Université Pierre & Marie Curie (Paris 6)

Licence (S5)

UE 3M245 "Probabilités élémentaires"

Année 2019–20

Probabilités sans Théorie de la Mesure

Amaury Lambert[1]

[1]Responsable de l'UE 3M245 de 2014–15 à 2019–20. Mél: amaury.lambert@upmc.fr

Contents

Chapter 1

Un peu de théorie des ensembles

1.1 La droite achevée

Définition 1.1 *On appelle* droite achevée *l'ensemble* $\bar{\mathbb{R}} := \mathbb{R} \cup \{-\infty\} \cup \{+\infty\}$*.* $\bar{\mathbb{R}}$ *est muni de la topologie usuelle de* $\mathbb{R}$*, complétée avec les notions usuelles de convergence vers* $+\infty$ *et vers* $-\infty$*.*

La droite achevée est également munie des opérations algébriques usuelles, avec les conventions suivantes :

$$+\infty + \infty = +\infty, \quad -\infty - \infty = -\infty, \quad a + \infty = +\infty, \quad a - \infty = -\infty,$$

pour tout $a \in \mathbb{R}$, ainsi que

$$0 \times \infty = 0,$$

et

$$a \in]0, \infty] \;\Rightarrow\; a \times \infty = +\infty, \quad a \in [-\infty, 0[\;\Rightarrow\; a \times \infty = -\infty.$$

Remarque 1.2 *Tout au long de ce cours, il faudra acquérir le réflexe de* NE JAMAIS ÉCRIRE *aucune des opérations interdites suivantes :* $(+\infty) - (+\infty)$*, ainsi que* $(-\infty) - (-\infty)$*, et encore* $(\pm\infty)/(\pm\infty)$*.*

1.2 Rappels sur les suites et séries numériques

Une suite numérique est une suite à valeurs dans $\mathbb{R}$ ou dans $\bar{\mathbb{R}}$.

Notation 1.3 *Lorsqu'une suite* (u_n) *est croissante (resp. décroissante), on notera souvent* $\lim_n \uparrow u_n$ *(resp.* $\lim_n \downarrow u_n$*) sa limite, pour signifier que la suite est monotone, et surtout pour indiquer que cette limite existe* donc *toujours (dans* $\bar{\mathbb{R}}$*).*

Définition 1.4 *On dit que la série de terme général* (u_n) *est* absolument convergente *si la suite des sommes partielles* $(\sum_{k=0}^{n} |u_k|)_n$ *converge dans* $\mathbb{R}$*, ce que l'on note également* $\sum_n |u_n| < \infty$*.*

Théorème 1.5 *Si la série de terme général (u_n) est absolument convergente, alors elle est convergente, c'est-à-dire que la suite des sommes partielles $(\sum_{k=0}^{n} u_k)_n$ converge dans $\mathbb{R}$.*

Proposition 1.6 *La* somme *de la série de terme général $u_n \geq 0$ (c'est-à-dire la limite de la suite des sommes partielles, qui existe toujours dans $\bar{\mathbb{R}}_+$) ne dépend pas de l'ordre de sommation.*

Démonstration. Soit une bijection $\varphi : \mathbb{N} \longrightarrow \mathbb{N}$. On veut montrer que la suite $S'_n := \sum_{k=0}^{n} u_{\varphi(k)}$ a même limite dans $\bar{\mathbb{R}}_+$ que $S_n := \sum_{k=0}^{n} u_k$.

Soit $n \geq 0$ et $N := \max\{\varphi(0), \dots, \varphi(n)\}$. Alors $S'_n = u_{\varphi(0)} + \cdots + u_{\varphi(n)} \leq \sum_{j=0}^{N} u_j = S_N$, donc $S'_n \leq S_N \leq S_\infty$. Faisant tendre $n \to \infty$ on obtient que $S'_\infty \leq S_\infty$. L'inégalité opposée s'obtient par symétrie. □

Proposition 1.7 *Soit $(a_{n,m})$ une suite doublement indicée de nombres réels positifs. Alors on a l'égalité suivante dans $\bar{\mathbb{R}}_+$*

$$\sum_{n=0}^{\infty} \sum_{m=0}^{\infty} a_{n,m} = \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} a_{n,m}.$$

Démonstration. Soient $A_j := \sum_{k=0}^{\infty} a_{j,k}$ et $B_k := \sum_{j=0}^{\infty} a_{j,k}$. Il faut montrer que $\sum_{j=0}^{\infty} A_j = \sum_{k=0}^{\infty} B_k$. Soit $S_{n,m} := \sum_{j=0}^{n} \sum_{k=0}^{m} a_{j,k}$. Comme on peut intervertir l'ordre des sommations pour les sommes finies et que le terme général de la suite est positif on a pour tous n et m l'inégalité

$$S_{n,m} = \sum_{k=0}^{m} \sum_{j=0}^{n} a_{j,k} \leq \sum_{k=0}^{m} \sum_{j=0}^{\infty} a_{j,k} = \sum_{k=0}^{m} B_k \leq \sum_{k=0}^{\infty} B_k.$$

D'autre part, on peut intervertir somme finie et passage à la limite, donc

$$\lim_{m\to\infty} S_{n,m} = \sum_{j=0}^{n} \lim_{m\to\infty} \sum_{k=0}^{m} a_{j,k} = \sum_{j=0}^{n} A_j.$$

On obtient donc l'inégalité $\sum_{j=0}^{n} A_j \leq \sum_{k=0}^{\infty} B_k$, et en passant à la limite en n,

$$\sum_{j=0}^{\infty} A_j \leq \sum_{k=0}^{\infty} B_k.$$

L'inégalité opposée s'obtient par symétrie, ce qui implique l'égalité cherchée. □

1.3 Parties et suites de parties

Soit Ω un ensemble. Un sous-ensemble de Ω est aussi appelé *partie* de Ω et sera dès le prochain chapitre appelé *évènement*. On note $\mathscr{P}(\Omega)$ l'ensemble des parties de Ω. Pour toute partie A de Ω, on note ${}^cA$ le *complémentaire* de A dans Ω, défini par ${}^cA := \{x \in \Omega : x \notin A\}$.

Définition 1.8 *Pour toute partie A de Ω, on appelle* indicatrice *ou* fonction indicatrice *de A, et l'on note $\mathbb{1}_A$, la fonction*

$$\begin{aligned} \mathbb{1}_A : \Omega &\longrightarrow \{0,1\} \\ x &\longmapsto \begin{cases} 0 & si \quad x \notin A \\ 1 & si \quad x \in A. \end{cases} \end{aligned}$$

Remarque 1.9 *Noter que $\mathbb{1}_{{}^cA} = 1 - \mathbb{1}_A$.*

Définition 1.10 *Soient E un ensemble et $f : \Omega \longrightarrow E$. Pour tout $B \in \mathscr{P}(E)$, on note $f^{-1}(B)$ l'*image réciproque *de B par f :*

$$f^{-1}(B) := \{x \in \Omega : f(x) \in B\}.$$

Notation 1.11 *On notera souvent $\{x \in \Omega : f(x) \in B\}$ sous la forme condensée $\{f \in B\}$, autrement dit*

$$\{f \in B\} := f^{-1}(B).$$

Remarque 1.12 *Noter que $\mathbb{1}_A^{-1}(\{1\}) = A$ et $\mathbb{1}_A^{-1}(\{0\}) = {}^cA$.*

Théorème 1.13 *L'ensemble $\{0,1\}^\Omega$ des fonctions de Ω à valeurs dans $\{0,1\}$ est en bijection avec $\mathscr{P}(\Omega)$.*

Démonstration. L'application

$$\begin{aligned} \psi : \mathscr{P}(\Omega) &\longrightarrow \{0,1\}^\Omega \\ A &\longmapsto \mathbb{1}_A \end{aligned}$$

est clairement une bijection, avec application réciproque[1] ψ^{-1} donnée par $\psi^{-1}(f) = f^{-1}(\{1\})$. □

Exercice 1.14 *Montrer les formules de Hausdorff suivantes. Pour tout J ensemble d'indices non vide, pour toute famille $(B_j)_{j\in J}$ de parties de E, pour toute fonction $f : \Omega \longrightarrow E$,*

$$f^{-1}\left(\bigcup_{j\in J} B_j\right) = \bigcup_{j\in J} f^{-1}(B_j),$$

$$f^{-1}\left(\bigcap_{j\in J} B_j\right) = \bigcap_{j\in J} f^{-1}(B_j),$$

et pour tout $B \subseteq E$,

$${}^c\big(f^{-1}(B)\big) = f^{-1}\left({}^cB\right).$$

[1] Attention, ici ψ^{-1} désigne l'application inverse de l'application bijective ψ, tandis que $f^{-1}(\{1\})$ désigne l'image réciproque de $\{1\}$. Dans la suite du cours, il n'y aura plus d'ambiguïté car seul le second usage de cette notation nous sera utile.

Recensons quelques opérations classiques sur les parties d'un ensemble E. Soient A_1 et A_2 deux parties de E.

- *La réunion* de A_1 et A_2, notée $A_1 \cup A_2$: $\forall x \in E$,
$$x \in A_1 \cup A_2 \Leftrightarrow \exists i \in \{1, 2\}, x \in A_i$$
- *L'intersection* de A_1 et A_2, notée $A_1 \cap A_2$: $\forall x \in E$,
$$x \in A_1 \cap A_2 \Leftrightarrow \forall i \in \{1, 2\}, x \in A_i$$
- *La différence* de A_1 avec A_2, notée $A_1 \setminus A_2$ et dite *différence propre* dans le cas où $A_2 \subseteq A_1$: $\forall x \in E$,
$$x \in A_1 \setminus A_2 \Leftrightarrow x \in A_1 \text{ et } x \notin A_2$$

Remarque 1.15 *Remarquer l'association de la réunion avec le quantificateur "$\exists$", de l'intersection avec le quantificateur "$\forall$", ainsi que l'association du passage au complémentaire avec la négation et de l'inclusion avec l'implication : $A_1 \subseteq A_2$ ssi $\forall x \in E$,*
$$x \in A_1 \Rightarrow x \in A_2.$$

Exercice 1.16 *Montrer les identités suivantes :*
$$^c(A_1 \cup A_2) = {}^cA_1 \cap {}^cA_2$$
$$^c(A_1 \cap A_2) = {}^cA_1 \cup {}^cA_2$$
$$A_1 \setminus A_2 = A_1 \cap {}^cA_2$$

Nous allons définir la notion de limite d'une suite de parties lorsque cette suite est monotone, c'est-à-dire croissante ou décroissante au sens de l'ordre partiel induit par la relation d'inclusion. Soit (A_n) une suite de parties de Ω.

Définition 1.17 *La suite (A_n) est dite* croissante *lorsque pour tout entier n, $A_n \subseteq A_{n+1}$.*

La suite (A_n) est dite décroissante *lorsque pour tout entier n, $A_{n+1} \subseteq A_n$.*

Dans les deux cas, on dira que la suite (A_n) est monotone.

Lorsque la suite (A_n) est croissante, la limite de (A_n) est définie naturellement comme la réunion *de tous les A_n :*
$$\lim_{n\to\infty} A_n := \bigcup_n A_n$$
Par analogie avec le cas réel, on notera cette limite $\lim_n \uparrow A_n$ pour faire référence au fait que la suite (A_n) est croissante et que donc sa limite existe[2]

*Lorsque la suite (A_n) est décroissante, la limite de (A_n) est définie naturellement comme l'*intersection *de tous les A_n :*
$$\lim_{n\to\infty} A_n := \bigcap_n A_n$$
Par analogie avec le cas réel, on notera cette limite $\lim_n \downarrow A_n$ pour faire référence au fait que la suite (A_n) est décroissante et que donc sa limite existe.

[2]La notion de limite de suite non monotone de parties existe mais n'est pas au programme. Quoiqu'il en soit, comme dans le cas réel, toutes les suites de parties n'admettent pas de limite.

1.4 Dénombrabilité

Définition 1.18 *L'ensemble Ω est dit* dénombrable *s'il existe une* injection $\varphi : \Omega \longrightarrow \mathbb{N}$.

Remarque 1.19 *L'idée derrière cette définition est de généraliser aux ensembles infinis la notion de cardinal. Si E et F sont deux ensembles finis,* $\mathrm{Card}(E) \leq \mathrm{Card}(F)$ *si et seulement si il existe une injection $\varphi : E \longrightarrow F$, autrement dit, E n'est 'pas plus gros' que F. Un ensemble est dénombrable s'il n'est 'pas plus gros' que $\mathbb{N}$.*

Remarque 1.20 *La définition d'ensemble infini peut également s'exprimer en ces termes. Un ensemble Ω est dit* infini *s'il existe une injection $\psi : \mathbb{N} \longrightarrow \Omega$. Le résultat suivant est une conséquence du théorème de Cantor–Bernstein*[3].

Définition 1.21 *L'ensemble Ω est infini et dénombrable si et seulement si il existe une* bijection $\varphi : \Omega \longrightarrow \mathbb{N}$.

Exemple 1.22 *L'ensemble* $2\mathbb{N}$ *des entiers pairs est infini dénombrable car l'application $n \mapsto 2n$ est une bijection de $\mathbb{N}$ sur $2\mathbb{N}$.*

Proposition 1.23 *Si Ω' est un ensemble dénombrable et s'il existe une injection $\varphi : \Omega \longrightarrow \Omega'$, alors Ω est dénombrable.*

Dém. Par hypothèse, Ω' est dénombrable, donc il existe une injection $\varphi' : \Omega' \longrightarrow \mathbb{N}$. Par conséquent $\varphi' \circ \varphi : \Omega \longrightarrow \mathbb{N}$ est une injection, ce qui prouve que Ω est dénombrable. □

Proposition 1.24 *Pour tout $n \in \mathbb{N}^\star$, le produit cartésien $\mathbb{N}^n$ est dénombrable.*

Dém. Pour $n = 2$, on peut exhiber une énumération $\varphi : \mathbb{N}^2 \longrightarrow \mathbb{N}$ en suivant du bas vers le haut les points successifs des droites d'équation $y = -x + k$, lorsque k croît dans $\mathbb{N}$, ce qui donne $\varphi(n, k) = k + (n + k)(n + k + 1)/2$. On vérifie facilement que cette application est bien injective, ce qui montre que $\mathbb{N}^2$ est dénombrable. On démontre ensuite par récurrence sur n que $\mathbb{N}^n$ est dénombrable. Soit $n \geq 3$ tel que $\mathbb{N}^{n-1}$ soit dénombrable. Alors la fonction

$$\begin{array}{rrcl} f_n : & \mathbb{N}^n & \longrightarrow & \mathbb{N}^{n-1} \\ & (x_1, \ldots, x_n) & \longmapsto & (\varphi(x_1, x_2), x_3, \ldots, x_n) \end{array}$$

est clairement une injection. D'après l'hypothèse de récurrence et la proposition précédente, $\mathbb{N}^n$ est dénombrable. □

La proposition suivante est une conséquence des deux précédentes.

Proposition 1.25 *Tout produit cartésien fini d'ensembles dénombrables est dénombrable.*

[3] Ce théorème classique assure que deux ensembles sont en bijection dès qu'ils s'injectent l'un dans l'autre.

Dém. Pour $i = 1, \dots, n$, soit Ω_i dénombrable et une injection $\varphi_i : \Omega_i \to \mathbb{N}$. Alors la fonction

$$\begin{aligned} \phi : \quad \Pi_{i=1}^n \Omega_i &\longrightarrow \mathbb{N}^n \\ (x_1, \dots, x_n) &\longmapsto (\varphi_1(x_1), \dots, \varphi_n(x_n)) \end{aligned}$$

est clairement injective donc $\Pi_i \Omega_i$ est dénombrable, car il s'injecte dans l'ensemble dénombrable $\mathbb{N}^n$. $\square$

Proposition 1.26 *Les ensembles $\mathbb{Z}$ et $\mathbb{Q}$ sont dénombrables.*

Dém. Pour ce qui est de $\mathbb{Z}$, la fonction

$$\begin{aligned} f : \mathbb{Z} &\longrightarrow \mathbb{N} \\ n &\longmapsto \begin{cases} -2n & \text{si} \quad n \leq 0 \\ 2n - 1 & \text{si} \quad n > 0 \end{cases} \end{aligned}$$

est une bijection.

Enfin, rappelons que pour tout $x \in \mathbb{Q}^\star$, $\exists!(p, q) \in \mathbb{Z}^\star \times \mathbb{N}^\star$ tel que $x = p/q$ et $p \wedge q = 1$. Ainsi la fonction qui à 0 associe $(0, 1)$ et qui est définie sur $\mathbb{Q}^\star$ par

$$\begin{aligned} f : \ \mathbb{Q}^\star &\longrightarrow \mathbb{Z} \times \mathbb{N}^\star \\ p/q &\longmapsto (p, q) \end{aligned}$$

est une injection, donc $\mathbb{Q}$ s'injecte dans $\mathbb{Z} \times \mathbb{N}^\star$, qui est dénombrable d'après la proposition précédente. $\square$

Proposition 1.27 *Toute réunion dénombrable d'ensembles dénombrables est dénombrable.*

Dém. Soit $\Omega = \bigcup_{n \in \mathbb{N}} \Omega_n$, où pour tout $n \in \mathbb{N}$, Ω_n est dénombrable. Alors par définition, pour tout $n \in \mathbb{N}$ il existe une injection $\varphi_n : \Omega_n \to \mathbb{N}$. Pour tout $x \in \Omega$ on définit alors

$$N(x) := \min\{n \geq 0 : x \in \Omega_n\} < \infty.$$

Alors la fonction

$$\begin{aligned} \phi : \Omega &\longrightarrow \mathbb{N}^2 \\ x &\longmapsto (N(x), \varphi_{N(x)}(x)) \end{aligned}$$

est une injection car pour tous $x, y \in \Omega$ tels que $\phi(x) = \phi(y)$, on a $N(x) = N(y) =: n$ puis $\varphi_{N(x)}(x) = \varphi_{N(y)}(y)$, c'est-à-dire $\varphi_n(x) = \varphi_n(y)$, donc $x = y$, puisque φ_n est injective. Par conséquent, Ω s'injecte dans $\mathbb{N}^2$ et est donc dénombrable. $\square$

Après cette avalanche de résultats exhibant tant d'ensembles dénombrables, on pourrait se demander si tous les ensembles ne sont pas dénombrables. La proposition suivante assure que ce n'est pas le cas.

Proposition 1.28 *Soit Ω un ensemble non vide et $f : \Omega \to \mathscr{P}(\Omega)$. Alors f n'est pas surjective.*

En particulier, $\mathbb{N}$ et $\mathscr{P}(\mathbb{N})$ ne sont pas en bijection, donc $\mathscr{P}(\mathbb{N})$ n'est pas dénombrable.

Dém. Soit $f : \Omega \to \mathscr{P}(\Omega)$. Montrons que f ne peut être surjective. Soit

$$A := \{x \in \Omega : x \notin f(x)\}.$$

Montrons par l'absurde que A ne peut avoir d'antécédent par f. Si $\exists z \in \Omega$ tel que $f(z) = A$ alors

- soit $z \in A$ alors $z \notin f(z)$, c'est-à-dire $z \notin A$;
- soit $z \notin A$ alors $z \in f(z)$, c'est-à-dire $z \in A$,

ce qui constitue une contradiction. Noter qu'il existe clairement une injection de Ω dans $\mathscr{P}(\Omega)$, par exemple celle qui à x associe $\{x\}$.

Pour conclure, rappelons qu'un ensemble infini comme $\mathscr{P}(\mathbb{N})$ est dénombrable ssi il est en bijection avec $\mathbb{N}$. Comme il n'y a pas de surjection de $\mathbb{N}$ dans $\mathscr{P}(\mathbb{N})$, il n'y a pas de bijection de $\mathbb{N}$ dans $\mathscr{P}(\mathbb{N})$, et donc $\mathscr{P}(\mathbb{N})$ n'est pas dénombrable. $\square$

Théorème 1.29 *Les ensembles $\mathbb{R}$, $\{0,1\}^{\mathbb{N}}$ et $\mathscr{P}(\mathbb{N})$ sont en bijection et ne sont donc pas dénombrables. On dit qu'ils ont la* puissance du continu.

Nous avons déjà vu que $\{0,1\}^{\mathbb{N}}$ et $\mathscr{P}(\mathbb{N})$ sont en bijection. La bijection avec $\mathbb{R}$ est admise.

Chapter 2

Probabilité, variables aléatoires discrètes

2.1 Probabilité

Définition 2.1 *Soit Ω un ensemble quelconque*[1] *et $\mathscr{A} := \mathscr{P}(\Omega)$. On appelle* probabilité *une fonction $P : \mathscr{A} \longrightarrow [0,1]$ telle que*

(i) $P(\varnothing) = 0$ et $P(\Omega) = 1$;

*(ii) P est σ-*additive *: pour toute suite $(A_n)_{n\in\mathbb{N}}$ d'éléments de $\mathscr{A}$ deux à deux disjoints,*

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

On dit que $(\Omega, \mathscr{A}, P)$ est un espace probabilisé *ou* espace de probabilité *et pour tout $A \in \mathscr{A}$, on appelle A un* évènement *et $P(A)$ la probabilité de A.*

Remarque 2.2 *Un élément ω de Ω est vu comme le* résultat d'une expérience*, ainsi un évènement A est un ensemble de résultats possibles.*

Notation 2.3 *Deux évènements A_1 et A_2 tels que $A_1 \cap A_2 = \varnothing$ sont indifféremment dits disjoints (comme dans l'énoncé du théorème) ou* incompatibles. *La réunion $\bigcup_{n\in\mathbb{N}} A_n$ mentionnée dans le théorème s'écrit parfois également*

$$\bigsqcup_{n} A_n$$

pour signifier que les A_n sont deux à deux incompatibles.

Remarque 2.4 *Rigoureusement, une probabilité n'est souvent définie que sur un sous-ensemble $\mathscr{A}$ de $\mathscr{P}(\Omega)$ appelé* tribu*, notion issue de la théorie de la mesure qui est ici hors programme. Dans ce cours on notera $\mathscr{A}$ à la place de $\mathscr{P}(\Omega)$ pour garder en mémoire que $\mathscr{A}$ ne contient pas en général toutes les parties de Ω. En pratique on prendra $\mathscr{A} = \mathscr{P}(\Omega)$, ce qui d'ailleurs est tout à fait possible dès que Ω est dénombrable.*

[1] par quelconque, on veut en fait dire : 'pas forcément dénombrable'...

Remarque 2.5 *L'égalité $P(\cup_n A_n) = \sum_n P(A_n)$ ne dépend pas de l'énumération des (A_n), autrement dit, pour toute bijection $\varphi : \mathbb{N} \to \mathbb{N}$, chaque membre de l'égalité reste inchangé si l'on remplace (A_n) par $(A_{\varphi(n)})$ (l'ordre de sommation dans le membre de droite n'intervient pas car la série est à termes positifs).*

Proposition 2.6 *Une probabilité P sur $(\Omega, \mathscr{A})$ vérifie pour tous $A, B \in \mathscr{A}$:*

(i) Additivité finie : $P(A) = P(A \setminus B) + P(A \cap B)$ et plus généralement, pour tout n-uplet $(A_i)_{i=1,\ldots,n}$ d'éléments de $\mathscr{A}$ deux à deux disjoints, $P(\sqcup_{i=1}^n A_i) = \sum_{i=1}^n P(A_i)$;

(ii) Additivité forte : $P(A \cup B) = P(A) + P(B) - P(A \cap B)$;

(iii) Croissance : si $A \subseteq B$, $P(A) \leq P(B)$.

Dém. (i) Montrons d'abord la seconde égalité grâce à la σ-additivité. Si l'on définit $A_i = \varnothing$ pour tout $i \geq n+1$, alors $P(\sqcup_{i=1}^n A_i) = P(\sqcup_{i=1}^\infty A_i) = \sum_{i \geq 1} P(A_i) = \sum_{i=1}^n P(A_i)$. Pour la première égalité, $A \setminus B$ et $A \cap B$ sont disjoints et leur réunion est A.

(ii) $A \setminus B$, $A \cap B$ et $B \setminus A$ sont disjoints et leur réunion est $A \cup B$, donc

$$P(A \setminus B) + P(A \cap B) + P(B \setminus A) = P(A \cup B),$$

donc en ajoutant $P(A \cap B)$ à chaque membre on obtient

$$P(A \setminus B) + P(A \cap B) + P(B \setminus A) + P(A \cap B) = P(A \cup B) + P(A \cap B),$$

mais dans le premier membre, grâce à (i), la somme des deux premiers termes vaut $P(A)$ et la somme des deux derniers termes vaut $P(B)$.

(iii) D'après (i) si $A \subseteq B$, alors

$$P(B) = P(B \setminus A) + P(B \cap A) = P(B \setminus A) + P(A) \geq P(A),$$

qui est l'inégalité souhaitée. $\square$

Proposition 2.7 *Soit $P : \mathscr{A} \longrightarrow [0,1]$ une probabilité. Alors,*

(i) P est continue à gauche[2] *: pour toute suite* croissante *$(A_n)_{n \in \mathbb{N}}$ d'évènements,*

$$P(\lim_n \uparrow A_n) = \lim_n \uparrow P(A_n).$$

(ii) P est continue à droite *: pour toute suite* décroissante *$(B_n)_{n \in \mathbb{N}}$ d'évènements,*

$$P(\lim_n \downarrow B_n) = \lim_n \downarrow P(B_n).$$

Remarque 2.8 *Pour (i), on se rappellera que comme la suite $(A_n)_{n \in \mathbb{N}}$ est croissante,* $\lim_n \uparrow A_n$ *n'est autre que $\cup_n A_n$. De plus, par la propriété de croissance des probabilités, la suite $(P(A_n))$ est une suite croissante de $[0,1]$, d'où la notation* $\lim_n \uparrow P(A_n)$*, qui ne sert qu'à rappeler que la suite est croissante et admet donc bien une limite dans* $[0,1]$.

Deux remarques analogues sont valables pour (ii).

[2]il s'agit d'une expression figurée qui signifie 'continue pour les suites croissantes' et est utilisée par analogie avec les fonctions : $\mathbb{R} \to \mathbb{R}$ pour qui ces deux expressions sont synonymes

Dém. Montrons la continuité à gauche. Soit (A_n) une suite croissante d'éléments de $\mathscr{A}$ et soient $B_0 := A_0$, et pour tout entier naturel non nul n, $B_n := A_n \setminus A_{n-1}$. Alors les (B_n) sont des éléments de $\mathscr{A}$ deux à deux disjoints, donc

$$P\left(\bigsqcup_n B_n\right) = \sum_n P(B_n).$$

Mais d'une part, $\sqcup_n B_n = \cup_n A_n = \lim_n \uparrow A_n$ et d'autre part,

$$\sum_n P(B_n) = \lim_n \sum_{k=0}^{n} P(B_k) = \lim_n P(\sqcup_{k=0}^n B_k) = \lim_n P(A_n).$$

Montrons la continuité à droite. Soit (B_n) une suite décroissante d'éléments de $\mathscr{A}$ et pour tout entier n $A_n := {}^c B_n$. Alors

$$P\left(\lim_n \downarrow B_n\right) = P\left(\bigcap_n B_n\right) = 1 - P\left({}^c\bigcap_n B_n\right) = 1 - P\left(\bigcup A_n\right).$$

Or (A_n) est une suite croissante donc d'après ce qui précède,

$$P\left(\bigcup A_n\right) = \lim_n \uparrow P(A_n) = \lim_n \uparrow (1 - P(B_n)) = 1 - \lim_n \downarrow P(B_n).$$

On obtient donc

$$P\left(\lim_n \downarrow B_n\right) = 1 - P\left(\bigcup A_n\right) = 1 - \left(1 - \lim_n \downarrow P(B_n)\right) = \lim_n \downarrow P(B_n),$$

ce qui achève la démonstration. □

Exemple 2.9 *Quelques exemples simples de probabilités :*

- *Soit $a \in \Omega$. La* mesure de Dirac *au point a est définie pour tout $A \in \mathscr{P}(\Omega)$ par*

$$P(A) := \begin{cases} 1 & si\ a \in A \\ 0 & sinon. \end{cases}$$

  Cette mesure est souvent notée δ_a;

- *Si Ω est fini, la* probabilité uniforme *sur Ω, ou* équiprobabilité, *est définie pour tout $A \in \mathscr{P}(\Omega)$ par*

$$P(A) := \frac{Card(A)}{Card(\Omega)}$$

  où $Card(A)$ désigne le cardinal (ou nombre d'éléments) de l'évènement A.

- *Si Ω est dénombrable, alors par σ-additivité, on peut écrire pour tout évènement A*

$$P(A) = \sum_{\omega \in A} P(\{\omega\}).$$

Autrement dit, une probabilité P sur un ensemble dénombrable Ω est caractérisée par la fonction $f : \Omega \longrightarrow [0,1]$ définie par $f(\omega) = P(\{\omega\})$. Réciproquement, toute fonction f vérifiant $\sum_{\omega\in\Omega} f(\omega) = 1$ définit une probabilité par l'égalité $P(A) = \sum_{\omega\in A} P(\{\omega\})$.

Le cas de la mesure de Dirac correspond à la fonction $f = \mathbb{1}_{\{a\}}$ et le cas de la probabilité uniforme correspond à la fonction constante égale partout à $1/Card(\Omega)$.

Remarque 2.10 *Attention, pour une probabilité sur un espace non dénombrable on ne peut pas écrire $P(A) = \sum_{\omega\in A} P(\{\omega\})$.*

Proposition 2.11 *Soit P une probabilité sur Ω. Alors l'ensemble $A := \{\omega \in \Omega : P(\{\omega\}) \neq 0\}$ est dénombrable.*

Dém. Soit $A_n := \{\omega \in \Omega : P(\{\omega\}) \geq \frac{1}{n}\}$. Si A_n était infini, il existerait une suite $(a_k)_{k\in\mathbb{N}}$ d'éléments de A_n deux à deux distincts et on aurait

$$1 \geq P(A_n) \geq P(\sqcup_{k\in\mathbb{N}}\{a_k\}) = \sum_{k\in\mathbb{N}} P(\{a_k\}) \geq \sum_{k\in\mathbb{N}} \frac{1}{n} = \infty.$$

Donc A_n est fini, et comme $A = \cup_{n\geq 1} A_n$, A est dénombrable. □

Remarque 2.12 *L'exemple type de probabilité sur un espace non dénombrable est la* mesure de Lebesgue *(définie ci-dessous) sur l'hypercube $[0,1]^d$. La mesure de Lebesgue est simplement la mesure de volume dans $\mathbb{R}^d$ (de longueur si $d = 1$, de surface si $d = 2$).*

Définition 2.13 *Il existe une unique probabilité P sur $[0,1]^d$ appelée* mesure de Lebesgue[3], *qui vérifie pour tout pavé $A = \prod_{i=1}^d [a_i, b_i]$,*

$$P(A) = \prod_{i=1}^{d} (b_i - a_i).$$

En particulier pour tout $\omega \in [0,1]^d$, $P(\{\omega\}) = 0$.

2.2 Variables aléatoires discrètes

2.2.1 Définition et exemples

Soit $(\Omega, \mathscr{A}, P)$ un espace de probabilité.

Définition 2.14 *On appelle* variable aléatoire réelle *toute fonction $X : \Omega \longrightarrow \mathbb{R}$. Si X prend ses valeurs dans une partie dénombrable de $\mathbb{R}$ on dira que X est une* variable aléatoire discrète. *En particulier, si elle ne prend que des valeurs entières, on dira que X est une* variable aléatoire entière.

[3]La mesure de Lebesgue est définie sur la tribu borélienne (notion hors programme) de $\mathbb{R}^d$, qui contient en particulier les pavés et (donc) les réunions dénombrables de pavés.

On appelle loi de X, et l'on note P_X la probabilité sur $\mathbb{R}$ définie pour toute partie B de $\mathbb{R}$ par

$$P_X(B) := P(X^{-1}(B)) = P(\{X \in B\}).$$

Notation 2.15 *Désormais on notera $P(X \in B)$ plutôt que $P(\{X \in B\})$.*

Remarque 2.16 *En toute rigueur, il faut aussi munir $\mathbb{R}$ d'une tribu $\mathscr{B}$ appelée tribu borélienne, ne considérer que des parties B de $\mathbb{R}$ telles que $B \in \mathscr{B}$ et vérifier que pour tout $B \in \mathscr{B}$, $\{X \in B\}$ est un évènement, c'est-à-dire $\{X \in B\} \in \mathscr{A}$. Toutes ces notions sont hors programme. Quoiqu'il en soit nous ne considérerons pour B que des intervalles, des ensembles dénombrables ou des réunions d'intervalles et d'ensembles dénombrables, qui sont tous bien dans la tribu borélienne.*

Dém. Il nous faut démontrer que P_X est bien une probabilité sur $\mathbb{R}$. Montrons d'abord que $P_X(\varnothing) = 0$ et que $P_X(\mathbb{R}) = 1$. On a $P_X(\varnothing) = P(X^{-1}(\varnothing)) = P(\varnothing) = 0$ et $P_X(\mathbb{R}) = P(X^{-1}(\mathbb{R})) = P(\Omega) = 1$. Montrons à présent la σ-additivité de P_X. Soit une suite (B_n) de parties de $\mathbb{R}$ disjointes deux à deux. On sait d'après les formules de Hausdorff que

$$X^{-1}\left(\bigcup_{n\in\mathbb{N}} B_n\right) = \bigcup_{n\in\mathbb{N}} X^{-1}(B_n),$$

et que pour tous entiers $i \neq j$, $X^{-1}(B_i) \cap X^{-1}(B_j) = X^{-1}(B_i \cap B_j) = X^{-1}(\varnothing) = \varnothing$, si bien que les $(X^{-1}(B_n))$ sont également deux à deux incompatibles. On peut donc écrire

$$\begin{aligned} P_X\left(\bigsqcup_{n\in\mathbb{N}} B_n\right) &= P\left(X^{-1}\left(\bigsqcup_{n\in\mathbb{N}} B_n\right)\right) \\ &= P\left(\bigsqcup_{n\in\mathbb{N}} X^{-1}(B_n)\right) = \sum_{n\in\mathbb{N}} P\left(X^{-1}(B_n)\right) = \sum_{n\in\mathbb{N}} P_X\left(B_n\right), \end{aligned}$$

ce qui prouve que P_X est bien σ-additive. □

Exemple 2.17 *Voici maintenant les exemples usuels de variables aléatoires à valeurs dans $\mathbb{N}$.*

- *Variable aléatoire* de Bernoulli. *Une v.a.* de Bernoulli *X de paramètre $p \in [0,1]$ est une v.a. à valeurs dans $\{0,1\}$ telle que*

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

- *Variable aléatoire* binomiale. *Une v.a.* binomiale *X de paramètres $n \in \mathbb{N}$ et $p \in [0,1]$ est une v.a. à valeurs dans $\{0,1,\ldots,n\}$ telle que*

$$P(X = k) = \binom{n}{k} p^k (1-p)^{n-k} \qquad k = 0, 1, \ldots, n.$$

- *Variable aléatoire* géométrique. *Une v.a.* géométrique X *de paramètre* $a \in]0,1[$ *est une v.a. à valeurs dans* $\mathbb{N}$ *telle que*

$$P(X = k) = (1-a)a^k \qquad k \in \mathbb{N}.$$

Noter qu'on qualifie invariablement de géométrique toute v.a. dont la loi est donnée par une suite géométrique, qu'elle parte de 0 ou de 1, et que l'on appelle paramètre la raison a *de la suite ou* $1-a$. *Il sera toujours précisé dans un énoncé d'exercice ou d'examen ce que l'on entend par variable géométrique de paramètre* a.

- *Variable aléatoire* de Poisson. *Une v.a.* de Poisson X *de paramètre* $\theta \in]0,\infty[$ *est une v.a. à valeurs dans* $\mathbb{N}$ *telle que*

$$P(X = k) = e^{-\theta}\frac{\theta^k}{k!} \qquad k \in \mathbb{N}.$$

- *Variable aléatoire de Cauchy discrète. On appellera ici v.a. de Cauchy discrète* X *une v.a. à valeurs dans* $\mathbb{N}$ *telle que*

$$P(X = k) = \frac{1}{k(k-1)} \qquad k \geq 2.$$

Exercice 2.18 *Pour chaque v.a. définie précédemment, montrer que les probabilités des évènements* $\{X = k\}$ *somment bien à 1.*

2.2.2 Espérance et variance

Définition 2.19 *Pour toute v.a.* X *entière positive, on définit l'*espérance *de* X, *notée* $E(X)$ *par*

$$E(X) := \sum_{n \in \mathbb{N}} P(X = n)\, n.$$

De manière générale, pour toute v.a. discrète positive X *à valeurs dans* $B \subseteq \mathbb{R}_+$ *dénombrable,*

$$E(X) := \sum_{b \in B} P(X = b)\, b.$$

Remarque 2.20 *Noter que l'espérance d'une v.a.* X *entière positive n'est pas forcément entière et peut même être infinie. C'est le cas de la v.a. de Cauchy discrète* X *qui vérifie*

$$E(X) = \sum_{n \geq 2} \frac{1}{n(n-1)}\, n = \sum_{n \geq 2} \frac{1}{n-1} = \infty.$$

Proposition 2.21 *Si* Ω *est dénombrable et que* X *est une variable discrète positive,*

$$E(X) = \sum_{\omega \in \Omega} P(\{\omega\})\, X(\omega).$$

Dém. Pour simplifier, supposons que X est entière, mais la démonstration se fait de la même manière dans le cas général. Comme les évènements $(\{X = n\})_{n\in\mathbb{N}}$ forment une partition de Ω,

$$\sum_{\omega\in\Omega} P(\{\omega\})\, X(\omega) = \sum_{n\in\mathbb{N}} \sum_{\omega\in\{X=n\}} P(\{\omega\})\, X(\omega).$$

Or pour tout $\omega \in \{X = n\}$, $X(\omega) = n$ par définition, donc

$$\sum_{\omega\in\Omega} P(\{\omega\})\, X(\omega) = \sum_{n\in\mathbb{N}} \sum_{\omega\in\{X=n\}} P(\{\omega\})n = \sum_{n\in\mathbb{N}} P(X = n)n,$$

où la dernière égalité est obtenue par σ-additivité. □

Exercice 2.22 *Linéarité de l'espérance pour les v.a. positives. En supposant que Ω est dénombrable, montrer que pour tout $a \geq 0$ et toutes v.a. discrètes positives X et Y, $E(aX + Y) = aE(X) + E(Y)$.*

Définition 2.23 *Pour toute v.a. X on définit les v.a. positives*

$$X^+ := X\mathbb{1}_{X\geq 0} \quad \text{et} \quad X^- := -X\mathbb{1}_{X<0},$$

de sorte que

$$X = X^+ - X^- \quad \text{et} \quad |X| = X^+ + X^-.$$

Par linéarité de l'espérance pour les v.a. positives, on a

$$E(|X|) = E(X^+) + E(X^-).$$

Définition 2.24 *On dit qu'une v.a. discrète X (qu'elle soit positive ou pas) est* intégrable *si $E(|X|) < \infty$.*

D'après la définition précédente, comme $E(|X|) = E(X^+) + E(X^-)$, si X est intégrable les deux espérances $E(X^+)$ et $E(X^-)$ sont finies et l'on peut définir l'espérance de X par

$$E(X) = E(X^+) - E(X^-)$$

qui est de signe quelconque mais forcément finie. En particulier, l'inégalité triangulaire $|E(X^+) - E(X^-)| \leq |E(X^+)| + |E(X^-)|$ s'écrit

$$|E(X)| \leq E(|X|) < \infty.$$

Comme pour une v.a. positive, on peut alors écrire si Ω est dénombrable

$$E(X) = \sum_{\omega\in\Omega} P(\{\omega\})\, X(\omega).$$

Remarque 2.25 *Pour toute v.a. discrète et intégrable X à valeurs dans l'ensemble dénombrable B, par définition*

$$E(X^+) := \sum_{b\in B\cap\mathbb{R}_+} P(X = b)b \quad \text{et} \quad E(X^-) := \sum_{b\in B\cap\mathbb{R}_-^\star} P(X = b)(-b),$$

où B est l'ensemble dénombrable dans lequel X prend ses valeurs. On a donc

$$E(X) = \sum_{b \in B \cap \mathbb{R}_+} P(X = b)b - \sum_{b \in B \cap \mathbb{R}_-^\star} P(X = b)(-b) = E(X) = \sum_{b \in B} P(X = b)b,$$

et en particulier si X est entière,

$$E(X) = \sum_{n \in \mathbb{Z}} P(X = n)n.$$

Proposition 2.26 *L'espérance a les propriétés suivantes :*

(i) Croissance. Si X et Y sont deux v.a. discrètes qui sont toutes deux positives ou toutes deux intégrables, et telles que $X(\omega) \leq Y(\omega)$ pour tout $\omega \in \Omega$, on a l'inégalité

$$E(X) \leq E(Y).$$

(ii) Linéarité. Si X et Y sont deux v.a. discrètes positives alors $E(X + Y) = E(X) + E(Y)$, et pour tout réel a, $E(aX) = aE(X)$.

Si X et Y sont deux v.a. discrètes intégrables, alors $X + Y$ est intégrable, $E(X + Y) = E(X) + E(Y)$ et

$$E(|X + Y|) \leq E(|X|) + E(|Y|).$$

De plus pour tout réel a, $E(aX) = aE(X)$.

(iii) Pour tout évènement A, $E(\mathbb{1}_A) = P(A)$. Pour toute v.a. X constante égale à $a \in \mathbb{R}$, $E(X) = a$.

Dém. Faisons les démonstrations dans le cas où Ω est dénombrable.

(i) Croissance. Dans le cas où X et Y sont toutes deux positives, puisque $X(\omega) \leq Y(\omega)$ pour tout ω, on a $P(\{\omega\})X(\omega) \leq P(\{\omega\})Y(\omega)$ et en sommant ces termes positifs sur ω on obtient l'inégalité $E(X) \leq E(Y)$. Lorsque X et Y sont intégrables, le fait que $X(\omega) \leq Y(\omega)$ pour tout ω implique que $X^+(\omega) \leq Y^+(\omega)$ et $X^-(\omega) \geq Y^-(\omega)$ pour tout ω, de sorte que $E(X^+) \leq E(Y^+)$ et $E(X^-) \geq E(Y^-)$. Comme tous ces termes sont finis, on peut en faire la différence pour obtenir

$$E(X^+) - E(X^-) \leq E(Y^+) - E(Y^-),$$

c'est-à-dire $E(X) \leq E(Y)$.

(ii) Linéarité. Le cas des v.a. positives a été vu en exercice. Dans le cas des v.a. intégrables, l'inégalité triangulaire implique $|X + Y|(\omega) \leq |X|(\omega) + |Y|(\omega)$ pour tout ω, donc par croissance de l'espérance

$$E(|X + Y|) \leq E(|X| + |Y|) = E(|X|) + E(|Y|) < \infty,$$

où l'on a utilisé la linéarité de l'espérance pour les v.a. positives. Ainsi $X+Y$ est intégrable, et comme $(X+Y)^+ - (X+Y)^- = X+Y = X^+ - X^- + Y^+ - Y^-$, on a

$$(X+Y)^+ + X^- + Y^- = (X+Y)^- + X^+ + Y^+,$$

qui implique par la linéarité de l'espérance des v.a. positives

$$E((X+Y)^+) + E(X^-) + E(Y^-) = E((X+Y)^-) + E(X^+) + E(Y^+).$$

Comme X, Y et $X+Y$ sont toutes intégrables, tous les termes de l'égalité précédente sont finis et l'on peut donc retrancher $E((X+Y)^-) + E(X^-) + E(Y^-)$ à chaque membre, ce qui donne

$$E((X+Y)^+) - E((X+Y)^-) = E(X^+) - E(X^-) + E(Y^+) - E(Y^-),$$

qui n'est autre que l'égalité $E(X+Y) = E(X) + E(Y)$.

(iii) Pour tout évènement A, la v.a. $\mathbb{1}_A$ est une v.a. de Bernoulli qui ne prend que les valeurs 0 ou 1, donc en la notant X

$$E(X) = 0 \cdot P(X=0) + 1 \cdot P(X=1) = P(X=1) = P(A),$$

car $\{X=1\} = A$. De la même manière, si X ne prend que la valeur a,

$$E(X) = aP(X=a) = a,$$

car $\{X=a\} = \Omega$. $\square$

Remarque 2.27 *Si X est une v.a. bornée par M, c'est-à-dire telle que $P(|X| \leq M) = 1$, alors X est intégrable et $E(|X|) \leq E(M) = M$.*

Définition 2.28 *Soit X une v.a. discrète telle que X^2 est intégrable. On dit que X* est de carré intégrable. *Alors X est intégrable, et si l'on note $m = E(X)$, on définit la variance de X, notée $Var(X)$ par*

$$Var(X) := E\left((X-m)^2\right) = E\left(X^2\right) - m^2.$$

Dém. Montrons que si X à valeurs dans B dénombrable est de carré intégrable, alors elle est intégrable. Comme $|b| \leq b^2$ dès que $|b| \geq 1$,

$$\begin{aligned}
E(|X|) &= \sum_{b\in B} |b|\, P(X=b) \\
&= \sum_{b\in B\cap[-1,1]} |b|\, P(X=b) + \sum_{b\in B\setminus[-1,1]} |b|\, P(X=b) \\
&\leq \sum_{b\in B\cap[-1,1]} P(X=b) + \sum_{b\in B\setminus[-1,1]} b^2 P(X=b) \\
&\leq 1 + \sum_{b\in B} b^2 P(X=b) \\
&= 1 + E\left(X^2\right) < \infty.
\end{aligned}$$

Plus directement, on peut remarquer que $|X| \leq 1 + X^2$ (c'est-à-dire $|X|(\omega) \leq 1 + X^2(\omega)$ pour tout ω) et utiliser la croissance de l'espérance.

Montrons l'égalité exposée dans la définition. Grâce à la linéarité de l'espérance,

$$
\begin{aligned}
E\left((X-m)^2\right) &= E\left(X^2 - 2mX + m^2\right) \\
&= E\left(X^2\right) - 2mE(X) + E\left(m^2\right) = E\left(X^2\right) - 2m^2 + m^2,
\end{aligned}
$$

ce qui donne l'égalité cherchée $E\left((X-m)^2\right) = E\left(X^2\right) - m^2$. □

Exemple 2.29 *Nous donnons maintenant l'espérance et la variances des variables aléatoires entières usuelles.*

- *Si X est une v.a. de Bernoulli de paramètre $p \in [0,1]$, alors*

$$E(X) = p \quad \text{et} \quad \mathit{Var}(X) = p(1-p).$$

- *Si X est une v.a. binomiale de paramètres $n \in \mathbb{N}$ et $p \in [0,1]$, alors*

$$E(X) = np \quad \text{et} \quad \mathit{Var}(X) = np(1-p).$$

- *Si X est une v.a. géométrique de paramètre $a \in [0,1]$, alors*

$$E(X) = \frac{a}{1-a} \quad \text{et} \quad \mathit{Var}(X) = \frac{a}{(1-a)^2}.$$

- *Si X est une v.a. de Poisson de paramètre $\theta \in]0,\infty[$, alors*

$$E(X) = \theta \quad \text{et} \quad \mathit{Var}(X) = \theta.$$

Exercice 2.30 *Démontrer les identités données dans le paragraphe précédent.*

Chapter 3

Probabilité conditionnelle, indépendance, Borel–Cantelli

Dans tout le chapitre, on se donne un espace de probabilité $(\Omega, \mathscr{A}, P)$.

3.1 Probabilité conditionnelle

Définition 3.1 *Soit B un évènement tel que $P(B) \neq 0$. Alors on définit la probabilité P_B, appelée* probabilité conditionnelle sachant B *par*

$$P_B(A) = \frac{P(A \cap B)}{P(B)} \qquad A \in \mathscr{A}.$$

On note habituellement $P_B(A)$ sous la forme $P(A|B)$.

Dém. Montrons que P_B est une probabilité. D'abord $P(\varnothing \cap B) = 0$ donc $P_B(\varnothing) = 0$. Ensuite $P(\Omega \cap B) = P(B)$ donc $P_B(\Omega) = 1$. Enfin, pour toute suite (A_n) d'évènements deux à deux incompatibles,

$$\begin{aligned} P_B\left(\bigsqcup_n A_n\right) &= \frac{P\left(\left(\bigcup_n A_n\right) \cap B\right)}{P(B)} \\ &= \frac{P_B\left(\bigsqcup_n (A_n \cap B)\right)}{P(B)} = \frac{\sum_n P(A_n \cap B)}{P(B)} = \sum_n \frac{P(A_n \cap B)}{P(B)} = \sum_n P_B(A_n), \end{aligned}$$

ce qui prouve la σ-additivité de P_B. $\square$

Exemple 3.2 *On modélise une famille de deux enfants par l'espace $\Omega = \{F, G\} \times \{F, G\}$. Par exemple, (G, F) représente une famille où l'aîné est un garçon et la cadette une fille. Montrer que la probabilité que les deux enfants sont des filles sachant que l'aîné des enfants est une fille vaut $1/2$, mais que la probabilité que les deux enfants sont des filles sachant qu'au moins un des deux enfants est une fille vaut $1/3$.*

Proposition 3.3 (Formule des probabilités totales) *Soit $(B_n)_{n\in\mathbb{N}}$ une suite d'évènements formant une partition de Ω (c'est-à-dire que $\Omega = \sqcup_n B_n$), telle que $P(B_n) \neq 0$ pour tout n. Alors pour tout évènement A,*

$$P(A) = \sum_{n\in\mathbb{N}} P(A \cap B_n) = \sum_{n\in\mathbb{N}} P(A|B_n)P(B_n).$$

Dém. Les $(A \cap B_n)$ sont deux à deux incompatibles et leur réunion vaut A, donc par σ-additivité,

$$P(A) = P(\sqcup_n(A \cap B_n)) = \sum_n P(A \cap B_n),$$

et l'on applique la formule $P(A \cap B_n) = P(A|B_n)P(B_n)$ pour conclure. □

Proposition 3.4 (Formule de Bayes) *Soit $(B_n)_{n\in\mathbb{N}}$ une suite d'évènements formant une partition de Ω telle que $P(B_n) \neq 0$ pour tout n. Alors pour tout évènement A tel que $P(A) \neq 0$ et pour tout $k \in \mathbb{N}$,*

$$P(B_k|A) = \frac{P(A|B_k)P(B_k)}{\sum_{n\in\mathbb{N}} P(A|B_n)P(B_n)}.$$

Dém. On écrit

$$P(B_k|A) = \frac{P(B_k \cap A)}{P(A)} = \frac{P(A \cap B_k)}{P(A)} = \frac{P(A|B_k)P(B_k)}{P(A)},$$

et l'on remplace $P(A)$ par l'expression donnée par la formule des probabilités totales. □

Exercice 3.5 *Un patient à l'hôpital peut être sain (S) ou malade (M), séropositif ($+$) ou séronégatif ($-$) pour un test de sérologie censé diagnostiquer la maladie. Calculer la probabilité qu'un patient est vraiment malade sachant qu'il est séropositif en supposant que sont connues la prévalence de la maladie $p := P(M)$, la sensibilité du test $\beta := P(+|M)$ et sa spécificité $\gamma = P(-|S)$. Faire l'application numérique avec $p = 10^{-4}$, $\beta = 0,9$ et $\gamma = 0,95$. Réponse*[1] *: $P(M|+) = 1,8.10^{-3}$.*

3.2 Indépendance

Définition 3.6 *Deux évènements A et B sont dits* indépendants *si $P(A\cap B) = P(A)\,P(B)$.*

Remarque 3.7 *Soient A et B deux évènements indépendants. Alors si $P(B) \neq 0$, $P(A|B) = P(A)$ et si $P(A) \neq 0$, $P(B|A) = P(B)$.*

Exemple 3.8 *Tirage uniforme dans un paquet de 52 cartes. Soit $A := \{$ cartes de cœur $\}$ et $B := \{$ valets $\}$. Alors A et B sont indépendants car $P(A) = 13/52$, $P(B) = 4/52$ et*

$$P(A \cap B) = P(\{\text{ valet de cœur }\}) = \frac{1}{52} = P(A)\,P(B).$$

Proposition 3.9 *Si A et B sont indépendants alors ${}^cA$ et B sont indépendants (et donc ${}^cB$ et A le sont, ${}^cB$ et ${}^cA$ le sont).*

[1] absolument surprenante, n'est-ce-pas ?

Dém. Par la formule des probabilités totales, $P(B) = P(B \cap A) + P(B \cap {}^cA)$, donc

$$P({}^cA \cap B) = P(B) - P(A \cap B) = P(B) - P(A)\, P(B) = (1 - P(A))\, P(B) = P({}^cA)\, P(B),$$

ce qui prouve que ${}^cA$ et B sont indépendants. $\square$

Définition 3.10 *On dit que les évènements $A_1, \dots, A_n$ sont indépendants si pour toute partie I de $\{1, \dots, n\}$,*

$$P\left(\bigcap_{i \in I} A_i\right) = \prod_{i \in I} P\left(A_i\right).$$

Remarque 3.11 *Si les $A_1, \dots, A_n$ sont indépendants, alors pour tous indices i et j, on obtient en prenant $I = \{i, j\}$ que $P(A_i \cap A_j) = P(A_i)\, P(A_j)$, c'est-à-dire que les (A_i) sont indépendants deux à deux. Attention, la réciproque n'est pas vraie.*

Exemple 3.12 *On modélise une série de lancers d'une pièce de monnaie par l'ensemble $\Omega = \{pile, face\}^{\mathbb{N}}$ des suites à valeurs dans $\{pile, face\}$, et l'on définit la projection X_k qui à $\omega \in \Omega$ associe sa k-ième coordonnée $\omega_k =: X_k(\omega)$ qui est donc le résultat du k-ième lancer. Le résultat du k-ième lancer est pile ssi $\omega \in A_k$, où*

$$A_k := \{\omega \in \Omega : \omega_k = pile\} = \{X_k = pile\}.$$

Dans ce cadre, on peut supposer que les évènements (A_k) sont indépendants. Si l'on définit la v.a. T comme le rang du premier 'pile'

$$T(\omega) := \min\{k \geq 1 : X_k(\omega) = pile\},$$

alors

$$\{T = n\} = {}^cA_1 \cap \cdots \cap {}^cA_{n-1} \cap A_n,$$

et par indépendance de tous ces évènements,

$$P(T = n) = P(A_n) \prod_{i=1}^{n-1} P({}^cA_i).$$

Supposons que l'on utilise la même pièce à chaque lancer, c'est-à-dire que la probabilité p de tomber sur face est la même pour tous les lancers, autrement dit $P(A_k) = 1 - p$ pour tout k. Alors l'égalité précédente implique

$$P(T = n) = p^{n-1}(1 - p) \qquad n \geq 1.$$

Cela montre que T suit la loi géométrique (décalée de 1, en fait, car par définition T ne prend pas la valeur 0) et l'on peut calculer

$$E(T) = \frac{1}{1 - p}.$$

3.3 Lemme de Borel–Cantelli

Dans cette section, on s'intéresse à une suite $(A_n)_{n\in\mathbb{N}}$ d'évènements et l'on cherche à connaître la probabilité qu'une infinité d'entre eux se produit. Plus précisément, on définit la variable aléatoire

$$Z(\omega) := \sum_{n\in\mathbb{N}} \mathbb{1}_{A_n}(\omega)$$

à valeurs dans $\bar{\mathbb{N}} := \mathbb{N} \cup \{+\infty\}$ et l'évènement $\bar{A} := \{Z = +\infty\}$, et l'on s'intéresse à $P(\bar{A})$.

Définition 3.13 *La suite $(\cup_{k\geq n} A_k)$ étant décroissante, on peut considérer sa limite, que l'on appelle* limite supérieure *de la suite (A_n) et que l'on note de deux manières*

$$\limsup_n A_n = \overline{\lim_n} A_n := \lim_n \downarrow \bigcup_{k\geq n} A_k.$$

Proposition 3.14 *On a l'égalité entre évènements $\bar{A} = \limsup_n A_n$.*

Dém. Pour tout $\omega \in \Omega$, on a

$$\begin{aligned}
\omega \in \limsup_{n\to\infty} A_n &\iff \omega \in \bigcap_{n\geq 0} \bigcup_{k\geq n} A_k \\
&\iff \forall n \ \exists k \geq n, \ \omega \in A_k \\
&\iff \{n : \omega \in A_n\} \text{ est infini} \\
&\iff \sum_{n\in\mathbb{N}} \mathbb{1}_{A_n}(\omega) = \infty \\
&\iff \omega \in \bar{A},
\end{aligned}$$

ce qui prouve l'égalité cherchée. □

Théorème 3.15 (Lemme de Borel–Cantelli) *Si $\sum_n P(A_n) < \infty$, alors $P(\bar{A}) = 0$.*
Si $\sum_n P(A_n) = \infty$ et que les (A_n) sont indépendants, alors $P(\bar{A}) = 1$.

Exemple 3.16 *Reprenons l'exemple des lancers de pièces vu en Exemple 3.12, mais supposons ici qu'on lance une pièce différente à chaque lancer, et que la n-ième pièce tombe sur pile avec probabilité p_n. Si l'on définit X_n le résultat du n-ième lancer et que l'on définit $A_n = \{X_n = pile\}$, alors les (A_n) sont indépendants et $P(A_n) = p_n$. Le lemme de Borel–Cantelli appliqué à la suite (A_n) assure alors que dans cette suite infinie de lancers, le nombre de lancers qui tombent sur pile est fini avec probabilité 1 ou infini avec probabilité 1 selon que la série de terme général converge ou diverge.*

Dém. Montrons d'abord la première assertion. Soit $\varepsilon > 0$. Comme la série de terme général $P(A_n)$ converge, son reste est inférieur à ε à partir d'un certain rang, autrement dit il existe $N \in \mathbb{N}$ tel que pour tout $n \geq N$, $\sum_{k \geq n} P(A_k) < \varepsilon$. Par la sous-$\sigma$-additivité de P,

$$P\left(\bigcup_{k \geq n} A_k\right) \leq \sum_{k \geq n} P(A_k) < \varepsilon.$$

Donc la suite $(P(\cup_{k \geq n} A_k))_n$ est majorée par ε et grâce à la continuité à droite de P, elle converge vers $P(\bar{A})$, donc $P(\bar{A}) < \varepsilon$. Comme ε est arbitraire, ceci montre que $P(\bar{A}) = 0$.

Montrons à présent la seconde assertion. Rappelons-nous que l'indépendance des (A_n) implique l'indépendance des $({}^c A_n)$, ainsi pour tous entiers $n \leq N$,

$$\begin{aligned} P\left(\bigcup_{k=n}^{N} A_k\right) &= 1 - P\left(\bigcap_{k=n}^{N} {}^c A_k\right) \\ &= 1 - \prod_{k=n}^{N} P({}^c A_k) = 1 - \exp\left(\sum_{k=n}^{N} \ln(P({}^c A_k))\right) = 1 - \exp\left(\sum_{k=n}^{N} \ln(1 - P(A_k))\right). \end{aligned}$$

Or pour tout $x \geq 0$, $\ln(1-x) \leq -x$, ainsi par croissance de l'exponentielle,

$$P\left(\bigcup_{k=n}^{N} A_k\right) \geq 1 - \exp\left(-\sum_{k=n}^{N} P(A_k)\right).$$

Chaque membre de la dernière inégalité est croissant en N. Par continuité à gauche de P, le membre de gauche converge vers $P\left(\cup_{k \geq n} A_k\right)$ lorsque $N \to \infty$. Parce que la série de terme général $P(A_n)$ diverge, le membre de gauche converge vers 1 lorsque $N \to \infty$. Ainsi, l'inégalité devient à la limite

$$P\left(\bigcup_{k \geq n} A_k\right) \geq 1,$$

et comme une probabilité est toujours inférieure à 1, on a $P\left(\cup_{k \geq n} A_k\right) = 1$. Par continuité à droite de P, le membre de gauche converge vers $P(\bar{A})$ lorsque $n \to \infty$, ce qui donne le résultat. $\square$

Remarque 3.17 *La définition d'espérance peut s'étendre au cas des variables aléatoires à valeurs dans $\bar{\mathbb{N}}$ par $E(Z) = +\infty$ si $P(Z = +\infty) \neq 0$. Alors en admettant que l'on peut intervertir sommation et espérance, on a*

$$E(Z) = E\left(\sum_n \mathbb{1}_{A_n}\right) = \sum_n E(\mathbb{1}_{A_n}) = \sum_n P(A_n).$$

Ainsi, si $\sum_n P(A_n) < \infty$, alors $E(Z) < \infty$, ce qui implique que $P(Z = +\infty) = 0$.

Chapter 4

Loi d'une variable aléatoire réelle

4.1 Préliminaires

On rappelle qu'une probabilité P sur $\mathbb{R}$ n'est pas définie en général sur $\mathscr{P}(\mathbb{R})$ mais sur un sous-ensemble de $\mathscr{P}(\mathbb{R})$ noté $\mathscr{B}$ ou $\mathscr{B}(\mathbb{R})$ et appelé tribu borélienne, dont les éléments sont appelés boréliens, ou parties boréliennes de $\mathbb{R}$, mais que cette notion est hors-programme. En pratique, nous ne considérerons de toutes façons que des parties B de $\mathbb{R}$ qui sont des réunions finies ou dénombrables d'intervalles (qui sont bien 'boréliennes').

Définition 4.1 *Pour toute probabilité P sur $\mathbb{R}$, on dit que x est un* atome *de P si $P(\{x\}) \neq 0$. On se rappelle que l'ensemble des atomes de P est toujours dénombrable.*

Théorème 4.2 (Théorème de classe monotone, admis) *On dit que deux probabilités P et Q sur $\mathbb{R}$ sont égales, et l'on note $P = Q$, si pour tout $B \in \mathscr{B}(\mathbb{R})$, on a l'égalité $P(B) = Q(B)$. Si $P(I) = Q(I)$ pour tout intervalle ouvert I de $\mathbb{R}$, alors $P = Q$.*

Remarque 4.3 *Le théorème précédent peut être vu comme une propriété de continuité (et de densité), de même que si deux fonctions continues f et g définies sur $\mathbb{R}$ coïncident sur les rationnels alors $f = g$.*

Remarque 4.4 *Noter que le théorème reste vrai si l'on remplace 'intervalle ouvert' par 'intervalle fermé'. En effet, si P et Q coïncident sur les intervalles fermés, elles coïncident sur les singletons. Ainsi pour tous $a < b$, $P(\{a\}) = Q(\{a\})$, $P(\{b\}) = Q(\{b\})$ et par hypothèse $P([a, b]) = Q([a, b])$. Mais comme*

$$P(]a, b[) = P([a, b]) - P(\{a\}) - P(\{b\}),$$

et de même pour Q, on a l'égalité $P(]a, b[) = Q(]a, b[)$ pour tous $a < b$, ce qui implique bien $P = Q$.

On se donne à présent un espace de probabilité $(\Omega, \mathscr{A}, P)$ et l'on rappelle qu'une variable aléatoire réelle, notée souvent v.a.r., sur Ω est une fonction $X : \Omega \to \mathbb{R}$. On rappelle également que la loi de X est une probabilité sur $\mathbb{R}$ définie par $P_X(B) = P(X \in B) = P(\{\omega \in \Omega : X(\omega) \in B\})$, où $B \in \mathscr{B}(\mathbb{R})$. Dans la suite on ne fera plus référence au Théorème 4.2, mais l'on se servira de la définition suivante.

Définition 4.5 *On dit que deux variables aléatoires réelles X et Y ont même loi, et l'on note $X \stackrel{\mathscr{L}}{=} Y$, si $P_X = P_Y$, c'est-à-dire si pour tout intervalle ouvert I de $\mathbb{R}$, $P_X(I) = P_Y(I)$.*

On rappelle l'existence d'une probabilité P sur $[0,1]$, appelée mesure de Lebesgue ou mesure de longueur, telle que pour tous $0 \leq a \leq b \leq 1$, $P([a,b]) = P(]a,b[) = b - a$.

Exemple 4.6 *On dit qu'une variable aléatoire réelle U suit la loi* uniforme *sur $[0,1]$ si P_U est la mesure de Lebesgue sur $[0,1]$. On dit aussi que U est une variable aléatoire uniforme.*

Plus généralement, une v.a.r. (variable aléatoire réelle) X est dite uniforme sur l'intervalle $[c,d]$ si pour tous éléments $a < b$ de $[c,d]$,

$$P(X \in [a,b]) = P_X([a,b]) = \frac{b-a}{d-c}$$

On voit que pour une v.a.r. uniforme X, l'ensemble des atomes de X est vide. On ne peut donc PAS écrire que $P_X([a,b])$ est égal à $\sum_{x \in [a,b]} P_X(\{x\})$ (car un intervalle non réduit à un singleton n'est pas dénombrable), qui vaut 0. Cela tranche avec le cas des variables discrètes, pour lesquelles on peut écrire une telle égalité parce que la somme est dénombrable.

4.2 Fonction de répartition

Définition 4.7 *Pour toute v.a.r. X, on définit sa* fonction de répartition *et l'on note F_X, la fonction $F_X : \mathbb{R} \to [0,1]$ définie par*

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

Exemple 4.8 *La fonction de répartition d'une v.a.r. U uniforme sur $[0,1]$ est la fonction F_U définie par*

$$F_U(x) = \begin{cases} 0 & si \quad x < 0 \\ x & si \quad x \in [0,1] \\ 1 & si \quad x > 1. \end{cases}$$

Proposition 4.9 *Pour toute v.a.r. X, sa fonction de répartition F_X vérifie les propriétés suivantes :*

(i) F_X *est croissante;*

(ii) F_X *est continue à droite;*

(iii) $\lim_{x \to -\infty} F_X(x) = 0$ *et* $\lim_{x \to +\infty} F_X(x) = 1$.

Dém. La propriété (i) est due à la croissance de P_X, en effet pour tous $x \leq y$, $]-\infty, x] \subset]-\infty, y]$ donc

$$P_X(]-\infty, x]) \leq P_X(]-\infty, y]),$$

ce qui montre que $F_X(x) \leq F_X(y)$. Pour (ii) on se donne $x \in \mathbb{R}$ et une suite décroissante (x_n) telle que $\lim_n \downarrow x_n = x$. Alors il est facile de voir que la suite $(]-\infty, x_n])$ est décroissante et que sa limite est $]-\infty, x]$, donc par continuité à droite de P_X,

$$\lim_n F_X(x_n) = \lim_n \downarrow P_X(]-\infty, x_n]) = P_X(\lim_n \downarrow]-\infty, x_n]) = P_X(]-\infty, x]) = F_X(x),$$

ce qui prouve la continuité à droite de F_X. Démontrons enfin (iii). Soit d'abord une suite décroissante (x_n) telle que $\lim_n \downarrow x_n = -\infty$. De la même manière que précédemment, la suite $(]-\infty, x_n])$ est décroissante de limite $\varnothing$, donc par continuité à droite de P_X,

$$\lim_n F_X(x_n) = \lim_n \downarrow P_X(]-\infty, x_n]) = P_X(\varnothing) = 0.$$

Soit maintenant une suite croissante (x_n) telle que $\lim_n \uparrow x_n = +\infty$. Cette fois, la suite $(]-\infty, x_n])$ est croissante de limite $\mathbb{R}$, donc par continuité à gauche de P_X,

$$\lim_n F_X(x_n) = \lim_n \uparrow P_X(]-\infty, x_n]) = P_X(\mathbb{R}) = 1,$$

ce qui achève la démonstration. □

On rappelle que toute fonction croissante F admet une limite à gauche en tout point x de $\mathbb{R}$ notée $F(x-)$ telle que

$$F(x-) = \lim_{n\to\infty} F(x_n),$$

pour toute suite croissante (x_n) de limite égale à x et telle que $x_n < x$ pour tout n.

Proposition 4.10 *Pour toute v.a.r. X,*

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

Ainsi $P_X(\{x\}) = F_X(x) - F_X(x-)$ et x est donc un atome de P_X si et seulement si x est un point de discontinuité de F_X.

Dém. Si (x_n) est une suite croissante de limite égale à x et telle que $x_n < x$ pour tout n, alors la suite $(]-\infty, x_n])$ est une suite croissante de limite $]-\infty, x[$. La continuité à gauche de P_X permet de conclure que

$$F_X(x-) = \lim_n F_X(x_n) = \lim_n \uparrow P_X(]-\infty, x_n]) = P_X(]-\infty, x[).$$

La dernière égalité de la proposition consiste simplement à écrire que $P_X(\{x\}) = P_X(]-\infty, x]) - P_X(]-\infty, x[)$. □

Remarque 4.11 *On sait que l'ensemble des points de discontinuité d'une fonction croissante est dénombrable. On retrouve ainsi le fait que l'ensemble des atomes de P_X est dénombrable.*

Corollaire 4.12 *On déduit de la proposition précédente les quatre égalités suivantes, valides pour tous* $a < b$

- $P_X(]a,b]) = F_X(b) - F_X(a)$
- $P_X(]a,b[) = F_X(b-) - F_X(a)$
- $P_X([a,b[) = F_X(b-) - F_X(a-)$
- $P_X([a,b]) = F_X(b) - F_X(a-)$.

Dém. Il suffit d'écrire $P_X(]a,b]) = P_X(]-\infty,b]) - P_X(]-\infty,a]) = F_X(b) - F_X(a)$ et d'utiliser le fait que $P_X(\{x\}) = F_X(x) - F_X(x-)$, pour $x = a$ ou b. □

Corollaire 4.13 *La fonction de répartition d'une v.a.r. caractérise sa loi. En d'autres termes, si* X *et* Y *sont deux v.a.r. telles que* $F_X = F_Y$ *alors* $P_X = P_Y$.

Dém. Si $F_X = F_Y$ alors pour tout $x \in \mathbb{R}$, $F_X(x) = F_Y(x)$ et aussi $F_X(x-) = F_Y(x-)$. Par conséquent pour tous $a < b$,

$$P_X(]a,b[) = F_X(b-) - F_X(a) = F_Y(b-) - F_Y(a) = P_Y(]a,b[),$$

ce qui montre que P_X et P_Y coïncident sur les intervalles ouverts, donc sont égales. □

Théorème 4.14 *Soit* F *une fonction qui vérifie les propriétés de la Proposition 4.9. Alors* F *est la fonction de répartition d'une certaine variable aléatoire. Plus précisément, si l'on définit* F^{-1} *l'inverse à droite de* F *par*

$$F^{-1}(u) := \inf\{y \in [0,1] : F(y) > u\} \qquad u \in [0,1[,$$

et si U *est une v.a.r. uniforme sur* $[0,1]$*, alors la fonction de répartition de la v.a.r.* $F^{-1}(U)$ *est* F.

Dém. Il n'est pas difficile de démontrer que pour tout $x \in \mathbb{R}$ et $u \in [0,1[$,

$$F(x) > u \ \Rightarrow\ x \geq F^{-1}(u) \ \Rightarrow\ F(x) \geq u.$$

Ainsi,

$$\{\omega : F(x) > U(\omega)\} \subset \{\omega : x \geq F^{-1}(U(\omega))\} \subset \{\omega : F(x) \geq U(\omega)\}$$

et par croissance de P,

$$P(U < F(x)) \leq P(F^{-1}(U) \leq x) \leq P(U \leq F(x)).$$

Or on se rappelle que pour tout $a \in [0,1]$, $P(U < a) = P(U \leq a) = a$, ce qui implique $P(F^{-1}(U) \leq x) = F(x)$. □

Remarque 4.15 *En vertu des deux résultats précédents, pour toute v.a.r.* X *de fonction de répartition* F_X*, si* F_X^{-1} *désigne l'inverse à droite de* X *et* U *une v.a.r. uniforme sur* $[0,1]$*, alors* $F_X^{-1}(U)$ *a même loi que* X.

4.3 Variables à densité

Définition 4.16 *Une fonction* $f : \mathbb{R} \to [0, \infty[$ *est appelée* densité de probabilité *ou tout simplement* densité, *si* $\int_{-\infty}^{+\infty} f(y)\, dy = 1$.

Si f est une densité, on dit qu'une v.a.r. X a pour densité f, ou suit la loi de densité f, si $P(X \in [a, b]) = \int_a^b f(y)\, dy$ pour tous $a < b$, ou de manière équivalente si

$$F_X(x) = \int_{-\infty}^{x} f(y)\, dy \qquad x \in \mathbb{R}.$$

Remarque 4.17 *Si X a pour densité f, (alors F_X est continue et donc) P_X n'a pas d'atome.*

Proposition 4.18 *Si X a pour densité f, et que f est continue en x alors F_X est dérivable en x et $F'_X(x) = f(x)$.*

Si X est une v.a.r. dont la fonction de répartition est dérivable et si $f := F'_X$ est intégrable, alors X a pour densité f.

Dém. Ces deux propriétés sont des théorèmes fondamentaux d'analyse. Pour la première assertion, on écrit pour tout x et $\varepsilon > 0$

$$\inf_{[x,x+\varepsilon]} f \leq f(y) \leq \sup_{[x,x+\varepsilon]} f \qquad y \in [x, x+\varepsilon],$$

puis on intègre en y pour obtenir

$$\varepsilon \inf_{[x,x+\varepsilon]} f \leq \int_x^{x+\varepsilon} f(y)\, dy \leq \varepsilon \sup_{[x,x+\varepsilon]} f.$$

Or $\int_x^{x+\varepsilon} f(y)\, dy = F_X(x+\varepsilon) - F_X(x)$, donc en divisant par ε,

$$\inf_{[x,x+\varepsilon]} f \leq \frac{1}{\varepsilon}(F_X(x+\varepsilon) - F_X(x)) \leq \sup_{[x,x+\varepsilon]} f,$$

et en faisant tendre ε vers 0, on obtient que F_X est dérivable à droite en x de dérivée égale à $f(x)$. On démontrerait de même que F_X est dérivable à gauche de même dérivée.

Pour la deuxième assertion, si F_X est dérivable de dérivée f intégrable, alors $\int_a^b f(y)\, dy = F_X(b) - F_X(a) = P_X([a, b])$, ce qui montre que X a pour densité f. □

Exemple 4.19 *Voici maintenant les exemples usuels de v.a.r. à densité.*

- *On dit qu'une v.a.r. X suit la* loi exponentielle *de paramètre $\lambda > 0$ si elle a pour densité f donnée par*

$$f(x) = \begin{cases} 0 & si \quad x < 0 \\ \lambda\, e^{-\lambda x} & si \quad x \geq 0. \end{cases}$$

On dit aussi que X est une v.a. exponentielle (de paramètre λ), ce que l'on note parfois
$$X \sim \mathscr{E}xp(\lambda).$$
Sa fonction de répartition F_X est donnée par
$$F_X(x) = \begin{cases} 0 & si \quad x < 0 \\ 1 - e^{-\lambda x} & si \quad x \geq 0. \end{cases}$$

- *On dit qu'une v.a.r. X suit la* loi de Cauchy *si elle a pour densité f donnée par*
$$f(x) = \frac{1}{\pi(1+x^2)} \qquad x \in \mathbb{R}.$$
On dit aussi que X est une v.a. de Cauchy.

- *On dit qu'une v.a.r. X suit la* loi normale *ou* gaussienne *de paramètres $m \in \mathbb{R}$ et $\sigma > 0$ si elle a pour densité f donnée par*
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}} \qquad x \in \mathbb{R}.$$
On dit aussi que X est une v.a. normale, ou gaussienne, d'espérance m et de variance σ^2, ce que l'on note parfois
$$X \sim \mathscr{N}(m, \sigma^2).$$

4.4 Changement de variable

Soit X une v.a.r. à valeurs dans un intervalle I de $\mathbb{R}$ et $g : I \to J$ une bijection, où J est un intervalle. On s'intéresse à la loi de la nouvelle v.a.r. $Y = g(X)$.

On rappelle que g est appelée homéomorphisme si g et g^{-1} sont continues, et qu'alors g et g^{-1} sont soit toutes deux croissantes soit toutes deux décroissantes. Si de plus g et g^{-1} sont de classe C^1, on dit que g est un C^1- difféomorphisme. Dans ce cas g' et $(g^{-1})'$ sont soit toutes deux strictement positives, soit toutes deux strictement négatives.

Proposition 4.20 *Soit F_Y la fonction de répartition de $Y = g(X)$. Si g est un homéomorphisme, alors si g est croissante $F_Y = F_X \circ g^{-1}$ tandis que si g est décroissante F_Y est donnée par*
$$F_Y(y) = 1 - F_X(g^{-1}(y)-) \qquad y \in J.$$

Dém. Si g est croissante, alors
$$F_Y(y) = P(g(X) \leq y) = P(X \leq g^{-1}(y)) = F_X(g^{-1}(y)),$$
alors que si g est décroissante
$$F_Y(y) = P(g(X) \leq y) = P(X \geq g^{-1}(y)) = 1 - P(X < g^{-1}(y)) = 1 - F_X(g^{-1}(y)-),$$
ce qui achève la démonstration. □

Théorème 4.21 *On suppose que X a une densité f continue et que g est un C^1-difféomorphisme. Alors Y a pour densité $\tilde{f}$, où $\tilde{f}$ est donnée par*

$$\tilde{f}(y) = \left|(g^{-1})'(y)\right| \, f \circ g^{-1}(y) = \frac{f \circ g^{-1}(y)}{|g' \circ g^{-1}(y)|} \qquad y \in J.$$

Dém. D'après la proposition précédente adaptée au cas particulier où X a une densité (et a donc une fonction de répartition continue), $F_Y = F_X \circ (g^{-1})$ si $g' > 0$ et $F_Y = 1 - F_X \circ (g^{-1})$ si $g' < 0$. D'après la section précédente, comme f est continue, F_X est dérivable et $F_X' = f$. Donc F_Y est dérivable et $F_Y' = (F_X \circ (g^{-1}))'$ si $g' > 0$, $F_Y' = -(F_X \circ (g^{-1}))'$ si $g' < 0$. La dérivation donne le résultat. □

Chapter 5

Moments d'une variable aléatoire et inégalités classiques

5.1 Espérance d'une v.a. réelle

Théorème 5.1 *Il existe une unique application, notée E et appelée* espérance *qui à toute v.a.r. positive X associe un élément de $\bar{\mathbb{R}}_+ := [0, +\infty]$ noté $E(X)$, telle que :*

(i) E est linéaire : pour tout $a \geq 0$, pour toutes v.a.r. positives X et Y, $E(aX+Y) = aE(X) + E(Y)$;

(ii) E est croissante : pour toutes v.a.r. positives X et Y telles que $X(\omega) \geq Y(\omega)$ pour tout $\omega \in \Omega$, $E(X) \geq E(Y)$;

(iii) Si $X = \mathbb{1}_A$ alors $E(X) = P(A)$.

De plus, si X est une v.a.r. positive qui a pour densité f sur $\mathbb{R}_+$ alors

$$E(X) = \int_0^\infty x\, f(x)\, dx.$$

Remarque 5.2 *Dans l'égalité (i), les opérations usuelles sont prolongées à $\bar{\mathbb{R}}_+$ par $c + (+\infty) = +\infty$ pour tout réel c et $c \times (+\infty) = +\infty$ pour tout réel strictement positif c. En revanche, il faut adopter la convention $0 \times (+\infty) = 0$ pour que l'égalité soit valide même lorsque $a = 0$ et $E(X) = +\infty$, ce que nous ferons dans toute la suite.*

Remarque 5.3 *D'après la théorie de la mesure, E est en fait ce qu'on appelle l'intégrale par rapport à la mesure de probabilité P, que l'on note dans ce cadre $E(X) = \int_\Omega X(\omega)\, dP(\omega)$ ou plus simplement $E(X) = \int X\, dP$. À nouveau, cette notion est hors programme.*

Théorème 5.4 (de Beppo Levi, de convergence monotone; admis) *Soit (X_n) une suite croissante de v.a.r. positives, c'est-à-dire que pour tout n et tout ω, $X_{n+1}(\omega) \geq X_n(\omega)$ (on dit parfois que (X_n) est croissante ω par ω). On peut alors définir pour tout ω*

$$X(\omega) := \lim_n \uparrow X_n(\omega) \in \bar{\mathbb{R}}_+,$$

et l'on peut intervertir limite et espérance :

$$E(X) = \lim_n \uparrow E(X_n),$$

où l'égalité a lieu dans $\bar{\mathbb{R}}_+$.

Le théorème de convergence monotone nous permet de connecter la notion d'espérance que nous venons de définir à celle que nous avons vue dans le cas des v.a. discrètes.

Corollaire 5.5 *Soit X une v.a.r. positive à valeurs dans un espace dénombrable. Alors*

$$E(X) = \sum_{x \in X(\Omega)} xP(X = x),$$

où l'on a noté $X(\Omega) = \{y \in \mathbb{R} : \exists \omega \in \Omega, y = X(\omega)\}$ *l'image de* Ω *par* X.

Dém. Comme $X(\Omega)$ est dénombrable, il existe une suite injective (x_n) de $\mathbb{R}_+$ telle que $X(\Omega) = \{x_n; n \in \mathbb{N}\}$. Si l'on note A_n l'évènement $\{X = x_n\} = X^{-1}(\{x_n\})$, alors on peut écrire X sous la forme

$$X = \sum_{n \in \mathbb{N}} x_n \mathbb{1}_{A_n}.$$

Soit alors X_n la v.a.r. définie par $X_n := \sum_{k=0}^n x_k \mathbb{1}_{A_k}$. Par linéarité (i) de l'espérance,

$$E(X_n) = \sum_{k=0}^n x_k E(\mathbb{1}_{A_k}) = \sum_{k=0}^n x_k P(A_k),$$

où l'on a utilisé (iii). Comme les x_n sont positifs, pour tout ω la suite $(X_n(\omega))$ est croissante : elle ne prend que deux valeurs, 0 puis $X(\omega)$ à partir du rang n tel que $X(\omega) = x_n$. Donc sa limite est $X(\omega)$ et d'après le théorème de convergence monotone,

$$E(X) = \lim_n \uparrow E(X_n) = \lim_n \uparrow \sum_{k=0}^n x_k P(A_k) = \sum_{k=0}^\infty x_k P(A_k).$$

Or $A_k = \{X = x_k\}$, donc $E(X) = \sum_{k=0}^\infty x_k P(X = x_k) = \sum_{x \in X(\Omega)} xP(X = x)$. □

Pour étendre l'espérance aux v.a.r. de signe quelconque, on utilise les mêmes notations que dans le cas des v.a. discrètes :

$$X^+ := X \mathbb{1}_{X \geq 0} \quad \text{et} \quad X^- := -X \mathbb{1}_{X<0},$$

de sorte que

$$X = X^+ - X^- \quad \text{et} \quad |X| = X^+ + X^-.$$

Par linéarité de l'espérance pour les v.a.r. positives, on a

$$E(|X|) = E(X^+) + E(X^-).$$

Définition 5.6 *On dit qu'une v.a.r. X de signe constant ou de signe quelconque est* intégrable *si $E(|X|) < \infty$.*

Exactement comme dans la démonstration de la Proposition 2.26(ii), on peut montrer que si X et Y sont deux v.a.r. intégrables alors $X + Y$ l'est aussi, ce qui justifie la proposition (et définition) suivante.

Proposition 5.7 *L'ensemble des v.a.r. intégrables sur Ω est un espace vectoriel noté $\mathscr{L}^1(\Omega)$. On peut prolonger l'espérance à $\mathscr{L}^1(\Omega)$ par la définition*

$$E(X) := E(X^+) - E(X^-),$$

qui est de signe quelconque mais forcément finie. L'espérance sur $\mathscr{L}^1(\Omega)$ vérifie toutes les propriétés énoncées dans la Proposition 2.26 comme la linéarité et la croissance.

Théorème 5.8 (de Lebesgue, ou de convergence dominée; admis) *Soit X une v.a.r. et (X_n) une suite de v.a.r. telle que pour tout $\omega \in \Omega$,*

$$\lim_n X_n(\omega) = X(\omega).$$

S'il existe une v.a.r. intégrable Z telle que pour tout n et tout ω, $|X_n(\omega)| \leq Z(\omega)$, alors X est intégrable, on a la convergence $\lim_n E(|X_n - X|) = 0$, et l'on peut intervertir limite et espérance :

$$E(X) = \lim_n E(X_n).$$

Proposition 5.9 *Soit X une v.a.r. de signe quelconque ayant pour densité f sur $\mathbb{R}$. Alors*

$$E(|X|) = \int_{-\infty}^{\infty} |x| \ f(x) \, dx,$$

donc $X \in \mathscr{L}^1(\Omega) \Leftrightarrow \int_{-\infty}^{\infty} |x| \ f(x) \, dx < \infty$, et dans ce cas

$$E(X) = \int_{-\infty}^{\infty} x \, f(x) \, dx$$

qui est de signe quelconque mais forcément finie.

Dém. Il suffit d'écrire $E(X^+) = \int_0^{\infty} x \, f(x) \, dx \in [0, \infty[$ et $E(X^-) = -\int_{-\infty}^0 x \, f(x) \, dx \in [0, \infty[$, d'en faire la somme pour montrer la première égalité et d'en faire la différence pour montrer la seconde. $\square$

Proposition 5.10 *Soit $h : \mathbb{R} \to \mathbb{R}$. Pour toute v.a.r. X, la v.a.r. $h(X)$ est*

(i) positive dès que h est positive, et alors $E(h(X))$ est bien définie dans $\bar{\mathbb{R}}_+$;

(ii) bornée dès que h est bornée, et alors $E(h(X))$ est bien définie dans $\mathbb{R}$ et l'on a

$$|E(h(X))| \leq E(|h(X)|) \leq \sup_{\mathbb{R}} |h| \, .$$

Dém. Il faut seulement démontrer (ii). Soit $A := \sup_{\mathbb{R}} |h|$. Alors A est une v.a. (constante) telle que $|h(X(\omega))| \leq A$ pour tout ω. Par croissance de l'espérance, on a $E(|h(X)|) \leq E(A) = A$. □

Proposition 5.11 (Formule de transfert) *Soit $h : \mathbb{R} \to \mathbb{R}$ positive ou bornée et X une v.a.r.*

(i) Si X est discrète alors

$$E(h(X)) = \sum_{x \in X(\Omega)} h(x)\, P(X = x).$$

(ii) Si X a pour densité f alors

$$E(h(X)) = \int_{-\infty}^{\infty} h(x)\, f(x)\, dx.$$

5.2 Moments d'une v.a.r.

Définition 5.12 *Pour tout $p > 0$, on appelle $E(|X|^p) \in \bar{\mathbb{R}}_+$ le* moment d'ordre p *de la v.a. X. Si $E(|X|^p) < \infty$, on dit que X admet un moment d'ordre p.*

Théorème 5.13 (Inégalité de Minkowski) *Soient X et Y deux v.a.r. admettant un moment d'ordre p. Alors c'est le cas de $X + Y$, ce qui fait de l'ensemble $\mathscr{L}^p(\Omega)$ des v.a.r. admettant un moment d'ordre p un espace vectoriel. De plus,*

$$\left(E(|X+Y|^p)\right)^{1/p} \leq \left(E(|X|^p)\right)^{1/p} + \left(E(|Y|^p)\right)^{1/p}.$$

Proposition 5.14 *Pour tous $q \leq p$, $\mathscr{L}^p(\Omega) \subset \mathscr{L}^q(\Omega)$.*

Dém. Soit $X \in \mathscr{L}^p(\Omega)$. Alors

$$\begin{aligned} |X|^q &= |X|^q\, \mathbb{1}_{\{|X| \leq 1\}} + |X|^q\, \mathbb{1}_{\{|X|>1\}} \\ &\leq 1 + |X|^q\, \mathbb{1}_{\{|X|>1\}} \\ &\leq 1 + |X|^p\, \mathbb{1}_{\{|X|>1\}} \\ &\leq 1 + |X|^p, \end{aligned}$$

donc $E(|X|^q) \leq E(1 + |X|^p) = 1 + E(|X|^p) < \infty$. □

Définition 5.15 *Si $X \in \mathscr{L}^2(\Omega)$, alors $X \in \mathscr{L}^1(\Omega)$ et l'on peut définir la variance de X comme dans le cas discret par*

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

On définit de même l'écart-type de X par $\sqrt{Var(X)}$ et l'on vérifie également que $Var(aX+b) = a^2\, Var(X)$.

Exemple 5.16 *Nous donnons l'espérance et la variances des v.a.r. usuelles.*

- *Si X suit la loi uniforme sur $[0,1]$, alors*

$$E(X) = \frac{1}{2} \quad \text{et} \quad Var(X) = \frac{1}{12}.$$

- *Si X suit la loi exponentielle $\mathscr{E}xp(\lambda)$, alors*

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

- *Si X suit la loi de Cauchy, X n'est pas intégrable donc n'admet pas de moment d'ordre 2 non plus.*
- *Si X suit la loi normale $\mathscr{N}(m,\sigma^2)$, alors*

$$E(X) = m \quad \text{et} \quad Var(X) = \sigma^2.$$

5.3 Inégalités classiques

Théorème 5.17 (Inégalité de Markov) *Soit $X \in \mathscr{L}^1(\Omega)$. Pour tout $a > 0$,*

$$P(|X| \geq a) \leq \frac{E(|X|)}{a}$$

Dém. On a

$$|X| \geq |X| \, \mathbb{1}_{\{|X|\geq a\}} \geq a \, \mathbb{1}_{\{|X|\geq a\}}.$$

Par la croissance de l'espérance, on obtient $E(|X|) \geq E(a\mathbb{1}_{\{|X|\geq a\}}) = a\,P(|X| \geq a)$. □

Théorème 5.18 (Inégalité de Bienaymé-Chebyshev) *Soit $X \in \mathscr{L}^2(\Omega)$. Pour tout $a > 0$,*

$$P(|X - E(X)| \geq a) \leq \frac{Var(X)}{a^2}$$

Dém. Par l'inégalité de Markov on a

$$P(|X - E(X)| \geq a) = P((X - E(X))^2 \geq a^2) \leq \frac{E((X - E(X))^2)}{a^2} = \frac{\mathrm{Var}(X)}{a^2},$$

ce qui achève la démonstration. □

Proposition 5.19 *Soit X une v.a.r. positive. Si $E(X) = 0$ alors $P(X = 0) = 1$.*

Dém. Pour tout entier non nul n, d'après l'inégalité de Markov,

$$P(X \geq 1/n) \leq \frac{E(X)}{1/n} = 0,$$

donc $P(A_n) = 0$ pour tout $n \geq 1$, où $A_n = \{X \geq 1/n\}$. Mais $\{X > 0\} = \lim_n \uparrow A_n$, donc par continuité à gauche de P, $P(X > 0) = \lim_n P(A_n) = 0$.

□

Proposition 5.20 *Soit* $X \in \mathscr{L}^2(\Omega)$ *et* $m := E(X)$. *Si* $Var(X) = 0$, *alors* $P(X = m) = 1$.

Dém. D'après l'inégalité de Bienaymé-Chebyshev,

$$P(|X - m| \geq 1/n) \leq \frac{\mathrm{Var}(X)}{(1/n)^2} = 0,$$

donc par le même raisonnement que précédemment, $P(|X - m| > 0) = \lim_n P(|X - m| \geq 1/n) = 0$. □

Théorème 5.21 (Inégalité de Jensen) *Soient* I *et* J *deux intervalles de* $\mathbb{R}$ *et* $\varphi : I \to J$ *une fonction convexe. Soit* X *une v.a.r. à valeurs dans* I *telle que* X *et* $\varphi(X)$ *sont intégrables. Alors* $E(X) \in I$ *et*

$$\varphi(E(X)) \leq E(\varphi(X)).$$

Dém. On note $a := \inf I$, $b := \sup I$ et $m := E(X)$, où $a \in [-\infty, +\infty[$, $b \in]-\infty, +\infty]$ et m est fini. Ayant pour tout ω l'inégalité $a \leq X(\omega) \leq b$, par croissance de l'intégrale, $a \leq E(X) \leq b$. Montrons que si I est ouvert en a, $m > a$ et que si I est ouvert en b, $m < b$, de sorte que $m \in I$. Si a ou b est infini, il n'y a rien à montrer car on sait que m est fini. Si a est fini et que I est ouvert en a, on peut montrer que $m > a$. En effet, si $E(X) = a$, alors la v.a. $Y = X - a$ est une variable positive d'espérance nulle, donc nulle avec probabilité 1, c'est-à-dire que $P(X = a) = 1$, mais cela est exclu car I est ouvert en a. La démonstration est analogue pour b. On a donc démontré que $m \in I$ et l'on peut donc lui appliquer φ.

Comme φ est convexe, il existe une droite passant par $(m, \varphi(m))$ qui est située entièrement sous le graphe de φ. Autrement dit si s désigne la pente d'une telle droite,

$$\varphi(x) \geq s(x - m) + \varphi(m) \qquad x \in I.$$

Donc pour tout ω, $\varphi(X(\omega)) \geq s(X(\omega) - m) + \varphi(m)$, et par croissance de l'intégrale,

$$E(\varphi(X)) \geq E(s(X - m) + \varphi(m)) = sE(X) - sm + \varphi(m) = \varphi(m),$$

ce qui est l'inégalité cherchée. □

Chapter 6

Caractérisation de la loi d'une variable aléatoire

6.1 Cas des variables aléatoires entières

On rappelle que se donner la loi d'une variable aléatoire discrète X revient à se donner la collection $(P(X = x))_{x \in X(\Omega)}$, car l'ensemble $X(\Omega)$ des valeurs prises par X est dénombrable. Dans le cas particulier où X est entière positive, c'est-à-dire lorsque $X(\Omega) = \mathbb{N}$, il suffit de se donner la collection $(P(X = n))_{n \in \mathbb{N}}$ ou de manière équivalente, comme nous allons le voir maintenant, sa fonction génératrice.

Noter que dans le cas des variables réelles, la collection $(P(X = x))_{x \in X(\Omega)}$ n'est pas dénombrable et le plus souvent, toutes ces probabilités sont nulles, si bien qu'au contraire cette collection ne contient pas d'information sur la loi de X.

Définition 6.1 *On appelle* fonction génératrice *d'une v.a. entière positive X, et l'on note f_X, la fonction $f_X : [0,1] \to [0,1]$ définie par*

$$f_X(s) = \sum_{n \geq 0} P(X = n)\, s^n = E\left(s^X\right),$$

où l'on a utilisé la formule de transfert dans la deuxième égalité.

Remarque 6.2 *Noter que $f_X(s)$ est une série entière en s qui converge absolument pour tout s dans $[-1,1]$, puisque que ses coefficients sont positifs et que $f_X(1) = \sum_{n \geq 0} P(X = n) = 1$. Ainsi son rayon de convergence est supérieur ou égal à 1. En particulier, f_X est de classe C^∞ sur $]-1,1[$ et la série est dérivable terme à terme, ce qui se traduit par la proposition suivante.*

Proposition 6.3 *Pour tout $k \geq 1$, pour tout $s \in [0,1[$,*

$$f_X^{(k)}(s) = \sum_{n \geq k} P(X = n)\, n(n-1)\cdots(n-k+1)\, s^{n-k} = E\left(X(X-1)\cdots(X-k+1)\, s^{X-k}\right), \tag{6.1}$$

où l'on a à nouveau utilisé la formule de transfert. En particulier, $f_X' \geq 0$ donc f_X est croissante, et $f_X'' \geq 0$, donc f_X est convexe.

Exemple 6.4 *Calculons la fonction génératrice des variables aléatoires entières usuelles.*

- *Si X est une v.a. de Bernoulli de paramètre $p \in [0,1]$, alors*

$$f_X(s) = 1 - p + ps.$$

- *Si X est une v.a. binomiale de paramètres $n \in \mathbb{N}$ et $p \in [0,1]$, alors*

$$f_X(s) = \sum_{k=0}^{n} \binom{n}{k} p^k (1-p)^{n-k} s^k = (1-p+ps)^n.$$

- *Si X est une v.a. géométrique de paramètre $a \in [0,1]$, alors*

$$f_X(s) = \sum_{n \geq 0} (1-a) a^n s^n = \frac{1-a}{1-as}.$$

- *Si X est une v.a. de Poisson de paramètre $\theta \in]0, \infty[$, alors*

$$f_X(s) = \sum_{n \geq 0} e^{-\theta} \frac{\theta^n}{n!} s^n = e^{-\theta(1-s)}.$$

Proposition 6.5 *Pour tout $k \geq 0$*

$$P(X = k) = \frac{f_X^{(k)}(0)}{k!}$$

En conséquence, deux v.a. entières positives qui ont la même fonction génératrice ont la même loi. On dit que la fonction génératrice d'une v.a. caractérise sa loi.

Dém. La première égalité vient de l'application de (6.1) à $s = 0$. En effet, le terme de degré 0 de la série $f_X^{(k)}$ est obtenu en prenant n égal à k dans la somme. Si X et Y sont deux v.a. entières positives telles que $f_X = f_Y$, alors pour tout entier k, $f_X^{(k)} = f_Y^{(k)}$ et en particulier $f_X^{(k)}(0) = f_Y^{(k)}(0)$. On a donc pour tout k,

$$P(X = k) = \frac{f_X^{(k)}(0)}{k!} = \frac{f_Y^{(k)}(0)}{k!} = P(Y = k),$$

donc X et Y ont la même loi. □

Proposition 6.6 *Pour tout $k \geq 0$*

$$f_X^{(k)}(1-) := \lim_{s \to 1, s < 1} f_X^{(k)}(s) = E\left(X(X-1)\cdots(X-k+1)\right),$$

où l'égalité a lieu dans $\bar{\mathbb{R}}_+ = [0, +\infty]$.

Dém. Soit (s_n) une suite croissante de $[0,1[$ telle que $\lim_n s_n = 1$. Soit

$$Z_n(\omega) := X(\omega)(X(\omega)-1)\cdots(X(\omega)-k+1)\, s_n^{X(\omega)-k},$$

de sorte que $f_X^{(k)}(s_n) = E(Z_n)$. On remarque que la suite (Z_n) est une suite croissante de variables aléatoires, au sens où $Z_{n+1}(\omega) \geq Z_n(\omega)$ pour tous n et ω. De plus par continuité en 1 des applications $s \mapsto s^k$ pour tout k, on a la convergence pour tout ω de $Z_n(\omega)$ vers

$$Z(\omega) = X(\omega)(X(\omega)-1)\cdots(X(\omega)-k+1).$$

Donc par le théorème de Beppo Levi,

$$\lim_n f_X^{(k)}(s_n) = \lim_n E(Z_n) = E(\lim_n Z_n) = E(Z),$$

autrement dit

$$\lim_{s\to 1, s<1} f_X^{(k)}(s) = E\left(Z\right),$$

ce qui est le résultat cherché. □

Corollaire 6.7 *Soit X une v.a. entière positive. Alors*

$$E(X) = f'_X(1-) \quad \textit{et} \quad E(X(X-1)) = f''_X(1-).$$

En particulier, $X \in \mathscr{L}^1$ est intégrable ssi $f'_X(1-) < \infty$ et $X \in \mathscr{L}^2$ ssi $f''_X(1-) < \infty$. Si $X \in \mathscr{L}^2$, alors

$$Var(X) = f''_X(1-) + f'_X(1-) - (f'_X(1-))^2.$$

Dém. Les deux premières égalités s'obtiennent simplement en prenant $k=1$ et $k=2$ dans la proposition précédente. La première équivalence est triviale puisque $E(X) = f'_X(1-)$. Si $X \in \mathscr{L}^2$, $f''_X(1-) = E(X(X-1)) < E(X^2) < \infty$. Si $f''_X(1-) < \infty$, alors f''_X est bornée donc f'_X, qui est croissante, est forcément bornée et donc $f'_X(1-) < \infty$. Ainsi $X(X-1)$ et X sont intégrables, donc $X^2 = X(X-1) + X$ est intégrable, c'est-à-dire $X \in \mathscr{L}^2$. Pour le calcul de la variance,

$$\mathrm{Var}(X) = E(X^2) - E(X)^2 = E(X(X-1)+X) - E(X)^2 = E(X(X-1)) + E(X) - E(X)^2,$$

et il suffit de remplacer $E(X(X-1))$ par $f''_X(1-)$ et $E(X)$ par $f'_X(1-)$. □

Exercice 6.8 *Retrouver l'espérance et la variance des variables aléatoires entières usuelles en utilisant le corollaire précédent.*

6.2 Cas des variables aléatoires réelles

6.2.1 Caractérisation par fonction test

Théorème 6.9 (admis) *Soient X et Y deux v.a. réelles. Si pour toute fonction h continue et bornée,*

$$E(h(X)) = E(h(Y)),$$

alors X et Y ont même loi. En particulier, s'il existe une densité f telle que pour toute fonction h continue et bornée,

$$E(h(X)) = \int_{-\infty}^{\infty} h(x)\, f(x)\, dx,$$

alors X admet pour densité f.

Remarque 6.10 *On dit que les fonctions continues bornées forment une classe suffisante de fonctions tests : il suffit que les lois de X et de Y soient indistinguables lorsqu'on les 'teste' contre toutes les fonctions de cette classe pour que les lois de X et Y soient égales.*

Le corollaire suivant renforce les résultats de la Section 4.4, dans le sens où l'on ne suppose plus que la densité f soit continue.

Corollaire 6.11 (Formule du changement de variable) *Soit X une v.a.r. de densité f sur un intervalle ouvert I de $\mathbb{R}$ et $g : I \to J$ un C^1-difféomorphisme, où J est un intervalle ouvert de $\mathbb{R}$, c'est-à-dire que g est bijective, et que g et g^{-1} sont dérivables. Alors $g(X)$ a pour densité $\tilde{f}$ sur J, où $\tilde{f}$ est donnée par*

$$\tilde{f}(y) = \left|(g^{-1})'(y)\right|\, f \circ g^{-1}(y) = \frac{f \circ g^{-1}(y)}{|g' \circ g^{-1}(y)|} \qquad y \in J.$$

Dém. On note $Y = g(X)$. Soit h continue et bornée. Alors en appliquant la formule de transfert à $h \circ g$,

$$E(h(Y)) = E(h \circ g(X)) = \int_I h \circ g(x)\, f(x)\, dx,$$

puis en faisant le changement de variable $y = g(x)$,

$$E(h(Y)) = \int_J h(y)\, f\left(g^{-1}(y)\right)\, \left|(g^{-1})'(y)\right|\, dy = \int_J h(y)\, \tilde{f}(y)\, dy,$$

et l'on conclut grâce au théorème précédent. □

6.2.2 Fonction caractéristique

La fonction caractéristique d'une variable aléatoire réelle est l'analogue de la fonction génératrice que nous avons introduite dans le cas des variables aléatoires entières positives. Avant de donner la définition de la fonction caractéristique, il nous faut préciser ce qu'est l'espérance d'une variable aléatoire complexe.

Définition 6.12 *Soit X une variable aléatoire complexe. On dit que X est intégrable si $E(|X|) < \infty$, où $|X|$ désigne le module de X. Alors $Re(X)$ et $Im(X)$ sont toutes deux intégrables et l'on peut définir*

$$E(X) := E(Re(X)) + iE(Im(X)).$$

L'espérance des v.a. complexes intégrables a toutes les propriétés de l'espérance des v.a. réelles intégrables, en particulier la linéarité, le théorème de convergence dominée et l'inégalité $|E(X)| \leq E(|X|)$.

Définition 6.13 *On appelle* fonction caractéristique *d'une v.a.r. X, et l'on note φ_X la fonction $\varphi_X : \mathbb{R} \to \mathbb{C}$ définie par*

$$\varphi_X(t) := E\left(e^{itX}\right).$$

En particulier si X a pour densité f, alors

$$\varphi_X(t) := \int_{-\infty}^{\infty} e^{itx} f(x)\, dx.$$

Remarque 6.14 *Noter que pour tout réel t, e^{itX} est de module 1, donc intégrable.*

Proposition 6.15 *Pour toute v.a.r. X, φ_X est une fonction continue et bornée en module par 1. De plus, $\varphi_X(0) = 1$ et pour tous réels λ et a,*

$$\varphi_{\lambda X + a}(t) = e^{ita}\, \varphi_X(\lambda t).$$

Dém. Tout d'abord $\left|e^{itX}\right| = 1$ donc

$$|\varphi_X(t)| = \left|E\left(e^{itX}\right)\right| \leq E(\left|e^{itX}\right|) = E(1) = 1.$$

Soit maintenant $t \in \mathbb{R}$. Montrons que φ_X est continue en t. Soit (t_n) une suite de réels convergeant vers t et soit Z_n la v.a. définie par

$$Z_n(\omega) := e^{it_n X(\omega)}.$$

Par continuité des applications $t \mapsto e^{itx}$ pour tous $x \in \mathbb{R}$, la suite $(Z_n(\omega))$ converge pour tout ω vers $e^{itX(\omega)}$. De plus, pour tous n et ω, $|Z_n(\omega)| = 1$, donc la suite (Z_n) est dominée en module uniformément en n par une v.a. intégrable (la v.a. constante à 1). Le théorème de convergence dominée appliqué aux suites de v.a. complexes permet donc de conclure

$$\lim_n \varphi_X(t_n) = \lim_n E(Z_n) = E(\lim_n Z_n) = E(e^{itX}) = \varphi_X(t),$$

ce qui prouve que φ_X est continue en t. Ensuite $\varphi_X(0) = E(e^0) = E(1) = 1$ et

$$\varphi_{\lambda X+a}(t) = E\left(e^{it(\lambda X+a)}\right) = E\left(e^{it\lambda X}\, e^{ita}\right) = e^{ita}\, E\left(e^{it\lambda X}\right) = e^{ita}\, \varphi_X(\lambda t),$$

ce qui achève la démonstration. □

Exemple 6.16 *Nous donnons sans démonstration la fonction caractéristique des variables aléatoires réelles usuelles.*

- *Si X est une v.a. uniforme sur $[a, b]$, alors*
$$\varphi_X(t) = \frac{e^{itb} - e^{ita}}{it(b-a)}.$$
- *Si X est une v.a. exponentielle de paramère λ, alors*
$$\varphi_X(t) = \frac{\lambda}{\lambda - it}.$$
- *Si X est une v.a. de Cauchy symétrique, alors*
$$\varphi_X(t) = e^{-|t|}.$$
- *Si X est une v.a. normale d'espérance m et de variance σ^2, alors*
$$\varphi_X(t) = \exp\left\{ itm - \frac{\sigma^2 t^2}{2} \right\}.$$

Proposition 6.17 *Si X admet un moment d'ordre k, alors φ_X est de classe C^k et*
$$\varphi_X^{(k)}(t) = E\left((iX)^k\, e^{itX}\right).$$

Dém. La formule suggère que l'on peut intervertir dérivation et espérance. C'est ce que nous allons montrer, mais seulement pour $k = 1$, en utilisant le théorème de convergence dominée et l'identité $|e^{ix} - 1| \leq |x|$ valable pour tout réel x. Supposons que X est intégrable. Soit $t \in \mathbb{R}$. Nous allons montrer que φ_X est dérivable en t. Soit (h_n) une suite de réels convergeant vers 0 et soit Z_n la v.a. définie par
$$Z_n(\omega) := \frac{e^{i(t+h_n)X(\omega)} - e^{itX(\omega)}}{h_n} = e^{itX(\omega)}\, \frac{e^{ih_nX(\omega)} - 1}{h_n}.$$
D'après l'identité rappelée plus haut, $|Z_n(\omega)| \leq |X(\omega)|$, de sorte que la suite (Z_n) est bornée en module, uniformément en n, par la variable intégrable X. De plus, pour tout ω, la suite $(Z_n(\omega))$ converge lorsque $n \to \infty$ vers $iX(\omega)\, e^{itX(\omega)}$. Donc par le théorème de convergence dominée,
$$\lim_n \frac{\varphi_X(t + h_n) - \varphi_X(t)}{h_n} = \lim_n E(Z_n) = E(\lim_n Z_n) = E\left(iX\, e^{itX}\right),$$
ce qui montre que φ_X est dérivable en t et que $\varphi'_X(t) = E\left(iX\, e^{itX}\right)$. La continuité de φ'_X se montre par le théorème de convergence dominée comme dans la démonstration de la proposition précédente. □

Théorème 6.18 (admis) *Si deux v.a.r. X et Y sont telles que $\varphi_X = \varphi_Y$ alors X et Y ont même loi. De plus, si*
$$\int_{-\infty}^{\infty} |\varphi_X(t)|\, dt < \infty,$$
alors X a une densité continue et bornée f donnée par
$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \varphi_X(t)\, e^{-itx}\, dt.$$

Remarque 6.19 *La première partie du théorème assure que, comme son nom l'indique, la fonction caractéristique d'une v.a. caractérise sa loi. On en déduit que la classe des fonctions s'écrivant sous la forme $x \mapsto e^{itx}$ est une classe suffisante de fonctions-tests. On remarquera que cette classe est beaucoup plus petite que la classe des fonctions continues bornées, puisqu'elle est indicée par un seul paramètre réel (le paramètre t de e^{itx}).*

La deuxième partie du théorème est encore plus incroyable. Elle assure que sous l'hypothèse d'intégrabilité de la fonction φ_X, on peut 'inverser' la fonction caractéristique, c'est-à-dire exprimer explicitement la densité de X à partir de sa fonction caractéristique.

Chapter 7

Variables indépendantes, produit de convolution

7.1 Définitions et premières propriétés

Définition 7.1 *Soient X et Y deux variables aléatoires réelles. On dit que X et Y sont indépendantes et l'on note $X \perp\!\!\!\perp Y$ si pour toutes parties*[1] *A et B de $\mathbb{R}$,*

$$P(X \in A, Y \in B) := P(\{X \in A\} \cap \{Y \in B\}) = P(X \in A)\, P(Y \in B).$$

Proposition 7.2 *Soient X et Y deux variables aléatoires réelles. Si $X \perp\!\!\!\perp Y$, alors pour toutes fonctions $f, g : \mathbb{R} \to \mathbb{R}$, $f(X)$ et $g(Y)$ sont indépendantes.*

Dém. Soient A et B deux parties de $\mathbb{R}$. Alors

$$\begin{aligned} P(f(X) \in A, g(Y) \in B) &= P\left(X \in f^{-1}(A), Y \in g^{-1}(B)\right) \\ &= P\left(X \in f^{-1}(A)\right)\, P\left(Y \in g^{-1}(B)\right) \\ &= P(f(X) \in A)\, P(g(Y) \in B). \end{aligned}$$

où la deuxième égalité est due à l'indépendance de X et Y. □

Le théorème suivant est admis.

Théorème 7.3 *Soient X et Y deux variables aléatoires réelles. Chacune des quatre conditions suivantes est équivalente à l'indépendance de X et Y :*

(i) $P(X \in A, Y \in B) = P(X \in A)\, P(Y \in B)$ *pour tous intervalles A et B;*

(ii) $P(X \in A, Y \in B) = P(X \in A)\, P(Y \in B)$ *pour toutes demie-droites A et B;*

(iii) $E\left(f(X)\, g(Y)\right) = E\left(f(X)\right)\, E\left(g(Y)\right)$ *pour toutes fonctions $f, g : \mathbb{R} \to \mathbb{R}$ toutes deux positives ou toutes deux bornées;*

[1] boréliennes

(iv) $E\left(e^{isX}e^{itY}\right) = E\left(e^{isX}\right)\, E\left(e^{itY}\right)$ *pour tous* $s, t \in \mathbb{R}$.

Définition 7.4 *Soient* $X_1, \ldots, X_n$ *n variables aléatoires réelles. On dit que les v.a.* (X_i) *sont indépendantes si pour toutes parties*[2] *de* $\mathbb{R}$ $A_1, \ldots, A_n$,

$$P(X_i \in A_i, i = 1, \ldots, n) := P\left(\cap_{i=1}^n \{X_i \in A_i\}\right) = \prod_{i=1}^n P(X_i \in A_i).$$

Le théorème précédent se généralise sous la forme suivante.

Théorème 7.5 *Soient* $X_1, \ldots, X_n$ *n variables aléatoires réelles. Chacune des quatre conditions suivantes est équivalente à l'indépendance des variables* (X_i) *:*

(i) $P(X_i \in A_i, i = 1, \ldots, n) = \prod_{i=1}^n P(X_i \in A_i)$ *pour tous intervalles* $A_1, \ldots, A_n$*;*

(ii) $P(X_i \in A_i, i = 1, \ldots, n) = \prod_{i=1}^n P(X_i \in A_i)$ *pour toutes demie-droites* $A_1, \ldots, A_n$*;*

(iii) $E\left(\prod_{i=1}^n f_i(X_i)\right) = \prod_{i=1}^n E\left(f_i(X_i)\right)$ *pour toutes fonctions* $f_i : \mathbb{R} \to \mathbb{R}$ *toutes positives ou toutes bornées;*

(iv) $E\left(\prod_{j=1}^n e^{it_jX_j}\right) = \prod_{j=1}^n E\left(e^{it_jX_j}\right)$ *pour tous réels* $t_1, \ldots, t_n$.

Exercice 7.6 *Montrer que les variables* (X_i) *sont indépendantes ssi les évènements* $(\{X_i \in A_i\})$ *sont indépendants pour tout n-uplet d'intervalles* $A_1, \ldots, A_n$.

Remarque 7.7 *La relation d'indépendance est symétrique mais elle n'est pas transitive. En effet si* $X \perp\!\!\!\perp Y$ *et* $Y \perp\!\!\!\perp Z$*, on n'a pas forcément* $X \perp\!\!\!\perp Z$*. Penser par exemple au cas où* $Z = f(X)$ *(ou tout simplement* $Z = X$*).*

Proposition 7.8 *Si* $X, Y \in \mathscr{L}^2$*, alors* $XY \in \mathscr{L}^1$*. Comme* $X + Y \in \mathscr{L}^2$*, sa variance est bien définie, donnée par*

$$Var\,(X + Y) = Var\,(X) + Var\,(Y) + 2Cov\,(X, Y),$$

où l'on a défini

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

Si de plus $X \perp\!\!\!\perp Y$*, alors* $Cov\,(X, Y) = 0$ *et donc* $Var\,(X + Y) = Var\,(X) + Var\,(Y)$.

Dém. On a l'inégalité $|XY| \leq (X^2 + Y^2)/2$. Comme X^2 et Y^2 sont intégrables $(X^2 + Y^2)/2$ l'est aussi donc XY également. Calculons la variance de $X + Y$. On a d'abord

$$E\left((X + Y)^2\right) = E\left(X^2 + Y^2 + 2XY\right) = E\left(X^2\right) + E\left(Y^2\right) + 2E\left(XY\right)$$

et parrallèlement

$$\left[E\left(X + Y\right)\right]^2 = \left[E(X) + E(Y)\right]^2 = \left[E(X)\right]^2 + \left[E(Y)\right]^2 + E(X)\,E(Y).$$

[2]boréliennes

L'expression cherchée s'obtient en faisant la différence des deux égalités précédentes.

Montrons enfin que si X et Y sont indépendantes leur covariance est nulle. On définit la fonction f_n par $f_n(x) = x\mathbb{1}_{|x|\leq n}$. Comme f est bornée et que $X \perp\!\!\!\perp Y$, on a

$$E(f_n(X)\, f_n(Y)) = E(f_n(X))\, E(f_n(Y)).$$

Montrons que les suites $(E(f_n(X)\, f_n(Y))$, $(E(f_n(X))$ et $(E(f_n(Y))$ convergent vers $E(XY)$, $E(X)$ et $E(Y)$ respectivement, ce qui donnera le résultat. D'abord $f_n(x) \to x$ lorsque $n \to \infty$, donc les suite de v.a. $(f_n(X)\, f_n(Y))$, $(f_n(X))$ et $(f_n(Y))$ convergent (ω par ω) vers XY, X et Y respectivement. De plus, ces suites sont dominées par $|XY|$, $|X|$ et $|Y|$ respectivement, qui sont toutes trois intégrables, donc par le théorème de convergence dominée, les suites $(E(f_n(X)\, f_n(Y))$, $(E(f_n(X))$ et $(E(f_n(Y))$ convergent vers $E(XY)$, $E(X)$ et $E(Y)$ respectivement. □

Remarque 7.9 *Nous avons vu que si $X \perp\!\!\!\perp Y$, alors Cov $(X, Y) = 0$. Attention, la réciproque n'est pas vraie en général, il existe des v.a. de carré intégrable dont la covariance est nulle mais qui ne sont pas indépendantes.*

Exercice 7.10 *Soient $X_1, \ldots, X_n$ n variables aléatoires réelles. Montrer que si $X_i \perp\!\!\!\perp X_j$ pour tous $i \neq j$, alors*

$$Var\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} Var(X_i)\,.$$

Noter que c'est le cas dès que les variables (X_i) sont indépendantes (ce qui est plus fort que l'hypothèse précédente d'indépendance paire par paire).

7.2 Cas des variables discrètes

Proposition 7.11 *Soient X et Y deux v.a. discrètes. Alors on a l'équivalence*

$$X \perp\!\!\!\perp Y \Leftrightarrow \text{ Pour tous } x \in X(\Omega) \text{ et } y \in Y(\Omega),\ P(X = x, Y = y) = P(X = x)\, P(Y = y).$$

Dém. Si $X \perp\!\!\!\perp Y$, il suffit de remplacer A par le singleton $\{x\}$ et B par le singleton $\{y\}$ dans la définition d'indépendance pour obtenir la première implication. Montrons à présent l'implication réciproque. Soient A et B deux intervalles de $\mathbb{R}$. Alors comme X et Y sont discrètes, par la formule des probabilités totales,

$$\begin{aligned} P(X \in A, Y \in B) &= \sum_{x\in A}\sum_{y\in B} P(X = x, Y = y) = \sum_{x\in A}\sum_{y\in B} P(X = x)\, P(Y = y) \\ &= \sum_{x\in A} P(X = x) \sum_{y\in B} P(Y = y) = P(X \in A)\, P(Y \in B), \end{aligned}$$

ce qui prouve l'indépendance de X et Y. □

Remarque 7.12 *Bien sûr, la proposition précédente se généralise facilement au cas de n variables aléatoires discrètes.*

Proposition 7.13 *Soient $X_1, \ldots, X_n$ n v.a. entières positives. Alors les v.a. (X_i) sont indépendantes ssi pour tous $s_1, \ldots, s_n \in [0,1]$,*

$$E\left(\prod_{i=1}^{n} s_i^{X_i}\right) = \prod_{i=1}^{n} E\left(s_i^{X_i}\right).$$

Dém. Si les v.a. (X_i), sont indépendantes, il suffit de remplacer les fonctions f_i du Théorème 7.5(iii) par la fonction qui à $x \in \mathbb{N}$ associe s_i^x pour obtenir la première implication. Montrons à présent l'implication réciproque dans le cas $n = 2$ pour simplifier. Soit

$$F(s,t) := \sum_{k \geq 0} \sum_{n \geq 0} P(X_1 = k, X_2 = n)\, s^k t^n.$$

Par la formule de transfert, on a $F(s,t) = E\left(s^{X_1} t^{X_2}\right)$. D'autre part, il est classique d'obtenir que

$$\frac{\partial^{k+n} F}{\partial s^k \partial t^n}(0,0) = k!\, n!\, P(X_1 = k, X_2 = n).$$

Or par hypothèse $F(s,t) = E\left(s^{X_1} t^{X_2}\right) = E\left(s^{X_1}\right) E\left(t^{X_2}\right) = f(s) g(t)$, où f est la fonction génératrice de X_1 et g celle de X_2, donc

$$\frac{\partial^{k+n} F}{\partial s^k \partial t^n}(0,0) = f^{(k)}(0)\, g^{(n)}(0) = k!\, P(X_1 = k)\, n!\, P(X_2 = n).$$

On a donc $P(X_1 = k, X_2 = n) = P(X_1 = k)\, P(X_2 = n)$ pour tous k, n, ce qui prouve l'indépendance de X_1 et X_2. □

Corollaire 7.14 *Si $X_1, \ldots, X_n$ sont des v.a. entières positives indépendantes, alors en définissant $Z := \sum_{i=1}^n X_i$, on a*

$$f_Z = \prod_{i=1}^{n} f_{X_i}$$

Dém. Il suffit de prendre tous les s_i égaux à s dans la proposition précédente, ce qui donne

$$f_Z(s) = E\left(s^{\sum_{i=1}^n X_i}\right) = \prod_{i=1}^{n} E\left(s^{X_i}\right) = \prod_{i=1}^{n} f_{X_i}(s),$$

ce qui prouve le résultat cherché. □

Définition 7.15 *Soient $X_1, \ldots, X_n$ n variables aléatoires. Si elles sont indépendantes et que de plus elles ont toutes la même loi, on dit qu'elle sont i.i.d. pour "indépendantes et identiquement distribuées".*

Exemple 7.16 *Soient $X_1, \ldots, X_n$ n variables aléatoires réelles i.i.d. suivant toutes la loi de Bernoulli de paramètre p. Alors $Z := \sum_{i=1}^n X_i$ est une variable binomiale de paramètres p et n. En effet, pour tout $s \in [0,1]$,*

$$f_Z(s) = \prod_{i=1}^{n} E\left(s^{X_i}\right) = \prod_{i=1}^{n} (1 - p + ps) = (1 - p + ps)^n,$$

qui n'est autre que la fonction génératrice cherchée. On retrouve alors très facilement l'espérance et la variance de Z, car

$$E(Z) = \sum_{i=1}^{n} E(X_i) = np \quad \text{et} \quad Var\,(Z) = \sum_{i=1}^{n} Var\,(X_i) = np(1-p).$$

Noter que l'on a utilisé l'indépendance des variables (X_i) seulement dans le calcul de la variance.

Définition 7.17 *Soient μ et ν deux probabilités sur $\mathbb{N}$. Le* produit de convolution *de μ avec ν est la probabilité sur $\mathbb{N}$ notée $\mu \star \nu$ définie par*

$$\mu \star \nu(\{n\}) = \sum_{k=0}^{n} \mu(\{k\})\,\nu(\{n-k\}).$$

Exercice 7.18 *Montrer que $\mu\star\nu$ est bien une probabilité et que le produit de convolution est commutatif et associatif. On rappelle que l'on peut toujours échanger l'ordre de sommation dans les séries doubles à termes positifs.*

Proposition 7.19 *Si X et Y sont deux v.a. entières positives indépendantes, alors*

$$P_{X+Y} = P_X \star P_Y.$$

Dém. Pour tout entier positif n, par la formule des probabilités totales,

$$\begin{aligned} P_{X+Y}(\{n\}) &= P(X+Y=n) = \sum_{k=0}^{n} P(X=k, Y=n-k) \\ &= \sum_{k=0}^{n} P(X=k)\,P(Y=n-k) = \sum_{k=0}^{n} P_X(\{k\})\,P_Y(\{n-k\}), \end{aligned}$$

ce qui prouve l'égalité attendue. □

7.3 Cas des variables à densité

Nous allons commencer cette section par quelques préliminaires concernant les intégrales multiples.

Théorème 7.20 (Fubini–Tonelli) *Soit $f : \mathbb{R}^n \to \mathbb{R}_+$ une fonction*[3] *positive. Alors l'intégrale de f par rapport à la mesure de Lebesgue, notée $\int_{\mathbb{R}} \cdots \int_{\mathbb{R}} f(x_1, \ldots, x_n)\,dx_1 \cdots dx_n$ ou plus simplement $\int_{\mathbb{R}^n} f(x)\,dx$ est toujours définie dans $[0, +\infty]$ et peut être calculée en intégrant successivement par rapport à chacune des n variables dans n'importe quel ordre.*

[3] borélienne

Théorème 7.21 (Fubini–Lebesgue) *Soit $f : \mathbb{R}^n \to \mathbb{R}$ une fonction[4] de signe quelconque. Sous l'hypothèse $\int_{\mathbb{R}^n} |f(x)|\, dx < \infty$, qu'en pratique on essaie d'obtenir en utilisant le théorème précédent, l'intégrale de f par rapport à la mesure de Lebesgue $\int_{\mathbb{R}^n} f(x)\, dx$ est bien définie dans $\mathbb{R}$ et peut être calculée en intégrant successivement par rapport à chacune des n variables dans n'importe quel ordre.*

Voyons à présent comment caractériser l'indépendance de variables à densité.

Proposition 7.22 (admise) *Soient X et Y deux variables aléatoires réelles de densités f et g respectivement. Alors*

$$\begin{aligned} X \perp\!\!\!\perp Y \quad &\Leftrightarrow \quad P(X \in A, Y \in B) = \int_A f(x)\, dx \int_B g(y)\, dy \text{ pour tous intervalles } A \text{ et } B \\ &\Leftrightarrow \quad E(h(X,Y)) = \int_{\mathbb{R}} dx \int_{\mathbb{R}} dy\, h(x,y)\ f(x)\, g(y) \text{ pour toute fonction } h \text{ positive ou bornée.} \end{aligned}$$

Définition 7.23 *Soient f et g deux densités sur $\mathbb{R}$. Le produit de convolution de f avec g, noté $f \star g$, est la fonction définie par*

$$f \star g(x) := \int_{\mathbb{R}} f(x-y)\, g(y)\, dy.$$

Exercice 7.24 *Grâce au Théorème de Fubini–Tonelli, montrer que $f \star g$ est une densité et que le produit de convolution est associatif (il est commutatif par un changement de variable immédiat).*

Proposition 7.25 *Soient X et Y deux variables aléatoires réelles indépendantes. Alors*

$$\varphi_{X+Y} = \varphi_X\, \varphi_Y,$$

et si X et Y ont pour densité f et g respectivement, alors $X+Y$ a pour densité $f \star g$.

Dém. La première implication s'obtient en prenant $s = t$ dans le Théorème 7.3(iv). Donc si X et Y ont pour densité f et g respectivement,

$$\varphi_{X+Y}(t) = \int_{\mathbb{R}} dx \int_{\mathbb{R}} dy\, e^{it(x+y)}\ f(x)\, g(y),$$

ce qui devient, grâce au Théorème de Fubini–Lebesgue (l'intégrale double est la succession de deux intégrales simples), après le changement de variable $z = x + y$

$$\varphi_{X+Y}(t) = \int_{\mathbb{R}} dx \int_{\mathbb{R}} dz\, e^{itz}\ f(x)\, g(z-x).$$

À nouveau grâce au Théorème de Fubini–Lebesgue (les deux intégrales successives se font dans n'importe quel ordre),

$$\varphi_{X+Y}(t) = \int_{\mathbb{R}} dz\, e^{itz} \int_{\mathbb{R}} dx\, f(x)\, g(z-x) = \int_{\mathbb{R}} dz\, e^{itz} f \star g(z).$$

[4] borélienne

Comme la fonction caractéristique d'une v.a. caractérise sa loi, la dernière égalité implique que $X + Y$ a $f \star g$ pour densité. $\square$

Proposition 7.26 *Soient X_1 suivant la loi $\mathcal{N}(m_1, \sigma_1^2)$ et X_2 suivant la loi $\mathcal{N}(m_2, \sigma_2^2)$ telles que $X_1 \perp\!\!\!\perp X_2$. Alors $X_1 + X_2$ suit la loi $\mathcal{N}(m_1 + m_2, \sigma_1^2 + \sigma_2^2)$.*

Dém. On calcule la fonction caractéristique de $X_1 + X_2$:

$$\varphi_{X_1+X_2}(t) = \varphi_{X_1}(t)\,\varphi_{X_2}(t) = e^{itm_1 - \frac{\sigma_1^2 t^2}{2}} e^{itm_2 - \frac{\sigma_2^2 t^2}{2}} = e^{it(m_1+m_2) - \frac{(\sigma_1^2+\sigma_2^2)t^2}{2}}$$

qui est bien la fonction caractéristique d'une variable aléatoire $\mathcal{N}(m_1 + m_2, \sigma_1^2 + \sigma_2^2)$. $\square$

Chapter 8

Vecteurs aléatoires

8.1 Définitions et propriétés

Définition 8.1 *On rappelle que pour toutes parties $A_1, \ldots, A_n$ de $\mathbb{R}$, le produit cartésien $A_1 \times \cdots \times A_n$ est la partie de $\mathbb{R}^n$, appelée* pavé, *définie par*

$$(x_1, \ldots, x_n) \in A_1 \times \cdots \times A_n \Leftrightarrow x_i \in A_i \text{ pour tout } i = 1, \ldots, n.$$

Il n'est peut-être pas inutile de rappeler également que toutes les parties de $\mathbb{R}^n$ ne sont pas des pavés (penser par exemple à une boule euclidienne).

Définition 8.2 *Un vecteur aléatoire $X = (X_1, \ldots, X_n)$ est une application*[1] *$X : \Omega \to \mathbb{R}^n$. Comme d'habitude, on appelle loi de X et l'on note P_X la probabilité définie sur $\mathbb{R}^n$ par $P_X(B) = P\left(X^{-1}(B)\right) = P(\{X \in B\})$ pour toute partie*[2] *de $\mathbb{R}^n$.*

En particulier, pour toutes parties[3] *$A_1, \ldots, A_n$ de $\mathbb{R}$, avec $B = A_1 \times \cdots \times A_n$,*

$$P_X(B) = P(X \in A_1 \times \cdots \times A_n) = P(X_1 \in A_1, \ldots, X_n \in A_n) = P\left(\cap_{i=1}^n \{X_i \in A_i\}\right).$$

On appelle aussi P_X la loi jointe *de $(X_1, \ldots, X_n)$ et on appelle P_{X_i} la i-*ème loi marginale *de X. Noter que les lois marginales peuvent s'obtenir à partir de la loi jointe (mais non réciproquement) par*

$$P_{X_i}(A) = P_X(A_1 \times \cdots \times A_n),$$

avec $A_i = A$ et $A_j = \mathbb{R}$ pour tout $j \neq i$.

Proposition 8.3 *Soient X et Y deux vecteurs aléatoires de $\mathbb{R}^n$. Comme d'habitude, on dit que X et Y ont même loi et l'on écrit $X \stackrel{\mathscr{L}}{=} Y$ si $P_X = P_Y$*[4]*. Les vecteurs X et Y ont même loi dès que $P_X(B) = P_Y(B)$ pour tout B de la forme $B = A_1 \times \cdots \times A_n$ où les (A_i) sont des intervalles. Cette dernière assertion reste vraie si l'on remplace 'intervalles' par 'intervalles ouverts', par 'intervalles fermés', ou même par 'intervalles de la forme $]-\infty, x]$'.*

[1] mesurable

[2] borélienne

[3] boréliennes

[4] c'est-à-dire ici si $P_X(B) = P_Y(B)$ pour toute partie borélienne de $\mathbb{R}^n$

Corollaire 8.4 *La loi d'un vecteur aléatoire X est caractérisée par sa fonction de répartition $F_X : \mathbb{R}^n \to [0,1]$ définie par*

$$F_X(x_1, \dots, x_n) := P(X_1 \leq x_1, \dots, X_n \leq x_n).$$

Corollaire 8.5 *Soient $X = (X_1, \dots, X_n)$ et $Y = (Y_1, \dots, Y_n)$ deux vecteurs aléatoires tels que les (X_i) sont indépendants et les (Y_i) sont aussi indépendants. Alors*

$$P_X = P_Y$$

$$\Updownarrow$$

$$P_{X_i} = P_{Y_i} \text{ pour tout } i = 1, \dots, n.$$

Dém. Nous avons déjà vu le sens $\Downarrow$ de l'équivalence, même sans l'hypothèse d'indépendance. Montrons $\Uparrow$. Soient $A_1, \dots, A_n$ des intervalles de $\mathbb{R}$. Par l'hypothèse d'indépendance,

$$P_X(A_1 \times \cdots \times A_n) = P\left(\cap_{i=1}^n \{X_i \in A_i\}\right) = \prod_{i=1}^n P(X_i \in A_i) = \prod_{i=1}^n P_{X_i}(A_i).$$

Comme il en est de même pour Y, et que $P_{X_i}(A_i) = P_{Y_i}(A_i)$ pour tout i par hypothèse, on a donc $P_X(A_1 \times \cdots \times A_n) = P_Y(A_1 \times \cdots \times A_n)$ pour tous intervalles $A_1, \dots, A_n$. D'après la proposition précédente, ceci implique que $P_X = P_Y$. $\square$

Définition 8.6 *On dit que $f : \mathbb{R}^n \to \mathbb{R}_+$ est une densité si $\int_{\mathbb{R}^n} f(x)\, dx = 1$.*

Exemple 8.7 *Soient $C_n := \{x \in \mathbb{R}^n : x_i \in [0,1]\}$ l'hypercube unité de $\mathbb{R}^n$ et $B_n := \{x \in \mathbb{R}^n : \sum_{i=1}^n x_i^2 \leq 1\}$ la boule euclidienne unité de $\mathbb{R}^n$. Soient $c_n := \int_{\mathbb{R}^n} \mathbb{1}_{C_n}(x)\, dx$ et $b_n := \int_{\mathbb{R}^n} \mathbb{1}_{B_n}(x)\, dx$. Alors $\frac{1}{c_n}\mathbb{1}_{C_n}$ et $\frac{1}{b_n}\mathbb{1}_{B_n}$ sont des densités, les densités uniformes sur l'hypercube unité et sur la boule euclidienne unité respectivement.*

Il n'est pas difficile de voir, grâce au Thèorème de Fubini–Tonelli que $c_n = 1$. Le calcul de b_n est hors programme, mais nous savons que $b_1 = 2$ est la longueur de l'intervalle $[-1,1]$, que $b_2 = \pi$ est la surface du disque unité et que $b_3 = 4\pi/3$ est le volume de la boule unité de $\mathbb{R}^3$.

Définition 8.8 *On dit qu'un vecteur aléatoire X a pour densité $f : \mathbb{R}^n \to \mathbb{R}_+$ si pour tout pavé A de $\mathbb{R}^n$, on a*

$$P_X(A) = \int_A f(x)\, dx.$$

Dans ce cas, l'égalité précédente est vraie pour toute partie[5] A de $\mathbb{R}^n$ et l'on a la formule de transfert suivante

$$E(h(X)) = \int_{\mathbb{R}^n} h(x)\, f(x)\, dx,$$

pour toute fonction[6] $h : \mathbb{R}^n \to \mathbb{R}$ positive ou bornée.

[5] borélienne

[6] borélienne

Remarque 8.9 *Le résultat suivant est une généralisation élémentaire de la Proposition 7.22.*

Proposition 8.10 *Soit $X = (X_1, \ldots, X_n)$ un vecteur aléatoire dont la i-ème marginale P_{X_i} a pour densité f_i, pour chaque $i = 1, \ldots, n$. Alors les (X_i) sont indépendantes ssi X admet pour densité la fonction f définie par*

$$f(x_1, \ldots, x_n) = \prod_{i=1}^{n} f_i(x_i) \qquad x_1, \ldots, x_n \in \mathbb{R}.$$

Corollaire 8.11 *Soit $X = (X_1, \ldots, X_n)$ un vecteur aléatoire de $\mathbb{R}^n$. Alors X a pour densité $\mathbb{1}_{C_n}$ ssi les (X_i) sont iid uniformes sur $[0, 1]$.*

Dém. Il suffit de voir que

$$\mathbb{1}_{C_n}(x) = \mathbb{1}_{\{x_i \in [0,1] \ \forall i\}} = \prod_{i=1}^{n} \mathbb{1}_{\{x_i \in [0,1]\}} = \prod_{i=1}^{n} \mathbb{1}_{[0,1]}(x_i) = \prod_{i=1}^{n} f(x_i),$$

où $f = \mathbb{1}_{[0,1]}(x)$ est la densité uniforme sur $[0, 1]$. $\square$

Remarque 8.12 *Appelons* support *d'un vecteur aléatoire X à densité la partie de $\mathbb{R}^n$ où sa densité est non nulle, et notons-le S_X. Supposons que les (X_i) sont indépendantes. Alors X admet pour densité f donnée par $f(x) = \prod_{i=1}^n f_i(x_i)$. Notons alors S_{X_i} la partie de $\mathbb{R}$ où f_i est non nulle. De plus,*

$$x \in S_X \Leftrightarrow f(x) \neq 0 \Leftrightarrow \prod_{i=1}^{n} f_i(x_i) \neq 0 \Leftrightarrow f_i(x_i) \neq 0 \ \forall i \Leftrightarrow x_i \in S_{X_i} \ \forall i \Leftrightarrow x \in \prod_{i=1}^{n} S_{X_i}$$

et donc S_X est égal au pavé $\prod_{i=1}^n S_{X_i}$. Par exemple, on voit que si X est uniforme dans l'hypercube unité, son support (l'hypercube unité) est bien un pavé, le produit cartésien des intervalles unités. En revanche, B_n n'est pas un pavé, donc si X est uniforme dans B_n, les (X_i) ne sont pas *indépendantes.*

8.2 Changement de variable

On commence par rappeler le Théorème d'Inversion Globale et le Théorème de Changement de Variable.

Théorème 8.13 (Inversion Globale) *Soit U ouvert de $\mathbb{R}^n$ et $g : U \to \mathbb{R}^n$ de classe C^1. Si g est injective et que sa jacobienne $g'(x)$ en tout point x de U est inversible, alors $V := g(U)$ est ouvert et $g : U \to V$ est un C^1-difféomorphisme. De plus la jacobienne de g^{-1} est donnée par*

$$\left(g^{-1}\right)'(y) = \left[g'\left(g^{-1}(y)\right)\right]^{-1} \qquad y \in V.$$

En particulier,

$$det\left[\left(g^{-1}\right)'(y)\right] = \frac{1}{det\ \left[g'\left(g^{-1}(y)\right)\right]} \qquad y \in V.$$

Théorème 8.14 (Changement de Variable) *Soit $g : U \to V$ un C^1-difféomorphisme, où U et V sont deux ouverts de $\mathbb{R}^n$. Soit $f : V \to \mathbb{R}$ positive ou bornée telle que $\int_V |f(y)|\, dy < \infty$. Alors*

$$\int_V f(y)\, dy = \int_U f \circ g(x)\ |det\ [g'(x)]|\ dx.$$

Proposition 8.15 (Formule du changement de variable) *Soit X un vecteur aléatoire à valeurs dans U ouvert de $\mathbb{R}^n$ et $g : U \to V$ un C^1-difféomorphisme. Si X a pour densité $f : U \to \mathbb{R}_+$ alors $g(X)$ a pour densité $\tilde{f}$ donnée par*

$$\tilde{f}(y) = \left| det\ \left[\left(g^{-1}\right)'(y) \right] \right|\ f \circ g^{-1}(y) = \frac{f \circ g^{-1}(y)}{|det\ [g' \circ g^{-1}(y)]|} \qquad y \in V.$$

Dém. Soit $h : \mathbb{R}^n \to \mathbb{R}$ positive ou bornée. Alors grâce au changement de variable $x = g^{-1}(y)$,

$$\begin{aligned} E(h(g(X))) &= \int_U h \circ g(x)\, f(x)\, dx \\ &= \int_V h \circ g\left(g^{-1}(y)\right)\ f\left(g^{-1}(y)\right)\ \left|\det \left[\left(g^{-1}\right)'(y)\right]\right|\ dy \\ &= \int_V h(y)\, \frac{f\left(g^{-1}(y)\right)}{|\det [g'(g^{-1}(y))]|}\, dy = \int_V h(y)\, \tilde{f}(y)\, dy. \end{aligned}$$

Comme d'habitude, en notant $Y = g(X)$, le fait que $E(h(Y)) = \int_V h(y)\, \tilde{f}(y)\, dy$ pour toute fonction h positive ou bornée implique que Y a pour densité $\tilde{f}$. □

Exemple 8.16 *Soit (X, Y) ayant pour densité $\frac{1}{\pi}\mathbb{1}_{B_2}$, c'est-à-dire que X et Y sont les coordonnées d'un point uniforme dans le disque unité. On rappelle que $g :]0, +\infty[\times]0, 2\pi[\to \mathbb{R} \setminus \Delta$, où $\Delta := \{(x, y) : y = 0, x > 0\}$ définie par*

$$g(r, \theta) := (r \cos\theta, r \sin\theta)$$

est un C^1-difféomorphisme tel que pour tous r, θ, $det\ [g'(r, \theta)] = r$.

Nous allons donner la loi de $(R, \Theta) := g^{-1}(X, Y)$ en appliquant la proposition précédente. Comme (X, Y) a pour densité $f(x, y) = \frac{1}{\pi}\mathbb{1}_{\{x^2+y^2\leq 1\}}$, et que g^{-1} est un C^1-difféomorphisme, (R, Θ) a pour densité sur $]0, +\infty[\times]0, 2\pi[$

$$\tilde{f}(r, \theta) = |det\ [g'(r, \theta)]|\ f \circ g(r, \theta) = \frac{r}{\pi}\mathbb{1}_{\{r^2\cos^2\theta + r^2\sin^2\theta \leq 1\}} = \frac{r}{\pi}\mathbb{1}_{\{r\leq 1\}}$$

Autrement dit, (R, Θ) a pour densité $f_1(r)\, f_2(\theta)$ où

$$f_1(r) = 2r\mathbb{1}_{]0,1]}(r) \quad \text{et} \quad f_2(\theta) = \frac{1}{2\pi}\mathbb{1}_{]0,2\pi[}(\theta).$$

En particulier R et Θ sont indépendantes et Θ est uniforme sur $]0, 2\pi[$.

8.3 Covariance

Définition 8.17 *Soit $X = (X_1, \dots, X_n)$ un vecteur aléatoire de $\mathbb{R}^n$. On dit que $X \in \mathscr{L}^p$ ssi $X_i \in \mathscr{L}^p$ pour tout i. Si $X \in \mathscr{L}^1$, on note*

$$E(X) := (E(X_1), \dots, E(X_n)).$$

On rappelle que lorsque $X \in \mathscr{L}^1$, on dit que X est intégrable, et lorsque $X \in \mathscr{L}^2$, on dit que X est de carré intégrable.

Définition 8.18 *Soit $(X, Y) \in \mathscr{L}^2$ un vecteur aléatoire de $\mathbb{R}^2$ de carré intégrable. On rappelle qu'alors $X + Y \in \mathscr{L}^2$ et $XY \in \mathscr{L}^1$, et que la covariance de X et Y est définie par*

$$Cov(X, Y) := E(XY) - E(X)\,E(Y) = E\left[(X - E(X))\,(Y - E(Y)\right].$$

L'application $(X, Y) \mapsto Cov(X, Y)$ est bilinéaire et symétrique.

Dém. Tout a été vu dans la Proposition 7.8, sauf la bilinéarité de la covariance. Soient X, Y et Z trois variables aléatoires réelles et $a \in \mathbb{R}$. Alors par linéarité de l'espérance,

$$\begin{aligned}
\mathrm{Cov}(aX + Y, Z) &= E((aX + Y)Z) - E(aX + Y)\,E(Z) \\
&= aE(XZ) + E(YZ) - (aE(X) + E(Y))\,E(Z) \\
&= a\,[E(XZ) - E(X)\,E(Z)] + [E(YZ) - E(Y)\,E(Z)] \\
&= a\mathrm{Cov}(X, Z) + \mathrm{Cov}(Y, Z),
\end{aligned}$$

ce qui prouve la linéarité par rapport à la première composante de la covariance. La linéarité par rapport à la seconde composante s'en déduit par symétrie. □

Exercice 8.19 *On généralise ici le résultat précédent et celui de l'Exercice 7.10. Soit $X = (X_1, \dots, X_n)$ un vecteur aléatoire de carré intégrable. Montrer que*

$$Var\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n Var(X_i) + \sum_{i=1}^n \sum_{j=1,\, j \neq i}^n Cov(X_i, X_j).$$

On pourra écrire

$$Var\left(\sum_{i=1}^n X_i\right) = Cov\left(\sum_{i=1}^n X_i, \sum_{j=1}^n X_j\right).$$

Exercice 8.20 *On se rappelle que deux variables aléatoires réelles X et Y peuvent avoir une covariance nulle sans être indépendantes. Montrer que c'est le cas du couple (X, Y) suivant la loi uniforme sur le disque unité.*

Théorème 8.21 (Inégalité de Cauchy–Schwarz) *Soient X et Y deux v.a.r. de carré intégrable. Alors*

$$|Cov(X, Y)| \leq \sqrt{Var(X)}\sqrt{Var(Y)}.$$

Définition 8.22 *Soient X et Y deux v.a.r. de carré intégrable. On définit la* corrélation *de X et Y par*

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

En particulier, pour tous $a, b > 0$, $Corr(aX, bY) = Corr(X,Y)$ si bien que la corrélation ne dépend pas de l'unité dans laquelle sont exprimées les variables X et Y. De plus par l'inégalité de Cauchy–Schwarz, la corrélation est toujours comprise entre -1 et 1.

Définition 8.23 *Soit $X = (X_1, \ldots, X_n)$ un vecteur aléatoire de $\mathscr{L}^2$. On appelle* matrice de variance-covariance *de X, que l'on note C_X, la matrice d'élément générique $Cov(X_i, X_j)$. En particulier C_X est une matrice symétrique.*

Exercice 8.24 *Montrer que pour tout vecteur u de $\mathbb{R}^n$, la variable aléatoire ${}^t\!uX = \sum_{i=1}^n u_i X_i$ a pour variance $Var({}^t\!uX) = {}^t\!u\, C_X\, u$.*

Exercice 8.25 *Soit A une matrice carrée et Y le vecteur aléatoire défini par $Y = AX$. Montrer que*

$$C_Y = A\, C_X\, {}^t\!A.$$

Définition 8.26 *Soit $X = (X_1, \ldots, X_n)$ un vecteur aléatoire de $\mathscr{L}^2$ d'espérance $m \in \mathbb{R}^n$ et de matrice de variance-covariance C_X. On dit que X est un* vecteur gaussien *si pour tout vecteur u de $\mathbb{R}^n$, ${}^t\!uX = \sum_{i=1}^n u_i X_i$ est une variable aléatoire gaussienne, c'est-à-dire suit $\mathscr{N}({}^t\!um, {}^t\!u\, C_X\, u)$. Si C_X est inversible, alors X est un vecteur gaussien ssi il a une densité f donnée par*

$$f(x_1, \ldots, x_n) = \frac{1}{\sqrt{(2\pi)^n \det C_X}} \exp\left\{-\frac{1}{2}\, {}^t(x-m) C_X^{-1} (x-m)\right\} \qquad x \in \mathbb{R}^n.$$

Corollaire 8.27 *Soit $X = (X_1, \ldots, X_n)$ un vecteur gaussien. Alors les (X_i) sont indépendantes ssi $Cov(X_i, X_j) = 0$ pour tous $i \neq j$.*

Chapter 9

Convergences p.s., en probabilité et dans $\mathscr{L}^p$

Dans ce chapitre, on se donne une v.a.r. X et une suite de v.a.r. (X_n) définies sur un **même** espace de probabilité $(\Omega, \mathscr{A}, P)$.

9.1 Convergence p.s. et en probabilité

Définition 9.1 *On dit que la suite* (X_n) converge presque sûrement *vers* X*, abrégé p.s., et l'on note* $\lim_n X_n \stackrel{p.s.}{=} X$ *si* $P(\lim_n X_n = X) = 1$*, c'est-à-dire*

$$P(\{\omega \in \Omega : \lim_n X_n(\omega) = X(\omega)\}) = 1.$$

Proposition 9.2 *Chacun des deux critères suivants implique la convergence p.s. de* (X_n) *vers* X*.*

(i) Il existe une suite de réels positifs (ε_n) *telle que* $\lim_n \varepsilon_n = 0$ *et*

$$\sum_n P(|X_n - X| > \varepsilon_n) < \infty.$$

(ii) Pour tout $\varepsilon > 0$*,*

$$\sum_n P(|X_n - X| > \varepsilon) < \infty.$$

Dém. (i) Soit A_n l'évènement $\{|X_n - X| > \varepsilon_n\}$ et $A := \limsup_n A_n$. Par le lemme de Borel–Cantelli, comme $\sum_n P(A_n) < \infty$, $P(A) = 0$, autrement dit $P({}^cA) = 1$. Or

$$\begin{aligned} {}^cA = \bigcup_N \bigcap_{n \geq N} {}^cA_n &= \{\omega : \exists N, \forall n \geq N, \omega \notin A_n\} \\ &= \{\omega : \exists N, \forall n \geq N, |X_n(\omega) - X(\omega)| \leq \varepsilon_n\}. \end{aligned}$$

Il est alors facile de voir que ${}^cA \subset \{\omega : \lim_n X_n(\omega) = X(\omega)\}$, et donc que $P(\lim_n X_n = X) = 1$.

(ii) Soit (ε_k) une suite de réels strictement positifs telle que $\lim_k \varepsilon_k = 0$. Pour tout entier k, soit $B_{k,n}$ l'évènement $\{|X_n - X| > \varepsilon_k\}$ et $B_k := \limsup_n B_{k,n}$. D'après ce qui précède, comme $\sum_n P(B_{k,n}) < \infty$, $P(B_k) = 0$ et

$$^c\!B_k = \{\omega : \exists N, \forall n \geq N, |X_n(\omega) - X(\omega)| \leq \varepsilon_k\}.$$

Soit maintenant $B := \cup_k B_k$. Alors $P(B) \leq \sum_k P(B_k) = 0$, donc $P(^c\!B) = 1$ et

$$^c\!B = \bigcap_k {^c\!B_k} = \{\omega : \forall k, \exists N, \forall n \geq N, |X_n(\omega) - X(\omega)| \leq \varepsilon_k\},$$

c'est-à-dire que $^c\!B = \{\omega : \lim_n X_n(\omega) = X(\omega)\}$. □

Définition 9.3 *On dit que la suite* (X_n) converge en probabilité *vers* X*, et l'on note* $\lim_n X_n \stackrel{P}{=} X$*, si pour tout* $\varepsilon > 0$*,*

$$\lim_n P(|X_n - X| > \varepsilon) = 0.$$

Remarque 9.4 *On peut montrer facilement que* $\lim_n X_n \stackrel{P}{=} X$ *ssi pour tout* $\varepsilon > 0$*,*

$$\lim_n P(|X_n - X| \geq \varepsilon) = 0,$$

autrement dit on peut indifféremment se servir d'une inégalité stricte ou large dans la probabilité $P(|X_n - X| > \varepsilon)$ *pour caractériser la convergence en probabilité.*

Proposition 9.5 *Si* (X_n) *converge p.s. vers* X *alors* (X_n) *converge en probabilité vers* X*.*

Si (X_n) *converge en probabilité vers* X *alors il existe une suite extraite* $(X_{\varphi(n)})$ *qui converge p.s. vers* X *(où* $\varphi : \mathbb{N} \to \mathbb{N}$ *ne dépend pas de* ω*).*

Dém. Démontrons que la convergence p.s. implique la convergence en probabilité. Soit $\varepsilon > 0$ et soit A_N l'évènement $\{\omega : \forall n \geq N, |X_n - X| \leq \varepsilon\} = \cap_{n \geq N}\{|X_n - X| \leq \varepsilon\}$. La suite (A_n) est croissante de limite

$$A = \lim_N \uparrow A_N = \cup_N A_N = \{\omega : \exists N, \forall n \geq N, |X_n - X| \leq \varepsilon\}.$$

Comme (X_n) converge p.s. vers X, $P(A) = 1$, donc par continuité à gauche de P, $\lim_N P(A_N) = P(A) = 1$. Or $A_N \subset \{|X_N - X| \leq \varepsilon\}$, donc

$$P(|X_N - X| \leq \varepsilon) \geq P(A_N) \to 1,$$

ce qui prouve que $P(|X_N - X| > \varepsilon) \to 0$.

Démontrons à présent la deuxième assertion. On suppose donc que (X_n) converge en probabilité vers X. On se donne une suite arbitraire (ε_n) de réels strictement positifs telle que $\sum_n \varepsilon_n < \infty$, à partir de laquelle on va construire l'injection croissante φ. D'abord $\varphi(0) = 0$. Puis on définit par récurrence sur $n \geq 1$

$$\varphi(n) := \min\{k > \varphi(n-1) : P(|X_k - X| > \varepsilon_n) \leq \varepsilon_n\}.$$

Puisque (X_n) converge en probabilité vers X, $\varphi(n)$ est fini et de plus

$$P(\left|X_{\varphi(n)} - X\right| > \varepsilon_n) \leq \varepsilon_n,$$

donc $\sum_n P(\left|X_{\varphi(n)} - X\right| > \varepsilon_n) < \infty$, ce qui implique la convergence p.s. de $(X_{\varphi(n)})$ vers X, grâce à la Proposition 9.2(i). □

Proposition 9.6 *Les propriétés usuelles de la convergence dans $\mathbb{R}$ sont également valables pour les convergences p.s. et en probabilité. En particulier,*

(i) (unicité de la limite) Si (X_n) converge p.s. (resp. en probabilité) vers X et si (X_n) converge aussi p.s. (resp. en probabilité) vers Y, alors $P(X = Y) = 1$.

(ii) Si (X_n) converge p.s. (resp. en probabilité) vers X alors pour toute fonction continue f, $(f(X_n))$ converge p.s. (resp. en probabilité) vers $f(X)$.

Dém. Montrons d'abord chacun des deux résultats dans le cas de la convergence p.s. Pour (i) définissons A l'évènement $\{\lim_n |X_n - X| = 0\}$ et B l'évènement $\{\lim_n |X_n - Y| = 0\}$. Par hypothèse, $P(A) = P(B) = 1$ donc $P(A \cap B) = 1$ car

$$1 - P(A \cap B) = P(^c A \cup {^c B}) \leq P(^c A) + P(^c B) = 0.$$

De plus pour tout $\omega \in A \cap B$, la suite $(X_n(\omega))$ converge vers $X(\omega)$ et vers $Y(\omega)$ donc par unicité de la limite, $X(\omega) = Y(\omega)$. On a donc $A \cap B \subset \{X = Y\}$ et comme $P(A \cap B) = 1$, $P(X = Y) = 1$.

Pour (ii), soit A l'évènement $\{\lim_n |X_n - X| = 0\}$ et B l'évènement $\{\lim_n |f(X_n) - f(X)| = 0\}$. Pour tout $\omega \in A$, $\lim_n |f(X_n(\omega)) - f(X(\omega))| = 0$, c'est-à-dire que $A \subset B$. Mais par hypothèse $P(A) = 1$, donc $P(B) = 1$.

Montrons à présent les deux résultats dans le cas de la convergence en probabilité. Pour (i), nous allons nous servir de la Proposition 9.5. On sait qu'il existe une injection croissante φ telle que $(X_{\varphi(n)})$ converge p.s. vers X. Mais comme (X_n) converge vers Y en probabilité, $(X_{\varphi(n)})$ converge aussi vers Y en probabilité. Donc il existe une injection croissante ψ telle que $(X_{\varphi\circ\psi(n)})$ converge p.s. vers Y. Mais $(X_{\varphi\circ\psi(n)})$ converge aussi p.s. vers X, donc grâce au résultat précédent, $P(X = Y) = 1$.

On peut montrer (i) de façon plus classique, sans utiliser les suites extraites. Soit $\varepsilon > 0$. Notons que par l'inégalité triangulaire,

$$|X - Y| \leq |X - X_n| + |X_n - Y|,$$

si bien que

$$\begin{aligned} P(|X - Y| > \varepsilon) &\leq P(|X - X_n| + |X_n - Y| > \varepsilon) \\ &\leq P\left(|X - X_n| > \frac{\varepsilon}{2} \text{ ou } |X_n - Y| > \frac{\varepsilon}{2}\right) \\ &\leq P\left(|X - X_n| > \frac{\varepsilon}{2}\right) + P\left(|X_n - Y| > \frac{\varepsilon}{2}\right). \end{aligned}$$

Or chacun des deux termes de la dernière somme tend vers 0 lorsque $n \to \infty$. Donc $P(|X - Y| > \varepsilon) = 0$ pour tout $\varepsilon > 0$. C'est un exercice classique d'en déduire que $P(X \neq Y) = 0$.

Pour (ii), fixons $\varepsilon > 0$. Soit d'abord $M > 0$ tel que $P(|X| > M) < \frac{\varepsilon}{3}$. Toute fonction continue f est uniformément continue sur l'intervalle borné (compact) $I := [-M-1, M+1]$. Donc il existe $\alpha > 0$ tel que pour tous $x, y \in I$,

$$|x - y| < \alpha \Rightarrow |f(x) - f(y)| < \frac{\varepsilon}{3}.$$

Enfin par hypothèse, (X_n) converge en probabilité vers X, donc il existe un entier N_1 tel que pour tout $n \geq N_1$, $P(|X_n - X| \geq \alpha) < \frac{\varepsilon}{3}$ et un entier N_2 tel que pour tout $n \geq N_2$, $P(|X_n - X| \geq 1) < \frac{\varepsilon}{3}$. En désignant par A_n l'évènement $\{|f(X_n) - f(X)| > \varepsilon\}$, on obtient alors que pour tout $n \geq \max(N_1, N_2)$, $P(A_n)$ vaut

$$\begin{aligned}
&= P(A_n, |X| \geq M) + P(A_n, |X| < M, |X_n| > M+1) + P(A_n, |X| < M, |X_n| \leq M+1) \\
&\leq P(|X| \geq M) + P(|X - X_n| > 1) + P(A_n, |X| < M, |X_n| \leq M+1, |X_n - X| > \alpha) \\
&+ P(A_n, |X| < M, |X_n| \leq M+1, |X_n - X| \leq \alpha) \\
&\leq P(|X| \geq M) + P(|X - X_n| > 1) + P(|X_n - X| > \alpha) + 0 \\
&\leq \frac{\varepsilon}{3} + \frac{\varepsilon}{3} + \frac{\varepsilon}{3} = \varepsilon,
\end{aligned}$$

ce qui montre la convergence en probabilité de $(f(X_n))$ vers $f(X)$. □

Proposition 9.7 *Soit* (X_n) *une suite de v.a.r. de carré intégrable telle que* $\lim_n E(X_n) = a$ *et* $\lim_n$ *Var*$(X_n) = 0$. *Alors* (X_n) *converge en probabilité vers* a.

Dém. Par l'inégalité de Markov,

$$\begin{aligned}
P(|X_n - a| > \varepsilon) &= P((X_n - a)^2 > \varepsilon^2) \\
&\leq \frac{E((X_n - a)^2)}{\varepsilon^2} \\
&= \frac{\mathrm{Var}(X_n - a) + (E(X_n - a))^2}{\varepsilon^2} \\
&= \frac{\mathrm{Var}(X_n) + (E(X_n) - a)^2}{\varepsilon^2}
\end{aligned}$$

qui tend vers 0 par les deux hypothèses de la proposition. □

Proposition 9.8 *Soit* (X_n) *une suite de v.a.r. de même loi et* (a_n) *une suite de réels telle que* $\lim_n a_n = 0$. *Alors* $(a_n X_n)$ *converge en probabilité vers* 0.

Dém. On a

$$P(|a_n X_n| > \varepsilon) = P(|a_n X_1| > \varepsilon),$$

qui tend vers 0 lorsque $n \to \infty$. □

9.2 Convergence dans $\mathscr{L}^p$

Définition 9.9 *On dit que la suite* (X_n) converge dans $\mathscr{L}^p$ *vers* X, *et l'on note* $\lim_n X_n \overset{\mathscr{L}^p}{=} X$ *si* $X \in \mathscr{L}^p$, *si pour tout* n $X_n \in \mathscr{L}^p$, *et si*

$$\lim_n E\left(|X_n - X|^p\right) = 0.$$

Dans ce cas, on a aussi la convergence

$$\lim_n E\left(|X_n|^p\right) = E\left(|X|^p\right),$$

et dans le cas où $p = 1$, *on a également* $\lim_n E(X_n) = E(X)$.

Dans le cas où $p = 1$, *on dit que* (X_n) *converge* en moyenne *et dans le cas où* $p = 2$, *on dit que* (X_n) *converge* en moyenne quadratique.

Dém. Démontrons la convergence des moments. Comme

$$\left(E\left(|Y|^p\right)\right)^{\frac{1}{p}} \leq \left(E\left(|Y - X|^p\right)\right)^{\frac{1}{p}} + \left(E\left(|X|^p\right)\right)^{\frac{1}{p}}$$

et l'inégalité symétrique obtenue en échangeant X et Y, on a

$$\left|\left(E\left(|Y|^p\right)\right)^{\frac{1}{p}} - \left(E\left(|X|^p\right)\right)^{\frac{1}{p}}\right| \leq \left(E\left(|Y - X|^p\right)\right)^{\frac{1}{p}}.$$

En remplaçant Y par X_n et en utilisant la convergence de (X_n) vers X dans $\mathscr{L}^p$, on obtient

$$\lim_{n\to\infty} \left|\left(E\left(|X_n|^p\right)\right)^{\frac{1}{p}} - \left(E\left(|X|^p\right)\right)^{\frac{1}{p}}\right| = 0,$$

autrement dit $\lim_n \left(E\left(|X_n|^p\right)\right)^{\frac{1}{p}} = \left(E\left(|X|^p\right)\right)^{\frac{1}{p}}$, d'où le résultat en prenant l'image des éléments de cette suite convergente par la fonction continue $x \mapsto x^p$. La convergence des espérances se montre de la même manière. $\square$

Proposition 9.10 *Si* (X_n) *converge vers* X *dans* $\mathscr{L}^p$, *alors* (X_n) *converge vers* X *en probabilité.*

Dém. Par l'inégalité de Markov, on a

$$P(|X_n - X| > \varepsilon) = P(|X_n - X|^p > \varepsilon^p) \leq \frac{E\left(|X_n - X|^p\right)}{\varepsilon^p},$$

qui par hypothèse tend vers 0 lorsque $n \to \infty$. $\square$

Remarque 9.11 *La réciproque n'est pas vraie, comme en témoigne le contre-exemple suivant. Soit* $p > 0$ *et* X_n *la v.a.r. prenant uniquement la valeur* 0 *ou la valeur* n, *dont la loi est donnée par*

$$P(X_n = n) = n^{-\frac{p}{2}} = 1 - P(X_n = 0).$$

Alors (X_n) tend vers 0 en probabilité car

$$P(|X_n| > \varepsilon) = n^{-\frac{p}{2}} \to 0$$

lorsque $n \to \infty$. Montrons par l'absurde que (X_n) ne converge pas dans $\mathscr{L}^p$. Si (X_n) converge vers une v.a. X dans $\mathscr{L}^p$, alors, par la proposition précédente, (X_n) converge vers X en probabilité, et donc par unicité de la limite $P(X = 0) = 1$. Mais $E(|X_n - 0|^p) = n^{\frac{p}{2}}$, qui ne tend pas vers 0.

Remarque 9.12 *Si la suite (X_n) converge vers Y dans $\mathscr{L}^p$ et converge vers X en probabilité alors $P(X = Y) = 1$ et (X_n) converge vers X dans $\mathscr{L}^p$.*

En effet, on sait d'après la proposition précédente que (X_n) converge en probabilité vers Y. Donc par unicité de la limite, $P(Y = X) = 1$. Ainsi $X \in \mathscr{L}^p$ et $E(|X_n - X|^p) = E(|X_n - Y|^p)$, qui tend vers 0 lorsque $n \to \infty$.

Proposition 9.13 *Soient $p > q > 0$. Si (X_n) converge vers X dans $\mathscr{L}^p$ alors (X_n) converge vers X dans $\mathscr{L}^q$. En particulier, si (X_n) converge dans $\mathscr{L}^2$ alors (X_n) converge dans $\mathscr{L}^1$.*

Dém. Soit $\varepsilon \in]0,1[$. Alors $E(|X_n - X|^q)$ vaut

$$\begin{aligned}
&= E\left(|X_n - X|^q \mathbb{1}_{|X_n - X|^q \leq \varepsilon/3}\right) + E\left(|X_n - X|^q \mathbb{1}_{|X_n - X|^q \in]\varepsilon/3, 1[}\right) + E\left(|X_n - X|^q \mathbb{1}_{|X_n - X|^q \geq 1}\right) \\
&\leq \frac{\varepsilon}{3} + P\left(|X_n - X| > (\varepsilon/3)^{1/q}\right) + E\left(|X_n - X|^p\right).
\end{aligned}$$

Comme (X_n) tend vers X dans $\mathscr{L}^p$ et donc aussi en probabilité, chacun des deux derniers termes tend vers 0 lorsque $n \to \infty$, donc il existe un rang N tel que pour tout $n \geq N$, chacun de ces termes est majoré par $\varepsilon/3$. Donc pour tout $n \geq N$, $E(|X_n - X|^q) \leq \varepsilon$. □

Proposition 9.14 (Convergence $\mathscr{L}^p$-dominée) *Si (X_n) converge vers X en probabilité, et s'il existe une v.a.r. positive $Y \in \mathscr{L}^p$ définie sur le même espace, telle que $|X_n| \leq Y$ pour tout n, alors (X_n) converge vers X dans $\mathscr{L}^p$.*

Dém. Soit $\varepsilon > 0$. On voit d'abord assez facilement que $P(|X| \leq Y) = 1$ et que $P(|X_n - X| \leq 2Y) = 1$. Soit M tel que

$$E\left((2Y)^p \mathbb{1}_{(2Y)^p \geq M}\right) < \frac{\varepsilon}{3},$$

dont l'existence est garantie par convergence dominée. Soit N tel que pour tout $n \geq N$,

$$P\left(|X_n - X| > \left(\frac{\varepsilon}{3}\right)^{\frac{1}{p}}\right) \leq \frac{\varepsilon}{3M}.$$

Alors pour tout $n \geq N$,

$$\begin{aligned}
E\left(|X_n - X|^p\right) &= E\left(|X_n - X|^p \, \mathbb{1}_{|X_n - X|^p \leq \frac{\varepsilon}{3}}\right) \\
&+ E\left(|X_n - X|^p \, \mathbb{1}_{|X_n - X|^p > M}\right) \\
&+ E\left(|X_n - X|^p \, \mathbb{1}_{|X_n - X|^p \in]\frac{\varepsilon}{3}, M]}\right) \\
&\leq \frac{\varepsilon}{3} + E\left((2Y)^p \mathbb{1}_{(2Y)^p \geq M}\right) + MP\left(|X_n - X|^p > \varepsilon/3\right) \\
&\leq \frac{\varepsilon}{3} + \frac{\varepsilon}{3} + \frac{\varepsilon}{3} = \varepsilon,
\end{aligned}$$

ce qui montre la convergence de (X_n) vers X dans $\mathscr{L}^p$. □

Exemple 9.15 *Soit une suite (X_n) de variables aléatoires indépendantes telles que X_n ne peut prendre que la valeur 0 ou la valeur $\beta_n > \eta > 0$. On définit $\alpha_n = P(X_n = \beta_n)$. Comme $P(|X_n| > \varepsilon) = \alpha_n$ dès que $\varepsilon < \eta$, on voit que (X_n) converge vers 0 en probabilité ssi $\lim_n \alpha_n = 0$. On suppose désormais que $\lim_n \alpha_n = 0$.*

Nous allons maintenant déterminer à quelle condition (X_n) converge p.s. (vers 0). D'après les hypothèses, on voit que pour tout ω, la suite $(X_n(\omega))$ converge vers 0 ssi elle est stationnaire à 0, c'est-à-dire ssi elle ne prend la valeur β_n qu'un nombre fini de fois, autrement dit,

$$\{\lim_n X_n = 0\} = \left\{\sum_n \mathbb{1}_{X_n \neq 0} < \infty\right\}.$$

Attention, cette égalité n'est vraie que parce qu'ici (X_n) ne peut tendre vers 0 en se rapprochant progressivement de 0 (car soit $X_n = 0$, soit $X_n > \eta$). Or les évènements $(\{X_n \neq 0\})$ sont indépendants, donc par le lemme de Borel–Cantelli, $P(\sum_n \mathbb{1}_{X_n \neq 0} = \infty) = 0$ ou 1 selon que $\sum_n \alpha_n$ converge ou diverge. Ainsi la suite (X_n) converge vers 0 p.s. ssi $\sum_n \alpha_n$ converge. On retrouve bien le fait que la convergence p.s. implique la convergence en probabilité puisque la convergence de la série de terme général (α_n) implique la converge de la suite (α_n) vers 0.

Si la suite (X_n) converge dans $\mathscr{L}^p$, on sait qu'elle doit converger vers 0. Or

$$E\left(|X_n - 0|^p\right) = \alpha_n \beta_n^p,$$

Donc la convergence a lieu dans $\mathscr{L}^p$ ssi $\lim_n \alpha_n \beta_n^p = 0$. Ici aussi, on retrouve que la convergence dans $\mathscr{L}^p$ implique la convergence en probabilité, car si la suite $(\alpha_n \beta_n^p)$ tend vers 0 alors la suite (α_n) tend vers 0.

Chapter 10

Convergence en loi

Ici on se donne $X, X_0, X_1, X_2, \ldots$ des v.a. qui ne sont pas forcément définies sur le même espace. En effet, la convergence en loi consiste à comparer P_{X_n} et P_X et non plus X_n et X.

10.1 Définition et propriétés

Définition 10.1 *On dit que la suite* (X_n) converge en loi *vers* X*, et l'on note* $\lim_n X_n \overset{\mathscr{L}}{=} X$ *si pour toute fonction* $f : \mathbb{R} \to \mathbb{R}$ *continue et bornée,*

$$\lim_n E(f(X_n)) = E(f(X)).$$

Proposition 10.2 *Si* X *et les éléments de la suite* (X_n) *sont définis sur le même espace, et si la suite* (X_n) *converge en probabilité vers* X *alors elle converge en loi vers* X*.*

Dém. Soit f continue et bornée. Comme (X_n) converge en probabilité vers X et comme f est continue, $(f(X_n))$ converge en probabilité vers $f(X)$. De plus f est bornée, donc il existe $M > 0$ tel que $|f(x)| \leq M$ pour tout x. Donc la suite $(f(X_n))$ est dominée par M qui est bien sûr intégrable ($E(M) = M$). Par convergence $\mathscr{L}^1$-dominée, $(f(X_n))$ converge donc dans $\mathscr{L}^1$ vers $f(X)$. Ceci implique la convergence des espérances, c'est-à-dire que $E(f(X_n))$ converge vers $E(f(X))$. □

Proposition 10.3 *Si la suite* (X_n) *converge en loi vers une constante* c*, alors elle converge en probabilité vers* c*.*

Dém. Soit $\varepsilon > 0$. Soit φ_ε une fonction continue telle que $\varphi_\varepsilon(c) = 1$ et telle que pour tout $x \in \mathbb{R}$, $0 \leq \varphi_\varepsilon(x) \leq \mathbb{1}_{|x-c| \leq \varepsilon}$. Il suffit par exemple de prendre la fonction en triangle définie par

$$\varphi_\varepsilon(x) = \begin{cases} \frac{x-c}{\varepsilon} + 1 & \text{si} \quad x \in [c-\varepsilon, c] \\ -\frac{x-c}{\varepsilon} + 1 & \text{si} \quad x \in [c, c+\varepsilon] \\ 0 & \text{sinon.} \end{cases}$$

On a alors

$$P(|X_n - c| \leq \varepsilon) = E(\mathbb{1}_{\{|X_n-c|\leq\varepsilon\}}) \geq E(\varphi_\varepsilon(X_n)).$$

Or φ_ε est continue et bornée et (X_n) converge en loi vers c, donc

$$\lim_n E(\varphi_\varepsilon(X_n)) = E(\varphi_\varepsilon(c)) = \varphi_\varepsilon(c) = 1.$$

L'inégalité qui précédait implique que $\lim_n P(|X_n - c| \leq \varepsilon) = 1$, c'est-à-dire que (X_n) converge en probabilité vers c. □

Définition 10.4 *On dit qu'une fonction $f : \mathbb{R} \to \mathbb{R}$ est* à support compact *s'il existe un intervalle fermé borné I tel que $f(x) = 0$ pour tout $x \notin I$. Noter qu'une fonction continue à support compact est bornée.*

Proposition 10.5 *La suite (X_n) converge en loi vers X ssi pour toute fonction $f : \mathbb{R} \to \mathbb{R}$ continue à support compact,*

$$\lim_n E(f(X_n)) = E(f(X)).$$

Dém. Le sens direct est évident. En effet, si f est continue à support compact, elle est continue et bornée, donc la convergence en loi de (X_n) vers X implique que $\lim_n E(f(X_n)) = E(f(X))$.

Réciproquement, montrons que si $\lim_n E(f(X_n)) = E(f(X))$ pour toute fonction f continue à support compact, alors c'est vrai également pour toute fonction f continue bornée. Soit donc f continue et bornée et soit $\varepsilon > 0$. Comme f est bornée, il existe C tel que $|f(x)| \leq C$ pour tout x. De plus, il existe M tel que

$$P(|X| > M) \leq \frac{\varepsilon}{4C},$$

puisque $\lim_{M\to\infty} P(|X| > M) = 0$. On peut facilement construire une fonction $\varphi : \mathbb{R} \to [0, 1]$ continue à support compact telle que $\varphi(x) = 1$ pour tout $x \in [-M, M]$. Alors

$$E(\varphi(X)) \geq E(\varphi(X)\mathbb{1}_{\{X\in[-M,M]\}}) = P(|X| \leq M) \geq 1 - \frac{\varepsilon}{4C}.$$

De plus, comme φ est continue à support compact, par hypothèse $\lim_n E(\varphi(X_n)) = E(\varphi(X))$, si bien qu'il existe un rang N_1 tel que pour tout $n \geq N_1$,

$$E(\varphi(X_n)) \geq E(\varphi(X)) - \frac{\varepsilon}{4C} \geq 1 - \frac{\varepsilon}{2C}.$$

Enfin, φf est continue à support compact donc il existe $N \geq N_1$ tel que pour tout $n \geq N$,

$$|E(\varphi f(X_n)) - E(\varphi f(X))| \leq \frac{\varepsilon}{4}.$$

Ainsi, par l'inégalité triangulaire, la différence $|E(f(X_n)) - E(f(X))|$ est majorée par

$$\begin{aligned}
&\leq \; |E(f(X_n)) - E(\varphi f(X_n))| + |E(\varphi f(X_n)) - E(\varphi f(X))| + |E(\varphi f(X)) - E(f(X))| \\
&\leq \; E(|f(X_n)|\,(1-\varphi(X_n))) + \frac{\varepsilon}{4} + E(|f(X)|\,(1-\varphi(X))) \\
&\leq \; C(1 - E(\varphi(X_n))) + \frac{\varepsilon}{4} + C(1 - E(\varphi(X))) \\
&\leq \; \frac{\varepsilon}{2} + \frac{\varepsilon}{4} + \frac{\varepsilon}{4} = \varepsilon,
\end{aligned}$$

et ce dès que $n \geq N$, ce qui achève de prouver que $\lim_n E(f(X_n)) = E(f(X))$. □

10.2 Caractérisations de la convergence en loi

Proposition 10.6 *Si pour tout n X_n est à valeurs dans $\mathbb{Z}$ et que X est également à valeurs dans $\mathbb{Z}$, alors la suite (X_n) converge en loi vers X si pour tout $k \in \mathbb{Z}$,*

$$\lim_n P(X_n = k) = P(X = k).$$

Dém. Sens direct. Supposons que (X_n) converge en loi. Soit $\varphi_k : \mathbb{R} \to \mathbb{R}$ la fonction en triangle définie par

$$\varphi_k(x) = \begin{cases} x - k + 1 & \text{si} \quad x \in [k-1, k] \\ -x + k + 1 & \text{si} \quad x \in [k, k+1] \\ 0 & \text{sinon.} \end{cases}$$

Alors φ_k est continue et bornée et pour toute v.a. entière Y, $E(\varphi_k(Y)) = \sum_{j \in \mathbb{Z}} P(Y = j)\varphi_k(j) = P(Y = k)$. Par la convergence en loi de (X_n) vers X, on a donc

$$\lim_n P(X_n = k) = \lim_n E(\varphi_k(X_n)) = E(\varphi_k(X)) = P(X = k).$$

Réciproquement, supposons que $\lim_n P(X_n = k) = P(X = k)$ pour tout entier k. Soit $f : \mathbb{R} \to \mathbb{R}$ continue à support compact. Il existe un entier $M > 0$ tel que f est nulle hors de $[-M, M]$. Donc pour toute v.a. entière Y

$$E(f(Y)) = \sum_{j \in \mathbb{Z}} P(Y = j) f(j) = \sum_{j=-M}^{M} P(Y = j) f(j).$$

Ainsi, comme cette somme est finie,

$$\lim_n E(f(X_n)) = \sum_{j=-M}^{M} \lim_n P(X_n = j) f(j) = \sum_{j=-M}^{M} P(X = j) f(j) = E(f(X)),$$

ce qui donne la convergence en loi de (X_n) vers X grâce à la proposition précédente. □

Proposition 10.7 (admise) *Soit X une v.a. entière positive et (X_n) une suite de v.a. entières positives. La suite (X_n) converge en loi vers X ssi les fonctions génératrices de X_n convergent simplement vers la fonction génératrice de X, c'est-à-dire*

$$\lim_n f_{X_n}(s) = f_X(s) \qquad s \in [0,1].$$

Corollaire 10.8 *Soit (X_n) une suite de v.a.e., où X_n suit la loi binomiale de paramètres p_n et n, avec $\lim_n np_n = \theta$. Alors (X_n) converge en loi vers une variable de Poisson de paramètre θ.*

Dém. Noter que $p_n \sim \theta/n$, de sorte qu'en particulier $\lim_n p_n = 0$. La fonction génératrice de X_n vaut

$$f_{X_n}(s) = (1 - p_n + p_n s)^n = e^{n\ln(1-p_n(1-s))} = e^{n(-p_n(1-s)+o(p_n))},$$

qui converge vers $e^{-\theta(1-s)}$ lorsque $n \to \infty$, qui est la fonction génératrice d'une variable de Poisson de paramètre θ. □

Proposition 10.9 (admise) *Les trois assertions suivantes sont équivalentes :*

(i) *La suite (X_n) converge en loi vers X.*

(ii) $\lim_n F_{X_n}(x) = F_X(x)$ *pour tout $x \in \mathbb{R}$ qui n'est pas un point de discontinuité de F_X.*

(iii) $\lim_n \varphi_{X_n}(t) = \varphi_X(t)$ *pour tout $t \in \mathbb{R}$.*

Corollaire 10.10 *Soit (X_n) une suite de v.a. normales telles que*

$$\lim_n E(X_n) = \mu \quad \text{et} \quad \lim_n Var(X_n) = \sigma^2.$$

Alors (X_n) converge en loi vers une v.a. normale d'espérance μ et de variance σ^2.

Dém. Il suffit d'appliquer le dernier critère de la proposition précédente (convergence des fonctions caractéristiques). □

Chapter 11

Loi des Grands Nombres et Théorème Central Limite

Dans ce chapitre, on se donne une suite (X_n) de v.a.r. i.i.d. intégrables définies sur un même espace $(\Omega, \mathscr{A}, P)$. On note $\mu := E(X_0)$ leur espérance commune et l'on s'intéresse à la moyenne

$$M_n := \frac{1}{n}\sum_{i=1}^{n} X_i.$$

11.1 Loi des Grands Nombres

Théorème 11.1 (Loi Forte des Grands Nombres) *La suite (M_n) converge p.s. et dans $\mathscr{L}^1$ vers μ.*

Dém. Dans cette démonstration, nous supposons pour simplifier l'argumentation que les (X_n) sont de carré intégrable et l'on note $\sigma^2 := \mathrm{Var}(X_0)$ leur variance commune. On rappelle que $E(M_n) = \mu$ et que

$$\mathrm{Var}(M_n) = \frac{\sigma^2}{n}.$$

Comme on peut écrire

$$M_n - \mu = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu),$$

et que les v.a.r. $(X_n - \mu)$ sont i.i.d. d'espérance nulle, il suffit de montrer la convergence vers 0 de (M_n) dans le cas d'une espérance nulle. On suppose donc que $\mu = 0$.

Dans une première étape, nous montrons que (M_n) converge vers 0 dans $\mathscr{L}^1$. Nous utilisons le fait que la variance de $|M_n|$ est positive, ce qui implique $[E(|M_n|)]^2 \leq E(|M_n|^2) = E(M_n^2) = \mathrm{Var}(M_n)$, puisque $E(M_n) = 0$. Ainsi,

$$E(|M_n|) \leq \sqrt{E(M_n^2)} = \sqrt{\mathrm{Var}(M_n)} = \frac{\sigma}{\sqrt{n}},$$

ce qui implique $\lim_n E(|M_n|) = 0$, c'est-à-dire la convergence de (M_n) vers 0 dans $\mathscr{L}^1$.

Dans une deuxième étape nous montrons que la sous-suite (M_{n^2}) converge p.s. vers 0. Nous savons déjà que la convergence dans $\mathscr{L}^1$ de (M_n) implique sa convergence en probabilité et donc l'existence d'une sous-suite $(M_{\varphi(n)})$ qui converge p.s., mais nous allons montrer que $\varphi(n) = n^2$ convient. Comme $E(M_n) = 0$, d'après l'inégalité de Bienaymé-Chebyshev,

$$P(|M_{n^2}| \geq \varepsilon) \leq \frac{\mathrm{Var}(M_{n^2})}{\varepsilon^2} = \frac{\sigma^2}{\varepsilon^2 n^2},$$

donc la série de terme général $P(|M_{n^2}| \geq \varepsilon)$ converge. D'après la Proposition 9.2, ceci implique la convergence p.s. de (M_{n^2}) vers 0.

Dans une troisième et dernière étape, nous montrons que chaque terme de la suite M_n est très proche du terme M_{p^2} dont l'indice est le plus proche. Pour cela nous définissons $p(n)$ la partie entière de $\sqrt{n}$, c'est-à-dire l'unique entier tel que

$$[p(n)]^2 \leq n < [p(n)+1]^2,$$

et nous définissons la v.a. $Z_n := M_n - V_n$ où

$$V_n := \frac{p(n)^2}{n} M_{p(n)^2}.$$

Comme $|V_n| \leq \left|M_{p(n)^2}\right|$, et que la suite $(M_{p(n)^2})$ converge p.s. vers 0 (il s'agit de la suite (M_{n^2}) où chaque terme est répété consécutivement $(n+1)^2 - n^2$ fois), la suite (V_n) converge p.s. vers 0. Ainsi il suffit de montrer que (Z_n) converge p.s. vers 0 pour obtenir la convergence p.s. de (M_n) vers 0. Or

$$Z_n = \frac{1}{n}\sum_{i=1}^{n} X_i - \frac{p(n)^2}{n}\,\frac{1}{p(n)^2}\sum_{i=1}^{p(n)^2} X_i = \frac{1}{n}\sum_{i=p(n)^2+1}^{n} X_i.$$

Donc $E(Z_n) = 0$ et

$$\begin{aligned}
\mathrm{Var}(Z_n) &= \left(\frac{1}{n}\right)^2 \sum_{i=p(n)^2+1}^{n} \sigma^2 = \frac{n - p(n)^2}{n^2}\sigma^2 \\
&\leq \frac{[p(n)+1]^2 - p(n)^2}{n^2}\sigma^2 = \frac{2p(n)+1}{n^2}\sigma^2 \\
&\leq \frac{2\sqrt{n}+1}{n^2}\sigma^2 = \frac{2\sigma^2}{n\sqrt{n}} + \frac{\sigma^2}{n^2}.
\end{aligned}$$

À nouveau l'inégalité de Bienaymé-Chebyshev donne

$$P(|Z_n| \geq \varepsilon) \leq \frac{\mathrm{Var}(Z_n)}{\varepsilon^2} = \frac{2\sigma^2}{\varepsilon^2 n\sqrt{n}} + \frac{\sigma^2}{\varepsilon^2 n^2},$$

donc la série de terme général $P(|Z_n| \geq \varepsilon)$ converge. D'après la Proposition 9.2, ceci implique la convergence p.s. de (Z_n) vers 0. □

Corollaire 11.2 *Soit* (Y_n) *une suite de v.a.r. i.i.d. définies sur le même espace et* $f : \mathbb{R} \to \mathbb{R}$ *une fonction bornée. Alors la convergence*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} f(Y_i) = E(f(Y_0))$$

a lieu p.s. et dans $\mathscr{L}^1$.

Dém. Il suffit d'appliquer la loi des grand nombre à la suite (X_n) où $X_n = f(Y_n)$ est bien intégrable puisque f est bornée. □

Remarque 11.3 *Mentionnons deux cas particuliers intéressants du dernier corollaire.*

- *Si* $f = \mathbb{1}_I$ *où* I *est un intervalle quelconque de* $\mathbb{R}$*, alors*

$$\lim_{n\to\infty} \frac{1}{n} \mathrm{Card}\{1 \leq i \leq n : Y_i \in I\} = P(Y_0 \in I) \qquad \textit{p.s. et dans } \mathscr{L}^1.$$

  Ce résultat est ce qu'on appelle un théorème ergodique : la moyenne du temps passé dans l'intervalle I *par la suite* (Y_n) *est asymptotiquement égale à la probabilité que* Y_0 *'tombe' dans* I.

- *Si* Y_0 *est uniforme dans l'intervalle* $[a, b]$*, alors*

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} f(Y_i) = \frac{1}{b-a} \int_a^b f(x)\,dx \qquad \textit{p.s. et dans } \mathscr{L}^1,$$

  ce qui suggère que l'on peut approcher le calcul d'une intégrale grâce à une suite de v.a. uniformes. Nous allons voir cependant dans la section suivante que cette approximation n'est pas excellente, dans la mesure où la convergence est très lente.

11.2 Théorème Central Limite

Dans cette section on cherche à savoir à quelle vitesse (M_n) converge vers μ, c'est-à-dire comment choisir une suite (a_n) qui tend vers 0 telle que $(M_n - \mu)/a_n$ converge, en un sens à déterminer, vers un réel W ni nul ni infini. Si l'on trouve une telle suite, on peut écrire de façon informelle $M_n - \mu \sim a_n W$, ce qui donne un développement asymptotique à l'ordre 2 pour le comportement de (M_n) lorsque $n \to \infty$, et une estimation de l'erreur que l'on commet en approximant μ par M_n (voir dernière remarque). Le théorème central limite affirme que dans le cas où les (X_n) sont de carré intégrable, $a_n = 1/\sqrt{n}$ convient, la convergence a lieu en loi, et W est une v.a. normale.

Théorème 11.4 (Théorème Central Limite) *On suppose que les* (X_n) *sont de carré intégrable et l'on définit* $\sigma^2 := \mathit{Var}(X_0)$*. Alors*

$$\sqrt{n}(M_n - \mu) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} (X_i - \mu)$$

converge en loi lorsque $n \to \infty$ *vers* N_σ*, où* N_σ *est une variable normale d'espérance nulle et de variance* σ^2.

Dém. Comme à la section précédente on peut supposer sans perte de généralité que $\mu = 0$. On utilise la caractérisation de la convergence en loi par la convergence simple des fonctions caractéristiques. Soit φ la fonction caractéristique de X_0 et φ_n celle de $\sqrt{n}M_n$. On cherche donc à montrer que φ_n converge simplement vers la fonction caractéristique de N_σ. Or

$$\begin{aligned}
\varphi_n(t) &= E\left(e^{it\sqrt{n}M_n}\right) \\
&= E\left(e^{i\frac{t}{\sqrt{n}}\sum_{j=1}^n X_j}\right) \\
&= E\left(\prod_{j=1}^n e^{i\frac{t}{\sqrt{n}}X_j}\right) \\
&= \prod_{j=1}^n E\left(e^{i\frac{t}{\sqrt{n}}X_j}\right) \\
&= \prod_{j=1}^n E\left(e^{i\frac{t}{\sqrt{n}}X_0}\right) \\
&= \left[\varphi\left(\frac{t}{\sqrt{n}}\right)\right]^n,
\end{aligned}$$

où l'on a utilisé l'indépendance des (X_j) (et donc celle des $e^{i\frac{t}{\sqrt{n}}X_j}$), puis le fait qu'elles ont même loi. Le comportement de $\varphi(t/\sqrt{n})^n$ lorsque $n \to \infty$ est indéterminé, puisque $\varphi(0) = 1$. Il faut donc utiliser le développement de Taylor de φ au voisinage de 0. Comme X_0 admet des moments d'ordre 2, d'après la Proposition 6.17, φ de classe C^2, avec

$$\varphi'(t) = E\left(iX\,e^{itX}\right) \quad \text{et} \quad \varphi''(t) = E\left(-X^2\,e^{itX}\right).$$

En particulier

$$\varphi'(0) = iE\left(X\right) = 0 \quad \text{et} \quad \varphi''(0) = -E\left(X^2\right) = -\mathrm{Var}(X) = -\sigma^2.$$

On a donc

$$\varphi(x) = 1 - \frac{\sigma^2 x^2}{2} + g(x)$$

où $g(x) = o(x^2)$, c'est-à-dire que $\lim_{x\to 0} g(x)/x^2 = 0$. Ainsi, pour tout t fixé, lorsque $n \to \infty$,

$$\begin{aligned}
\varphi_n(t) &= \left[1 - \frac{\sigma^2 t^2}{2n} + g\left(\frac{t}{\sqrt{n}}\right)\right]^n \\
&= \exp\left[n\ln\left(1 - \frac{\sigma^2 t^2}{2n} + g\left(\frac{t}{\sqrt{n}}\right)\right)\right] \\
&= \exp\left[n\left(-\frac{\sigma^2 t^2}{2n} + o\left(\frac{1}{n}\right)\right)\right] \\
&\longrightarrow \exp\left[-\frac{\sigma^2 t^2}{2}\right],
\end{aligned}$$

qui est bien la fonction caractéristique de N_σ. $\square$

Remarque 11.5 *Noter que* $\sqrt{n}(M_n - \mu)/\sigma$ *converge en loi vers* N_σ/σ*. Comme* N_σ/σ *a même loi que* N_1*,* $\sqrt{n}(M_n - \mu)/\sigma$ *converge en loi vers* N_1*.*

Remarque 11.6 *Le théorème central limite constitue le premier outil de base de la statistique, en particulier pour quantifier l'erreur que l'on commet en approximant* μ *par* M_n*. On peut par exemple montrer que pour tout* $a > 0$*,*

$$\lim_{n\to\infty} P\left(\mu \in \left[M_n - \frac{a\sigma}{\sqrt{n}}, M_n + \frac{a\sigma}{\sqrt{n}}\right]\right) = P(N_1 \in [-a, a]).$$

Si l'on choisit a *tel que* $P(N_1 \in [-a, a]) = 1 - \varepsilon$ *pour un niveau de confiance fixé* $1 - \varepsilon$ *(par exemple, si* $\varepsilon = 0,05$*,* a *vaut à peu près* $1,96$*), l'intervalle exhibé ci-dessus est appelé* intervalle de confiance*.*