X suit la loi $\mathcal{N}(0, 1)$, alors $\mathbb{E}[X] = 0$ et $\text{Var}(X) = 1$;
(iii) la fonction caractéristique de la loi $\mathcal{N}(0, 1)$ est donnée, pour tout $t \in \mathbb{R}$, par

$$\varphi(t) := \mathbb{E}[e^{itX}] = \exp\left(-\frac{t^2}{2}\right).$$

Avec le point (ii), on parle aussi de *loi normale centrée réduite*.

Pour montrer qu'une variable aléatoire suit une loi Gaussienne, on dispose donc de deux outils : calculer la densité, ou bien calculer sa fonction caractéristique : en effet Y suit la loi $\mathcal{N}(0, 1)$ si et seulement si pour tout $t \in \mathbb{R}$ on a $\mathbb{E}[\exp(itY)] = \varphi(t)$.

On finit cette section par quelques résultats.

Les moments de la loi normale $\mathcal{N}(0, 1)$ vérifient les identités suivantes : si $X \sim \mathcal{N}(0, 1)$, pour tout $n \in \mathbb{N}$,

$$\begin{aligned}\mathbb{E}[X^{2n}] &= \frac{(2n)!}{2^n n!}; \\ \mathbb{E}[X^{2n+1}] &= 0.\end{aligned}$$

Enfin, on dispose d'un moyen très simple d'estimer la *queue de la distribution*, c'est-à-dire $\mathbb{P}(X \geq a)$, pour $a > 0$:

$$\begin{aligned}\mathbb{P}(X \geq a) &= \frac{1}{\sqrt{2\pi}} \int_a^{+\infty} \exp\left(-\frac{x^2}{2}\right) dx \\ &\leq \frac{1}{\sqrt{2\pi}} \int_a^{+\infty} \exp\left(-\frac{xa}{2}\right) dx \\ &\leq \frac{2}{a\sqrt{2\pi}} \exp\left(-\frac{a^2}{2}\right).\end{aligned}$$

Des estimations plus fines sont possibles, donnant un équivalent de $\mathbb{P}(X \geq a)$, ou des développements asymptotiques plus précis ; en pratique, le calcul ci-dessus est plus simple, et peut (doit) se refaire directement si besoin.

Exercice 35. En utilisant une intégration par parties, montrer une relation de récurrence entre les moments d'ordre pair de la loi normale, puis obtenir la formule annoncée.

Pourquoi les moments d'ordre impair sont-ils nuls ?

Justifier les étapes du calcul précédent.

4.1.2 Loi normale, cas général

Définition 4.3. Soit $m \in \mathbb{R}$ et $\sigma \geq 0$. Une variable aléatoire X suit une loi normale $\mathcal{N}(m, \sigma^2)$ si elle s'écrit $X = m + \sigma Y$, où Y suit une loi normale centrée réduite $\mathcal{N}(0, 1)$. Lorsque $\sigma = 0$, X suit la loi $\mathcal{N}(m, \sigma^2)$ si $X = m$ (variable aléatoire constante, toujours égale à m).

Dans le cas $\sigma = 0$, on parle de loi normale dégénérée.

On a le résultat suivant :

Théorème 4.4. Supposons $\sigma > 0$, et soit X une variable aléatoire de loi $\mathcal{N}(m, \sigma^2)$.

(i) On a $\mathbb{E}[X] = m$ et $\text{Var}(X) = \sigma^2$;

(ii) X a pour densité de probabilité la fonction ϕ_{m, σ^2} , telle que pour tout $x \in \mathbb{R}$

$$\phi_{m, \sigma^2}(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right);$$

(iii) la fonction caractéristique de la loi $\mathcal{N}(m, \sigma^2)$ est donnée, pour tout $t \in \mathbb{R}$, par

$$\varphi_{m, \sigma^2}(t) := \mathbb{E}[\exp itX] = \exp(itm - \sigma^2 \frac{t^2}{2}).$$

Exercice 36. Soit $(X_n)_{n \geq 1}$ une suite de variables aléatoires iid, de loi $\mathcal{N}(0, 1)$.

Montrer que pour tout $n \geq 1$, $\frac{X_1 + \dots + X_n}{n}$ suit une loi normale. (Indication : utiliser le critère "fonction caractéristique")

Quelle est son espérance, sa variance en fonction de n ?

Montrer que pour tout $\epsilon > 0$, $\mathbb{P}(|\frac{X_1 + \dots + X_n}{n}| \geq \epsilon) \rightarrow 0$ quand $n \rightarrow +\infty$.

4.1.3 Le théorème central de la limite

On rappelle ici l'importance de la loi normale, comme étant la loi limite apparaissant lorsqu'on examine les fluctuations de la moyenne empirique d'une suite de va iid autour de l'espérance.

Théorème 4.5 (TCL). Soit $(X_n)_{n \in \mathbb{N}}$ une suite de va réelles iid, de carré intégrable. On note $m = \mathbb{E}[X]$ et $\sigma^2 = \text{Var}(X)$, où $X = X_i$ pour i quelconque. On définit pour tout $n \in \mathbb{N}^*$ la variable aléatoire suivante :

$$Z_n = \sqrt{n} \left(\frac{X_1 + \dots + X_n}{n} - m \right).$$

Alors la suite $(Z_n)_{n \in \mathbb{N}^*}$ converge en loi vers la loi normale $\mathcal{N}(0, \sigma^2)$, au sens où si Z suit cette loi,

$$\mathbb{P}(Z_n \leq x) \rightarrow \mathbb{P}(Z \leq x) \quad \text{quand } n \rightarrow +\infty, \quad \text{pour tout } x \in \mathbb{R};$$

de façon équivalente la suite des fonctions caractéristiques $(\varphi_{Z_n})_{n \in \mathbb{N}^*}$ converge vers φ_Z : pour tout $t \in \mathbb{R}$,

$$\varphi_{Z_n}(t) = \mathbb{E}[\exp(itZ_n)] \rightarrow \mathbb{E}[\exp(itZ)] = \varphi_Z(t) = \exp(-\sigma^2 t^2/2), \quad \text{quand } n \rightarrow +\infty.$$

Rappelons que ce résultat est la base de la construction d'*intervalles de confiance asymptotiques* : par exemple,

$$\lim_{n \rightarrow +\infty} \mathbb{P} \left(\frac{X_1 + \dots + X_n}{n} \in \left[m - \frac{1.96\sigma}{\sqrt{n}}; m + \frac{1.96\sigma}{\sqrt{n}} \right] \right) \geq 0.95.$$

4.2 Vecteurs Gaussiens

Soit un entier $n \in \mathbb{N}$. On va maintenant considérer des vecteurs aléatoires $X = (X_1, \dots, X_n)$, à valeurs dans \mathbb{R}^n , chaque X_i étant une variable aléatoire réelle.

Définition 4.6. *Un vecteur aléatoire $X = (X_1, \dots, X_n)$ est un vecteur gaussien si toute combinaison linéaire des variables X_i suit une loi normale : pour tout n -uplet de nombres réels $(\lambda_1, \dots, \lambda_n) \in \mathbb{R}^n$, la variable aléatoire $\lambda_1 X_1 + \dots + \lambda_n X_n$ suit une loi normale.*

En choisissant $\lambda_1 = \dots = \lambda_n = 0$, on a une loi normale dégénérée, c'est pour cette raison qu'on a inclus ce cas dans la définition de la section précédente. En choisissant $\lambda_j = 1$, et $\lambda_i = 0$ si $i \neq j$, on voit que si X est un vecteur gaussien alors chacune de ses composantes X_j suit une loi normale. Mais cette condition nécessaire n'est pas suffisante !

Exercice 37. Soit X une variable aléatoire de loi $\mathcal{N}(0, 1)$, et soit $a > 0$ un réel strictement positif. Posons Y la variable aléatoire suivante :

$$Y = \begin{cases} X & \text{si } |X| \leq a \\ -X & \text{si } |X| > a \end{cases}.$$

Montrer que Y suit la loi normale $\mathcal{N}(0, 1)$, mais que (X, Y) n'est pas un vecteur gaussien.

On a quand même un exemple important pour lequel on a un vecteur gaussien facilement.

Théorème 4.7. *Si X_1, \dots, X_n sont des variables aléatoires (réelles) suivant une loi normale et qu'elles sont indépendantes, alors (X_1, \dots, X_n) est un vecteur gaussien.*

Le théorème suivant traite des liens entre vecteurs gaussiens et applications linéaires.

Théorème 4.8. *Soit $X = (X_1, \dots, X_n)$ un vecteur aléatoire de taille n , et $Y = (Y_1, \dots, Y_m)$ un vecteur gaussien, de taille m . Si $X = F(Y)$, où F est une application linéaire de \mathbb{R}^m dans \mathbb{R}^n , alors X est aussi un vecteur gaussien.*

Exercice 38. Démontrer ce théorème (par exemple en explicitant composante par composante l'hypothèse $X = F(Y)$).

Ce théorème est absolument essentiel dans la pratique. Il arrive fréquemment que le vecteur étudié soit construit de façon linéaire à partir de variables aléatoires qui forment un vecteur gaussien.

Pour terminer, on donne les définitions importantes suivantes.

Définition 4.9. *Soit $X = (X_1, \dots, X_n)$ un vecteur aléatoire gaussien. Son espérance $m = \mathbb{E}[X] \in \mathbb{R}^n$ est le vecteur $(m_1, \dots, m_n) := (\mathbb{E}[X_1], \dots, \mathbb{E}[X_n])$. Sa matrice de covariance $C = \text{Cov}(X)$ est la matrice carrée de terme général $C_{i,j} := \text{Cov}(X_i, X_j) = \mathbb{E}[X_i X_j] - \mathbb{E}[X_i] \mathbb{E}[X_j]$, pour tout $1 \leq i, j \leq n$.*

Le vecteur m est quelconque dans \mathbb{R}^n ; par contre, une matrice de covariance C est nécessairement symétrique et positive.

Exercice 39. Montrer que pour tout vecteur $\lambda \in \mathbb{R}^n$ - identifié à un vecteur colonne - on a

$$\langle \lambda, C\lambda \rangle := \sum_{i,j=1}^n C_{i,j} \lambda_i \lambda_j = \text{Var}\left(\sum_{i=1}^n \lambda_i X_i\right) \geq 0.$$

En profiter pour revoir l'algèbre bilinéaire, et la théorie des matrices symétriques.

Remarque 4.10. Ces notions d'espérance et de matrice de covariance peuvent être définies de la même manière hors du cas gaussien. Néanmoins, le théorème suivant n'est valable que dans le cas gaussien.

Théorème 4.11. *La donnée du vecteur espérance m et de la matrice de covariance C suffit à caractériser un vecteur gaussien; plus précisément, si on suppose*

1. X et Y sont deux vecteurs gaussiens (de même taille n);
2. $\mathbb{E}[X] = m = \mathbb{E}[Y]$;
3. $\text{Cov}(X) = C = \text{Cov}(Y)$;

alors X et Y ont même loi, notée $\mathcal{N}(m, C)$. C'est une loi de probabilité sur \mathbb{R}^n .

De plus, étant donné un vecteur $m \in \mathbb{R}^n$, et une matrice C **symétrique positive**, alors la loi $\mathcal{N}(m, C)$ est bien définie : il existe un vecteur aléatoire gaussien, d'espérance m et de matrice de covariance C .

Exemple 4.12. Si X_1, \dots, X_n sont indépendantes, de lois respectives $\mathcal{N}(m_j, \sigma_j^2)$, pour $1 \leq j \leq n$, alors $m = (m_1, \dots, m_n)$ et $C = \text{Diag}(\sigma_1^2, \dots, \sigma_n^2)$, matrice diagonale.

Par conséquent, on a la caractérisation de l'indépendance **pour des vecteurs gaussiens**, par le fait que la matrice de covariance est une matrice diagonale.

4.3 Processus Gaussiens

Soit $\mathbb{T} = \mathbb{N}$ ou \mathbb{R}^+ (ou un sous-intervalle $\{0, \dots, N\}$, ou $[0, T]$), et $X = (X_t)_{t \in \mathbb{T}}$ un processus stochastique.

Définition 4.13. Il s'agit d'un processus aléatoire gaussien si pour tout entier $n \geq 1$ et tout n -uplet d'instant $(t_1, \dots, t_n) \in \mathbb{T}^n$, le vecteur aléatoire $(X_{t_1}, \dots, X_{t_n})$ est un vecteur gaussien.

En particulier, à chaque instant $t \in \mathbb{T}$, la variable aléatoire X_t suit une loi normale. Comme pour le cas des vecteurs gaussiens, il s'agit d'une condition nécessaire, mais non suffisante.

On dispose, comme pour le cas des vecteurs gaussiens, des notions d'espérance, de fonction de covariance, et d'un théorème de caractérisation de la loi d'un processus gaussien, et d'existence étant données une fonction d'espérance et une fonction de covariance.

Définition 4.14. La fonction d'espérance du processus gaussien $X = (X_t)_{t \in \mathbb{T}}$ est la fonction $m : \mathbb{T} \rightarrow \mathbb{R}$, définie par

$$m(t) = \mathbb{E}[X_t], \quad \text{pour tout } t \in \mathbb{T}.$$

La fonction de covariance d'un processus gaussien $(X_t)_{t \in \mathbb{T}}$ est la fonction $C : \mathbb{T} \times \mathbb{T} \rightarrow \mathbb{R}$, définie par

$$C(t, s) = \text{cov}(X_t, X_s), \quad \text{pour tous } t, s \in \mathbb{T}.$$

Il s'agit d'une fonction symétrique et positive :

– pour tous $t, s \in \mathbb{T}$, $C(t, s) = C(s, t)$;

- pour tous $n \geq 1$, $(t_1, \dots, t_n) \in \mathbb{T}^n$ et $(\lambda_1, \dots, \lambda_n) \in \mathbb{R}^n$, on a $\sum_{i,j=1}^n C(t_i, t_j) \lambda_i \lambda_j \geq 0$.

Théorème 4.15. *La donnée de la fonction d'espérance m et de la fonction de covariance C suffit à caractériser un processus gaussien. Plus précisément, si on suppose*

1. $X = (X_t)_{t \in \mathbb{T}}$ et $Y = (Y_t)_{t \in \mathbb{T}}$ sont deux processus gaussiens (indexés par \mathbb{T});
2. pour tout $t \in \mathbb{T}$, $\mathbb{E}[X_t] = m(t) = \mathbb{E}[Y_t]$;
3. pour tous $t, s \in \mathbb{T}$, $\text{Cov}(X_t, X_s) = C(t, s) = \text{Cov}(Y_t, Y_s)$;

alors les processus X et Y ont même loi, au sens qu'ils ont le mêmes lois finidimensionnelles : c'est-à-dire que pour tout $n \in \mathbb{N}$ et tout $(t_1, \dots, t_n) \in \mathbb{T}^n$, les vecteurs (gaussiens) $(X_{t_1}, \dots, X_{t_n})$ et $(Y_{t_1}, \dots, Y_{t_n})$ ont la même loi.

De plus, étant donnés une fonction $m : \mathbb{T} \rightarrow \mathbb{R}$ et une fonction $C : \mathbb{T} \times \mathbb{T} \rightarrow \mathbb{R}$ symétrique positive, alors il existe un processus gaussien $X = (X_t)_{t \in \mathbb{T}}$, dont la fonction d'espérance est m et la fonction de covariance est C .

On a deux exemples fondamentaux.

Le premier est le *bruit blanc gaussien*, noté $\eta = (\eta_t)_{t \in \mathbb{T}}$. Il s'agit d'un processus gaussien centré (i.e. $m(t) = 0$ pour tout $t \in \mathbb{T}$), dont la fonction de covariance est donnée par $C(t, s) = 1$ si $t = s$, 0 si $t \neq s$. En fait, le processus est constitué de variables aléatoires de même loi gaussienne standard, toutes indépendantes.

Le deuxième cas est plus intéressant, c'est le *mouvement brownien*, avec $\mathbb{T} = \mathbb{R}^+$ ou $[0, T]$. On le note $(B_t)_{t \in \mathbb{T}}$ (parfois aussi avec la lettre W pour Wiener au lieu de B pour Brown). Ce processus gaussien est centré, et sa fonctionnelle de covariance est donnée par $C(t, s) = \min(s, t)$. On vérifiera par la suite que C satisfait bien à la condition de positivité.

4.4 Différente caractérisation du mouvement brownien

Le mouvement brownien, défini au dessus comme un processus gaussien centré de fonction de covariance $C(s, t) = \min(s, t)$, peut aussi être défini et caractérisé de la façon suivant.

Proposition 4.16. *Le mouvement brownien $(B_t)_{t \in \mathbb{R}^+}$ est un P.A.I.S. caractérisé par le fait que $B_0 = 0$ et pour tout $s < t \in \mathbb{R}_+$, $B_t - B_s$ suit la loi gaussienne $\mathcal{N}(0, t - s)$.*

Montrons que sa fonction de covariance est alors bien déterminée : soit $0 \leq s \leq t$, on doit calculer

$$\begin{aligned}\text{Cov}(X_t, X_s) &= \mathbb{E}[X_t X_s] \\ &= \mathbb{E}[(X_s + X_t - X_s)X_s] \\ &= \mathbb{E}[X_s^2] + \mathbb{E}[(X_t - X_s)X_s] \\ &= s + \mathbb{E}[X_t - X_s]\mathbb{E}[X_s] \\ &= s = \min(s, t).\end{aligned}$$

Comme promis plus haut, vérifions que $C(s, t) = \min(s, t)$ définit bien une fonction de covariance. Pour cela, inspirons-nous du calcul précédent : on a écrit X_t comme la somme de deux variables indépendantes... Donc si on part d'une famille finie de variables aléatoires gaussiennes, et qu'on définit comme nouveau processus la marche aléatoire associée, on va trouver un processus gaussien, qui aura comme fonction de covariance la fonction dont on cherche à prouver la positivité!

Proposition 4.17. *La fonction $C : \mathbb{R}^+ \rightarrow \mathbb{R}$, avec $C(s, t) = \min(s, t)$, est symétrique positive. Par conséquent, le mouvement brownien est bien défini, en tant que processus gaussien centré dont C est la fonction de covariance.*

Démonstration. Il faut prouver que pour tout $n \geq 1$, tout n -uplet (ordonné) $0 \leq t_1 \leq \dots \leq t_n$, et tout n -uplet $(\lambda_1, \dots, \lambda_n)$, on a

$$\sum_{i,j=1}^n C(t_i, t_j) \lambda_i \lambda_j \geq 0,$$

ce qui est équivalent au fait que la matrice $(C(t_i, t_j))_{1 \leq i, j \leq n}$ est symétrique positive.

Considérons une famille (X_1, \dots, X_n) de n variables indépendantes gaussiennes, de loi $\mathcal{N}(0, t_1), \mathcal{N}(0, t_2 - t_1), \dots, \mathcal{N}(0, t_n - t_{n-1})$. Considérons alors les variables aléatoires

$$\begin{aligned}Y_1 &= X_1 \\ Y_2 &= X_1 + X_2 \\ &\vdots \\ Y_n &= X_1 + \dots + X_{n-1} + X_n.\end{aligned}$$

On voit que le vecteur $Y = (Y_1, \dots, Y_n)$ s'écrit comme une transformation linéaire du vecteur $X = (X_1, \dots, X_n)$ (Exercice : écrire la matrice A telle que $Y^T = AX^T$).

Donc Y est un vecteur gaussien, centré. Calculons sa matrice de covariance : pour tout $1 \leq i \leq j \leq n$

$$\begin{aligned}\text{Cov}(Y_i, Y_j) &= \mathbb{E}[Y_i Y_j] - \mathbb{E}[Y_i] \mathbb{E}[Y_j] = \mathbb{E}[Y_i Y_j] \\ &= \mathbb{E}[Y_i(Y_i + Y_j - Y_i)] \\ &= \mathbb{E}[Y_i^2] + \mathbb{E}[Y_i(Y_j - Y_i)] \\ &= \mathbb{E}[Y_i^2] + \mathbb{E}[Y_i] \mathbb{E}[Y_j - Y_i] \\ &= \mathbb{E}[Y_i^2] = \text{Var}(Y_i).\end{aligned}$$

On va calculer par récurrence $\text{Var}(Y_i)$: pour $2 \leq i \leq n$,

$$\begin{aligned}\text{Var}(Y_i) &= \text{Var}(Y_{i-1} + X_i) \\ &= \text{Var}(Y_{i-1}) + 2\text{Cov}(Y_{i-1}, X_i) + \text{Var}(X_i) \\ &= \text{Var}(Y_{i-1}) + 0 + (t_{i+1} - t_i),\end{aligned}$$

grâce à l'indépendance des variables aléatoires X_j . On en déduit alors que

$$\text{Cov}(Y_i, Y_j) = \text{Var}(Y_i) = t_i = \min(t_i, t_j).$$

(la formule finale étant symétrique en (i, j) , il n'y a pas besoin de considérer le cas $j < i$.)

Ceci prouve que $(C(t_i, t_j))_{1 \leq i, j \leq n}$ est la matrice de covariance d'un vecteur gaussien, et est en particulier symétrique positive. \square

4.5 Quelques exercices autour du mouvement brownien

Exercice 40. Soit $B = (B_t)_{t \geq 0}$ un mouvement brownien, et soit $T > 0$ un instant fixé. Montrer que le processus $B^{(T)}$ tel que $B_t^{(T)} = B_{T+t} - B_T$, est un autre mouvement brownien, indépendant du processus $(B_s)_{0 \leq s \leq T}$ - c'est-à-dire de tout vecteur aléatoire $(B_{s_1}, \dots, B_{s_n})$ avec $0 \leq s_1 \leq \dots \leq s_n \leq T$.

Exercice 41. Soit $B = (B_t)_{t \geq 0}$ un mouvement brownien.

1. Soit $a > 0$; montrer que le processus $B^{a,1}$ tel que $B^{a,1}(t) = B_{at}$ est un processus gaussien, préciser sa fonction d'espérance et sa fonction de covariance.
2. Soit $b \in \mathbb{R}^*$; montrer que le processus $B^{1,b}$ tel que $B^{1,b}(t) = bB_t$ est un processus gaussien, préciser sa fonction d'espérance et sa fonction de covariance.

3. Étant donnés $a > 0$ et $b \in \mathbb{R}^*$, à quel condition le processus $B^{a,b}$ tel que $B_t^{a,b} = bB_{at}$ est-il un mouvement brownien ?

Exercice 42. On considère une marche aléatoire dont les incréments sont des va iid $(X_n)_{n \in \mathbb{N}^*}$, de telle sorte que $S_n = \sum_{k=1}^n X_k$. On définit une suite de processus à temps continu : pour tout $n \in \mathbb{N}$ fixé, on pose pour tout $t \in \mathbb{R}$ $X_t^n = S_{[nt]}$, où $[\cdot]$ est la notation pour la partie entière d'un nombre réel (le plus grand entier qui lui est inférieur ou égal). On a ainsi effectué un changement d'échelle de temps pour observer la marche aléatoire.

1. Que se passe-t-il pour $n = 0$?
2. Soit $n \in \mathbb{N}^*$ fixé ; que vaut X_0^n ?
3. Supposons que les va X_i suivent une loi gaussienne $\mathcal{N}(0, 1)$; montrer que pour tout $n \in \mathbb{N}^*$ le processus $X^n = (X_t^n)_{t \geq 0}$ est un processus gaussien ; déterminer ses caractéristiques.
4. On opère désormais un changement d'échelle en espace : déterminer la bonne fonction de zoom z définie sur les entiers, telle qu'on est susceptible d'obtenir à la limite un mouvement brownien, pour le processus B^n tel que

$$B_t^n = z(n)X_t^n = z(n)S_{[nt]}.$$

5. Ce résultat ne dépend en fait pas de la nature gaussienne des accroissements de la marche. Interpréter le cas de la convergence à $t = 1$ fixé.

On ne demande pas de prouver les convergences évoquées ci-dessus !