

Introduction aux Processus Stochastiques

Cours ESIEA 4A

septembre-octobre 2013

Plan du cours

1	Introduction: premiers exemples de processus stochastiques	1
1.1	Marche aléatoire	1
1.2	Ruine du joueur	2
1.3	Arbre de Galton-Watson	2
1.4	Mouvement brownien	3
1.5	Une file d'attente	3
2	Généralités sur les processus stochastiques	4
2.1	Rappels de probabilités	4
2.1.1	Espace de probabilité, variables aléatoires, loi	4
2.1.2	Exemples de variables aléatoires discrètes et continues	6
2.1.3	Espérance	7
2.1.4	Caractérisation d'une loi sur \mathbb{R}	8
2.1.5	Probabilités conditionnelles	9
2.2	Processus stochastiques	11
2.2.1	Définitions	11
2.2.2	Filtrations, temps d'arrêt	12
2.2.3	Exemple: variables aléatoires indépendantes	13
2.2.4	Quelques quantités importantes	15
2.2.5	Exercices	16
3	Chaînes de Markov	17
3.1	Définition, exemples	17
3.2	Noyau de transition	18
3.3	Graphe de transition	19
3.4	Distribution	20
3.5	Lois invariantes	22
3.6	Chaînes de Markov et temps d'arrêt	23
3.7	Classification des états	24
3.7.1	Communication	24
3.7.2	Réurrence-transience	26
3.7.3	Espérance du temps de retour	27
3.7.4	Périodicité	28

3.8	Calcul pratique de la probabilité et de l'espérance du temps d'absorption . .	28
3.9	Comportement en temps long	30
3.10	Complément: Chaînes de Markov non homogènes	32
4	Processus Gaussiens	34
4.1	Loi normale	34
4.1.1	Loi standard	34
4.1.2	Loi normale, cas général	35
4.1.3	Le Théorème Central Limite	36
4.2	Vecteurs gaussiens	37
4.3	Processus gaussiens	38
4.4	Indépendance des accroissements	40
4.5	Stationnarité des accroissements	40
4.6	Mouvement Brownien	41
4.7	Quelques exercices autour du mouvement brownien	42
5	Processus de Poisson	44
5.1	La loi de Poisson	44
5.2	La loi exponentielle	45
5.3	Processus de comptage	45
5.4	Processus de Poisson	46
5.5	Quelques propriétés du processus de Poisson	47
5.6	Processus de Poisson marqué	47
5.7	Processus de Poisson composé	48
5.8	Processus Markovien de saut	48
5.8.1	Description chaîne de Markov induite	49
5.8.2	Description "générateur"	50
5.8.3	Lien entre les deux descriptions	50
5.8.4	Lois invariantes	50
5.9	Exemples de modèles de files d'attente	51
5.9.1	$M/M/0$	51
5.9.2	$M/M/1$	51
5.9.3	$M/M/s$	52
5.9.4	$M/M/\infty$	52
5.9.5	$M/M/1/k$	53
5.9.6	$M/M/s/0$	53

Avertissement

Ces notes contiennent un peu plus de matériel que ce qui a été vu en cours. A certains endroits, il ne s'agit pas de retenir le détail des calculs, mais plutôt leur principe. Il pourra être demandé de refaire le raisonnement dans un cadre plus simple.

Par contre, les définitions et théorèmes sont à connaître!

Et les exercices sont à faire!

Quelques références:

Markov Chains, Brémaud.

Probability and Random Processes, Grimmett & Ztirzaker.

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

Chapitre 1

Introduction: premiers exemples de processus stochastiques

1.1 Marche aléatoire

Fixons une dimension $d \in \mathbb{N}^*$, par exemple $d \in \{1, 2, 3\}$ pour pouvoir faire des dessins.

On considère une suite de variables aléatoires, indépendantes, $(X_n)_{n \in \mathbb{N}^*}$, toutes de même loi, pouvant prendre uniformément (c'est-à-dire avec la même probabilité $1/(2d)$) les $2d$ valeurs suivantes:

$$\begin{aligned} & (1, 0, \dots, 0), (-1, 0, \dots, 0) \\ & \vdots \\ & (0, \dots, 0, 1, 0, \dots, 0), (0, \dots, 0, -1, 0, \dots, 0) \\ & \vdots \\ & (0, \dots, 0, 1), (0, \dots, 0, -1). \end{aligned}$$

On se déplace en fait sur un réseau, \mathbb{Z}^d .

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 $n+1$ dépend de la valeur de X_{n+1} (dans laquelle des d directions possibles, puis dans lequel des deux sens possibles).

Une question naturelle qu'on peut se poser est la possibilité de retour à l'origine: se produit-il presque sûrement? si oui, quel est la loi du temps de retour? quelle est sa moyenne?

Les outils présentés dans ce cours permettent d'aborder ces questions, et d'y répondre (si on fait quelques calculs supplémentaires...). Il se trouve que le retour presque sûr n'est vrai que pour $d = 1$ ou $d = 2$, et que dans ce cas le temps de retour est infini en moyenne!

La marche aléatoire présentée ci-dessus est souvent qualifiée de *simple*; des lois de mouvements plus générales peuvent également servir à définir des marches aléatoires.

1.2 Ruine du joueur

On considère la situation d'un joueur contre la banque: 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.

La première question à se poser est la suivante: le jeu se termine-t-il ou continue-t-il indéfiniment?

Une seconde question très naturelle est alors de déterminer, 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.

1.3 Arbre 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.

1.4 Mouvement brownien

Reprenons l'exemple de la marche aléatoire (simple) en dimension $d = 1$. En changeant convenablement l'échelle de temps et l'échelle d'espace à laquelle on considère le processus (on verra plus précisément de quoi il s'agit dans un chapitre ultérieur), avec un zoom de plus en plus précis, on obtient un processus évoluant continûment dans les deux variables de temps et d'espace, avec une distribution de probabilité particulière (et universelle, au sens où le choix de la marche aléatoire "simple" n'est pas fondamental): il s'agit d'un processus gaussien, à accroissements indépendants et stationnaires, appelé mouvement brownien.

Le nom fait référence à sa découverte expérimentale au début du 19ème siècle, dans des recherches en botanique (par un certain R. Brown). Les premières études mathématiques datent du début du 20ème siècle (Einstein, Bachelier, Lévy, Wiener...).

Aujourd'hui, c'est un processus stochastique "de base", aux applications multiples: finance, physique, chimie, biologie...

1.5 Une file d'attente

Dans le dernier 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 la aussi assez universel) est donné par le processus de Poisson (par référence au mathématicien français Poisson, du début du 19ème 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.

Chapitre 2

Généralités sur les processus stochastiques

2.1 Rappels de probabilités

2.1.1 Espace de probabilité, variables aléatoires, loi

La phrase typique dans la théorie et les exercices est la suivante: *Soit X une variable aléatoire de loi μ sur un espace de probabilités $(\Omega, \mathcal{F}, \mathbb{P})$.*

Essayons d'explicitier chacun des termes importants. C'est un peu abstrait, on verra les exemples plus usuels dans le paragraphe suivant.

Ω est un ensemble. Un élément $\omega \in \Omega$ est appelé une éventualité.

\mathcal{F} est une tribu. C'est un sous-ensemble de l'ensemble $\mathcal{P}(\Omega)$ des parties de Ω ; un sous-ensemble $A \in \mathcal{F}$ est appelé un événement.

La définition rigoureuse d'une tribu est hors programme, mais donnons-la quand même:

- $\emptyset \in \mathcal{F}, \Omega \in \mathcal{F}$;
- \mathcal{F} est stable par passage au complémentaire: $A \in \mathcal{F} \Rightarrow A^C \in \mathcal{F}$;
- \mathcal{F} est stable par réunion et intersection dénombrable:
 $(A_i \in \mathcal{F} \forall i \in \mathbb{N}) \Rightarrow \bigcup_{i \in \mathbb{N}} A_i \in \mathcal{F}, \bigcap_{i \in \mathbb{N}} A_i \in \mathcal{F}$.

Il y a trois exemples importants:

1. $\mathcal{F} = \{\emptyset, \Omega\}$;
2. si $\Omega \subset \mathbb{N}$, $\mathcal{F} = \mathcal{P}(\Omega)$;
3. si $\Omega = \mathbb{R}$, $\mathcal{F} = \mathcal{B}(\mathbb{R})$, tribu *borélienne*; retenir qu'elle contient tous les intervalles (ouverts, semi-ouverts, fermés, bornés ou non...), mais qu'elle contient également bien plus d'événements.

Mais il y en a d'autres, notamment lorsqu'on étudie des processus stochastiques...

\mathbb{P} est une mesure de probabilités: c'est une application définie sur \mathcal{F} une tribu, à valeurs dans $[0, 1]$, qui vérifie:

- $\mathbb{P}(\emptyset) = 0, \mathbb{P}(\Omega) = 1$;
- si $(A_i)_{i \in \mathbb{N}}$ est une famille **dénombrable** d'ensembles **deux à deux disjoints** (c'est-à-dire $A_i \cap A_j = \emptyset$ si $i \neq j$), alors

$$\mathbb{P}\left(\bigcup_{i \in \mathbb{N}} A_i\right) = \sum_{i \in \mathbb{N}} \mathbb{P}(A_i).$$

Une définition importante: on dit qu'une propriété est presque sûre si elle correspond à un événement de probabilité 1. A l'inverse, on parle d'événement négligeable lorsque sa probabilité est égale à 0.

Des exemples seront donnés dans le paragraphe qui suit. Une variable aléatoire (abrégé par la suite en v.a.) X est une application de Ω dans \mathbb{R} (variables aléatoires dites *réelles*) ou \mathbb{N} (variables aléatoires dites *discrètes*). De plus, on suppose une condition supplémentaire de *mesurabilité*, liée à la tribu \mathcal{F} sur Ω :

1. pour une v.a. discrète X , on demande que pour tout $n \in \mathbb{N}$, $X^{-1}(\{n\}) \in \mathcal{F}$;
2. pour une v.a. réelle X , on demande que pour tout $a \in \mathbb{R}$, $X^{-1}((-\infty, a]) \in \mathcal{F}$ (on note aussi cet événement $\{X \leq a\}$).

On rappelle que pour tout ensemble B , $X^{-1}(B) = \{\omega \in \Omega; X(\omega) \in B\}$.

Un exemple de propriété presque sûre: deux variables aléatoires X et Y sont dites égales presque sûrement si l'événement $\{X = Y\}$ est de probabilité 1.

Parfois, on aura des variables aléatoires à valeurs dans $\mathbb{N} \cup \{\infty\}$; il suffit juste de vérifier pour tout $n \in \mathbb{N}$, $X^{-1}(\{n\}) \in \mathcal{F}$, la propriété $X^{-1}(\{+\infty\}) \in \mathcal{F}$ étant automatiquement vérifiée (pourquoi?).

Il sera aussi utile de considérer des variables aléatoires discrètes prenant des valeurs autres que des entiers naturels: elles prennent leurs valeurs dans des ensembles $(x_i)_{i \in I}$ avec $x_i \in \mathbb{R}$ pour tout $i \in I$, et I un ensemble d'indices inclus dans \mathbb{N} (fini ou dénombrable).

Si $\omega \in \Omega$, on dit que $X(\omega)$ est une réalisation de la variable aléatoire X .

La morale est la suivante: la tribu \mathcal{F} représente l'ensemble des événements A dont on a le droit de mesurer la probabilité $\mathbb{P}(A)$ - d'où la terminologie "mesurable". Lorsqu'on considère une variable aléatoire X , on va s'intéresser aux ensembles $\{X \leq a\}$; pour mesurer leur probabilité, il faut bien qu'ils soient mesurables!

On a dit qu'il est important de préciser la tribu pour définir la notion de variable aléatoire. En effet, dans le cas discret par exemple, la mesurabilité par rapport à la tribu $\{\emptyset, \Omega\}$ signifie être une fonction constante, alors que toute fonction de \mathbb{N} dans \mathbb{N} est mesurable par rapport à la tribu $\mathcal{P}(\mathbb{N})$!

On termine avec la notion de loi d'une variable aléatoire X : c'est l'unique mesure de probabilité notée μ_X ou \mathbb{P}_X , telle que:

1. pour une v.a. discrète X , pour tout $n \in \mathbb{N}$, on a $\mu_X(\{n\}) := \mathbb{P}(X^{-1}(\{n\}))$;
2. pour une v.a. réelle X , pour tout $a \in \mathbb{R}$, on a $\mu_X((-\infty, a]) := \mathbb{P}(X^{-1}(a))$.

Plus généralement, pour tout ensemble B mesurable (sous-ensemble de \mathbb{N} ou \mathbb{R} selon les cas), $\mu_X(B) = \mathbb{P}(X^{-1}(B))$.

Les remarques suivantes sont assez fondamentales, mais un peu abstraites... Ce n'est pas forcément utile de les retenir; essayez de les comprendre, mais ne vous posez surtout pas de questions, il n'y aura jamais de problème dans la suite!!

Remarque 2.1 Dans la pratique, le choix de l'espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$ importe peu, et on se contente de le mentionner quand on définit la (ou les) variable(s) aléatoire(s) utile(s) au problème. Ce qui sera important, c'est que la loi de cette (ou ces) variable(s) aléatoire(s) correspondent à ce qu'on souhaite modéliser; leur réalisation ne compte pas.

Par exemple (attention c'est un peu abstrait) on a le choix canonique: étant donnée une loi μ , un choix possible est $(\Omega = \mathbb{R}, \mathcal{F} = \mathcal{B}(\mathbb{R}), \mathbb{P} = \mu)$, et $X(\omega) = \omega$. Mais ce n'est pas le seul.

Remarque 2.2 Le choix de modélisation par un espace de probabilité n'est pas unique, pour une loi donnée. Par exemple, si $\Omega = \{0, 1\}$, la loi des variables aléatoires définies par $X_1(\omega) = \omega$ et $X_2(\omega) = 1 - \omega$ est la même:

$$\mathbb{P}(X_1 = 0) = \mathbb{P}(X_2 = 0) = 1/2, \quad \mathbb{P}(X_1 = 1) = \mathbb{P}(X_2 = 1) = 1/2.$$

Alors qu'on a toujours $X_1(\omega) \neq X_2(\omega)$, pour tout ω (réalisations différentes, mais lois identiques).

Remarque 2.3 Pourquoi définir plusieurs variables aléatoires sur le même espace de probabilité?

Réponse: ça permet de les additionner, multiplier, comparer, ω par ω .

Plus généralement, si X est une variable aléatoire $\Omega \rightarrow \mathbb{R}$, et si $\phi : \mathbb{R} \rightarrow \mathbb{R}$ est continue (même mesurable, au sens où c'est une variable aléatoire), alors $\phi(X)$ est aussi une variable aléatoire.

Attention: on peut additionner des variables aléatoires, mais surtout pas des lois!!

2.1.2 Exemples de variables aléatoires discrètes et continues

Lois discrètes

1. Loi uniforme sur $\{1, \dots, N\}$: $\mu(\{i\}) = \frac{1}{N}$ pour tout $1 \leq i \leq N$.
2. Loi de Bernouilli, notée $\mathcal{B}(p)$, pour $0 \leq p \leq 1$: $\mu(0) = 1 - p$, $\mu(1) = p$.
3. Loi binomiale $\mathcal{B}(n, p)$, $0 \leq p \leq 1$: $\mu(k) = \binom{n}{k} p^k (1 - p)^{n-k}$, pour tout $0 \leq k \leq n$.
4. Loi géométrique $\mathcal{G}(p)$: pour tout $k \in \mathbb{N}^*$, $\mu(\{k\}) = p(1 - p)^{k-1}$.

Lois continues à densité

On va tout d'abord considérer des lois μ à densité: il existe une fonction f positive d'intégrale 1, telle que pour tout $x \in \mathbb{R}$,

$$\mathbb{P}((-\infty, x]) = \int_{-\infty}^x f(t)dt.$$

Dans ce cas, on a toujours $\mathbb{P}(\{x\}) = 0$, pour tout $x \in \mathbb{R}$; par conséquent, pour $a < b$ donnés, les intervalles (a, b) , $(a, b]$, etc... ont la même probabilité.

1. Loi uniforme sur (a, b) : pour tout $a < x < y < b$, $\mu([x, y]) = \frac{y-x}{b-a}$, rapport des longueurs des intervalles. Exercice: mesure de $(-\infty, x)$, pour tout $x \in \mathbb{R}$? Densité: $f(t) = \frac{1}{b-a}$ si $a < t < b$, $= 0$ sinon.
2. Loi exponentielle de paramètre $\lambda > 0$: de densité $f(x) = \lambda \exp(-\lambda x)$ si $x \geq 0$, 0 sinon. Voir le Chapitre 5.
3. Loi normale/gaussienne: de densité $f(x) = \frac{1}{\sqrt{2\pi}} \exp(-x^2/2)$, pour tout $x \in \mathbb{R}$. Voir le Chapitre 4.

2.1.3 Espérance

Soit X une variable aléatoire définie sur l'espace de probabilités $(\Omega, \mathcal{F}, \mathbb{P})$, à valeurs dans l'espace d'états \mathbb{N} ou \mathbb{R} ; on peut en fait toujours supposer qu'il s'agit de \mathbb{R} (vu que $\mathbb{N} \subset \mathbb{R}$!).

Si $\phi : \mathbb{R} \rightarrow \mathbb{R}$ est une fonction, disons continue, alors $\phi(X)$ est une variable aléatoire, définie aussi sur $(\Omega, \mathcal{F}, \mathbb{P})$, à valeurs dans \mathbb{R} ; précisément, la notation $\phi(X)$ représente la composée $\phi \circ X$ des fonctions ϕ et X , de telle sorte que pour tout $\omega \in \Omega$ on a $\phi(X)(\omega) = \phi(X(\omega))$.

Remarque 2.4 *La bonne notion pour la régularité de ϕ est la mesurabilité plutôt que la continuité...*

Supposons de plus que ϕ est positive: alors on peut définir

$$\mathbb{E}[\phi(X)] = \int_{\Omega} \phi(X(\omega))\mathbb{P}(d\omega) = \int_{\mathbb{R}} \phi(x)\mu_X(dx) \in [0, +\infty],$$

où μ_X est la loi de X . A noter en particulier: cette quantité ne dépend que de la loi de la variable aléatoire, pas de ses réalisations. De plus, elle peut être infinie.

Si ϕ n'est pas positive, on peut définir tout d'abord $\mathbb{E}[|\phi(X)|] \in [0, +\infty]$; si en fait $\mathbb{E}[|\phi(X)|] < +\infty$ (on dit que $\phi(X)$ est *intégrable*), on peut ensuite définir $\mathbb{E}[\phi(X)] \in \mathbb{R}$; dans le cas contraire, on ne peut pas. Une condition suffisante (mais non nécessaire) d'intégrabilité est de supposer ϕ bornée.

Détaillons les cas des lois discrètes et des lois à densité; c'est ce qu'il faut retenir (notamment pour faire des calculs), la formule précédente contient les deux cas (et est en fait plus générale): sous réserve d'intégrabilité

1. Lois discrètes:

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

2. Lois à densité:

$$\mathbb{E}[\phi(X)] = \int_{-\infty}^{+\infty} \phi(x) f(x) dx.$$

Exemple 2.5 Pour tout événement A , on a $\mathbb{P}(X \in A) = \mu_X(A) = \mathbb{E}[\mathbb{1}_A(X)]$, où $\mathbb{1}_A$ est la fonction indicatrice de A : elle vaut 1 sur A et 0 sur son complémentaire A^c .

La définition suivante donne quelques cas importants:

Définition 2.6 • Avec le choix $\phi(x) = x$: si X est intégrable (c'est-à-dire $\mathbb{E}|X| < +\infty$), on définit l'espérance de X comme étant la quantité $\mathbb{E}[X]$.

- Avec le choix $\phi(x) = (x - \mathbb{E}[X])^2$: si $\phi(X)$ est intégrable (c'est-à-dire $\mathbb{E}(X - \mathbb{E}[X])^2 < +\infty$), on définit la variance de X comme étant la quantité $\mathbb{E}(X - \mathbb{E}[X])^2 = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$.
- Plus généralement, on définit lorsque c'est possible les moments de X d'ordre n quelconque comme étant $\mathbb{E}[X^n]$.

Exercice 1 Calculer l'espérance et la variance pour les lois données dans la section précédente. Et essayer de les retenir...

Parfois, au lieu d'espérance on parle de moyenne, par référence au cas des lois uniformes, où cette interprétation est la plus claire. Dans le cas général, il s'agit de moyennes *pondérées* par la connaissance de la loi de la variable aléatoire.

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

Commençons par noter que la caractérisation d'une loi discrète est en théorie très facile: il suffit de connaître $\mu(n)$ pour tout entier n . Le cas d'une variable à densité (ou même plus générale, un peu en dehors du cadre du cours...) est moins clair. On dispose essentiellement de trois moyens: le premier est théorique; les deux suivants sont plus calculatoires, lequel choisir dépendant des situations.

- Une loi de probabilité μ , associée (par exemple canoniquement, voir la remarque 2.1) à une variable aléatoire X , est caractérisée par la connaissance de $\mathbb{E}[\phi(X)] = \int \phi(x) \mu(dx)$ pour toute fonction continue bornée ϕ .
- **Fonction de répartition**: c'est la fonction

$$\begin{aligned} F_X : \mathbb{R} &\rightarrow \mathbb{R} \\ x &\mapsto \mathbb{P}(X \leq x) = \mathbb{E}[\mathbb{1}_{(-\infty, x]}]. \end{aligned}$$

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 à densité, F_X est toujours continue, et vérifie $F_X(x) = \int_{-\infty}^x f(t)dt$.

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.

Exercice 2 Dessiner la fonction de répartition pour une loi uniforme discrète et une loi uniforme continue, ainsi que pour la loi exponentielle.

- **Fonction caractéristique:** c'est la fonction

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

On peut vérifier par exemple que sous réserve de validité du calcul on a $\mathbb{E}[X] = \frac{d}{dt}\big|_{t=0}\varphi_X(t)$.

Remarque 2.7 Dans le cas des lois discrètes, on dispose également de la **série génératrice**:

$$g_X(x) = \sum_{n \in \mathbb{N}} \mathbb{P}(X = n)x^n.$$

Cette série de fonctions converge uniformément sur $[-1, 1]$ (en la variable réelle x , on aurait aussi pu considérer la variable complexe), avec entre autres comme propriétés $g_X(1) = 1$, $\mathbb{P}(X = n) = n!g_X^{(n)}(0)$, $\mathbb{E}[X] = g_X'(1)$ lorsque $\mathbb{E}|X| < +\infty$.

Remarque 2.8 Plus généralement, de façon similaire on peut caractériser la loi de vecteurs aléatoires $X = (X_1, \dots, X_d)$ à valeurs dans \mathbb{R}^d , chaque X_i étant une variable aléatoire réelle - on parle de loi jointe. Il faut considérer des fonctions de répartition et caractéristique jointes, dépendant de n variables.

2.1.5 Probabilités conditionnelles

Définition 2.9 Soit A et B deux événements.

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'événement B ; la division correspond à une normalisation: l'application

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

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 $\mathbb{P}(A|B) = \mathbb{P}(A)$;
- $\mathbb{P}(A \cap B) = \mathbb{P}(A|B)\mathbb{P}(B)$; 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 $j \in I$,

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

Pour terminer cette section de "rappels et compléments", voici la notion d'espérance conditionnelle dans le cas de variables aléatoires discrètes; la construction générale est plus complexe, et on n'en aura pas besoin ici.

Considérons deux variables aléatoires discrètes X et Y , prenant les valeurs respectives (potentiellement non entières, mais ça ne change rien au problème), respectivement $(x_i)_{i \in I}$ et $(y_j)_{j \in J}$, avec I et J deux ensembles d'indices finis ou dénombrables, inclus dans \mathbb{N} .

On note $A_i = \{X = x_i\}$ et $B_j = \{Y = y_j\}$, pour tous $i \in I, j \in J$.

On peut supposer que pour tout $j \in J$ $\mathbb{P}(B_j) > 0$, quitte à supprimer les valeurs non prises par Y .

Fixons d'abord $j \in J$. On peut alors définir pour tout $i \in I$, la probabilité conditionnelle $\mathbb{P}(A_i|B_j)$. En remarquant que X est une variable aléatoire prenant la valeur i avec la probabilité $\mathbb{P}(X = i) = \mathbb{P}(A_i)$, si bien que $\mathbb{E}[X] = \sum_{i \in I} i\mathbb{P}(A_i)$, on pose (par envie de linéarité) que

$$\mathbb{E}[X|B_j] = \sum_{i \in I} i\mathbb{P}(A_i|B_j).$$

Maintenant, on définit la variable aléatoire $\mathbb{E}[X|Y]$ par sa valeur $\mathbb{E}[X|B_j]$ sur l'ensemble B_j :

$$\mathbb{E}[X|Y] = \begin{cases} \mathbb{E}[X|B_1] & \text{si } Y = y_1 \\ \vdots \\ \mathbb{E}[X|B_j] & \text{si } Y = y_j \\ \vdots \end{cases}.$$

On note deux propriétés simples donnant des possibilités de calcul:

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

2.2 Processus stochastiques

2.2.1 Définitions

On introduit une dimension *dynamique*, par rapport à la situation *statique* de la section précédente (où on considère chaque variable aléatoire indépendamment des autres). On fait en fait dépendre les variables aléatoires d'une variable supplémentaire, interprétée comme une variable *temporelle*.

Définition 2.10 Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace de probabilités.

- Un processus stochastique à temps discret est une famille $\mathbb{X} = (X_n)_{n \in \mathbb{N}}$ de variables aléatoires. Dans ce cas, on note $\mathbb{T} = \mathbb{N}$.
- Un processus stochastique à temps continu est une famille $\mathbb{X} = (X_t)_{t \in \mathbb{R}^+}$ de variables aléatoires. Dans ce cas, on note $\mathbb{T} = \mathbb{R}^+$.

Toutes les variables aléatoires sont définies sur le même espace de probabilités; ce sont des variables aléatoires réelles (par défaut, dans certains cas ce sont même des va discrètes, voir le Chapitre sur les chaînes de Markov).

Plus généralement, l'espace d'états E (noté parfois S) est \mathbb{R} dans le cas de va réelles ou \mathbb{N} dans le cas de va discrètes; il est muni de la tribu canonique \mathcal{E} , respectivement $\mathcal{B}(\mathbb{R})$ ou $\mathcal{P}(\mathbb{N})$.

On rappelle que \mathbb{R}^+ est l'ensemble des nombres réels positifs (ou nuls). Dans la suite, en toute généralité on notera le processus $\mathbb{X} = (X_t)_{t \in \mathbb{T}}$, si on n'a pas à distinguer entre temps discret ou temps continu.

Remarque 2.11 Dans les deux cas de la définition précédente, l'ensemble \mathbb{T} est muni d'une relation d'ordre \leq , qui de plus est totale: étant donnés $(s, t) \in \mathbb{T}^2$, on a $s \leq t$ ou $t \leq s$.

On peut aussi considérer des processus sur un horizon de temps fini: dans le cas discret, on considère $\mathbb{T} = \{0, \dots, N\}$, pour un certain instant final N ; dans le cas continu, on pose $\mathbb{T} = [0, T]$.

Exemple 2.12 *Les exemples considérés dans le chapitre introductif sont des processus stochastiques.*

Une notion importante est celle de trajectoire, à mettre en parallèle avec celle de *réalisation* pour une variable aléatoire.

Définition 2.13 *À éventualité $\omega \in \Omega$ fixée, la trajectoire (ou ω -trajectoire) du processus \mathbb{X} est la courbe suivante:*

$$t \in \mathbb{T} \mapsto X_t \in \mathbb{R}.$$

Remarque 2.14 *C'est pour cette notion qu'il est important de définir les variables aléatoires du processus sur un seul espace de probabilités fixé - et c'est toujours possible, on n'a pas besoin de le mentionner.*

Pour comprendre la dynamique, il est important (et on le fera souvent) de tracer des exemples de trajectoires - même si on ne peut pas toutes les dessiner, ça donnera une bonne idée du comportement typique du processus.

2.2.2 Filtrations, temps d'arrêt

Les définitions rigoureuses et générales de ces notions (notamment dans le cas des processus à temps continu) sont totalement **hors programme** - même si elles sont données. Cependant, elles permettent de formaliser le phénomène d'augmentation de l'information disponible à propos du processus au cours du temps, ainsi que les conditions à vérifier pour pouvoir stopper l'évolution d'un processus.

Considérons l'exemple de la ruine du joueur. On souhaite introduire un critère afin de s'arrêter de jouer avant d'avoir tout perdu. Par exemple, si on a une fortune initiale égale à a , on pourrait décider de s'arrêter soit quand on a touché le jackpot, soit, si on sait qu'on va perdre à la fin, de s'arrêter au dernier instant où la fortune est égale à a - on a joué un bon moment, mais on s'arrête juste avant qu'il ne soit trop tard.

Mais, à un instant donné quelconque, on ne peut pas déterminer que cette situation se produit, en connaissant uniquement ce qui s'est produit jusqu'à cet instant, il faudrait connaître le futur!

Le cas à temps discret

On commence par ce cas, pour lequel on peut décrire précisément la filtration.

On a donc $\mathbb{T} = \mathbb{N}$ (ou un sous-intervalle fini); chaque X_n est une variable aléatoire à valeurs dans E , muni de sa tribu \mathcal{E} . E est \mathbb{R} ou \mathbb{N} , et il suffit de tester les ensembles $(-\infty, a)$ et $\{n\}$ respectivement.

Pour tout $n \in \mathbb{N}$, on définit

$$\mathcal{F}_n = \{X_0^{-1}(A_0) \cap \dots \cap X_n^{-1}(A_n); (A_0, \dots, A_n) \in \mathcal{E}^{n+1}\}.$$

C'est l'ensemble de tous les événements qui sont définies à partir des valeurs possibles des variables aléatoires X_0, \dots, X_n .

Quelques remarques: en fait pour chaque $n \in \mathbb{N}$, \mathcal{F}_n est une tribu, et la suite des $(\mathcal{F}_n)_{n \in \mathbb{N}}$ est croissante pour l'inclusion.

De plus, si $k \leq n$, X_k est une variable aléatoire \mathcal{F}_n -mesurable: pour tout $A \in \mathcal{E}$, $X_k^{-1}(A) \in \mathcal{F}_n$. En fait, \mathcal{F}_n est la plus petite tribu qui vérifie cette propriété.

À retenir: **\mathcal{F}_n contient les événements possibles dans le passé, jusqu'à l'instant n compris. Quand n augmente, il y en a de plus en plus.**

Introduisons la notion de temps d'arrêt: il s'agit d'une variable aléatoire à valeurs dans $\mathbb{T} \cup \{+\infty\}$, telle que chaque événement $\{T = n\}$ ne doit dépendre que des événements possibles dans le passé jusqu'à l'instant n : c'est-à-dire que pour tout $n \in \mathbb{N}$, on a $\{T = n\} \in \mathcal{F}_n$. On dit que $T < +\infty$ presque sûrement si $\mathbb{P}(T = +\infty) = 0$, ou de façon équivalente $\sum_{n=0}^{+\infty} \mathbb{P}(T = n) = 1$.

On peut remarquer qu'une définition équivalente d'un temps d'arrêt est la suivante: pour tout $n \in \mathbb{N}$, $\{T \leq n\} \in \mathcal{F}_n$: c'est l'événement "l'arrêt se produit avant l'instant n " qui ne dépend que du passé. C'est cette propriété qui se généralise au cas continu.

L'exemple fondamental est le suivant: soit X un processus à temps discret et soit A un ensemble (avec $A \in \mathcal{E}$); alors

$$T := \inf \{n \in \mathbb{N}; X_n \in A\}$$

est un temps d'arrêt, c'est le temps d'atteinte de l'ensemble A .

Un autre exemple est le temps de retour en la position initiale:

$$T := \inf \{n \in \mathbb{N}^*; X_n = X_0\}.$$

Le cas à temps continu

Cette section est hors programme, elle est juste donnée pour présenter comment le cas à temps discret se généralise.

2.2.3 Exemple: variables aléatoires indépendantes

On présente ici un exemple classique et fondamental, qu'on interprète ici comme un processus.

Définition 2.15 • Soit $n \in \mathbb{N}$, et X_0, \dots, X_n des variables aléatoires à valeurs dans l'espace d'états $S = \mathbb{R}$ ou \mathbb{N} . On dit qu'elles sont indépendantes si pour tout $(A_0, \dots, A_n) \in \mathcal{E}^{n+1}$ on a

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

- Une famille de variables aléatoires $(X_t)_{t \in \mathbb{T}}$ 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 $(t_0, \dots, t_n) \in \mathbb{T}^{n+1}$, distincts deux à deux, les variables aléatoires X_{t_0}, \dots, X_{t_n} sont indépendantes.

Rappelons que la notation $\mathbb{P}(X_0 \in A_0, \dots, X_n \in A_n)$ correspond à: $\mathbb{P}(\{X_0 \in A_0\} \cap \dots \cap \{X_n \in A_n\})$.

Rappelons la notion d'indépendance 2 à 2 (pour une famille finie ou infinie indexée par \mathbb{T}): pour tous s, t distincts, X_t et X_s sont indépendantes. Et que l'indépendance 2 à 2 n'entraîne pas l'indépendance, comme le démontre le contre-exemple suivant!

Exemple 2.16 On considère les lancers respectifs de deux dés à 6 faces, considérés indépendants. 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 2.17 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 $S \rightarrow \mathbb{R}$ ($S = \mathbb{R}$ ou \mathbb{N}) (mesurables) 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.

Remarque 2.18 La bonne notion d'indépendance est en fait définie mathématiquement en termes de tribus plutôt que de variables aléatoires...

L'exercice 3 ci-dessous donne un exemple d'étude d'un processus à temps discret constitué de variables aléatoires indépendantes. Un autre exemple intéressant est donné par des processus définis comme sommes de variables indépendantes, comme la marche aléatoire (voir aussi la section sur les processus à accroissements indépendants dans le Chapitre sur les processus gaussiens). Dans le chapitre sur les chaînes de Markov, l'indépendance aura aussi un grand rôle.

2.2.4 Quelques quantités importantes

Soit $\mathbb{X} = (X_t)_{t \in \mathbb{T}}$ un processus stochastique.

- Fonctionnelle d'espérance: (sous réserve d'intégrabilité) c'est la fonction

$$\begin{aligned} m : \mathbb{T} &\rightarrow \mathbb{R} \\ t &\mapsto m(t) := \mathbb{E}X_t. \end{aligned}$$

- Fonctionnelle de covariance: (si de carré intégrable)

$$\begin{aligned} c : \mathbb{T}^2 &\rightarrow \mathbb{R} \\ (s, t) &\mapsto c(s, t) = \text{Cov}(X_s, X_t). \end{aligned}$$

- Autres possibilités: fonctionnelle de variance, de corrélation...

On rappelle que la corrélation entre deux variables aléatoires X et Y non constantes (presque sûrement) est définie par

$$\text{corr}(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)}\sqrt{\text{Var}(Y)}}.$$

C'est un réel appartenant à l'intervalle $[-1, 1]$.

- Fonction de répartition uni-dimensionnelle: pour tout $t \in \mathbb{T}$, c'est

$$\begin{aligned} F_t : \mathbb{R} &\rightarrow \mathbb{R} \\ x &\mapsto F_t(x) := \mathbb{P}(X_t \leq x). \end{aligned}$$

- Fonction de répartition bi-dimensionnelle: pour tout $(s, t) \in \mathbb{T}^2$, c'est

$$\begin{aligned} F_{s,t} : \mathbb{R}^2 &\rightarrow \mathbb{R} \\ (x, y) &\mapsto F_{s,t}(x, y) := \mathbb{P}(X_s \leq x, X_t \leq y). \end{aligned}$$

Remarque 2.19 *En fait la loi d'un processus stochastique (vu comme variable aléatoire dans l'espace produit $S^{\mathbb{T}}$) est caractérisée par ses lois fini-dimensionnelles, ou de façon équivalente par les fonctions de répartition F_{t_1, \dots, t_n} , n -dimensionnelles, pour tout entier n et pour tout choix de n -uplet (t_1, \dots, t_n) , avec $F_{t_1, \dots, t_n}(x_1, \dots, x_n) = \mathbb{P}(X_{t_1} \leq x_1, \dots, X_{t_n} \leq x_n)$.*

2.2.5 Exercices

Exercice 3 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 4 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 3

Chaînes de Markov

3.1 Définition, exemples

Espace d'états: ensemble S fini ou dénombrable.

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 plus attention, et s'adapter!

Définition 3.1 Une chaîne de Markov à valeurs dans S est un **processus stochastique à temps discret** noté $(X_n)_{n \in \mathbb{N}}$, tel que:

- pour tout entier $n \in \mathbb{N}$, X_n est une variable aléatoire dans S ;
- pour tout $n \in \mathbb{N}$, pour tout $(n+1)$ -uplet à valeurs dans S , (x_0, \dots, x_n) , vérifiant $\mathbb{P}(X_0 = x_0, \dots, X_{n+1} = x_{n+1}) \neq 0$, on a

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

- de plus, pour tout $1 \leq k \leq n$, on a

$$\mathbb{P}(X_k = x_k | X_{k-1} = x_{k-1}) = p(x_{k-1}, x_k),$$

la fonction p étant indépendante de k .

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

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.

Remarque 3.2 La troisième condition dans la définition précédente est une hypothèse d'homogénéité en temps; dans ce cours, elle sera toujours satisfaite, même si certains points de la théorie restent valables dans le cas non-homogène.

Proposition 3.3 *On dispose d'une formulation équivalente du deuxième point de la définition d'une chaîne de Markov:*

pour tout $n \in \mathbb{N}$, pour tout $(n+1)$ -uplet à valeurs dans S , (x_0, \dots, x_n) , vérifiant $\mathbb{P}(X_0 = x_0, \dots, X_n = x_n) \neq 0$, on a

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

Exercice 5 (Théorique) *Prouver l'équivalence annoncée dans la proposition ci-dessus.*

Interprétation: un processus est une chaîne de Markov si la transition d'un état x_{n-1} à l'instant $n-1$ à un autre x_n à un instant n ne dépend que de:

- la position x_{n-1} à l'instant $n-1$, et pas du reste de la trajectoire passée (c'est-à-dire de x_0, \dots, x_{n-2});
- le résultat d'une expérience aléatoire, distribuée selon la probabilité conditionnelle sachant $X_{n-1} = x_{n-1}$ (notée $\mathbb{P}(\cdot | X_{n-1} = x_{n-1})$).

Exemple 3.4 *Une suite de variables aléatoires iid.*

Exemple 3.5 *Une marche aléatoire simple.*

Exercice 6 *Vérifier que les deux exemples précédents conviennent.*

3.2 Noyau de transition

Définition 3.6 *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: $\sum_{y \in S} P(x, y) = 1$.

Lien avec les chaînes de Markov: la fonction donnant les probabilités conditionnelles $\mathbb{P}(X_k = x_k | X_{k-1} = x_{k-1})$ 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 3.7 *On dit que $(X_n)_{n \in \mathbb{N}}$ est une chaîne de Markov de noyau P et de loi initiale μ si:*

- c'est une chaîne de Markov au sens de la définition précédente;
- la variable aléatoire X_0 suit la loi μ , c'est-à-dire $\mathbb{P}(X_0 = x_0) = \mu(x_0)$, pour tout $x_0 \in S$;
- les probabilités conditionnelles sont données par la fonction $p = P$.

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 7 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).

3.3 Graphe de transition

On représente très souvent une chaîne de Markov à l'aide d'un graphe 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 (x, y) tels que $P(x, y) > 0$.

De plus, sur les arêtes orientées, on indique la probabilité de transition correspondante (non nulle).

Exercice 8 Représenter le graphe dans le cas de la marche aléatoire sur \mathbb{Z} , dans le cas de la ruine du joueur.

3.4 Distribution

Soit $(X_n)_{n \in \mathbb{N}}$ une chaîne de Markov, de noyau de transition P , sur un espace d'états S , et de loi initiale μ : pour tout état $x_0 \in S$ on a $\mathbb{P}(X_0 = x_0) = \mu(x_0)$.

Par la formule des probabilités totales, puis en utilisant la propriété de Markov (c'est-à-dire la définition de chaîne de Markov): pour tout entier $n \geq 1$ et tout état $x_n \in S$,

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

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 9 *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 3.8 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 3.9 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 10 Démontrer le théorème à l'aide des formules données plus haut.

Bonus:

Théorème 3.10 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 11 Démonstration?

Un dernier résultat:

Théorème 3.11 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 3.12 (Equation de Chapman-Kolmogorov) Pour tous entiers $n, m \in \mathbb{N}$ et tous états $x, y \in S$,

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

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

3.5 Lois invariantes

On commence par une définition générale:

Définition 3.13 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 (et c'est ce qu'il faut retenir):

Théorème 3.14 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 $\mu_n = \mu_0$.

Définition 3.15 Une loi initiale telle que la chaîne dont elle est issue est stationnaire est appelée loi invariante. En d'autres termes, une loi μ est 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 a priori difficiles, 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):

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

Exercice 12 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.

La situation de l'exercice ci-dessus se généralise en la notion de loi réversible (**Hors programme**); 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 13 Montrer cette propriété.

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

Définition 3.16 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.

3.6 Chaînes de Markov et temps d'arrêt

Voici une explication de l'intérêt de la notion de temps d'arrêt, dans le cadre des chaînes de Markov.

Pour commencer, donnons une nouvelle caractérisation de la propriété de Markov.

Soit $(X_n)_{n \in \mathbb{N}}$ un processus à valeurs dans l'espace d'états S .

Il s'agit d'une chaîne de Markov si **conditionnellement au présent, futur et passé sont indépendants**: quels que soient les entiers $n, m \geq 1$, et les états x_0, \dots, x_{n+m} ,

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

Ici:

présent = instant n ;

passé = instants $0, \dots, n-1$;

futur = instants $n+1, \dots$

Proposition 3.17 Dans ce cas, conditionnellement aux événements $\{X_n = x_n\}$, le processus $(X_{n+m})_{m \in \mathbb{N}}$ est une chaîne de Markov:

- de position initiale x_n ;
- de noyau de transition P (ne dépendant pas de n ; c'est le même noyau que la chaîne initiale);
- de plus, indépendante de (X_0, \dots, X_{n-1}) .

On peut en fait remplacer l'entier n par un temps aléatoire T , à deux conditions:

- on conditionne par rapport aux événements $\{T < +\infty\} \cap \{X_T = x\}$;

- T doit être un temps d'arrêt.

En fait, on peut décomposer selon les valeurs finies n prises par la variable T , et appliquer la proposition précédente.

C'est un peu compliqué. Ce qu'il faut retenir, c'est plutôt ce qui suit (et c'est important!). Fixons un état $a \in S$.

On pose la position initiale $X_0 = a$, si bien qu'on regarde des trajectoires de la chaîne de Markov toutes issues de a .

On pose pour tout entier $k \in \mathbb{N}$

$$\begin{aligned} T_0(a) &= 0; \\ T_1(a) &= \inf \{n \leq T_0(a) + 1; X_n = a\}; \\ &\vdots \\ T_{k+1}(a) &= \inf \{n \leq T_k(a) + 1; X_n = a\}; \\ &\vdots \end{aligned}$$

avec la convention suivante: si pour un entier k on a $\{n \leq T_k(a) + 1; X_n = a\} = \emptyset$, alors on pose $T_l(a) = +\infty$ pour tout $l \geq k + 1$: si à partir d'un moment la chaîne ne revient pas en a , on décide que le temps de retour est $+\infty$.

La suite aléatoire $(T_k(a))_{k \in \mathbb{N}}$ décrit les temps de retour successifs de la chaîne à l'état a . Ce sont tous des temps d'arrêts.

La propriété fondamentale est la suivante:

$$(X_{T_0(a)} = a, \dots, X_{T_1(a)-1}), (X_{T_1(a)} = a, \dots, X_{T_2(a)-1}), \dots, (X_{T_k(a)} = a, \dots, X_{T_{k+1}(a)-1}), \dots$$

sont des morceaux de trajectoires **indépendants** et identiquement distribués (la longueur aléatoire fait partie de la loi), avec la propriété que chaque début de trajectoire est en a , et que a n'est plus visité ensuite.

En d'autres termes: à chaque retour en a , on calcule une nouvelle trajectoire issue de a , indépendamment de tout le passé, et du numéro du retour (premier, second,... retour), toujours avec la même distribution de probabilité. On s'arrête au retour suivant, et on enlève le a à la fin de la trajectoire.

3.7 Classification des états

3.7.1 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 3.18 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.

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 3.19 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 3.20 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 3.21 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 3.22 La marche aléatoire simple sur \mathbb{Z}^d est irréductible.

Exercice 14 Déterminer les classes de communication du noyau de transition associé au problème de la ruine du joueur.

Deux notions pour terminer:

Définition 3.23 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.

Question: quel critère peut assurer qu'une classe de communication est fermée?

3.7.2 Récurrence-transience

Définition 3.24 Un état $x \in S$ est dit récurrent (pour le noyau P) si

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

Dans le cas contraire, x est dit transient.

On définit deux quantités importantes: conditionnellement à $X_0 = x$,

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

$T(x)$ 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:

$$x \text{ récurrent} \Leftrightarrow \mathbb{P}(N(x) = \infty \mid X_0 = x) = 1.$$

Mieux on a le théorème suivant:

Théorème 3.25 • *Caractérisation de la récurrence:*

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

• *Caractérisation de la transience:*

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

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

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

Remarque 3.26 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), ce qui se fait avec probabilité $\mathbb{P}(T(x) < +\infty | X_0 = x)$, puis ne pas revenir.

Proposition 3.27 *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 3.28 (Culture) *La marche aléatoire simple sur \mathbb{Z}^d est récurrente si $d = 1$ ou $d = 2$, et est transiente si $d \geq 3$. En fait, on a un équivalent*

$$P^n(0, 0) \sim \frac{C_d}{n^{d/2}}.$$

On peut maintenant répondre à la question posée à la fin de la section concernant les classes de communication.

Théorème 3.29 *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.

Remarque 3.30 FAIRE DES DESSINS.

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 15 *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.*

3.7.3 Espérance du temps de retour

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 3.31 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 $\mathbb{E}[T(x) | X_0 = x] = \infty$.

Proposition 3.32 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.

3.7.4 Périodicité

Définition 3.33 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 3.34 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 3.35 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 3.36 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.

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

Soit P un noyau de transition sur l'espace d'états S .

Soit a un état absorbant.

Etant donnée une condition initiale $x_0 \in S$, on veut calculer

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

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

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 $x \in S$

$$\begin{aligned} p(x) &= \mathbb{P}(\{\exists n \in \mathbb{N}; X_n = a\} \mid X_0 = x) \in [0, 1], \\ t(x) &= \mathbb{E}[\inf \{\exists n \in \mathbb{N}; X_n = a\} \mid X_0 = x] \in [0, +\infty]. \end{aligned}$$

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 16 Reprendre la situation de l'exercice 4: 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.

3.9 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 3.37 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 3.38 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 17 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 3.39 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 3.40 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 particulier important:

Théorème 3.41 Si l'espace d'états S est **fini**, une chaîne irréductible est nécessairement récurrente positive.

Remarque 3.42 Dans le cas non-irréductible, pour comprendre la dynamique (en temps long), on utilise le théorème de Décomposition de l'espace d'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 3.43 *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'une 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 $\nu = \mu$.

3.10 Complément: Chaînes de Markov non homogènes

Dans la définition de chaîne de Markov, on a demandé que pour tout $1 \leq k \leq n$

$$\mathbb{P}(X_k = x_k | X_{k-1} = x_{k-1}) = p(x_{k-1}, x_k),$$

la fonction p étant indépendante de k .

On peut ôter cette hypothèse, et ainsi considérer des chaînes de Markov non-homogènes:

Définition 3.44 Une chaîne de Markov non-homogène à valeurs dans S est un **processus stochastique à temps discret** noté $(X_n)_{n \in \mathbb{N}}$, tel que:

- pour tout entier $n \in \mathbb{N}$, X_n est une variable aléatoire dans S ;
- pour tout $n \in \mathbb{N}$, pour tout $(n+1)$ -uplet à valeurs dans S , (x_0, \dots, x_n) , vérifiant $\mathbb{P}(X_0 = x_0, \dots, X_{n-1} = x_{n-1}) \neq 0$, on a

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

- de plus, pour tout $1 \leq k \leq n$, on a

$$\mathbb{P}(X_k = x_k | X_{k-1} = x_{k-1}) = P_k(x_{k-1}, x_k),$$

P_k étant un noyau de transition sur S .

A chaque instant, on a ainsi une matrice de transition différente (mais indépendante de la loi initiale).

Chapitre 4

Processus Gaussiens

4.1 Loi normale

C'est une loi fondamentale en probabilités, liée au fameux Théorème Central Limite. On parle indifféremment de loi normale, de loi gaussienne, de loi de Laplace-Gauss...

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 la fonction suivante: pour tout $x \in \mathbb{R}$:

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

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 $x \in \mathbb{R}$:

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

Remarque 4.2 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 4.3 (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 par: pour tout $t \in \mathbb{R}$

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

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 $a > 0$:

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

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 18 *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 4.4 *Soit $m \in \mathbb{R}$ et $\sigma > 0$.*

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).

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

On a le résultat suivant:

Théorème 4.5 *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 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 par: pour tout $t \in \mathbb{R}$,

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

Exercice 19 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 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 4.6 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 $X = X_i$, 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_1, \dots, X_n) , à valeurs dans \mathbb{R}^n , chaque X_i étant une variable aléatoire réelle.

Définition 4.7 *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.*

Remarque 4.8 *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.

Cette condition nécessaire n'est pas suffisante!

Exercice 20 *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 4.9 *Si X_1, \dots, X_n sont des variables aléatoires (réelles) suivant une loi normale et **indépendantes**, alors (X_1, \dots, X_n) est un vecteur gaussien.*

Pour prouver qu'un vecteur aléatoire est un vecteur gaussien, on dispose du critère suivant:

Théorème 4.10 *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 21 *Démontrer ce théorème (par exemple en explicitant composante par composante l'hypothèse $X = F(Y)$).*

Ce théorème est essentiel dans la pratique: dans l'énoncé on donnera un premier vecteur gaussien, et par transformations linéaires on pourra en obtenir des nouveaux.

Pour terminer, on donne les définitions suivantes:

Définition 4.11 Soit $X = (X_1, \dots, X_n)$ un vecteur aléatoire gaussien. On définit:
son espérance $m = \mathbb{E}[X]$: c'est le vecteur $(m_1, \dots, m_n) := (\mathbb{E}[X_1], \dots, \mathbb{E}[X_n])$;
sa matrice de covariance $C = \text{Cov}(X)$: c'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 22 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 4.12 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 4.13 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; on note cette loi $\mathcal{N}(m, C)$, 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 4.14 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 $\mathbb{T} = \mathbb{R}^+$ (ou un sous-intervalle $\{0, \dots, N\}$, respectivement $[0, T]$), et $X = (X_t)_{t \in \mathbb{T}}$ un processus aléatoire.

Définition 4.15 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 4.16 La fonction d'espérance d'un processus gaussien $(X_t)_{t \in \mathbb{T}}$ est la fonction $m : \mathbb{T} \rightarrow \mathbb{R}$, définie par

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

Définition 4.17 La matrice 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 4.18 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 fini-dimensionnelles: 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 données 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, dont la fonctionnelle de covariance est donnée par $C(t, s) = \begin{cases} 1 & \text{si } t = s \\ 0 & \text{si } t \neq s \end{cases}$. 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 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. On note qu'à chaque instant t , la loi de B_t est une gaussienne de variance t ; comme on va le voir maintenant, ce ne sont pas les composantes du processus qui sont indépendantes, mais leurs *accroissements*.

4.4 Indépendance des accroissements

Les définitions de cette section et de la suivante sont valables en fait pour n'importe quelle sorte de processus, l'hypothèse gaussienne n'étant exploitée qu'ensuite. On les retrouvera dans le chapitre suivant.

Définition 4.19 Soit $\mathbb{X} = (X_t)_{t \in \mathbb{T}}$ 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 4.20 La marche aléatoire simple est un P.A.I. à temps discret.

4.5 Stationnarité des accroissements

Définition 4.21 Soit $\mathbb{X} = (X_t)_{t \in \mathbb{T}}$ un processus. On dit qu'il est à accroissements stationnaires (noté P.A.S.) si pour tous $s \leq t$ deux instants dans \mathbb{T} , la loi de l'accroissement $X_t - X_s$ ne dépend que de $t - s$: en particulier c'est la loi de $X_{t-s} - X_0$.

Exemple 4.22 La marche aléatoire simple est un P.A.S. à temps discret.

Définition 4.23 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.

4.6 Mouvement Brownien

Définition 4.24 *Le mouvement Brownien $(B_t)_{t \in \mathbb{R}^+}$ est un P.A.I.S. Gaussien, caractérisé par le fait que pour tout $t \in \mathbb{R}^+$, B_t suit la loi gaussienne $\mathcal{N}(0, t)$. En particulier, $B_0 = 0$.*

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 4.25 *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.

Preuve 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.

4.7 Quelques exercices autour du mouvement brownien

Exercice 23 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 24 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 25 Soit $B = (B_t)_{t \geq 0}$ un mouvement brownien.

Définissons le processus \bar{B} sur $[0, 1]$, de telle sorte que $\bar{B}_t = B_t - tB_1$.

1. Que vaut \overline{B}_0 ? Que vaut \overline{B}_1 ? Justifier la terminologie de "pont brownien" pour ce processus.
2. Montrer qu'il s'agit d'un processus gaussien, déterminer ses caractéristiques.

Exercice 26 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 $[.]$ 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!

Chapitre 5

Processus de Poisson

5.1 La loi de Poisson

On introduit une nouvelle loi de probabilité sur \mathbb{N} , dépendant d'un paramètre $\lambda > 0$, appelée loi de Poisson et notée $\mathcal{P}(\lambda)$: une variable aléatoire discrète suit cette loi si on a pour tout $n \in \mathbb{N}$

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

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 27 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 28 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 de l'exercice 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.

5.2 La loi exponentielle

Elle a été définie lors des rappels de probabilités: une variable aléatoire réelle suit la loi exponentielle de paramètre $\lambda > 0$, notée $\mathcal{E}(\lambda)$, si pour tout $t > 0$, on a

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

En particulier, on a $\mathbb{P}(X \leq 0) = 0$: X ne prend que des valeurs positives.

Exercice 29 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.

5.3 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 $\mathbb{T} = \mathbb{R}^+$, 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 5.1 *Presque sûrement on a en fait un ordre strict: $0 = T_0 < T_1 < \dots < T_n < \dots$; de plus on a $T_n \rightarrow +\infty$, presque sûrement, quand $n \rightarrow +\infty$.*

On notera $T_0 = 0$ l'instant initial.

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 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 $N = (N_t)_{t \in \mathbb{N}}$ 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 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.

5.4 Processus de Poisson

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

Hypothèse 5.2 *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 5.3 *Le processus de comptage $(N_t)_{t \in \mathbb{R}^+}$ est un P.A.I.S: pour tout n -uplet (n quelconque) $t_1 < t_2 < \dots < t_n$, les accroissements $N_{t_1}(-N_0), N_{t_2} - N_{t_1}, \dots, N_{t_n} - N_{t_{n-1}}$ sont indépendants et suivent pour lois respectives, celle de $N_{t_1(-0)}, N_{t_2-t_1}, \dots, N_{t_n-t_{n-1}}$.*

On a alors le résultat suivant (admis):

Théorème 5.4 *On suppose l'hypothèse 5.1 vérifiée.*

Si l'hypothèse 5.2 est satisfaite, alors le processus de comptage associé est un P.A.I.S.

Réciproquement, si l'hypothèse 5.3 est satisfaite, alors il existe $\lambda > 0$ tel que l'hypothèse 5.2 est vérifiée.

Dans ce cas, pour tout $t \in \mathbb{R}^+$, la variable aléatoire N_t suit une loi de Poisson de paramètre λt .

On parle de processus de Poisson (homogène) d'intensité λ .

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

Remarque 5.5 *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 5.6 *tirer un nombre aléatoirement selon une loi donnée=générer (par un certain mécanisme) une réalisation d'un variable aléatoire, de la loi en question.*

5.5 Quelques propriétés du processus de Poisson

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.

5.6 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 ça 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 5.2, 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.

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 $p\lambda$.

5.7 Processus de Poisson composé

Dans la 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.

5.8 Processus Markovien de saut

Dans les deux raffinements 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".

Un cas particulier est celui des processus de naissance et de mort: à valeurs dans \mathbb{N} , dont tous les sauts sont $+1$ ou -1 .

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.

5.8.1 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).

5.8.2 Description "générateur"

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

$$Q(x, y) \geq 0 \text{ pour tout } y \in S, y \neq x;$$

$$Q(x, x) = - \sum_{y \in S, y \neq x} Q(x, y) \leq 0.$$

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 $\frac{Q(x, y)}{-Q(x, x)}$.

5.8.3 Lien entre les deux descriptions

Les deux descriptions précédentes sont équivalentes, avec les relations:

$$q(x) = -Q(x, x),$$

$$R(x, y) = \frac{Q(x, y)}{q(x)} \text{ pour tout } y \in S, y \neq x, \text{ si } q(x) \neq 0,$$

$$R(x, y) = 0 \text{ pour tout } y \in S, y \neq x, \text{ si } q(x) = 0.$$

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

5.8.4 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 $\mu Q = 0$.

5.9 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.

5.9.1 $M/M/0$

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.

5.9.2 $M/M/1$

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 $\frac{\rho}{1-\rho}$.

5.9.3 $M/M/s$

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

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.

5.9.4 $M/M/\infty$

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 $\rho := \frac{\lambda}{\mu}$.

5.9.5 $M/M/1/k$

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

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

5.9.6 $M/M/s/0$

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

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