

COURS DE PROBABILITÉS (SEMESTRE 6) - MTH333

KOSSI ESSONA GNEYOU
Professeur à l'Université de Lomé

Faculté des Sciences - Département de Mathématiques. E-mail: kgneyou@gmail.com

Avril 2020

Introduction

Le calcul des probabilités semble avoir son origine dans les jeux de hasard. Il fallait savoir décider des mises et des gains pour que les jeux soient équitables ou non. Aujourd'hui, il a investi de très nombreuses activités humaines et son utilisation est nécessaire dans la plupart des activités socio-économiques et industrielles. Il est utile par exemple de savoir si la construction d'un appareil est raisonnable, s'il a une chance de fonctionner ou si l'assemblage d'un très grand nombre de composants réduira son espérance de vie de manière inacceptable (navette spatiale par exemple).

Introduction

Le calcul des probabilités et les statistiques sont utilisées pour augmenter, souvent de manière importante, la qualité des produits manufacturés, les coûts de production s'en trouvent réduits car le risque de panne est diminué.

Un autre domaine du calcul des probabilités et des statistiques est celui de la compression des données. Ici, il faut utiliser les structures particulières des données à stocker pour en minimiser la taille. Mais ces structures ne sont pas parfaitement établies, elles varient statistiquement, c'est ce qui rend difficile le travail de compression. Les travaux de Shannon et de Kolmogorov sur la théorie de l'information ont ouvert de nouveaux domaines de recherche en probabilité et statistique.

Introduction

Ce cours est organisé en plusieurs sections progressives. Il sera complété par des TD et des exercices à chercher par les étudiants en vue du renforcement de leurs capacité à se servir des modèles probabilistes et statistiques. Les théorèmes, propositions lemmes et définitions ont une numérotation dépendant de la section ou de la sous section dans lesquelles ils apparaissent.

Comme recommandation, les étudiants doivent chercher à comprendre le cours, reprendre les exemples et/ou exercices traités ou proposés en cours. Mais cela ne suffit pas. Il faudra aussi chercher à renforcer ses capacités par des exercices et problèmes d'entraînement que l'on trouvera facilement dans le net sous forme de documents numériques gratuits ou à la bibliothèque universitaire.

Sommaire

1 Rappels de dénombrement

- Les cardinaux remarquables

2 Rappels de probabilités

Les ensembles et sous-ensembles

1- Soit E un ensemble à p éléments et F un ensemble à n éléments. Alors

$$E \times F = \{(x, y) \mid x \in E \text{ et } y \in F\}$$

et

$$\text{Card}(E \times F) = p \times n = np.$$

2-

$$F^p = \overbrace{F \times F \times \cdots \times F}^{p \text{ facteurs}} = \{(x_1, x_2, \dots, x_p) \mid x_1 \in F, x_2 \in F, \dots, x_p \in F\}$$

Alors

$$\text{Card } F^p = n \times n \times \cdots \times n = n^p$$

.

Les listes ou suites

Un élément $(x_1, x_2, \dots, x_p)$ de F^p est une suite ordonnée de p éléments de F appelé un p -uplet ou une p -liste d'éléments de F .

C'est aussi une application de l'ensemble E à p éléments vers l'ensemble F à n éléments.

Remarques

- 1) L'ordre est important dans une p -liste.
- 2) Une p -liste peut contenir plusieurs fois un même élément.

Les suites à éléments distincts

Définition 1.1: arrangement

- 1) Une p -liste à éléments tous distincts est appelé un arrangement de p éléments pris parmi les n éléments. C'est aussi une application injective d'un ensemble E à p éléments vers un ensemble F à n éléments.
- 2) Lorsque $p = n$, un arrangement de n éléments pris dans un ensemble à n éléments est une permutation des n éléments. C'est aussi une application bijective de d'un ensemble E à n éléments dans un ensemble F à n éléments.

exemples de listes ou suites

Exemple

- une p -liste de boules est un tirage successif avec remise de p boules dans une urne contenant n boules.
- un arrangement de p boules est un tirage successif sans remise de p boules dans une urne contenant n boules.
- une permutation de n boules est un tirage successif sans remise de n boules dans une urne contenant n boules.

Les parties ou sous-ensemble d'éléments distincts

Définition 1.2: combinaison

Une partie ou sous-ensemble de p éléments distincts d'un ensemble à n éléments est appelé une combinaison de p éléments pris parmi les n éléments.

Notons que contrairement à une p -liste, l'ordre des éléments ne compte pas dans une combinaison et une combinaison ne peut contenir un même élément plus d'une fois.

Un tirage simultané de p boules dans une urne contenant n boules est une combinaison des p boules prises parmi les n boules.

Proposition 1.1

Soit E un ensemble à p éléments et F un ensemble à n éléments.

- ❶ Le nombre de p -listes d'un ensemble à n éléments ou nombre d'applications de E vers F est égal à n^p ,
- ❷ Le nombre de p -listes d'éléments tous distincts ou nombre d'arrangements de p éléments d'un ensemble à n éléments ou encore le nombre d'applications injectives de E vers F est égal à $A_n^p = \frac{n!}{(n-p)!}$,
- ❸ Le nombre de n -listes d'éléments tous distincts ou nombre d'arrangements de n éléments d'un ensemble à n éléments ou encore le nombre d'applications bijectives d'un ensemble E à n éléments vers un ensemble F à n éléments est égal à $A_n^n = n!$,
- ❹ Le nombre de combinaisons de p éléments distincts d'un ensemble à n éléments ou encore le nombre de parties (sous-ensembles) à p éléments pris dans un ensemble à n éléments est égal à $C_n^p = \frac{A_n^p}{p!} = \frac{n!}{p!(n-p)!}$.

Proposition 1.2

Soit E un ensemble à n éléments. On note par $\mathcal{P}(E)$ l'ensemble des parties de E . Alors

$$\text{Card}(\mathcal{P}(E)) = 2^n$$

Démonstration.

$\mathcal{P}(E)$ est l'ensemble formé de l'ensemble vide, des singletons, des paires, des sous-ensembles à 3 éléments, ..., des sous ensembles à p éléments, etc., et de E lui-même. Donc

$$\text{Card}(\mathcal{P}(E)) = C_n^0 + C_n^1 + \cdots + C_n^n = (1 + 1)^n = 2^n$$

On rappelle que $(a + b)^n = \sum_{k=0}^n C_n^k a^k b^{n-k}$.



Proposition 1.3

- 1- Le nombre de parties ou combinaisons de p éléments avec répétitions possible d'un ensemble F à n éléments est égal à Γ_n^p avec $\Gamma_n^p = C_{n+p-1}^p$.
- 2- Le nombre de suites de n entiers naturels dont la somme est p est aussi égal à $\Gamma_n^p = C_{n+p-1}^p$ c'-à-d. si $E = \{(x_1, x_2, \dots, x_n) \in \mathbb{N}^n \mid x_1 + x_2 + \dots + x_n = p\}$ alors $\text{Card}(E) = \Gamma_n^p = C_{n+p-1}^p$.

Démonstration.

Soit $\{x_1, x_2, \dots, x_p\}$ une combinaison avec répétition (CAR) de p éléments de F . Représentons le terme x_i par 0 autant de fois qu'il se répète puis par 1, pour $i = 1, 2, \dots, n$ en convenant de ne pas mettre le dernier 1. Alors Γ_n^p est le nombre de façons de choisir $n - 1$ places de "1" parmi les $n + p - 1$ places total de 0 et de 1, soit $\Gamma_n^p = C_{n+p-1}^{n-1} = C_{n+p-1}^p$. □

Sommaire

1 Rappels de dénombrement

2 Rappels de probabilités

- le vocabulaire
- Langage des événements
- Notion intuitive de probabilité

- Notion axiomatique de la probabilité
- Probabilité Conditionnelle
- Événements indépendants
- Indépendance d'une famille de tribus
- Application : Loi de tout et de rien

Vocabulaire des probabilités

La théorie de la probabilité a un langage qui lui est propre. On appelle :

- *expérience aléatoire* (e.a.) un fait dont le résultat dépend du hasard.

L'ensemble des résultats possibles d'une e.a. est appelé *univers* ou ensemble fondamental noté Ω . On appelle ensuite :

- *Épreuve*, tout essai d'une e.a.

- *évènement*, toute partie (ou sous-ensemble) A de Ω , c'est-à-dire une collection d'éléments ω de Ω dont la réalisation de l'un décrit un fait possible de l'expérience aléatoire. Un évènement réduit à un singleton $\{\omega\}$, $\omega \in \Omega$, est appelé *évènement élémentaire* ou *éventualité*.

Vocabulaire des probabilités

La classe de tous les évènements possibles liés à une e.a. est notée $\mathcal{A}$. On impose à $\mathcal{A}$ d'être contenu dans l'ensemble $\mathcal{P}(\Omega)$ des parties de Ω et d'être une *tribu*, c-à-d. un ensemble contenant Ω , stable par complémentation et par réunion dénombrable.

On appelle espace probabilisable, tout couple $(\Omega, \mathcal{A})$ lié à une e.a. où Ω est l'ensemble des résultats possibles et $\mathcal{A}$ est une tribu sur Ω dite tribu des événements de l'expérience aléatoire.

Vocabulaire des probabilités

On dit qu'un événement A lié à une expérience aléatoire $\mathcal{E}$ est réalisé si le résultat ω obtenu est un élément de A . Il vient :

- . Ω est l'évènement certain (car il est toujours réalisé au cours d'une e.a.)
- . $\emptyset$ est l'évènement impossible (car $\emptyset$ n'est jamais réalisé).
- . À chaque événement A correspond l'évènement contraire de A noté "non A " ou A^c ou encore $\overline{A}$.

Vocabulaire des probabilités

Soient A et B deux événements liés à une même e.a., lorsqu'ils se produisent, on note :

- $A \subset B$ ou $A \Rightarrow B$ l'évènement " A implique B ";
- $A \cap B$ l'évènement " A et B se produisent simultanément". Si $A \cap B = \emptyset$, les événements A et B sont dits incompatibles;
- $A \cup B$ l'évènement " A ou B se produit";
- $A \setminus B$ ou $A \cap \overline{B}$ l'évènement " A se produit et non B ";
- $A \Delta B$ l'évènement "un et un seul des deux A et B se produit".

Rappelons que $A \Delta B = (A \setminus B) \cup (B \setminus A) = (A \cup B) \setminus (A \cap B)$.

Vocabulaire des probabilités

Si (A_n) est une suite d'évènements, on note :

- $\cup A_n$ l'évènement "l'un au moins des A_n se produit" ;
- $\cap A_n$ l'évènement "tous les A_n se produisent simultanément" ;
- $\limsup A_n$ ou $\overline{\lim} A_n$ l'évènement "une infinité des A_n se produit" ;
- $\liminf A_n$ ou $\underline{\lim} A_n$ l'évènement "tous les A_n se produisent sauf peut-être un nombre fini d'entre eux".

Les évènements $\overline{\lim} A_n$ et $\underline{\lim} A_n$ sont définis par :

$$\overline{\lim} A_n = \bigcap_{n \geq 1} \bigcup_{k \geq n} A_k \quad \text{et} \quad \underline{\lim} A_n = \bigcup_{n \geq 1} \bigcap_{k \geq n} A_k.$$

Notion intuitive de probabilité

Une probabilité est la chance a priori qu'un événement aléatoire se produise.

Soit $\mathcal{E}$ une e.a. Tous les évènements liés à $\mathcal{E}$ n'ont pas la même "chance" de se produire.

Soit A un évènement lié à $\mathcal{E}$ c-à-d. $A \in \mathcal{A}$. Répétons n fois l'expérience $\mathcal{E}$ dans les mêmes conditions de façon indépendante. Soit r le nombre de fois que A est réalisé au cours des n épreuves. Alors le nombre $f(A) = \frac{r}{n}$ est appelé fréquence d'apparition de A et on a $f(A) \in [0, 1]$.

Notion intuitive de probabilité

Si B est un autre évènement tels que A et B soient incompatibles et si B s'est produit s fois au cours des n épreuves, $A \cup B$ s'est produit $r + s$ fois (puisque A et B ne se produisent jamais simultanément). On a donc

$$f(A \cup B) = \frac{r + s}{n} = \frac{r}{n} + \frac{s}{n} = f(A) + f(B).$$

L'évènement certain Ω s'est produit n fois (puisque'il se produit à chaque épreuve de $\mathcal{E}$). On a donc $f(\Omega) = \frac{n}{n} = 1$. On a également

$$f(\overline{A}) = \frac{n - r}{n} = 1 - \frac{r}{n} = 1 - f(A)$$

et $f(\emptyset) = 0 = 1 - f(\Omega)$.

Notion intuitive de probabilité

Intuitivement, lorsque n tend vers $+\infty$, $f(A) = \frac{r}{n}$ admet une limite qui ne dépend pas de l'évènement A . Cette limite est appelée probabilité de l'évènement A et est notée $P(A)$.

Remarquons que la fréquence $f(A)$ est une probabilité empirique (évaluée a posteriori sur un échantillon de taille n d'expériences aléatoires $\mathcal{E}$ identiques). C'est donc un point de vue statistique de la probabilité.

Notion axiomatique de la probabilité

Définition 2.1: définition axiomatique de la probabilité

Soit $(\Omega, \mathcal{A})$ un espace probabilisable. On appelle probabilité sur $(\Omega, \mathcal{A})$, toute application P de $\mathcal{A}$ dans $[0, 1]$ possédant les propriétés suivantes :

- ① $P(\Omega) = 1$
- ② $\forall A, B \in \mathcal{A}, A \cap B = \emptyset, P(A \cup B) = P(A) + P(B).$
- ③ Si $A_1, A_2, \dots$ est une suite d'évènements deux à deux incompatibles,

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n).$$

On appelle espace probabilisé, tout triplet $(\Omega, \mathcal{A}, P)$ où $(\Omega, \mathcal{A})$ est un espace probabilisable et P est une probabilité sur $(\Omega, \mathcal{A})$.

Proposition 2.1

Soit $(\Omega, \mathcal{A}, P)$ un espace probabilisé, A et B deux évènements (i.e. $A, B \in \mathcal{A}$).

On a alors

- 1) $P(\emptyset) = 0$
- 2) $P(\overline{A}) = 1 - P(A)$
- 3) Si $A \subset B$, $P(A) \leq P(B)$
- 4) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$
- 5) Si (A_n) est une suite dans $\mathcal{A}$, $P(\bigcup_{n \geq 1} A_n) \leq \sum_{n=1}^{\infty} P(A_n)$.

Démonstration.

1) évident. 2) On écrit $\Omega = A \cup \bar{A}$ donc $1 = P(A) + P(\bar{A})$ (car $A \cap \bar{A} = \emptyset$).

3) On écrit $B = (B \setminus A) \cup A$ d'où $P(B) = P(B \setminus A) + P(A) \geq P(A)$.

4) On écrit $A = (A \setminus B) \cup (A \cap B)$, $B = (B \setminus A) \cup (A \cap B)$ d'où

$$\begin{aligned} P(A) + P(B) &= P(A \setminus B) + P(A \cap B) + P(B \setminus A) + P(A \cap B) \\ &= P\left((A \setminus B) \cup (B \setminus A) \cup (A \cap B)\right) + P(A \cap B) \\ &= P(A \cup B) + P(A \cap B) \end{aligned}$$

5) On écrit $\cup A_n = \cup B_n$ avec

$B_1 = A_1$, $B_2 = A_2 \setminus A_1$, $\dots$, $B_n = A_n \setminus (\cup_{k=1}^{n-1} A_k)$ d'où

$P(\cup_{n \geq 1} A_n) = P(\cup_{n \geq 1} B_n) = \sum_{n=1}^{\infty} P(B_n) \leq \sum_{n=1}^{\infty} P(A_n)$. □

Remarques

- ① Lorsque Ω est fini, toutes les éventualités ont la même probabilité égale à $\frac{1}{\text{Card } \Omega}$ de se produire. On dit qu'elles sont équiprobables. Dans ce cas $\mathcal{A} = \mathcal{P}(\Omega)$ et $\forall A \in \mathcal{P}(\Omega)$,

$$P(A) = \frac{\text{Card } A}{\text{Card } \Omega} = \frac{\text{nbre de cas favorables à } A}{\text{nbre de cas possibles}}$$

et on retrouve la formule classique rencontrée au moins en terminale.

- ② Si on a une probabilité sur $(\Omega, \mathcal{A})$, on peut aussi définir d'autres probabilités sur $(\Omega, \mathcal{A})$ appelées probabilités conditionnelles.

Probabilité conditionnelle

Définition 2.2: Probabilité conditionnelle

Soit $(\Omega, \mathcal{A}, P)$ un espace probabilisé, B est un évènement de probabilité non nulle, (c'-à-d $B \in \mathcal{A}$ tel que $P(B) > 0$) et soit A un évènement (i.e. $A \in \mathcal{A}$). On appelle probabilité conditionnelle de A sachant B , la quantité notée $P(A | B)$ donnée par

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

Exercice 2.1:

Montrer que la fonction d'ensemble sur $\mathcal{A}$, $A \mapsto P(A \mid B)$ est une probabilité.

Remarque

1) Si $A \cap B = \emptyset$, $P(A \cap B) = 0$;

2) Si $B \subset A$, $P(A|B) = 1$.

La probabilité conditionnelle sachant B : $P^B(\cdot) = P(\cdot|B)$ est donc portée par B .

Les deux propriétés suivantes sont des plus importantes propriétés de la probabilité conditionnelle :

Théorème 2.1: (Règle de la multiplication)

Pour tous n évènements $A_1, A_2, \dots, A_n$ telle que $P(A_1 A_2 \cdots A_n) > 0$, on a

$$P(A_1 A_2 \cdots A_n) = P(A_1).P(A_2 \mid A_1)P(A_3 \mid A_1 A_2) \cdots P(A_n \mid A_1 A_2 \cdots A_{n-1})$$

où AB signifie $A \cap B$ pour des évènements A et B .

Démonstration.

Par récurrence immédiate sur n . e.g., pour $n = 2$, on a si $P(AB) > 0$

$$P(AB) = P(A \mid B)P(B).$$



Théorème 2.2: (Théorème de la partition ou règle des probabilités totales)

Soit $(C_i)_{i=1}^n$ une partition de Ω ou encore un système complet d'évènements tous de probabilité non nulle c'-à-d : $C_i \cap C_j = \emptyset, \forall i \neq j, \Omega = \bigcup_{i=1}^n C_n$ et $\forall i = 1, \dots, n, P(C_i) > 0$. Alors pour tout évènement $A \in \mathcal{A}$, on a

$$P(A) = \sum_{i=1}^n P(A \cap C_i) = \sum_{i=1}^n P(A|C_i)P(C_i).$$

Si de plus $P(A) > 0$ alors (règle de **Bayes**) : $\forall k \geq 1$,

$$P(C_k | A) = \frac{P(A | C_k)P(C_k)}{\sum_{i=1}^n P(A | C_i)P(C_i)}.$$

Événements indépendants

Définition 2.3: Indépendance de deux événements

Soient A et B deux événements ($A, B \in \mathcal{A}$). On dit que A et B sont indépendants et on note $A \perp B$ si $P(A \cap B) = P(A)P(B)$.

On voit immédiatement que si A et B sont indépendants, il en est de même pour $\overline{A}$ et B , A et $\overline{B}$, $\overline{A}$ et $\overline{B}$.

Événements indépendants

Définition 2.4: indépendants d'une famille d'événements

Soit (A_n) une suite d'évènements. On dit que les A_i sont indépendants (mutuellement) si et ssi pour toute suite finie $A_{i_1}, A_{i_2}, \dots, A_{i_p}$ d'entre eux, $2 \leq p \leq n$, on a

$$P(A_{i_1} A_{i_2} \cdots A_{i_p}) = \prod_{k=1}^p P(A_{i_k}). \quad (1)$$

On remarque que dans le cas d'une suite finie de n événements, il y a $C_n^2 + C_n^3 + \cdots + C_n^n = 2^n - n - 1$ conditions (1) à vérifier.

Le Lemme suivant permet de démontrer des résultats événements presque surs.

Lemme 2.1: (Borel-Cantelli)

Soit $(A_n)_{n \in \mathbb{N}^*}$ une suite d'évènements. Deux cas sont possibles en ce qui concerne l'évènement $\overline{\lim} A_n$:

- a) Si $\sum_{n=1}^{+\infty} P(A_n) < +\infty$ alors $P[\overline{\lim} A_n] = 0$. En d'autres termes, avec probabilité 1, au plus un nombre fini des A_n se produisent si la série précédente converge.
- b) Si les (A_n) sont indépendants et si $\sum_{n=1}^{+\infty} P(A_n) = +\infty$ alors $P[\overline{\lim} A_n] = 1$. En d'autres termes, avec probabilité 1, une infinité des A_n se produisent si la série diverge.

Indépendance d'une famille de collection

Définition 2.5: Indépendance d'une famille de collections

Soit $(\Omega, \mathcal{A})$ un espace probabilisable et $(\mathcal{A}_i)_{i \in I}$ une collection de famille d'évènements telles que $\forall i \in I, \mathcal{A}_i \in \mathcal{A}$.

On dit que cette collection est indépendante si pour toute partie finie J de I , pour tout $A_j \in \mathcal{A}_j, j \in J : P[\bigcap_{j \in J} A_j] = \prod_{j \in J} P(A_j)$.

Lemme 2.2

Soit $(\mathcal{C}_i)_{i \in I}$ une collection de famille d'évènements indépendantes. Alors les classes σ -additives $\mathcal{D}_i$ engendrées par les $\mathcal{C}_i$ sont encore indépendantes.

Indépendance d'une famille de collection

Théorème 2.3

Si $(\mathcal{C}_i)_{i \in I}$ est une collection de familles indépendantes et si chaque $\mathcal{C}_i$ est stable par intersection finie, alors les tribus engendrées par les $\mathcal{C}_i$ sont encore indépendantes.

Corollaire 2.1

Soit $(\mathcal{A}_i)_{i \in I}$ est une famille de tribus indépendantes. Soit $(I_j)_{j \in J}$ une partition de I . Notons $\mathcal{A}_{I_j} = \vee_{i \in I_j} \mathcal{A}_i = \sigma(\bigcup_{i \in I_j} \mathcal{A}_i)$. Alors les $\mathcal{A}_{I_j}$, $j \in J$ sont encore indépendantes (principe de regroupement).

Indépendance d'une famille de collection

Corollaire 2.2

Soit $(I, \leq)$ un ensemble préordonné filtrant à droite : $\forall i, j \in I \quad \exists l \in I$ tq $i \leq l$ et $j \leq l$ (e.g. $(\mathbb{N}, \leq)$, $(\mathbb{R}, \leq)$). Soit $(\mathcal{A}_i)_{i \in I}$ une famille croissante de sous-tribus de $\mathcal{A}$: $\forall i, j \in I, \quad i \leq j \Rightarrow \mathcal{A}_i \subset \mathcal{A}_j \subset \mathcal{A}$ et soit $\mathcal{C}$ une sous-tribu de $\mathcal{A}$. Si $\forall i \in I, \quad \mathcal{A}_i \perp \mathcal{C}$ alors $\lim \mathcal{A}_i = \bigvee_{i \in I} \mathcal{A}_i \perp \mathcal{C}$.

Loi de tout et de rien

Soit $(\Omega, \mathcal{A}, P)$ un e.p. , $(\mathcal{A}_n)$ une suite indépendante de sous-tribus de $\mathcal{A}$. Posons

$$\mathcal{F}_n = \bigvee_{k \leq n} \mathcal{A}_k$$

e.g. si $(\mathcal{E}_n)$ est une suite d'expériences aléatoires indépendantes et $\mathcal{A}_n$ est la tribu d'évènements relatifs à $\mathcal{E}_n$ alors $\mathcal{F}_n$ est la tribu de tous les évènements relatifs aux n premières expériences) et $\mathcal{G}_n = \bigvee_{k \geq n} \mathcal{A}_k$ est la tribu des évènements relatifs à toutes les expériences aléatoires de rang supérieure ou égale à n .

Alors $(\mathcal{F}_n)$ est croissante et $(\mathcal{G}_n)$ est décroissante vers $\bigcap_n \mathcal{G}_n$.

Loi de tout et de rien

La tribu

$$\mathcal{A} = \bigcap_n \mathcal{G}_n$$

est appelée tribu asymptotique ou tribu de queue. Les évènements de $\mathcal{A}$ sont appelés évènements asymptotiques. Ce sont les évènements qui peuvent s'exprimer par le comportement à l'infini des $\mathcal{E}_n$. On a

Théorème 2.4: (Loi de tout et de rien ou loi 0 , 1 de Kolmogorov)

$$\forall A \in \mathcal{A}, P(A) = 0 \text{ ou } 1.$$

Loi de tout et de rien

Démonstration.

Il suffit de montrer que $\mathcal{A} \perp \mathcal{A}$. D'après le principe de regroupement,

$\mathcal{G}_{n+1} = \bigvee_{k \geq n+1} \mathcal{A}_k$ est indépendante de $\mathcal{F}_n = \bigvee_{k \leq n} \mathcal{A}_k$. Donc

$\mathcal{A} \subset \mathcal{G}_{n+1} \implies \mathcal{A} \perp \mathcal{F}_n$ c'est-à-dire $\mathcal{A} \perp \bigvee_{k \leq n} \mathcal{A}_k$.

Or $(\mathcal{F}_n)$ converge vers $\bigvee_n \mathcal{F}_n = \sigma(\bigcup_n \mathcal{F}_n) = \sigma(\bigcup_n \mathcal{A}_n) = \bigvee_n \mathcal{A}_n$. Donc en passant à la limite, $\mathcal{A} \perp \bigvee_n \mathcal{A}_n = \mathcal{G}_1$. Comme $\mathcal{A} \subset \mathcal{G}_1$ alors $\mathcal{A} \perp \mathcal{A}$. Par suite,

$\forall A \in \mathcal{A}, A \perp A$. D'où $P(A \cap A) = (P(A))^2 = P(A)$ soit

$P(A) = 0$ ou $P(A) = 1$. □

Sommaire

- 1 Variables aléatoires – lois de probabilité.
 - Inégalités fondamentales
 - Lois multivariées

- fonctions caractéristiques
- Formule de changement de variables
- Indépendance de variables aléatoires

Variables aléatoires

Dans une expérience aléatoire, il est possible de faciliter la description d'un évènement A en lui associant une fonction X qui prend une valeur aléatoire x si et seulement si l'évènement A se produit.

Définition 1.1: variable aléatoire

Soit $(\Omega, \mathcal{A}, P)$ un espace probabilisé. On appelle variable aléatoire sur $(\Omega, \mathcal{A})$, toute application X de $(\Omega, \mathcal{A})$ dans un espace probabilisable $(\mathcal{X}, \mathcal{B})$ telle que, pour tout évènement $B \in \mathcal{B}$, $X^{-1}(B) \in \mathcal{A}$.

En particulier $\mathcal{X} = X(\Omega)$ est l'ensemble des valeurs possibles prises par la variable aléatoire X définie sur $(\Omega, \mathcal{A})$.

Variables aléatoires

Lorsque :

- $\mathcal{X} = \mathbb{R}$, X est dite variable aléatoire réelle (v.a.r.),
- $\mathcal{X} = \mathbb{R}^n$, X est un n -uplet de v.a.r. dit vecteur aléatoire dans $\mathbb{R}^n$,
- $\mathcal{X} = E$ espace vectoriel quelconque, X est dit vecteur aléatoire,
- $\mathcal{X} = \mathbb{N}$, $\mathbb{Z}$ ou D avec D dénombrable, X est dite variable aléatoire réelle discrète.

Notons que

$$X^{-1}(B) = \{\omega \in \Omega \mid X(\omega) \in B\} \stackrel{\text{notation}}{=} \{X \in B\}.$$

Par exemple, si $x \in \mathbb{R}$ ou $]a, b[$ est un intervalle de $\mathbb{R}$,

$$\begin{aligned} X^{-1}(]a, b[) &= \{a < X < b\} \\ X^{-1}(]-\infty, x]) &= \{X \leq x\} \\ X^{-1}(\{x\}) &= \{X = x\} \end{aligned}$$

Variables aléatoires

NB : Toutes les opérations sur les fonctions sont valables pour les variables aléatoires.

Exemple 1.1:

- ❶ Soit $A \in \mathcal{A}$. On définit la fonction indicatrice de A notée I_A par

$$I_A(\omega) = \begin{cases} 1 & \text{si } \omega \in A \\ 0 & \text{sinon} \end{cases}$$

Alors I_A est une v.a.r.

- ❷ Soit $X : (\Omega, \mathcal{A}) \longrightarrow (\mathcal{X}, \mathcal{B})$ une v.a. et $f : (\mathcal{X}, \mathcal{B}) \longrightarrow \mathbb{R}$ une fonction numérique mesurable. Alors $Y = f(X) = f \circ X$ est une v.a.r. sur $(\Omega, \mathcal{A})$.

Loi de probabilité

Définition 1.2: Loi de probabilité, fonction de répartition d'une v.a.

Soit $X : (\Omega, \mathcal{A}) \longrightarrow (\mathcal{X}, \mathcal{B})$ une v.a.

- ① On appelle loi de probabilité de la v.a. X , la probabilité P_X image de P par X définie par

$$\forall B \in \mathcal{B}, P_X(B) = P[X^{-1}(B)] = P[X \in B].$$

- ② Dans le cas où X est une v.a.r. sur $(\Omega, \mathcal{A}, P)$, on appelle fonction de répartition (f.r.) de X (ou de la loi de X), la fonction F_X définie sur $\mathbb{R}$ par

$$F_X(x) = P_X(]-\infty, x]) = P(X \in]-\infty, x]) = P[X \leq x], \quad \forall x \in \mathbb{R}.$$

Remarque 1.1: Propriétés immédiates

On vérifie que F_X caractérise la loi de la v.a.r. X et a les propriétés suivantes :

- ❶ F_X est croissante.
- ❷ $\lim_{x \rightarrow +\infty} F_X(x) = 1$ et $\lim_{x \rightarrow -\infty} F_X(x) = 0$.
- ❸ F_X est continue à droite, c'est-à-dire $\lim_{h \rightarrow 0^+} F_X(x+h) = F_X(x)$, $\forall x \in \mathbb{R}$.

Réciproquement si une fonction réelle F sur $\mathbb{R}$ vérifie les trois propriétés ci-dessus, il existe une v.a.r. X sur un espace probabilisé $(\Omega, \mathcal{A}, P)$ dont la loi a pour f.r. F .

Autres propriétés de F_X

- Si F_X est continue, donc strictement croissante, la loi de X est dite continue ou diffuse, ce qui est équivalent à $P[X = x] = 0$, $\forall x \in \mathbb{R}$.
- S'il existe une fonction positive f ayant un nombre fini de points de discontinuité telle que

$$\forall x \in \mathbb{R}, F_X(x) = \int_{-\infty}^x f(t) dt$$

alors on dit que X ou la loi de X est absolument continue (par rapport à la mesure de Lebesgue dx sur $\mathbb{R}$) admettant f pour densité de probabilité.

Notons que la densité de probabilité joue le même rôle que la probabilité élémentaire $P_x = P[X = x]$ dans le cas X discrète, mais $f \not\leq 1$ en général.

Réciproquement si g est une fonction positive et si $\int_{-\infty}^{+\infty} g(x) dx = 1$, alors g est la densité de probabilité d'une v.a.r. dont la fonction de répartition continue sur $\mathbb{R}$ est G définie par $G(x) = \int_{-\infty}^x g(t) dt$.

Théorème 1.1

Si X est une v.a.r. continue, ayant pour densité de probabilité f alors

- ① $P[X = x] = 0, \forall x \in \mathbb{R};$
- ② $P[a \leq X \leq b] = P[a < X \leq b] = P[a < X < b] = \int_a^b f(t) dt, \forall a, b \in \mathbb{R}.$

La v.a.r. X étant continue, les inégalités peuvent être remplacées par des inégalités strictes.

Moments d'une v.a.r.

Définition 1.3: Moment d'une v.a.

Soit X une v.a.r. sur un espace probabilisé $(\Omega, \mathcal{A}, P)$.

- ① On appelle espérance mathématique de X lorsqu'elle existe, la quantité réelle notée $E(X)$ ou $\overline{X}$ et donnée par

$$\overline{X} = E(X) = \int_{\Omega} X \, dP.$$

- ② Soit $k \in \mathbb{N}^*$. On appelle moment d'ordre k de X lorsqu'il existe, la quantité réelle notées $m_k(X)$ et donnée par

$$m_k(X) = E(X^k) = \int_{\Omega} X^k \, dP.$$

Moments d'une v.a.r.

Définition 1.4: Moment d'une v.a.

- ① On appelle variance (respectivement moment centré d'ordre $k \geq 1$) de X notée $Var(X)$ ou σ_X^2 (respectivement $\mu_k(X)$) lorsqu'ils existent, les quantités réels donnés par

$$\sigma_X^2 = Var(X) = E(X - E(X))^2 = \int_{\Omega} (X - E(X))^2 dP,$$

$$\mu_k(X) = E(X - E(X))^k = \int_{\Omega} (X - E(X))^k dP.$$

- ② On appelle écart-type de X , le réel noté $\sigma_X \in \mathbb{R}^+$ donné par $\sigma_X = \sqrt{Var(X)}$.

Moments d'une v.a.r.

Notons que par linéarité de l'opérateur $E(.) = \int_{\Omega} . dP$, on a

$$\sigma_X^2 = Var(X) = E(X - E(X))^2 = E(X^2) - (E(X))^2.$$

On a le résultat suivant permettant de calculer facilement l'espérance d'une fonction de la v.a.r. X :

Théorème 1.2: (de transfert)

Soit X une v.a.r. sur un espace probabilisé $(\Omega, \mathcal{A}, P)$ et g une fonction de $\mathbb{R}$ dans $\mathbb{R}$. Alors $g(X)$ est P -intégrale si, et seulement si, g est P_X -intégrable et on a dans ce cas

$$E(g(X)) = \int_{\Omega} g(X) dP = \int_{\mathbb{R}} g dP_X = \int_{-\infty}^{+\infty} g(x) dP_X(x).$$

Moments d'une v.a.r.

Remarque 1.2

- ❶ Si la loi P_X de X est absolument continue, admettant pour densité de probabilité f , alors $dP_X(x) = f(x)dx$ et

$$\int_{-\infty}^{+\infty} g(x) dP_X(x) = \int_{-\infty}^{+\infty} g(x)f(x) dx.$$

- ❷ On dit que X a un moment d'ordre $p \in \mathbb{N}^*$ ou une espérance mathématique (cas $p = 1$), si $E(|X|^p) < +\infty$. On écrit $X \in L^p = L^p(\Omega, \mathcal{A}, P)$.

Inégalité fondamentales

Théorème 1.3: (Inégalité de Markov)

Soit X une v.a.r. telle que $|X|^r \in L^1$, $r > 0$. Alors

$$\forall \varepsilon > 0, P[|X| > \varepsilon] \leq \frac{1}{\varepsilon^r} E(|X|^r).$$

Corollaire 1.1: (Inégalité de Bienamey-Tchebyshev)

Si $X \in L^2(\Omega, \mathcal{A}, P)$ alors $\forall \varepsilon > 0, P[|X - E(X)| > \varepsilon] \leq \frac{\text{var}(X)}{\varepsilon^2}$.

Lois multivariées

Les définitions et résultats précédents se généralisent au cas de vecteurs aléatoires dans $\mathbb{R}^n$. Soit $X = (X_1, \dots, X_n)$ un v.a. sur un espace probabilisé $(\Omega, \mathcal{A}, P)$ à valeurs dans $\mathbb{R}^n$.

Définition 1.5: Loi d'un v.a., fonction de répartition

On appelle fonction de répartition du v.a. X , la fonction F_X sur $\mathbb{R}^n$ définie par

$$F_X(x_1, \dots, x_n) = P[X_1 \leq x_1, \dots, X_n \leq x_n], \quad \forall (x_1, \dots, x_n) \in \mathbb{R}^n.$$

F_X détermine la loi de probabilité de X dite loi conjointe du n -uplet $(X_1, \dots, X_n)$.

Lois multivariées

Définition 1.6: Loi d'un v.a., loi marginale, absolue continuité

- 1) On appelle loi marginale de la v.a.r. X_i du n -uplet, la loi de fonction de répartition

$$F_{X_i}(x) = \lim_{\substack{x_1, \dots, x_{i-1}, x_{i+1}, \dots, x_n \rightarrow +\infty \\ x_i \rightarrow x}} F_X(x_1, \dots, x_n).$$

- 2) La loi du v.a. X est dite absolument continue s'il existe une fonction positive f sur $\mathbb{R}^n$ telle que $\forall (x_1, \dots, x_n) \in \mathbb{R}^n$,

$$F_X(x_1, \dots, x_n) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} f(t_1, \dots, t_n) dt_1 \cdots dt_n.$$

Lois multivariées

Définition 1.7: Loi d'un v.a., densité de probabilité

S'il en est ainsi, on dit que la v.a. X a pour densité

$$f_X(x_1, \dots, x_n) = \frac{\partial^n F_X(x_1, \dots, x_n)}{\partial x_1 \cdots \partial x_n}.$$

On appelle densité marginale du p -uplet $(X_{i_1}, \dots, X_{i_p})$ extrait d'un n -uplet $(X_1, \dots, X_n)$, $1 \leq p \leq n$, la fonction

$$f_{(X_{i_1}, \dots, X_{i_p})}(x_{i_1}, \dots, x_{i_p}) = \int_{\mathbb{R}^{n-p}} f(x_{i_1}, \dots, x_{i_p}, u_{i_{p+1}}, \dots, u_{i_n}) du_{i_{p+1}} \cdots du_{i_n}.$$

Lois multivariées

Exemple

- Soit (X, Y) un couple de *v.a.r.* de fonction de répartition conjointe

$$F_{(X,Y)}(x, y) = \begin{cases} 1 - e^{-x} - e^{-y} + e^{-x-y}, & \text{si } (x, y) \in \mathbb{R}_+^2 \\ 0, & \text{sinon} \end{cases}.$$

Alors les lois marginales de X et de Y ont pour fonction de répartition

$$F_X(x) = (1 - e^{-x})\mathbb{I}_{\mathbb{R}_+}(x) \quad \text{et} \quad F_Y(y) = (1 - e^{-y})\mathbb{I}_{\mathbb{R}_+}(y).$$

Lois multivariées

Exemple suite

La densité conjointe est donnée par

$$f_{(X,Y)}(x,y) = \frac{\partial^2 F_{(X,Y)}(x,y)}{\partial x \partial y} = \begin{cases} e^{-x-y}, & \text{si } (x,y) \in \mathbb{R}_+^2 \\ 0, & \text{sinon} \end{cases}$$

et les densités marginales de X et Y sont données par

$$f_X(x) = \int_{-\infty}^{+\infty} f_{(X,Y)}(x,y) \, dy = e^{-x} \mathbb{I}_{\mathbb{R}_+}(x) = F'_X(x)$$

et

$$f_Y(y) = \int_{-\infty}^{+\infty} f_{(X,Y)}(x,y) \, dx = e^{-y} \mathbb{I}_{\mathbb{R}_+}(y) = F'_Y(y).$$

Théorème 1.4

Si $X = (X_1, \dots, X_n)$ est un v.a. dans $\mathbb{R}^n$ admettant pour densité f_X et si A est un ouvert de $\mathbb{R}^n$ alors

$$P[X = (X_1, \dots, X_n) \in A] = \int_{\{(x_1, \dots, x_n) \in A\}} f_X(x_1, \dots, x_n) dx_1 \cdots dx_n.$$

Corollaire 1.2

Si $X = (X_1, \dots, X_n)$ est un v.a. dans $\mathbb{R}^n$ admettant pour densité f_X et $g : \mathbb{R}^n \rightarrow \mathbb{R}$ est une fonction positive ou bornée, alors

$$\begin{aligned} E[g(X_1, \dots, X_n)] &= \int_{\Omega} g(X_1, \dots, X_n) dP \\ &= \int_{\mathbb{R}^n} g(x_1, \dots, x_n) f_X(x_1, \dots, x_n) dx_1 \cdots dx_n. \end{aligned}$$

Définition 1.8: Covariance, corrélation linéaire

Soit $X = (X_1, \dots, X_n)$ un v.a. de $\mathbb{R}^n$. Alors l'espérance mathématique de X est le vecteur de $\mathbb{R}^n$ donné par

$$E(X) = (E(X_1), \dots, E(X_n))'$$

où prime désigne la transposé d'un vecteur ou d'une matrice.

- Pour $n = 2$, on appelle covariance du couple (X, Y) le nombre noté $Cov(X, Y)$ ou σ_{XY} donné par

$$Cov(X, Y) = E[(X - E(X))(Y - E(Y))] = E(XY) - E(X)E(Y).$$

- On appelle coefficient de corrélation de X et Y , le nombre noté r_{XY} ou ρ_{XY}

Définition 1.9: matrice de covariance

donné par

$$r_{XY} = \frac{Cov(X, Y)}{\sqrt{Var(x)}\sqrt{Var(Y)}} = \frac{Cov(X, Y)}{\sigma_x \sigma_Y}.$$

- On appelle matrice de covariance du v.a. $X = (X_1, \dots, X_n)$ de $\mathbb{R}^n$, la matrice carrée d'ordre n notée Σ_X appelée aussi matrice var-covar ou matrice de dispersion de X donnée par

$$\Sigma_X = [\sigma_{i,j}] \quad \text{avec} \quad \sigma_{i,j} = Cov(X_i, X_j), \quad 1 \leq i, j \leq n.$$

Remarque 1.3

- On démontre que

$$-1 \leq R_{XY} \leq 1.$$

Par conséquent $\exists \theta \in [0, \pi]$ tels que $R_{XY} = \cos \theta$. On vérifie aussi que

- Σ_X est une matrice symétrique et positive :

$$\forall a \in \mathbb{R}^n, a' \Sigma_X a \geq 0$$

et invariante par translation par un vecteur constant :

$$\forall a \in \mathbb{R}^n, \Sigma_{X+a} = \Sigma_X.$$

- si $M \in \mathcal{M}_{p,n}(\mathbb{R})$ et $Y = MX$ alors

$$E(Y) = ME(X) \in \mathcal{M}_{p,1}(\mathbb{R}) \quad \text{et} \quad \Sigma_Y = M\Sigma_X M' \in \mathcal{M}_p(\mathbb{R}).$$

Définition 1.10: Fonctions caractéristiques

Soit $X = (X_1, \dots, X_n)$ un v.a. de $\mathbb{R}^n$, $n \geq 1$.

- On appelle fonction caractéristique du v.a. X , la fonction sur $\mathbb{R}^n$, $t \mapsto \Phi_X(t)$ donnée par

$$\Phi_X(t) = E[e^{i\langle t, X \rangle}] = \int_{\mathbb{R}^n} e^{i\langle t, x \rangle} dP_X(x)$$

où pour tous $n \geq 1$, pour tous $t, x \in \mathbb{R}^n$, $\langle t, x \rangle = \sum_{i=1}^n t_i x_i$ désigne le produit scalaire canonique des vecteurs x et t de $\mathbb{R}^n$.

- Lorsque X est une v.a.r. à valeurs dans $\mathbb{N}$, on définit la fonction génératrice G_X de X par

$$G_X(s) = E[s^X] = \sum_{k=0}^{\infty} s^k P[X = k], \quad \forall s \in [-1, 1].$$

Remarque 1.4

- Si la loi de X a une densité de probabilité f_X , $dP_X(x) = f_X(x)dx$.
- Si X est une v.a.r.d., $dP_X(x) = P[X = x]$ et l'intégrale est une somme.

Théorème 1.5: Propriétés

1) La fonction caractéristique $\Phi_X(t)$ caractérise la loi de la v.a. X :

$$\Phi_X(t) = \Phi_Y(t), \quad \forall t \in \mathbb{R}^n, \quad \Longleftrightarrow \quad X \text{ et } Y \text{ ont même loi.}$$

2) Si $a \in \mathbb{R}$, $b \in \mathbb{R}^n$, on a $\Phi_{aX+b}(t) = e^{i\langle t, b \rangle} \Phi_X(at)$.

3) Si X est une v.a.r. (cas $n = 1$) et si $|X|^p \in L^1$, alors Φ_X est de classe C^p et pour k , $1 \leq k \leq p$ on a

$$\Phi_X^{(k)}(0) = i^k E[X^k]$$

Théorème 1.6: (Formule d'inversion)

Soit X est une v.a.r. telle que $\Phi_X(t) \in L^1(\mathbb{R})$. Alors la loi de X est absolument continue admettant pour densité de probabilité donnée par

$$f_X(x) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{-itx} \Phi_X(t) dt.$$

Changement de variables

En calculs de probabilités on est dès fois amené à effectuer un changement de variables. Le théorème suivant permet de d'en déterminer la loi.

Théorème 1.7: (Formule de changement de variable)

Soit $\mathcal{D}$ et Δ deux ouverts de $\mathbb{R}^n$, $n \geq 1$ et φ un C^1 -difféomorphisme de $\mathcal{D}$ sur Δ . Alors pour toute fonction borélienne positive ou intégrable g sur Δ on a

$$\int_{\Delta} g(y) \, dy = \int_{\mathcal{D}} g(\varphi(x)) |J\varphi(x)| \, dx \quad (1)$$

où $J\varphi(x)$ désigne le déterminant de la matrice jacobienne de $\varphi = (\varphi_1, \dots, \varphi_n)$.

Cas des variables aléatoires

Soit X un v.a. à valeurs dans un ouvert $\mathcal{D}$ de $\mathbb{R}^n$ ayant une densité f_X et soit ψ un C^1 -difféomorphisme de l'ouvert $\mathcal{D}$ dans un ouvert Δ de $\mathbb{R}^n$. On pose $Y = \psi(X)$ et on cherche la loi de Y . On a

pour toute fonction positive g quelconque sur Δ :

$$\begin{aligned} E(g(Y)) &= E(g(\psi(X))) = \int_{\Omega} g(\psi(X)) \, dP = \int_{\mathcal{D}} g(\psi(x)) f_X(x) \, dx \\ &= \int_{\Delta} g(y) f_X(\psi^{-1}(y)) |J\psi^{-1}(y)| \, dy \\ &= \int_{\Delta} g(y) f_X(\psi^{-1}(y)) |J\psi^{-1}(y)| \mathbb{I}_{\Delta}(y) \, dy. \end{aligned}$$

On en déduit que la v.a. Y a pour loi de densité donnée par

$$f_Y(y) = f_X(\psi^{-1}(y)) |J\psi^{-1}(y)| \mathbb{I}_{\Delta}(y).$$

Exemple

Exemple

Soit $X = (X_1, X_2)$ un couple de v.a.r. de loi conjointe sur $\mathbb{R}^2$ définie par sa densité $f_X(x_1, x_2) = \frac{1}{2\pi} e^{-\frac{1}{2}(x_1^2 + x_2^2)}$. On pose $Y = \frac{X_1}{X_2}$ avec $Y = 0$ si $X_2 = 0$. cherchons la loi de Y .

Pour toute g positive sur $\mathbb{R}$ on a

$$E(g(Y)) = E\left[g\left(\frac{X_1}{X_2}\right)\right] = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} g\left(\frac{x_1}{x_2}\right) e^{-\frac{1}{2}(x_1^2 + x_2^2)} dx_1 dx_2.$$

Soit le difféomorphisme $\psi : \mathbb{R}^2 \mapsto \mathbb{R}^2$ qui à (x_1, x_2) associe (y, z) avec $y = \frac{x_1}{x_2}$ et $z = x_2$. Alors ψ^{-1} est le difféomorphisme réciproque qui à (y, z) associe $x_1 = yz$, $x_2 = z$.

Exemple suite

On a $J\psi^{-1}(y, z) = \frac{D(x_1, x_2)}{D(y, z)} = \begin{vmatrix} z & y \\ 0 & 1 \end{vmatrix} = z$. D'où

$$\begin{aligned} E(g(Y)) &= \frac{1}{2\pi} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} g(y) e^{-\frac{1}{2}(1+y^2)z^2} |z| dy dz \\ &\stackrel{\text{Fubini}}{=} \int_{\mathbb{R}} g(y) \left\{ \frac{1}{2\pi} \int_{\mathbb{R}} e^{-\frac{1}{2}(1+y^2)z^2} |z| dy dz \right\} \\ &= \int_{\mathbb{R}} g(y) \frac{dy}{\pi(1+y^2)}. \end{aligned}$$

La loi de Y est donc la loi de densité $f(y) = \frac{1}{\pi(1+y^2)}$, $\forall y \in \mathbb{R}$.

Indépendance des variables aléatoires

Notons par $\mathcal{B}(\mathbb{R}^n)$ la tribu sur $\mathfrak{X} = \mathbb{R}^n$ (tribu engendrée par les boréliens de $\mathbb{R}^n$).

Définition 1.11: Indépendance des variables aléatoires

- ① Soit (X, Y) un couple de v.a. sur $(\Omega, \mathcal{A}, P)$ à valeurs dans $\mathbb{R}^n$, $n \geq 1$.
On dit que X et Y sont indépendantes et on note $X \perp Y$ si

$$P[X \in A, Y \in B] = P[X \in A]P[Y \in B], \text{ pour tous } A, B \in \mathcal{B}(\mathbb{R}^n).$$

- ② On dit que les v.a. $X_1, X_2, \dots, X_n$ sur $(\Omega, \mathcal{A}, P)$ à valeurs dans $\mathbb{R}^n$, $n \geq 1$ sont (mutuellement) indépendantes et on note $X_1, \dots, X_n \perp$ si pour tous $A_1, \dots, A_n \in \mathcal{B}(\mathbb{R}^n)$

$$P[X_1 \in A_1, \dots, X_n \in A_n] = P[X_1 \in A_1] \cdots P[X_n \in A_n]. \quad (2)$$

Indépendance des variables aléatoires

Définition 1.12: Indépendance des variables aléatoires

- ① On dit que $(X_n)_{n \in \mathbb{N}}$ est une suite de v.a. indépendante si (2) est vérifiée pour tout $n \in \mathbb{N}^*$
- ② On dit que les v.a. $X_1, X_2, \dots, X_n$ sont i.i.d. (indépendantes et identiquement distribuées) si elles sont (mutuellement) indépendantes et elles sont toutes de même loi.

Remarque 1.5

- Le fait que les X_i soient indépendantes deux à deux ($X_i \perp X_j \forall i \neq j$) n'implique pas que les v.a. $X_1, X_2, \dots, X_n$ sont mutuellement indépendantes. En revanche, l'indépendance mutuelle implique l'indépendance deux à deux.
- Si X et Y sont deux v.a. indépendantes on a $\text{Cov}(X, Y) = 0$ ou encore $E(XY) = E(X)E(Y)$. Mais la réciproque est fausse. Notons que si $\text{Cov}(X, Y) = 0$, on dit que les v.a. X et Y sont non corrélées ou sont orthogonales.
- Les transformations mesurables préservent l'indépendance : si $X \perp Y$, alors $f(X) \perp g(Y)$, quelles que soient les fonctions boréliennes f et g .

Indépendance des variables aléatoires

Proposition 1.1: (Critères)

1- **Critère des fonctions de répartitions** : Les v.a.r. $X_1, \dots, X_n$ sont indépendantes si et seulement si,

$$F_{(X_1, \dots, X_n)}(x_1, \dots, x_n) = \prod_{i=1}^n F_{X_i}(x_i), \quad \forall (x_1, \dots, x_n) \in \mathbb{R}^n. \quad (3)$$

2- **Critère des fonctions des densités** : Si les n v.a.r. $X_1, \dots, X_n$ ont pour densités $f_1, \dots, f_n$ respectivement alors $(X_1, \dots, X_n) \perp$ si et seulement si, $\forall (x_1, \dots, x_n) \in \mathbb{R}^n$,

$$f_{(X_1, \dots, X_n)}(x_1, \dots, x_n) = \prod_{i=1}^n f_{X_i}(x_i) \quad (4)$$

Indépendance des variables aléatoires

Proposition 1.2: (Critères suite)

3- **Critère des espérances** : Les v.a.r. $X_1, \dots, X_n$ sur $(\Omega, \mathcal{A}, P)$ sont indépendantes si et ssi, pour toutes fonctions réelles ou complexes positives ou bornées $h_1, \dots, h_n$

$$E(\Pi_{i=1}^n h_i(X_i)) = \Pi_{i=1}^n E(h_i(X_i)). \quad (5)$$

4- **Critère des fonctions caractéristiques** : Soient $\Phi_{X_1}, \dots, \Phi_{X_n}$ les fonctions caractéristiques respectives des v.a.r. $X_1, \dots, X_n$. Alors $(X_1, \dots, X_n)$ indépendante si et seulement si, $\forall (t_1, \dots, t_n) \in \mathbb{R}^n$,

$$\Phi_{(X_1, \dots, X_n)}(t_1, \dots, t_n) = \Pi_{i=1}^n \Phi_{X_i}(t_i). \quad (6)$$

Indépendance des variables aléatoires

Proposition 1.3: (Loi d'une somme de v.a.r.)

Si $X_1, \dots, X_n$ sont n v.a.r. sur $(\Omega, \mathcal{A}, P)$ indépendantes alors la loi de $S_n = X_1 + \dots + X_n$ est égale au produit de convolution des lois des X_i :

$$P_{X_1 + \dots + X_n} = P_{X_1} * P_{X_2} * \dots * P_{X_n}.$$

En particulier si les X_i ont des densités f_i , $i = 1, 2, \dots, n$ et si les X_i sont mutuellement indépendante alors

$$f_{S_n} = f_1 * f_2 * \dots * f_n.$$

Indépendance des variables aléatoires

Notons que par définition, la convolution $*$ de deux fonctions f et g positives ou intégrables est définie par

$$f * g(x) = \int_{-\infty}^{+\infty} f(t)g(x - t)dt.$$

Cette loi $*$ est commutative et associative. La proposition précédente dit que si X et Y sont deux v.a.r. admettant pour densités f_X et f_Y respectivement, et si X et Y sont indépendantes, alors la loi de la v.a.r. $S = X + Y$ a pour densité de probabilité

$$f_S(x) == \int_{-\infty}^{+\infty} f_X(t)f_Y(x - t)dt.$$

Sommaire

- 1 Lois conditionnelles
 - Cas de variables aléatoires réelles

discrètes

- Cas de variables aléatoires réelles continues
- Espérance conditionnelle

Lois conditionnelles

Soit X et Y deux v.a.r. discrètes. Par analogie à la définition de la probabilité conditionnelle de A sachant B avec $P(B) > 0$, on définit la loi conditionnelle de X sachant $Y = y$ pour tout $y \in \mathbb{R}$ telle que $P(Y = y) \neq 0$ par :

$$P^{Y=y}(X = x) = P(X = x|Y = y) = \frac{P(X = x, Y = y)}{P(Y = y)}.$$

La fonction de répartition conditionnelle de X sachant $Y = y$ est définie par

$$F_{X/Y}(x|y) = P^{Y=y}(X \leq x) = \sum_{s \leq x} P^{Y=y}(X = s).$$

Lois conditionnelles

Exemple

Soient X et Y deux v.a. indépendantes de lois de Poisson de paramètres λ et μ respectivement. On cherche la loi de X sachant $X + Y = n$, $n \in \mathbb{N}$. On a, pour tout $0 \leq k \leq n$,

$$P(X = k | X + Y = n) = \frac{P(X = k, X + Y = n)}{P(X + Y = n)} = \frac{P(X = k, Y = n - k)}{P(X + Y = n)}.$$

Or X et Y sont indépendantes, donc $X + Y$ suit la loi de Poisson de paramètre $\lambda + \mu$ et $P(X = k, Y = n - k) = P(X = k)P(Y = n - k)$. Par conséquent,

$$P(X = k | X + Y = n) = \frac{P(X = k)P(Y = n - k)}{P(X + Y = n)} = \frac{\frac{\lambda^k e^{-\lambda}}{k!} \frac{\mu^{(n-k)} e^{-\mu}}{(n-k)!}}{\frac{(\lambda + \mu)^n e^{-(\lambda + \mu)}}{n!}}.$$

Après simplification, on obtient

$$P(X = k | X + Y = n) = C_n^k \left(\frac{\lambda}{\lambda + \mu} \right)^k \left(\frac{\mu}{\lambda + \mu} \right)^{n-k}.$$

On en conclut que la loi de X sachant $X + Y = n$ est la loi binomiale de paramètres n et $p = \frac{\lambda}{\lambda + \mu}$.

Remarque 1.1

Lorsque X et Y sont indépendantes, $P(X = x|Y = y) = P(X = x)$ et $F_{X/Y}(x|y) = F_X(x) = P(X \leq x)$.

Lois conditionnelles

Si les v.a.r. X et Y sont continues, soit $f_{X,Y}(x,y)$ la densité conjointe du couple (X,Y) , $f_X(x)$ et $f_Y(y)$ leurs densités marginales respectives. Alors il existe des distributions conditionnelles de X sachant Y et de Y sachant X données par

- densité conditionnelle de X sachant $Y = y$:

$$f_X^{Y=y}(x) = \frac{f_{X,Y}(x,y)}{f_Y(y)}, \quad \forall y \quad \text{tq} \quad f_Y(y) \neq 0, \quad (1)$$

- densité conditionnelle de Y sachant $X = x$:

$$f_Y^{X=x}(y) = \frac{f_{X,Y}(x,y)}{f_X(x)}, \quad \forall x \quad \text{tq} \quad f_X(x) \neq 0. \quad (2)$$

Exemple

Soit X et Y de densité conjointe

$$f_{X,Y}(x,y) = \begin{cases} 2e^{-x-y}, & \text{si } 0 \leq x \leq y < +\infty \\ 0, & \text{sinon.} \end{cases}$$

$I(A)$ désignant ici la fonction indicatrice de l'événement A , on a

$$f_X(x) = 2e^{-2x}I(x \geq 0), \quad f_Y(y) = 2e^{-y}(1 - e^{-y})I(y \geq 0).$$

La densité conditionnelle de Y sachant $X = x$ est

$$f_Y^{X=x}(y) = \frac{2e^{-x-y}I(0 \leq x \leq y)}{2e^{-2x}I(x \geq 0)} = e^{x-y}I(0 \leq x \leq y).$$

Exemple

La densité conditionnelle de X sachant $Y = y$ est

$$f_X^{Y=y}(x) = \frac{2e^{-x-y}I(0 \leq x \leq y)}{2e^{-y}(1 - e^{-y})I(y \geq 0)} = \frac{e^{-x}}{1 - e^{-y}}I(0 \leq x \leq y).$$

Espérance conditionnelles

Cas discret :

Pour tout y telle que $P(Y = y) > 0$, on définit l'espérance conditionnelle de X sachant $Y = y$ par

$$E(X|Y = y) = \sum_x x P^{Y=y}(X = x) = \sum_x x P(X = x|Y = y) \quad (3)$$

cas continu :

Lorsque les v.a. X et Y sont continues, on définit l'espérance conditionnelle de X sachant $Y = y$ pour tout y telle que $f_Y(y) > 0$ par

$$E(X|Y = y) = \int_{\mathbb{R}} x f_X^{Y=y}(x) dx. \quad (4)$$

Exemple

Soit X et Y de densité conjointe $f_{X,Y}(x, y) = \frac{e^y e^{-x/y}}{y} I(x > 0, y > 0)$.

On a d'abord, en se rappelant que $\mathbb{I}_{A \cap B} = \mathbb{I}_A \mathbb{I}_B$,

$$\begin{aligned} f_X^{Y=y}(x) &= \frac{f_{X,Y}(x, y)}{f_Y(y)} = \frac{f_{X,Y}(x, y)}{\int_0^{+\infty} f_{X,Y}(x, y) dx} \\ &= \frac{y^{-1} e^y e^{-x/y} I(x > 0, y > 0)}{\int_{-\infty}^{+\infty} y^{-1} e^y e^{-x/y} I(x > 0, y > 0) dx} = \frac{1}{y} e^{-\frac{x}{y}} I(y > 0). \end{aligned}$$

Donc pour tout $y > 0$, la loi conditionnelle de X sachant $Y = y$ est la loi exponentielle de paramètre $\frac{1}{y}$ et cette loi a pour espérance y . Ce qui veut dire que

$$E(X|Y = y) = \int_0^{+\infty} xy^{-1} e^y e^{-x/y} dx = y.$$

Espérance conditionnelles

Notons que $E(X|Y)$ est une variable aléatoire réelle qui est de la forme $E(X|Y) = f \circ Y = f(Y)$, f étant la fonction réelle $x \mapsto f(y) = E(X|Y = y)$. On a

Théorème 1.1: (Espérance successive)

$$E(E(X|Y)) = E(X) \quad (5)$$

Si Y est discrète, ce théorème dit que

$$E(X) = E(E(X|Y)) = \sum_y E(X|Y = y)P(Y = y) \quad (6)$$

et dans le cas cas continu, il dit que

$$E(X) = E(E(X|Y)) = \int_{\mathbb{D}} E(X|Y = y)f_Y(y)dy. \quad (7)$$

Sommaire

1 Lois de probabilités usuelles

- Lois discrètes courantes
- Lois continues courantes

Lois discrètes courantes

Définition 1.1: Loi de Bernoulli $\mathcal{B}(1, p)$

On dit qu'une v.a.r. X suit une loi de Bernoulli de paramètre p , $0 < p < 1$ et on note $X \sim \mathcal{B}(1, p)$ si

$$\mathcal{X} = X(\Omega) = \{0, 1\} \quad \text{et on a} \quad P[X = 1] = p \text{ et } P[X = 0] = q = 1 - p.$$

On vérifie que $E(X) = p$, $Var(X) = p(1 - p) = pq$ et $\Phi_X(t) = q + pe^{it}$.

La loi de Bernoulli est utilisée pour modéliser des expériences ayant seulement deux résultats possibles, en particulier des situations classifiées en échec ($X = 0$) ou succès ($X = 1$).

Lois discrètes courantes

Définition 1.2: Loi binomiale $\mathcal{B}(n, p)$

On dit qu'une v.a.r. X suit une loi binomiale de paramètres n et p , $0 < p < 1$ et on note $X \rightsquigarrow \mathcal{B}(n, p)$ si

$\mathcal{X} = X(\Omega) = \{0, 1, \dots, n\}$ et on a $P[X = k] = C_n^k p^k q^{n-k}$, $\forall k = 0, 1, \dots, n$.

On vérifie que $E(X) = np$, $Var(X) = npq$ et $\Phi_X(t) = (q + pe^{it})^n$.

La loi binomiale $\mathcal{B}(n, p)$ modélise la loi du nombre de succès parmi n expériences i.i.d. de Bernoulli $\mathcal{B}(1, p)$.

Lois discrètes courantes

Définition 1.3: Loi de Poisson $\mathcal{P}(\lambda)$

Une v.a. X suit une loi de Poisson de paramètre $\lambda > 0$ et on note $X \sim \mathcal{P}(\lambda)$ si $X(\Omega) = \mathbb{N}$ et $\forall k \in \mathbb{N}$,

$$P[X = k] = \frac{\lambda^k e^{-\lambda}}{k!}.$$

On vérifie que $E(X) = \lambda$, $Var(X) = \lambda$ et $\Phi_X(t) = e^{\lambda(e^{it}-1)}$.

La loi de Poisson est une sorte de loi binomiale pour laquelle n est assez grand, p petite voisine de 0 et np voisin de λ .

Lois discrètes courantes

Définition 1.4: Loi de Pascal de paramètre $p : \mathcal{G}(p)$

Une v.a. X suit une loi de Pascal ou loi géométrique de paramètre p , $0 < p < 1$ et on note $X \sim \mathcal{G}(p)$ si $X(\Omega) = \mathbb{N}^*$ et $\forall k \in \mathbb{N}^*$,

$$P[X = k] = p(1 - p)^{k-1} = pq^{k-1}.$$

$$\text{On a } E(X) = \frac{1}{p}, \quad \text{Var}(X) = \frac{q}{p^2} \quad \text{et} \quad \Phi_X(t) = \frac{pe^{it}}{1 - qe^{it}}.$$

La loi géométrique peut être définie sur $\mathbb{N}$. Dans ce cas la v.a. X est par exemple le nombre d'échecs nécessaires pour obtenir pour la première fois un succès.

Lois discrètes courantes

Définition 1.5: Loi binomiale négative de paramètres n et p :
 $\mathcal{BN}(n, p)$

Une v.a. X suit une loi binomiale négative de paramètres n et p , $0 < p < 1$, si $X(\Omega) = \{n, n+1, \dots\}$ et $\forall k \geq n$,

$$P[X = k] = C_{k-1}^{n-1} p^n (1-p)^{k-n} = C_{n+k-1}^{n-1} p^n (1-p)^k \mathbb{I}_{\{k \geq 0\}}.$$

Notons qu'on peut écrire $X = \sum_{i=1}^n \zeta_i$ avec ζ_i , $i = 1, \dots, n$ i.i.d. avec $\zeta_1 \sim \mathcal{G}(p)$

Lois discrètes courantes

Définition 1.6: Loi Hypergéométrique de paramètres N, M, n :
 $\mathcal{H}(N, M, n)$

Une v.a. X suit une loi hypergéométrique de paramètres N , M et n si $X(\Omega) = \{0, 1, \dots, n\}$, $n \leq N$ et $\forall k \in X(\Omega)$,

$$P(X = k) = \frac{C_M^k C_{N-M}^{n-k}}{C_N^n}.$$

Notons que parfois les paramètres sont N, n, p où $p = \frac{M}{N}$. On a $E(X) = np$ et $Var(X) = \frac{N-n}{N-1}npq$, avec $q = 1 - p$.

Lois continues courantes

Définition 1.7: Loi uniforme sur $[a, b]$: $\mathcal{U}[a, b]$

Une v.a.r. X est dite suivre une loi uniforme sur $[0, 1]$ ou simplement une loi uniforme, si X est continue et a pour densité

$$f_X(x) = \mathbb{I}_{[0,1]}(x) = \begin{cases} 1, & \text{si } 0 \leq x \leq 1 \\ 0, & \text{sinon} \end{cases}$$

$$\text{On a } E(X) = \frac{1}{2}, \quad \text{Var}(X) = \frac{1}{12} \text{ et } \Phi_X(t) = \frac{e^{it} - 1}{it}.$$

Lois continues courantes

Définition 1.8: Loi uniforme sur $[a, b]$: $\mathcal{U}[a, b]$ suite

Plus généralement, on dit que X suit une loi uniforme sur l'intervalle $[a, b]$, $a < b$ et on note $X \rightsquigarrow \mathcal{U}[a, b]$ si la v.a. $U = \frac{X - a}{b - a}$ suit la loi $\mathcal{U}[0, 1]$. On en déduit que

$$f_X(x) = \frac{1}{b - a} \mathbb{I}_{[a, b]}(x), \quad E(X) = \frac{a + b}{2} \quad \text{et} \quad \text{var}(X) = \frac{(b - a)^2}{12}.$$

Notons que si $\Phi_X(t)$ est la f.c. d'une v.a.r. X alors $Y = aX + b$ a pour f.c. $\Phi_Y(t) = e^{itb} \Phi_X(at)$.

Lois continues courantes

Définition 1.9: Loi normale de moyenne m et d'écart-type σ : $\mathcal{N}(m, \sigma)$

$X \sim \mathcal{N}(m, \sigma)$, $m \in \mathbb{R}$ et $\sigma > 0$, si X est continue et admet pour densité de probabilité

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}(x-m)^2}, \quad \forall x \in \mathbb{R}.$$

La courbe représentative de $f_X(x)$ est symétrique par rapport à la droite d'équation $x = m$.

Dans le cas particulier $m = 0$ et $\sigma = 1$, la loi $\mathcal{N}(0, 1)$ est appelée loi normale centrée réduite ou loi de Laplace-Gauss.

Lois continues courantes

Définition 1.10: Loi normale $\mathcal{N}(0, 1)$

Sa densité de probabilité est donnée par

$$f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \quad \forall x \in \mathbb{R}.$$

Soit $\Phi(x)$ la f.r. de la loi $\mathcal{N}(0, 1)$. $\Phi(x) = \int_{-\infty}^x f_X(t) dt$. On a par symétrie de la CR de $\Phi(x)$: $\Phi(-u) = 1 - \Phi(u)$. Donc si $u_\alpha \in \mathbb{R}$ vérifie $\Phi(u_\alpha) = \alpha \in [0, 1]$, on a en valeur absolue $u_\alpha = u_{1-\alpha}$. Notons que $X \rightsquigarrow \mathcal{N}(m, \sigma)$ si et ssi,

$U = \frac{X - m}{\sigma} \rightsquigarrow \mathcal{N}(0, 1)$ ou encore si et ssi $X = \sigma U + m$ avec $U \rightsquigarrow \mathcal{N}(0, 1)$. D'où $E(X) = m$ et $\sigma_X^2 = \text{Var}(X) = \sigma^2$.

Lois continues courantes

Définition 1.11: Loi log-normale de paramètres m et σ : $\mathcal{LN}(m, \sigma)$

$X \sim \mathcal{LN}(m, \sigma)$, $m \in \mathbb{R}$ et $\sigma > 0$, si $Y = \log(X)$ suit la loi normale $\mathcal{N}(m, \sigma)$.
 X admet pour densité de probabilité

$$f_X(t) = \frac{\alpha}{\sqrt{2\pi t}} e^{-\frac{\alpha^2 (\log(\lambda t))^2}{2}}$$

avec $m = -\log(\lambda)$ et $\sigma = \frac{1}{\alpha}$.

Lois continues courantes

Définition 1.12: Loi gamma de paramètres p et $\theta : \gamma(p, \theta)$

X suit une loi de gamma de paramètres $p > 0$ et $\theta > 0$ et on note $X \sim \gamma(p, \theta)$, si X est continue et a pour densité

$$f_X(x) = \frac{\theta^p}{\Gamma(p)} x^{p-1} e^{-\theta x} \mathbb{I}_{\mathbb{R}^+}(x)$$

où $\Gamma(s) = \int_0^{+\infty} x^{s-1} e^{-x} dx$, $\forall s > 0$. On a

$$E(X) = \frac{p}{\theta}, \quad Var(X) = \frac{p}{\theta^2} \quad \text{et} \quad \forall k \in \mathbb{N}^*, \quad E(X^k) = \frac{\Gamma(p+k)}{\theta^k \Gamma(p)}.$$

Lois continues courantes

Définition 1.13: Loi exponentielle de paramètre $\theta > 0$: $\mathcal{E}(\theta)$

Dans le cas particulier $p = 1$, la loi $\gamma(1, \theta)$ est appelée loi exponentielle de paramètre $\theta > 0$ notée $\mathcal{E}(\theta)$. $X \sim \mathcal{E}(\theta)$ si X est continue et a pour densité

$$f_X(x) = \theta e^{-\theta x} \mathbb{I}_{\mathbb{R}^+}(x).$$

$$E(X) = \frac{1}{\theta}, \quad \sigma_X^2 = \text{Var}(X) = \frac{1}{\theta^2} \quad \text{et} \quad \Phi_X(t) = \frac{\theta}{\theta - it}.$$

Lois continues courantes

Définition 1.14: Loi de Weibull de paramètres λ et $\alpha : \mathcal{W}(\lambda, \alpha)$

X suit une loi de Weibull de paramètres $\lambda > 0$ et $\alpha > 0$ et on note $X \sim \mathcal{W}(\lambda, \alpha)$, si X est continue et a pour densité

$$f_X(t) = \lambda \alpha (\lambda t)^{\alpha-1} e^{-(\lambda t)^\alpha}$$

On a

$$E(X) = \frac{1}{\lambda} \Gamma(1 + \frac{1}{\alpha}) \quad \text{et} \quad Var(X) = \frac{1}{\lambda^2} \Gamma(1 + \frac{2}{\alpha}) - \frac{1}{\lambda^2} (\Gamma(1 + \frac{1}{\alpha}))^2.$$

Notons que pour $\alpha = 1$, la loi $\mathcal{W}(\lambda, 1)$ est identique à la loi $\mathcal{E}(\lambda)$ et que les lois gamma, exponentielle et de Weibull servent à modéliser les durées.

Lois continues courantes

Définition 1.15: Loi de Cauchy de paramètres a et b : $\text{Cauchy}(a, b)$

$X \rightsquigarrow \text{Cauchy}(a, b)$ si X est continue et a pour densité

$$f_X(x) = \frac{b}{\pi(b^2 + (x - a)^2)}.$$

Les cas $a = 0$ et $b = 1$ sont les plus courants. Comme on le voit bien, la loi de Cauchy n'a aucun moment. On a

$$\Phi_X(t) = e^{-|t|}, \quad \forall t \in \mathbb{R}.$$

Lois continues courantes

Définition 1.16: Loi de Chi-carré à ν d.d.l. : χ_ν^2

$X \sim \chi_\nu^2$ si $X = \sum_{i=1}^{\nu} U_i^2$ avec $U_1, \dots, U_\nu$ iid $\mathcal{N}(0, 1)$. On peut prouver aussi que X a la loi $\gamma(\frac{\nu}{2}, \frac{1}{2})$ et donc que

$$E(X) = \nu, \quad Var(X) = 2\nu.$$

Lois continues courantes

Définition 1.17: Loi de Student à ν d.d.l. : $ST(\nu)$

$X \rightsquigarrow ST(\nu)$ si $X = \frac{U}{\sqrt{\frac{\chi^2_\nu}{\nu}}}$ avec $U \rightsquigarrow \mathcal{N}(0, 1)$ et U indépendante de χ^2_ν . On a

$$E(X) = 0.$$

Lois continues courantes

Définition 1.18: Loi de Fisher-Snedecor à ν_1 et ν_2 d.d.l. : $F(\nu_1, \nu_2)$

$$X \rightsquigarrow F(\nu_1, \nu_2) \text{ si } X = \frac{\chi_{\nu_1}^2 / \nu_1}{\chi_{\nu_2}^2 / \nu_2}.$$

On démontre que si $X \rightsquigarrow F(\nu_1, \nu_2)$ alors $Y = \frac{1}{X} \rightsquigarrow F(\nu_2, \nu_1)$.

On notera que les lois de Chi-carré, de Student et de Fisher-Snedecor sont de la famille des lois normales et elles sont toutes tabulées.

Loi dégénérée

Définition 1.19: Loi dégénérée ou loi de Dirac en x_0 : δ_{x_0}

$X \rightsquigarrow \delta_{x_0}$ si $P[X = x_0] = 1$.

On a

$$\Phi_X(t) = e^{-itx_0}, \quad \forall t \in \mathbb{R}.$$

Sommaire

1 Vecteurs Gaussiens

- Représentation de lois normales
- Singularité et absolue continuité

Vecteurs Gaussiens

Définition 1.1: Lois normales sur $\mathbb{R}^n$

Soit X un v.a. dans $\mathbb{R}^n$. On dit que X est normal ou gaussien s'il existe $m \in \mathbb{R}^n$ et $\Gamma \in \mathcal{M}_n(\mathbb{R})$ symétrique et positive telle que

$$\Phi_X(t) = e^{i\langle t, m \rangle} e^{-\frac{1}{2}\langle \Gamma t, t \rangle}, \quad \forall t \in \mathbb{R}^n$$

où $\langle t, m \rangle = \sum_{i=1}^n t_i m_i$ et $\langle \Gamma t, t \rangle = t' \Gamma t = \sum_{1 \leq i, j \leq n} \gamma_{ij} t_i t_j$, γ_{ij} étant les éléments de la matrice Γ . On note $X \sim \mathcal{N}(\vec{m}, \Gamma)$.

Vecteurs Gaussiens

Théorème 1.1

Un v.a. $X = (X_1, \dots, X_n)$ dans $\mathbb{R}^n$ est gaussien si et seulement si, toute combinaison linéaire ou affine des X_i suit une loi normale sur $\mathbb{R}$.

Corollaire 1.1

Si $X = (X_1, \dots, X_n) \rightsquigarrow \mathcal{N}(\vec{m}, \Gamma)$ alors $\vec{m} = (E(X_i))_{1 \leq i \leq n}$ et $\Gamma = [\text{Cov}(X_i, X_j)]_{1 \leq i, j \leq n}$.

Vecteurs Gaussiens

Théorème 1.2

Soit X un vecteur gaussien de $\mathbb{R}^n$: $X \sim \mathcal{N}(\vec{m}, \Gamma)$ et soit $g : \mathbb{R}^n \longrightarrow \mathbb{R}^p$ une fonction affine : $g(x) = \varphi(x) + b$ avec $\varphi \in L(\mathbb{R}^n)$ et $b \in \mathbb{R}^p$. Alors $g(X)$ est gaussien de loi $\mathcal{N}(g(m), \varphi \Gamma {}^t\varphi)$.

Vecteurs Gaussiens

Soit $m \in \mathbb{R}^n$, Γ une matrice symétrique d'ordre n et $X \sim \mathcal{N}(\vec{m}, \Gamma)$. Si Γ est du type positif de rang $p < n$, il existe une base orthonormée de $\mathbb{R}^n$ dans laquelle Γ est semblable à $\tilde{\Gamma} = \text{diag}(\lambda_1, \dots, \lambda_p, 0, \dots, 0)$ avec $\lambda_i > 0, \forall i = 1, 2, \dots, p$. Soit P la matrice de passage. P est orthogonale donc $P^{-1} = {}^tP$ et on a $\Gamma = P \tilde{\Gamma} {}^tP$. Posons $Y = {}^tP(X - m)$. Y est gaussienne comme fonction linéaire d'un vecteur gaussien. On a $E(Y) = 0$ et

$$\begin{aligned} \Phi_Y(t) &= e^{-\frac{1}{2} \langle {}^tP \Gamma P t, t \rangle} = e^{-\frac{1}{2} \langle \tilde{\Gamma} t, t \rangle} = e^{-\frac{1}{2} \sum_{i=1}^p \lambda_i t_i^2} = \prod_{i=1}^p e^{-\frac{1}{2} \lambda_i t_i^2} \\ &= \Phi_{\mathcal{N}(0, \lambda_1)}(t_1) \times \Phi_{\mathcal{N}(0, \lambda_2)}(t_2) \times \dots \times \Phi_{\mathcal{N}(0, \lambda_p)}(t_p) \times \Phi_{\delta_0}(t_{p+1}) \times \dots \times \Phi_{\delta_0}(t_n) \\ &= \Phi_{\mathcal{N}(0, \lambda_1) \otimes \mathcal{N}(0, \lambda_2) \otimes \dots \otimes \mathcal{N}(0, \lambda_p) \otimes \delta_0 \otimes \dots \otimes \delta_0}(t), \quad \forall t \in \mathbb{R}^n. \end{aligned}$$

Vecteurs Gaussiens

Donc $Y = (Y_1, \dots, Y_p, 0, \dots, 0)$ avec $Y_1, Y_2, \dots, Y_p$ indépendantes telles que pour tout $i = 1, 2, \dots, p$, $Y_i \sim \mathcal{N}(0, \sigma_i^2)$, où les $\sigma_i^2 = \lambda_i$ sont les valeurs propres deux à deux distinctes Γ . On a donc $X = PY + m$.

En conclusion, si X est un v.a. gaussien dans $\mathbb{R}^n$: $X \sim \mathcal{N}(\vec{m}, \Gamma_X)$, $\vec{m} \in \mathbb{R}^n$ et $\Gamma_X \in \mathcal{M}_n(\mathbb{R})$ symétrique positive, de rang p , il existe une suite de p v.a.r. $Y_1, Y_2, \dots, Y_p$ indépendantes, de loi $\mathcal{N}(0, \lambda_i)$, $i = 1, 2, \dots, p$ respectivement où les λ_i sont les valeurs propres non nulles de Γ_X et une matrice orthogonale P telles que $Y = (Y_1, \dots, Y_p, 0, \dots, 0)$ et $X = PY + m$. Réciproquement, on a

Théorème 1.3

$\forall \vec{m} \in \mathbb{R}^n$ et $\forall \Gamma \in \mathcal{M}_n(\mathbb{R})$ symétrique positive, il existe un v.a. X dans $\mathbb{R}^n$ telle que $X \sim \mathcal{N}(\vec{m}, \Gamma)$.

Vecteurs Gaussiens

Théorème 1.4

Soit X un vecteur gaussien dans $\mathbb{R}^n$ de moyenne $\vec{m}$ et de variance Γ_X .

- ① Si $rg(\Gamma_X) < n$, alors X reste presque sûrement dans un sous-espace affine de $\mathbb{R}^n$ de dimension $rg(\Gamma_X)$.
- ② Si $rg(\Gamma_X) = n$ alors X est continue admettant pour densité de probabilité f_X donnée par

$$f_X(x) = \frac{1}{(2\pi)^{\frac{n}{2}} \sqrt{\det \Gamma_X}} \exp \left\{ -\frac{1}{2} \langle \Gamma_X^{-1}(x - m), (x - m) \rangle \right\}, \quad \forall x \in \mathbb{R}^n.$$

Vecteurs Gaussiens

Remarque 1.1

- ❶ Pour $n = 1$, X gaussienne sur $\mathbb{R}$ de moyenne $m \in \mathbb{R}$ et de variance $\Gamma = \sigma^2$ si et seulement, $\Phi_X(t) = e^{itm} e^{-\frac{1}{2}\sigma^2 t^2}$, $\forall t \in \mathbb{R}$ et
$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(x-m)^2}, \quad \forall x \in \mathbb{R}.$$
- ❷ Pour $\sigma = 0$, $\Phi_X(t) = e^{itm} = \Phi_{\delta_m}(t)$ et $X \rightsquigarrow \delta_m$ où δ_m est la mesure de Dirac en m , par conséquent $X = m$ p.s. On dit que X ou la loi de X est dégénérée.

Vecteurs Gaussiens

Exercice 1.1: Soit X un vecteur gaussien dans $\mathbb{R}^n$, $X = (X_1, \dots, X_n)$ et soit I et J deux parties disjointes de $\{1, 2, \dots, n\}$. Montrer que si $\forall i \in I$ et $\forall j \in J$, $\text{Cov}(X_i, X_j) = 0$ alors $(X_i)_{i \in I}$ et $(X_j)_{j \in J}$ sont indépendantes.

indication

Posons $Y = (X_1, X_2, \dots, X_p, X_{p+1}, \dots, X_n) = (U, V)$ avec $U = (X_1, X_2, \dots, X_p)$ et $V = (X_{p+1}, \dots, X_n)$. Alors $\Gamma_Y = \begin{bmatrix} \Gamma_U & 0 \\ 0 & \Gamma_V \end{bmatrix}$ et $\forall t \in \mathbb{R}^n$, $t = (u, v)$, $u \in \mathbb{R}^p$, $v \in \mathbb{R}^{n-p}$, $\Phi_Y(t) = \Phi_Y(u, v) = \Phi_U(u) \cdot \Phi_V(v)$.

Sommaire

1 Suite de variables aléatoires réelles - Convergence

- Les quatre types de convergence
- Approximations usuelles
- Une application de la loi des grands nombres : Intégrale de Monte-Carlo

Suite de variables aléatoires réelles - Convergence

Définition 1.1: Suite de variables aléatoires réelles - Convergence

Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires réelles définies sur un espace probabilisé $(\Omega, \mathcal{A}, P)$ et X une variable aléatoire réelle. Lorsque n tend vers $+\infty$, on dit que X_n tend ou converge vers X :

- ① en probabilité et on note $X_n \xrightarrow[n \rightarrow \infty]{P} X$ si et seulement si $\forall \varepsilon > 0$,
$$\lim_{n \rightarrow \infty} P[|X_n - X| > \varepsilon] = 0.$$
- ② presque sûrement et on note $X_n \xrightarrow[n \rightarrow \infty]{p.s.} X$ si et seulement si $\forall \varepsilon > 0$,
$$P\left[\limsup_{n \rightarrow \infty} |X_n - X| > \varepsilon\right] = 0.$$
- ③ en moyenne d'ordre p et on note $X_n \xrightarrow[n \rightarrow \infty]{L^p} X$ si et seulement si
$$\lim_{n \rightarrow \infty} E(|X_n - X|^p) = 0.$$

Suite de variables aléatoires réelles - Convergence

Définition 1.2: Convergence en loi

Soit $(X_n)_{n \in \mathbb{N}}$ une suite de variables aléatoires réelles définies sur un espace probabilisé $(\Omega, \mathcal{A}, P)$ et X une variable aléatoire réelle. On dit que X_n tend ou converge vers X en loi lorsque n tend vers $+\infty$ et on note $X_n \xrightarrow[n \rightarrow \infty]{\mathcal{L}} X$ si et seulement si pour toute fonction réelle ou complexe continue et bornée f sur $\mathbb{R}$,

$$\lim_{n \rightarrow \infty} E(f(X_n)) = E(f(X)).$$

Suite de v.a.r., convergence

Remarque 1.1

- i) Pour la convergence en probabilité, l'inégalité de Markov est souvent utilisée pour obtenir une borne qui converge vers 0. D'après cette inégalité, il apparaît que la convergence en moyenne entraîne la convergence en probabilité.
- ii) Pour la convergence presque sûre, il apparaît que le Lemme de Borel-Cantelli est un très bon outil d'étude.
- iii) Pour la convergence en loi, il apparaît que les fonctions caractéristiques sont un très bon outil d'étude et la convergence de la suite des fonctions de répartitions $F_n(x) = P[X_n \leq x]$ dans certains cas.

La convergence p.s. $\Rightarrow$ la convergence en probabilité $\Rightarrow$ la convergence en loi.

Suite de v.a.r., convergence

Théorème 1.1: (Slutsky)

Soit $(X_n)_{n \in \mathbb{N}}$ une suite de v.a.r. et $(Y_n)_{n \in \mathbb{N}}$ une suite de v.a.r.

① Si $X_n \xrightarrow[n \rightarrow \infty]{\mathcal{L}} X$ et f est une fonction continue alors $f(X_n) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} f(X)$.

② Si $X_n \xrightarrow[n \rightarrow \infty]{\mathcal{L}} X$ et $Y_n \xrightarrow[n \rightarrow \infty]{\mathcal{L}} Y$ alors $X_n + Y_n \xrightarrow[n \rightarrow \infty]{\mathcal{L}} X + Y$ et
 $X_n Y_n \xrightarrow[n \rightarrow \infty]{\mathcal{L}} XY$.

Les résultats précédents étant aussi vrais pour la convergence presque sûre et la convergence en probabilité.

③ Si $X_n \xrightarrow[n \rightarrow \infty]{\mathcal{L}} a$ et $Y_n \xrightarrow[n \rightarrow \infty]{\mathcal{L}} Y$ avec $a \in \mathbb{R}$ est une constante, alors
 $X_n + Y_n \xrightarrow[n \rightarrow \infty]{\mathcal{L}} Y + a$ et $X_n Y_n \xrightarrow[n \rightarrow \infty]{\mathcal{L}} aY$.

Suite de v.a.r., convergence

Théorème 1.2: (Paul-Levy)

Soit $(X_n)_{n \in \mathbb{N}}$ une suite de v.a.r. et X une v.a.r. définies sur $(\Omega, \mathcal{A}, P)$. $\Phi_Y(t) = E[e^{itY}]$ désignant la f.c. d'une v.a.r. Y , on a les assertions suivantes.

- ① Si $X_n \xrightarrow[n \rightarrow \infty]{\mathcal{L}} X$ alors $\Phi_{X_n}(t) \longrightarrow \Phi_X(t)$ simplement.
- ② Si $\Phi_{X_n}(t) \longrightarrow \varphi(t)$ simplement et si φ est continue en 0 alors il existe une v.a.r. X tel que $X_n \xrightarrow[n \rightarrow \infty]{\mathcal{L}} X$ et on a $\varphi = \Phi_X$.

En particulier $X_n \xrightarrow[n \rightarrow \infty]{\mathcal{L}} X$ si et seulement si $\Phi_{X_n}(t) \xrightarrow{s} \Phi_X(t)$.

Suite de v.a.r., convergence

Théorème 1.3: (Loi des grands nombres)

Soit $(X_n)_{n \in \mathbb{N}}$ une suite de v.a.r. indépendantes et identiquement distribuées (iid) définies sur $(\Omega, \mathcal{A}, P)$ telles que $m = E(X_1)$ et $\sigma^2 = Var(X_1)$ existent. Soit $S_n = X_1 + X_2 + \cdots + X_n$. Alors lorsque n tend vers $+\infty$

- 1) $\frac{S_n}{n} \xrightarrow{P} m$ (Loi faible des grands nombres : LfGN).
- 2) $\frac{S_n}{n} \xrightarrow{p.s.} m$ (Loi forte des grands nombres : LFGN).

Théorème Central Limite

Théorème 1.4: (TCL)

Soit $(X_n)_{n \in \mathbb{N}}$ une suite de v.a.r. indépendantes et identiquement distribuées (iid) définies sur $(\Omega, \mathcal{A}, P)$ telles que $m = E(X_1)$ et $\sigma^2 = Var(X_1)$ existent.

On pose $S_n = X_1 + X_2 + \cdots + X_n$ et $\bar{X}_n = \frac{S_n}{n}$. Alors

$$\frac{\sqrt{n}(\bar{X}_n - m)}{\sigma} \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, 1).$$

En pratique on fait l'approximation $\frac{\sqrt{n}(\bar{X}_n - m)}{\sigma} \simeq \mathcal{N}(0, 1)$ dès que $n \geq 30$.

Suite de v.a.r., convergence

Proposition 1.1: Convergence en loi

Approximation de la loi binomiale par la loi de Poisson

On montre théoriquement que pour n assez grand (en pratique $n \geq 30$), la loi binomiale $\mathfrak{B}(n, p)$ peut être approximée par la loi de Poisson de paramètre $np : \mathcal{P}(np)$.

validité de l'approximation : $0 < p \leq 0.1$, $n \geq 30$ et $np < 15$, $npq \leq 5$.

Exemple

Soit $X \sim \mathfrak{B}(40, 0.03)$. On a $P[X = 2] = C_{40}^2 (0.03)^2 (0.97)^{38} \# 0.2206$.

Par l'approximation, $\mathfrak{B}(40, 0.03) \simeq \mathcal{P}(40 \cdot 0.03) = \mathcal{P}(1.2)$.

On a, si $Y \sim \mathcal{P}(1.2)$, $P[X = 2] \simeq P[Y = 2] = e^{-1.2} \frac{(1.2)^2}{2!} \# 0.2169 \simeq 0.22$.

Suite de v.a.r., convergence

Proposition 1.2: Convergence en loi

Approximation de la loi binomiale par la loi normale

On établit aussi que pour n assez grand la loi $\mathfrak{B}(n, p)$ peut être approximée par la loi $\mathcal{N}(np, \sqrt{npq})$, valable si $n \geq 30$, $np > 15$ et $npq > 5$.

Correction de continuité : Vu que $\mathfrak{B}(n, p)$ est discrète et $\mathcal{N}(np, \sqrt{npq})$ est continue, on a, en notant F la fonction de répartition de la loi $\mathcal{N}(np, \sqrt{npq})$: $P[X = k] = P[k - 0.5 < X \leq k + 0.5] = F(k + 0.5) - F(k - 0.5)$ (car les décimaux qui s'arrondissent à l'entier k sont ceux de l'intervalle $[k - 0.5, k + 0.5]$)

$$P[X = 0] = P[X < 0.5] = F(0.5)$$

$$P[X = n] = P[X \geq n - 0.5] = 1 - F(n - 0.5).$$

Suite de v.a.r., convergence

Approximation de la loi binomiale par la loi normale

Exemple

Soit $X \sim \mathfrak{B}(40, 0.5)$. On peut faire l'approximation $\mathfrak{B}(40, 0.5) \simeq \mathcal{N}(20, \sqrt{10})$. D'où

$$\begin{aligned} P[X = 20] &= F(20.5) - F(19.5) = \Phi\left(\frac{20.5-20}{\sqrt{10}}\right) - \left(\frac{19.5-20}{\sqrt{10}}\right) \\ &= 2\Phi\left(\underbrace{\frac{0.5}{\sqrt{10}}}_{0.158}\right) - 1 \simeq 2\Phi(0.16) - 1 \simeq 0.1272 \end{aligned}$$

Un calcul direct donne $P[X = 20] = C_{40}^{20}(0.5)^{20}(0.5)^{20} \# 0.1254$.

Suite de v.a.r., convergence

Proposition 1.3: Convergence en loi

Approximation de la loi de Poisson par la loi normale

La loi de Poisson $\mathcal{P}(\lambda)$ peut être approchée par la loi normale $\mathcal{N}(\lambda, \sqrt{\lambda})$ si $\lambda \geq 15$.

Exemple

Soit $X \sim \mathcal{P}(16)$. Les calculs des probabilités concernant X peuvent être effectués en utilisant la loi normale $\mathcal{N}(16, 4)$ car $16 \geq 15$.

e.g.

$$P[X = 16] = F(16.5) - F(15.5) = \Phi\left(\frac{0.5}{4}\right) - \Phi\left(\frac{-0.5}{4}\right) = 2\Phi(0.125) - 1 \approx 0.0995.$$

Calcul direct :

$$P[X = 16] = \frac{e^{-16} 16^{16}}{16!} \approx 0.0992.$$

Intégrale de Monte-Carlo

La méthode de Monte-Carlo est initialement développée par les physiciens qui utilisaient les nombres au hasard pour programmer le calcul d'une intégrale complexe

$$\int_a^b h(x)dx.$$

Supposons que l'on puisse écrire $h(x)$ comme le produit d'une fonction $g(x)$ et d'une densité de probabilité $f(x)$ définie sur l'intervalle $[a, b]$. Alors

$$\int_a^b h(x)dx = \int_a^b g(x)f(x)dx = E_f(g(X))$$

où X est la v.a.r. dont la loi a pour densité f .

Intégrale de Monte-Carlo

Alors la méthode de Monte-Carlo (MMC) consiste à tirer au hasard n nombres $x_1, x_2, \dots, x_n$ selon la loi de X et de faire l'approximation

$$\int_a^b h(x)dx = \int_a^b g(x)f(x)dx = E_f(g(X)) \simeq \frac{1}{n} \sum_{i=1}^n g(x_i).$$

La MMC peut aussi être utilisée pour approximer une distribution marginale lorsque la densité a posteriori $f(y|x)$ est connue :

$$f_Y(y) = \int f(y|x)f_X(x)dx \simeq \frac{1}{n} \sum_{i=1}^n f(y|x_i).$$

Intégrale de Monte-Carlo

Notons que dans la cas où la densité $f(x)$ est approximée par une densité $\hat{f}(x)$ alors

$$\int g(x) f(x) dx = \int g(x) \frac{f(x)}{\hat{f}(x)} \hat{f}(x) dx = E_{\hat{f}} \left(g \frac{f}{\hat{f}}(X) \right) \simeq \frac{1}{n} \sum_{i=1}^n g(x_i) \frac{f(x_i)}{\hat{f}(x_i)}$$

et pour la densité marginale

$$f_Y(y) = \int f(y|x) f_X(x) dx \simeq \frac{1}{n} \sum_{i=1}^n f(y|x_i) \frac{f(x_i)}{\hat{f}(x_i)}$$

où les x_i sont tirées au hasard selon la loi de la v.a.r. X de densité $\hat{f}$.

Sommaire

- 1 Processus aléatoires, chaînes de Markov
 - Ergodicité
 - Chaîne de Markov continues

Introduction

on appelle processus aléatoire une famille de v.a. $X = \{X(t) : t \in T\}$ à valeurs dans un espace $\mathcal{X}$ où T est un ensemble quelconque. Lorsque $T \subset \mathbb{N}$ on dit que le processus X est à temps discret et lorsque T est un intervalle de $\mathbb{R}^+$ on dit que X est un processus à temps continu. Lorsque $\mathcal{X} = \mathbb{R}$, X est dit processus réel. L'application $T \rightarrow \mathcal{X}$, $t \mapsto X(t)$ est appelée trajectoire du processus et pour tout $t \in T$, $\omega \mapsto X(t, \omega)$ est appelé état du processus au temps t . Une suite de v.a.r. $(X_n)_{n \in \mathbb{N}}$ est un exemple simple de processus aléatoire avec $T = \mathbb{N}$ et $\mathcal{X} = \mathbb{R}$. Nous étudierons ici un exemple de processus aléatoire courant : chaîne de Markov classique.

Processus aléatoires, chaînes de Markov

Définition 1.1: Processus aléatoires, chaînes de Markov

Soit $X = (X_n)_{n \geq 0}$ un processus aléatoire à valeurs dans un espace fini ou dénombrable E servant à modéliser l'état d'un système au cours du temps, X_n désignant cet état au temps n . Dans ce cadre, nous dirons que le système se trouve dans l'état $i \in E$ au temps n si $X_n = i$.

On appelle chaîne de Markov, un processus dont l'état futur ne dépend du passé que par le présent. En particulier, on dira qu'un processus aléatoire $X = (X_n)_{n \geq 0}$ est une chaîne de Markov homogène à espace d'états E , si, partant de tout état i de E , la probabilité que la chaîne passe immédiatement à l'état j de E est constante au cours du temps. Plus formellement :

Processus aléatoires, chaînes de Markov

Définition 1.2: Chaînes de Markov homogène

Soit $X = (X_n)_{n \geq 0}$ un processus aléatoire à valeurs dans un espace fini ou dénombrable E . On dira que X est une chaîne de Markov homogène (CMH) si, pour tous états $i, j, i_0, i_1, \dots, i_{n-1} \in E$ et pour tout $n \in N^*$,

$$P[X_{n+1} = j | X_n = i, X_{n-1} = i_{n-1}, \dots, X_0 = i_0] = P[X_{n+1} = j | X_n = i] \quad (1)$$

Les probabilités $P(i, j) = P[X_{n+1} = j | X_n = i] = P[X_1 = j | X_0 = i]$, $(i, j) \in E^2$ sont appelées **probabilités de transition** de la CMH.

Processus aléatoires, chaînes de Markov

Les probabilités de transitions $P_{i,j} = P(i, j)$ vérifient

$$P_{i,j} \geq 0 \quad \text{et} \quad \forall i \in E, \quad \sum_{j \in E} P_{i,j} = 1 \quad (2)$$

Lorsque E est fini, avec $\text{Card}(E) = M$, il est commode de disposer de la matrice de transition $\mathbb{P} = [P_{i,j}]$, $1 \leq i, j \leq M$ de la CMH X . La connaissance de transitions $\mathbb{P}$ d'une CMH $(X_n)_{n \in \mathbb{N}}$ et de la loi initiale de X_0 détermine la loi de toute la chaîne.

Processus aléatoires, chaînes de Markov

Exemple

On admet que le fait qu'il ait plu ou non un jour donné est la seule considération à prendre en compte pour prévoir s'il pleuvra le lendemain. Plus précisément s'il ne pleut pas aujourd'hui, il pleuvra demain avec probabilité α et s'il ne pleut pas aujourd'hui, il ne pleuvra pas demain avec probabilité β , $0 < \alpha, \beta < 1$. En convenant de dire que le système est dans l'état 1 s'il ne pleut pas et dans l'état 2 s'il pleut, la situation peut être représentée par une CMH à deux états $E = \{1, 2\}$ et les probabilités de transitions $P_{11} = P[X_{n+1} = 1 | X_n = 1] = 1 - \alpha$, $P_{12} = P[X_{n+1} = 2 | X_n = 1] = \alpha$, $P_{21} = P[X_{n+1} = 1 | X_n = 2] = \beta$, $P_{22} = P[X_{n+1} = 2 | X_n = 2] = 1 - \beta$. La matrice de transitions est donc

$$\mathbb{P} = \begin{bmatrix} 1 - \alpha & \alpha \\ \beta & 1 - \beta \end{bmatrix}$$

Processus aléatoires, chaînes de Markov

Soit $(X_n)_{n \in \mathbb{N}}$ une CMH. La probabilité de passer d'un état i à un état j en une transition est $P_{i,j} = P[X_{n+1} = j | X_n = i]$. Mais la probabilité que la chaîne passe de i à j en m transitions s'écrit

$$P_{ij}^{(m)} = P^{(m)}(i, j) = P[X_{n+m} = j | X_n = i] \quad (3)$$

e.g.

$$\begin{aligned} P_{ij}^{(2)} &= P^{(2)}(i, j) = P[X_{n+2} = j | X_n = i] = P[X_2 = j | X_0 = i] \\ &= \sum_{k \in E} P[X_2 = j | X_1 = k] P[X_1 = k | X_0 = i] \\ &= \sum_{k \in E} P(k, j) P(i, k) = \sum_{k \in E} P_{ik} P_{kj} \end{aligned}$$

Processus aléatoires, chaînes de Markov

Donc $P_{ij}^{(2)}$ est l'élément de la ligne i et de la colonne j de la matrice $\mathbb{P}^2$. Plus généralement

$$\begin{aligned} P_{ij}^{(m)} &= P^{(m)}(i, j) = P[X_{n+m} = j | X_n = i] = P[X_m = j | X_0 = i] \\ &= \sum_{k \in E} P(k, j) P^{(m-1)}(i, k) = \sum_{k \in E} P_{ik}^{(m-1)} P_{kj} \end{aligned}$$

l'élément de la ligne i et de la colonne j de la matrice $\mathbb{P}^m = \mathbb{P}^{m-1}\mathbb{P} = \mathbb{P}\mathbb{P}^{m-1}$.

Théorème 1.1: (Chapmann-Kolmogorov)

La matrice de transition en n étapes est égale à la puissance n -ième de la matrice de transition en une étape.

Processus aléatoires, chaînes de Markov

Remarque 1.1

Avec l'intérêt des puissances m -ième de la matrice de transition via le Théorème de Chapman-Kolmogorov, on a, pour tout $n \in \mathbb{N}$,

$$P[X_n = j] = \sum_{i \in E} P[X_n = j | X_0 = i] P[X_0 = i] \quad (4)$$

Donc en notant par μ la loi de X_0 (loi initiale de la chaîne), l'équation précédente s'écrit

$$P[X_n = j] = \sum_{i \in E} \mu_i P_{ij}^{(n)} \stackrel{\text{Déf}}{=} \mu \mathbb{P}^n(j), \quad \forall n \geq 1 \quad (5)$$

où $\mu_i = \mu(i) = P[X_0 = i]$, $\forall i \in E$. En d'autres termes $\mathcal{L}(X_n) = \mu \mathbb{P}^n$.

Chaînes de Markov, Ergodicité

Pour une chaîne de Markov homogène, il est intéressant de savoir dans quelle mesure le passé influe sur le futur. Cela revient à s'intéresser au comportement limite des $P_{ij}^{(n)}$ lorsque $n \rightarrow +\infty$. Il apparaît que pour un grand nombre de CMH, la suite $P_{ij}^{(n)}$ converge vers une limite π_j ne dépendant que de l'état j . En d'autres termes, la probabilité de se trouver à l'état j après n transitions pour n assez grand est approximativement π_j , indépendamment de l'état i de départ. On peut montrer qu'une condition suffisante pour qu'une CMH possède cette propriété est qu'il existe un entier $n > 0$ telles que

$$P_{ij}^{(n)} > 0, \quad \forall (i, j) \in E^2 \quad (6)$$

Chaînes de Markov, Ergodicité

Les chaînes de Markov qui vérifient la condition (6) sont dites **ergodiques**. Or d'après le Théorème de Chapman-Kolmogorov

$$P_{ij}^{(n+1)} = \sum_{k \in E} P_{ik}^n P_{kj},$$

Donc lorsque $n \rightarrow +\infty$, les chaînes de Markov ergodiques vérifient $\pi_j = \sum_{k \in E} \pi_k P_{kj}$, $\forall j \in E$, c'est-à-dire

$$\pi \mathbb{P} = \pi \tag{7}$$

Comme $\sum_{j \in E} P_{ij}^n = 1$, on a aussi $\sum_{j \in E} \pi_j = 1$.

Chaînes de Markov, ergodicité

Les résultats précédent sont rassemblés dans le théorème suivant

Théorème 1.2: (Ergodique)

Pour toute chaîne de Markov ergodique $\pi_j = \lim_{n \rightarrow +\infty} P_{ij}^n$ existe et les π_j sont les seules solutions non négatives du système d'équation

$$\begin{cases} \pi = \pi \mathbb{P} \\ \sum_{j \in E} \pi_j = 1 \end{cases} \quad (8)$$

π est dite loi stationnaire de la chaîne de Markov.

Chaînes de Markov, ergodicité

Remarque 1.2

L'étude de la convergence des P_{ij}^n ou d'une manière équivalente, de la matrice $\mathbb{P}^n$ revient à étudier le comportement des valeurs propres de $\mathbb{P}$. En effet $\mathbb{P}$ est semblable à une matrice diagonale ou triangulaire. Donc si $\mathbb{P}$ a une valeur propre λ avec $|\lambda| > 1$, la matrice $\mathbb{P}^n$ ne converge pas. On notera que 1 est valeur propre (vp) triviale de $\mathbb{P}$. La chaîne est donc ergodique si et ssi toutes les valeurs propres de $\mathbb{P}$ sont à l'intérieur du cercle unité.

Chaînes de Markov ergodiques

Exemple 1

Soit la chaîne dont la matrice de transition est $\mathbb{P} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}$.

La matrice semblable à $\mathbb{P}$ est $\mathbb{P}' = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}$. Donc les valeurs propres sont toutes de module 1. $\mathbb{P}^n$ ne peut donc converger et la chaîne n'est pas ergodique. On a d'ailleurs $\mathbb{P}^{2n} = I$ et $\mathbb{P}^{2n+1} = \mathbb{P}$.

2) Soit une chaîne à 4 états 1, 2, 3, et 4, dont la matrice de transition est

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

Chaînes de Markov ergodiques

Exemple 1 suite

Alors $\mathbb{P} = Q\mathbb{P}'Q^{-1}$ avec $\mathbb{P}' = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \frac{1}{2} & 0 \\ 0 & 0 & 0 & -\frac{1}{2} \end{pmatrix}$ et $Q = \begin{pmatrix} 1 & 0 & 0 & 0 \\ \frac{2}{3} & \frac{1}{3} & 1 & 1 \\ \frac{1}{3} & \frac{2}{3} & 1 & -1 \\ 0 & 1 & 0 & 0 \end{pmatrix}$

avec $Q^{-1} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ -\frac{1}{2} & \frac{1}{2} & \frac{1}{2} & -\frac{1}{2} \\ -\frac{1}{6} & \frac{1}{2} & -\frac{1}{2} & \frac{1}{6} \end{pmatrix}$. On a donc

$$\mathbb{P}^n = Q\mathbb{P}'^nQ^{-1} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ \frac{2}{3} & 0 & 0 & \frac{1}{3} \\ \frac{1}{3} & 0 & 0 & \frac{2}{3} \\ 0 & 0 & 0 & 1 \end{pmatrix} + \frac{1}{2^{n+1}} \begin{pmatrix} 0 & 0 & 0 & 0 \\ -1 & 1 & 1 & -1 \\ -1 & 1 & 1 & -1 \\ 0 & 0 & 0 & 0 \end{pmatrix} + \frac{(-1)^n}{2^{n+1}} \begin{pmatrix} 0 & 0 & 0 & 0 \\ -\frac{1}{3} & 1 & -1 & \frac{1}{3} \\ \frac{1}{3} & -1 & 1 & -\frac{1}{3} \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

Chaînes de Markov ergodiques

Exemple 1 suite

Par conséquent

$$\pi = \lim_{n \rightarrow +\infty} \mathbb{P}^n = \begin{pmatrix} 1 & 0 & 0 & 0 \\ \frac{2}{3} & 0 & 0 & \frac{1}{3} \\ \frac{1}{3} & 0 & 0 & \frac{2}{3} \\ 0 & 0 & 0 & 1 \end{pmatrix}.$$

Interprétation :

Si la chaîne se trouve initialement à l'une des extrémités du segment $[1, 4]$, elle y reste presque sûrement. Si elle se trouve à l'état 2, elle a deux chances sur trois d'aboutir tôt ou tard à l'état 1, et une chance sur trois d'aboutir à l'état extrémité 4. Par contre, elle n'a aucune chance de se stabiliser en un état intermédiaire. On a un résultat analogue si le point de départ est l'état 3.

Chaînes de Markov ergodiques

Interprétation suite :

On peut aussi dire que la chaîne est attirée par les extrémités absorbantes 1 et 4, qu'elle finit toujours par rejoindre. Elle a naturellement plus de chance d'aboutir à l'extrémité la plus proche de sa position initiale.

Exemple 2

Soit une chaîne de Markov à trois états 1, 2 et 3, dont la matrice de transitions est

$$\mathbb{P} = \begin{pmatrix} \frac{1}{2} & \frac{1}{4} & \frac{1}{4} \\ \frac{1}{3} & \frac{2}{3} & \frac{1}{6} \\ \frac{1}{5} & \frac{1}{10} & \frac{1}{10} \end{pmatrix}.$$

La diagonalisation de $\mathbb{P}$ donne comme valeurs propres $\lambda_1 = 1$, $\lambda_2 = \frac{1}{2}$ et $\lambda_3 = \frac{1}{5}$. La matrice de passage Q et son inverse sont données par :

Chaînes de Markov ergodiques

Exemple 2 suite

$$Q = \begin{pmatrix} 1 & 1 & -5 \\ 1 & 2 & 5 \\ 1 & -2 & 1 \end{pmatrix} \quad \text{et} \quad Q^{-1} = \begin{pmatrix} \frac{4}{12} & \frac{3}{12} & \frac{5}{12} \\ \frac{2}{18} & \frac{3}{18} & \frac{-5}{18} \\ \frac{-4}{36} & \frac{3}{36} & \frac{1}{36} \end{pmatrix}.$$

On a donc

$$\lim_{n \rightarrow +\infty} \mathbb{P}^n = \lim_{n \rightarrow +\infty} Q \mathbb{P}'^n Q^{-1} = \begin{pmatrix} \frac{4}{12} & \frac{3}{12} & \frac{5}{12} \\ \frac{4}{12} & \frac{3}{12} & \frac{5}{12} \\ \frac{4}{12} & \frac{3}{12} & \frac{5}{12} \end{pmatrix}.$$

Chaînes de Markov ergodiques

Exemple 2, interprétation

La chaîne en question comporte trois éventualités. Notons les E_1 , E_2 et E_3 .

Lorsque la chaîne est en E_1 elle a les probabilités $\frac{1}{2}$, $\frac{1}{4}$ et $\frac{1}{4}$ de rester en E_1 ou d'aller à E_2 ou à E_3 respectivement, etc.

Lorsque la chaîne fait un grand nombre n de transitions à partir d'un instant k , les probabilités que la chaîne soit en E_1 , E_2 et E_3 respectivement à l'instant $n + k$, sont lorsque $n \rightarrow +\infty$, $\frac{4}{12}$, $\frac{3}{12}$, $\frac{5}{12}$ respectivement. Ces probabilités sont bien indépendantes des états de départ à l'instant k considéré.

Chaînes de Markov continues

On dit que $X = (X_n)_{n \in \mathbb{N}}$ est une chaîne de Markov homogène continue si en plus d'avoir les propriétés de Markov homogène précédentes, les X_n sont des variables aléatoires réelles continues et de même loi admettant une densité de probabilité notée $f(x)$. La loi de probabilité conditionnelle de X_{n+k} sachant $X_n = x$ admet donc une densité $f^{(k)}(y|x)$ indépendante de n . Enfin la loi conditionnelle de X_{n+k} ($k \geq 1$) sachant les valeurs prise par les X_i , $i \leq n$ dépend uniquement de la valeur récente prise par X_n . La loi du couple (X_n, X_{n+k}) est alors définie par sa densité

$$f^{(k)}(y|x)f(x). \quad (9)$$

De même par la règle de la multiplication, celle du triplet $(X_n, X_{n+k}, X_{n+k+m})$ est définie par la densité

$$f^{(m)}(z|y)f^{(k)}(y|x)f(x). \quad (10)$$

Chaînes de Markov continues

Il en résulte que la loi conditionnelle de X_{n+k+m} sachant $X_n = x$ est donnée par

$$f^{(k+m)}(z|x) = \int_{-\infty}^{+\infty} f^{(m)}(z|y) f^{(k)}(y|x) dy. \quad (11)$$

En particulier $f^{(1)}(y|x) = f(y|x)$ est la densité (ou noyau) de transition de x à y en une étape et plus généralement

$$f^{(n+1)}(z|x) = \int_{-\infty}^{+\infty} f^{(n)}(y|x) f(z|y) dy. \quad (12)$$

Exemple

Supposons que le passage de X_n à X_{n+1} suive une loi normale et que chacun des X_n obéisse à une loi normale réduite.

Chaînes de Markov continues

Posons

$$f(y|x) = \frac{1}{\sqrt{2\pi}\sqrt{1-r^2}} e^{-\frac{1}{2} \frac{(y-rx)^2}{1-r^2}}.$$

Cherchons $f^{(k)}(z|x)$. On a d'abord

$$f^{(2)}(z|x) = \int_{-\infty}^{+\infty} f(z|y)f(y|x)dy = \frac{1}{2\pi(1-r^2)} \int_{-\infty}^{+\infty} e^{-\frac{1}{2} \frac{(y-rx)^2 + (z-ry)^2}{1-r^2}} dy.$$

On obtient avec les résultats sur l'intégrale de Gauss

$$f^{(2)}(z|x) = \frac{1}{\sqrt{2\pi}\sqrt{1-r^2}} e^{-\frac{1}{2} \frac{(z-r^2x)^2}{1-r^2}}.$$

Chaînes de Markov continues

On établit enfin par récurrence sur n que $f^{(n)}(z|x)$ s'obtient en remplaçant r par r^n dans l'expression de $f(z|x)$, r étant le coefficient de corrélation entre X_n et X_{n+1} . Si $|r| < 1$, $r^n \rightarrow 0$ quand $n \rightarrow +\infty$ et donc

$$f^{(n)}(z|x) \rightarrow f(z) = \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}.$$

On retrouve la propriété ergodique étudiée dans le cas discret, mais cette propriété ergodique n'est pas assurée si $r = \pm 1$.

Si $r = 1$ par exemple, $f^{(n)}(z|x)$ ne dépend plus de n . Dans ce cas, on $X_{n+1} = X_n$ presque sûrement. X_1 est aléatoire mais son choix étant fait, les X_n successifs sont bien déterminés. De même si $r = -1$, on a $X_2 = -X_1$, $X_3 = X_1$, $X_4 = -X_1$, etc.

Processus de Markov

Notons que l'espace d'états E d'une chaîne de Markov au lieu d'être discret, peut dépendre d'un temps continu t . Au lieu donc de X_n , on écrit X_t ou $X(t)$ et $X(t)$ est une fonction aléatoire de t . $X(t)_{t \in T}$ est un processus de Markov si la loi de $X(t+h)$ sachant $X(s)$, $s \leq t$ est identique à la loi de $X(t+h)$ sachant $X(t)$, $\forall h > 0$. Les relations de Chapman-Kolmogorov

$$\begin{aligned} P_{ij}^{n+m} &= \sum_{k \in E} P_{ik}^n P_{kj}^m \\ f^{\alpha+\beta}(z|x) &= \int_{-\infty}^{+\infty} f^{\alpha}(z|y) f^{\beta}(y|x) dy \end{aligned} \quad (13)$$

restent valables, mais ici α et β sont des variables non négatives continues qui peuvent être aussi dérivables. Moyennant quelques hypothèses de régularité, la relation fonctionnelle intégrale ci-dessus peut alors être remplacée par une équation aux dérivées partielles de Kolmogorov. Ce qui est du niveau Master 2.