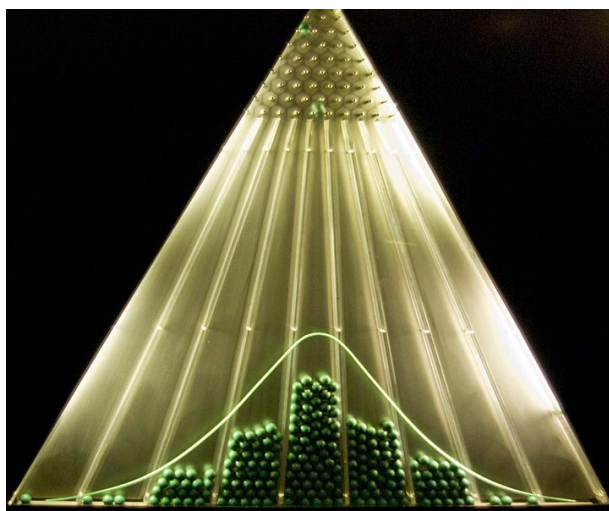


Cours de théorie des probabilités

avec exercices corrigés et devoirs



Licence de mathématiques, 3^{ième} année

Bruno Saussereau ¹

Année universitaire 2013-2014

¹Bruno Saussereau, Laboratoire de Mathématiques de Besançon, UFR Sciences & Techniques, 16, route de Gray, 25030 Besançon cedex, France. Courriel : bruno.saussereau@univ-fcomte.fr

Présentation du cours

Ce cours correspond à l'unité d'enseignement de théorie des probabilités dispensée dans le cadre du semestre 5 de l'enseignement à distance de la Licence de Mathématiques.

La diffusion de ce cours est strictement limitée aux étudiants régulièrement inscrits à l'unité d'enseignement correspondante du Centre de Télé-enseignement Universitaire.

Public visé

Cet enseignement par correspondance s'adresse en priorité aux étudiants désireux de poursuivre des études de Master en vue de la recherche, de passer le concours de l'agrégation externe de mathématiques ou à ceux qui se destinent à des études de mathématiques appliquées en vue de devenir ingénieurs-mathématiciens.

Pré-requis et révisions

Ce cours ne suppose aucun pré-requis sur le formalisme des probabilités. Tout le formalisme et le vocabulaire des probabilités est défini et introduit au fur et à mesure des besoins. Il suppose juste une sensibilisation aux phénomènes aléatoires et à leur étude élémentaire telle qu'elle est enseignée depuis quelques années au lycée et dans le semestre 4 de la Licence. Pour une rapide mise à niveau sur l'approche élémentaire des probabilités on peut se reporter aux deux ouvrages classiques [11] et [12]. Certains des exercices proposés dans cette unité sont inspirés de ces deux ouvrages moyennant quelques adaptations de vocabulaire dues au formalisme introduit dans le cours.

En revanche ce cours suppose connus les concepts classiques de la théorie de la mesure et de l'intégration, dite **intégrale de Lebesgue**. Ces concepts seront souvent rappelés dans ce cours de façon à rendre sa lecture autonome. Ces résultats seront énoncés sous leur version la plus utile pour les applications en probabilités, ils seront admis et ne feront donc pas l'objet d'une démonstration sauf cas particuliers. Pour leur version générale et leurs démonstrations, on pourra se reporter à l'ouvrage [8].

Outre ces résultats spécifiques, le cours nécessitera la connaissance de résultats et de techniques classiques de mathématiques générales. C'est donc l'occasion, dès maintenant, de réviser également ces notions mathématiques indispensables qui seront supposées connues. A cet effet, on pourra se reporter à un cours classique de mathématiques générales, par exemple [1], largement suffisant pour revoir ces notions. Il s'agit en particulier de bien connaître

1. les notions et résultats élémentaires de la théorie des ensembles : ensembles, parties d'un ensemble, inclusion, appartenance, partition d'un ensemble, intersection et réunion de plusieurs sous-ensembles, différence de deux sous-ensembles, complémentaire d'un sous-ensemble, applications, bijections, image-réciproque d'une application, opérations sur les applications, composition de deux applications,...
2. les éléments de théorie de la mesure et de l'intégrale de Lebesgue
3. le calcul des sommes de séries : série géométrique, série exponentielle, dérivation des séries entières, ...
4. quelques éléments d'algèbre générale et multilinéaire en dimension finie : binôme de Newton, nombre de combinaisons, espaces vectoriels, produit scalaire euclidien, norme euclidienne, calcul matriciel, transposé d'une matrice, opérations élémentaires sur les matrices, diagonalisation d'une matrice symétrique, ...

Conseils de travail

Le cours proprement dit comprendra des définitions, des propositions (théorèmes, lemmes, formules, ...), des démonstrations, des exemples et des exercices corrigés. Les démonstrations doivent être connues, elles sont exigibles lors des épreuves d'évaluation.

Les démonstrations développées dans le cours sont choisies en fonction de l'intérêt pédagogique du raisonnement qu'elles mettent en oeuvre. Il faut les étudier, crayon en main, essayer de les refaire en mettant en évidence les deux ou trois axes de la démonstration qu'il convient de retenir pour être capable de la restituer sans document. C'est à ce critère que vous pourrez mesurer si vous avez compris quelque chose. Il est conseillé aussi de bien mettre en évidence dans ces démonstrations, en les énonçant complètement et en vérifiant que leurs hypothèses de validité sont satisfaites, les résultats antérieurs sur lesquels elles prennent appui. Certaines démonstration seront détaillées, d'autres seront volontairement plus succinctes afin de vous entraîner à détailler par vous-même les passages rapides de la démonstration.

Les exemples du cours servent à illustrer une définition sur un cas particulier ou à montrer une application concrète d'une proposition. Leur rédaction est aussi parfois volontairement succincte. Il convient alors d'en détailler les calculs, de vérifier les résultats annoncés, et d'essayer de noter les astuces ou techniques utilisées et transposables dans d'autres situations, éventuellement moyennant certaines adaptations. Ce qui est dit pour les exemples est aussi valable pour tous les exercices proposés en auto-correction et leurs corrigés.

Les exercices sont divisés en deux catégories :

1. **Les exercices de la première catégorie** sont les exercices insérés dans le texte du cours proprement dit. Ils sont assez simples et sont conçus comme des applications directes du cours et de ce qui vient d'être vu.
2. **Les exercices de la seconde catégorie**, dits de révision, sont placés en fin de chaque chapitre à partir du chapitre III. Ils sont, quant à eux, de difficultés variables et font appel aux diverses notions mises en place dans les chapitres antérieurs y compris le chapitre étudié.

Vous devez essayer de chercher à résoudre le maximum d'exercices, en vous aidant du cours. Pour les exercices que vous ne savez pas résoudre ou que vous n'avez pas pu chercher, par exemple par manque de temps, il faut au moins étudier leurs solutions en vous reportant au chapitre VIII.

Ce qui a été dit, plus haut, pour l'étude des démonstrations s'applique également pour étudier **la correction d'un exercice**. Encore une fois, après avoir étudié une démonstration ou la solution d'un exercice, vous devez être capable de refaire cette démonstration ou cet exercice, sans regarder le cours, trois ou quatre jours plus tard. C'est là un bon test pour savoir si vous avez compris la démonstration ou la solution de l'exercice. Il ne faut pas hésiter à réviser les chapitres déjà travaillés c'est-à-dire à revenir plusieurs fois, après de longs intervalles de temps, sur les démonstrations ou exercices étudiés auparavant.

Trois devoirs à rédiger et à retourner à la correction sont proposés dans le cadre de cet enseignement afin de vous permettre de tester vos connaissances et de vous inciter à un travail régulier. Ces devoirs permettent aussi de montrer au correcteur que vous avez compris le cours, que vous connaissez les résultats vus en cours et les hypothèses qui les commandent, et que vous savez les mobiliser pour répondre à une question ou démontrer un résultat nouveau. Il est donc recommander de tout mettre en œuvre pour atteindre cet objectif.

Il est bon de porter son attention, en particulier, sur les conseils suivants :

Un devoir de mathématiques est un devoir de français qui traite de mathématiques, c'est donc avant tout un texte de français. Il doit donc être rédigée de façon correcte en français. Les hypothèses spécifiques justifiant l'utilisation de chaque théorème doivent être correctement explicitées et le résultat du cours utilisé doit être clairement identifié voire explicitement énoncé. Les résultats intermédiaires et les conclusions obtenues doivent être mis en évidence. Les notations utilisées ou introduites, surtout si elles sont nouvelles par rapport au cours, doivent être clairement annoncées. La rédaction du cours peut être considérée comme un guide de rédaction d'un texte mathématique.

Les épreuves d'examen comporteront des exercices et des questions portant sur l'ensemble du cours. Elles peuvent également comprendre des questions de cours proprement dites : énoncer un ou plusieurs résultats du cours, refaire une ou plusieurs démonstrations vues en cours, traiter un exemple ou un exercice corrigé proposés dans les documents fournis dans le cadre de cette unité d'enseignement. La table de la loi normale standard de l'annexe B (sans les explications sur son utilisation), ainsi que le formulaire de l'annexe A, seront disponibles avec les sujets lors des épreuves d'évaluation. Lors de ces épreuves, l'utilisation d'une calculatrice est autorisée.

Certaines propositions du cours concernent des résultats mentionnés "hors programme". Ils sont simplement donnés dans un but de culture mathématique, mais ne feront donc pas l'objet d'évaluation et leur connaissance n'est pas exigible dans les évaluations. Souvent ils apportent des compléments ou des précisions sur un résultat ou une remarque qui viennent d'être faits.

Enfin, il est évident que l'appréciation d'une copie par le correcteur, que ce soit celle d'un devoir ou d'une épreuve d'examen, accordera une place importante à la rédaction, à la clarté des justifications et de l'argumentation ainsi qu'à la présentation globale de la copie. Une copie illisible ou mal rédigée pourra ne pas être corrigée et sera sanctionnée en conséquence.

Annexes

Ce document comprend cinq annexes :

1. **L'annexe A**, page 205, est un rappel des principales relations mathématiques utilisées dans les calculs de probabilités et des lois de probabilités classiques à connaître. Ce formulaire sera disponible lors des épreuves de contrôles ou d'examens.
2. **L'annexe B**, page 211, explique l'usage de la table statistique de la loi normale centrée-réduite reproduite en fin de l'annexe. La table de la loi normale standard, sans les explications qui l'accompagnent, sera disponible lors des épreuves d'examen.
3. **L'annexe C**, page 215, comprend les sujets des trois devoirs qui devront être envoyés à la correction et précise les dates de ces trois envois. Les corrigés de ces devoirs seront retournés avec la copie corrigée.

Bibliographie

Pour le cours, et surtout pour apporter des compléments à ce cours, on pourra utiliser avec profit le livre de [4]. Pour les exercices on pourra se reporter à [2] pour ceux relevant de la théorie de la mesure et de l'intégration, et à [3] où on trouvera des exercices supplémentaires concernant la théorie des probabilités.

Pour une justification du choix du formalisme et de sa signification en tant que modèle de la "réalité", on pourra consulter avec profit en première lecture le chapitre I de [5] et [7] puis, en seconde lecture, [4] pages 93 et 132, et [13] page 56.

Une approche historique et épistémologique en liaison avec les questions d'enseignement des concepts probabilistes peut être trouvée dans [6].

Calendrier de travail

Le cours lui-même est divisé en sept chapitres auxquels s'ajoute un huitième chapitre regroupant les corrections des exercices proposés dans les chapitres précédents.

Les trois premiers chapitres sont principalement destinés à mettre en place le formalisme des probabilités en transcrivant dans le langage des probabilités les notions de théorie de la mesure et de l'intégration vues dans l'unité correspondante : tribu, application mesurable, mesure, image d'une mesure, règles d'intégration, théorèmes de Lebesgue, ... etc. Normalement ces notions ont été vues dans l'unité d'intégration qui est conseillée pour suivre cet enseignement de probabilité. Elles doivent être étudiées assez rapidement de façon à faire porter votre travail sur les autres chapitres. Dans ces trois premiers chapitres la notion de loi de probabilité, le théorème du transfert, la notion de fonction caractéristique et les critères d'identification des lois, doivent être bien assimilés et maîtrisés.

Les concepts vraiment nouveaux et propres à la théorie des probabilités : indépendance, vecteurs gaussiens, convergences, théorèmes-limites, ... etc, sont vues dans les quatre derniers chapitres et constituent le noyau de l'unité de probabilités.

Il faut consacrer en gros un tiers du temps de travail de l'unité à l'étude des chapitres 1, 2 et 3. Un tiers du temps aux chapitres 4 et 5, et un tiers du temps aux chapitres 6 et 7.

Vous avez à rédiger trois devoirs à envoyer pour correction à l'adresse suivante :

Bruno Sausseureau, Laboratoire de Mathématiques de Besançon, UFR des Sciences et Techniques, 16, route de Gray, 25030 Besançon cedex, France.

1. Le devoir 1, dont le texte se trouve en annexe [C](#), page [216](#), porte sur les chapitres I, II et III. Il doit être envoyé au plus tard pour le 21 février 2014.
2. Le devoir 2, dont le texte se trouve en annexe [C](#), page [218](#), porte principalement sur le chapitre IV et V mais pourra bien sûr faire appel à des résultats des chapitres précédents. Il doit être envoyé au plus tard pour le 28 mars 2014.
3. Le devoir 3, dont le texte se trouve en annexe [C](#), page [220](#), porte principalement sur les chapitres VI et VII, mais pourra bien sûr faire appel à des résultats des chapitres précédents. Il doit être envoyé au plus tard pour le 18 avril 2014.

Le calendrier ci-dessus est donné à titre indicatif. Bien entendu, j'accepterai de corriger vos devoirs à n'importe quel moment. Cependant je vous conseille d'essayer de travailler régulièrement et de suivre ce calendrier.

Remarque finale

Comme pour tout document, des erreurs ou des coquilles peuvent s'être glissées lors de sa rédaction, merci de me signaler celles que vous pourriez relever. Plus généralement, si vous avez des remarques sur le document, n'hésitez pas à m'en faire part.

Besançon, le 10 janvier 2014,

Bruno Sausseureau

Table des matières.

Présentation du cours	i
Notations	xi
1 Modèles probabilistes	1
1.1 Préliminaires	1
1.2 Tribu sur un ensemble	3
1.3 Mesures et probabilités	6
1.3.1 Mesure	6
1.3.2 Probabilités et événements	7
1.3.3 Propriétés élémentaires des probabilités	11
1.4 Fonctions de répartition	13
2 Loi d'un vecteur aléatoire	19
2.1 Remarques sur la modélisation de l'aléatoire	19
2.1.1 Cas discret	19
2.1.2 Cas continu	20
2.1.3 Principe de modélisation	20
2.2 Applications mesurables	21
2.3 Loi d'une variable aléatoire	23
2.3.1 Variables aléatoires	23
2.3.2 Loi d'une variable aléatoire	24
3 Moments d'un vecteur aléatoire	29
3.1 Rappels sur l'intégration des applications mesurables	29
3.1.1 Intégration des fonctions positives	29
3.1.2 Intégration des fonctions numériques	33
3.1.3 Intégration des fonctions vectorielles	36
3.1.4 Propriétés de l'intégrale	37
3.1.5 Espaces de Lebesgue d'ordre p	38
3.2 Théorème du transfert et moments d'une v.a.	40
3.2.1 Théorème du transfert et identification de lois	40
3.2.2 Moments d'une variable aléatoire	45
3.3 Fonction caractéristique et loi d'une v.a.	49
3.4 Exercices de révision sur les chapitres I à III	59

4	Indépendance stochastique	61
4.1	Intégration sur \mathbb{R}^{n+p}	61
4.2	Indépendance de vecteurs aléatoires, d'événements, de tribus	66
4.2.1	Indépendance de vecteurs aléatoires	66
4.2.2	Critères d'indépendance de vecteurs aléatoires	68
4.2.3	Indépendance d'événements, de tribus	77
4.3	Tribu et événements asymptotiques	80
4.4	Somme de v.a.r. indépendantes	84
4.5	Exercices de révision sur les chapitres I à IV	90
5	Vecteurs aléatoires gaussiens	95
5.1	Vecteur gaussien	95
5.2	Loi d'un vecteur gaussien	98
5.3	Exercices de révision sur les chapitres I à V	103
6	Lois des grands nombres et convergences de v.a.r.	105
6.1	Convergence en probabilité d'une suite de v.a.r.	105
6.1.1	Loi faible des grands nombres	105
6.1.2	Convergence en probabilité	109
6.2	Convergence presque-sûre d'une suite de v.a.r.	112
6.2.1	Loi forte des grands nombres	112
6.2.2	Convergence presque-sûre	113
6.3	Convergence dans $L^p(\Omega, \mathcal{F}, \mathbb{P})$ où $p \in [1, +\infty]$	118
6.4	Comparaison des convergences dans $L^0(\Omega, \mathcal{F}, \mathbb{P})$	119
6.5	Exercices de révision sur les chapitres I à VI	121
7	Théorème-limite central et convergence de lois	123
7.1	Théorème-limite central (TLC)	123
7.1.1	Énoncé du théorème-limite central (TLC)	123
7.1.2	Cas particuliers du théorème-limite central (TLC)	126
7.1.3	Correction de continuité	128
7.2	Convergence d'une suite de probabilités, convergence en loi	129
7.3	Exercices de révision sur les chapitres I à VII	141
8	Corrigés des exercices	145
8.1	Corrigés des exercices du chapitre I	145
8.2	Corrigés des exercices du chapitre II	152
8.3	Corrigés des exercices du chapitre III	155
8.4	Corrigés des exercices du chapitre IV	165
8.5	Corrigés des exercices du chapitre V	183
8.6	Corrigés des exercices du chapitre VI	190
8.7	Corrigés des exercices du chapitre VII	196
A	Formulaire	205
A.1	Rappels de notations	205
A.2	Quelques relations à connaître en probabilités	205
A.3	Probabilités usuelles discrètes	207
A.4	Probabilités usuelles à densité	208

B	Table de la loi normale standard	211
B.1	Calculs avec des v.a.r. normales centrées-réduites	211
B.2	Calculs avec des v.a.r. normales de paramètres quelconques	212
C	Devoirs à envoyer à la correction	215
C.1	Devoir 1 à renvoyer le 21 février 2014 au plus tard	216
C.2	Devoir 2 à renvoyer le 28 mars 2014 au plus tard	218
C.3	Devoir 3 à renvoyer le 18 avril 2014 au plus tard	220
	Bibliographie.	221

Notations

Nous répertorions ici quelques notations générales qui seront utilisées dans l'ensemble du cours.

On note de façon classique respectivement par les lettres, \mathbb{N} , \mathbb{Z} , \mathbb{Q} , \mathbb{R} , \mathbb{C} , les ensembles des nombres entiers naturels, relatifs, rationnels, réels, complexes.

Les lettres \mathbb{P} et \mathbb{E} seront introduites dans le cours mais ne devront pas être confondues avec les notations d'ensembles de nombres.

On pose $\overline{\mathbb{R}} := \mathbb{R} \cup \{+\infty, -\infty\}$. On étend l'ordre usuel de \mathbb{R} à $\overline{\mathbb{R}}$ en posant, pour tout $x \in \mathbb{R}$, $-\infty < x < +\infty$. On prolonge les opérations classiques sur \mathbb{R} de la façon suivante : pour tout $x \in \mathbb{R} \cup \{+\infty\}$, $x + (+\infty) = +\infty$, $x - (-\infty) = +\infty$; pour tout $x \in \mathbb{R} \cup \{-\infty\}$, $x + (-\infty) = -\infty$, $x - (+\infty) = -\infty$. On remarquera que $(+\infty) + (-\infty)$ et $(+\infty) - (+\infty)$ ne sont pas définis.

On suppose connues les notations classiques de la théorie élémentaire des ensembles : intersection \cap , réunion \cup , différence de deux ensembles \setminus , ensemble vide \emptyset , passage au complémentaire \complement ou plus fréquemment c , inclusion (au sens large) \subseteq .

Le symbole de Halmos, \square , désignera la fin d'une démonstration.

Le symbole $:=$ signifie "est égal par définition". Il indique que le membre de gauche de $:=$ est une notation pour le membre de droite.

Chaque proposition, exemple, exercice, est numérotée par deux nombres séparés par un point. Par exemple "proposition 5.12" désigne la proposition 12 du chapitre 5.

Chapitre 1

Modèles probabilistes

Le formalisme de la théorie des probabilités utilise les outils de la théorie de la mesure en adoptant un vocabulaire spécifique aux probabilités.

1.1 Préliminaires

Certaines définitions et notations de la théorie élémentaire des ensembles seront constamment utilisées dans la suite. Afin d'éviter toute ambiguïté nous les rappelons rapidement dans ce paragraphe.

Définition 1.1.

Dans ce cours un ensemble sera dit **dénombrable** s'il est en bijection avec une partie (finie ou infinie) de \mathbb{N} .

(Attention : dans certains ouvrages, un tel ensemble est dit **au-plus-dénombrable**, le qualificatif dénombrable désignant alors les ensembles possédant un nombre fini d'éléments.)

Si A et B sont deux parties d'un ensemble E , on note $A^c := \{x \in E / x \notin A\}$, ou aussi $\complement_E A$ si on souhaite préciser l'ensemble de référence E , le complémentaire de A dans E et

$$A \setminus B := A \cap B^c = \{x \in A / x \notin B\}.$$

Définition 1.2.

Soit f une application d'un ensemble E dans un ensemble F . Si A est une partie de F , l'**image-réciproque** de A par f est l'ensemble, noté par les probabilistes $\{f \in A\}$, défini par $\{f \in A\} := \{x \in E / f(x) \in A\}$.

L'ensemble $\{f \in A\}$ est donc une partie de E .

Exemples 1.1.

Si f et g sont deux applications de E dans \mathbb{R} et a un réel,

$$\{f = g\} := \{x \in E / f(x) = g(x)\}, \quad \{f \leq g\} := \{x \in E / f(x) \leq g(x)\}, \\ \{f = a\} := \{x \in E / f(x) = a\}.$$

En vue de la proposition suivante, rappelons que si $(A_i)_I$ est une famille quelconque de parties d'un ensemble F , $\bigcup_{i \in I} A_i$ désigne la partie de F constituée des éléments x de F tels qu'il existe au moins un indice $k \in I$, $x \in A_k$. De même, $\text{bigcap}_{i \in I} A_i$ désigne la partie de F constituée des éléments x de F tels que, pour tout indice $k \in I$, $x \in A_k$.

Voici quelques propriétés classiques de l'image-réciproque :

Proposition 1.1.

Avec les notations introduites ci-dessus,

1. $\{f \in \emptyset\} := \{x \in E \mid f(x) \in \emptyset\} = \emptyset$.
2. Si A et B sont des parties de F avec $A \subseteq B$ alors, $\{f \in A\} \subseteq \{f \in B\}$.
3. Si $(A_i)_I$ est une famille quelconque de parties de F ,

$$\left\{f \in \bigcup_{i \in I} A_i\right\} = \bigcup_{i \in I} \{f \in A_i\} \text{ et } \left\{f \in \bigcap_{i \in I} A_i\right\} = \bigcap_{i \in I} \{f \in A_i\}.$$

4. $\{f \in A\}^c = E \setminus \{f \in A\} = \{f \in F \setminus A\} = \{f \in A^c\}$.

On fera attention à l'ambiguïté de la notation c pour le complémentaire d'un ensemble dans l'assertion 4 de cette proposition : $\{f \in A\}^c$ signifie $\complement_E \{f \in A\}$ et $\{f \in A^c\}$ signifie $\{f \in \complement_F A\}$.

Exercice 1.1. (Corrigé de l'exercice : page 145)

Démontrer la proposition précédente.

Définition 1.3.

L'indicatrice d'une partie A de E est l'application, notée $\mathbb{1}_A$, de E dans \mathbb{R} définie, pour tout $x \in E$, par $\mathbb{1}_A(x) := 0$ si $x \notin A$ et $\mathbb{1}_A(x) := 1$ si $x \in A$.

Exercice 1.2. (Corrigé de l'exercice : page 145)

Soient A , B et C trois parties d'un ensemble Ω .

1. Écrire $\mathbb{1}_{A \cap B}$ et $\mathbb{1}_{A \cup B}$ en fonction de $\mathbb{1}_A$ et $\mathbb{1}_B$ lorsque :
 - (a) A et B sont disjoints (i.e. $A \cap B = \emptyset$).
 - (b) A et B sont quelconques.
2. Exprimer, en fonction des indicatrices de A , B et C , les indicatrices des ensembles suivants : A^c , $A \setminus B$, $A \cup B \cup C$.

Exercice 1.3. (Corrigé de l'exercice : page 146)

Représenter graphiquement les fonctions définies sur \mathbb{R} :

1. $\sum_{n \geq 0} \mathbb{1}_{[n, +\infty[}$.
2. $\sum_{n \geq 0} \mathbb{1}_{[0, n]}$.

$$3. \sum_{n \geq 0} (n+1) \mathbb{1}_{[n, n+1[}.$$

Enfin, rappelons que, si f et g sont deux applications d'un ensemble E dans $\overline{\mathbb{R}}$, la notation $f \leq g$ signifie que, pour tout $x \in E$, $f(x) \leq g(x)$.

1.2 Tribu sur un ensemble

Définition 1.4.

Une famille \mathcal{A} de parties d'un ensemble E est appelée une **tribu sur E** , (ou dans certains ouvrages une **σ -algèbre sur E**), si elle vérifie les trois axiomes suivants :

1. $E \in \mathcal{A}$,
2. Si $A \in \mathcal{A}$, alors $A^c \in \mathcal{A}$,
3. Si $(A_n)_{n \in \mathbb{N}}$ est une suite d'éléments de \mathcal{A} , alors $\bigcup_{n \in \mathbb{N}} A_n \in \mathcal{A}$.

Définition 1.5.

Le couple (E, \mathcal{A}) s'appelle un **espace mesurable** et les éléments de \mathcal{A} sont appelés les **parties mesurables de E relativement à la tribu \mathcal{A}** ou **parties \mathcal{A} -mesurables de E** .

On notera bien que \mathcal{A} est un ensemble constitué de parties de E et donc une partie de $\mathcal{P}(E)$, l'ensemble de toutes les parties de E .

Exemples 1.2.

les familles de parties de E , $\{\emptyset, E\}$ et $\mathcal{P}(E)$, sont des tribus sur E appelées **tribus triviales** de E . On peut donc définir au moins une tribu sur tout ensemble E .

Exercice 1.4. (Corrigé de l'exercice : page 147)

Soient n un entier strictement positif et (A_1, A_2, \dots, A_n) une partition de E , i.e. une suite de parties non vides de E , deux à deux disjointes, dont la réunion est égale à E . Soit \mathcal{A} la famille des réunions qu'on peut fabriquer à partir de toutes les sous-familles de la suite (A_1, A_2, \dots, A_n) , c'est-à-dire la famille des parties de E de la forme $\bigcup_{i \in K} A_i$ où K parcourt

l'ensemble des parties de $\{1, 2, \dots, n\}$. Montrer que la famille \mathcal{A} est une tribu sur E .

Pour une généralisation de ce résultat, on pourra consulter [2] exercice I-7 question 2.

Exercice 1.5. (Corrigé de l'exercice : page 147)

Montrer que l'intersection d'une famille quelconque de tribus est une tribu. En est-il de même pour la réunion ?

La proposition suivante donne un procédé de construction de parties mesurables à partir d'autres éléments de la tribu :

Proposition 1.2.

Soit \mathcal{A} une tribu sur E .

1. $\emptyset \in \mathcal{A}$.
2. Si $(A_i)_{i \in I}$, où $I \subseteq \mathbb{N}$, est une suite (finie ou infinie) d'éléments de \mathcal{A} , alors

$$\bigcap_{i \in I} A_i \in \mathcal{A} \text{ et } \bigcup_{i \in I} A_i \in \mathcal{A}.$$

Démonstration : $\emptyset = E^c$, on conclut par les axiomes 1 et 2 de la définition des tribus.

On pose $A'_i := A_i$ pour tout entier $i \in I$ et $A'_i := \emptyset$ pour tout entier $i \in \mathbb{N} \setminus I$. On applique le résultat précédent et l'axiome 3 de la définition pour montrer que $\bigcup_{i \in I} A_i = \bigcup_{n \in \mathbb{N}} A'_n \in \mathcal{A}$.

Pour montrer $\bigcap_{i \in I} A_i \in \mathcal{A}$, on remarque que

$$\bigcap_{i \in I} A_i = \left(\bigcup_{i \in I} A_i^c \right)^c$$

On utilise alors le résultat précédent et l'axiome 2. \square

Exercice 1.6. (Corrigé de l'exercice : page 148)

Montrer que si A et B sont deux parties mesurables de E relativement à la tribu \mathcal{A} , alors $A \setminus B \in \mathcal{A}$.

Pour des raisons techniques qui seront précisées plus loin, lorsqu'on travaille sur $E := \mathbb{R}$ ou plus généralement $E := \mathbb{R}^d$ avec $d \in \mathbb{N}^*$, il n'est pas possible d'utiliser la tribu $\mathcal{P}(\mathbb{R})$ ou $\mathcal{P}(\mathbb{R}^d)$ car elle est trop "grosse". Pour des explications plus détaillées consulter l'annexe ??, page ?? . On doit donc définir une tribu plus "petite" (au sens de l'inclusion des ensembles) mais suffisamment riche en éléments pour contenir au moins les ensembles utilisés dans les applications pratiques de la théorie des probabilités, comme les intervalles de \mathbb{R} ou les pavés de \mathbb{R}^d , et leurs réunions ou intersections dénombrables.

Pour cela on définit la **tribu borélienne** ou **tribu de Borel de \mathbb{R}** notée $\mathcal{B}(\mathbb{R})$. C'est la plus petite des tribus sur \mathbb{R} contenant tous les intervalles de la forme $]a, b]$, où a et b sont des réels tels que $a < b$. Cette dernière phrase signifie que si \mathcal{A} est une tribu sur \mathbb{R} contenant tous les intervalles de la forme $]a, b]$, où a et b sont des réels tels que $a < b$, alors tout élément de la tribu $\mathcal{B}(\mathbb{R})$ est un élément de la tribu \mathcal{A} .

Plus généralement,

Définition 1.6.

La **tribu borélienne** ou **tribu de Borel de \mathbb{R}^d** , notée $\mathcal{B}(\mathbb{R}^d)$, est la plus petite des tribus sur \mathbb{R}^d contenant tous les **pavés de \mathbb{R}^d** i.e. les parties de la forme $]a_1, b_1] \times]a_2, b_2] \times \cdots \times]a_d, b_d]$ où, pour tout entier $1 \leq i \leq d$, a_i et b_i sont des réels tels que $a_i < b_i$.

On peut de même définir la tribu borélienne sur $\mathcal{B}(\overline{\mathbb{R}})$:

Définition 1.7.

La tribu borélienne sur $\overline{\mathbb{R}}$, notée $\mathcal{B}(\overline{\mathbb{R}})$, est la plus petite des tribus sur $\overline{\mathbb{R}}$ contenant tous les intervalles de la forme $]a, b]$, où a et b sont des réels tels que $a < b$, et les intervalles $]a, +\infty]$ où $a \in \mathbb{R}$.

Définition 1.8.

Les éléments des tribus $\mathcal{B}(\overline{\mathbb{R}})$, resp. $\mathcal{B}(\mathbb{R}^d)$, sont appelés **boréliens** de $\overline{\mathbb{R}}$, resp. \mathbb{R}^d .

Exercice 1.7. (Corrigé de l'exercice : page 148)

Prouver l'existence de la tribu de Borel de \mathbb{R} . Pour cela, montrer que $\mathcal{B}(\mathbb{R})$ est l'intersection de la famille (non vide car la tribu $\mathcal{P}(\mathbb{R})$ en fait partie) des tribus contenant tous les intervalles de la forme $]a, b]$ où a et b sont des réels tels que $a < b$.

Plus généralement :

Définition 1.9.

Soit \mathcal{C} une famille de partie d'un ensemble E . On appelle **tribu engendrée par \mathcal{C} sur E** , et on note $\sigma(\mathcal{C})$, la plus petite tribu (au sens de l'inclusion) définie sur E contenant la famille \mathcal{C} .

On vérifiera aisément que la tribu $\sigma(\mathcal{C})$ est l'intersection de toutes les tribus sur E qui contiennent \mathcal{C} .

Exemples 1.3.

On montre en théorie de la mesure que la tribu borélienne de \mathbb{R}^d est engendrée par la famille constituée des parties ouvertes de \mathbb{R}^d .

Exercice 1.8. (Corrigé de l'exercice : page 148)

Soient n un entier strictement positif et (A_1, A_2, \dots, A_n) une partition de E . Montrer que la tribu construite dans l'exercice 1.4 est la tribu sur E engendrée par la famille (A_1, A_2, \dots, A_n) .

Dans la suite du cours les ensembles $\overline{\mathbb{R}}$ et \mathbb{R}^d seront toujours supposés munis de leurs tribus boréliennes.

La proposition suivante donne des exemples de boréliens de \mathbb{R} . Pratiquement ceux-ci correspondent à la plupart des ensembles qui seront manipulés dans la suite :

Proposition 1.3.

1. Tout singleton de \mathbb{R} est un borélien.
2. Toute partie dénombrable de \mathbb{R} est un borélien.
3. Tous les intervalles de \mathbb{R} , quelle que soit leur forme, sont des boréliens de \mathbb{R} .
4. Toutes les réunions dénombrables ou intersections dénombrables d'intervalles de \mathbb{R} , ou plus généralement de boréliens, sont des boréliens.

Démonstration : Pour le singleton, on remarque que si $a \in \mathbb{R}$, on peut écrire

$$\{a\} = \bigcap_{k=1}^{+\infty} \left] a - \frac{1}{k}, a \right].$$

On conclut alors avec la proposition 1.2 de la page 4. Pour l'assertion 2, on utilise l'axiome 3 de la définition des tribus. Pour démontrer 3, on montre que tout intervalle peut être écrit comme réunion ou intersection dénombrable d'intervalles de la forme $]a, b]$ ou de singletons. Par exemple $[a, b] =]a, b] \cup \{a\}$ ce qui prouve que tout intervalle fermé borné est borélien. Autres exemples : $]a, b[=]a, b] \setminus \{b\}$ ou encore

$$]a, b[= \bigcup_{k=1}^{+\infty}]a, b - \frac{1}{k}] , \quad]-\infty, b] = \bigcup_{k > -b}^{+\infty}]-k, b]. \square$$

On notera que si toute réunion dénombrable ou intersection dénombrable d'intervalles de \mathbb{R} est un borélien, cela ne signifie pas pour autant que tous les boréliens de \mathbb{R} sont de cette forme.

De plus on montre que $\mathcal{B}(\mathbb{R}^d)$ est strictement incluse dans l'ensemble des parties de \mathbb{R}^d . Il existe donc des parties de \mathbb{R}^d qui ne sont pas mesurables pour la tribu de Borel. Mais dans la pratique, tous les ensembles que nous serons amenés à utiliser dans \mathbb{R}^d seront en fait des boréliens.

1.3 Mesures et probabilités

1.3.1 Mesure

Définition 1.10.

Soit (E, \mathcal{A}) un espace mesurable. Une **mesure** sur (E, \mathcal{A}) est une application μ de \mathcal{A} dans $[0, +\infty]$ vérifiant les axiomes :

1. $\mu(\emptyset) = 0$,
2. **σ -additivité** : pour toute suite $(A_n)_{n \in \mathbb{N}}$ d'éléments de \mathcal{A} deux à deux disjoints

$$\mu\left(\bigcup_{k=0}^{+\infty} A_k\right) = \sum_{k=0}^{+\infty} \mu(A_k).$$

Le triplet (E, \mathcal{A}, μ) s'appelle un **espace mesuré**.

La σ -additivité entraîne la **simple-additivité** i.e. pour toute suite finie A_1, \dots, A_n d'éléments de \mathcal{A} deux à deux disjoints $\mu\left(\bigcup_{k=0}^{+\infty} A_k\right) = \sum_{k=0}^{+\infty} \mu(A_k)$. Mais la réciproque est fautive, c'est-à-dire qu'il ne suffit pas que le deuxième axiome de la définition précédente soit vrai pour les suites finies deux à deux disjointes pour qu'il le soit pour les suites infinies deux à deux disjointes. Un contre-exemple est proposé dans l'exercice suivant.

Exercice 1.9. (Corrigé de l'exercice : page 148)

On considère l'application μ de $\mathcal{P}(\mathbb{N})$ dans $[0, +\infty]$ définie, pour tout $A \subseteq \mathbb{N}$, par $\mu(A) := \sum_{n \in A} \frac{1}{n^2}$ (avec la convention $\frac{1}{0} = +\infty$) si A est fini, $\mu(A) = +\infty$ si A est infini, et $\mu(\emptyset) = 0$. Montrer que

1. μ est simplement-additive sur \mathbb{N} , i.e. pour toute suite finie A_1, \dots, A_n de parties de \mathbb{N} , deux à deux disjointes, $\mu \left(\bigcup_{k=1}^n A_k \right) = \sum_{k=1}^n \mu(A_k)$.
2. μ n'est pas σ -additive sur \mathbb{N} .

On admettra qu'il existe une unique mesure sur $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$, notée $\lambda^{(d)}$ et appelée **mesure de Lebesgue sur \mathbb{R}^d** , telle que, pour tout pavé de la forme $]a_1, b_1] \times]a_2, b_2] \times \dots \times]a_d, b_d]$ où pour tout entier $1 \leq i \leq d$, les réels a_i et b_i vérifient $a_i < b_i$,

$$\lambda^{(d)} (]a_1, b_1] \times]a_2, b_2] \times \dots \times]a_d, b_d]) = (b_1 - a_1)(b_2 - a_2) \dots (b_d - a_d).$$

La mesure de Lebesgue étend donc les notions de mesure de longueur (cas $d = 1$), mesure d'aire (cas $d = 2$), mesure de volume (cas $d = 3$) à toutes les parties de \mathbb{R}^d qui sont des boréliens. Dans le cas $d = 1$ on notera, pour simplifier, $\lambda := \lambda^{(1)}$.

On montre que $\lambda^{(d)}(\mathbb{R}^d) = +\infty$. On dit que la mesure de Lebesgue est une mesure **non finie** contrairement aux probabilités que nous allons définir ci-dessous et qui sont des cas particuliers de mesures **finies**.

Exercice 1.10. (Corrigé de l'exercice : page 149)

Vérifier que $\mathbb{R} = \bigcup_{k \in \mathbb{Z}}]k, k+1]$ et en déduire que $\lambda(\mathbb{R}) = +\infty$ en appliquant l'axiome de σ -additivité de la mesure de Lebesgue.

1.3.2 Probabilités et événements

Définition 1.11.

Une **probabilité sur (E, \mathcal{A})** est une mesure μ sur (E, \mathcal{A}) telle que $\mu(E) = 1$. Le triplet (E, \mathcal{A}, μ) s'appelle alors un **espace de probabilité**, les parties mesurables s'appellent les **événements relatifs à μ** . E est l'événement **certain**, \emptyset est l'événement **impossible**. Deux événements disjoints sont dits **incompatibles**.

Dorénavant, sauf indication contraire, (E, \mathcal{A}, μ) désignera un espace de probabilité.

Définition 1.12.

Une partie A de E est dite **négligeable pour μ** , s'il existe un événement B tel que $A \subseteq B$ avec $\mu(B) = 0$. Une propriété $\mathcal{P}(x)$, dépendant de l'élément $x \in E$, est dite **μ -presque-sûre** (en abrégé **μ -p.s.**) si l'ensemble des $x \in E$ pour lesquels la propriété $\mathcal{P}(x)$ n'est pas vérifiée est négligeable pour μ .

Définition 1.13.

Deux événements A et B sont dits **μ -presque-sûrement égaux** si l'événement $(A \setminus B) \cup (B \setminus A)$ est négligeable pour μ .

Un événement négligeable pour μ est **μ -presque-sûrement vide**, c'est-à-dire **μ -presque-sûrement impossible**.

Exemples 1.4.

1) Donnons un premier exemple de probabilité sur $E := \mathbb{R}^d$. Comme convenu on sous-entend $\mathcal{A} := \mathcal{B}(\mathbb{R}^d)$. Soit $a \in \mathbb{R}^d$ fixé, on note δ_a l'application de $\mathcal{B}(\mathbb{R}^d)$ dans $\{0, 1\}$ définie, pour tout borélien A , par $\delta_a(A) = \mathbb{1}_A(a)$ c-à-d $\delta_a(A) = 1$ si $a \in A$ et $\delta_a(A) = 0$ sinon.

δ_a est une probabilité sur $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ appelée **probabilité de Dirac au point a** sur \mathbb{R}^d . On vérifie aisément que toute partie de \mathbb{R}^d ne contenant pas a est négligeable pour δ_a . Le singleton $\{a\}$ est un événement δ_a -presque-sûrement égal à l'événement certain \mathbb{R}^d .

2) D'après le résultat de l'exercice 1.10, la mesure de Lebesgue n'est pas une probabilité sur \mathbb{R}^d .

Exercice 1.11. (Corrigé de l'exercice : page 149)

Vérifier que δ_a , où a est un réel, est bien une probabilité sur \mathbb{R} .

Donnons sous forme de proposition un exemple générateur de mesures et en particulier de probabilités :

Proposition 1.4.

Soient $(\mu_k)_{k \in \mathbb{N}}$ une suite de mesures sur (E, \mathcal{A}) et $(\alpha_k)_{k \in \mathbb{N}}$ une suite de réels positifs. Alors l'application

$$\mu : A \in \mathcal{A} \mapsto \mu(A) := \sum_{k=0}^{+\infty} \alpha_k \mu_k(A)$$

est une mesure sur (E, \mathcal{A}) notée

$$\mu := \sum_{k=0}^{+\infty} \alpha_k \mu_k.$$

Démonstration : On vérifie aisément que $\mu(\emptyset) = 0$. La σ -additivité de μ découle immédiatement du lemme suivant sur l'interversion des indices, souvent utile dans les calculs : Si $(a_{i,j})_{(i,j) \in \mathbb{N}^2}$ est une suite-double de réels positifs, alors

$$\sum_{i=0}^{+\infty} \sum_{j=0}^{+\infty} a_{i,j} = \sum_{j=0}^{+\infty} \sum_{i=0}^{+\infty} a_{i,j}.$$

Cette somme peut être éventuellement infinie. Pour une démonstration du lemme se reporter à [1] tome 2, p. 306. \square

On notera que si les mesures μ_k sont des probabilités sur (E, \mathcal{A}) et si $\sum_{k=0}^{+\infty} \alpha_k = 1$, alors la

mesure $\sum_{k=0}^{+\infty} \alpha_k \mu_k$ est une probabilité sur (E, \mathcal{A}) .

Exemples 1.5.

Appliqué au cas particulier où les probabilités μ_k sont les probabilités sur \mathbb{R} de Dirac au point $k \in \mathbb{N}$, le procédé précédent permet de construire d'autres exemples classiques de probabilités. Si $n \in \mathbb{N}^*$, $\alpha \in]0, +\infty[$, $p \in]0, 1[$ et $q := 1 - p$, on définit :

1. la probabilité **binomiale de paramètres n et p** :

$$\mathcal{B}(n, p) := \sum_{k=0}^n C_n^k p^k q^{n-k} \delta_k.$$

2. la probabilité **de Poisson de paramètre α** :

$$\mathcal{P}(\alpha) := \sum_{k=0}^{\infty} e^{-\alpha} \frac{\alpha^k}{k!} \delta_k.$$

3. la probabilité **géométrique de paramètre p** :

$$\mathcal{G}(p) := \sum_{k=1}^{+\infty} p q^{k-1} \delta_k.$$

4. la probabilité **uniforme-discrète de paramètre n** ou **équiprobabilité sur $\{1, 2, \dots, n\}$** :

$$\mathcal{U}(n) := \frac{1}{n} \sum_{k=1}^n \delta_k.$$

La probabilité $\mathcal{B}(1, p)$ est appelée probabilité **de Bernoulli de paramètre p** et se note simplement $\mathcal{B}(p)$.

Exercice 1.12. (Corrigé de l'exercice : page 149)

Vérifier que les probabilités introduites dans l'exemple précédent sont bien des probabilités construites suivant le procédé de la proposition 1.4.

Exercice 1.13. (Corrigé de l'exercice : page 149)

1) Expliciter les expressions analytiques, pour tout $i \in \mathbb{N}$, de $\mathcal{B}(n, p)(\{i\})$ et $\mathcal{P}(\alpha)(\{i\})$.

2) Expliciter et calculer $\mathcal{P}(\frac{1}{10})(\{1, 3, 5, 7\})$ et $\mathcal{B}(7, \frac{3}{10})(\{0, 3, 5\})$.

Définition 1.14.

Une probabilité μ sur \mathbb{R}^d est dite **discrète et portée par l'ensemble F** si elle peut s'écrire sous la forme $\mu = \sum_{n \in \mathbb{N}} p_n \delta_{a_n}$ où $(p_n)_{n \in \mathbb{N}}$ est une suite de réels positifs ou nuls, $(a_n)_{n \in \mathbb{N}}$ est une suite de vecteurs de \mathbb{R}^d et F désigne l'ensemble des $a_n \in \mathbb{R}^d$ pour lesquels $p_n > 0$.

Exemples 1.6.

Les probabilités binomiale $\mathcal{B}(n, p)$, de Poisson $\mathcal{P}(\alpha)$, de Dirac δ_a sont discrètes et portées respectivement par les ensembles $\{0, 1, \dots, n\}$, \mathbb{N} , $\{a\}$.

Il ne faut pas croire que toutes les probabilités soient discrètes. Par exemple on admettra qu'il existe une unique probabilité sur \mathbb{R} , notée $\mathcal{N}_1(0, 1)$ et appelée **probabilité de Gauss-Laplace standard**, ou **probabilité normale standard**, telle que pour tout $x \in \mathbb{R}$,

$$\mathcal{N}_1(0, 1)(]-\infty, x]) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{1}{2}t^2} dt.$$

On verra un peu plus loin que cette probabilité ne peut pas s'écrire sous la forme d'une combinaison linéaire de probabilités de Dirac et n'est donc pas discrète.

Remarquons que le nombre réel $\mathcal{N}_1(0, 1)(]-\infty, x])$ représente la mesure de l'aire délimitée par l'axe des abscisses t , la courbe d'équation $y = \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}t^2}$, et la droite d'équation $t = x$. On dira pour simplifier qu'il s'agit de la mesure de l'aire sous la courbe d'équation $y = \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}t^2}$, comprise entre $-\infty$ et x .

On peut généraliser un peu la construction précédente.

Définition 1.15.

Nous appellerons **densité de probabilité sur \mathbb{R}** toute application ρ positive de \mathbb{R} dans $[0, +\infty]$, continue sur \mathbb{R} , sauf éventuellement en un nombre fini de points où la courbe présente des sauts finis, telle que $\int_{-\infty}^{+\infty} \rho(t)dt = 1$.

On montre alors qu'il existe une unique probabilité μ sur \mathbb{R} telle que, pour tout $x \in \mathbb{R}$,

$$\mu(]-\infty, x]) = \int_{-\infty}^x \rho(t)dt.$$

On dit que μ est une probabilité à **densité sur \mathbb{R}** . On écrit $\mu = \rho \cdot \lambda$ pour exprimer que μ admet ρ pour densité. Nous généraliserons de façon définitive la définition de densité d'une probabilité sur \mathbb{R}^d au chapitre III par la définition 3.2, page 35.

Comme précédemment, le réel $\mu(]-\infty, x])$ représente la mesure de l'aire sous la courbe d'équation $y = \rho(t)$, comprise entre $-\infty$ et x .

On peut de façon plus générale définir des **mesures à densité**, qui ne sont plus nécessairement des probabilités, en remplaçant dans la définition de la densité ci-dessus, la condition $\int_{-\infty}^{+\infty} \rho(t)dt = 1$ par la condition plus faible $\int_{-\infty}^{+\infty} \rho(t)dt < +\infty$.

L'existence des mesures à densité résulte d'un théorème de prolongement assez technique que nous n'énoncerons pas. Nous nous contenterons d'admettre l'existence de telles mesures.

Exemples 1.7.

1. La probabilité de Gauss-Laplace standard vue plus haut admet la densité ρ définie sur \mathbb{R} par $\rho(t) := \frac{1}{\sqrt{2\pi}}e^{-\frac{1}{2}t^2}$.
2. L'application $\rho := \frac{1}{b-a}\mathbb{1}_{[a,b]}$, avec $a < b$, est la densité d'une probabilité sur \mathbb{R} appelée **probabilité uniforme-continue sur $[a, b]$** et notée $\mathcal{U}([a, b])$.
3. L'application ρ , définie sur \mathbb{R} par $\rho(t) := \alpha e^{-\alpha t}\mathbb{1}_{]0, +\infty[}(t)$, est la densité d'une probabilité sur \mathbb{R} appelée **probabilité exponentielle de paramètre $\alpha > 0$** et notée $\mathcal{E}(\alpha)$.

On pourrait se demander pourquoi on ne définit pas les mesures comme des applications μ σ -additives de l'ensemble des parties de E dans $[0, +\infty]$ avec $\mu(\emptyset) = 0$. Cela reviendrait à prendre toujours $\mathcal{A} := \mathcal{P}(E)$ et éviterait le recours à la notion de tribu. En fait, on montre que certaines probabilités, comme celle de Gauss définie plus haut, ne peuvent pas être définies pour toutes les parties de \mathbb{R} . Plus précisément, on montre que, toujours dans le cas de $E := \mathbb{R}$, les seules probabilités qui satisferaient à cette nouvelle définition seraient les probabilités discrètes. Malheureusement cette famille n'est pas assez riche pour permettre de modéliser grand nombre des situations aléatoires qui se présentent dans les applications concrètes de la théorie. Pour plus de développements se reporter à l'annexe ??, page ??, de ce cours.

1.3.3 Propriétés élémentaires des probabilités

(E, \mathcal{A}, μ) désigne un espace de probabilité.

Proposition 1.5.

1. Pour tout $A, B \in \mathcal{A}$ tel que $A \subseteq B$, $\mu(A) \leq \mu(B)$. En particulier pour tout $A \in \mathcal{A}$, $\mu(A) \leq 1$.
2. Pour tout $A, B \in \mathcal{A}$, $\mu(B \setminus A) = \mu(B) - \mu(A \cap B)$. En particulier si $A \subseteq B$, $\mu(B \setminus A) = \mu(B) - \mu(A)$.
3. Pour tout $A, B \in \mathcal{A}$, $\mu(A \cup B) = \mu(A) + \mu(B) - \mu(A \cap B)$.
4. Pour tout $A \in \mathcal{A}$, $\mu(A^c) = 1 - \mu(A)$.

Démonstration : 1) On remarque que $B = (B \setminus A) \cup A$ car $A \subseteq B$. De plus $(B \setminus A) \cap A = \emptyset$. D'où $\mu(B) = \mu(A) + \mu(B \setminus A) \geq \mu(A)$ l'égalité résultant de ce que l'union est disjointe. On conclut par l'axiome 2 des mesures. Pour la deuxième partie prendre $B = E$.

2) résulte de l'égalité ensembliste $(B \setminus A) \cup (A \cap B) = B$ avec $(B \setminus A) \cap (A \cap B) = \emptyset$. Pour la deuxième partie remarquer que si $A \subseteq B$, $A \cap B = A$.

3) résulte de $A \cup B = (A \setminus B) \cup B$ où l'union est disjointe.

4) résulte de $A^c = E \setminus A$ avec $A \subseteq E$. \square

Pour démontrer l'inégalité de Bonferroni nous aurons besoin du résultat ensembliste suivant laissé en exercice :

Proposition 1.6.

Soit $(A_n)_{n \in \mathbb{N}}$ une suite de parties d'un ensemble E . Posons $B_0 := A_0$ et, pour tout entier $k \geq 1$, $B_k := A_k \setminus (A_0 \cup A_1 \cup \dots \cup A_{k-1})$. Alors, pour tout entier $n \geq 0$, $B_n \subseteq A_n$ et la suite $(B_n)_{n \in \mathbb{N}}$ est formée de parties deux à deux disjointes vérifiant, pour tout entier naturel n , $\bigcup_{k=0}^n B_k = \bigcup_{k=0}^n A_k$,

$$\text{et } \bigcup_{k=0}^{+\infty} B_k = \bigcup_{k=0}^{+\infty} A_k.$$

Exercice 1.14. (Corrigé de l'exercice : page 150)

Démontrer la proposition 1.6.

Ce résultat est souvent utile pour se ramener à des familles de parties deux à deux disjointes car, du fait de l'axiome de σ -additivité, il est beaucoup plus facile de manipuler des réunions

de parties de E deux à deux disjointes. Voici une illustration de cette remarque dans la démonstration ci-dessous de l'inégalité de Bonferroni.

Proposition 1.7.

Inégalité de Bonferroni ou propriété de sous-additivité

Pour toute suite $(A_n)_{\mathbb{N}}$ d'éléments de \mathcal{A} ,

$$\mu \left(\bigcup_{k=0}^{+\infty} A_k \right) \leq \sum_{k=0}^{+\infty} \mu(A_k).$$

En conséquence, une réunion dénombrable d'événements négligeables pour μ est négligeable.

Démonstration : On applique la σ -additivité des mesures à la suite $(B_n)_{\mathbb{N}}$ de la proposition 1.6 et on utilise les propriétés élémentaires des probabilités énoncées dans la proposition 1.5. \square

La formule de Poincaré ci-dessous, qu'on admettra, généralise au cas de n événements la relation 3) de la proposition 1.5, page 11, établie pour deux événements (la démonstration peut se faire par récurrence sur l'entier n , une autre démonstration utilisant les propriétés de l'intégrale sera proposée au chapitre III dans l'exemple 3.9, page 37).

Proposition 1.8.

Formule de Poincaré

Pour toute suite (A_1, A_2, \dots, A_n) d'éléments de \mathcal{A} ,

$$\mu \left(\bigcup_{k=1}^{k=n} A_k \right) = \sum_{k=1}^{k=n} (-1)^{k+1} \left(\sum_{1 \leq i_1 < i_2 < \dots < i_k \leq n} \mu(A_{i_1} \cap A_{i_2} \cap \dots \cap A_{i_k}) \right).$$

Proposition 1.9.

Théorème de continuité monotone

1. Pour toute suite $(A_n)_{\mathbb{N}}$ d'éléments de \mathcal{A} , croissante au sens de l'inclusion, $(\mu(A_n))_{\mathbb{N}}$ est une suite réelle croissante convergeant vers $\mu \left(\bigcup_{k=0}^{+\infty} A_k \right)$ c-à-d

$$\mu \left(\bigcup_{k=0}^{+\infty} A_k \right) = \lim_{n \rightarrow +\infty} \mu(A_n).$$

2. Pour toute suite $(A_n)_{\mathbb{N}}$ d'éléments de \mathcal{A} , décroissante au sens de l'inclusion, la suite $(\mu(A_n))_{\mathbb{N}}$ est une suite réelle décroissante convergeant vers $\mu \left(\bigcap_{k=0}^{+\infty} A_k \right)$ c-à-d

$$\mu \left(\bigcap_{k=0}^{+\infty} A_k \right) = \lim_{n \rightarrow +\infty} \mu(A_n).$$

Démonstration : 1) Soit $(A_n)_{\mathbb{N}}$ suite croissante d'éléments de \mathcal{A} . Utilisons la suite construite dans la proposition 1.6. Comme la suite $(A_n)_{\mathbb{N}}$ est croissante, pour tout entier $n \geq 1$,

$B_n = A_n \setminus A_{n-1}$ et $B_0 = A_0$. $(B_n)_{\mathbb{N}}$ est une suite d'éléments de \mathcal{A} deux à deux disjoints avec $\bigcup_{k=0}^{+\infty} A_k = \bigcup_{k=0}^{+\infty} B_k$. Il vient

$$\begin{aligned} \mu \left(\bigcup_{k=0}^{+\infty} A_k \right) &= \mu \left(\bigcup_{k=0}^{+\infty} B_k \right) = \sum_{k=0}^{+\infty} \mu(B_k) = \mu(A_0) + \sum_{k=1}^{+\infty} (\mu(A_k) - \mu(A_{k-1})) \\ &= \mu(A_0) + \lim_n \sum_{k=1}^{k=n} (\mu(A_k) - \mu(A_{k-1})) = \lim_n \mu(A_n). \end{aligned}$$

D'où la première partie.

2) Comme μ est une probabilité,

$$\mu \left(\bigcap_{k=0}^{+\infty} A_k \right) = 1 - \mu \left(\bigcup_{k=0}^{+\infty} A_k^c \right).$$

Or $(A_n^c)_{\mathbb{N}}$ est une suite croissante d'éléments de \mathcal{A} . D'après la première partie de la démonstration, $\mu \left(\bigcup_{k=0}^{+\infty} A_k^c \right) = \lim_n \mu(A_n^c)$. Par suite

$$\mu \left(\bigcap_{k=0}^{+\infty} A_k \right) = \lim_n (1 - \mu(A_n^c)) = \lim_n \mu(A_n).$$

D'où la deuxième partie. \square

1.4 Fonctions de répartition

La possibilité de définir une probabilité sur une tribu à partir de la connaissance des valeurs de cette mesure sur une sous-famille de la tribu, résulte d'un théorème de prolongement assez technique que nous n'énoncerons pas. En revanche, il est souvent utile de montrer qu'il existe une unique probabilité sur la tribu qui prend des valeurs seulement connues sur une sous-famille de la tribu.

L'unicité dans le cas des probabilités résulte d'un théorème, appelé **théorème d'unicité**, qui découle lui-même du **théorème des classes monotones** qu'on admettra, dont il est utile de connaître l'énoncé. Commençons tout d'abord par donner deux définitions :

Définition 1.16.

Une famille \mathcal{M} de parties de E est appelée une **classe monotone** sur E si elle vérifie les trois axiomes suivants :

1. $E \in \mathcal{M}$.
2. Si $A \in \mathcal{M}$ et $B \in \mathcal{M}$ avec $B \subseteq A$, alors $A \setminus B \in \mathcal{M}$.
3. Si $(A_n)_{\mathbb{N}}$ est une suite croissante au sens de l'inclusion d'éléments de la famille \mathcal{M} , alors $\bigcup_{n=0}^{+\infty} A_n \in \mathcal{M}$.

De façon analogue à la définition correspondante pour les tribus, si \mathcal{J} est une famille de parties de E , on appellera **classe monotone engendrée par \mathcal{J}** la plus petite classe monotone sur E contenant tous les éléments de la famille \mathcal{J} . On vérifie aisément que la classe monotone engendrée par \mathcal{J} est l'intersection de toutes les classes monotones sur E contenant tous les éléments de la famille \mathcal{J} .

Moyennant ces deux définitions le théorème des classes monotones s'énonce :

Proposition 1.10.

Théorème des classes monotones (admis)

Soit \mathcal{J} une famille, stable par intersections finies, de parties d'un ensemble E , alors la classe monotone engendrée par \mathcal{J} coïncide avec la tribu engendrée par \mathcal{J} .

Une application importante de ce théorème est le théorème d'unicité sur les probabilités :

Proposition 1.11.

Théorème d'unicité pour les probabilités

Soit \mathcal{C} une famille, stable par intersections finies, de parties d'un ensemble E . Soit \mathcal{A} la tribu engendrée par \mathcal{C} , i.e. $\mathcal{A} = \sigma(\mathcal{C})$. Si μ et ν sont deux probabilités définies sur l'espace (E, \mathcal{A}) telles que, pour tout $A \in \mathcal{C}$, $\mu(A) = \nu(A)$, alors, pour tout $A \in \mathcal{A}$, $\mu(A) = \nu(A)$, i.e. $\mu = \nu$.

Démonstration : Notons \mathcal{H} la famille des événements $A \in \mathcal{A}$ tels que $\mu(A) = \nu(A)$. D'après l'item 1 de la proposition 1.9, on vérifie aisément que \mathcal{H} est une classe monotone qui contient la famille \mathcal{C} . Donc \mathcal{H} contient la classe monotone engendrée par \mathcal{C} . Comme, par hypothèse \mathcal{C} est stable par intersections finies, d'après le théorème des classes monotones, la classe monotone engendrée par \mathcal{C} coïncide avec la tribu engendrée par \mathcal{C} , c'est-à-dire \mathcal{A} . Finalement, pour tout $A \in \mathcal{A} \subseteq \mathcal{H}$, $\mu(A) = \nu(A)$. \square

Ce résultat montre que pour prouver que deux probabilités sont égales, il suffit de mettre en évidence qu'elles coïncident sur une famille engendrant la tribu, stable par intersections finies. Cette remarque justifie la définition suivante :

Définition 1.17.

Une famille de parties d'un ensemble non vide E stable par intersections finies est appelée un π -système de parties de E .

Par exemple, si nous prenons $E = \mathbb{R}$, la famille \mathcal{C} de tous les intervalles de \mathbb{R} de la forme $] - \infty, x]$ où x parcourt \mathbb{R} , est un π -système de parties de \mathbb{R} . De plus \mathcal{C} engendre la tribu borélienne de \mathbb{R} car, pour tout réel a et b avec $a < b$, $]a, b] =] - \infty, b] \setminus] - \infty, a]$. Le théorème d'unicité 1.11 appliqué à \mathcal{C} et \mathbb{R} muni de la tribu de Borel devient alors :

Proposition 1.12.

Lemme d'unicité pour les probabilités sur \mathbb{R}

Soient μ et ν deux probabilités sur \mathbb{R} .

Si pour tout $x \in \mathbb{R}$ $\mu(] - \infty, x]) = \nu(] - \infty, x])$, alors $\mu = \nu$.

Ce résultat a pour conséquence que pour identifier une probabilité μ sur \mathbb{R} , il suffit d'identifier l'application F_μ de \mathbb{R} dans $[0, 1]$, définie, pour tout $x \in \mathbb{R}$, par $F_\mu(x) := \mu(] - \infty, x])$.

Définition 1.18.

On dit que F_μ est la **fonction de répartition** de la probabilité μ , en abrégé **f.r.** .

Avec ces notations on peut énoncer autrement le lemme d'unicité pour les probabilités sur \mathbb{R} :

Proposition 1.13.

Deux probabilités sur \mathbb{R} sont identiques si, et seulement si, elles ont la même fonction de répartition.

Exemples 1.8.

1) La f.r. de δ_a , où $a \in \mathbb{R}$, est $\mathbb{1}_{[a, +\infty[}$.

2) La f.r. de $\mathcal{B}(p)$ est $p\mathbb{1}_{[1, +\infty[} + (1-p)\mathbb{1}_{[0, +\infty[}$.

3) La f.r. de $\frac{1}{4}\delta_0 + \frac{3}{4}\mathcal{N}_1(0, 1)$ est $\frac{1}{4}\mathbb{1}_{[0, +\infty[} + \frac{3}{4}F$ où F désigne la f.r. de la probabilité $\mathcal{N}_1(0, 1)$.

Les valeurs de la fonction de répartition de la probabilité $\mathcal{N}_1(0, 1)$ sont "tabulées". On trouvera la tables des valeurs de la fonction de répartition de la probabilité normale standard, appelée communément **table de la loi normale centrée-réduite**, avec un mode d'emploi dans l'annexe B, page 211, de ce cours.

A titre d'entraînement, on pourra également chercher à exprimer les fonctions de répartition des probabilités $\mathcal{E}(\alpha)$ et $\mathcal{U}([a, b])$ (On trouvera leur expression dans le formulaire reproduit en annexe A, page 205).

Exercice 1.15. (Corrigé de l'exercice : page 150)

Soit F l'application de \mathbb{R} dans \mathbb{R} définie, pour tout réel x , par $F(x) := 1 - \frac{1}{2}e^{-x}$ si $x \geq 0$, et $F(x) := \frac{1}{2}e^x$ si $x \leq 0$. Montrer que F est la fonction de répartition d'une probabilité à densité qu'on déterminera.

Proposition 1.14.

Soit μ une probabilité sur \mathbb{R} de fonction de répartition F . Alors

1. F est croissante sur \mathbb{R} et admet des limites à droite en tout point de $\mathbb{R} \cup \{-\infty\}$ et à gauche en tout point de $\mathbb{R} \cup \{+\infty\}$. De plus F est continue-à-droite sur \mathbb{R} ,

$$\lim_{x \rightarrow -\infty} F(x) = 0 \text{ et } \lim_{x \rightarrow +\infty} F(x) = 1.$$

2. Pour tous réels a, b avec $a < b$:

- (a) $\mu([a, b]) = F(b) - F(a)$ et $\mu(]-\infty, a]) = F(a-)$ où $F(a-)$ désigne la limite-à-gauche de F au point a .
- (b) $\mu(\{a\}) = F(a) - F(a-)$.
- (c) F est continue en a si, et seulement si, $\mu(\{a\}) = 0$.

Démonstration : 1) Soient x, y des réels vérifiant $x \leq y$. Comme $]-\infty, x] \subseteq]-\infty, y]$ il vient $F(x) = \mu(]-\infty, x]) \leq \mu(]-\infty, y]) = F(y)$. Donc F est croissante sur \mathbb{R} .

Pour montrer que F admet une limite-à-gauche, considérons un point a de $\mathbb{R} \cup \{+\infty\}$ et posons $l := \sup_{x < a} F(x)$. l est dans \mathbb{R} puisque F est bornée par 1. Soit $\varepsilon > 0$, il existe $x_0 < a$ tel que $l \geq F(x_0) > l - \varepsilon$. Donc, pour tout $x \in]x_0, a[$, $l \geq F(x) \geq F(x_0) > l - \varepsilon$ c-à-d $|F(x) - l| < \varepsilon$, ce qui donne l'existence de la limite-à-gauche en a pour F .

On montre de même l'existence d'une limite-à-droite $F(a+) := \inf_{x > a} F(x)$.

La suite d'intervalles $\left(] - \infty, a + \frac{1}{n}] \right)_{n \in \mathbb{N}^*}$ est décroissante et $\bigcap_{k=0}^{+\infty}] - \infty, a + \frac{1}{n}] =] - \infty, a]$, donc par le théorème de continuité monotone 1.9 de la page 12

$$\mu(] - \infty, a]) = \lim_n \mu\left(] - \infty, a + \frac{1}{n}]\right)$$

c-à-d $F(a) = \lim_n F\left(a + \frac{1}{n}\right) = F(a+)$ car la limite-à-droite existe au point a . F est donc continue-à-droite en tout point de \mathbb{R} .

La suite d'intervalles $(] - \infty, -n])_{n \in \mathbb{N}}$ est décroissante et $\bigcap_{n=0}^{+\infty}] - \infty, -n] = \emptyset$. La suite $(] - \infty, n])_{n \in \mathbb{N}}$ est croissante et $\bigcup_{n=0}^{+\infty}] - \infty, n] = \mathbb{R}$. Par application du théorème de continuité monotone à ces deux dernières suites, on obtient les valeurs des limites de F en $-\infty$ et $+\infty$.

2-a) $\mu([a, b]) = \mu(] - \infty, b]) - \mu(] - \infty, a])$ car $]a, b] =] - \infty, b] \setminus] - \infty, a]$. D'où le premier résultat.

Comme $] - \infty, a[= \bigcup_{n \geq 1}] - \infty, a - \frac{1}{n}]$ et que F admet une limite-à-gauche en a d'après la première partie,

$$\mu(] - \infty, a[) = \lim_{n \rightarrow +\infty} \mu(] - \infty, a - \frac{1}{n}]) = \lim_{n \rightarrow +\infty} F\left(a - \frac{1}{n}\right) = \lim_{x \rightarrow a, x < a} F(x) = F(a-).$$

ce qui donne le second résultat.

2-b) On peut écrire $\{a\} =] - \infty, a] \setminus] - \infty, a[$. Par suite, $\mu(\{a\}) = \mu(] - \infty, a]) - \mu(] - \infty, a[) = F(a) - F(a-)$.

2-c) F est continue en a si, et seulement si, $F(a) = F(a-)$ c-à-d $\mu(\{a\}) = 0$, d'après 2-b. \square

Exercice 1.16. (Corrigé de l'exercice : page 150)

Montrer que, pour tout réel x , $\mathcal{N}_1(0, 1)(\{x\}) = 0$ et en déduire que la probabilité $\mathcal{N}_1(0, 1)$ ne peut pas s'écrire comme combinaison linéaire de probabilités de Dirac.

Exercice 1.17. (Corrigé de l'exercice : page 150)

1. Montrer que si μ est une probabilité admettant une densité sur \mathbb{R} , alors pour tout réel a , $\mu(\{a\}) = 0$, c'est-à-dire tout singleton est négligeable pour μ . On dit dans ce cas que la probabilité μ est **diffuse** sur \mathbb{R} .
2. Avec les notations de la proposition précédente, montrer que pour tous réels a, b vérifiant $a < b$, $\mu([a, b]) = F(b-) - F(a)$ et $\mu([a, b]) = F(b-) - F(a-)$.

Exercice 1.18. (Corrigé de l'exercice : page 151)

Calculer $\mathcal{U}([0, 1]) \left(\left[\frac{1}{6}, \frac{4}{3} \right] \right)$, $\mathcal{U}([0, 1]) (\mathbb{Q})$, $\mathcal{E}(2) \left(\{\pi\} \cup \left[\frac{9}{2}, 7 \right] \right)$.

On admettra le résultat suivant, réciproque de l'item 1 de la proposition 1.14, qui prouve qu'il y a bijection entre l'ensemble des probabilités sur \mathbb{R} et l'ensemble des fonctions sur \mathbb{R} , croissantes, continues-à-droite sur \mathbb{R} , telles que $\lim_{x \rightarrow -\infty} F(x) = 0$ et $\lim_{x \rightarrow +\infty} F(x) = 1$ (cf. [3] exercice I-16) :

Proposition 1.15.

Si F est une application croissante de \mathbb{R} dans $[0, 1]$, continue-à-droite sur \mathbb{R} avec

$$\lim_{x \rightarrow -\infty} F(x) = 0 \text{ et } \lim_{x \rightarrow +\infty} F(x) = 1,$$

alors il existe une unique probabilité sur \mathbb{R} dont F est la fonction de répartition.

Exercice 1.19. (Corrigé de l'exercice : page 151)

Donner une représentation graphique de l'application

$$F : t \in \mathbb{R} \mapsto F(t) = \frac{1}{4}(t+2)\mathbb{1}_{[-1,0[\cup]1,2[}(t) + \frac{3}{4}\mathbb{1}_{[0,1]}(t) + \mathbb{1}_{[2,+\infty[}(t),$$

et montrer que F est la fonction de répartition d'une probabilité à densité qu'on précisera.

Chapitre 2

Loi d'un vecteur aléatoire

2.1 Remarques sur la modélisation de l'aléatoire

Le but de ce premier paragraphe est de fournir quelques éléments de réflexion sur la modélisation mathématique de phénomènes aléatoires. Pour une analyse plus approfondie sur l'intérêt d'introduire la notion de variable aléatoire et de loi de probabilité, on pourra consulter l'annexe ??, page ??.

Considérons les deux situations suivantes :

2.1.1 Cas discret

Une personne s'intéresse à la somme des valeurs obtenues dans le lancer simultané de deux dés équilibrés. On modélisera l'ensemble des issues possibles de cette expérience aléatoire par

$$\Omega := \{(i, j) \in \mathbb{N}^2 / 1 \leq i, j \leq 6\}.$$

Les événements peuvent être modélisés par des parties de Ω . On peut prendre comme tribu des événements l'ensemble $\mathcal{P}(\Omega)$ de toutes les parties de Ω . Les dés étant équilibrés, on choisira pour probabilité \mathbb{P} sur $(\Omega, \mathcal{P}(\Omega))$ l'équiprobabilité sur Ω i.e. pour tout $(i, j) \in \Omega$, $\mathbb{P}(\{(i, j)\}) = \frac{1}{36}$ ou encore $\mathbb{P} = \frac{1}{36} \sum_{1 \leq i, j \leq 6} \delta_{(i, j)}$.

Le triplet $(\Omega, \mathcal{P}(\Omega), \mathbb{P})$ représente le modèle mathématique permettant de traiter la situation. Cependant comme on s'intéresse plutôt à la somme des valeurs obtenues, l'événement "*La somme des valeurs obtenues appartient à A*", où A est un borélien de \mathbb{R} , se modélise par la partie e_A de Ω formée des couples (i, j) tels que $i + j \in A$. On peut aussi écrire l'événement e_A grâce au langage des applications en notant X l'application de Ω dans \mathbb{R} qui, à tout $\omega = (i, j)$, associe $X(\omega) = i + j$ et en remarquant que $e_A = \{\omega \in \Omega / X(\omega) \in A\} = \{X \in A\}$ c-à-d que e_A est l'image-réciproque de A par l'application X . On remarque enfin que ce qui est important pour notre étude du phénomène c'est de connaître la valeur de $\mathbb{P}(e_A) = \mathbb{P}(X \in A)$ pour tout borélien A de \mathbb{R} .

2.1.2 Cas continu

Envisageons maintenant le cas d'un ingénieur hydraulicien qui s'intéresse aux risques d'inondation par un fleuve dans l'intention de construire une digue protectrice. Pour cela il va considérer l'évolution de la hauteur du niveau de l'eau sur l'année. Cela revient à considérer la hauteur sur une année comme une application continue de $[0, 1]$ dans \mathbb{R}^+ . L'ensemble des issues possibles de ce phénomène aléatoire peut être modélisé par $\Omega := \mathcal{C}([0, 1], \mathbb{R}^+)$ ensemble des applications continues de $[0, 1]$ dans \mathbb{R}^+ . Comme pour \mathbb{R} , et contrairement à ce qu'on a fait pour le cas précédent, il n'est pas possible de prendre $\mathcal{P}(\Omega)$ comme tribu sur Ω . On considérera une tribu \mathcal{F} plus petite qu'on ne précise pas pour l'instant. De même on ne précisera pas la probabilité \mathbb{P} définie sur \mathcal{F} . On verra plus loin qu'au fond ce n'est pas nécessaire, seule l'existence du triplet $(\Omega, \mathcal{F}, \mathbb{P})$ devant être assurée.

En fait l'ingénieur s'intéressera surtout aux événements de la forme "*La hauteur maximale du niveau du fleuve sur une année appartient à A* " où A est un intervalle de \mathbb{R} . Cet événement se modélise par la partie e_A de Ω formée des fonctions $\omega \in \mathcal{C}([0, 1], \mathbb{R}^+)$ telles que $\sup_{0 \leq t \leq 1} \omega(t) \in A$. On peut aussi écrire l'événement e_A grâce au langage des applications en notant X l'application de Ω dans \mathbb{R} qui à tout ω associe $X(\omega) = \sup_{0 \leq t \leq 1} \omega(t)$ et en remarquant que $e_A = \{\omega \in \Omega / X(\omega) \in A\} = \{X \in A\}$ c-à-d que e_A est l'image-réciproque de A par l'application X .

Pour que l'expression $\mathbb{P}(X \in A)$ ait un sens, il faudra s'assurer (ou imposer) plus généralement que, pour tout borélien A de \mathbb{R} , l'image-réciproque de A par l'application X soit un élément de \mathcal{F} . Car, comme dans la situation précédente, c'est la valeur de $\mathbb{P}(X \in A)$ qui intéressera l'ingénieur, c-à-d l'application $\mathbb{P}_X : A \in \mathcal{F} \mapsto \mathbb{P}(X \in A)$. \mathbb{P}_X est une probabilité sur \mathbb{R} donc un objet mathématique beaucoup plus facile à manipuler qu'une probabilité sur une tribu de $\mathcal{C}([0, 1], \mathbb{R}^+)$.

2.1.3 Principe de modélisation

En conclusion de ces deux exemples on notera que, en pratique, modéliser mathématiquement un phénomène aléatoire revient à introduire :

1. un triplet $(\Omega, \mathcal{F}, \mathbb{P})$, sans en préciser davantage les termes, comme un espace de probabilité abstrait,
2. une application $X : \Omega \mapsto \mathbb{R}^d$ telle que, pour tout borélien A de \mathbb{R}^d , l'image-réciproque de A par l'application X soit un élément de \mathcal{F} .

C'est alors l'application $\mathbb{P}_X : A \in \mathcal{F} \mapsto \mathbb{P}(X \in A)$ qui sera l'objet important du modèle, celui qui traduira mathématiquement le problème particulier qui intéresse l'ingénieur au sein de la situation aléatoire globale.

Dans la suite de l'ouvrage le triplet $(\Omega, \mathcal{F}, \mathbb{P})$ désignera un espace de probabilité pris comme référence et quelquefois appelé **espace de base**. Les ensembles mesurables relativement à \mathcal{F} seront appelés **événements** de Ω .

2.2 Applications mesurables

Définition 2.1.

Soient (E, \mathcal{A}) et (F, \mathcal{B}) deux espaces mesurables, une application f de E dans F est dite $(\mathcal{A}, \mathcal{B})$ -mesurable si, pour tout $B \in \mathcal{B}$, $\{f \in B\} \in \mathcal{A}$.

Dans les cas où (E, \mathcal{A}) est quelconque et $(F, \mathcal{B}) := (\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$, on dit simplement \mathcal{A} -mesurable au lieu de $(\mathcal{A}, \mathcal{B}(\mathbb{R}^k))$ -mesurable.

Une application \mathcal{A} -mesurable à valeurs dans $\overline{\mathbb{R}}$ est une application $(\mathcal{A}, \mathcal{B}(\overline{\mathbb{R}}))$ -mesurable.

La proposition suivante donne un premier exemple d'applications mesurables :

Proposition 2.1.

Soit A une partie de E . Alors $\mathbb{1}_A$ est \mathcal{A} -mesurable si, et seulement si, $A \in \mathcal{A}$.

Démonstration : On remarque que si B est un borélien de \mathbb{R} , l'image réciproque de B par $\mathbb{1}_A$ est l'un des ensembles \emptyset , A , A^c , ou E . Ce qui prouve par définition de la mesurabilité que $\mathbb{1}_A$ est \mathcal{A} -mesurable si, et seulement si, A est \mathcal{A} -mesurable. \square

Définition 2.2.

Dans les cas où $(E, \mathcal{A}) := (\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$ et $(F, \mathcal{B}) := (\mathbb{R}^k, \mathcal{B}(\mathbb{R}^k))$ on dit que f est borélienne pour exprimer qu'elle est $(\mathcal{B}(\mathbb{R}^n), \mathcal{B}(\mathbb{R}^k))$ -mesurable.

La proposition suivante donne des classes importantes de fonctions boréliennes qui correspondent à la plupart des cas qu'on considérera par la suite. Pour une démonstration d'une partie de la proposition on pourra consulter [2] exercice I-10.

Proposition 2.2.

(admis)

Toute application continue de \mathbb{R}^n dans \mathbb{R}^k est borélienne. Toute application monotone de \mathbb{R} dans \mathbb{R} est borélienne. Toute dérivée d'une application dérivable de \mathbb{R} dans \mathbb{R} est borélienne.

Exercice 2.1. (Corrigé de l'exercice : page 152)

Soit f une application borélienne de \mathbb{R}^k dans \mathbb{R}^d et φ une application \mathcal{A} -mesurable de E dans \mathbb{R}^k . Montrer que l'application $f \circ \varphi$ est une application \mathcal{A} -mesurable de E dans \mathbb{R}^d .

Comme pour la notion d'ensemble mesurable, les applications mesurables correspondent aux applications sur lesquelles la théorie de la mesure permet de dire quelque chose d'intéressant. On doit s'attendre à ce que toutes les applications qu'on est amené à manipuler dans la pratique soient mesurables.

Introduisons la notation suivante qui est utile pour étendre une propriété, vraie pour la classe des fonctions positives, à la classe des fonctions de signe quelconque :

Définition 2.3.

Si f est une application d'un ensemble E dans $\overline{\mathbb{R}}$ notons $f^+ := \sup(f, 0)$ et $f^- := \sup(-f, 0)$. Les applications f^+ et f^- sont appelées respectivement la partie positive et la partie négative de f .

On vérifie aisément que ce sont des applications à valeurs dans $[0, +\infty]$ telles que $|f| = f^+ + f^-$ et $f = f^+ - f^-$.

Exemples 2.1.

Supposons $E := \mathbb{R}$, si $f(x) = x$, $f^+(x) = x\mathbb{1}_{[0, +\infty[}(x)$ et $f^-(x) = -x\mathbb{1}_{]-\infty, 0]}(x)$.

Grosso modo les opérations classiques sur les applications mesurables conservent la mesurabilité. Plus précisément, on admettra :

Proposition 2.3.

1. Si f et g sont des applications \mathcal{A} -mesurables d'un ensemble E dans \mathbb{R}^d et α un réel, alors αf , $\langle f, g \rangle$, $f + g$, $|f|$ sont des applications \mathcal{A} -mesurables, où $\langle \cdot, \cdot \rangle$ et $|\cdot|$ désignent respectivement les produit scalaire et norme usuels de \mathbb{R}^d .
2. Si f et g sont des applications \mathcal{A} -mesurables d'un ensemble E dans $\overline{\mathbb{R}}$, alors f^+ , f^- sont des applications \mathcal{A} -mesurables.
3. Si $(f_n)_{n \in \mathbb{N}}$ est une suite d'applications \mathcal{A} -mesurables d'un ensemble E dans $\overline{\mathbb{R}}$, Alors $\sup_{n \in \mathbb{N}}(f_n)$, $\inf_{n \in \mathbb{N}}(f_n)$ sont des applications \mathcal{A} -mesurables.
4. Si $(f_n)_{n \in \mathbb{N}}$ est une suite d'applications \mathcal{A} -mesurables d'un ensemble E dans \mathbb{R}^d convergeant simplement vers une application f , alors sa limite f est \mathcal{A} -mesurable.

Définition 2.4.

Une application \mathcal{A} -mesurable f est dite **étagée sur E** si elle est à valeurs dans \mathbb{R} et si elle ne prend qu'un nombre fini de valeurs distinctes.

Si on note $\alpha_1, \alpha_2, \dots, \alpha_n$ les valeurs deux à deux distinctes d'une application étagée f et si on pose, pour tout entier k vérifiant $1 \leq k \leq n$, $A_k := \{x \in E / f(x) = \alpha_k\}$, alors f s'écrit sous la forme

$$f = \sum_{k=1}^n \alpha_k \mathbb{1}_{A_k}.$$

Cette écriture s'appelle la **décomposition canonique** de f . On vérifie aisément que la décomposition canonique d'une application étagée est unique.

L'intérêt de cette définition réside dans la proposition suivante. Pour la démonstration on pourra consulter [2] exercice I-13.

Proposition 2.4.

Lemme fondamental (admis)

Toute application \mathcal{A} -mesurable de E dans $[0, +\infty]$ est la limite d'une suite croissante d'applications \mathcal{A} -mesurables étagées et positives.

Ce lemme est à la base d'une technique de démonstration utilisée en probabilités lorsqu'on veut montrer que les applications \mathcal{A} -mesurables possèdent une certaine propriété \mathcal{P} . Pour cela, on montre que les indicatrices $\mathbb{1}_A$, où $A \in \mathcal{A}$, vérifient \mathcal{P} , puis on montre qu'il en est de même pour les applications \mathcal{A} -mesurables de la forme $\sum_{i=1}^n \alpha_i \mathbb{1}_{A_i}$ où $\alpha_i \in \mathbb{R}^+$ et $A_i \in \mathcal{A}$, $1 \leq i \leq n$.

On montre ensuite, en utilisant le lemme fondamental, que la propriété \mathcal{P} est encore vérifiée par les applications \mathcal{A} -mesurables positives, puis par les applications \mathcal{A} -mesurables quelconques f en remarquant que $f = f^+ - f^-$ où $f^+ := \sup(f, 0)$ et $f^- := \sup(-f, 0)$ sont des applications \mathcal{A} -mesurables positives. Cette technique de démonstration est souvent appelée "**technique des fonctions étagées**".

2.3 Loi d'une variable aléatoire

2.3.1 Variables aléatoires

Parallèlement aux définitions introduites ci-dessus, une terminologie différente est utilisée en probabilité pour les applications mesurables dans le cas où (E, \mathcal{A}) est l'espace mesurable de base (Ω, \mathcal{F}) .

Définition 2.5.

Si $(E, \mathcal{A}) := (\Omega, \mathcal{F})$ et $(F, \mathcal{B}) = (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$, une application $(\mathcal{F}, \mathcal{B}(\mathbb{R}^d))$ -mesurable s'appelle un **vecteur aléatoire**, ou **variable aléatoire vectorielle**, de dimension d .

Un vecteur aléatoire de dimension $d = 1$ s'appelle aussi une **variable aléatoire réelle** en abrégé **v.a.r.**.

On peut être quelquefois amené à considérer des variables aléatoires à valeurs dans $\overline{\mathbb{R}}$, ce sont les applications $(\mathcal{F}, \mathcal{B}(\overline{\mathbb{R}}))$ -mesurables de Ω dans $\overline{\mathbb{R}}$.

Les variables aléatoires sont traditionnellement notées par des lettres majuscules X, Y, \dots

La proposition suivante est l'énoncé avec un vocabulaire différent du résultat de l'exercice 2.1 de la page 21 sur la composition des applications mesurables.

Proposition 2.5.

Si f est une application borélienne de \mathbb{R}^k dans \mathbb{R}^d et X un vecteur aléatoire de dimension k , alors l'application $f \circ X$ est un vecteur aléatoire de dimension d .

Démonstration : Il suffit pour cela de remarquer que si B est un borélien de \mathbb{R}^d , alors l'image-réciproque de B par $f \circ X$ est $(f \circ X)^{-1}(B) = X^{-1}[(f^{-1}(B))]$ et d'appliquer ensuite la définition de la mesurabilité de f et X . \square

On notera dans la suite par abus $f(X)$ au lieu de $f \circ X$. Par exemple, on écrira e^X pour exprimer l'application composée de l'application exponentielle et de la variable aléatoire réelle X .

Proposition 2.6.

$X = (X_1, X_2, \dots, X_d)$ un vecteur aléatoire de dimension k si, et seulement si, pour tout $i = 1, 2, \dots, d$, X_i est une variable aléatoire réelle.

Démonstration : La démonstration est une conséquence directe de la proposition 2.5 où on prend pour f les projections de \mathbb{R}^d sur \mathbb{R} . \square

Deux vecteurs aléatoires X et Y de dimension d sont égaux presque-sûrement si, et seulement si, $\mathbb{P}(X \neq Y) = 0$. L'égalité presque-sûre est une relation d'équivalence sur l'ensemble des vecteurs aléatoires de dimension d .

2.3.2 Loi d'une variable aléatoire

Proposition 2.7.

Soit X un vecteur aléatoire de dimension d . L'application

$$\mathbb{P}_X : B \in \mathcal{B}(\mathbb{R}^d) \mapsto \mathbb{P}_X(B) := \mathbb{P}(\{X \in B\}) \in [0, 1]$$

est une probabilité sur \mathbb{R}^d .

Démonstration : On rappelle la notation $\{X \in B\} := \{\omega \in \Omega / X(\omega) \in B\}$ et on notera que $\{X \in B\} \in \mathcal{F}$, ce qui donne bien un sens à $\mathbb{P}(\{X \in B\})$.

Soit $B \in \mathcal{B}(\mathbb{R}^d)$, $\mathbb{P}_X(B) = \mathbb{P}(\{X \in B\}) \in [0, 1]$. De plus, comme $\{X \in \mathbb{R}^d\} = \Omega$, $\mathbb{P}_X(\mathbb{R}^d) = \mathbb{P}(\{X \in \mathbb{R}^d\}) = \mathbb{P}(\Omega) = 1$. Soit $(A_n)_{n \in \mathbb{N}}$ une suite deux à deux disjointe de boréliens de \mathbb{R}^d , alors

$$\left\{ X \in \bigcup_{k \in \mathbb{N}} A_k \right\} = \bigcup_{k \in \mathbb{N}} \{X \in A_k\}$$

l'union du second membre étant deux à deux disjointe. Par suite

$$\begin{aligned} \mathbb{P}_X \left(\bigcup_{k \in \mathbb{N}} A_k \right) &= \mathbb{P} \left(X \in \bigcup_{k \in \mathbb{N}} A_k \right) = \mathbb{P} \left(\bigcup_{k \in \mathbb{N}} \{X \in A_k\} \right) \\ &= \sum_{k \in \mathbb{N}} \mathbb{P}(X \in A_k) = \sum_{k \in \mathbb{N}} \mathbb{P}_X(A_k) \end{aligned}$$

d'où la σ -additivité de \mathbb{P}_X . \square

Définition 2.6.

La probabilité \mathbb{P}_X est appelée la loi de probabilité relativement à \mathbb{P} du vecteur aléatoire X ou plus simplement la loi de X .

On notera que cette loi dépend de X mais aussi de la probabilité \mathbb{P} de l'espace de probabilité de base.

On admettra le résultat théorique suivant démontré dans [3] exercice I-16 :

Proposition 2.8.

Si μ est une probabilité sur \mathbb{R} , alors il existe un espace de probabilité de base $(\Omega, \mathcal{F}, \mathbb{P})$ et une variable aléatoire réelle X sur cet espace telle que $\mathbb{P}_X = \mu$.

Exemples 2.2.

Soit X une variable aléatoire réelle de loi $\mathcal{N}_1(0, 1)$ (on notera qu'une telle affirmation a un sens d'après la proposition précédente). Déterminons la loi de la variable aléatoire réelle $Y := X^2$.

Pour cela il suffit d'identifier la fonction de répartition F_Y de la variable aléatoire réelle Y i.e. la f.r. de la probabilité \mathbb{P}_Y . Soit $y \in \mathbb{R}$,

$$F_Y(y) = \mathbb{P}_Y([-\infty, y]) = \mathbb{P}(Y \in [-\infty, y]) = \mathbb{P}(Y \leq y).$$

Remarquons que, si $y < 0$, $\{Y \leq y\} = \{X^2 \leq y\} = \emptyset$ et, si $y \geq 0$,

$$\{Y \leq y\} = \{X^2 \leq y\} = \{-\sqrt{y} \leq X \leq \sqrt{y}\}.$$

Par suite si $y < 0$, $F_Y(y) = 0$, et si $y \geq 0$,

$$\begin{aligned} F_Y(y) &= \mathbb{P}(-\sqrt{y} \leq X \leq \sqrt{y}) = \mathbb{P}_X([-\sqrt{y}, \sqrt{y}]) = F_X(\sqrt{y}) - F_X(-\sqrt{y}) \\ &= \int_{-\sqrt{y}}^{\sqrt{y}} \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}t^2} dt = \int_0^y \frac{1}{\sqrt{2\pi x}} e^{-\frac{x}{2}} dx. \end{aligned}$$

On notera qu'on a utilisé dans la troisième égalité la continuité (à gauche) de F_X . Le résultat précédent montre que la f.r. de \mathbb{P}_Y peut s'écrire

$$F_Y(y) = \mathbb{P}_Y([-\infty, y]) = \int_{-\infty}^y \rho(x) dx \text{ avec } \rho(x) := \frac{1}{\sqrt{2\pi x}} e^{-\frac{x}{2}} \mathbb{1}_{]0, +\infty[}(x).$$

La loi de Y , \mathbb{P}_Y , est la probabilité sur \mathbb{R} admettant ρ pour densité. Elle appartient à la famille des **lois gamma**, voir la définition dans le formulaire de l'annexe A, page 205. On la note $\gamma(\frac{1}{2}, \frac{1}{2})$. \square

Exercice 2.2. (Corrigé de l'exercice : page 152)

Soit X une variable aléatoire réelle de loi $\mathcal{N}_1(0, 1)$. En utilisant une démarche analogue à celle adoptée dans l'exemple précédent, montrer que la loi de la variable aléatoire réelle $Y := e^X$ admet pour densité la fonction ρ définie sur \mathbb{R} par

$$\rho(x) := \frac{1}{x\sqrt{2\pi}} \exp\left(-\frac{1}{2}(\ln x)^2\right) \mathbb{1}_{]0, +\infty[}(x).$$

On dit que Y suit la **loi Log-normale standard**.

Exercice 2.3. (Corrigé de l'exercice : page 153)

On considère une variable aléatoire réelle X dont la fonction de répartition est donnée par

$$F_X(t) := \frac{1}{2} [e^t \mathbb{1}_{]-\infty, 0]}(t) + (2 - e^{-t}) \mathbb{1}_{]0, +\infty[}(t)].$$

Déterminer la loi de la variable aléatoire réelle $Y = |X|$.

Définition 2.7.

Soient $m \in \mathbb{R}$ et $\sigma > 0$. Nous appellerons **loi de Gauss-Laplace de paramètres m et σ^2** , et noterons $\mathcal{N}_1(m, \sigma^2)$, la probabilité sur \mathbb{R} admettant pour densité la fonction ρ définie sur \mathbb{R} , pour tout réel x , par

$$\rho(x) := \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right).$$

Le résultat suivant est souvent utile dans les calculs pratiques en permettant de se ramener à des variables de loi de Gauss-Laplace standard :

Proposition 2.9.

Procédé de standardisation

Avec les notations précédentes, une variable aléatoire réelle X suit la loi $\mathcal{N}_1(m, \sigma^2)$ si, et seulement si, la variable aléatoire réelle $Z := \frac{X - m}{\sigma}$ suit la loi $\mathcal{N}_1(0, 1)$.

Exercice 2.4. (Corrigé de l'exercice : page 153)

Démontrer le résultat précédent.

Définition 2.8.

Une v.a. X à valeurs dans \mathbb{R}^d est dite **discrète** si sa loi est discrète.

Les variables aléatoires réelles discrètes constituent une famille de v.a.r importante dans les applications des probabilités, une autre classe de v.a.r. très importante aussi est celle des v.a.r. à densité.

Définition 2.9.

Une variable aléatoire réelle est dite à **densité** (ou **absolument continue**) sur \mathbb{R} si sa loi est à densité sur \mathbb{R} .

Exemples 2.3.

- 1) Les variables aléatoires réelles de Poisson, de Bernoulli, binomiale, hypergéométrique, géométrique, uniforme-discrète sont des exemples de v.a.r. discrètes.
- 2) Les v.a.r. de Gauss-Laplace, exponentielle, uniforme sur un intervalle de \mathbb{R} sont des exemples de v.a.r. à densité sur \mathbb{R} .

Pour les définitions des lois usuelles (discrètes ou à densité), on pourra se reporter au formulaire de l'annexe A, page 205, de ce cours.

Exercice 2.5. (Corrigé de l'exercice : page 153)

Montrer que si X est une variable aléatoire réelle de loi

$$\mathbb{P}_X := \sum_{n \in \mathbb{N}} p_n \delta_n$$

où $(p_n)_{\mathbb{N}}$ est une suite de réels positifs ou nuls, alors $p_n = \mathbb{P}(X = n)$ pour tout $n \in \mathbb{N}$.

On notera qu'on peut avoir affaire à des probabilités qui ne sont ni discrètes ni à densité. Par exemple, on peut avoir des probabilités μ définies sur \mathbb{R} , telles $\mu = \mu_1 + \mu_2$ où μ_1 est une mesure à densité (mais pas une probabilité) et μ_2 une mesure discrète (mais pas une probabilité), c'est-à-dire qu'il existe une application f (par exemple positive et continue sur \mathbb{R}), et une suite de

réels positifs $(\alpha_n)_{\mathbb{N}}$, avec $\int_{-\infty}^{+\infty} f(t) dt + \sum_{n=0}^{+\infty} \alpha_n = 1$, telle que, pour tout intervalle $]a, b[$ de \mathbb{R} ,

$$\mu(]a, b[) = \int_a^b f(t) dt + \sum_{n=0}^{+\infty} \alpha_n \delta_n(]a, b[).$$

Dans ce cas $\mu_1([a, b]) = \int_a^b f(t) dt$ et $\mu_2([a, b]) = \sum_{n=0}^{+\infty} \alpha_n \delta_n([a, b])$. Comme on n'a pas nécessairement $\int_{-\infty}^{+\infty} f(t) dt = 1$ et $\sum_{n=0}^{+\infty} \alpha_n = 1$, les mesures μ_1 et μ_2 ne sont pas des probabilités.

Les variables aléatoires réelles discrètes sont les variables aléatoires réelles à valeurs presque-sûrement dans un ensemble dénombrable. De façon précise :

Proposition 2.10.

Un vecteur aléatoire X de dimension d est discret si, et seulement si, il existe une partie $D := \{e_k, k \in K \subseteq \mathbb{N}\}$ de \mathbb{R}^d telle que $\mathbb{P}(X \in D) = 1$. Dans ce cas la loi du vecteur aléatoire X s'écrit

$$\mathbb{P}_X = \sum_{k \in K} \mathbb{P}(X = e_k) \delta_{e_k}.$$

Démonstration : Soit X une v.a. telle qu'il existe une partie dénombrable D de \mathbb{R}^d avec $X \in D := \{e_k, k \in K \subseteq \mathbb{N}\}$ presque-sûrement i.e. $\mathbb{P}(X \in D) = 1$.

Soit A un borélien de \mathbb{R}^d , on a $\mathbb{P}_X(A) = \mathbb{P}_X(A \cap D) = \mathbb{P}(X \in A \cap D)$. Comme

$$\{X \in A \cap D\} = \bigcup_{x \in A \cap D} \{X = x\}$$

et que l'union est mutuellement disjointe, on peut écrire

$$\mathbb{P}_X(A) = \sum_{x \in A \cap D} \mathbb{P}(X = x) = \sum_{x \in D} \mathbb{P}(X = x) \mathbf{1}_A(x) = \sum_{x \in D} \mathbb{P}(X = x) \delta_x(A) = \sum_{k \in K} \mathbb{P}(X = e_k) \delta_{e_k}(A).$$

La v.a. X est donc discrète et sa loi est $\mathbb{P}_X = \sum_{k \in K} \mathbb{P}(X = e_k) \delta_{e_k}$.

Réciproquement soit X une v.a. de loi $\mu = \sum_{n \in K} p_n \delta_{e_n}$ où $(p_n)_K$ est une suite (finie ou infinie) de

réels strictement positifs avec $K \subseteq \mathbb{N}$, et $(e_n)_K$ une suite (finie ou infinie) d'éléments de \mathbb{R}^d .

Prenant $D := \{e_n/n \in K\}$, on a $\mathbb{P}(X \in D) = 1$ et, pour tout $n \in K$, $\mathbb{P}(X = e_n) = p_n$. \square

On dit aussi dans ce cas que la loi de X est **portée par** D , ou encore que X a ses valeurs **presque-sûrement dans** D , pour exprimer $\mathbb{P}(X \in D) = 1$. On notera que D est une partie dénombrable (finie ou infinie) de \mathbb{R}^d .

Ce résultat ramène alors la détermination de la loi d'une variable aléatoire réelle discrète au calcul des coefficients $\mathbb{P}(X = a_k)$ qui interviennent dans son écriture. Il explique aussi le choix de certains auteurs de manuels scolaires de définir la loi d'une variable aléatoire réelle à valeurs dans \mathbb{N} comme étant l'application $n \in \mathbb{N} \mapsto \mathbb{P}(X = n)$. En fait cette définition n'est pas judicieuse car elle ne se généralise pas au cas des variable aléatoire réelle à densité. En effet, pour une variable aléatoire réelle à densité, pour tout réel x , $\mathbb{P}(X = x) = \mathbb{P}_X(\{x\}) = 0$ d'après ce qui a été vu au premier chapitre. Par suite l'application $x \in \mathbb{R} \mapsto \mathbb{P}(X = x)$ est l'application-nulle pour toute variable aléatoire réelle X admettant un densité, ce qui ne présente plus d'intérêt.

La proposition précédente sera notamment appliquée dans le cas où les variable aléatoire réelle sont **entières** i.e. prennent leurs valeurs dans \mathbb{N} ou \mathbb{Z} .

Proposition 2.11.

Toute variable aléatoire réelle X à valeurs dans \mathbb{N} , resp. \mathbb{Z} , est discrète. Sa loi s'écrit alors

$$\mathbb{P}_X = \sum_{k \in \mathbb{N}} \mathbb{P}(X = k) \delta_k, \quad \text{resp.} \quad \mathbb{P}_X = \sum_{k \in \mathbb{Z}} \mathbb{P}(X = k) \delta_k.$$

Démonstration : Prendre $D = \mathbb{N}$ (resp. $D = \mathbb{Z}$). \square

Exercice 2.6. (Corrigé de l'exercice : page [154](#))

Soit X une variable aléatoire dont la fonction de répartition F est définie pour tout nombre réel x , par $F(x) = 0$ si $x < 1$ et, si $x \geq 1$, par $F(x) = 1 - \frac{1}{n(n+1)}$ où n est l'unique entier strictement positif (dépendant de x) tel que $n \leq x < n+1$.

1. Donner une représentation graphique de la fonction F et expliquer brièvement pourquoi c'est bien une fonction de répartition.
2. Calculer, pour tout entier naturel n , $\mathbb{P}(X = n)$.
3. Calculer l'espérance de la variable aléatoire X .
4. Que peut-on dire de la variance de la variable aléatoire X ?

Travail conseillé : Étudier dans [\[11\]](#), pages 147 à 171, l'interprétation probabiliste à l'aide de tirages dans une urne, des variable aléatoire réelle de lois de Bernoulli, binomiale, géométrique, de Pascal, binomiale-négative, hypergéométrique.

Chapitre 3

Moments d'un vecteur aléatoire

3.1 Rappels sur l'intégration des applications mesurables

(E, \mathcal{A}, μ) désigne un espace mesuré quelconque.

3.1.1 Intégration des fonctions positives

Notons $\mathcal{M}^+(E, \mathcal{A})$ l'ensemble des applications \mathcal{A} -mesurables (positives) d'un ensemble E dans $[0, +\infty]$.

La première proposition de ce chapitre est fondamentale pour la suite. Elle affirme l'existence et l'unicité d'un opérateur d'intégration, qu'on notera \mathbb{E}_μ , défini sur $\mathcal{M}^+(E, \mathcal{A})$.

Cet opérateur est construit d'après le procédé suivant :

Dans un premier temps, on définit cet opérateur sur l'ensemble \mathcal{E}^+ des applications \mathcal{A} -mesurables étagées et positives de E dans \mathbb{R}^+ . Pour cela, si $\varphi \in \mathcal{E}^+$ on considère sa

décomposition canonique $\varphi = \sum_{k=1}^n \alpha_k \mathbf{1}_{A_k}$ où $\alpha_1 \geq 0, \alpha_2 \geq 0, \dots, \alpha_n \geq 0$ sont les valeurs deux

à deux distinctes de φ et, pour tout entier k vérifiant $1 \leq k \leq n$, $A_k := \{x \in E / \varphi(x) = \alpha_k\}$.

Remarquons que $A_k \in \mathcal{A}$. On pose alors, avec la convention $0 \times (+\infty) := 0$,

$$\mathbb{E}_\mu(\varphi) := \sum_{k=1}^n \alpha_k \mu(A_k).$$

On remarquera que, par sa définition, $\mathbb{E}_\mu(\varphi)$ est un nombre positif éventuellement infini (par exemple si un des $\mu(A_k)$ est infini avec $\alpha_k > 0$) c-à-d $\mathbb{E}_\mu(\varphi) \in [0, +\infty]$. Dans un deuxième temps, on prolonge cet opérateur aux applications de $\mathcal{M}^+(E, \mathcal{A})$ en posant, pour tout $f \in \mathcal{M}^+(E, \mathcal{A})$,

$$\mathbb{E}_\mu(f) := \sup \{ \mathbb{E}_\mu(\varphi) / \varphi \in \mathcal{E}^+ \text{ et } \varphi \leq f \}.$$

Pour une démonstration détaillée, on se reportera à [8], pages 79 à 85. On remarquera encore que, par sa définition, $\mathbb{E}_\mu(f)$ est un nombre positif éventuellement infini c-à-d $\mathbb{E}_\mu(f) \in [0, +\infty]$.

Exemples 3.1.

Considérons $E = \mathbb{R}$ et $\mathcal{A} = \mathcal{B}(\mathbb{R})$ et μ la mesure de Lebesgue sur \mathbb{R} i.e. $\mu = \lambda$. Soit $\varphi = \sum_{k=1}^n \alpha_k \mathbb{1}_{]a_{k-1}, a_k]}$ où $a_0 < a_1 < a_2 < \dots < a_n$ est une suite strictement croissante de $n+1$ réels, les réels $\alpha_1, \alpha_2, \dots, \alpha_n$ n'étant pas nécessairement deux à deux distincts. On dit que φ est une **fonction en escalier** sur \mathbb{R} . Alors

$$\mathbb{E}_\lambda(\varphi) := \sum_{k=1}^n \alpha_k \lambda(]a_{k-1}, a_k]) = \sum_{k=1}^n \alpha_k (a_k - a_{k-1}).$$

Dans ce cas, $\mathbb{E}_\lambda(\varphi)$ représente la mesure de l'aire située sous la courbe représentative de φ . \square

On notera qu'une fonction en escalier est borélienne et étagée sur \mathbb{R} mais que, par exemple, $\mathbb{1}_{\mathbb{Q}}$ est borélienne et étagée sur \mathbb{R} sans être en escalier.

La proposition suivante caractérise l'opérateur \mathbb{E}_μ par trois propriétés fondamentales (pour la démonstration voir [8], page 84).

Proposition 3.1.

Théorème fondamental de l'intégration par rapport à une mesure (admis)

1. Si μ est une mesure sur (E, \mathcal{A}) , il existe une application notée \mathbb{E}_μ , et une seule, de $\mathcal{M}^+(E, \mathcal{A})$ dans $[0, +\infty]$ possédant les trois propriétés suivantes :
 - (a) Pour tout $A \in \mathcal{A}$, $\mathbb{E}_\mu(\mathbb{1}_A) = \mu(A)$.
 - (b) Pour tous f et g appartenant à $\mathcal{M}^+(E, \mathcal{A})$ et tout réel $\alpha \geq 0$.

$$\mathbb{E}_\mu(f + g) = \mathbb{E}_\mu(f) + \mathbb{E}_\mu(g) \text{ et } \mathbb{E}_\mu(\alpha f) = \alpha \mathbb{E}_\mu(f),$$

avec la convention $0 \times (+\infty) := 0$.

(c) Propriété de convergence monotone de Beppo-Lévi

Pour toute suite croissante $(f_n)_{n \in \mathbb{N}}$ d'éléments de $\mathcal{M}^+(E, \mathcal{A})$,

$$\lim_{n \rightarrow +\infty} \mathbb{E}_\mu(f_n) = \mathbb{E}_\mu \left(\lim_{n \rightarrow +\infty} f_n \right).$$

2. Soient f et g deux éléments de $\mathcal{M}^+(E, \mathcal{A})$. Si $f \leq g$, alors $\mathbb{E}_\mu(f) \leq \mathbb{E}_\mu(g)$.

On notera bien qu'on peut avoir $\mathbb{E}_\mu(f) = +\infty$ et qu'on ne parle dans cette proposition que d'applications mesurables et positives. Celles de signe quelconque seront considérées plus loin.

On trouve, suivant les ouvrages ou les usages, différentes notations pour $\mathbb{E}_\mu(f)$:

$$\mathbb{E}_\mu(f) = \int_E f d\mu = \int_E f(x) d\mu(x) = \int_E f(x) \mu(dx).$$

\mathbb{E}_μ s'appelle l'**opérateur d'intégration sur E suivant μ** . $\mathbb{E}_\mu(f)$ s'appelle l'**intégrale de f sur E suivant μ** .

La proposition précédente est un théorème d'existence et d'unicité mais ne permet pas d'explicitier directement le nombre $\mathbb{E}_\mu(f)$ si ce n'est dans des cas simples. Par exemple :

Exemples 3.2.

1) Si f est l'application $f = \sum_{i=1}^n a_i \mathbb{1}_{A_i}$, où (A_1, A_2, \dots, A_n) est une famille finie d'éléments de \mathcal{A} et (a_1, a_2, \dots, a_n) une famille de réels positifs, alors

$$\mathbb{E}_\mu(f) = \sum_{i=1}^n a_i \mu(A_i).$$

2) Si $\mu := \delta_0 + \delta_5 + \lambda$ et $f := \pi \mathbb{1}_{[0, \frac{1}{3}]} + \mathbb{1}_{[6, 10]} + 3 \mathbb{1}_{\{5\}}$, alors $\int_{\mathbb{R}} f d\mu = \frac{4}{3}\pi + 7$. \square

Exercice 3.1. (Corrigé de l'exercice : page 155)

Vérifier les affirmations de l'exemple précédent.

Les propositions admises suivantes donnent quelques "**règles d'intégration**" suivant la mesure considérée. Ces règles seront suffisantes pour la suite et seront constamment utilisées. Elles diffèrent bien sûr en fonction des mesures utilisées. Commençons par le cas de la mesure de Lebesgue sur \mathbb{R} . Le cas de la mesure de Lebesgue sur \mathbb{R}^d avec $d \geq 2$ sera traité au chapitre suivant.

Proposition 3.2.

Cas de la mesure de Lebesgue sur \mathbb{R} pour les fonctions positives (admis)

On suppose $E := \mathbb{R}$, $\mathcal{A} := \mathcal{B}(\mathbb{R})$, $\mu := \lambda$ où λ désigne la mesure de Lebesgue sur \mathbb{R} .

Si f est une application borélienne de \mathbb{R} dans $[0, +\infty]$ intégrable au sens de Riemann sur tout intervalle fermé borné de \mathbb{R} , alors son intégrale sur \mathbb{R} suivant λ est égale à son intégrale généralisée au sens de Riemann *c-à-d*

$$\mathbb{E}_\mu(f) = \int_{\mathbb{R}} f(t) d\lambda(t) = \int_{-\infty}^{+\infty} f(t) dt.$$

Exemples 3.3.

$$\int_{\mathbb{R}} e^{-x} \mathbb{1}_{[0, +\infty[}(x) d\lambda(x) = \int_0^{+\infty} e^{-x} dx = 1 \text{ et } \int_{\mathbb{R}} x^2 \mathbb{1}_{[0, +\infty[}(x) d\lambda(x) = +\infty.$$

Proposition 3.3.

Cas de la mesure de Dirac sur \mathbb{R}^d (admis)

On suppose $E := \mathbb{R}^d$, $\mathcal{A} := \mathcal{B}(\mathbb{R}^d)$, $\mu := \delta_a$ où $a \in \mathbb{R}^d$.

Si f est une application borélienne de \mathbb{R}^d dans $[0, +\infty]$, alors

$$\mathbb{E}_\mu(f) = \int_{\mathbb{R}^d} f(t) d\mu(t) = f(a).$$

La proposition qui suit généralise la précédente :

Proposition 3.4.**Cas des mesures discrètes sur \mathbb{R}^d (admis)**

On suppose $E := \mathbb{R}^d$, $\mathcal{A} := \mathcal{B}(\mathbb{R}^d)$, $\mu := \sum_{i=0}^{+\infty} \alpha_i \delta_{a_i}$ où $(a_k)_{\mathbb{N}}$ est une suite de vecteurs de \mathbb{R}^d et $(\alpha_k)_{\mathbb{N}}$ une suite de réels positifs ou nuls.
Si f est une application borélienne de \mathbb{R}^d dans $[0, +\infty]$, alors

$$\mathbb{E}_{\mu}(f) = \int_{\mathbb{R}^d} f(t) d\mu(t) = \sum_{i=0}^{+\infty} \alpha_i f(a_i).$$

Exemples 3.4.

Soient $\mu = \mathcal{P}(\alpha) := \sum_{k=0}^{+\infty} e^{-\alpha} \frac{\alpha^k}{k!} \delta_k$ la probabilité de Poisson sur \mathbb{R} où $\alpha > 0$.

i) Si f est une application borélienne de \mathbb{R} dans $[0, +\infty]$, alors

$$\mathbb{E}_{\mu}(f) = \sum_{k=0}^{+\infty} e^{-\alpha} \frac{\alpha^k}{k!} f(k)$$

ii)

$$\int_{\mathbb{R}} x(x-1) \mathbb{1}_{[1, +\infty[}(x) d\mu(x) = \sum_{k=2}^{+\infty} e^{-\alpha} \frac{\alpha^k}{k!} k(k-1) = \alpha^2.$$

Exemples 3.5.

Soit $(u_n)_{\mathbb{N}}$ une suite de réels positifs ou nuls.

Considérons l'application $f := \sum_{k=0}^{+\infty} u_k \mathbb{1}_{\{k\}}$ et la mesure $\mu := \sum_{i=0}^{+\infty} \delta_i$. On vérifie aisément que

$$\mathbb{E}_{\mu}(f) = \sum_{k=0}^{+\infty} u_k. \square$$

Ce dernier exemple montre que la théorie des séries à termes réels positifs peut être considérée comme une théorie de l'intégration suivant la mesure sur \mathbb{R} dite **de dénombrement**

$\mu := \sum_{i=0}^{+\infty} \delta_i$. La théorie de l'intégration permet ainsi d'unifier dans un même formalisme l'étude des probabilités discrètes, qui font intervenir des séries dans les calculs, et celle des probabilités à densité où pratiquement interviennent des intégrales de Riemann classiques.

La proposition suivante ramène le calcul d'intégrales suivant les mesures à densité au calcul d'une intégrale de Lebesgue sur \mathbb{R} qu'on effectue alors par application de la proposition 3.2.

Proposition 3.5.**Cas des mesures à densité sur \mathbb{R}** (admis)*On suppose $E := \mathbb{R}$, $\mathcal{A} := \mathcal{B}(\mathbb{R})$, μ une mesure admettant une densité ρ sur \mathbb{R} .**Si f est une application borélienne de \mathbb{R} dans $[0, +\infty]$, alors,*

$$\mathbb{E}_\mu(f) = \int_{\mathbb{R}} f(t) d\mu(t) = \int_{\mathbb{R}} f(t) \rho(t) d\lambda(t) = \mathbb{E}_\lambda(f\rho).$$

Exemples 3.6.

Soit $\mu := \mathcal{N}_1(0, 1)$, $\int_{\mathbb{R}} x^2 d\mu(x) = \int_{\mathbb{R}} x^2 \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} d\lambda(x)$. On est ramené au calcul d'une intégrale suivant la mesure de Lebesgue. D'où

$$\int_{\mathbb{R}} x^2 \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} d\lambda(x) = \int_{-\infty}^{+\infty} x^2 \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} dx = 1,$$

$$\text{c-à-d } \int_{\mathbb{R}} x^2 d\mu(x) = 1. \quad \square$$

Exercice 3.2. (Corrigé de l'exercice : page 155)

1. Soient μ et ν deux mesures sur (E, \mathcal{A}) . A l'aide du théorème fondamental de l'intégration montrer que, pour tout $f \in \mathcal{M}^+(E, \mathcal{A})$, $\mathbb{E}_{\mu+\nu}(f) = \mathbb{E}_\mu(f) + \mathbb{E}_\nu(f)$.
2. Soient $\alpha > 0$, μ la mesure sur \mathbb{R} de densité ρ définie par $\rho(x) := \alpha e^{-\alpha x} \mathbb{1}_{[0,1]}(x)$ et $\nu := e^{-\alpha} \delta_1$. Montrer que $\int_{\mathbb{R}} e^{\alpha x} d(\mu + \nu)(x) = \alpha + 1$.
3. Soient $\alpha > 0$, μ la mesure sur \mathbb{R} de densité ρ définie par $\rho(x) := e^{-\alpha x} \mathbb{1}_{[-1,1]}(x)$ et $\nu := \sum_{k=0}^{+\infty} \frac{\alpha^k}{k!} \delta_k$. Calculer $\int_{\mathbb{R}} e^{\alpha x} d(\mu + \nu)(x)$ et $\int_{\mathbb{R}} \mathbb{1}_{\mathbb{R}} d(\mu + \nu)$.

La mesure $\mu + \nu$ est-elle une probabilité ?

3.1.2 Intégration des fonctions numériques

Soit f une application \mathcal{A} -mesurable de E dans $[-\infty, +\infty]$. Les applications $f^+ := \sup(f, 0)$ et $f^- := \sup(-f, 0)$ sont des applications \mathcal{A} -mesurables de E dans $[0, +\infty]$. D'après ce qui précède les quantités $\mathbb{E}_\mu(f^+)$ et $\mathbb{E}_\mu(f^-)$ sont des éléments de $[0, +\infty]$ éventuellement infinies. La différence $\mathbb{E}_\mu(f^+) - \mathbb{E}_\mu(f^-)$ aura un sens si $\mathbb{E}_\mu(f^+)$ et $\mathbb{E}_\mu(f^-)$ sont toutes les deux finies. On pourra alors poser $\mathbb{E}_\mu(f) := \mathbb{E}_\mu(f^+) - \mathbb{E}_\mu(f^-)$, d'où les définitions :

Définition 3.1.

Une application f de E dans $[-\infty, +\infty]$ est dite **intégrable sur E suivant μ** ou plus simplement **μ -intégrable** si elle est \mathcal{A} -mesurable et si les quantités $\mathbb{E}_\mu(f^+)$ et $\mathbb{E}_\mu(f^-)$ sont toutes les deux finies. Dans ce cas on appelle **intégrale de f sur E suivant μ** le réel $\mathbb{E}_\mu(f) := \mathbb{E}_\mu(f^+) - \mathbb{E}_\mu(f^-)$.

On remarquera que $\mathbb{E}_\mu(f) \in \mathbb{R}$.

On utilise aussi les mêmes notations que celles déjà vues dans le cas des fonctions positives pour $\mathbb{E}_\mu(f)$.

Proposition 3.6.

Soit f une application \mathcal{A} -mesurable de E dans $[-\infty, +\infty]$. Alors f est intégrable si, et seulement si, $\mathbb{E}_\mu(|f|)$ est fini.

Démonstration : $|f| = f^+ + f^-$ est une application \mathcal{A} -mesurable positive et, d'après l'item 1 de la proposition 3.1, $\mathbb{E}_\mu(|f|) = \mathbb{E}_\mu(f^+) + \mathbb{E}_\mu(f^-)$. Par suite, f est intégrable si, et seulement si, $\mathbb{E}_\mu(f^+)$ et $\mathbb{E}_\mu(f^-)$ sont toutes les deux finies c-à-d si, et seulement si, $\mathbb{E}_\mu(|f|) = \mathbb{E}_\mu(f^+) + \mathbb{E}_\mu(f^-)$ est fini. \square

Exemples 3.7.

1) Soit $\mu := \mathcal{N}_1(0, 1)$, $\int_{\mathbb{R}} x d\mu(x) = 0$.

En effet, $f^+(x) = x\mathbb{1}_{[0, +\infty[}(x)$ et $f^-(x) = -x\mathbb{1}_{]-\infty, 0]}(x)$. On vérifie les deux suites d'égalités

$$\begin{aligned} \int_{\mathbb{R}} f^+(x) d\mu(x) &= \int_{\mathbb{R}} x\mathbb{1}_{[0, +\infty[}(x) d\mu(x) = \frac{1}{\sqrt{2\pi}} \int_0^{+\infty} te^{-\frac{1}{2}t^2} dt < +\infty \\ \text{et } \int_{\mathbb{R}} f^-(x) d\mu(x) &= \int_{\mathbb{R}} -x\mathbb{1}_{]-\infty, 0]}(x) d\mu(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^0 -te^{-\frac{1}{2}t^2} dt \\ &= \frac{1}{\sqrt{2\pi}} \int_0^{+\infty} te^{-\frac{1}{2}t^2} dt < +\infty, \end{aligned}$$

en vertu de la convergence de la dernière intégrale (généralisée au sens de Riemann).

D'après la définition de l'intégrale $\int_{\mathbb{R}} x d\mu(x) := \int_{\mathbb{R}} f^+(x) d\mu(x) - \int_{\mathbb{R}} f^-(x) d\mu(x) = 0$.

2) Soit $\mu := \delta_a$ où $a \in \mathbb{R}$. Les applications boréliennes de \mathbb{R} dans $[-\infty, +\infty]$ intégrables suivant δ_a sont celles qui prennent une valeur finie au point a .

En effet, f est δ_a -intégrable si, et seulement si, $\mathbb{E}_{\delta_a}(|f|) = |f(a)| < +\infty$ c-à-d si, et seulement si, $f(a) \in \mathbb{R}$. \square

Exercice 3.3. (Corrigé de l'exercice : page 156)

Soit $\mu := \sum_{i=0}^{+\infty} \alpha_i \delta_{a_i}$ où, pour tout $i \in \mathbb{N}$, $a_i \in \mathbb{R}^d$ et $(\alpha_k)_{\mathbb{N}}$ est une suite de réels positifs.

Montrer que les applications boréliennes f de \mathbb{R}^d dans $[-\infty, +\infty]$ intégrables suivant μ sont celles pour lesquelles la série numérique $\sum_{i=0}^{+\infty} \alpha_i f(a_i)$ est absolument convergente.

Les règles d'intégration des fonctions de signe quelconque intégrables sont les mêmes que celles pour les fonctions positives vues dans le cas des mesures discrètes ou à densité. On peut démontrer cela en écrivant les fonctions comme différence de leur partie positive et de leur partie négative. Par contre dans le cas de la mesure de Lebesgue sur \mathbb{R} la proposition 3.2 devient fausse pour les fonctions qui ne sont pas de signe constant. Dans ce cas on utilise si possible la proposition suivante :

Proposition 3.7.

Cas de la mesure de Lebesgue sur \mathbb{R} pour les fonctions réelles (admis)

On suppose $E := \mathbb{R}$, $\mathcal{A} := \mathcal{B}(\mathbb{R})$, $\mu := \lambda$ où λ désigne la mesure de Lebesgue sur \mathbb{R} .

1. Si f est une application borélienne de \mathbb{R} dans \mathbb{R} nulle en dehors d'un intervalle fermé borné $[a, b]$ et intégrable au sens de Riemann sur $[a, b]$, alors son intégrale sur \mathbb{R} suivant λ est égale à son intégrale au sens de Riemann sur $[a, b]$, c-à-d

$$\mathbb{E}_\mu(f) = \int_{\mathbb{R}} f(t) d\lambda(t) = \int_a^b f(t) dt.$$

2. Si f est une application borélienne de \mathbb{R} dans \mathbb{R} intégrable au sens de Riemann sur tout intervalle fermé borné de \mathbb{R} et telle que $\int_{-\infty}^{+\infty} |f(t)| dt < +\infty$, alors son intégrale sur \mathbb{R} suivant λ est égale à son intégrale généralisée au sens de Riemann c-à-d

$$\mathbb{E}_\mu(f) = \int_{\mathbb{R}} f(t) d\lambda(t) = \int_{-\infty}^{+\infty} f(t) dt.$$

Le théorème précédent s'applique en particulier lorsque f est continue ou monotone sur un nombre fini d'intervalles de \mathbb{R} et nulle en dehors de la réunion de ces intervalles.

Nous sommes en mesure maintenant d'étendre la définition des probabilités à densité, considérées jusqu'à présent uniquement sur \mathbb{R} , au cas des probabilités sur les espaces \mathbb{R}^d avec $d \geq 1$. Cette définition sera en particulier utile dans les chapitres IV et V.

Définition 3.2.

On appelle **densité de probabilité sur \mathbb{R}^d** toute application borélienne positive ρ de \mathbb{R}^d dans $[0, +\infty]$ vérifiant

$$\int_{\mathbb{R}^d} \rho d\lambda^{(d)} = 1.$$

La proposition suivante (admise) montre que la règle d'intégration suivant une mesure de probabilité à densité est tout à fait analogue à celle déjà vue pour les fonctions positives.

Proposition 3.8.

Soit ρ une densité de probabilité sur \mathbb{R}^d .

1. L'application

$$\nu : A \in \mathcal{B}(\mathbb{R}^d) \mapsto \nu(A) := \int_{\mathbb{R}^d} \mathbb{1}_A \rho d\lambda^{(d)} \in [0, 1]$$

est une probabilité sur \mathbb{R}^d .

On dit que ν **admet ρ pour densité sur \mathbb{R}^d** et on note $\nu = \rho \cdot \lambda^{(d)}$.

2. Si f est une application de \mathbb{R}^d dans $\overline{\mathbb{R}}$ borélienne positive, resp. intégrable suivant ν , alors l'application $f\rho$ est borélienne positive, resp. intégrable suivant $\lambda^{(d)}$, et

$$\mathbb{E}_\nu(f) = \int_{\mathbb{R}^d} f(t) d\nu(t) = \int_{\mathbb{R}^d} f(t) \rho(t) d\lambda^{(d)}(t).$$

Exercice 3.4. (Corrigé de l'exercice : page 166)

Démontrer la première partie de la proposition précédente.

On remarquera que les définitions 1.15, page 10, de densité et de probabilité à densité introduites au chapitre I sont bien des cas particuliers de la définition donnée ci-dessus.

3.1.3 Intégration des fonctions vectorielles

Soit f une application \mathcal{A} -mesurable de E dans \mathbb{R}^d . Dans la base canonique de \mathbb{R}^d , f admet les composantes f_1, f_2, \dots, f_d , où, pour tout entier $k \leq d$, f_k est une application \mathcal{A} -mesurable de E dans \mathbb{R} . On écrira $f := (f_1, f_2, \dots, f_d)$.

Définition 3.3.

On dit que f est **intégrable sur E suivant μ** si toutes les applications-composantes f_1, f_2, \dots, f_d sont intégrables sur E suivant μ . Dans ce cas on appelle **intégrale de f sur E suivant μ** le vecteur de \mathbb{R}^d de composantes dans la base canonique $\mathbb{E}_\mu(f_1), \mathbb{E}_\mu(f_2), \dots, \mathbb{E}_\mu(f_d)$, et on note $\mathbb{E}_\mu(f) := (\mathbb{E}_\mu(f_1), \mathbb{E}_\mu(f_2), \dots, \mathbb{E}_\mu(f_d))$.

Exercice 3.5. (Corrigé de l'exercice : page 156)

Montrer que $f := (f_1, f_2, \dots, f_d)$ est intégrable suivant μ si, et seulement si, $\mathbb{E}_\mu(|f|) < +\infty$ où $|\cdot|$ désigne la norme usuelle de \mathbb{R}^d .

Un cas particulier intéressant pour la suite est le cas où f est à valeurs dans le plan complexe \mathbb{C} qu'on identifie à \mathbb{R}^2 . On écrit alors $f := f_1 + if_2$ identifié à $f := (f_1, f_2)$ et on pose $\mathbb{E}_\mu(f) := \mathbb{E}_\mu(f_1) + i\mathbb{E}_\mu(f_2)$.

A titre d'exemple, développons une application de la notion de fonction vectorielle μ -intégrable. Soient $E := \mathbb{R}^n$, μ une probabilité sur \mathbb{R}^n et f l'application définie sur \mathbb{R}^n à valeurs dans \mathbb{C} , par $f(x) := \exp(i\langle x, u \rangle)$ où u est un vecteur fixé de \mathbb{R}^n , $\langle \cdot, \cdot \rangle$ et $|\cdot|$ les produit scalaire et norme usuels de \mathbb{R}^n .

Alors $f_1(x) := \cos\langle x, u \rangle$ et $f_2(x) := \sin\langle x, u \rangle$. D'où, pour $k = 1$ ou 2 ,

$$\mathbb{E}_\mu(|f_k|) = \int_{\mathbb{R}^n} |f_k| d\mu \leq \int_{\mathbb{R}^n} \mathbf{1}_{\mathbb{R}^n} d\mu = \mu(\mathbb{R}^n) = 1.$$

f_1 et f_2 sont donc μ -intégrables, par définition il en est de même de f . On peut donc définir $\mathbb{E}_\mu(f) \in \mathbb{C}$ pour tout vecteur u de \mathbb{R}^n .

Définition 3.4.

L'application

$$\Phi_\mu : u \in \mathbb{R}^n \mapsto \Phi_\mu(u) := \int_{\mathbb{R}^n} \exp(i\langle x, u \rangle) d\mu(x)$$

s'appelle la **fonction caractéristique de μ** , en abrégé **f.c.**. Si X est un vecteur aléatoire, on appelle **fonction caractéristique de X** , et on note Φ_X , la f.c. de la loi de X .

Exemples 3.8.

Si μ est la probabilité de Bernoulli de paramètres p .

$\Phi_\mu(t) = \int_{\mathbb{R}} e^{itx} d(p\delta_1 + (1-p)\delta_0)(x) = pe^{it} + (1-p)$ d'après les règles d'intégration par rapport à une mesure de Dirac.

Exercice 3.6. (Corrigé de l'exercice : page 156)

Expliciter les fonctions caractéristiques des probabilités de Dirac, binomiale, de Poisson.

On montrera dans l'exercice 3.20, page 60, que la fonction caractéristique de la probabilité normale $\mathcal{N}_1(m, \sigma^2)$ est définie sur \mathbb{R} , pour tout réel t , par $\Phi(t) := \exp(imt - \frac{1}{2}t^2\sigma^2)$.

Plus généralement, on trouvera la liste des fonctions caractéristiques des probabilités usuelles sur \mathbb{R} dans le formulaire donné dans l'annexe A, page 205, de ce cours.

3.1.4 Propriétés de l'intégrale

L'intégrale d'une fonction suivant une mesure μ possède toutes les propriétés des intégrales classiques vues en premier cycle universitaire, on admettra :

Proposition 3.9.

Soient f et g deux applications de E dans \mathbb{R}^d intégrables suivant μ , a et b deux réels, alors

1. $\mathbb{E}_\mu(af + bg) = a\mathbb{E}_\mu(f) + b\mathbb{E}_\mu(g)$.
2. $|\mathbb{E}_\mu(f)| \leq \mathbb{E}_\mu(|f|)$ où $|\cdot|$ est la norme euclidienne sur \mathbb{R}^d .
3. Si de plus $d = 1$ et $f \leq g$, alors $\mathbb{E}_\mu(f) \leq \mathbb{E}_\mu(g)$.

Exemples 3.9.

Comme l'indicatrice de $A_1 \cup A_2 \cup \dots \cup A_n$, où A_1, A_2, \dots, A_n , sont des parties de E , est donnée par

$$\mathbb{1}_{A_1 \cup A_2 \cup \dots \cup A_n} = 1 - (1 - \mathbb{1}_{A_1})(1 - \mathbb{1}_{A_2}) \cdots (1 - \mathbb{1}_{A_n}),$$

en développant le second membre de cette égalité et en utilisant les propriétés 1)a du théorème fondamental de l'intégration et 1) de la proposition précédente, on obtient aisément une autre démonstration de la formule de Poincaré 1.8, page 12, énoncée dans le premier chapitre.

Les énoncés de théorèmes permettant d'intervertir les symboles \mathbb{E}_μ et $\lim_{n \rightarrow +\infty}$ sont particulièrement simples dans cette théorie de l'intégration. Commençons par rappeler (cf. [8], page 82) :

Proposition 3.10.

Théorème de convergence monotone de Beppo-Lévi (admis)

Pour toute suite croissante $(f_n)_{\mathbb{N}}$ d'applications \mathcal{A} -mesurables positives,

$$\mathbb{E}_\mu \left(\lim_{n \rightarrow +\infty} f_n \right) = \lim_{n \rightarrow +\infty} \mathbb{E}_\mu(f_n).$$

Ce résultat a pour corollaire :

Proposition 3.11.

Théorème d'interversion de \mathbb{E}_μ et \sum_0^∞

Pour toute suite $(f_n)_\mathbb{N}$ d'applications \mathcal{A} -mesurables positives,

$$\mathbb{E}_\mu \left(\sum_{n=0}^{\infty} f_n \right) = \sum_{n=0}^{\infty} \mathbb{E}_\mu(f_n).$$

Démonstration : Posons, pour tout entier n , $g_n := \sum_{k=0}^n f_k$. On applique alors le théorème de Beppo-Lévi à la suite croissante d'applications \mathcal{A} -mesurables positives $(g_n)_\mathbb{N}$. \square

On en déduit aussi la proposition suivante (cf. [8], page 83) :

Proposition 3.12.

Lemme de Fatou

Si $(f_n)_\mathbb{N}$ est une suite d'applications \mathcal{A} -mesurables positives, alors

$$\mathbb{E}_\mu \left(\liminf_{n \rightarrow +\infty} f_n \right) \leq \liminf_{n \rightarrow +\infty} \mathbb{E}_\mu(f_n).$$

Notons que ces trois résultats précédents sont faux si les fonctions f_n ne sont plus supposées positives.

Terminons par un théorème valable pour les fonctions (à valeurs réelles) de signe quelconque à la condition d'être intégrables (cf. [8], page 102). Ce théorème ainsi que celui de Beppo-Lévi sont des théorèmes fondamentaux de la théorie de l'intégration. Ce sont principalement ces résultats qui font la supériorité de la théorie de Lebesgue sur celle de Riemann vue en premier cycle universitaire.

Proposition 3.13.

Théorème de convergence dominée de Lebesgue (admis)

Si $(f_n)_\mathbb{N}$ est une suite d'applications \mathcal{A} -mesurables convergeant presque-partout vers une application \mathcal{A} -mesurable f et s'il existe une application intégrable φ telle que, pour tout $k \in \mathbb{N}$, $|f_k| \leq \varphi$, alors f est intégrable et $\mathbb{E}_\mu(f) = \lim_{n \rightarrow +\infty} \mathbb{E}_\mu(f_n)$.

3.1.5 Espaces de Lebesgue d'ordre p

Pour les démonstrations et énoncés plus généraux de ce paragraphe, on pourra se reporter à [8], pages 149 à 156.

Soit (E, \mathcal{A}, μ) un espace de probabilité. Soit $1 \leq p < \infty$, si f est une application de E dans $\overline{\mathbb{R}}$, \mathcal{A} -mesurable, définie et finie μ -presque sûrement, on posera $\|f\|_p = [\mathbb{E}(|f|^p)]^{\frac{1}{p}}$ si l'application

$|f|^p$ est intégrable pour μ , et $\|f\|_p = +\infty$ sinon.

On note $\mathcal{L}^p(E, \mathcal{A}, \mu)$ l'ensemble des applications f de E dans $\overline{\mathbb{R}}$, \mathcal{A} -mesurables, définies et finies μ -presque sûrement, telles que $\|f\|_p < +\infty$. $\mathcal{L}^p(E, \mathcal{A}, \mu)$ est un espace vectoriel, et $\|f\|_p$ est une semi-norme sur cet espace. On note alors $L^p(E, \mathcal{A}, \mu)$ l'espace des classes d'équivalence des applications \mathcal{A} -mesurables de $\mathcal{L}^p(E, \mathcal{A}, \mu)$ pour l'égalité μ -presque-sûre. En particulier, l'espace $L^1(E, \mathcal{A}, \mu)$ est constitué des classes d'équivalence des applications de E dans $\overline{\mathbb{R}}$, \mathcal{A} -mesurables, définies et finies μ -presque sûrement, intégrables pour μ , égales μ -presque sûrement.

On définit également $\|f\|_\infty = \sup \{ \alpha \in \mathbb{R} / \mu[|f| \geq \alpha] > 0 \}$, et $L^\infty(E, \mathcal{A}, \mu)$ comme l'ensemble des classes d'équivalence, pour l'égalité μ -presque-sûre, des applications de E dans $\overline{\mathbb{R}}$, \mathcal{A} -mesurables, définies et finies μ -presque sûrement, telles que $\|f\|_\infty < +\infty$.

Définition 3.5.

Les espaces $L^p(E, \mathcal{A}, \mu)$ pour tout réel p tel que $1 \leq p \leq \infty$ sont appelés les espaces de Lebesgue d'ordre p .

Moyennant ces définitions, on rappelle le résultat suivant sur les espaces de Lebesgue :

Proposition 3.14.

Soit (E, \mathcal{A}, μ) un espace de probabilité,

1. Pour tout réel p tel que $1 \leq p \leq \infty$, la semi-norme $\|\cdot\|_p$ induit une norme sur l'espace de Lebesgue $L^p(E, \mathcal{A}, \mu)$, encore notée $\|\cdot\|_p$.
2. Pour tout réel p tel que $1 \leq p \leq \infty$, l'espace de Lebesgue $L^p(E, \mathcal{A}, \mu)$ muni de la norme $\|\cdot\|_p$ est un espace de Banach.
3. Pour tout p et q tels que $1 \leq p < q \leq \infty$, on a la suite des inégalités

$$\|\cdot\|_1 \leq \|\cdot\|_p < \|\cdot\|_q \leq \|\cdot\|_\infty,$$

et la suite d'inclusions

$$L^\infty(E, \mathcal{A}, \mu) \subseteq L^q(E, \mathcal{A}, \mu) \subseteq L^p(E, \mathcal{A}, \mu) \subseteq L^1(E, \mathcal{A}, \mu).$$

4. Si p et q sont tels que $1 \leq p, q \leq \infty$ avec $\frac{1}{p} + \frac{1}{q} = 1$, alors, pour tout $f \in L^p(E, \mathcal{A}, \mu)$ et $g \in L^q(E, \mathcal{A}, \mu)$, l'application $fg \in L^1(E, \mathcal{A}, \mu)$ et on a $\|fg\|_1 \leq \|f\|_p \|g\|_q$.

Remarque : L'inégalité de l'item 4 s'appelle l'**inégalité de Hölder**. Pour le cas particulier, $p = q = 2$, cette inégalité se ramène à l'inégalité de Schwarz car l'espace $L^2(E, \mathcal{A}, \mu)$, muni du produit scalaire $\langle f, g \rangle = \mathbb{E}(fg)$ est un espace de Hilbert.

L'item 3 de la proposition précédente n'est plus vrai dès que μ n'est plus une mesure finie.

Désormais dans la suite du cours, les fonctions utilisées seront souvent définies seulement presque-partout. Nous écrirons par abus, $f \in L^q(E, \mathcal{A}, \mu)$ pour exprimer que f est une application de E dans $\overline{\mathbb{R}}$, \mathcal{A} -mesurable, définie et finie μ -presque sûrement, et que sa classe d'équivalence pour l'égalité presque-partout est dans $L^q(E, \mathcal{A}, \mu)$.

3.2 Théorème du transfert et moments d'une v.a.

$(\Omega, \mathcal{F}, \mathbb{P})$ désigne l'espace de probabilité de base.

Les variables aléatoires utilisées seront souvent définies seulement presque-sûrement. En appliquant aux variables aléatoires la convention d'écriture de la fin du paragraphe précédent, nous écrirons par abus, $X \in L^p(\Omega, \mathcal{F}, \mathbb{P})$ pour exprimer que X est une variable aléatoire définie et finie presque-sûrement sur Ω et que sa classe d'équivalence pour l'égalité presque-sûre est dans $L^p(\Omega, \mathcal{F}, \mathbb{P})$.

Suivant l'usage on notera dorénavant, sauf cas exceptionnels, l'opérateur d'intégration (sur l'espace de probabilité de base) \mathbb{E} au lieu de $\mathbb{E}_{\mathbb{P}}$. On appelle \mathbb{E} **l'espérance mathématique suivant \mathbb{P}** ou, plus simplement s'il n'y a pas de risque de confusion, **espérance**. Ainsi si X est une variable aléatoire positive, resp. vectorielle intégrable, on utilisera indifféremment les notations $\mathbb{E}(X)$ ou $\int_{\Omega} X(\omega) d\mathbb{P}(\omega)$ ou $\int_{\Omega} X d\mathbb{P}$ pour désigner $\mathbb{E}_{\mathbb{P}}(X)$.

Si h est une application de \mathbb{R}^d dans \mathbb{R}^n et X un vecteur aléatoire de dimension d , on rappelle la notation abusive déjà introduite $h(X) := h \circ X$.

3.2.1 Théorème du transfert et identification de lois

Le théorème du transfert est d'un usage constant en probabilité. Donnons-en deux versions, une pour les fonctions positives (c'est la plus utile), l'autre pour les fonctions vectorielles intégrables.

Proposition 3.15.

Théorème du transfert (cas positif)

Soient h une application borélienne positive de \mathbb{R}^d dans $[0, +\infty]$ et X un vecteur aléatoire de dimension d , alors

$$\mathbb{E}[h(X)] = \int_{\mathbb{R}^d} h d\mathbb{P}_X = \mathbb{E}_{\mathbb{P}_X}(h)$$

qu'on écrit également sous la forme :

$$\int_{\Omega} h(\omega) d\mathbb{P}_X(\omega) = \int_{\mathbb{R}^d} h(x) d\mathbb{P}_X(x).$$

Démonstration : Cette proposition se démontre à l'aide de la technique des fonctions étagées. \square

Proposition 3.16.**Théorème du transfert (cas vectoriel)** (admis)

Soient h une application borélienne de \mathbb{R}^d dans \mathbb{R}^n et X un vecteur aléatoire de dimension d . Alors h est intégrable sur \mathbb{R}^d suivant \mathbb{P}_X si, et seulement si, $h(X)$ est intégrable sur Ω suivant \mathbb{P} , et dans ce cas

$$\mathbb{E}[h(X)] = \int_{\mathbb{R}^d} h d\mathbb{P}_X = \mathbb{E}_{\mathbb{P}_X}(h)$$

qu'on écrit également sous la forme :

$$\int_{\Omega} h(\omega) d\mathbb{P}_X(\omega) = \int_{\mathbb{R}^d} h(x) d\mathbb{P}_X(x).$$

Exemples 3.10.

Soient X un vecteur aléatoire de dimension d et Φ_X sa fonction caractéristique. Par application du théorème du transfert (cas vectoriel) on obtient, pour tout élément u de \mathbb{R}^d ,

$$\Phi_X(u) := \int_{\mathbb{R}^d} \exp(i\langle x, u \rangle) d\mathbb{P}_X(x) = \mathbb{E}[\exp(i\langle X, u \rangle)]. \square$$

Exemples 3.11.

Soit X une variable aléatoire réelle de loi $\mathcal{N}_1(0, 1)$ i.e. $\mathbb{P}_X = \mathcal{N}_1(0, 1)$. Calculons $\mathbb{E}(X^2)$. Donnons deux méthodes.

- Première méthode : $\mathbb{E}(X^2)$ est de la forme $\mathbb{E}[h(X)]$ avec $h(t) := t^2$. On applique le théorème du transfert (cas positif), on remarque que h est continue donc borélienne. On doit donc calculer à l'aide d'une intégration par parties,

$$\mathbb{E}(X^2) = \int_{\mathbb{R}} t^2 d\mathbb{P}_X(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} t^2 e^{-\frac{1}{2}t^2} dt = 1.$$

- Deuxième méthode : On a vu au chapitre II dans l'exemple 2.2, page 24, que la variable aléatoire réelle $Y := X^2$ suit la loi $\gamma(\frac{1}{2}, \frac{1}{2})$ de densité

$$\rho(x) := \frac{1}{\sqrt{2\pi x}} e^{-\frac{x}{2}} \mathbb{1}_{]0, +\infty[}(x).$$

On cherche à calculer

$$\begin{aligned} \mathbb{E}(X^2) &= \mathbb{E}(Y) = \int_{\mathbb{R}} t d\mathbb{P}_Y(t) = \int_{\mathbb{R}} x \frac{1}{\sqrt{2\pi x}} e^{-\frac{x}{2}} \mathbb{1}_{]0, +\infty[}(x) d\lambda(x) \\ &= \int_{-\infty}^{+\infty} \sqrt{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{x}{2}} \mathbb{1}_{]0, +\infty[}(x) dx = \int_0^{+\infty} \frac{\sqrt{x}}{\sqrt{2\pi}} e^{-\frac{x}{2}} dx = 1. \end{aligned}$$

Dans ces calculs nous avons utilisé les règles d'intégration suivant une mesure à densité et une mesure de Lebesgue, puis effectué un changement de variable pour calculer l'intégrale généralisée finale. \square

Exercice 3.7. (Corrigé de l'exercice : page 157)

Soit X une variable aléatoire réelle de loi $\mathcal{N}_1(0, 1)$.

a) Calculer de deux façons différentes $\mathbb{E}(e^X)$.

b) Montrer que X^3 est intégrable (suivant la mesure \mathbb{P}) et calculer $\mathbb{E}(X^3)$.

Le théorème du transfert permet d'établir un critère d'identification des lois utilisant les fonctions boréliennes positives :

Proposition 3.17.

Critère des fonctions boréliennes positives

Soient X un vecteur aléatoire de dimension d et μ une probabilité sur \mathbb{R}^d . Alors le vecteur aléatoire X a pour loi μ si, et seulement si, pour toute application borélienne positive h de \mathbb{R}^d dans $[0, +\infty]$,

$$\mathbb{E}[h(X)] = \int_{\mathbb{R}^d} h d\mu,$$

qui peut aussi s'écrire

$$\int_{\Omega} h(X) d\mathbb{P} = \int_{\mathbb{R}^d} h(x) d\mu(x).$$

Démonstration : • C.N. - Si $\mathbb{P}_X = \mu$, d'après le théorème du transfert, pour toute application borélienne positive h de \mathbb{R}^d dans $[0, +\infty]$, $\mathbb{E}[h(X)] = \int_{\mathbb{R}^d} h d\mathbb{P}_X = \int_{\mathbb{R}^d} h d\mu$.

• C.S. - Supposons que, pour toute application borélienne positive h de \mathbb{R}^d dans $[0, +\infty]$, $\mathbb{E}[h(X)] = \int_{\mathbb{R}^d} h d\mu$. Alors, comme pour tout $B \in \mathcal{B}(\mathbb{R}^d)$ $h := \mathbb{1}_B$ est une application borélienne positive de \mathbb{R}^d dans $[0, +\infty]$, par hypothèse d'une part

$$\mathbb{E}[h(X)] = \mathbb{E}[\mathbb{1}_B(X)] = \mathbb{E}_{\mu}(\mathbb{1}_B) = \mu(B),$$

et par le théorème de transfert d'autre part

$$\mathbb{E}[h(X)] = \mathbb{E}[\mathbb{1}_B(X)] = \int_{\mathbb{R}^d} \mathbb{1}_B d\mathbb{P}_X = \mathbb{P}_X(B).$$

D'où, pour tout $B \in \mathcal{B}(\mathbb{R}^d)$, $\mathbb{P}_X(B) = \mu(B)$ ce qui signifie que $\mathbb{P}_X = \mu$. \square

Exemples 3.12.

Soit X un vecteur aléatoire de dimension 2 de loi $\mathbb{P}_X := \sum_{k \geq 1, l \geq 1} \frac{1}{2^{k+l}} \delta_{(k,l)}$. On note X_1, X_2

les composantes de X dans la base canonique de \mathbb{R}^2 . Déterminons la loi de la variable aléatoire réelle $Y := \sup(X_1, X_2)$. Pour cela notons $A := \{(x, y) \in \mathbb{R}^2 / x < y\}$. Soit h une application borélienne de \mathbb{R} dans $[0, +\infty]$. En remarquant que, pour tout $(x, y) \in \mathbb{R}^2$, $h(\sup(x, y)) = h(y)\mathbb{1}_A(x, y) + h(x)\mathbb{1}_{A^c}(x, y)$, il vient

$$\begin{aligned}
\mathbb{E}[h(Y)] &= \mathbb{E}[h(\sup(X_1, X_2))] = \int_{\mathbb{R}^2} h(\sup(x, y)) d\mathbb{P}_X(x, y) \\
&= \int_{\mathbb{R}^2} h(y) \mathbb{1}_A(x, y) d\mathbb{P}_X(x, y) + \int_{\mathbb{R}^2} h(x) \mathbb{1}_{A^c}(x, y) d\mathbb{P}_X(x, y) \\
&= \sum_{i=1}^{+\infty} \sum_{j=1}^{+\infty} \frac{1}{2^{i+j}} h(j) \mathbb{1}_A(i, j) + \sum_{i=1}^{+\infty} \sum_{j=1}^{+\infty} \frac{1}{2^{i+j}} h(i) \mathbb{1}_{A^c}(i, j) \\
&= \sum_{j=1}^{+\infty} \sum_{i=1}^{j-1} \frac{1}{2^{i+j}} h(j) + \sum_{i=1}^{+\infty} \sum_{j=1}^i \frac{1}{2^{i+j}} h(i) \\
&= \sum_{j=1}^{+\infty} \frac{1}{2^j} \left(1 - \frac{1}{2^{j-1}}\right) h(j) + \sum_{i=1}^{+\infty} \frac{1}{2^i} \left(1 - \frac{1}{2^i}\right) h(i) = \sum_{i=1}^{+\infty} \frac{1}{2^i} \left(2 - \frac{3}{2^i}\right) h(i) \\
&= \sum_{i=1}^{+\infty} \frac{1}{2^i} \left(2 - \frac{3}{2^i}\right) \int_{\mathbb{R}} h(z) d\delta_i(z).
\end{aligned}$$

On notera que, pour obtenir le premier terme de la quatrième égalité, il a été fait usage du lemme de permutation des symboles \sum_i et \sum_j pour une suite-double de réels positifs.

On a donc $\mathbb{E}[h(Y)] = \int_{\mathbb{R}} h d\mu$ avec $\mu := \sum_{i=1}^{+\infty} \frac{1}{2^i} \left(2 - \frac{3}{2^i}\right) \delta_i$, ce qui prouve que μ est la loi de la variable aléatoire réelle Y .

On pourra se reporter à [3] exercice I-4 question 2 pour trouver une autre démonstration utilisant la remarque suivant la proposition 2.10, page 27, sur le calcul de lois de variable aléatoire réelle discrètes. \square

Un autre critère fait intervenir plus particulièrement les fonctions continues positives à support compact (qui forment une sous-classe des fonctions boréliennes positives).

Proposition 3.18.

Critère des fonctions à support compact

Soient X un vecteur aléatoire de dimension d et μ une probabilité sur \mathbb{R}^d . Alors le vecteur aléatoire X a pour loi μ si, et seulement si, pour toute application positive h de \mathbb{R}^d dans $[0, +\infty]$, continue et à support compact,

$$\mathbb{E}[h(X)] = \int_{\mathbb{R}^d} h d\mu.$$

Ce qui peut aussi s'écrire avec la notation des opérateurs d'intégration

$$\mathbb{E}_{\mathbb{P}}[h(X)] = \mathbb{E}_{\mu}(h).$$

Démonstration : • - C.N. Supposons que $\mathbb{P}_X = \mu$. Si f est une fonction continue positive à support compact, elle est en particulier une fonction positive borélienne. Donc d'après le critère des fonctions positives boréliennes vu précédemment, on a $\mathbb{E}[f(X)] = \int_{\mathbb{R}^d} f d\mu$.

• - C.S. Réciproquement, supposons que, pour toute application positive h de \mathbb{R}^d dans $[0, +\infty]$, continue et à support compact, $\mathbb{E}[h(X)] = \int_{\mathbb{R}^d} h d\mu$. Soit A une partie ouverte de \mathbb{R}^d . D'après un résultat d'analyse fonctionnelle, il existe une suite croissante $(f_n)_{\mathbb{N}}$ de fonctions positives,

continues et à support compact sur \mathbb{R}^d qui converge vers la fonction indicatrice de A . On a d'une part $\mu(A) = \int_{\mathbb{R}^d} \mathbf{1}_A d\mu$ et $\mathbb{P}_X(A) = \mathbb{E}(\mathbf{1}_A)$, et d'autre part, par le théorème de convergence monotone de Beppo-Lévi, on obtient : $\int_{\mathbb{R}^d} \mathbf{1}_A d\mu = \int_{\mathbb{R}^d} (\lim_{n \rightarrow +\infty} f_n) d\mu = \lim_{n \rightarrow +\infty} \int_{\mathbb{R}^d} f_n d\mu = \lim_{n \rightarrow +\infty} \mathbb{E}(f_n(X)) = \lim_{n \rightarrow +\infty} \int_{\mathbb{R}^d} f_n d\mathbb{P}_X = \int_{\mathbb{R}^d} (\lim_{n \rightarrow +\infty} f_n) d\mathbb{P}_X = \int_{\mathbb{R}^d} \mathbf{1}_A d\mathbb{P}_X$. Par suite, pour tout ouvert A de \mathbb{R}^d , $\mu(A) = \mathbb{P}_X(A)$. Les probabilités μ et \mathbb{P}_X coïncident sur une famille de parties de \mathbb{R}^d stable par intersection finie (π -système) qui engendre la tribu borélienne de \mathbb{R}^d , donc elles sont égales en vertu du théorème d'unicité 1.11 de la page 14. \square

Les deux critères des fonctions positives expriment qu'un vecteur aléatoire X , de dimension d , a pour loi la probabilité μ si, et seulement si, la relation $\mathbb{E}_{\mathbb{P}}[h(X)] = \mathbb{E}_{\mu}(h)$ est vérifiée pour tout fonction h de la famille \mathcal{C} des applications boréliennes positives définies sur \mathbb{R}^d (ou de la famille \mathcal{C} des applications continues positives à support compact sur \mathbb{R}^d).

Le critère d'identification de lois utilisant les fonctions de répartition (lemme d'unicité) peut aussi s'énoncer sous cette forme. Ainsi, ces trois critères peuvent se formuler en un seul énoncé :

Proposition 3.19.

Critères d'identification de lois

Soient X un vecteur aléatoire de dimension d et μ une probabilité sur \mathbb{R}^d . Alors le vecteur aléatoire X a pour loi μ si, et seulement si, la relation $\mathbb{E}_{\mathbb{P}}[h(X)] = \mathbb{E}_{\mu}(h)$ est vérifiée pour tous les éléments h d'un des ensembles \mathcal{C} suivants :

1. Si $d \geq 1$, \mathcal{C} est l'ensemble des applications boréliennes de \mathbb{R}^d dans $[0, +\infty]$.
2. Si $d \geq 1$, \mathcal{C} est l'ensemble des applications positives continues et à support compact de \mathbb{R}^d dans $[0, +\infty]$.
3. Si $d = 1$, \mathcal{C} est l'ensemble des indicatrices $\mathbf{1}_{]-\infty, u]}$ lorsque u parcourt \mathbb{R} .

Démonstration : 1) A déjà été vu dans la proposition 3.17.

2) A déjà été vu dans la proposition 3.18.

3) On se limite au cas $d = 1$. On remarque que, pour tout $u \in \mathbb{R}$,

$$\begin{aligned} \mathbb{P}_X(]-\infty, u]) &= \mathbb{E}(\mathbf{1}_{]-\infty, u]}(X)) \\ \text{et } \mu(]-\infty, u]) &= \mathbb{E}_{\mu}(\mathbf{1}_{]-\infty, u]}). \end{aligned}$$

On conclut par le lemme d'unicité 1.12, page 14. \square

Exercice 3.8. (Corrigé de l'exercice : page 158)

Avec les notations de l'exemple précédent 3.12, montrer que $\mathbb{E}(X) = (2, 2)$ et que la loi de la variable aléatoire réelle $Z := X_1 + X_2$ est

$$\mathbb{P}_Z = \sum_{i=1}^{+\infty} \frac{i-1}{2^i} \delta_i.$$

3.2.2 Moments d'une variable aléatoire

Définition 3.6.

Soit X est une v.a.r. sur $(\Omega, \mathcal{F}, \mathbb{P})$. On appelle **espérance mathématique de X suivant \mathbb{P}** ou quelquefois **moyenne de X** , et on note $\mathbb{E}(X)$, la quantité (si elle est définie)

$$\mathbb{E}(X) = \int_{\Omega} X d\mathbb{P}.$$

Plus généralement si $p \in \mathbb{N}^*$, on appelle **moment d'ordre p de X** , resp. **moment centré d'ordre p de X** , le nombre réel (s'il est défini) $m_p := \mathbb{E}(X^p)$, resp. $m'_p := \mathbb{E}[(X - m_1)^p]$.

Définition 3.7.

Le moment centré d'ordre 2 s'appelle aussi **la variance de X** et se note $\text{Var}(X)$. Sa racine carrée positive s'appelle **l'écart-type de X** et se note σ_X .

Si X et Y sont deux v.a.r. on appelle **covariance de X et Y** , le réel (s'il est défini) $\text{Cov}(X, Y) := \mathbb{E}[(X - \mathbb{E}(X))[Y - \mathbb{E}(Y)]]$.

La proposition suivante donne une condition suffisante d'existence des moments d'une v.a.r. .

Proposition 3.20.

Existence des moments de v.a.r.

1. Soit X une v.a.r. telle qu'il existe un entier naturel non nul p vérifiant $\mathbb{E}(|X|^p) < +\infty$, i.e. $X \in L^p(\Omega, \mathcal{F}, \mathbb{P})$. Alors, pour tout entier k vérifiant $1 \leq k \leq p$, les moments d'ordre k , $m_k := \mathbb{E}(X^k)$ et $m'_k := \mathbb{E}[(X - m_1)^k]$, existent dans \mathbb{R} .
2. Si X et Y sont deux variables aléatoires réelles vérifiant $\mathbb{E}(X^2) < +\infty$ et $\mathbb{E}(Y^2) < +\infty$, i.e. X et Y sont dans $L^2(\Omega, \mathcal{F}, \mathbb{P})$, alors la covariance de X et Y , $\text{Cov}(X, Y)$, existe dans \mathbb{R} .

Démonstration : 1) Pour tout $k \leq p$, $|X^k| \leq 1 + |X|^p$. D'où

$$\mathbb{E}(|X^k|) \leq 1 + \mathbb{E}(|X|^p) < +\infty,$$

ce qui prouve que la variable aléatoire réelle X^k est intégrable et donc que $\mathbb{E}(X^k)$ est bien défini dans \mathbb{R} .

De même, $|X - m_1|^k \leq (|m_1| + |X|)^k$. En développant le second membre, en prenant l'espérance de l'expression et en utilisant le résultat démontré juste avant, on obtient que $\mathbb{E}(|X - m_1|^k) < +\infty$. Par suite $\mathbb{E}[(X - m_1)^k]$ est bien défini dans \mathbb{R} .

2) Si X et Y sont de carré intégrable, d'après l'inégalité $|XY| \leq X^2 + Y^2$ déduite du développement de $(|X| - |Y|)^2 \geq 0$, on obtient $\mathbb{E}(|XY|) \leq \mathbb{E}(X^2) + \mathbb{E}(Y^2) < +\infty$. La variable aléatoire réelle XY est donc intégrable ainsi que la variable aléatoire réelle $Z := [X - \mathbb{E}(X)][Y - \mathbb{E}(Y)]$, ce qui donne bien un sens à la covariance de X et de Y . \square

Par application du théorème de transfert il vient aisément :

Proposition 3.21.

Sous les conditions d'existence des différents moments,

$$\begin{aligned} m_1 &:= \mathbb{E}(X) = \int_{\Omega} X(\omega) d\mathbb{P}(\omega) = \int_{\mathbb{R}} x d\mathbb{P}_X(x). \\ \sigma_X^2 &:= \int_{\Omega} (X(\omega) - m_1)^2 d\mathbb{P}(\omega) = \int_{\mathbb{R}} (x - m_1)^2 d\mathbb{P}_X(x). \\ m_p &:= \int_{\Omega} X^p(\omega) d\mathbb{P}(\omega) = \int_{\mathbb{R}} x^p d\mathbb{P}_X(x). \\ m'_p &:= \int_{\Omega} (X(\omega) - m_1)^p d\mathbb{P}(\omega) = \int_{\mathbb{R}} (x - m_1)^p d\mathbb{P}_X(x). \end{aligned}$$

Par commodité on pose la définition suivante :

Définition 3.8.

Une variable aléatoire réelle X est dite **de carré intégrable** si $\mathbb{E}(X^2) < +\infty$, i.e. $X \in L^2(\Omega, \mathcal{F}, \mathbb{P})$.

Exemples 3.13.

Soit X une variable aléatoire normale standard. Le calcul développé dans l'exemple 3.7, page 34, prouve que $\mathbb{E}(X) = 0$ et celui développé dans l'exemple 3.6, page 33, que $\text{Var}(X) = 1$. Plus généralement, on vérifie aisément par un calcul élémentaire que, si X est une variable de loi normale $\mathcal{N}(m, \sigma^2)$, alors $\mathbb{E}(X) = m$ et $\text{Var}(X) = \sigma^2$. \square

Proposition 3.22.**Formules de König-Huygens**

Soient X et Y deux variable aléatoire réelle de carré intégrable. Alors

1. $\text{Var}(X) = \text{Cov}(X, X) = \mathbb{E}(X^2) - [\mathbb{E}(X)]^2$.
2. $\text{Cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$.

Exercice 3.9. (Corrigé de l'exercice : page 158)

Vérifier les formules de König-Huygens.

On vérifie aisément qu'on retrouve les définitions classiques de l'espérance pour les v.a.r. discrètes ou à densité comme l'indique le résultat suivant :

Proposition 3.23.

1. Si X est une variable aléatoire réelle intégrable discrète de loi

$$\mathbb{P}_X := \sum_{k=0}^{+\infty} \mathbb{P}(X = a_k) \delta_{a_k}, \text{ alors}$$

$$\mathbb{E}(X) = \sum_{k=0}^{+\infty} a_k \mathbb{P}(X = a_k).$$

2. Si X est une variable aléatoire réelle intégrable à densité ρ continue sur \mathbb{R} , alors

$$\mathbb{E}(X) = \int_{-\infty}^{+\infty} t \rho(t) dt.$$

On trouvera la liste des valeurs de l'espérance et de la variance des v.a.r. de lois usuelles dans le formulaire de l'annexe A, page 205, de ce cours.

Exercice 3.10. (Corrigé de l'exercice : page 159)

Soit X une variable aléatoire réelle suivant la **loi de Rayleigh** de densité définie sur \mathbb{R} par $\rho(x) := x e^{-\frac{x^2}{2}} \mathbb{1}_{\mathbb{R}^+}(x)$. Montrer que, pour tout entier $k \geq 1$,

$$\mathbb{E}(X^{2k-1}) = \frac{(2k)!}{2^k k!} \sqrt{\frac{\pi}{2}} \text{ et } \mathbb{E}(X^{2k}) = 2^k k!.$$

Expliciter l'espérance et la variance de la variable aléatoire réelle X .

Définition 3.9.

Soit X un vecteur aléatoire de dimension d de composantes X_1, X_2, \dots, X_d intégrables suivant \mathbb{P} . On appelle **espérance de X suivant \mathbb{P}** , et on note $\mathbb{E}(X)$, le vecteur de \mathbb{R}^d ,

$$\mathbb{E}(X) := (\mathbb{E}(X_1), \mathbb{E}(X_2), \dots, \mathbb{E}(X_d)).$$

Si X est une v.a. à valeurs matricielles, on appelle **espérance de X** , et on note $\mathbb{E}(X)$, la matrice dont les coefficients sont les espérances de ceux de X .

Définition 3.10.

Une variable aléatoire vectorielle ou matricielle d'espérance nulle est dite **centrée**. Une variable aléatoire réelle de carré intégrable et de variance égale à 1 est dite **réduite**.

Exercice 3.11. (Corrigé de l'exercice : page 159)

Montrer que les variable aléatoire réelle X_1, X_2, \dots, X_d sont de carré intégrable si, et seulement si, $\mathbb{E}(|X|^2) < +\infty$ où $X := (X_1, X_2, \dots, X_d)$.

Définition 3.11.

Soit X un vecteur aléatoire de dimension d de composantes X_1, X_2, \dots, X_d de carré intégrable sur Ω . On appelle **matrice de dispersion de X ou matrice des covariances de X** , et on la note D_X , l'espérance de la matrice carrée aléatoire $[X - \mathbb{E}(X)][X - \mathbb{E}(X)]^*$ d'ordre d où $*$ désigne l'opération de transposition des matrices, c'est-à-dire

$$D_X = \mathbb{E}([X - \mathbb{E}(X)][X - \mathbb{E}(X)]^*) .$$

La terminologie est justifiée par la troisième assertion de la proposition suivante :

Proposition 3.24.

Si X est un vecteur aléatoire de dimension d tel que $\mathbb{E}(|X|^2) < +\infty$ et M une matrice (déterministe) à coefficients réels à c lignes et d colonnes, alors

1. Le coefficient d'indice (i, j) de D_X est la covariance $\text{Cov}(X_i, X_j)$ des v.a.r. X_i et X_j . Les éléments diagonaux de D_X sont les variances des composantes de X .
- 2.

$$D_{[X - \mathbb{E}(X)]} = D_X, \quad \mathbb{E}(MX) = M\mathbb{E}(X), \quad D_{MX} = MD_X M^*.$$

3. D_X est une matrice symétrique. D_X est de **type positif** i.e., pour tout $u \in \mathbb{R}^d$, $u^* D_X u \geq 0$. En particulier, D_X est une matrice diagonalisable sur \mathbb{R} dont les valeurs propres sont des réels positifs ou nuls.

Exemples 3.14.

Soit X un vecteur aléatoire de dimension 2 de loi

$$\mathbb{P}_X := \sum_{k \geq 1, l \geq 1} \frac{1}{2^{k+l}} \delta_{(k,l)}.$$

Notons X_1, X_2 les composantes de X dans la base canonique de \mathbb{R}^2 . Par définition $\mathbb{E}(X) = (\mathbb{E}(X_1), \mathbb{E}(X_2))$ et d'après la proposition précédente,

$$D_X = \begin{pmatrix} \text{Var}(X_1) & \text{Cov}(X_1, X_2) \\ \text{Cov}(X_1, X_2) & \text{Var}(X_2) \end{pmatrix}.$$

En calculant $\mathbb{E}(X_1) = \mathbb{E}(X_2) = 2$ puis $\mathbb{E}(X_1^2) = \mathbb{E}(X_2^2) = 6$ et $\text{Cov}(X_1, X_2) = 0$ on obtient, par application de la formule de König-Huygens,

$$\mathbb{E}(X) = (2, 2) \text{ et } D_X = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}. \square$$

Exercice 3.12. (Corrigé de l'exercice : page 159)

Démontrer la proposition 3.24, page 48.

Exercice 3.13. (Corrigé de l'exercice : page 160)

Soient X et Y deux variables aléatoires réelles telles que $\mathbb{E}(X^2) < +\infty$ et $\mathbb{E}(Y^2) < +\infty$.

1. Montrer que $|YX| \leq X^2 + Y^2$. En déduire que les variables aléatoires réelles X , Y et XY sont intégrables suivant \mathbb{P} .
2. En étudiant le signe de l'expression $\mathbb{E}[(X + \alpha Y)^2]$ pour tout $\alpha \in \mathbb{R}$, prouver l'**inégalité de Cauchy-Schwarz** : $|\mathbb{E}(XY)| \leq \sqrt{\mathbb{E}(X^2)}\sqrt{\mathbb{E}(Y^2)}$. En donner une interprétation géométrique.

Exercice 3.14. (Corrigé de l'exercice : page 160)

Soit $X := (X_1, X_2, \dots, X_n)$ un vecteur aléatoire tel que $\mathbb{E}(|X|^2) < +\infty$.

1. Montrer que la variable aléatoire réelle $Y := \sum_{k=1}^n X_k$ est de carré intégrable.
2. Démontrer la relation

$$\text{Var}\left(\sum_{k=1}^n X_k\right) = \sum_{k=1}^n \text{Var}(X_k) + 2 \sum_{1 \leq i < j \leq n} \text{Cov}(X_i, X_j).$$

3.3 Fonction caractéristique et loi d'une v.a.

Dans le sous-paragraphe 3 du paragraphe 1, nous avons introduit la notion de fonction caractéristique.

On rappelle que :

Définition 3.12.

Si μ est une probabilité sur \mathbb{R}^d , l'application

$$\phi_\mu : u \in \mathbb{R}^d \mapsto \phi_\mu(u) := \int_{\mathbb{R}^d} \exp(i\langle x, u \rangle) d\mu(x) \in \mathbb{C},$$

s'appelle la **fonction caractéristique de μ** , (en abrégé **f.c.**).

Si X est un vecteur aléatoire de dimension d , la **fonction caractéristique de X** est l'application de \mathbb{R}^d dans \mathbb{C} définie, pour tout vecteur u de \mathbb{R}^d , par

$$\phi_X(u) := \int_{\mathbb{R}^d} \exp(i\langle x, u \rangle) d\mathbb{P}_X(x) = \mathbb{E}(\exp(i\langle X, u \rangle)).$$

Donnons quelques propriétés classiques des fonctions caractéristiques :

Proposition 3.25.

1. $\phi(0) = 1$
2. Pour tout $u \in \mathbb{R}^d$, $\phi(-u) = \overline{\phi(u)}$ et $|\phi(u)| \leq 1$. Une fonction caractéristique est donc une application bornée sur \mathbb{R}^d .
3. La fonction caractéristique ϕ d'un vecteur aléatoire X de dimension d est une fonction uniformément continue sur \mathbb{R}^d . En particulier, une fonction caractéristique ϕ est continue en 0.
4. Si μ et ν sont deux probabilités sur \mathbb{R}^d , alors

$$\int_{\mathbb{R}^d} \phi_\mu d\nu = \int_{\mathbb{R}^d} \phi_\nu d\mu.$$

Démonstration : • $\Phi(0) = 1$ est immédiat à vérifier.

• Pour tout $u \in \mathbb{R}^d$, $\Phi(-u) = \overline{\Phi(u)}$ est immédiat à vérifier. Pour tout vecteur u de \mathbb{R}^d , $|\Phi_X(u)| = \left| \int_{\mathbb{R}^d} \exp(i\langle x, u \rangle) d\mathbb{P}_X(x) \right| \leq \int_{\mathbb{R}^d} |\exp(i\langle x, u \rangle)| d\mathbb{P}_X(x) \leq \int_{\mathbb{R}^d} d\mathbb{P}_X(x) = 1$, donc Φ_X est bornée sur \mathbb{R}^d .

• Pour tous vecteurs u et v de \mathbb{R}^d , $|\Phi_X(u) - \Phi_X(v)| \leq \int_{\mathbb{R}^d} |\exp(i\langle x, u \rangle) - \exp(i\langle x, v \rangle)| d\mathbb{P}_X(x)$.

Or, pour tout réel t , $e^{it} - 1 = \int_0^t ie^{ix} dx$, d'où $|e^{it} - 1| \leq \inf(2, |t|)$. Par suite, pour tous vecteurs u et v de \mathbb{R}^d , par application de l'inégalité de Cauchy-Schwarz, il vient

$$\begin{aligned} |\exp(i\langle x, u \rangle) - \exp(i\langle x, v \rangle)| &\leq |\exp[i(\langle x, u \rangle - \langle x, v \rangle)] - 1| = |\exp(i\langle x, u - v \rangle) - 1| \\ &\leq \inf(2, |\langle x, u - v \rangle|) \leq \inf(2, |x||u - v|). \end{aligned}$$

D'où, pour tous vecteurs u et v de \mathbb{R}^d , $|\Phi_X(u) - \Phi_X(v)| \leq \int_{\mathbb{R}^d} \inf(2, |x||u - v|) d\mathbb{P}_X(x)$. En parti-

culier, pour tout entier naturel non nul n , et pour tous vecteurs u et v de \mathbb{R}^d , tels que $|u - v| \leq \frac{1}{n}$, on a $|\Phi_X(u) - \Phi_X(v)| \leq \int_{\mathbb{R}^d} \inf\left(2, \frac{|x|}{n}\right) d\mathbb{P}_X(x)$. La suite de fonctions $\left(\inf\left(2, \frac{|x|}{n}\right)\right)_{n \in \mathbb{N}^*}$ est dominée par la fonction constante qui vaut 2 sur \mathbb{R}^d , et converge vers la fonction-nulle sur \mathbb{R}^d . Par le théorème de convergence dominée de Lebesgue, $\lim_{n \rightarrow +\infty} \int_{\mathbb{R}^d} \inf\left(2, \frac{|x|}{n}\right) d\mathbb{P}_X(x) = 0$.

Soit $\varepsilon > 0$ donné, on peut donc trouver un entier n tel que $\left| \int_{\mathbb{R}^d} \inf\left(2, \frac{|x|}{n}\right) d\mathbb{P}_X(x) \right| \leq \varepsilon$. Par

suite, pour tout $\varepsilon > 0$, il existe η (prendre $\eta = \frac{1}{n}$) tel que, pour tous vecteurs u et v de \mathbb{R}^d , $|u - v| \leq \eta$ implique $|\Phi_X(u) - \Phi_X(v)| \leq \varepsilon$, ce qui prouve l'uniforme continuité de la fonction caractéristique.

• Pour prouver que $\int_{\mathbb{R}^d} \Phi_\mu d\nu = \int_{\mathbb{R}^d} \Phi_\nu d\mu$, il suffit d'appliquer le théorème de Fubini vu en théorie de la mesure et de l'intégration et rappelé dans le chapitre suivant, proposition 4.3, page 63. \square

La fonction caractéristique présente plusieurs points d'intérêt :

1. pour identifier la loi d'un vecteur aléatoire,
2. pour calculer les moments d'une variable aléatoire,
3. pour étudier l'indépendance d'une suite de v.a.r..

Nous allons nous intéresser dans ce paragraphe aux deux premiers points. Le dernier sera traité au chapitre IV.

Commençons par le premier point, qui est dû au résultat suivant :

Proposition 3.26.

Théorème d'injectivité des f.c.

Deux probabilités sur \mathbb{R}^d sont identiques si, et seulement si, elles ont la même fonction caractéristique.

Démonstration : • En préambule à la démonstration, montrons que, pour tout réel $a > 0$,

$$I(x - y) = \int_{\mathbb{R}^d} e^{i\langle u, y-x \rangle} e^{-a(|u_1|+|u_2|+\dots+|u_d|)} d\lambda^{(d)}(u) = \frac{2^d a^d}{\prod_{k=1}^{k=d} (a^2 + (y_k - x_k)^2)}.$$

En effet, en intégrant séparément sur $] -\infty, 0]$, puis sur $[0, +\infty[$, on obtient

$$K(x) = \int_{\mathbb{R}} e^{itx} e^{-a|t|} d\lambda(t) = \int_{-\infty}^0 e^{itx} e^{at} dt + \int_0^{+\infty} e^{itx} e^{-at} dt = \frac{1}{a - ix} + \frac{1}{a + ix} = \frac{2a}{a^2 + x^2}.$$

Par suite, $I(x - y) = \int_{\mathbb{R}^d} e^{i[u_1(y_1-x_1)+u_2(y_2-x_2)+\dots+u_d(y_d-x_d)]-a(|u_1|+|u_2|+\dots+|u_d|)} d\lambda^{(d)}(u)$ ou encore

$$I(x - y) = \int_{\mathbb{R}^d} \prod_{k=1}^{k=d} e^{i[u_k(y_k-x_k)-a|u_k|]} d\lambda^{(d)}(u).$$

Par le théorème de Fubini (son énoncé est rappelé dans le prochain chapitre, proposition 4.3, page 63), compte tenu de ce que la fonction à intégrer est à variables séparées, il vient

$$I(x - y) = \prod_{k=1}^{k=d} \int_{\mathbb{R}} e^{i[u_k(y_k-x_k)-a|u_k|]} d\lambda(u_k) = \prod_{k=1}^{k=d} K(y_k - x_k) = \prod_{k=1}^{k=d} \frac{2a}{a^2 + (y_k - x_k)^2}, \text{ d'où}$$

$$I(x - y) = \frac{2^d a^d}{\prod_{k=1}^{k=d} (a^2 + (y_k - x_k)^2)}.$$

• Montrons maintenant que :

$$J_{\mu}(x) = \int_{\mathbb{R}^d} \frac{2^d a^d}{\prod_{k=1}^{k=d} (a^2 + (y_k - x_k)^2)} d\mu(y) = \int_{\mathbb{R}^d} e^{-i\langle u, x \rangle} e^{-a(|u_1|+|u_2|+\dots+|u_d|)} \Phi_{\mu}(u) d\lambda^{(d)}(u).$$

En effet, en utilisant le résultat précédent et en reportant dans l'intégrale à calculer, on a $J_{\mu}(x) = \int_{\mathbb{R}^d} I(x - y) d\mu(y) = \int_{\mathbb{R}^d} \left(\int_{\mathbb{R}^d} e^{i\langle u, y-x \rangle} e^{-a(|u_1|+|u_2|+\dots+|u_d|)} d\lambda^{(d)}(u) \right) d\mu(y)$. Par application du théorème de Fubini aux mesures $\lambda^{(d)}$ et μ (vérifier que les hypothèses du théorème sont bien satisfaites), cette intégrale peut s'écrire :

$$J_{\mu}(x) = \int_{\mathbb{R}^d} e^{i\langle u, -x \rangle} e^{-a(|u_1|+|u_2|+\dots+|u_d|)} \left(\int_{\mathbb{R}^d} e^{i\langle u, y \rangle} d\mu(y) \right) d\lambda^{(d)}(u),$$

ou encore, par définition de la fonction caractéristique de μ ,

$$J_{\mu}(x) = \int_{\mathbb{R}^d} e^{-i\langle u, x \rangle} e^{-a(|u_1|+|u_2|+\dots+|u_d|)} \Phi_{\mu}(u) d\lambda^{(d)}(u).$$

• Soit f une fonction de \mathbb{R}^d dans \mathbb{R} , continue et à support compact. Calculons $H_{\mu}(a) = \int_{\mathbb{R}^d} f(x) J_{\mu}(x) d\lambda^{(d)}(x)$.

On a $H_{\mu}(a) = \int_{\mathbb{R}^d} f(x) \left[\int_{\mathbb{R}^d} \frac{2^d a^d}{\prod_{k=1}^{k=d} (a^2 + (y_k - x_k)^2)} d\mu(y) \right] d\lambda^{(d)}(x)$. Par le théorème de Fubini, $H_{\mu}(a) = \int_{\mathbb{R}^d} \left[\int_{\mathbb{R}^d} f(x) \frac{2^d a^d}{\prod_{k=1}^{k=d} (a^2 + (y_k - x_k)^2)} d\lambda^{(d)}(x) \right] d\mu(y)$. Notons $G_a(y)$ l'intégrale

entre crochets. Elle peut s'écrire, après le changement de variables dans \mathbb{R}^d , $(u_1, u_2, \dots, u_d) = (y_1 - x_1, y_2 - x_2, \dots, y_d - x_d)$, ou $u = y - x$, dont le jacobien est de valeur absolue égale à 1,

$$G_a(y) = \int_{\mathbb{R}^d} f(x) \frac{2^d a^d}{\prod_{k=1}^{k=d} (a^2 + (y_k - x_k)^2)} d\lambda^{(d)}(x) = \int_{\mathbb{R}^d} f(y-u) \frac{2^d a^d}{\prod_{k=1}^{k=d} (a^2 + (u_k)^2)} d\lambda^{(d)}(u).$$

Par suite, $H_\mu(a) = \int_{\mathbb{R}^d} G_a(y) d\mu(y)$.

• *Étudions maintenant, pour tout vecteur $y \in \mathbb{R}^d$, la limite $\lim_{a \rightarrow 0} G_a(y)$.*

Pour tout vecteur $y \in \mathbb{R}^d$, $G_a(y) = \int_{\mathbb{R}^d} f(y-u) \frac{2^d a^d}{\prod_{k=1}^{k=d} (a^2 + (u_k)^2)} d\lambda^{(d)}(u)$. Par le changement de variables dans \mathbb{R}^d , $(x_1, x_2, \dots, x_d) = (\frac{u_1}{a}, \frac{u_2}{a}, \dots, \frac{u_d}{a})$, qu'on peut écrire de façon plus synthétique, $x = \frac{1}{a}u$, dont le jacobien est de valeur absolue égale à a^d , l'intégrale devient

$$\begin{aligned} G_a(y) &= \int_{\mathbb{R}^d} f(y-ax) \frac{2^d a^d}{\prod_{k=1}^{k=d} (a^2 + (ax_k)^2)} a^d d\lambda^{(d)}(x) \\ &= \int_{\mathbb{R}^d} f(y-ax) \frac{2^d}{\prod_{k=1}^{k=d} (1 + x_k^2)} d\lambda^{(d)}(x). \end{aligned}$$

Comme f est continue à support compact, elle est bornée par un réel M . La famille, indexée par a , des fonctions $x \mapsto f(y-ax) \frac{2^d}{\prod_{k=1}^{k=d} (1 + x_k^2)}$ est dominée par la fonction $x \mapsto M \frac{2^d}{\prod_{k=1}^{k=d} (1 + x_k^2)}$ intégrable sur \mathbb{R}^d pour la mesure de Lebesgue. En effet, par le théorème de Fubini,

$$\int_{\mathbb{R}^d} M \frac{2^d}{\prod_{k=1}^{k=d} (1 + x_k^2)} d\lambda^{(d)}(x) = M 2^d \left(\int_{-\infty}^{+\infty} \frac{1}{1+t^2} dt \right)^d = M 2^d ([\text{Arctan}(x)]_{-\infty}^{+\infty})^d = M 2^d \pi^d,$$

ce qui prouve que la fonction $x \mapsto M \frac{2^d}{\prod_{k=1}^{k=d} (1 + x_k^2)}$ est intégrable sur \mathbb{R}^d .

Par suite en faisant tendre a vers 0, en vertu du théorème de convergence dominée, l'intégrale

$$G_a(y) = \int_{\mathbb{R}^d} f(y-ax) \frac{2^d}{\prod_{k=1}^{k=d} (1 + x_k^2)} d\lambda^{(d)}(x) \text{ tend vers l'intégrale}$$

$$\int_{\mathbb{R}^d} f(y) \frac{2^d}{\prod_{k=1}^{k=d} (1 + x_k^2)} d\lambda^{(d)}(x) = f(y) \int_{\mathbb{R}^d} \frac{2^d}{\prod_{k=1}^{k=d} (1 + x_k^2)} d\lambda^{(d)}(x) = 2^d \pi^d f(y),$$

d'après le calcul qui vient d'être fait plus haut, c'est-à-dire $\lim_{a \rightarrow 0} G_a(y) = 2^d \pi^d f(y)$.

• *Nous pouvons maintenant en déduire la limite de l'intégrale $H_\mu(a)$ lorsque a tend vers 0.*
En effet, pour tout réel a et tout $y \in \mathbb{R}^d$,

$$\begin{aligned} |G_a(y)| &\leq \int_{\mathbb{R}^d} |f(y-ax)| \frac{2^d a^d}{\prod_{k=1}^{k=d} (a^2 + (ax_k)^2)} a^d d\lambda^{(d)}(x) \\ &\leq M \int_{\mathbb{R}^d} \frac{2^d a^d}{\prod_{k=1}^{k=d} (a^2 + (ax_k)^2)} a^d d\lambda^{(d)}(x) = M 2^d \pi^d. \end{aligned}$$

La famille de fonctions G_a indexées par a , est donc dominée par la constante $M 2^d \pi^d$ intégrable par rapport à la mesure de probabilité μ . Par le théorème de convergence dominée de Lebesgue,

il vient $\lim_{a \rightarrow 0} H_\mu(a) = \int_{\mathbb{R}^d} \lim_{a \rightarrow 0} G_a(y) d\mu(y) = \int_{\mathbb{R}^d} 2^d \pi^d f(y) d\mu(y)$.

• Pour conclure, rassemblons les résultats précédents, en tenant compte que μ et ν sont deux probabilités sur \mathbb{R}^d ayant la même fonction caractéristique $\Phi_\mu = \Phi_\nu$. On remarque alors que $J_\mu(x) = J_\nu(x)$ et, par suite, pour tout réel $a > 0$, $H_\mu(a) = H_\nu(a)$. On en déduit alors, en faisant tendre a vers 0, que $\lim_{a \rightarrow 0} H_\mu(a) = \lim_{a \rightarrow 0} H_\nu(a)$, c'est-à-dire que

$$\int_{\mathbb{R}^d} 2^d \pi^d f(y) d\mu(y) = \int_{\mathbb{R}^d} 2^d \pi^d f(y) d\nu(y), \text{ et finalement } \int_{\mathbb{R}^d} f(y) d\mu(y) = \int_{\mathbb{R}^d} f(y) d\nu(y),$$

pour toute fonction f de \mathbb{R}^d dans \mathbb{R} , continue et à support compact. Par le critère des fonctions à support compact (proposition 3.18, page 43), on en conclut $\mu = \nu$. \square

Le théorème d'injectivité s'énonce alors avec les vecteurs aléatoires :

Proposition 3.27.

Critère d'identification de lois par les f.c.

Deux vecteurs aléatoires sur \mathbb{R}^d ont la même loi si, et seulement si, ils ont la même fonction caractéristique.

Nous pouvons maintenant regrouper tous les critères d'identification de lois vus jusqu'à présent sous une formulation unique qui complète la proposition 3.19, page 44 :

Proposition 3.28.

Critères d'identification de lois

Soient X un vecteur aléatoire de dimension d et μ une probabilité sur \mathbb{R}^d . Alors le vecteur aléatoire X a pour loi μ si, et seulement si, la relation

$$\mathbb{E}[h(X)] = \int_{\mathbb{R}^d} h(t) d\mu(t)$$

est vérifiée pour tous les éléments h d'un des ensembles \mathcal{C} suivants :

1. Si $d \geq 1$, \mathcal{C} est l'ensemble des applications boréliennes de \mathbb{R}^d dans $[0, +\infty]$.
2. Si $d \geq 1$, \mathcal{C} est l'ensemble des applications positives continues et à support compact de \mathbb{R}^d dans $[0, +\infty]$.
3. Si $d \geq 1$, \mathcal{C} est l'ensemble des applications

$$h_u : x \in \mathbb{R}^d \mapsto h_u(x) = \exp(i\langle u, x \rangle) \in \mathbb{C}$$

lorsque u parcourt \mathbb{R}^d .

4. Si $d = 1$, \mathcal{C} est l'ensemble des indicatrices $\mathbb{1}_{]-\infty, u]}$ lorsque u parcourt \mathbb{R} .

Démonstration : Il suffit de prouver l'item 3, les autres ayant déjà été vus dans le théorème 3.19 de la page 44. On remarque que, pour tout $u \in \mathbb{R}^d$, $\Phi_X(u) = \mathbb{E}(\exp(i\langle u, X \rangle)) = \mathbb{E}(h_u(X))$ et $\Phi_\mu(u) = \mathbb{E}_\mu(h_u)$. On conclut par le théorème d'injectivité 3.26. \square

Définition 3.13.

Les familles de fonctions qui apparaissent dans les différents item de la proposition précédente sont souvent appelées **familles de fonctions-test**.

Dans le cas où $d = 1$, les formules d'inversion données ci-dessous précisent le lien entre la

probabilité et sa fonction caractéristique. Elles permettent de retrouver la probabilité à partir de sa f.c. :

Proposition 3.29.

Formules d'inversion ou de réciprocity de Perron-Stieltjès

Si μ est une probabilité sur \mathbb{R} de f.c. Φ , a et b deux réels tels que $a < b$, alors :

1. La suite d'intégrales au sens de Riemann $\left(\frac{1}{2\pi} \int_{-n}^n \frac{e^{-ita} - e^{-itb}}{it} \Phi(t) dt \right)_{n \in \mathbb{N}}$ est convergente dans \mathbb{R} et $\lim_{n \rightarrow +\infty} \frac{1}{2\pi} \int_{-n}^n \frac{e^{-ita} - e^{-itb}}{it} \Phi(t) dt = \mu(]a, b[) + \frac{1}{2}\mu(\{a, b\})$.
2. La suite d'intégrales au sens de Riemann $\left(\frac{1}{2n} \int_{-n}^n e^{-ita} \Phi(t) dt \right)_{n \in \mathbb{N}^*}$ est convergente dans \mathbb{R} et $\lim_{n \rightarrow +\infty} \frac{1}{2n} \int_{-n}^n e^{-ita} \Phi(t) dt = \mu(\{a\})$.

Indications pour la démonstration (laissée en exercice non corrigé).

Vérifier que les intégrales qui interviennent dans la proposition peuvent bien être prises au sens de Riemann car les fonctions complexes à intégrer sont continues (ou prolongeables par continuité) sur $[-n, n]$. On rappelle que $\int_{-\infty}^{+\infty} \frac{\sin t}{t} dt = \pi$.

- Premier item : Justifier qu'on peut appliquer le théorème de Fubini pour obtenir :

$$I_n := \int_{-n}^n \frac{e^{-ita} - e^{-itb}}{it} \Phi(t) dt = \int_{\mathbb{R}} \left\{ \int_{-n}^n \frac{\sin(t(x-a))}{t} dt - \int_{-n}^n \frac{\sin(t(x-b))}{t} dt \right\} d\mu(x).$$

Posons, pour tout $u \in \mathbb{R}$, $S(u) := \int_0^u \frac{\sin t}{t} dt$ et $\operatorname{sgn}(u) := \frac{u}{|u|}$ si $u \neq 0$ avec $\operatorname{sgn}(0) := 0$.

Montrer que, pour tout $u \in \mathbb{R}$, $\int_{-n}^n \frac{\sin tu}{t} dt = 2\operatorname{sgn}(u)S(n|u|)$ et que $I_n = \int_{\mathbb{R}} f_n d\mu$ où, pour tout $x \in \mathbb{R}$, $f_n(x) := 2\operatorname{sgn}(x-a)S(n|x-a|) - 2\operatorname{sgn}(x-b)S(n|x-b|)$. Montrer que la suite $(f_n)_{n \in \mathbb{N}}$ converge vers $\pi \mathbb{1}_{\{a, b\}} + 2\pi \mathbb{1}_{]a, b[}$ et qu'il existe $M > 0$ tel que, pour tout $n \in \mathbb{N}$, $|f_n| \leq M$. Conclure à l'aide du théorème de convergence dominée.

- Second item : Utiliser une démarche analogue. \square

Dans le cas où la f.c. est intégrable au sens de Lebesgue sur \mathbb{R} , on peut préciser la connaissance de μ :

Proposition 3.30.

Soit μ une probabilité sur \mathbb{R} de f.c. Φ . Si Φ est intégrable au sens de Lebesgue sur \mathbb{R} , alors μ admet une densité f par rapport à la mesure de Lebesgue sur \mathbb{R} . L'application f est une fonction à valeurs réelles, positive, bornée, continue sur \mathbb{R} et, pour tout $x \in \mathbb{R}$, $f(x)$ s'exprime à l'aide de l'intégrale généralisée au sens de Riemann $f(x) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} e^{-itx} \Phi(t) dt$.

Démonstration : • Remarquons que, pour tous réels a et b avec $a < b$,

$$\left| \frac{e^{-ita} - e^{-itb}}{it} \right| = \left| \int_a^b e^{-itx} d\lambda(x) \right| \leq \int_a^b |e^{-itx}| d\lambda(x) \leq |b - a|.$$

Par suite, $\int_{-n}^n \left| \frac{e^{-ita} - e^{-itb}}{it} \Phi(t) \right| dt \leq |b - a| \int_{-n}^n |\Phi(t)| dt$. Comme Φ est supposée intégrable au sens de Lebesgue, on a $\int_{-\infty}^{+\infty} |\Phi(t)| dt < +\infty$, ce qui prouve que l'intégrale $\int_{-n}^n \left| \frac{e^{-ita} - e^{-itb}}{it} \Phi(t) \right| dt$ est absolument convergente. On a donc d'après la proposition 3.7, page 35,

$$\lim_{n \rightarrow +\infty} \int_{-n}^n \frac{e^{-ita} - e^{-itb}}{it} \Phi(t) dt = \int_{-\infty}^{+\infty} \frac{e^{-ita} - e^{-itb}}{it} \Phi(t) dt = \int_{\mathbb{R}} \frac{e^{-ita} - e^{-itb}}{it} \Phi(t) d\lambda(t).$$

- D'après le premier item de la proposition 3.29, nous pouvons donc écrire

$$\mu([a, b]) + \frac{1}{2}\mu(\{a, b\}) = \frac{1}{2\pi} \int_{\mathbb{R}} \frac{e^{-ita} - e^{-itb}}{it} \Phi(t) d\lambda(t).$$

On a alors, pour tous réels a et b avec $a < b$,

$$\begin{aligned} \frac{1}{2}\mu(\{a\}) &\leq \mu([a, b]) + \frac{1}{2}\mu(\{a, b\}) = \left| \frac{1}{2\pi} \int_{\mathbb{R}} \frac{e^{-ita} - e^{-itb}}{it} \Phi(t) d\lambda(t) \right| \\ &\leq \frac{1}{2\pi} \int_{\mathbb{R}} \left| \frac{e^{-ita} - e^{-itb}}{it} \Phi(t) \right| d\lambda(t) \leq \frac{1}{2\pi} |b - a| \int_{\mathbb{R}} |\Phi(t)| d\lambda(t) = C |b - a|, \end{aligned}$$

où $C = \frac{1}{2\pi} \int_{\mathbb{R}} |\Phi(t)| d\lambda(t)$ est une constante finie car Φ est intégrable au sens de Lebesgue.

Fixons a , en faisant tendre b vers a , dans l'inégalité précédente, on en déduit que la mesure μ est diffuse sur \mathbb{R} , c'est-à-dire que $\mu(\{x\}) = 0$, pour tout réel x .

- En revenant à la première formule d'inversion de la proposition 3.29, on obtient alors en tenant compte de la remarque précédente et du théorème de Fubini, pour tous réels a et b avec $a < b$,

$$\begin{aligned} \mu([a, b]) &= \frac{1}{2\pi} \int_{\mathbb{R}} \frac{e^{-ita} - e^{-itb}}{it} \Phi(t) d\lambda(t) = \frac{1}{2\pi} \int_{\mathbb{R}} \left(\int_{[a, b]} e^{-itx} d\lambda(x) \right) \Phi(t) d\lambda(t) \\ &= \int_{[a, b]} \left(\frac{1}{2\pi} \int_{\mathbb{R}} e^{-itx} \Phi(t) d\lambda(t) \right) d\lambda(x) = \int_{[a, b]} f(x) d\lambda(x) \end{aligned}$$

où on a posé, pour tout réel x , $f(x) = \frac{1}{2\pi} \int_{\mathbb{R}} e^{-itx} \Phi(t) d\lambda(t)$.

- On vérifie aisément à l'aide du théorème de convergence dominée que f est une fonction continue. De plus f est bornée sur \mathbb{R} . Comme, pour tout réel t , $\Phi(-t) = \overline{\Phi(t)}$, on en déduit en prenant le conjugué de l'intégrale suivi d'un changement de variable élémentaire que, pour tout réel x , $f(x) = \overline{f(x)}$, ce qui prouve que f est une fonction à valeurs réelles. Pour tous réels a et b avec $a < b$, $\mu([a, b]) = \int_{[a, b]} f(x) d\lambda(x) \geq 0$ et f est continue, on en déduit que f est positive et que c'est bien une densité de probabilité sur \mathbb{R} . Comme la fonction sous l'intégrale $t \mapsto e^{-itx} \Phi(t)$ est elle-même continue et bornée et intégrable au sens de Lebesgue, l'intégrale au sens de Lebesgue $\frac{1}{2\pi} \int_{\mathbb{R}} e^{-itx} \Phi(t) d\lambda(t)$ peut s'écrire sous la forme d'une intégrale généralisée au sens de Riemann (cf. proposition 3.7, page 35).

- La probabilité μ coïncide donc avec la probabilité de densité f sur la famille \mathcal{C} des intervalles ouverts de \mathbb{R} . On vérifie aisément que \mathcal{C} est un π -système qui engendre la tribu borélienne de \mathbb{R} . On conclut alors par le théorème d'unicité pour les probabilités (cf. proposition 1.11, page 14) que μ est la probabilité de densité f . \square

La réciproque de ce résultat est fautive. En effet on vérifie aisément que la loi exponentielle de paramètre 1 sur \mathbb{R} fournit un contre-exemple. Cependant, sous certaines hypothèses, on peut avoir des renseignements sur le comportement à l'infini de la fonction caractéristique Φ , comme le prouve la proposition suivante :

Proposition 3.31.

Soit μ une probabilité sur \mathbb{R} de f.c. Φ . On suppose que μ admet une densité f de classe \mathcal{C}^n telle que, pour tout entier $0 \leq k \leq n$, la dérivée $f^{(k)}$ d'ordre k de f soit intégrable au sens de Lebesgue sur \mathbb{R} (on pose $f^{(0)} := f$). Alors $\lim_{u \rightarrow +\infty} u^{n-1} \Phi(u) = 0$.

Indications pour la démonstration (laissée en exercice non corrigé) : Soit k un entier, $1 \leq k \leq n$. Prouver la relation, pour tout $x \in \mathbb{R}$:

$$f^{(k-1)}(x) = \int_0^x f^{(k)}(u) du + f^{(k-1)}(0).$$

En déduire que $\lim_{u \rightarrow +\infty} |f^{(k-1)}(u)| = 0$.

Par une intégration par parties, montrer que, pour tout $u \neq 0$,

$$\int_{-\infty}^{+\infty} e^{-iux} f^{(k-1)}(x) dx = \frac{i}{u} \int_{-\infty}^{+\infty} e^{-iux} f^{(k)}(x) dx.$$

En déduire que $\Phi(u) = \left(\frac{i}{u}\right)^n \int_{-\infty}^{+\infty} e^{-iux} f^{(n)}(x) dx$. Conclure. \square

Avec les notations de Landau, le résultat démontré peut s'écrire $\Phi(u) = o\left(\frac{1}{u^{n-1}}\right)$ au voisinage de $+\infty$. En fait, si f a ses n dérivées premières qui existent et sont intégrables au sens de Lebesgue sur \mathbb{R} , en utilisant le lemme de Riemann-Lebesgue (cf [4], exercice VI-30), on démontre que $\Phi(u) = o\left(\frac{1}{u^n}\right)$ au voisinage de $+\infty$.

Passons maintenant au second point d'intérêt de la notion de f.c.. Les propositions ci-dessous, 3.33, 3.34 et 3.35, donnent un procédé de calcul des moments d'une variable aléatoire réelle à l'aide de sa fonction caractéristique. Ces résultats se démontrent en utilisant le théorème de dérivation sous le signe \int vu en théorie de l'intégration, que nous rappelons sans le démontrer dans le cas particulier qui nous intéresse (cf. [8], page 105) :

Proposition 3.32.**Théorème de dérivation sous le signe \int**

Soit (E, \mathcal{E}, μ) un espace de probabilité, f une application de $E \times \mathbb{R}$ dans \mathbb{R} (ou \mathbb{C}). Si f vérifie les trois hypothèses suivantes :

1. pour μ -presque-tout $x \in E$, $t \in \mathbb{R} \mapsto f(x, t) \in \mathbb{R}$ est dérivable sur \mathbb{R} ,
2. pour tout $t \in \mathbb{R}$, $x \in E \mapsto f(x, t) \in \mathbb{R}$ est intégrable par rapport à μ ,
3. il existe une application g , intégrable par rapport à μ , telle que, pour μ -presque-tout $x \in E$ et pour tout $t \in \mathbb{R}$, $\left| \frac{\partial f}{\partial t}(x, t) \right| \leq g(x)$,

alors,

1. L'application $F : t \in \mathbb{R} \mapsto F(t) := \int_E f(x, t) d\mu(x) \in \mathbb{R}$ est dérivable sur \mathbb{R} ,
2. Pour tout $t \in \mathbb{R}$, $x \in E \mapsto \frac{\partial f}{\partial t}(x, t) \in \mathbb{R}$ est intégrable par rapport à μ ,
3. Pour tout $t \in \mathbb{R}$, $F'(t) = \int_E \frac{\partial f}{\partial t}(x, t) d\mu(x)$.

En appliquant ce résultat à l'application $(x, u) \in \mathbb{R} \times \mathbb{R} \mapsto \exp(i\langle x, u \rangle)$, et en raisonnant par récurrence, on obtient :

Proposition 3.33.

Si X est une variable aléatoire réelle telle que $\mathbb{E}(|X|^n) < +\infty$, i.e. $X \in L^n(\Omega, \mathcal{F}, \mathbb{P})$, où n est un entier naturel non nul, alors la fonction caractéristique Φ_X de X est continûment dérivable jusqu'à l'ordre n et on a, pour tout réel u :

$$\Phi_X^{(n)}(u) = i^n \int_{\mathbb{R}} x^n e^{iux} d\mathbb{P}_X(x) = i^n \mathbb{E}(X^n e^{iuX}).$$

En particulier $\Phi_X^{(n)}(0) = i^n \int_{\mathbb{R}} x^n d\mathbb{P}_X(x) = i^n \mathbb{E}(X^n)$.

Pour les vecteurs aléatoires, nous avons en particulier :

Proposition 3.34.

Si $X = (X_1, X_2, \dots, X_d)$ est un vecteur aléatoire de dimension d tel que $\mathbb{E}(|X|^2) < +\infty$, i.e. $X \in L^2(\Omega, \mathcal{F}, \mathbb{P})$, où d est un entier naturel non nul, alors, pour tout $k = 1, 2, \dots, d$ et $j = 1, 2, \dots, d$ on a $\mathbb{E}(X_k) = -i \frac{\partial \Phi_X}{\partial u_k}(0)$ et $\mathbb{E}(X_k X_j) = -\frac{\partial^2 \Phi_X}{\partial u_k \partial u_j}(0)$.

Pour le calcul des moments d'une variable aléatoire réelle, la proposition suivante est parfois utile :

Proposition 3.35.

Soient X une variable aléatoire réelle et n un entier naturel tels que la f.c. de X , Φ_X , soit dérivable en 0 à l'ordre n . Alors

1. $\mathbb{E}(X^{2p}) < +\infty$ où $2p$ est le plus grand entier pair inférieur à n . En particulier, X admet des moments jusqu'à l'ordre $2p$.
2. De plus, si le développement limité de Φ_X au voisinage de 0 à l'ordre n s'écrit

$$\Phi_X(u) = 1 + \sum_{k=1}^{k=n} a_k u^k + o(u^n) \text{ alors, pour tout } 1 \leq k \leq 2p, \mathbb{E}(X^k) = (-i)^k a_k k!.$$

Indications pour la démonstration (laissée en exercice non corrigé) : Supposons que Φ_X soit n -fois dérivable en 0. Faisons un raisonnement par récurrence finie sur k tel que $2k \leq n$, en prenant pour propriété de récurrence à l'ordre k , (\mathcal{P}_k) : " $M_{2k} < +\infty$."

• Montrons d'abord (\mathcal{P}_1) .

Pour simplifier notons f la partie-réelle de Φ_X et g sa partie-imaginaire. Montrer que g est impaire et f paire. En déduire $\Phi_X^{(2n)}(0) = f^{(2n)}(0)$. Montrer que la suite de fonctions $\left(2n^2(1 - f(\frac{1}{n}))\right)_{n \in \mathbb{N}}$ converge dans \mathbb{R} vers $-f''(0)$. En appliquant le lemme de Fatou (proposition 3.12, page 38) à la suite de fonctions $\left(2n^2(1 - \cos \frac{x}{n})\right)_{n \in \mathbb{N}}$, prouver que $M_2 \leq -f''(0) < +\infty$.

• Montrons l'hérédité de la propriété.

Supposons (\mathcal{P}_{k-1}) , pour un entier naturel k tel que $2k \leq n$, et montrons (\mathcal{P}_k) :

Notons h la partie-réelle de $\Phi_X^{(2k-2)}$. Montrer que $h(t) = (-1)^{k-1} \int_{\mathbb{R}} x^{2k-2} \cos(tx) d\mathbb{P}_X(x)$

et que $h''(0) = \Phi_X^{(2k)}(0)$. En déduire, en utilisant la convergence de la suite de fonctions $\left(2n^2(1 - h(\frac{1}{n}))\right)_{n \in \mathbb{N}}$, que $\lim_{n \rightarrow +\infty} (-1)^{k-1} \int_{\mathbb{R}} 2n^2 x^{2k-2} (1 - \cos \frac{x}{n}) d\mathbb{P}_X(x) = -\Phi_X^{(2k)}(0)$. En appli-

quant le lemme de Fatou, prouver que $M_{2k} \leq (-1)^k \Phi_X^{(2k)}(0) < +\infty$. Appliquer le théorème de récurrence finie sur k pour conclure.

La deuxième partie résulte de la proposition 3.33 précédente. \square

On notera que la connaissance des moments ne détermine pas en général la probabilité μ (cf. [14], page 89, exemple 11.1). Cependant, sous certaines hypothèses, la connaissance des moments détermine la probabilité μ . C'est le cas notamment dans le cas précisé par le résultat suivant :

Proposition 3.36.

Soit μ une probabilité sur \mathbb{R} de f.c. Φ . Supposons qu'il existe un réel $\alpha > 0$ tel que, pour tout entier naturel n , $\int_{\mathbb{R}} |x^n| d\mu(x) \leq \alpha^n$. Alors la f.c. Φ est développable en série entière sur \mathbb{R} au voisinage de 0.

Indications pour la démonstration (laissée en exercice non corrigé) : Justifier l'application de la formule de Taylor-Lagrange à l'ordre n dont on majorera le reste pour prouver que la f.c. est développable en série entière en 0 de rayon de convergence $+\infty$. Conclure. \square

Une conséquence immédiate de ce résultat est :

Proposition 3.37.

Soit μ une probabilité sur \mathbb{R} . Supposons qu'il existe un réel $\alpha > 0$ tel que, pour tout entier naturel n , $\int_{\mathbb{R}} |x^n| d\mu(x) \leq \alpha^n$. Si ν est une probabilité sur \mathbb{R} ayant les mêmes moments que μ , alors $\mu = \nu$.

Démonstration : D'après la proposition 3.36, les f.c. de μ et ν sont développables chacune en série entière. Comme, d'après la proposition 3.35, les coefficients des ces deux séries entières sont complètement déterminés par les moments des probabilités μ et ν , et comme ces deux probabilités ont les mêmes moments par hypothèse, on en déduit que les deux séries entières sont identiques et par suite que leurs f.c., qui sont respectivement égales aux sommes de ces deux séries, sont égales. On conclut par le théorème d'injectivité des f.c.. \square

3.4 Exercices de révision sur les chapitres I à III

Exercice 3.15. (Corrigé de l'exercice : page 161)

Soit X une v.a.r. de loi $\mathcal{N}_1(0, 1)$. Pour tout réel $a > 0$, on pose

$$X_a := X \mathbb{1}_{\{|X| \leq a\}} - X \mathbb{1}_{\{|X| > a\}}.$$

1. Vérifier que, pour tout réel $a > 0$ et toute application h de \mathbb{R} dans \mathbb{R} ,

$$h(X_a) = h(X) \mathbb{1}_{[0, a]}(|X|) + h(-X) \mathbb{1}_{]a, +\infty[}(|X|).$$

2. En déduire que, pour tout réel positif $a > 0$, la v.a.r. X_a suit la loi normale réduite centrée.

Exercice 3.16. (Corrigé de l'exercice : page 162)

Soit X une variable aléatoire de densité f définie pour tout nombre réel x , par

$$f(x) = \begin{cases} \frac{a}{x} & \text{si } -e \leq x < -1 \\ x + 1 - a & \text{si } -1 \leq x < 0 \\ 0 & \text{sinon} \end{cases}$$

où a est un nombre réel.

1. Calculer a et déterminer la fonction de répartition F de X .
2. Calculer l'espérance de la variable aléatoire X .

Exercice 3.17. (Corrigé de l'exercice : page 162)

Soit X une variable aléatoire normale centrée réduite. Préciser, dans chacun des cas ci-dessous, la loi de probabilité de la variable aléatoire Y définie en fonction de X

1. $Y = X^3$.
2. $Y = F(X)$ où F est la fonction de répartition de la variable X .

Exercice 3.18. (Corrigé de l'exercice : page 163)

Soit α un paramètre réel strictement positif et f l'application définie sur \mathbb{R} par

$$f(x) = \begin{cases} \frac{1}{2}\alpha e^{-\alpha x} & \text{si } x \geq 0 \\ \frac{1}{2}\alpha e^{\alpha x} & \text{si } x \leq 0 \end{cases}$$

1. Vérifier que f est bien une densité de probabilité.
2. Soit X une variable aléatoire réelle de densité f , quelle est la loi de la variable aléatoire $Y = |X|$? En déduire la variance de la variable X .

Exercice 3.19. (Corrigé de l'exercice : page 163)

Soit X une variable aléatoire réelle uniforme continue sur $[0, 1]$. Montrer que la relation $Y = -\frac{1}{\alpha} \ln(1 - X)$, où α est un paramètre réel strictement positif, définit presque-sûrement une variable aléatoire réelle. Quelle est la loi de la variable aléatoire Y ?

Exercice 3.20. (Corrigé de l'exercice : page 163)

Soit X une v.a.r. normale de loi $\mathcal{N}(m, \sigma^2)$, où m et σ sont des réels avec $\sigma > 0$.

1. Montrer que la fonction caractéristique de X peut s'exprimer à l'aide de la fonction caractéristique Φ de la loi de Gauss-Laplace standard $\mathcal{N}(0, 1)$.
2. En utilisant le théorème 3.32, page 57, de dérivation sous le signe \int , montrer que Φ est une solution particulière de l'équation différentielle du premier ordre $y'(t) + ty(t) = 0$. En déduire l'expression analytique de la fonction Φ , puis celle de la fonction caractéristique de la variable X .

Exercice 3.21. (Corrigé de l'exercice : page 164)

1. Montrer que si X et Y sont deux v.a.r. presque-sûrement égales, alors elles ont la même loi. Montrer que la réciproque est fausse.
2. Soient X et Y deux v.a.r. de même loi, g une application borélienne de \mathbb{R} dans \mathbb{R} . Montrer que les v.a.r. $g(X)$ et $g(Y)$ ont la même loi. Soit Z une autre v.a.r., montrer que les v.a.r. XZ et YZ n'ont pas nécessairement la même loi.

Exercice 3.22. (Corrigé de l'exercice : page 165)

Calculer la fonction caractéristique des variables aléatoires suivantes (pour les définitions se reporter au formulaire de l'annexe A, page 205) :

1. X suit une loi géométrique $\mathcal{G}(p)$.
2. Y suit une loi uniforme $\mathcal{U}([a, b])$.
3. Z suit une loi exponentielle $\mathcal{E}(\alpha)$.

Exercice 3.23. (Corrigé de l'exercice : page 165)

Soit μ une mesure de probabilité sur \mathbb{R} et D une partie partout dense dans \mathbb{R} (i.e. l'adhérence de D est égale à \mathbb{R}) telle que, pour tout $t \in D$, $F_\mu(t) \in \{0, 1\}$. Montrer qu'il existe $a \in \mathbb{R}$ tel que $\mu = \delta_a$.

Chapitre 4

Indépendance stochastique

Dans la suite, si A et B sont respectivement des parties de \mathbb{R}^n et \mathbb{R}^p , on posera avec un léger abus,

$$A \times B := \{(x_1, \dots, x_{n+p}) \in \mathbb{R}^{n+p} / (x_1, \dots, x_n) \in A \text{ et } (x_{n+1}, \dots, x_{n+p}) \in B\},$$

c-à-d $A \times B := (A \times \mathbb{R}^p) \cap (\mathbb{R}^n \times B)$. De même, si $a := (a_1, \dots, a_n) \in \mathbb{R}^n$ et $b := (b_1, \dots, b_p) \in \mathbb{R}^p$, on notera $(a, b) := (a_1, \dots, a_n, b_1, \dots, b_p)$ considéré comme élément de \mathbb{R}^{n+p} . On dit que (a, b) est obtenu par **concaténation** de a et b .

4.1 Intégration sur \mathbb{R}^{n+p}

Pour introduire la problématique de l'intégration sur \mathbb{R}^{n+p} , commençons par considérer les deux situations suivantes :

1) Soient $a := (a_1, \dots, a_n) \in \mathbb{R}^n$ et $b := (b_1, \dots, b_p) \in \mathbb{R}^p$. On remarque qu'avec les notations précisées en préliminaires on peut écrire, pour tous boréliens A et B respectivement de \mathbb{R}^n et \mathbb{R}^p ,

$$\delta_{(a,b)}(A \times B) = \mathbb{1}_{A \times B}(a, b) = \mathbb{1}_A(a) \mathbb{1}_B(b) = \delta_a(A) \delta_b(B).$$

2) De même, considérons la mesure de Lebesgue sur \mathbb{R}^2 . Par définition pour tous réels a, b, c, d , avec $a < b$ et $c < d$, $\lambda^{(2)}([a, b] \times [c, d]) = \lambda([a, b]) \lambda([c, d])$ où λ désigne la mesure de Lebesgue sur \mathbb{R} . Plus généralement on montre que, pour tous boréliens A et B de \mathbb{R} , $\lambda^{(2)}(A \times B) = \lambda(A) \lambda(B)$.

Généralisons ces situations en considérant le problème suivant : *Étant donné deux mesures μ et ν respectivement sur \mathbb{R}^n et \mathbb{R}^p , existe-t-il une mesure α sur \mathbb{R}^{n+p} telle que, pour tous boréliens A et B respectivement de \mathbb{R}^n et \mathbb{R}^p , $\alpha(A \times B) = \mu(A) \nu(B)$? Si oui, y a-t-il unicité de la mesure α ?*

On montre que dans le cas où μ et ν sont des probabilités la réponse est positive. Dans le cas des mesures plus générales ce n'est plus nécessairement vrai, cependant c'est encore vrai pour les mesures de Lebesgue. Plus précisément, on admettra le résultat suivant :

Proposition 4.1.

Soit μ (respectivement ν) une probabilité ou la mesure de Lebesgue sur \mathbb{R}^n (respectivement sur \mathbb{R}^p) alors il existe une unique mesure sur \mathbb{R}^{n+p} , notée $\mu \otimes \nu$, telle que, pour tous boréliens A et B respectivement de \mathbb{R}^n et \mathbb{R}^p , $\mu \otimes \nu(A \times B) = \mu(A) \nu(B)$.

Définition 4.1.

On dit que $\mu \otimes \nu$ est le **produit des mesures** μ et ν . On dit aussi que $\mu \otimes \nu$ est une **mesure-produit** sur \mathbb{R}^{n+p} .

On notera que toutes les mesures sur \mathbb{R}^{n+p} ne sont pas nécessairement des mesures-produit.

Exemples 4.1.

On admettra que, pour tous entiers $n \geq 1$ et $p \geq 1$, $\lambda^{(n+p)} = \lambda^{(n)} \otimes \lambda^{(p)}$. En particulier on utilisera souvent la relation $\lambda^{(2)} = \lambda \otimes \lambda$. \square

Le théorème suivant, énoncé sous la forme la plus utilisée dans ce cours, est un cas particulier du théorème de Tonelli vu en théorie de la mesure et de l'intégration. Il donne un procédé de calcul des intégrales sur \mathbb{R}^{n+p} . Il permettra par un procédé de récurrence d'en déduire une méthode de calcul des intégrales sur \mathbb{R}^d .

Proposition 4.2.**Théorème de Tonelli** (admis)

Soient μ une probabilité ou la mesure de Lebesgue sur \mathbb{R}^n , ν une probabilité ou la mesure de Lebesgue sur \mathbb{R}^p et f une application borélienne positive de \mathbb{R}^{n+p} dans $[0, +\infty]$, alors

1. Pour tous $y \in \mathbb{R}^p$ et $x \in \mathbb{R}^n$, les applications partielles

$$u \in \mathbb{R}^n \mapsto f(u, y) \in [0, +\infty] \text{ et } v \in \mathbb{R}^p \mapsto f(x, v) \in [0, +\infty]$$

sont boréliennes positives.

2. Les applications

$$x \in \mathbb{R}^n \mapsto \int_{\mathbb{R}^p} f(x, v) d\nu(v) \in [0, +\infty] \text{ et } y \in \mathbb{R}^p \mapsto \int_{\mathbb{R}^n} f(u, y) d\mu(u) \in [0, +\infty]$$

sont boréliennes positives,

3. On a les égalités

$$\begin{aligned} \int_{\mathbb{R}^{n+p}} f d(\mu \otimes \nu) &= \int_{\mathbb{R}^p} \left(\int_{\mathbb{R}^n} f(x, y) d\mu(x) \right) d\nu(y) \\ &= \int_{\mathbb{R}^n} \left(\int_{\mathbb{R}^p} f(x, y) d\nu(y) \right) d\mu(x). \end{aligned}$$

Le théorème de Tonelli permet en pratique de ramener le calcul d'une **intégrale multiple**, i.e. sur \mathbb{R}^d , au calcul d'une succession de **d intégrales simples**, i.e. sur \mathbb{R} , pour lesquelles on peut appliquer séparément les règles d'intégration déjà vues au chapitre III.

Ce résultat est encore vrai pour les applications f de signe quelconque à condition qu'elles soient supposées intégrables sur \mathbb{R}^{n+p} suivant la mesure-produit $\mu \otimes \nu$. Il est alors connu sous le nom de **théorème de Fubini** :

Proposition 4.3.**Théorème de Fubini**

Soient μ une probabilité ou la mesure de Lebesgue sur \mathbb{R}^n , ν une probabilité ou la mesure de Lebesgue sur \mathbb{R}^p et f une application $(\mu \otimes \nu)$ -intégrable de \mathbb{R}^{n+p} dans $\overline{\mathbb{R}}$, alors,

1. Pour μ -presque-tout $x \in \mathbb{R}^n$, l'application $y \in \mathbb{R}^p \mapsto f(x, y) \in \overline{\mathbb{R}}$ est ν -intégrable, et, pour ν -presque-tout $y \in \mathbb{R}^p$, l'application $x \in \mathbb{R}^n \mapsto f(x, y) \in \overline{\mathbb{R}}$ est μ -intégrable.
2. L'application $y \in \mathbb{R}^p \mapsto \int_{\mathbb{R}^n} f(x, y) d\mu(x) \in \mathbb{R}$ est définie ν -presque-partout et ν -intégrable, et l'application $x \in \mathbb{R}^n \mapsto \int_{\mathbb{R}^p} f(x, y) d\nu(y) \in \mathbb{R}$ est définie μ -presque-partout et μ -intégrable.
3. L'intégrale de f par rapport à la mesure-produit $\mu \otimes \nu$ est donnée par :

$$\begin{aligned} \int_{\mathbb{R}^{n+p}} f d(\mu \otimes \nu) &= \int_{\mathbb{R}^n} \left[\int_{\mathbb{R}^p} f(x, y) d\nu(y) \right] d\mu(x) \\ &= \int_{\mathbb{R}^p} \left[\int_{\mathbb{R}^n} f(x, y) d\mu(x) \right] d\nu(y). \end{aligned}$$

D'autres notations sont utilisées dans les ouvrages pour noter une intégrale suivant une mesure-produit. On trouvera indifféremment

$$\begin{aligned} \int_{\mathbb{R}^{n+p}} f d(\mu \otimes \nu) &= \int_{\mathbb{R}^{n+p}} f(x, y) d(\mu \otimes \nu)(x, y) = \int_{\mathbb{R}^{n+p}} f(x, y) d\mu(x) \otimes d\nu(y) \\ &= \int_{\mathbb{R}^{n+p}} f(x, y) d\mu(x) d\nu(y). \end{aligned}$$

Exemples 4.2.

Considérons l'application f définie sur \mathbb{R}^2 par $f(x, y) := \mathbb{1}_{]0, +\infty[}(x) \mathbb{1}_{]a, b[}(y) e^{-xy}$ où a et b sont des réels tels que $0 < a < b$. C'est une fonction borélienne positive sur \mathbb{R}^2 . Appliquons le théorème de Tonelli à f et à la mesure $\lambda^{(2)} = \lambda \otimes \lambda$,

$$\begin{aligned} \int_{\mathbb{R}^2} f d\lambda^{(2)} &= \int_{\mathbb{R}} \mathbb{1}_{]0, +\infty[}(x) \left(\int_{\mathbb{R}} \mathbb{1}_{]a, b[}(y) e^{-xy} d\lambda(y) \right) d\lambda(x) \\ &= \int_{\mathbb{R}} \mathbb{1}_{]a, b[}(y) \left(\int_{\mathbb{R}} \mathbb{1}_{]0, +\infty[}(x) e^{-xy} d\lambda(x) \right) d\lambda(y). \end{aligned}$$

Ce qui donne en utilisant la règle d'intégration suivant la mesure de Lebesgue sur \mathbb{R} des fonctions boréliennes positives,

$$\begin{aligned} \int_{\mathbb{R}^2} f d\lambda^{(2)} &= \int_0^{+\infty} \frac{e^{-ax} - e^{-bx}}{x} dx \\ &= \int_{\mathbb{R}} \mathbb{1}_{]a, b[}(y) \frac{1}{y} d\lambda(y) = \int_a^b \frac{1}{y} dy = \ln \left(\frac{b}{a} \right). \end{aligned}$$

On a ainsi par la même occasion établi la valeur de l'intégrale généralisée au sens de Riemann

$$\int_0^{+\infty} \frac{e^{-ax} - e^{-bx}}{x} dx = \ln \left(\frac{b}{a} \right). \quad \square$$

Pour compléter les techniques d'intégration suivant la mesure de Lebesgue sur \mathbb{R}^d , le théorème suivant, dit **de changement de variable**, est très utile.

Rappelons qu'un **difféomorphisme** F de classe \mathcal{C}^1 d'un ouvert U sur un ouvert V de \mathbb{R}^d est une application bijective, différentiable de U sur V telle que F^{-1} soit différentiable avec les applications différentielles de F et F^{-1} continues. De plus si f_1, \dots, f_d sont les applications-composantes de F , on appelle **jacobien** de F au point $u \in U$ le déterminant de la matrice carrée d'ordre d de coefficient général $\partial_j f_i(u)$ où $\partial_j f_i(u)$ désigne la dérivée-partielle de l'application f_i par rapport à la $j^{\text{ième}}$ variable calculée au point $u \in U$.

Proposition 4.4.

Théorème de changement de variable dans \mathbb{R}^d (admis)

Soient T un difféomorphisme de classe \mathcal{C}^1 d'un ouvert U sur un ouvert V de \mathbb{R}^d et f une application borélienne de \mathbb{R}^d dans $\overline{\mathbb{R}}$. Notons $J(v)$ le jacobien de T^{-1} au point $v \in V$.

1. Si f est à valeurs dans $[0, +\infty]$, alors $x \in \mathbb{R}^d \mapsto \mathbb{1}_V(x)f(T^{-1}(x))|J(x)|$ est une application borélienne positive.
2. Si l'application $x \in \mathbb{R}^d \mapsto \mathbb{1}_U(x)f(x)$ est $\lambda^{(d)}$ -intégrable sur \mathbb{R}^d , alors l'application $x \in \mathbb{R}^d \mapsto \mathbb{1}_V(x)f(T^{-1}(x))|J(x)|$ est $\lambda^{(d)}$ -intégrable sur \mathbb{R}^d .

De plus, dans les deux cas ci-dessus

$$\int_{\mathbb{R}^d} \mathbb{1}_U(u)f(u)d\lambda^{(d)}(u) = \int_{\mathbb{R}^d} \mathbb{1}_V(v)f(T^{-1}(v))|J(v)|d\lambda^{(d)}(v),$$

qu'on écrit encore :

$$\int_U f(u)d\lambda^{(d)}(u) = \int_V f(T^{-1}(v))|J(v)|d\lambda^{(d)}(v).$$

On dit alors qu'on a effectué le **changement de variable** $v := T(u)$ où v désigne le vecteur des "nouvelles" variables et u celui des "anciennes". On remarquera surtout que le jacobien utilisé est celui de la transformation exprimant les "anciennes" variables en fonction des "nouvelles" c-à-d de T^{-1} .

Illustrons par un exemple l'utilisation des théorèmes de Tonelli et de changement de variable.

Exemples 4.3.

Montrons que

$$I := \int_{\mathbb{R}^2} \mathbb{1}_{[0, +\infty]^2}(x, y)e^{-(x^2+y^2)}d\lambda^{(2)}(x, y) = \frac{\pi}{4}.$$

En effet, ici $U :=]0, +\infty[^2$, $u := (x, y)$. Considérons le changement de variable $x = r \cos \theta$

et $y = r \sin \theta$ où $v := (r, \theta) \in V :=]0, +\infty[\times]0, \frac{\pi}{2}[$. Comme $|J(v)| = r$, il vient

$$\begin{aligned} I &= \int_{\mathbb{R}^2} \mathbb{1}_{]0, +\infty[\times]0, \frac{\pi}{2}[}(r, \theta) e^{-r^2} r d\lambda^{(2)}(r, \theta) \\ &= \int_{\mathbb{R}} \mathbb{1}_{]0, +\infty[}(r) e^{-r^2} r \left(\int_{\mathbb{R}} \mathbb{1}_{]0, \frac{\pi}{2}[}(\theta) d\lambda(\theta) \right) d\lambda(r) \\ &= \left(\int_{\mathbb{R}} \mathbb{1}_{]0, +\infty[}(r) e^{-r^2} r d\lambda(r) \right) \left(\int_{\mathbb{R}} \mathbb{1}_{]0, \frac{\pi}{2}[}(\theta) d\lambda(\theta) \right) = \left(\int_0^{+\infty} e^{-r^2} r dr \right) \frac{\pi}{2} = \frac{\pi}{4}. \end{aligned}$$

Dans le calcul précédent on a utilisé le théorème de Tonelli et la règle d'intégration suivant la mesure de Lebesgue sur \mathbb{R} . \square

Exercice 4.1. (Corrigé de l'exercice : page 165)

1) Dédurre du résultat de l'exemple précédent les deux relations

$$\int_0^{+\infty} e^{-x^2} dx = \frac{1}{2} \sqrt{\pi} \text{ et } \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{1}{2}x^2} dx = 1.$$

2) En remarquant que $x^2 + 2xy + 2y^2 = (x + y)^2 + y^2$, déduire de la question précédente que $\int_{\mathbb{R}^2} e^{-(x^2+2xy+2y^2)} d\lambda^{(2)}(x, y) = \pi$.

Le théorème de Fubini ou de Tonelli sont en particulier très utiles dans les calculs faisant intervenir des probabilités à densité définies sur \mathbb{R}^d . Donnons un exemple d'utilisation de la proposition 3.8, page 35, dans le calcul de lois de probabilité sur \mathbb{R}^2 .

Exemples 4.4.

Soit (X, Y) un vecteur aléatoire de \mathbb{R}^2 dont la loi admet pour densité la fonction définie sur \mathbb{R}^2 $\rho := \frac{1}{2} \mathbb{1}_{\Delta}$ où $\Delta := \{(x, y) \in \mathbb{R}^2 / |x| + |y| \leq 1\}$. Cherchons la loi de la variable aléatoire réelle X .

Soit A un borélien de \mathbb{R} . Calculons $\mathbb{P}_X(A)$. Par définition de la notion de loi et comme $\{X \in A\} = \{(X, Y) \in A \times \mathbb{R}\}$,

$$\begin{aligned} \mathbb{P}_X(A) &= \mathbb{P}(X \in A) = \mathbb{P}((X, Y) \in A \times \mathbb{R}) \\ &= \mathbb{P}_{(X, Y)}(A \times \mathbb{R}) = \int_{\mathbb{R}^2} \mathbb{1}_A(x) \mathbb{1}_{\mathbb{R}}(y) d\mathbb{P}_{(X, Y)}(x, y). \end{aligned}$$

D'après la règle d'intégration des mesures à densité sur \mathbb{R}^2 puis par application du théorème de Tonelli à $\lambda^{(2)} = \lambda \otimes \lambda$,

$$\begin{aligned} \int_{\mathbb{R}^2} \mathbb{1}_A(x) \mathbb{1}_{\mathbb{R}}(y) d\mathbb{P}_{(X, Y)}(x, y) &= \int_{\mathbb{R}^2} \mathbb{1}_A(x) \mathbb{1}_{\mathbb{R}}(y) \frac{1}{2} \mathbb{1}_{\Delta}(x, y) d\lambda^{(2)}(x, y) \\ &= \int_{\mathbb{R}} \mathbb{1}_A(x) \left(\int_{\mathbb{R}} \frac{1}{2} \mathbb{1}_{\Delta}(x, y) d\lambda(y) \right) d\lambda(x) \\ &= \int_{\mathbb{R}} \mathbb{1}_A(x) \chi(x) d\lambda(x) \end{aligned}$$

où on a posé après le calcul de l'intégrale

$$\chi(x) := \int_{\mathbb{R}} \frac{1}{2} \mathbb{1}_{\Delta}(x, y) d\lambda(y) = (1 - |x|) \mathbb{1}_{[-1, 1]}(x).$$

Ce qui prouve que la variable aléatoire réelle X a pour loi la probabilité définie sur \mathbb{R} par la densité χ . On trouve par symétrie des rôles joués par X et Y que Y a même loi que X (cela ne signifie pas que $X = Y$!). \square

Exercice 4.2. (Corrigé de l'exercice : page 166)

Citer un exemple simple de variables aléatoires réelles X et Y distinctes et de même loi.

En fait on vient de montrer, sur un cas particulier, le résultat important ci-dessous qui affirme que si un vecteur aléatoire admet une densité, alors ses composantes sont des variables aléatoires réelles à densité. La réciproque est fautive en général comme le montre le contre-exemple proposé dans l'exercice 4.4 ci-dessous. Plus précisément :

Proposition 4.5.

Si $X := (X_1, \dots, X_d)$ est un vecteur aléatoire de densité ρ sur \mathbb{R}^d , alors, pour tout entier $1 \leq k \leq d$, la v.a.r. X_k admet pour densité l'application χ_k définie sur \mathbb{R} par

$$\chi_k(t) := \int_{\mathbb{R}^{d-1}} \rho(x_1, \dots, x_{k-1}, t, x_{k+1}, \dots, x_d) d\lambda^{(d-1)}(x_1, \dots, x_{k-1}, x_{k+1}, \dots, x_d).$$

La réciproque est fautive.

Exercice 4.3. (Corrigé de l'exercice : page 166)

Démontrer cette proposition pour $d = 3$ en adaptant la démarche développée dans l'exemple précédent dans lequel on avait $d = 2$.

Exercice 4.4. (Corrigé de l'exercice : page 167)

Soit X une variable aléatoire réelle de loi normale standard $\mathcal{N}_1(0, 1)$. On pose $\Delta := \{(x, y) \in \mathbb{R}^2, y = x\}$, Prouver que $\mathbb{P}_{(X, X)}(\Delta) = 1$. En supposant que le vecteur aléatoire (X, X) , admette une densité sur \mathbb{R}^2 , prouver que, sous cette hypothèse, $\mathbb{P}_{(X, X)}(\Delta) = 0$. En déduire que le vecteur aléatoire (X, X) de dimension 2 n'admet pas de densité sur \mathbb{R}^2 et que la réciproque de la proposition 4.5 est fautive.

4.2 Indépendance de vecteurs aléatoires, d'événements, de tribus

4.2.1 Indépendance de vecteurs aléatoires

Exemples 4.5.

Revenons à l'exemple 4.4 de la page 65 pour montrer que, dans ce cas, $\mathbb{P}_{(X, Y)} \neq \mathbb{P}_X \otimes \mathbb{P}_Y$.

En effet, considérons le borélien de \mathbb{R}^2 , $G :=]\frac{1}{2}, 1[\times]\frac{1}{2}, 1[$, et comparons $\mathbb{P}_{(X, Y)}(G)$ avec $\mathbb{P}_X \otimes \mathbb{P}_Y(G)$. Il vient d'une part,

$$\mathbb{P}_{(X, Y)}(G) = \int_{\mathbb{R}^2} \mathbb{1}_G(x, y) d\mathbb{P}_{(X, Y)}(x, y) = \int_{\mathbb{R}^2} \mathbb{1}_G(x, y) \frac{1}{2} \mathbb{1}_{\Delta}(x, y) d\lambda^{(2)}(x, y) = 0$$

car $\mathbb{1}_G \mathbb{1}_\Delta$ est l'application nulle sur \mathbb{R}^2 , et d'autre part,

$$\mathbb{P}_X \otimes \mathbb{P}_Y(G) = \mathbb{P}_X \left(\left] \frac{1}{2}, 1[\right) \mathbb{P}_Y \left(\left] \frac{1}{2}, 1[\right) .$$

Or

$$\mathbb{P}_X \left(\left] \frac{1}{2}, 1[\right) = \int_{\mathbb{R}} \mathbb{1}_{\left] \frac{1}{2}, 1[} d\mathbb{P}_X = \int_{\mathbb{R}} (1 - |x|) \mathbb{1}_{[-1, 1]}(x) \mathbb{1}_{\left] \frac{1}{2}, 1[}(x) d\lambda(x) = \int_{\frac{1}{2}}^1 (1 - x) dx = \frac{1}{8} .$$

Ce qui montre que $0 = \mathbb{P}_{(X, Y)}(G) \neq \mathbb{P}_X \otimes \mathbb{P}_Y(G) = \left(\frac{1}{8}\right)^2$ et prouve que le produit des lois de X et Y , qui est une probabilité sur \mathbb{R}^2 , n'est pas égal à la loi du vecteur (X, Y) i.e. $\mathbb{P}_{(X, Y)} \neq \mathbb{P}_X \otimes \mathbb{P}_Y$. \square

Les situations pour lesquelles on aura l'égalité seront celles où on dira qu'il y a indépendance des variables suivant la définition :

Définition 4.2.

Une suite finie de vecteurs aléatoires (X_1, \dots, X_n) , de dimensions quelconques (éventuellement distinctes), est dite **indépendante (relativement à \mathbb{P})** si

$$\mathbb{P}_{(X_1, \dots, X_n)} = \mathbb{P}_{X_1} \otimes \mathbb{P}_{X_2} \otimes \dots \otimes \mathbb{P}_{X_n} .$$

On dit aussi, par abus, que les vecteurs aléatoires X_1, \dots, X_n sont indépendants. Il s'agit là d'un abus car l'indépendance est une propriété de la suite (X_1, \dots, X_n) et non de chacune des variables aléatoires réelles X_k .

Définition 4.3.

La loi du vecteur aléatoire concaténé $X := (X_1, \dots, X_n)$ est dite aussi **loi conjointe** des vecteurs X_1, \dots, X_n . Pour tout entier $k = 1, \dots, n$, la loi du vecteur aléatoire X_k s'appelle alors la **loi marginale** de X de rang k .

Exemples 4.6.

Avec cette terminologie, on peut énoncer le résultat de l'exemple 4.5 précédent en exprimant que le couple de variables aléatoires réelles (X, Y) n'est pas indépendant.

Dans le chapitre sur les convergences de suite de variables aléatoires réelles on aura besoin de la définition suivante :

Définition 4.4.

Une suite infinie $(X_k)_{k \in \mathbb{N}}$ de vecteurs aléatoires est dite **indépendante (relativement à \mathbb{P})** si toute sous-famille finie est indépendante relativement à \mathbb{P} .

On montre que si $(\mu_k)_{k \in I}$, où $I \subseteq \mathbb{N}$, est une suite (finie ou infinie) de probabilités sur \mathbb{R} , on peut toujours construire un espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$ et une suite indépendante $(X_k)_{k \in I}$ de variables aléatoires réelles définies sur $(\Omega, \mathcal{F}, \mathbb{P})$ telle que, pour tout $k \in I$, μ_k soit la loi de la variable aléatoire réelle X_k .

On peut vérifier aisément qu'une suite infinie de vecteurs aléatoires $(X_k)_{k \in \mathbb{N}}$ est indépendante si, et seulement si, pour tout entier $n \in \mathbb{N}$, la suite finie (X_0, \dots, X_n) est indépendante.

Exercice 4.5. (Corrigé de l'exercice : page 167)

Soit (U, V) un couple indépendant de variables aléatoires réelles dont la loi de chaque composante est la loi uniforme $\mathcal{U}([0, 1])$ de densité $\mathbb{1}_{[0,1]}$. On définit les variables aléatoires $X := \sqrt{-2 \ln U} \cos(2\pi V)$ et $Y := \sqrt{-2 \ln U} \sin(2\pi V)$.

1. Montrer que le vecteur aléatoire (X, Y) de dimension 2 admet pour densité l'application définie sur \mathbb{R}^2 par $\rho(x, y) := \frac{1}{2\pi} e^{-\frac{x^2+y^2}{2}}$. On dit qu'il suit une **loi normale de dimension 2**. Les vecteurs de ce type seront étudiés plus en détail au chapitre suivant.
2. Montrer que les variables aléatoires réelles X et Y ont toutes les deux pour loi $\mathcal{N}_1(0, 1)$.
3. Dédurre des questions précédentes que le couple de variables aléatoires réelles (X, Y) est indépendant.

4.2.2 Critères d'indépendance de vecteurs aléatoires

Les propositions qui suivent ont pour but de donner des critères permettant de reconnaître si une suite finie de vecteurs aléatoires est indépendante.

Proposition 4.6.

Soit (X_1, \dots, X_n) une suite de vecteurs aléatoires de dimensions respectives d_1, \dots, d_n . Les assertions suivantes sont équivalentes :

1. La suite (X_1, \dots, X_n) est indépendante.
2. Pour tout entier $1 \leq k \leq n$ et tout borélien B_k de \mathbb{R}^{d_k} ,

$$\mathbb{P}_{(X_1, \dots, X_n)}(B_1 \times B_2 \times \dots \times B_n) = \mathbb{P}_{X_1}(B_1) \mathbb{P}_{X_2}(B_2) \dots \mathbb{P}_{X_n}(B_n).$$

3. Pour tout entier $1 \leq k \leq n$ et tout borélien B_k de \mathbb{R}^{d_k} ,

$$\mathbb{P}[X_1, \dots, X_n] \in B_1 \times \dots \times B_n = \mathbb{P}(X_1 \in B_1) \dots \mathbb{P}(X_n \in B_n).$$

4. Pour tout entier $1 \leq k \leq n$ et tout borélien B_k de \mathbb{R}^{d_k} ,

$$\mathbb{P}[\{X_1 \in B_1\} \cap \dots \cap \{X_n \in B_n\}] = \mathbb{P}(X_1 \in B_1) \dots \mathbb{P}(X_n \in B_n).$$

Démonstration : "(1) implique (2)" résulte de la définition des produits de lois. "(2) implique (3)" résulte de la définition de la notion de loi et des notations. "(3) implique (4)" résulte de la relation ensembliste immédiate à vérifier :

$$\{(X_1, \dots, X_n) \in B_1 \times \dots \times B_n\} = \{X_1 \in B_1\} \cap \{X_2 \in B_2\} \cap \dots \cap \{X_n \in B_n\}.$$

L'implication "(4) implique (1)" résulte de ce que la probabilité $\mathbb{P}_{(X_1, \dots, X_n)}$ vérifie la propriété caractéristique des mesures-produits : Pour tout entier $1 \leq k \leq n$ et tout borélien B_k de \mathbb{R}^{d_k} , $\mathbb{P}_{(X_1, \dots, X_n)}(B_1 \times B_2 \times \dots \times B_n) = \mathbb{P}_{X_1}(B_1) \mathbb{P}_{X_2}(B_2) \dots \mathbb{P}_{X_n}(B_n)$. Par unicité de la mesure-produit, on en déduit que $\mathbb{P}_{(X_1, \dots, X_n)} = \mathbb{P}_{X_1} \otimes \mathbb{P}_{X_2} \otimes \dots \otimes \mathbb{P}_{X_n}$. Ce qui prouve (1). \square

Dans les cas où on manipule des variables aléatoires réelles discrètes la proposition précédente a pour corollaire le critère d'indépendance ci-dessous. Pour simplifier on supposera que les variables aléatoires réelles sont portées par \mathbb{N} mais le résultat se généralise aux variables aléatoires réelles portées par un ensemble dénombrable quelconque de \mathbb{R} , notamment \mathbb{Z} .

Proposition 4.7.

Critère des v.a.r. discrètes

Soit (X_1, \dots, X_n) une suite de variables aléatoires réelles discrètes portées par \mathbb{N} , alors la suite (X_1, \dots, X_n) est indépendante si, et seulement si, pour tout $(k_1, \dots, k_n) \in \mathbb{N}^n$,

$$\mathbb{P}(\{X_1 = k_1\} \cap \dots \cap \{X_n = k_n\}) = \mathbb{P}(X_1 = k_1) \dots \mathbb{P}(X_n = k_n).$$

Démonstration : La condition nécessaire résulte de l'implication "(1) implique (4)" de la proposition précédente où on a pris $B_i := \{k_i\}$, $i = 1, \dots, n$, qui sont bien des boréliens de \mathbb{R} .

La condition suffisante résulte du fait que $\mathbb{P}_{(X_1, \dots, X_n)}$ est une probabilité discrète sur \mathbb{R}^n , elle est donc entièrement déterminée par la connaissance des nombres

$$\mathbb{P}_{(X_1, \dots, X_n)}(\{k_1\} \times \{k_2\} \times \dots \times \{k_n\})$$

qui sont, par hypothèse, égaux à

$$\mathbb{P}(X_1 = k_1) \mathbb{P}(X_2 = k_2) \dots \mathbb{P}(X_n = k_n).$$

De même on vérifie que $\mathbb{P}_{X_1} \otimes \dots \otimes \mathbb{P}_{X_n}$ est une probabilités portée par \mathbb{N}^n donc discrète. Elle est donc entièrement déterminée par la connaissance des nombres

$$\mathbb{P}_{X_1} \otimes \dots \otimes \mathbb{P}_{X_n}(\{k_1\} \times \{k_2\} \times \dots \times \{k_n\}) = \mathbb{P}(X_1 = k_1) \mathbb{P}(X_2 = k_2) \dots \mathbb{P}(X_n = k_n).$$

Par suite

$$\mathbb{P}_{(X_1, \dots, X_n)} = \mathbb{P}_{X_1} \otimes \dots \otimes \mathbb{P}_{X_n}.$$

Ce qui prouve l'indépendance de la suite (X_1, \dots, X_n) . \square

Exemples 4.7.

Reprenons les notations de l'exemple 3.12, page 42. $X := (X_1, X_2)$ est un vecteur aléatoire de dimension 2 de loi

$$\mathbb{P}_X := \sum_{k \geq 1, l \geq 1} \frac{1}{2^{k+l}} \delta_{(k,l)}.$$

Par suite, pour tous $k \in \mathbb{N}^*$ et $l \in \mathbb{N}^*$,

$$\mathbb{P}[\{X_1 = k\} \cap \{X_2 = l\}] = \frac{1}{2^{k+l}} \text{ et } \mathbb{P}(X_1 = k) = \mathbb{P}(X_2 = k) = \frac{1}{2^k}.$$

Ce qui montre que, pour tous $k \in \mathbb{N}^*$ et $l \in \mathbb{N}^*$,

$$\mathbb{P}[\{X_1 = k\} \cap \{X_2 = l\}] = \mathbb{P}(X_1 = k) \mathbb{P}(X_2 = l)$$

et prouve ainsi l'indépendance de la suite de variables aléatoires réelles (X_1, X_2) . \square

On vient d'établir un critère d'indépendance pour une grande famille de probabilités, celle des probabilités discrètes, donnons maintenant un critère pour une autre grande famille de probabilités, celle des probabilités à densité.

Proposition 4.8.

Critère des v.a.r. à densité

1. Si (X_1, \dots, X_n) est une suite indépendante de variables aléatoires réelles telle que, pour tout $k = 1, \dots, n$, la variable aléatoire réelle X_k soit de densité ρ_k sur \mathbb{R} , alors le vecteur aléatoire $X := (X_1, \dots, X_n)$ de dimension n admet pour densité sur \mathbb{R}^n l'application

$$\rho : (x_1, \dots, x_n) \in \mathbb{R}^n \mapsto \rho(x_1, \dots, x_n) := \rho_1(x_1)\rho_2(x_2) \cdots \rho_n(x_n) \in [0, +\infty].$$

2. Réciproquement, si $X := (X_1, \dots, X_n)$ est un vecteur aléatoire de dimension n admettant pour densité sur \mathbb{R}^n une application ρ définie par une relation de la forme $\rho(x_1, \dots, x_n) := g_1(x_1)g_2(x_2) \cdots g_n(x_n)$, où, pour tout $k = 1, \dots, n$, g_k est une application borélienne positive de \mathbb{R} dans $[0, +\infty]$, alors la suite de variables aléatoires réelles (X_1, \dots, X_n) est indépendante et, pour tout $k = 1, \dots, n$, la variable aléatoire réelle X_k admet pour densité sur \mathbb{R} l'application

$$\chi_k : t \in \mathbb{R} \mapsto \chi_k(t) := \frac{g_k(t)}{\int_{\mathbb{R}} g_k d\lambda} \in [0, +\infty].$$

Démonstration : Pour simplifier les notations, faisons la démonstration dans le cas $n = 3$.

1) Soient (X, Y, Z) une suite indépendante de variables aléatoires réelles et h une application borélienne positive de \mathbb{R}^3 dans $[0, +\infty]$. Par le théorème du transfert, l'indépendance de la suite (X, Y, Z) i.e. $\mathbb{P}_{(X,Y,Z)} = \mathbb{P}_X \otimes \mathbb{P}_Y \otimes \mathbb{P}_Z$, puis en appliquant le théorème de Tonelli à $\mathbb{P}_X \otimes \mathbb{P}_Y \otimes \mathbb{P}_Z$, la définition des lois à densité et à nouveau le théorème de Tonelli en sens inverse à $\lambda \otimes \lambda \otimes \lambda = \lambda^{(3)}$, il vient successivement

$$\begin{aligned} \mathbb{E}[h(X, Y, Z)] &= \int_{\mathbb{R}^3} h(u, v, w) d\mathbb{P}_X \otimes \mathbb{P}_Y \otimes \mathbb{P}_Z(u, v, w) \\ &= \int_{\mathbb{R}} \left(\int_{\mathbb{R}} \left(\int_{\mathbb{R}} h(u, v, w) d\mathbb{P}_X(u) \right) d\mathbb{P}_Y(v) \right) d\mathbb{P}_Z(w) \\ &= \int_{\mathbb{R}} \left(\int_{\mathbb{R}} \left(\int_{\mathbb{R}} h(u, v, w) \rho_1(u) d\lambda(u) \right) \rho_2(v) d\lambda(v) \right) \rho_3(w) d\lambda(w) \\ &= \int_{\mathbb{R}^3} h(u, v, w) \rho_1(u) \rho_2(v) \rho_3(w) d\lambda^{(3)}(u, v, w). \end{aligned}$$

Ce qui prouve que la loi du vecteur aléatoire (X, Y, Z) admet pour densité sur \mathbb{R}^3 l'application ρ définie par $\rho(u, v, w) := \rho_1(u)\rho_2(v)\rho_3(w)$.

2) Supposons que le vecteur aléatoire (X, Y, Z) admette pour densité sur \mathbb{R}^3 l'application ρ définie par $\rho(u, v, w) := g_1(u)g_2(v)g_3(w)$.

Commençons par étudier la loi de la première composante X .

D'après la proposition 4.5, la variable aléatoire réelle X admet pour densité sur \mathbb{R} l'application χ_1 définie par $\chi_1(t) := \int_{\mathbb{R}^2} \rho(t, u, v) d\lambda^{(2)}(u, v)$. Par application du théorème de Tonelli à

$\lambda^{(2)} = \lambda \otimes \lambda$, il vient

$$\begin{aligned}\chi_1(t) &= \int_{\mathbb{R}^2} g_1(t)g_2(u)g_3(v)d\lambda^{(2)}(u, v) = g_1(t) \int_{\mathbb{R}} \left(g_2(u) \int_{\mathbb{R}} g_3(v)d\lambda(v) \right) d\lambda(u) \\ &= g_1(t) \left(\int_{\mathbb{R}} g_2 d\lambda \right) \left(\int_{\mathbb{R}} g_3 d\lambda \right) = \frac{g_1(t)}{\int_{\mathbb{R}} g_1 d\lambda},\end{aligned}$$

car, comme ρ est une densité de probabilité,

$$\begin{aligned}1 &= \int_{\mathbb{R}^3} \rho(t, u, v)d\lambda^{(3)}(t, u, v) = \int_{\mathbb{R}^3} g_1(t)g_2(u)g_3(v)d\lambda^{(3)}(t, u, v) \\ &= \left(\int_{\mathbb{R}} g_1 d\lambda \right) \left(\int_{\mathbb{R}} g_2 d\lambda \right) \left(\int_{\mathbb{R}} g_3 d\lambda \right).\end{aligned}$$

On notera que cela implique en particulier que $\int_{\mathbb{R}} g_1 d\lambda \neq 0$. On trouve de la même manière un résultat analogue pour Y et Z .

Montrons l'indépendance de la suite (X, Y, Z) .

Soient A, B, C des boréliens de \mathbb{R} . En notant que, pour tout $\omega \in \Omega$,

$$\mathbf{1}_{\{(X,Y,Z) \in A \times B \times C\}}(\omega) = \mathbf{1}_{\{X \in A\} \cap \{Y \in B\} \cap \{Z \in C\}}(\omega) = \mathbf{1}_{\{X \in A\}}(\omega) \mathbf{1}_{\{Y \in B\}}(\omega) \mathbf{1}_{\{Z \in C\}}(\omega)$$

et $\mathbf{1}_{\{X \in A\}}(\omega) = \mathbf{1}_A(X(\omega))$, en utilisant les propriétés de l'opérateur d'intégration donnée dans la proposition 3.1, page 30, puis en appliquant les théorèmes du transfert et de Tonelli, il vient

$$\begin{aligned}\mathbb{P}_{(X,Y,Z)}(A \times B \times C) &= \mathbb{E}[\mathbf{1}_{\{(X,Y,Z) \in A \times B \times C\}}] = \mathbb{E}[\mathbf{1}_A(X) \mathbf{1}_B(Y) \mathbf{1}_C(Z)] \\ &= \int_{\mathbb{R}^3} \mathbf{1}_A(x) \mathbf{1}_B(y) \mathbf{1}_C(z) g_1(x) g_2(y) g_3(z) d\lambda^{(3)}(x, y, z) \\ &= \left(\int_{\mathbb{R}} \mathbf{1}_A g_1 d\lambda \right) \left(\int_{\mathbb{R}} \mathbf{1}_B g_2 d\lambda \right) \left(\int_{\mathbb{R}} \mathbf{1}_C g_3 d\lambda \right).\end{aligned}$$

De plus, comme les variables aléatoires réelles X, Y , et Z ont pour densités respectives χ_1, χ_2 , et χ_3 ,

$$\begin{aligned}\mathbb{P}_X(A) \mathbb{P}_Y(B) \mathbb{P}_Z(C) &= \left(\int_{\mathbb{R}} \mathbf{1}_A \chi_1 d\lambda \right) \left(\int_{\mathbb{R}} \mathbf{1}_B \chi_2 d\lambda \right) \left(\int_{\mathbb{R}} \mathbf{1}_C \chi_3 d\lambda \right) \\ &= \left(\int_{\mathbb{R}} \mathbf{1}_A(t) \frac{g_1(t)}{\int_{\mathbb{R}} g_1 d\lambda} d\lambda(t) \right) \left(\int_{\mathbb{R}} \mathbf{1}_B(t) \frac{g_2(t)}{\int_{\mathbb{R}} g_2 d\lambda} d\lambda(t) \right) \left(\int_{\mathbb{R}} \mathbf{1}_C(t) \frac{g_3(t)}{\int_{\mathbb{R}} g_3 d\lambda} d\lambda(t) \right) \\ &= \left(\int_{\mathbb{R}} \mathbf{1}_A g_1 d\lambda \right) \left(\int_{\mathbb{R}} \mathbf{1}_B g_2 d\lambda \right) \left(\int_{\mathbb{R}} \mathbf{1}_C g_3 d\lambda \right),\end{aligned}$$

toujours en vertu de

$$\left(\int_{\mathbb{R}} g_1 d\lambda \right) \left(\int_{\mathbb{R}} g_2 d\lambda \right) \left(\int_{\mathbb{R}} g_3 d\lambda \right) = 1.$$

Ce qui montre que, pour tous boréliens A, B, C de \mathbb{R} ,

$$\mathbb{P}_{(X,Y,Z)}(A \times B \times C) = \mathbb{P}_X(A) \mathbb{P}_Y(B) \mathbb{P}_Z(C)$$

et prouve ainsi l'indépendance de la suite de variables aléatoires réelles (X, Y, Z) . \square

Donnons un autre énoncé, beaucoup plus utile dans la pratique, du critère d'indépendance des variable aléatoire réelle à densité :

Proposition 4.9.

Soit (X_1, \dots, X_n) une suite de variables aléatoires réelles telle que, pour tout $k = 1, \dots, n$, la variable aléatoire réelle X_k soit de densité ρ_k sur \mathbb{R} . Alors la suite de variables aléatoires réelles (X_1, \dots, X_n) est indépendante si, et seulement si, le vecteur aléatoire $X := (X_1, \dots, X_n)$ de dimension n admet pour densité sur \mathbb{R}^n l'application

$$\rho : (x_1, \dots, x_n) \in \mathbb{R}^n \mapsto \rho(x_1, \dots, x_n) := \rho_1(x_1)\rho_2(x_2) \cdots \rho_n(x_n) \in [0, +\infty].$$

Démonstration : La condition nécessaire est l'assertion 1) de la proposition 4.8 précédente et la condition suffisante résulte de l'assertion 2) avec, pour tout $k = 1, \dots, n$, $g_k := \rho_k$ en remarquant que, par hypothèse, $\int_{\mathbb{R}} \rho_k d\lambda = 1$. \square

Exemples 4.8.

- 1) Reprenons les notations de l'exercice 4.5, page 68. On y établit que le vecteur aléatoire (X, Y) de dimension 2 admet pour densité sur \mathbb{R}^2 l'application $\rho(x, y) := \frac{1}{2\pi} e^{-\frac{x^2+y^2}{2}}$ et que les variables aléatoires réelles X et Y suivent la même loi $\mathcal{N}_1(0, 1)$. Leurs densités ρ_X et ρ_Y sur \mathbb{R} sont définies sur \mathbb{R} par $\rho_X(t) = \rho_Y(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}$ et vérifient la relation $\rho(x, y) = \rho_X(x)\rho_Y(y)$, ce qui prouve l'indépendance de la suite de variables aléatoires réelles (X, Y) en vertu de la proposition précédente.
- 2) Dans l'exemple 4.4, page 65, on vérifie aisément que $\rho(x, y) \neq \chi(x)\chi(y)$, ce qui est une autre façon de prouver que la suite de variables aléatoires réelles (X, Y) n'est pas indépendante. \square

Donnons un critère valable pour des vecteurs aléatoires généraux sans hypothèses sur le type de loi qu'ils satisfont.

Proposition 4.10.

Critère des fonctions positives

Soit (X_1, \dots, X_n) une suite de vecteurs aléatoires de dimensions respectives d_1, \dots, d_n . Alors, la suite (X_1, \dots, X_n) est indépendante si, et seulement si, pour tout entier $1 \leq k \leq n$ et toute application borélienne positive f_k de \mathbb{R}^{d_k} dans $[0, +\infty]$,

$$\mathbb{E} [f_1(X_1)f_2(X_2) \cdots f_n(X_n)] = \mathbb{E} [f_1(X_1)] \mathbb{E} [f_2(X_2)] \cdots \mathbb{E} [f_n(X_n)].$$

Démonstration : Faisons la démonstration pour $n = 2$.

- - C.N. Par application des théorèmes du transfert et de Tonelli où on utilise l'indépendance

en écrivant $\mathbb{P}_{(X_1, X_2)} = \mathbb{P}_{X_1} \otimes \mathbb{P}_{X_2}$, il vient

$$\begin{aligned} \mathbb{E}[f_1(X_1)f_2(X_2)] &= \int_{\mathbb{R}^{d_1+d_2}} f_1(x)f_2(y)d\mathbb{P}_{(X_1, X_2)}(x, y) \\ &= \int_{\mathbb{R}^{d_1}} f_1(x) \left(\int_{\mathbb{R}^{d_2}} f_2(y)d\mathbb{P}_{X_2}(y) \right) d\mathbb{P}_{X_1}(x) \\ &= \left(\int_{\mathbb{R}^{d_1}} f_1 d\mathbb{P}_{X_1} \right) \left(\int_{\mathbb{R}^{d_2}} f_2 d\mathbb{P}_{X_2} \right) \\ &= \mathbb{E}[f_1(X_1)] \mathbb{E}[f_2(X_2)]. \end{aligned}$$

• - C.S. Il suffit de prendre $f_1 := \mathbb{1}_A$ et $f_2 := \mathbb{1}_B$ où A et B sont des boréliens respectivement de \mathbb{R}^{d_1} et \mathbb{R}^{d_2} . En explicitant la relation de l'hypothèse

$$\mathbb{E}[\mathbb{1}_A(X_1)\mathbb{1}_B(X_2)] = \mathbb{E}[\mathbb{1}_A(X_1)] \mathbb{E}[\mathbb{1}_B(X_2)]$$

on obtient

$$\mathbb{P}[(X_1, X_2) \in A \times B] = \mathbb{P}[X_1 \in A] \mathbb{P}[X_2 \in B],$$

ce qui prouve que $\mathbb{P}_{(X_1, X_2)} = \mathbb{P}_{X_1} \otimes \mathbb{P}_{X_2}$. \square

A titre d'exemple d'utilisation de cette proposition, donnons un corollaire très utile dans les calculs faisant intervenir des variables aléatoires réelles indépendantes et intégrables :

Proposition 4.11.

Soit (X_1, \dots, X_n) une suite de variables aléatoires réelles intégrables. Si la suite (X_1, \dots, X_n) est indépendante, alors la variable aléatoire réelle produit $X_1 X_2 \dots X_n$ est intégrable et

$$\mathbb{E}[X_1 X_2 \dots X_n] = \mathbb{E}(X_1) \mathbb{E}(X_2) \dots \mathbb{E}(X_n).$$

La réciproque est fausse.

Démonstration : Faisons-la pour deux variables aléatoires réelles X et Y indépendantes et intégrables i.e. $\mathbb{E}(|Y|) < +\infty$ et $\mathbb{E}(|X|) < +\infty$. On peut écrire

$$\mathbb{E}(|XY|) = \mathbb{E}(|X||Y|) = \mathbb{E}(|X|)\mathbb{E}(|Y|) < +\infty,$$

où on a appliqué, dans la deuxième égalité, la proposition 4.10 avec les fonctions positives $x \mapsto |x|$ grâce à l'indépendance de (X, Y) . On a donc prouvé que la variable aléatoire réelle XY est intégrable.

Montrons la deuxième relation. Remarquons qu'en introduisant les parties positives et négatives des v.a.r. , on peut écrire

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

qui donne, en prenant les espérances de chaque membres de l'égalité précédente et en appliquant la proposition 4.10 aux fonctions boréliennes positives $x \mapsto x^+$ et $x \mapsto x^-$,

$$\begin{aligned} \mathbb{E}(XY) &= \mathbb{E}[X^+Y^+] + \mathbb{E}[X^-Y^-] - \mathbb{E}[X^-Y^+] - \mathbb{E}[X^+Y^-] \\ &= \mathbb{E}(X^+)\mathbb{E}(Y^+) + \mathbb{E}(X^-)\mathbb{E}(Y^-) - \mathbb{E}(X^-)\mathbb{E}(Y^+) - \mathbb{E}(X^+)\mathbb{E}(Y^-) \\ &= \mathbb{E}(X)\mathbb{E}(Y). \end{aligned}$$

d'où la relation cherchée. \square

L'exercice suivant propose un contre-exemple prouvant que la réciproque de la proposition précédente 4.11 est fautive. Cet exercice est souvent à la base de nombreux contre-exemples concernant l'indépendance des variables aléatoires.

Exercice 4.6. (Corrigé de l'exercice : page 168)

Soient X une variable aléatoire réelle de loi $\mathcal{N}_1(0, 1)$ et ε une variable aléatoire réelle indépendante de X de loi $\frac{1}{2}(\delta_{-1} + \delta_1)$. Montrer que la variable aléatoire réelle $Y := \varepsilon X$ est de loi $\mathcal{N}_1(0, 1)$. Prouver que les v.a.r. X, Y vérifient la relation $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$ mais que le couple (X, Y) n'est pas indépendant (pour cela on calculera $\mathbb{E}(X^2 Y^2)$ et on utilisera la proposition 4.10).

Proposition 4.12.

Critère des fonctions bornée

Soit (X_1, \dots, X_n) une suite de vecteurs aléatoires de dimensions respectives d_1, \dots, d_n . Alors, la suite (X_1, \dots, X_n) est indépendante si, et seulement si, pour tout entier $1 \leq k \leq n$ et toute application borélienne bornée f_k de \mathbb{R}^{d_k} dans $[0, +\infty]$,

$$\mathbb{E}[f_1(X_1)f_2(X_2)\cdots f_n(X_n)] = \mathbb{E}[f_1(X_1)]\mathbb{E}[f_2(X_2)]\cdots\mathbb{E}[f_n(X_n)].$$

Démonstration : Faisons la démonstration pour $n = 2$.

• - C.N. Par application des théorèmes du transfert et de Fubini (car les fonctions boréliennes bornées sont intégrables) où on utilise l'indépendance en écrivant $\mathbb{P}_{(X_1, X_2)} = \mathbb{P}_{X_1} \otimes \mathbb{P}_{X_2}$, il vient

$$\begin{aligned} \mathbb{E}[f_1(X_1)f_2(X_2)] &= \int_{\mathbb{R}^{d_1+d_2}} f_1(x)f_2(y)d\mathbb{P}_{(X_1, X_2)}(x, y) \\ &= \int_{\mathbb{R}^{d_1}} f_1(x) \left(\int_{\mathbb{R}^{d_2}} f_2(y)d\mathbb{P}_{X_2}(y) \right) d\mathbb{P}_{X_1}(x) \\ &= \left(\int_{\mathbb{R}^{d_1}} f_1 d\mathbb{P}_{X_1} \right) \left(\int_{\mathbb{R}^{d_2}} f_2 d\mathbb{P}_{X_2} \right) \\ &= \mathbb{E}[f_1(X_1)]\mathbb{E}[f_2(X_2)]. \end{aligned}$$

• - C.S. Il suffit de prendre $f_1 := \mathbb{1}_A$ et $f_2 := \mathbb{1}_B$ où A et B sont des boréliens respectivement de \mathbb{R}^{d_1} et \mathbb{R}^{d_2} . En explicitant la relation de l'hypothèse

$$\mathbb{E}[\mathbb{1}_A(X_1)\mathbb{1}_B(X_2)] = \mathbb{E}[\mathbb{1}_A(X_1)]\mathbb{E}[\mathbb{1}_B(X_2)]$$

on obtient

$$\mathbb{P}[(X_1, X_2) \in A \times B] = \mathbb{P}[X_1 \in A]\mathbb{P}[X_2 \in B],$$

ce qui prouve que $\mathbb{P}_{(X_1, X_2)} = \mathbb{P}_{X_1} \otimes \mathbb{P}_{X_2}$. \square

Exercice 4.7. (Corrigé de l'exercice : page 169)

Soit (X, Y) un couple indépendant de variables aléatoires réelles de même loi $\mathcal{N}_1(m, \sigma^2)$, montrer que $\mathbb{E}[(X + Y)^2] = 4m^2 + 2\sigma^2$.

Comme corollaire de la proposition 4.11,

Proposition 4.13.

Soit $X := (X_1, X_2, \dots, X_d)$ un vecteur aléatoire de carré intégrable de dimension d . Si la suite de variables aléatoires réelles (X_1, X_2, \dots, X_d) est indépendante, alors la matrice de dispersion de X est diagonale.

La réciproque est fausse.

Démonstration : La proposition 4.11 entraîne que, pour tout couple d'entiers (i, j) avec $i \neq j$, $\text{Cov}(X_i, X_j) = 0$. On conclut en utilisant l'assertion 3) de la proposition 3.24, page 48. \square

Donnons un procédé simple de construction de suites indépendantes de vecteurs aléatoires à partir de deux fonctions de v.a.r.. Ce procédé peut se généraliser à un nombre quelconque de fonctions.

Proposition 4.14.

Indépendance de fonctions de v.a.r.

Si $(X_1, \dots, X_n, Y_1, \dots, Y_p)$ est une suite indépendante de v.a.r., alors, pour toutes applications boréliennes φ de \mathbb{R}^n dans \mathbb{R}^{d_1} et ψ de \mathbb{R}^p dans \mathbb{R}^{d_2} , le couple de vecteurs aléatoires $(\varphi(X_1, \dots, X_n), \psi(Y_1, \dots, Y_p))$ est indépendant.

Démonstration : • Considérons les vecteurs aléatoires $X := (X_1, \dots, X_n)$ et $Y := (Y_1, \dots, Y_p)$. Commençons par montrer que le couple de vecteurs aléatoires (X, Y) est indépendant. Comme $(X_1, \dots, X_n, Y_1, \dots, Y_p)$ est une suite indépendante, pour tous boréliens de \mathbb{R} , A_1, \dots, A_n , il vient

$$\begin{aligned} \mathbb{P}_X[A_1 \times \dots \times A_n] &= \mathbb{P}_{(X_1, \dots, X_n)}[A_1 \times \dots \times A_n] \\ &= \mathbb{P}_{(X_1, \dots, X_n, Y_1, \dots, Y_p)}[A_1 \times \dots \times A_n \times \mathbb{R}^p] \\ &= \mathbb{P}_{X_1}(A_1) \cdots \mathbb{P}_{X_n}(A_n). \end{aligned}$$

Ce qui prouve que $\mathbb{P}_X = \mathbb{P}_{(X_1, \dots, X_n)} = \mathbb{P}_{X_1} \otimes \dots \otimes \mathbb{P}_{X_n}$, et par suite

$$\begin{aligned} \mathbb{P}_{(X, Y)} &= \mathbb{P}_{(X_1, \dots, X_n, Y_1, \dots, Y_p)} \\ &= \mathbb{P}_{X_1} \otimes \dots \otimes \mathbb{P}_{X_n} \otimes \mathbb{P}_{Y_1} \otimes \dots \otimes \mathbb{P}_{Y_p} \\ &= \mathbb{P}_X \otimes \mathbb{P}_Y. \end{aligned}$$

Le couple de vecteurs aléatoires (X, Y) est donc indépendant.

• Considérons maintenant deux applications boréliennes positives f_1 et f_2 définies respectivement sur \mathbb{R}^{d_1} et \mathbb{R}^{d_2} . Comme $f_1 \circ \varphi$ et $f_2 \circ \psi$ sont des fonctions boréliennes positives, en appliquant la condition nécessaire du critère des fonctions positives à ces fonctions et à la suite indépendante (X, Y) , il vient

$$\mathbb{E}[f_1(\varphi(X))f_2(\psi(Y))] = \mathbb{E}[f_1(\varphi(X))]\mathbb{E}[f_2(\psi(Y))].$$

On applique alors la condition suffisante du critère des fonctions positives aux fonctions f_1 , f_2 et à la suite $(\varphi(X), \psi(Y))$ pour conclure qu'elle est indépendante. \square

On pourra vérifier que si un couple de vecteurs aléatoires (X, Y) de dimensions respectives d et k est indépendant avec $X := (X_1, \dots, X_d)$ et $Y := (Y_1, \dots, Y_k)$ alors, pour tout (i, j) avec

$1 \leq i \leq d$ et $1 \leq j \leq k$, le couple de v.a.r. (X_i, Y_j) est indépendant.

Donnons un cas particulier de la proposition précédente, très utilisé en pratique :

Proposition 4.15.

Indépendance de fonctions de v.a.r. (cas usuel)

Si (X_1, \dots, X_n) est une suite indépendante de v.a.r., alors, pour toute suite d'applications boréliennes (f_1, \dots, f_n) de \mathbb{R} dans \mathbb{R} , la suite de variables aléatoires réelles $(f_1(X_1), \dots, f_n(X_n))$ est indépendante.

Terminons par un critère d'indépendance, simple d'application, utilisant les fonctions caractéristiques :

Proposition 4.16.

Critère d'indépendance par les f.c.

Soit (X_1, \dots, X_n) une suite de vecteurs aléatoires de dimensions respectives d_1, \dots, d_n . Alors, la suite (X_1, \dots, X_n) est indépendante si, et seulement si, pour tout $u_1 \in \mathbb{R}^{d_1}, \dots, u_n \in \mathbb{R}^{d_n}$,

$$\begin{aligned} \Phi_{(X_1, \dots, X_n)}(u_1, \dots, u_n) &= \Phi_{X_1}(u_1) \cdots \Phi_{X_n}(u_n) \\ \text{i.e. } \mathbb{E} \left[\exp \left(i \sum_{k=1}^n \langle u_k, X_k \rangle \right) \right] &= \mathbb{E} (e^{i \langle u_1, X_1 \rangle}) \cdots \mathbb{E} (e^{i \langle u_n, X_n \rangle}). \end{aligned}$$

Démonstration : • - C.N. Supposons que les vecteurs aléatoires (X_1, \dots, X_n) sont indépendants. Alors, pour tout $u_1 \in \mathbb{R}^{d_1}, \dots, u_n \in \mathbb{R}^{d_n}$,

$$\Phi_{(X_1, \dots, X_n)}(u_1, \dots, u_n) = \mathbb{E} (e^{i[\langle u_1, X_1 \rangle + \langle u_2, X_2 \rangle + \cdots + \langle u_n, X_n \rangle]}) = \mathbb{E} \left(\prod_{k=1}^{k=n} e^{i \langle u_k, X_k \rangle} \right).$$

D'après la propriété 4.12, page 74, appliquée aux fonctions boréliennes bornées $e^{i \langle u_1, X_1 \rangle}, e^{i \langle u_2, X_2 \rangle}, \dots, e^{i \langle u_n, X_n \rangle}$, on a

$$\Phi_{(X_1, \dots, X_n)}(u_1, \dots, u_n) = \prod_{k=1}^{k=n} \mathbb{E} (e^{i \langle u_k, X_k \rangle}) = \Phi_{X_1}(u_1) \cdots \Phi_{X_n}(u_n).$$

• - C.S. Soit $u_1 \in \mathbb{R}^{d_1}, \dots, u_n \in \mathbb{R}^{d_n}$, et $u = (u_1, u_2, \dots, u_n) \in \mathbb{R}^{d_1+d_2+\cdots+d_n}$. Soit $X = (X_1, X_2, \dots, X_n)$ un vecteur aléatoire de dimension $d_1 + d_2 + \cdots + d_n$ de loi \mathbb{P}_X , la condition $\Phi_{(X_1, \dots, X_n)}(u_1, \dots, u_n) = \Phi_{X_1}(u_1) \cdots \Phi_{X_n}(u_n)$ s'écrit, en appliquant le théorème du transfert et celui de Fubini,

$$\int_{\mathbb{R}^{d_1+d_2+\cdots+d_n}} e^{i \langle u, x \rangle} d\mathbb{P}_{(X_1, \dots, X_n)}(x) = \int_{\mathbb{R}^{d_1+d_2+\cdots+d_n}} e^{i \langle u, x \rangle} d[\mathbb{P}_{X_1} \otimes \mathbb{P}_{X_2} \otimes \cdots \otimes \mathbb{P}_{X_n}](x).$$

Ce qui prouve que les probabilités $\mathbb{P}_{X_1} \otimes \mathbb{P}_{X_2} \otimes \cdots \otimes \mathbb{P}_{X_n}$ et $\mathbb{P}_{(X_1, \dots, X_n)}$ ont les mêmes fonctions caractéristiques, donc sont égales en vertu du critère d'identification des lois par les fonctions caractéristiques. □

4.2.3 Indépendance d'événements, de tribus

On se donne comme référence un espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$. Les résultats précédents incitent à élargir la notion d'indépendance aux familles d'événements $(A_i)_I$ et aux familles $(\mathcal{F}_i)_I$ de sous-tribus de \mathcal{F} , où I est un ensemble quelconque.

Définition 4.5.

Une famille quelconque $(A_i)_I$ d'événements, est dite **(mutuellement) indépendante** pour \mathbb{P} si, pour toute sous-famille finie $(A_i)_K$, $K \subseteq I$ et K fini, on a :

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

Définition 4.6.

Une famille quelconque $(\mathcal{F}_i)_I$ de sous-tribus de \mathcal{F} , est dite **(mutuellement) indépendante** pour \mathbb{P} si toute famille d'événements $(A_i)_I$ avec $A_i \in \mathcal{F}_i$, pour tout $i \in I$, est indépendante pour \mathbb{P} .

On dit aussi plus fréquemment, et par abus de langage, que les événements $(A_i)_I$, sont **indépendants** (resp. les tribus $(\mathcal{F}_i)_I$, sont **indépendantes**).

On remarquera bien que la notion d'indépendance dépend de la probabilité \mathbb{P} choisie sur (Ω, \mathcal{F}) . De plus si l'indépendance mutuelle d'une famille d'événements entraîne leur indépendance deux à deux, il faut noter que la réciproque est fautive (cf. [14] ex. 3-1 à 3-3).

La preuve de l'indépendance de tribus peut s'établir en considérant des π -systèmes de générateurs d'après la proposition :

Proposition 4.17.

Soient \mathcal{C} et \mathcal{D} deux familles d'événements stables par intersection finie (π -systèmes). Si, pour tout $(A, B) \in \mathcal{C} \times \mathcal{D}$, les événements A et B sont indépendants, alors les sous-tribus engendrées respectivement par \mathcal{C} et \mathcal{D} sont indépendantes.

Démonstration : Il suffit de montrer que, pour tout $(A, B) \in \sigma(\mathcal{C}) \times \sigma(\mathcal{D})$, $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$.

• Montrons d'abord que, pour tout $B \in \mathcal{D}$ et pour tout $A \in \sigma(\mathcal{C})$, on a $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$. Soit $B \in \mathcal{D}$ fixé.

Si $\mathbb{P}(B) = 0$, alors pour tout $A \in \sigma(\mathcal{C})$, $A \cap B \subseteq B$ et par suite $\mathbb{P}(A \cap B) \leq \mathbb{P}(B) = 0$. Donc, dans ce cas, pour tout $A \in \sigma(\mathcal{C})$, on a bien $\mathbb{P}(A \cap B) = 0 = \mathbb{P}(A)\mathbb{P}(B)$.

Si $\mathbb{P}(B) > 0$, Considérons l'application $A \in \sigma(\mathcal{C}) \mapsto \mathbb{P}_B(A) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} \in [0, 1]$. On vérifie

aisément que c'est une probabilité sur $(\Omega, \sigma(\mathcal{C}))$. De plus la probabilité \mathbb{P}_B , coïncide avec la probabilité \mathbb{P} sur le π -système \mathcal{C} , car en vertu de l'hypothèse, pour tout $(A, B) \in \mathcal{C} \times \mathcal{D}$, les événements A et B sont indépendants, on a bien $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$, c'est-à-dire $\mathbb{P}_B(A) = \mathbb{P}(A)$. Donc d'après le théorème d'unicité pour les probabilités (cf. proposition 1.11, page 14) on en déduit que la probabilité \mathbb{P}_B coïncide avec la probabilité \mathbb{P} sur la tribu $\sigma(\mathcal{C})$ engendrée par le π -système \mathcal{C} c'est-à-dire que, pour tout $A \in \sigma(\mathcal{C})$, $\mathbb{P}_B(A) = \mathbb{P}(A)$, ou encore

pour tout $A \in \sigma(\mathcal{C})$, $\frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)} = \mathbb{P}(A)$, ou $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$.

En résumé, on a bien montré que, pour tout $B \in \mathcal{D}$ et pour tout $A \in \sigma(\mathcal{C})$, on a $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$.

• *Montrons maintenant que, pour tout $B \in \sigma(\mathcal{D})$ et pour tout $A \in \sigma(\mathcal{C})$, on a $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$.*

Utilisons la même démarche. Soit $A \in \sigma(\mathcal{C})$ fixé.

Si $\mathbb{P}(A) = 0$, alors pour tout $B \in \sigma(\mathcal{D})$, $A \cap B \subseteq A$ et par suite $\mathbb{P}(A \cap B) \leq \mathbb{P}(A) = 0$. Donc, dans ce cas, pour tout $B \in \sigma(\mathcal{D})$, on a bien $\mathbb{P}(A \cap B) = 0 = \mathbb{P}(A)\mathbb{P}(B)$.

Si $\mathbb{P}(A) > 0$, considérons l'application $B \in \sigma(\mathcal{D}) \mapsto \mathbb{P}_A(B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)} \in [0, 1]$. On vérifie de même que c'est une probabilité sur $(\Omega, \sigma(\mathcal{D}))$. De plus la probabilité \mathbb{P}_A , coïncide avec la probabilité \mathbb{P} sur le π -système \mathcal{D} , car en vertu de l'hypothèse, pour tout $(A, B) \in \mathcal{C} \times \mathcal{D}$, les événements A et B sont indépendants, on a $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$. Donc d'après le théorème d'unicité pour les probabilités (cf. proposition 1.11, page 14) on en déduit que la probabilité \mathbb{P}_A coïncide avec la probabilité \mathbb{P} sur la tribu $\sigma(\mathcal{D})$ engendrée par le π -système \mathcal{D} c'est-à-dire que, pour tout $B \in \sigma(\mathcal{D})$, $\mathbb{P}_A(B) = \mathbb{P}(B)$, ou encore pour tout $B \in \sigma(\mathcal{D})$, $\frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)} = \mathbb{P}(B)$, ou

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

Enfin, on a bien montré que, pour tout $B \in \sigma(\mathcal{D})$ et pour tout $A \in \sigma(\mathcal{C})$, on a $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$. \square

En utilisant la même démarche, on peut établir un critère d'indépendance des vecteurs aléatoires, qui prouve que dans le critère 4.6, page 68, on peut se limiter qu'à certains boréliens de \mathbb{R}^d :

Proposition 4.18.

Une suite (X_1, \dots, X_n) de vecteurs aléatoires de dimensions respectives d_1, \dots, d_n , est indépendante si, et seulement si,

$$\mathbb{P} \left[\bigcap_{k=1}^{k=n} \{X_k \in A_k\} \right] = \prod_{k=1}^{k=n} \mathbb{P}(X_k \in A_k),$$

pour tous $A_k \in \mathcal{C}_k$, où, pour tous $k = 1, 2, \dots, n$, \mathcal{C}_k est un π -système engendrant la tribu borélienne $\mathcal{B}(\mathbb{R}^{d_k})$.

Démonstration : Pour la démonstration remarquer que la relation $\mathbb{P} \left[\bigcap_{k=1}^{k=n} \{X_k \in A_k\} \right] = \prod_{k=1}^{k=n} \mathbb{P}(X_k \in A_k)$, peut s'écrire encore $\mathbb{P}_{(X_1, \dots, X_n)}(A_1 \times \dots \times A_n) = \prod_{k=1}^{k=n} \mathbb{P}_{X_k}(A_k)$. On montre la proposition pour $n = 2$ en raisonnant comme dans la proposition 4.17, puis on généralise par récurrence au cas n quelconque. \square

En particulier, la proposition précédente 4.17 a pour corollaire le théorème :

Proposition 4.19.

Soit $(\mathcal{F}_i)_I$ une famille indépendante de sous-tribus de \mathcal{F} . Si K et J sont deux parties disjointes et non vides de I , alors les tribus engendrées respectivement par les familles $\bigcup_{i \in J} \mathcal{F}_i$ et $\bigcup_{i \in K} \mathcal{F}_i$ sont indépendantes.

Démonstration : Notons \mathcal{C}_J la famille des intersections de familles finies d'événements de $\bigcup_{i \in J} \mathcal{F}_i$.

Un élément de \mathcal{C}_J est de la forme $\bigcap_{i \in J'} A_i$ où J' est une partie finie de J et, pour tout $i \in J'$, $A_i \in \mathcal{F}_i$.

On définit de même \mathcal{C}_K la famille des intersections de familles finies d'événements de $\bigcup_{i \in K} \mathcal{F}_i$.

Un élément de \mathcal{C}_K est de la forme $\bigcap_{i \in K'} A_i$ où K' est une partie finie de K et, pour tout $i \in K'$, $A_i \in \mathcal{F}_i$.

• On a $\sigma(\mathcal{C}_J) = \sigma\left(\bigcup_{i \in J} \mathcal{F}_i\right)$ et $\sigma(\mathcal{C}_K) = \sigma\left(\bigcup_{i \in K} \mathcal{F}_i\right)$.

En effet, en considérant les familles à un seul élément, on a facilement l'inclusion $\bigcup_{i \in J} \mathcal{F}_i \subseteq \mathcal{C}_J$,

d'où l'inclusion des tribus engendrées correspondantes $\sigma\left(\bigcup_{i \in J} \mathcal{F}_i\right) \subseteq \sigma(\mathcal{C}_J)$. Réciproque-

ment, par la stabilité de l'intersection finie dans les tribus, on a $\mathcal{C}_J \subseteq \sigma\left(\bigcup_{i \in J} \mathcal{F}_i\right)$, et

par suite $\sigma(\mathcal{C}_J) \subseteq \sigma\left(\bigcup_{i \in J} \mathcal{F}_i\right)$. D'où l'égalité. On montre de la même façon la relation

$\sigma(\mathcal{C}_K) = \sigma\left(\bigcup_{i \in K} \mathcal{F}_i\right)$. De plus, on vérifie aisément que les familles \mathcal{C}_J et \mathcal{C}_K sont des π -systèmes.

• Montrons que, pour tout $(A, B) \in \mathcal{C}_J \times \mathcal{C}_K$, les événements A et B sont indépendants.

En effet, soit $A \in \mathcal{C}_J$ et $B \in \mathcal{C}_K$. Alors $A = \bigcap_{i \in J'} A_i$ où J' est une partie finie de J et, pour

tout $i \in J'$, $A_i \in \mathcal{F}_i$, et $B = \bigcap_{i \in K'} A_i$ où K' est une partie finie de K et, pour tout $i \in K'$,

$A_i \in \mathcal{F}_i$. Par suite, comme $J \cap K = \emptyset$, on a $J' \cap K' = \emptyset$, et $A \cap B = \bigcap_{i \in J' \cup K'} A_i$. Il vient alors,

$\mathbb{P}(A \cap B) = \mathbb{P}\left(\bigcap_{i \in J' \cup K'} A_i\right) = \prod_{i \in J' \cup K'} \mathbb{P}(A_i)$, car la suite $(\mathcal{F}_i)_i$ est une famille indépendante de

sous-tribus de \mathcal{F} . Par la commutativité et l'associativité du produit, et puis par l'indépendance des sous-tribus, on peut alors écrire $\prod_{i \in J' \cup K'} \mathbb{P}(A_i) = \prod_{i \in J'} \mathbb{P}(A_i) \times \prod_{i \in K'} \mathbb{P}(A_i) = \mathbb{P}(A) \times \mathbb{P}(B)$.

Ce qui prouve que, pour tout $(A, B) \in \mathcal{C}_J \times \mathcal{C}_K$, les événements A et B sont indépendants.

• On peut alors appliquer la proposition 4.17 aux π -systèmes \mathcal{C}_J et \mathcal{C}_K pour obtenir le résultat recherché. \square

Comme corollaire, nous avons un résultat similaire avec les événements :

Proposition 4.20.

Soit $(A_i)_i$ une famille indépendante d'événements de \mathcal{F} . Si K et J sont deux parties disjointes et non vides de I , alors les tribus engendrées $\sigma((A_i)_{i \in J})$ et $\sigma((A_i)_{i \in K})$ sont indépendantes.

Démonstration : Il suffit d'appliquer la proposition 4.19 à la famille des sous-tribus $(\sigma(A_i))_i$ de \mathcal{F} , et de remarquer que $\sigma((A_i)_{i \in J})$ (resp. $\sigma((A_i)_{i \in K})$) est aussi la tribu engendrée par la famille $\bigcup_{i \in J} \sigma(A_i)$ (resp. $\bigcup_{i \in K} \sigma(A_i)$). \square

On peut vérifier facilement que :

Proposition 4.21.

Si la famille d'événements $(A_i)_I$ est indépendante, il en est de même de la famille d'événements $(B_i)_I$ où, pour tout $i \in I$, $B_i := A_i$ ou $B_i := A_i^c$.

Définition 4.7.

*Si X est un vecteur aléatoire de dimension d , on note $\sigma(X)$ la plus petite sous-tribu de \mathcal{F} rendant mesurable l'application X . $\sigma(X)$ s'appelle la **tribu engendrée par la variable X** .*

On vérifie facilement que $\sigma(X)$ est l'ensemble des images-réciproques de tous les boréliens de \mathbb{R}^d . (cf. aussi exercice 4.20, page 93, pour une généralisation)

Exemples 4.9.

Si A est un événement de \mathcal{F} , $\sigma(\mathbb{1}_A) = \sigma(A) = \{\emptyset, A, A^c, \Omega\}$.

Avec ces notations, le lien entre la notion d'indépendance pour les événements, celle pour les vecteurs aléatoires et celle pour les tribus est mis en évidence par les propositions suivantes dont les démonstrations sont élémentaires et laissées en exercice :

Proposition 4.22.

Une famille quelconque $(X_i)_I$ de vecteurs aléatoires X_i de dimension d_i , $i \in I$, est indépendante si, et seulement si, la famille de sous-tribus $(\sigma(X_i))_I$ est indépendante.

Proposition 4.23.

La famille d'événements $(A_i)_I$ est indépendante si, et seulement si, la famille des sous-tribus $(\sigma(A_i))_I$ est indépendante.

Proposition 4.24.

La famille d'événements $(A_i)_I$ est indépendante si, et seulement si, la famille des v.a.r. $(\mathbb{1}_{A_i})_I$ est indépendante.

Proposition 4.25.

Si \mathcal{A} et \mathcal{B} sont deux sous-tribus de \mathcal{F} indépendantes, X et Y deux vecteurs aléatoires respectivement \mathcal{A} -mesurable et \mathcal{B} -mesurable, alors les vecteurs aléatoires X et Y sont indépendants.

4.3 Tribu et événements asymptotiques

Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace de probabilité et $(A_n)_{\mathbb{N}}$ une suite d'événements de \mathcal{F} . Pour tout $n \in \mathbb{N}$, notons \mathcal{A}_n la tribu engendrée par la suite d'événements $(A_k)_{k \geq n}$ i.e. $\mathcal{A}_n := \sigma(A_n, A_{n+1}, \dots, A_{n+k}, \dots)$.

Définition 4.8.

Un événement A de \mathcal{F} est dit **événement asymptotique (relativement à la suite d'événements $(A_n)_{\mathbb{N}}$)** si A est mesurable par rapport à toutes les tribus de la suite $(\mathcal{A}_n)_{\mathbb{N}}$. Cela équivaut à dire que A est mesurable par rapport à la tribu $\bigcap_{n=0}^{\infty} \mathcal{A}_n$ appelée **tribu asymptotique relative à la suite d'événements $(A_n)_{\mathbb{N}}$** .

La loi du Tout ou Rien de Kolmogorov donne des informations sur la valeur de la probabilité d'un événement asymptotique relativement à une suite indépendante d'événements $(A_n)_{\mathbb{N}}$:

Proposition 4.26.**Loi du Tout ou Rien ou du Zéro-Un de Kolmogorov**

Soit $(A_n)_{\mathbb{N}}$ une suite indépendante d'événements. Si A est un événement asymptotique relativement à la suite d'événements $(A_n)_{\mathbb{N}}$, alors $\mathbb{P}(A) = 0$ ou $\mathbb{P}(A) = 1$.

Démonstration : On note, pour tout entier naturel n , $\mathcal{A}_n := \sigma(A_n, A_{n+1}, \dots, A_{n+k}, \dots)$, et $\mathcal{A}'_n := \sigma(A_0, A_1, A_2, \dots, A_n)$. On note \mathcal{A} la tribu asymptotique $\bigcap_{n=0}^{\infty} \mathcal{A}_n$.

L'idée de la démonstration est de montrer que la tribu asymptotique \mathcal{A} est indépendante d'elle-même.

• Montrons que, pour tout entier naturel n , $\mathcal{A}_{n+1} = \sigma(A_{n+1}, A_{n+2}, \dots, A_{n+k}, \dots)$, et $\mathcal{A}'_n := \sigma(A_0, A_1, A_2, \dots, A_n)$ sont des tribus indépendantes.

Pour cela il suffit d'appliquer la proposition 4.20, en reprenant ses notations, au cas où $I = \mathbb{N}$ avec $J = \{0, 1, \dots, n\}$ et $K = \{n+1, n+2, \dots, n+k, \dots\}$.

• Montrons que, pour tout $A \in \mathcal{A}$ et pour tout $B \in \mathcal{A}_0$, on a $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$ (ce qui exprime que la tribu asymptotique \mathcal{A} est indépendante de la tribu $\mathcal{A}_0 := \sigma(A_0, A_1, A_2, \dots, A_k, \dots)$.)

Remarquons tout d'abord que, suivant une démarche déjà utilisée dans la démonstration de la proposition 4.19, \mathcal{A}_0 est également engendrée par le π -système \mathcal{C} constitué des intersections de la forme $\bigcap_{i \in I} B_i$ où I est une partie finie de \mathbb{N} avec, pour tout $i \in I$, $B_i \in \sigma(A_i)$. Donc

$$\mathcal{A}_0 = \sigma(\mathcal{C}).$$

Il reste à montrer que, pour tout $A \in \mathcal{A}$ et pour tout $B \in \mathcal{A}_0$, on a $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$. Raisonnons comme dans la démonstration de la proposition 4.18.

Soit $A \in \mathcal{A}$ fixé.

Si $\mathbb{P}(A) = 0$, alors pour tout $B \in \mathcal{A}_0$, $A \cap B \subseteq A$ et par suite $\mathbb{P}(A \cap B) \leq \mathbb{P}(A) = 0$. Donc, dans ce cas, pour tout $B \in \mathcal{A}_0$, on a bien $\mathbb{P}(A \cap B) = 0 = \mathbb{P}(A)\mathbb{P}(B)$.

Si $\mathbb{P}(A) > 0$, considérons l'application $B \in \mathcal{A}_0 \mapsto \mathbb{P}_A(B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)} \in [0, 1]$. C'est une probabilité sur (Ω, \mathcal{A}_0) .

La probabilité \mathbb{P}_A , coïncide avec la probabilité \mathbb{P} sur le π -système \mathcal{C} . En effet, soit $B \in \mathcal{C}$, alors $B = \bigcap_{i \in I} B_i$ où I est une partie finie de \mathbb{N} avec, pour tout $i \in I$, $B_i \in \sigma(A_i)$. Posons

$n_0 = \max(I)$, alors $B \in \mathcal{A}'_{n_0} = \sigma(A_0, A_1, A_2, \dots, A_{n_0})$. mais $A \in \mathcal{A} \subseteq \mathcal{A}_{n_0+1}$. Comme d'après le premier point de la démonstration, les tribus \mathcal{A}_{n_0+1} et \mathcal{A}'_{n_0} sont indépendantes, on a $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$, c'est-à-dire $\mathbb{P}_A(B) = \mathbb{P}(B)$.

Donc d'après le théorème d'unicité pour les probabilités (cf. proposition 1.11, page 14) on en déduit que la probabilité \mathbb{P}_A coïncide avec la probabilité \mathbb{P} sur la tribu $\mathcal{A}_0 = \sigma(\mathcal{C})$ engendrée

par le π -système \mathcal{C} c'est-à-dire que, pour tout $B \in \mathcal{A}_0 = \sigma(\mathcal{C})$, $\mathbb{P}_A(B) = \mathbb{P}(B)$, ou encore pour tout $B \in \mathcal{A}_0 = \sigma(\mathcal{C})$, $\frac{\mathbb{P}(A \cap B)}{\mathbb{P}(A)} = \mathbb{P}(B)$, et par suite $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$.

Finalement, on a bien montré que, pour tout $B \in \mathcal{A}_0$ et pour tout $A \in \mathcal{A}$, on a $\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$.

• Soit A un événement asymptotique relativement à la suite indépendante d'événements $(A_n)_{\mathbb{N}}$, i.e. $A \in \mathcal{A}$. Comme $\mathcal{A} \subseteq \mathcal{A}_0$, on peut appliquer alors le résultat du point précédent à $A \in \mathcal{A}$ et à $B = A \in \mathcal{A}_0$, ce qui donne $\mathbb{P}(A \cap A) = \mathbb{P}(A)\mathbb{P}(A)$, ou encore $\mathbb{P}(A) = [\mathbb{P}(A)]^2$, ce qui implique que $\mathbb{P}(A)$ ne peut prendre que la valeur 0 ou la valeur 1. \square

Donnons deux exemples importants d'événements asymptotiques :

Si $(A_n)_{\mathbb{N}}$ est une suite d'événements, on notera $\limsup A_n$ l'ensemble des $\omega \in \Omega$ tels que $\{n \in \mathbb{N} / \omega \in A_n\}$ est infini. En conséquence, l'événement $\limsup A_n$ est réalisé si, et seulement si, une infinité d'événements de la suite $(A_n)_{\mathbb{N}}$ sont réalisés.

Si $(A_n)_{\mathbb{N}}$ est une suite d'événements, on notera $\liminf A_n$ l'ensemble des $\omega \in \Omega$ tels que $\{n \in \mathbb{N} / \omega \notin A_n\}$ est fini. En conséquence, l'événement $\liminf A_n$ est réalisé si, et seulement si, tous les événements de la suite $(A_n)_{\mathbb{N}}$, sauf éventuellement un nombre fini d'entre eux, sont réalisés.

La proposition suivante affirme que les événements $\limsup(A_n)$ et $\liminf(A_n)$ sont bien des événements asymptotiques :

Proposition 4.27.

Si $(A_n)_{\mathbb{N}}$ est une suite d'événements, alors

1. $\liminf A_n \subseteq \limsup A_n$.

2. $\limsup A_n = \bigcap_{p=0}^{\infty} \left(\bigcup_{k=p}^{\infty} A_k \right)$.

3. $\liminf A_n = \bigcup_{p=0}^{\infty} \left(\bigcap_{k=p}^{\infty} A_k \right)$.

4. $\limsup A_n$ et $\liminf A_n$ sont des événements asymptotiques relativement à la suite d'événements $(A_n)_{\mathbb{N}}$.

Démonstration : Les propriétés 1), 2) et 3) résultent directement des définitions de $\limsup A_n$ et $\liminf A_n$.

Montrons la propriété 4) pour l'événement $\limsup A_n$. Le raisonnement est analogue pour $\liminf A_n$. Posons, pour tout entier naturel p , $B_p = \bigcup_{k=p}^{\infty} A_k$. La suite $(B_n)_{\mathbb{N}}$ est une suite décroissante pour l'inclusion. Par suite, pour tout entier naturel p , $\bigcap_{k=0}^{\infty} B_k = \bigcap_{k=p}^{\infty} B_k$. La suite des tribus $(\mathcal{A}_n := \sigma(A_n, A_{n+1}, \dots, A_{n+k}, \dots))_{\mathbb{N}}$ est une suite décroissante, donc pour tout entier naturel p et pour tout entier naturel $k \geq p$, $B_k \in \mathcal{A}_k \subseteq \mathcal{A}_p$, ce qui implique que, pour tout entier naturel p , $\bigcap_{k=p}^{\infty} B_k \in \mathcal{A}_p$. Par suite, pour tout entier naturel p , $\limsup A_n = \bigcap_{k=0}^{\infty} B_k = \bigcap_{k=p}^{\infty} B_k \in \mathcal{A}_p$. Ce

qui prouve que $\limsup A_n \in \bigcap_{p=0}^{\infty} \mathcal{A}_p$. \square

En combinant l'item 4) de la proposition 4.27 précédente et la loi du Tout ou Rien, on obtient aisément le corollaire suivant :

Proposition 4.28.

Si $(A_n)_{\mathbb{N}}$ est une suite indépendante d'événements, alors

1. $\mathbb{P}(\limsup A_n) = 0$ ou $\mathbb{P}(\limsup A_n) = 1$.
2. $\mathbb{P}(\liminf A_n) = 0$ ou $\mathbb{P}(\liminf A_n) = 1$.

Le lemme suivant donne des conditions suffisantes permettant de préciser laquelle des deux valeurs possibles est la bonne :

Proposition 4.29.

Lemme de Borel-Cantelli

1. Soit $(A_n)_{\mathbb{N}}$ une suite d'événements (non nécessairement indépendante). Si la série $\sum \mathbb{P}(A_n)$ de terme général positif $\mathbb{P}(A_n)$ converge dans \mathbb{R} , alors $\mathbb{P}(\limsup A_n) = 0$, c'est-à-dire presque-sûrement seul un nombre fini des événements A_n est réalisé.
2. Soit $(A_n)_{\mathbb{N}}$ une suite d'événements indépendante. Si la série $\sum \mathbb{P}(A_n)$ de terme général positif $\mathbb{P}(A_n)$ diverge dans \mathbb{R} , alors $\mathbb{P}(\limsup A_n) = 1$, c'est-à-dire presque-sûrement un nombre infini des événements A_n est réalisé.

Démonstration :

1. Posons, pour tout entier naturel m , $B_m = \bigcup_{k=m}^{k=+\infty} A_k$. La suite ensembliste $(B_m)_{\mathbb{N}}$ est une suite décroissante (au sens de l'inclusion). D'après le théorème de continuité monotone des probabilités (cf. proposition 1.9, page 12), $\mathbb{P}(\limsup A_n) = \lim_{m \rightarrow +\infty} \mathbb{P}(B_m)$.

Or $\mathbb{P}(B_m) = \mathbb{P}\left(\bigcup_{k=m}^{k=+\infty} A_k\right) \leq \sum_{k=m}^{k=+\infty} \mathbb{P}(A_k)$, en vertu de l'inégalité de Bonferroni. Mais

$\sum_{k=m}^{+\infty} \mathbb{P}(A_k)$ est le reste de rang m de la série $\sum \mathbb{P}(A_n)$ convergente par hypothèse, donc

$\lim_{m \rightarrow +\infty} \sum_{k=m}^{+\infty} \mathbb{P}(A_k) = 0$, et par suite $\lim_{m \rightarrow +\infty} \mathbb{P}(B_m) = 0$. Donc $\mathbb{P}(\limsup A_n) = 0$, ce qu'il fallait démontrer.

2. Posons, pour simplifier les écritures, $A = \limsup A_n$. Comme $A = \bigcap_{p=0}^{p=+\infty} \left(\bigcup_{k=p}^{k=+\infty} A_k \right)$,

il vient en passant au complémentaire, $A^c = \bigcup_{p=0}^{p=+\infty} \left(\bigcap_{k=p}^{k=+\infty} A_k^c \right)$. Posons, pour tout

entier naturel p , $B_p^c = \bigcap_{k=p}^{k=+\infty} A_k^c$. Notons que la suite ensembliste $(B_p^c)_{\mathbb{N}}$ est crois-

sante. On peut aussi écrire $B_p^c = \bigcap_{m=p}^{m=+\infty} \left(\bigcap_{k=p}^{k=m} A_k^c \right)$. Donc, pour tout entier naturel

$p, \mathbb{P}(B_p^c) = \mathbb{P} \left[\bigcap_{m=p}^{m=+\infty} \left(\bigcap_{k=p}^{k=m} A_k^c \right) \right]$. Comme la suite ensembliste $\left(\bigcap_{k=p}^{k=m} A_k^c \right)_{m \in \mathbb{N}}$ est décroissante, par le théorème de continuité monotone des probabilités, il vient $\mathbb{P}(B_p^c) = \lim_{m \rightarrow +\infty} \mathbb{P} \left(\bigcap_{k=p}^{k=m} A_k^c \right)$. Comme la suite des événements $(A_n)_{\mathbb{N}}$ est indépendante, il en est de même de la suite des événements $(A_n^c)_{\mathbb{N}}$ (cf. proposition 4.21, page 80). Par suite, pour tout entier naturel p , $\mathbb{P}(B_p^c) = \lim_{m \rightarrow +\infty} \mathbb{P} \left(\bigcap_{k=p}^{k=m} A_k^c \right) = \lim_{m \rightarrow +\infty} \prod_{k=p}^{k=m} \mathbb{P}(A_k^c)$. Donc finalement $\mathbb{P}(B_p^c) = \lim_{m \rightarrow +\infty} \prod_{k=p}^{k=m} (1 - \mathbb{P}(A_k))$. La série $\sum \ln(1 - \mathbb{P}(A_n))$ a son terme général négatif qui vérifie, pour tout entier naturel n , $\ln(1 - \mathbb{P}(A_n)) \leq -\mathbb{P}(A_n)$. Comme, par hypothèse, la série $\sum \mathbb{P}(A_n)$ de terme général positif $\mathbb{P}(A_n)$ diverge dans \mathbb{R} , la série $\sum \ln(1 - \mathbb{P}(A_n))$ de terme général négatif $\ln(1 - \mathbb{P}(A_n))$ diverge vers $-\infty$, et donc, pour tout entier naturel n , $\lim_{m \rightarrow +\infty} \prod_{k=p}^{k=m} (1 - \mathbb{P}(A_k)) = 0$. D'où, pour tout entier naturel p , $\mathbb{P}(B_p^c) = 0$, et par suite $\mathbb{P}(A^c) = \mathbb{P} \left(\bigcup_{p=0}^{p=+\infty} B_p^c \right) = \lim_{p \rightarrow +\infty} \mathbb{P}(B_p^c) = 0$, par le théorème de continuité monotone appliqué à la suite croissante $(B_p^c)_{\mathbb{N}}$. Finalement $\mathbb{P}(A) = 1$. \square

On notera que si l'hypothèse d'indépendance n'est pas utile dans l'item 1) du lemme de Borel-Cantelli, elle est par contre nécessaire dans l'item 2) car, sans cette hypothèse, on peut construire des contre-exemples où $\mathbb{P}(\limsup A_n) = 0$ avec la série de terme général $\mathbb{P}(A_n)$ divergente. En effet, considérons l'exemple suivant :

Exemples 4.10.

Soit l'espace de probabilité $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \lambda)$ où λ désigne la mesure de Lebesgue sur \mathbb{R} . Posons, pour tout entier naturel n , $A_n =]0, \frac{1}{n+1}]$. Alors on vérifie aisément que $\limsup A_n = \emptyset$, d'où $\mathbb{P}(\limsup A_n) = 0$, mais, pour tout entier naturel n , $\mathbb{P}(A_n) = \frac{1}{n+1}$, ce qui entraîne que la série $\sum \mathbb{P}(A_n)$ à terme général positif diverge. Ici la suite d'événements $(A_n)_{\mathbb{N}}$ n'est pas indépendante car par exemple $\mathbb{P}(A_1 \cap A_2) = \mathbb{P}(A_2) = \frac{1}{3}$ alors que $\mathbb{P}(A_1) \cdot \mathbb{P}(A_2) = \frac{1}{2} \times \frac{1}{3} = \frac{1}{6} \neq \mathbb{P}(A_1 \cap A_2)$. \square

4.4 Somme de v.a.r. indépendantes

Démontrons d'abord un important corollaire de l'exercice 3.14, page 49 :

Proposition 4.30.

Si (X_1, X_2, \dots, X_n) est une suite indépendante de variables aléatoires réelles de carré intégrable, alors

$$\mathbb{E} \left[\left(\sum_{k=1}^n X_k \right)^2 \right] < +\infty \text{ et } \text{Var} \left(\sum_{k=1}^n X_k \right) = \sum_{k=1}^n \text{Var}(X_k).$$

La réciproque est fausse.

Démonstration : Montrons l'intégrabilité. On effectue les majorations

$$\left(\sum_{k=1}^n X_k \right)^2 \leq \left(\sum_{k=1}^n |X_k| \right)^2 = \sum_{k=1}^n |X_k|^2 + 2 \sum_{1 \leq k < l \leq n} |X_k| |X_l|.$$

Mais $|X_k| |X_l| \leq \frac{1}{2}(|X_k|^2 + |X_l|^2)$, d'où

$$\sum_{k=1}^n |X_k|^2 + 2 \sum_{1 \leq k < l \leq n} |X_k| |X_l| \leq K \sum_{k=1}^n |X_k|^2$$

où K est une constante. Par suite, grâce aux hypothèses d'intégrabilité sur les variables aléatoires réelles

$$\mathbb{E} \left[\left(\sum_{k=1}^n X_k \right)^2 \right] \leq K \mathbb{E} \left(\sum_{k=1}^n |X_k|^2 \right) = K \sum_{k=1}^n \mathbb{E} (|X_k|^2) < +\infty.$$

On vérifie aisément que

$$\text{Var} \left(\sum_{k=1}^n X_k \right) = \sum_{k=1}^n \text{Var}(X_k) + 2 \sum_{1 \leq i < j \leq n} \text{Cov}(X_i, X_j)$$

et on conclut en remarquant que, par indépendance des v.a.r. , $\text{Cov}(X_i, X_j) = 0$ pour tout couple d'entiers (i, j) tel que $i \neq j$. \square

Exercice 4.8. (Corrigé de l'exercice : page 169)

Trouver, parmi les exercices ou exemples déjà proposés, un contre-exemple prouvant que la réciproque de l'implication précédente est fausse.

On peut généraliser la proposition 4.30 aux vecteurs aléatoires en montrant que, si (X, Y) est un couple indépendant de vecteurs aléatoires de dimension d et de carré intégrable, alors $D_{X+Y} = D_X + D_Y$, où D_X désigne la matrice de dispersion de X .

Nous allons maintenant donner quelques résultats sur la somme de v.a.r. indépendantes suivant des lois classiques. Auparavant énonçons un corollaire du critère des fonctions caractéristiques qui sera commode dans la recherche des lois de sommes de variables aléatoires réelles indépendantes.

Proposition 4.31.

Si (X_1, X_2, \dots, X_n) est une suite indépendante de v.a.r., alors, pour tout $u \in \mathbb{R}$,

$$\begin{aligned} \phi_{X_1+\dots+X_n}(u) &= \phi_{X_1}(u) \cdots \phi_{X_n}(u) \\ \text{i.e. } \mathbb{E} \left[\exp \left(iu \sum_{k=1}^n X_k \right) \right] &= \mathbb{E}(e^{iuX_1}) \cdots \mathbb{E}(e^{iuX_n}). \end{aligned}$$

La réciproque est fausse.

Démonstration : Il suffit de prendre $u := u_1 = u_2 = \dots = u_n$ dans la condition nécessaire du critère des fonctions caractéristiques. \square

Exercice 4.9. (Corrigé de l'exercice : page 169)

Pour prouver que la réciproque est fausse, construire un contre-exemple en considérant les deux variables aléatoires réelles X et $Y := X$ où X est une variable aléatoire réelle de

Cauchy de paramètre 1 i.e. de densité ρ définie sur \mathbb{R} par $\rho(x) := \frac{1}{\pi(1+x^2)}$ dont la

fonction caractéristique ϕ est définie sur \mathbb{R} par $\phi(t) := e^{-|t|}$. Pour montrer que (X, Y) est non indépendant, on considérera le borélien $A = [-1, 1]$ et on calculera $\mathbb{P}_{(X,Y)}(A \times A^c)$ qu'on comparera au produit $\mathbb{P}_X(A)\mathbb{P}_Y(A^c)$.

A titre d'application de la proposition précédente, voici un résultat qui sera fondamental dans le chapitre sur les vecteurs gaussiens.

Proposition 4.32.**Stabilité des lois normales**

Si (X_1, X_2, \dots, X_p) est une suite indépendante de v.a.r. normales de lois respectives $\mathcal{N}_1(m_1, \sigma_1^2), \dots, \mathcal{N}_1(m_p, \sigma_p^2)$, alors la v.a.r. $S_p := X_1 + \dots + X_p$ est une v.a.r. normale d'espérance $m_1 + \dots + m_p$ et de variance $\sigma_1^2 + \dots + \sigma_p^2$.

Démonstration : On applique le résultat précédent 4.31 en notant que la fonction caractéristique d'une variable aléatoire réelle X de loi $\mathcal{N}_1(m, \sigma^2)$ est $\phi_X(t) = \exp(imt - \frac{1}{2}t^2\sigma^2)$. On applique ensuite le théorème d'injectivité 3.26 en remarquant que la fonction caractéristique de la variable aléatoire réelle S_p est celle de la loi $\mathcal{N}_1(m_1 + \dots + m_p, \sigma_1^2 + \dots + \sigma_p^2)$. \square

Citons, à titre d'exemple, un autre résultat fondamental dont la démonstration est laissée en exercice :

Proposition 4.33.

Si (X_1, X_2, \dots, X_n) est une suite indépendante de variables aléatoires réelles de Bernoulli de même paramètre $p \in]0, 1[$, alors la variable aléatoire réelle $S_n := X_1 + \dots + X_n$ suit la loi binomiale $\mathcal{B}(n, p)$.

Exercice 4.10. (Corrigé de l'exercice : page 169)

Démontrer la proposition précédente.

Exercice 4.11. (Corrigé de l'exercice : page 169)

Soit (X_1, \dots, X_n) une suite indépendante de n variables aléatoires réelles de même densité

ρ définie sur \mathbb{R} par $\rho(x) := \alpha e^{-\alpha x} \mathbb{1}_{[0, +\infty[}(x)$ où $\alpha > 0$. Pour tout $\omega \in \Omega$, on range les nombres réels $X_1(\omega), \dots, X_n(\omega)$ dans l'ordre décroissant et on note $X_{(k)}(\omega)$ le $k^{\text{ième}}$ de ces nombres ainsi rangés où $k = 1, \dots, n$.

1. Vérifier que $X_{(1)} = \max(X_1, \dots, X_n)$ et calculer sa fonction de répartition.
2. Que représente la variable aléatoire réelle $X_{(n)}$? Calculer sa fonction de répartition.
3. Soit $t > 0$. Pour tout $k = 1, \dots, n$, on pose $Y_k := \mathbb{1}_{]t, +\infty[}(X_k)$.
 - (a) Prouver que la loi de la variable aléatoire réelle Y_k est $\mathbb{P}_{Y_k} = e^{-\alpha t} \delta_1 + (1 - e^{-\alpha t}) \delta_0$.
Quelle est la loi de la variable aléatoire réelle $Y_1 + Y_2 + \dots + Y_n$?
 - (b) Comparer les événements $\{X_{(k)} \leq t\}$ et $\{Y_1 + Y_2 + \dots + Y_n \leq k - 1\}$.
4. Déterminer, pour tout $k = 1, \dots, n$, la fonction de répartition de la variable aléatoire réelle $X_{(k)}$.

Dans certains cas on peut directement calculer la loi de la variable aléatoire réelle "somme". En voici deux exemples énoncés sous forme de propositions 4.34 et 4.36, la démonstration de la seconde proposition 4.36 est laissée en exercice.

Proposition 4.34.

Stabilité des lois binomiales

Si (X, Y) est un couple indépendant de variables aléatoires réelles de lois binomiales respectives $\mathcal{B}(n, p)$ et $\mathcal{B}(m, p)$ de même paramètre $p \in]0, 1[$, alors la variable aléatoire réelle $X + Y$ suit la loi binomiale $\mathcal{B}(n + m, p)$.

Démonstration : Dans ce qui suit on adoptera la convention d'écriture : $C_n^j := 0$ pour tout entier $j > n$ ou $j < 0$. X et Y sont des lois discrètes portées respectivement par $\{0, 1, \dots, n\}$ et $\{0, 1, \dots, m\}$. La loi de la variable aléatoire réelle $X + Y$ sera aussi discrète et portée par $\{0, 1, \dots, n + m\}$. D'après la proposition 2.10, page 27, il suffit de calculer, pour tout entier $0 \leq k \leq n + m$, le nombre $\mathbb{P}(X + Y = k)$. Par l'égalité ensembliste facile à vérifier

$$\{X + Y = k\} = \bigcup_{j=0}^{j=k} (\{X = j\} \cap \{Y = k - j\})$$

et comme l'union est deux à deux disjointe il vient, en appliquant le critère d'indépendance des variables aléatoires réelles discrètes au couple (X, Y) ,

$$\begin{aligned} \mathbb{P}(X + Y = k) &= \sum_{j=0}^{j=k} \mathbb{P}(X = j) \mathbb{P}(Y = k - j) = \sum_{j=0}^{j=k} C_n^j C_m^{k-j} p^k (1 - p)^{n+m-k} \\ &= C_{n+m}^k p^k (1 - p)^{n+m-k}. \end{aligned}$$

La dernière égalité du calcul précédent résulte de la formule de Vandermonde rappelée ci-dessous (proposition 4.35). La loi de la variable aléatoire réelle $X + Y$ s'écrit alors

$$\mathbb{P}_{X+Y} = \sum_{k=0}^{n+m} C_{n+m}^k p^k (1 - p)^{n+m-k} \delta_k.$$

On reconnaît la loi binomiale $\mathcal{B}(n + m, p)$. \square

Proposition 4.35.**Formule de Vandermonde**

Soient n et m deux entiers naturels non nuls. Pour tout entier naturel k vérifiant $0 \leq k \leq n+m$, on a

$$\sum_{j=0}^{j=k} C_n^j C_m^{k-j} = C_{n+m}^k,$$

avec la convention habituelle sur les C_n^i (cf. formulaire de l'annexe A, page 205)

Démonstration : Il suffit de développer de deux manières différentes l'égalité $(1+X)^{n+m} = (1+X)^n(1+X)^m$ et d'égaliser le coefficient de X^k des deux expressions, où $0 \leq k \leq n+m$. \square

Proposition 4.36.**Stabilité des lois de Poisson**

Si (X, Y) est un couple indépendant de variables aléatoires réelles de lois de Poisson respectives $\mathcal{P}(\alpha)$ et $\mathcal{P}(\beta)$, où les réels α et β sont strictement positifs, alors la variable aléatoire réelle $X+Y$ suit la loi de Poisson $\mathcal{P}(\alpha+\beta)$.

Exercice 4.12. (Corrigé de l'exercice : page 170)

Donner deux démonstrations de la proposition précédente, l'une directe en s'inspirant de la démonstration pour le cas des variables binomiales et l'autre en utilisant les fonctions caractéristiques.

Voici quelques résultats plus généraux sur les sommes de variables aléatoires réelles indépendantes. Plus que de retenir des formules, Il faut surtout être capable de refaire directement les calculs dans chaque cas particulier.

Dans la proposition suivante la notation $*$ désigne le **produit de convolution** de deux fonctions f et g positives boréliennes de \mathbb{R} dans \mathbb{R} défini comme l'application

$$f * g : x \in \mathbb{R} \mapsto f * g(x) := \int_{\mathbb{R}} f(x-u)g(u)d\lambda(u) \in [0, +\infty].$$

Proposition 4.37.

Soit (X, Y) un couple indépendant de variables aléatoires réelles admettant pour densités respectives ρ_X et ρ_Y , alors la variable aléatoire réelle $X+Y$, admet pour densité l'application $\rho_{X+Y} := \rho_X * \rho_Y$.

Démonstration : Soit h une application positive borélienne définie sur \mathbb{R} . Par les théorèmes de

transfert, de Tonelli et l'indépendance de (X, Y) , il vient

$$\begin{aligned}
 \mathbb{E}[h(X + Y)] &= \int_{\mathbb{R}^2} h(x + y) d\mathbb{P}_{(X, Y)}(x, y) = \int_{\mathbb{R}^2} h(x + y) d\mathbb{P}_X \otimes \mathbb{P}_Y(x, y) \\
 &= \int_{\mathbb{R}} \left(\int_{\mathbb{R}} h(x + y) d\mathbb{P}_X(x) \right) d\mathbb{P}_Y(y) \\
 &= \int_{\mathbb{R}} \left(\int_{\mathbb{R}} h(x + y) \rho_X(x) d\lambda(x) \right) \rho_Y(y) d\lambda(y) \\
 &= \int_{\mathbb{R}^2} h(x + y) \rho_X(x) \rho_Y(y) d\lambda \otimes \lambda(x, y) \\
 &= \int_{\mathbb{R}^2} h(x + y) \rho_X(x) \rho_Y(y) d\lambda^{(2)}(x, y)
 \end{aligned}$$

Effectuons le changement de variables, $(u, v) := (x + y, x)$. Le jacobien de l'application inverse en (u, v) est $J(u, v) = -1$ d'où, par application du théorème de changement de variable à la dernière intégrale précédente,

$$\begin{aligned}
 \mathbb{E}[h(X + Y)] &= \int_{\mathbb{R}^2} h(u) \rho_X(v) \rho_Y(u - v) d\lambda^{(2)}(u, v) \\
 &= \int_{\mathbb{R}} h(u) \left(\int_{\mathbb{R}} \rho_X(v) \rho_Y(u - v) d\lambda(v) \right) d\lambda(u) \\
 &= \int_{\mathbb{R}^d} h(u) (\rho_X * \rho_Y)(u) d\lambda^{(d)}(u),
 \end{aligned}$$

ce qui prouve que $\rho_X * \rho_Y$ est bien la densité de $X + Y$. \square

Exemples 4.11.

Soit (X, Y) un couple indépendant de v.a.r.. On suppose que la variable aléatoire réelle X suit la loi uniforme $\mathcal{U}([0, 1])$ définie par la densité $\rho_X := \mathbb{1}_{[0, 1]}$ et Y suit la loi exponentielle de paramètre 1 de densité ρ_Y définie sur \mathbb{R} par $\rho_Y(t) := \mathbb{1}_{[0, +\infty[}(t) e^{-t}$. La densité de la variable aléatoire réelle $X + Y$ est définie sur \mathbb{R} par $\rho_{X+Y}(t) := \int_{-\infty}^{+\infty} \mathbb{1}_{[0, 1]}(t - x) \mathbb{1}_{[0, +\infty[}(x) e^{-x} dx$. Ce qui après le calcul de l'intégrale donne $\rho_{X+Y}(t) = (1 - e^{-t}) \mathbb{1}_{[0, 1]}(t) + e^{-t}(e - 1) \mathbb{1}_{1, +\infty[}(t)$. \square

Proposition 4.38.

Soit (X, Y) un couple indépendant de v.a.r.. On suppose que X admet une densité ρ_X et que Y est une variable discrète portée par \mathbb{N} de loi $\sum_{k=0}^{+\infty} p_k \delta_k$. Alors la variable aléatoire réelle $Z := X + Y$ admet pour densité l'application

$$\rho_{X+Y} : x \in \mathbb{R} \mapsto \rho_{X+Y}(x) := \sum_{k=0}^{+\infty} p_k \rho_X(x - k) \in [0, +\infty].$$

Démonstration : Soit h une application de \mathbb{R} dans \mathbb{R} borélienne positive. En utilisant le théorème du transfert puis l'indépendance de (X, Y) avec le théorème de Tonelli, il vient

$$\begin{aligned}
 \mathbb{E}[h(X + Y)] &= \int_{\mathbb{R}^2} h(x + y) d\mathbb{P}_{(X, Y)}(x, y) = \int_{\mathbb{R}^2} h(x + y) d(\mathbb{P}_X \otimes \mathbb{P}_Y)(x, y) \\
 &= \int_{\mathbb{R}} \left(\int_{\mathbb{R}} h(x + y) d\mathbb{P}_X(x) \right) d\mathbb{P}_Y(y) \\
 &= \int_{\mathbb{R}} \left(\int_{\mathbb{R}} h(x + y) \rho_X(x) d\lambda(x) \right) d \left(\sum_{k=0}^{+\infty} p_k \delta_k \right) (y) \\
 &= \sum_{k=0}^{+\infty} p_k \int_{\mathbb{R}} h(x + k) \rho_X(x) d\lambda(x).
 \end{aligned}$$

Appliquons alors à $\int_{\mathbb{R}} h(x + k) \rho_X(x) d\lambda(x)$ le changement de variable défini sur \mathbb{R} , pour $k \in \mathbb{N}$ fixé, par $u := x + k$. Il vient

$$\int_{\mathbb{R}} h(x + k) \rho_X(x) d\lambda(x) = \int_{\mathbb{R}} h(u) \rho_X(u - k) d\lambda(u).$$

Par suite, revenant aux égalités précédentes,

$$\begin{aligned}
 \mathbb{E}[h(X + Y)] &= \sum_{k=0}^{+\infty} p_k \int_{\mathbb{R}} h(x + k) \rho_X(x) d\lambda(x) = \sum_{k=0}^{+\infty} p_k \int_{\mathbb{R}} h(u) \rho_X(u - k) d\lambda(u) \\
 &= \int_{\mathbb{R}} \left(\sum_{k=0}^{+\infty} p_k \rho_X(u - k) \right) h(u) d\lambda(u)
 \end{aligned}$$

La dernière égalité se justifie par la propriété de Beppo-Lévi vue au chapitre III. D'où le résultat. \square

Exemples 4.12.

Soit (X, Y) un couple indépendant de v.a.r. . On suppose que la v.a.r. X suit la loi de Gauss-Laplace $\mathcal{N}_1(0, 1)$, et Y suit la loi de Poisson de paramètre $\alpha > 0$. Alors la densité de la v.a.r. $X + Y$ est définie sur \mathbb{R} par

$$\rho_{X+Y}(t) := \sum_{k=0}^{+\infty} \frac{e^{-\alpha} \alpha^k}{k! \sqrt{2\pi}} \exp\left(-\frac{1}{2}(t - k)^2\right) = \frac{e^{-\alpha}}{\sqrt{2\pi}} \sum_{k=0}^{+\infty} \frac{\alpha^k}{k!} \exp\left(-\frac{1}{2}(t - k)^2\right). \square$$

4.5 Exercices de révision sur les chapitres I à IV

Exercice 4.13. (Corrigé de l'exercice : page 171)

Soient A et B deux v.a.r. indépendantes de loi uniforme $\mathcal{U}([0, 1])$. Quelle est la probabilité que le polynôme $x^2 - 2Ax + B$ ait :

1. deux racines réelles distinctes,
2. deux racines complexes et non réelles,

3. une racine double.
4. Traiter les questions précédentes en utilisant la loi de la v.a.r. $\Delta := A^2 - B$.

Exercice 4.14. (Corrigé de l'exercice : page 172)

On considère une variable aléatoire (X, Y) à valeurs dans \mathbb{R}^2 dont la loi $\mathbb{P}_{(X,Y)}$ admet la densité

$$f(x, y) := \alpha(1 - x^2)\mathbb{1}_{[0,1]}(x)ye^{-3y}\mathbb{1}_{[0,+\infty[}(y),$$

où α est un réel, par rapport à la mesure de Lebesgue $\lambda^{(2)}$ sur \mathbb{R}^2 .

1. Déterminer la valeur du réel α .
2. Déterminer les lois marginales du couple (X, Y) .
3. Calculer $\mathbb{P}(0 < X \leq 2, Y \geq 1)$.
4. Calculer la matrice de dispersion D de (X, Y) .

Exercice 4.15. (Corrigé de l'exercice : page 174)

Soient n et m deux entiers naturels non nuls, X et Y deux variables aléatoires réelles indépendantes. On suppose que la variable aléatoire réelle X est binomiale de paramètres n et $\frac{1}{2}$, et que la variable aléatoire réelle Y est binomiale de paramètres m et $\frac{1}{2}$. Calculer la probabilité que $X = Y$.

Exercice 4.16. (Corrigé de l'exercice : page 174)

Soit $(X_k)_{k \in \mathbb{N}}$ une suite indépendante de v.a.r. de Bernoulli toutes de même paramètre $0 < p < 1$. Soit un entier $r \geq 1$, on définit deux nouvelles v.a.r. , en posant pour tout $\omega \in \Omega$,

$$\tau_r(\omega) := \inf\{n \in \mathbb{N}^* / X_1(\omega) + X_2(\omega) + \dots + X_n(\omega) = r\}$$

et

$$\theta_r(\omega) := \inf\{n \in \mathbb{N}^* / X_1(\omega) + X_2(\omega) + \dots + X_{n+r}(\omega) = r\}$$

avec la convention $\inf \emptyset := +\infty$.

1. Montrer, pour tout $x \in]0, 1[$, la relation

$$\sum_{k=r-1}^{+\infty} C_k^{r-1} x^{k-r+1} = \frac{1}{(1-x)^r}.$$

2. Montrer que la variable aléatoire réelle τ_r est une variable aléatoire réelle discrète de loi (dite **loi de Pascal de paramètres r et p**)

$$\mathcal{P}(r, p) := \sum_{k=r}^{+\infty} C_{k-1}^{r-1} p^r (1-p)^{k-r} \delta_k.$$

Vérifier que $\mathbb{P}(\tau_r = +\infty) = 0$.

3. Montrer que la variable aléatoire réelle θ_r est une variable aléatoire réelle discrète de loi (dite **loi binomiale-négative de paramètres r et p**)

$$\mathcal{I}(r, p) := \sum_{k=0}^{+\infty} C_{k+r-1}^{r-1} p^r (1-p)^k \delta_k.$$

Vérifier que $\mathbb{P}(\theta_r = +\infty) = 0$.

4. Donner une interprétation des variables aléatoires réelles τ_r et θ_r en terme de jeu de Pile-ou-Face.

5. Montrer qu'un des deux modèles précédents permet de formaliser le problème dit **des boîtes d'allumettes de Stephan Banach** :

Un fumeur a dans chacune de ses deux poches une boîte contenant au départ N allumettes. Chaque fois qu'il désire fumer une cigarette, il choisit une poche au hasard. Quelle est la probabilité que, le fumeur se rendant compte pour la première fois qu'une boîte est vide, l'autre boîte contienne k allumettes où k est un entier naturel inférieur ou égal à N ?

Exercice 4.17. (Corrigé de l'exercice : page 176)

On considère une v.a.r. réelle positive X de fonction de répartition F_X .

1. Pour tout entier naturel n non nul, en considérant sur l'espace mesuré $(\Omega \times \mathbb{R}^+, \mathcal{F} \otimes \mathcal{B}(\mathbb{R}^+), \mathbb{P} \otimes \lambda)$, où λ est la mesure de Lebesgue, la fonction à valeurs réelles définie par $H(\omega, t) := nt^{n-1} \mathbb{1}_{t, +\infty[}(X(\omega))$, montrer que

$$\mathbb{E}[X^n] = \int_0^{+\infty} nt^{n-1} \mathbb{P}(X > t) dt = \int_0^{+\infty} nt^{n-1} (1 - F_X(t)) dt.$$

2. Montrer par un exemple que l'hypothèse X positive est nécessaire.

3. En utilisant le résultat de la première question, calculer l'espérance et la variance des variables aléatoires suivantes :

(a) X de fonction de répartition $F_X(t) := t \mathbb{1}_{[0,1]}(t) + \mathbb{1}_{1,+\infty[}(t)$.

(b) Y de fonction de répartition $F_Y(t) := (1 - e^{-\alpha t}) \mathbb{1}_{[0,+\infty[}(t)$ où $\alpha > 0$.

(c) Z de fonction de répartition $F_Z(t) := \sum_{n=0}^{+\infty} e^{-\alpha} \frac{\alpha^n}{n!} \mathbb{1}_{[n,+\infty[}(t)$

Exercice 4.18. (Corrigé de l'exercice : page 177)

On considère $(X_n)_{\mathbb{N}^*}$ une suite indépendante de v.a.r. de même loi de Bernoulli $\mathcal{B}(p)$ définies sur un même espace de Probabilité $(\Omega, \mathcal{F}, \mathbb{P})$. Pour $\omega \in \Omega$ on définit

$$T(\omega) := \inf\{n \in \mathbb{N} / X_n(\omega) = 1\}$$

avec la convention $\inf \emptyset = +\infty$. Montrer que T est une v.a. à valeurs dans $\overline{\mathbb{N}}^*$, déterminer sa loi et calculer son espérance.

Exercice 4.19. (Corrigé de l'exercice : page 178)

Le but de cet exercice est de montrer qu'il n'existe pas de probabilité \mathbb{P} sur l'espace $(\mathbb{N}^*, \mathcal{P}(\mathbb{N}^*))$ telle que, pour tout $n \geq 1$, $\mathbb{P}(n\mathbb{N}^*) = \frac{1}{n}$ où $n\mathbb{N}^* = \{nk, k \in \mathbb{N}^*\}$.

Supposons qu'une telle probabilité existe. Soit $(p_k)_{\mathbb{N}}$ la suite des nombres entiers premiers rangés en ordre croissant.

1. Par un raisonnement simple montrer que $\mathbb{P}(\limsup_k (p_k \mathbb{N}^*)) = 0$.

2. Montrer que la suite $(p_k \mathbb{N}^*)_{\mathbb{N}}$ est indépendante. En déduire, en utilisant le fait que la série $\sum_k \frac{1}{p_k} = +\infty$, une autre valeur de $\mathbb{P}(\limsup_k (p_k \mathbb{N}^*))$. Conclure que la probabilité \mathbb{P} n'existe pas.

Exercice 4.20. (Corrigé de l'exercice : page 178)

Soit $(X_i)_I$ une famille quelconque de v.a.r. sur un même espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$. On note $\sigma(X_i, i \in I)$ la plus petite (au sens de l'inclusion) des sous-tribus \mathcal{A} de \mathcal{F} telles que, pour tout $i \in I$, X_i est \mathcal{A} -mesurable (Cette définition est une généralisation, au cas de plusieurs v.a.r., de la définition 4.7, page 80)

1. Justifier l'existence de $\sigma(X_i, i \in I)$.
2. Soit \mathcal{C} la famille des intersections finies d'événements de la forme $\{X_i \in B\}$ où $i \in I$ et $B \in \mathcal{B}(\mathbb{R})$. Montrer que $\sigma(X_i, i \in I)$ est la tribu sur Ω engendrée par \mathcal{C} .
3. On suppose que la famille de v.a.r. $(X_i)_I$ est indépendante. Montrer que, si J et K sont deux parties disjointes et non vides de I , alors les tribus $\sigma(X_i, i \in J)$ et $\sigma(X_i, i \in K)$ sont indépendantes.

Exercice 4.21. (Corrigé de l'exercice : page 179)

Déterminer la loi d'une somme de variables aléatoires indépendantes dans les cas suivants :

1. $Y := X_1 + X_2$ où X_1 suit une loi gamma $\gamma(a, \alpha)$ et X_2 une loi gamma $\gamma(b, \alpha)$.
2. $Z := X_1 + \dots + X_n$ où, pour tout $k = 1, \dots, n$, X_k suit la loi exponentielle $\mathcal{E}(\alpha)$.

Exercice 4.22. (Corrigé de l'exercice : page 179)

On considère $(X_j)_{j \geq 1}$ une suite indépendante de v.a.r. de même loi sur un même espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$, et donc de même fonction caractéristique Φ . Soit Y une v.a.r. discrète à valeurs dans \mathbb{N}^* et indépendante de la suite $(X_j)_{j \geq 1}$. En utilisant la fonction définie sur $] - 1, 1[$ par $\psi(t) := \sum_{n \geq 1} t^n \mathbb{P}(Y = n)$, déterminer la fonction caractéristique Φ_Z de la v.a.r. Z définie, pour presque tout $\omega \in \Omega$, par

$$Z(\omega) := \sum_{j=1}^{j=Y(\omega)} X_j(\omega).$$

Exercice 4.23. (Corrigé de l'exercice : page 180)

Soient X et Y deux v.a.r. à valeurs dans \mathbb{N}^* , indépendantes et de même loi sur un espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$. On définit les v.a.r. $D := X - Y$ et $M := \min(X, Y)$.

1. On suppose que X et Y suivent la **loi géométrique** de paramètre $p \in]0, 1[$ i.e.

$$\mathbb{P}_X = \mathbb{P}_Y := \sum_{k=1}^{\infty} p q^{k-1} \delta_k \text{ où } q := 1 - p.$$

- (a) Montrer que, pour tout $(i, j) \in \mathbb{Z} \times \mathbb{N}^*$,

$$\{D = i\} \cap \{M = j\} = \begin{cases} \{X = i + j\} \cap \{Y = j\} & \text{si } i \geq 0, \\ \{X = j\} \cap \{Y = j - i\} & \text{si } i < 0, \end{cases}$$

et en déduire que

$$\mathbb{P}(D = i) = \frac{p^2}{1 - q^2} q^{|i|} \text{ et } \mathbb{P}(M = j) = \frac{p}{q^2} q^{2j} (q + 1).$$

- (b) Démontrer que les v.a.r. D et M sont indépendantes.

2. Réciproquement, on suppose que les v.a.r. D et M sont indépendantes, montrer que, pour tout entier $n \geq 1$,

$$\frac{\mathbb{P}[(X = n+1) \cap (Y = n)]}{\mathbb{P}[(X = n) \cap (Y = n)]} = \frac{\mathbb{P}(X = n+1)}{\mathbb{P}(X = n)} = \frac{\mathbb{P}(D = 1)}{\mathbb{P}(D = 0)}.$$

En déduire que les v.a.r. X et Y suivent une loi géométrique dont on déterminera le paramètre.

Exercice 4.24. (Corrigé de l'exercice : page [181](#))

Soit X et Y deux variables aléatoires uniformes sur l'intervalle $[0, \alpha]$ où α est un réel strictement positif. On suppose que les variables X et Y sont indépendantes. On pose $Z = \sup(X, Y) - \inf(X, Y)$ et $T = \frac{\inf(X, Y)}{\sup(X, Y)}$. Déterminer les lois des variables Z et T .

Chapitre 5

Vecteurs aléatoires gaussiens

Rappelons que, si m et σ sont des réels avec $\sigma > 0$, $\mathcal{N}_1(m, \sigma^2)$ désigne la mesure de probabilité sur \mathbb{R} admettant la densité ρ définie sur \mathbb{R} par

$$\rho(x) := \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right).$$

Afin de simplifier les énoncés des théorèmes sur les vecteurs gaussiens, on est amené à considérer la probabilité de Dirac au point $m \in \mathbb{R}$ comme un cas, dit **dégénéré**, de loi gaussienne et par suite on pose $\mathcal{N}_1(m, 0) := \delta_m$.

Dans la suite, M^* désignera la matrice-transposée, $\det(M)$ le déterminant de la matrice M , $\langle \cdot, \cdot \rangle$ et $\|\cdot\|$ le produit-scalaire et la norme usuels de \mathbb{R}^d où $d \in \mathbb{N}^*$. Sauf précision contraire, les vecteurs de \mathbb{R}^d seront repérés par leurs composantes dans la base canonique de \mathbb{R}^d . La base canonique de \mathbb{R}^d est le système de vecteurs (e_1, e_2, \dots, e_d) où, pour tout $i = 1, \dots, d$, e_i est le d -uplet dont tous les termes prennent la valeur 0 sauf le terme de rang i qui prend la valeur 1.

5.1 Vecteur gaussien

Définition 5.1.

Une variable aléatoire réelle de loi $\mathcal{N}_1(m, \sigma^2)$, où m est un réel et σ un réel positif ou nul, est dite **gaussienne**.

En clair, une variable aléatoire réelle gaussienne (de loi $\mathcal{N}_1(m, \sigma^2)$) est soit une variable aléatoire réelle normale d'espérance m et de variance $\sigma^2 > 0$, soit une variable aléatoire réelle de Dirac au point m (constante déterministe égale à m , c'est le cas où $\sigma^2 = 0$).

Définition 5.2.

Soient X un vecteur aléatoire de dimension d et (X_1, X_2, \dots, X_d) ses composantes dans la base canonique de \mathbb{R}^d . On dit que X est un **vecteur (aléatoire) gaussien de dimension d** si, pour tous réels a_1, a_2, \dots, a_d , la variable aléatoire réelle $a_1X_1 + a_2X_2 + \dots + a_dX_d$ est une variable aléatoire réelle gaussienne.

Exemples 5.1.

Une variable aléatoire réelle gaussienne est un vecteur gaussien de dimension 1.

Compte tenu que, dans un changement de base dans \mathbb{R}^d , les composantes dans une nouvelle base sont des combinaisons linéaires des composantes dans l'ancienne base, la propriété d'être gaussien pour un vecteur ne dépend pas de la base choisie pour exprimer les composantes du vecteurs.

La proposition suivante est un corollaire immédiat, dont la démonstration est laissée en exercice, de la définition.

Proposition 5.1.

Soit (X_1, X_2, \dots, X_d) une suite de v.a.r. . Si le vecteur aléatoire $X := (X_1, X_2, \dots, X_d)$ est un vecteur gaussien de dimension d alors, pour tout $k = 1, 2, \dots, d$, X_k est une variable aléatoire réelle gaussienne.

La réciproque est fausse.

Exemples 5.2.

Reprenons les hypothèses et notations de l'exercice 4.6, page 74. On vérifie facilement que $\mathbb{P}(X + Y = 0) = \frac{1}{2}$, et donc que la variable aléatoire réelle $X + Y$ n'est pas une v.a.r. gaussienne, sinon comme $X + Y$ n'est pas une variable déterministe (i.e. de loi une probabilité de Dirac), cela veut dire que $X + Y$ serait une v.a.r. normale et on aurait $\mathbb{P}(X + Y = 0) = 0$. Le vecteur aléatoire (X, Y) n'est donc pas gaussien (sinon $X + Y$ serait une v.a.r. gaussienne) alors que ses composantes sont des v.a.r. gaussiennes, ce qui donne un contre-exemple à la réciproque de la proposition 5.1 précédente. \square

En revanche si on rajoute une hypothèse d'indépendance sur la suite des composantes du vecteur X , on obtient un procédé simple de construction de vecteurs gaussiens de dimensions $d \geq 2$ grâce à la proposition :

Proposition 5.2.

Soit (X_1, X_2, \dots, X_d) une suite indépendante de v.a.r. . Le vecteur aléatoire $X := (X_1, X_2, \dots, X_d)$ est un vecteur gaussien de dimension d si, et seulement si, pour tout $k = 1, 2, \dots, d$, X_k est une variable aléatoire réelle gaussienne.

Démonstration : • La condition nécessaire résulte de la définition des vecteurs gaussiens et n'utilise pas l'hypothèse d'indépendance. C'est un cas particulier de la proposition précédente.
• La condition suffisante résulte de ce que, si (X_1, X_2, \dots, X_d) est une suite indépendante de v.a.r. , alors pour tous réels a_1, a_2, \dots, a_d , la suite $(a_1X_1, a_2X_2, \dots, a_dX_d)$ est indépendante. De plus si la variable aléatoire réelle X_k a pour loi $\mathcal{N}_1(m_k, \sigma_k^2)$, la variable aléatoire réelle a_kX_k a pour loi $\mathcal{N}_1(a_k m_k, a_k^2 \sigma_k^2)$. D'après la proposition 4.32, page 86, la variable aléatoire réelle $a_1X_1 + a_2X_2 + \dots + a_dX_d$ est alors une variable aléatoire réelle gaussienne comme somme de variables aléatoires réelles gaussiennes indépendantes. \square

La proposition 5.1 a aussi pour conséquence que si $X = (X_1, X_2, \dots, X_d)$ est un vecteur gaussien de dimension d , alors, pour tout $k = 1, 2, \dots, d$, X_k est une variable aléatoire réelle de carré intégrable car de loi gaussienne. Par suite on peut définir l'espérance $m := \mathbb{E}(X)$ et la matrice de dispersion $D_X := \mathbb{E}([X - \mathbb{E}(X)][X - \mathbb{E}(X)]^*)$ du vecteur gaussien X . L'espérance m est

un vecteur de \mathbb{R}^d et D_X est une matrice carrée d'ordre d à coefficients réels, symétrique et de type positif d'après la proposition 3.24, page 48, sur les propriétés des matrices de dispersion.

Exercice 5.1. (Corrigé de l'exercice : page 183)

Soit $(U_k)_{\mathbb{N}}$ une suite indépendante de variable aléatoire réelle de même loi normale centrée et de variance $\sigma^2 > 0$. Pour tout $\theta \in \mathbb{R}$, on définit la suite $(X_k)_{\mathbb{N}^*}$ où $X_k = \theta U_{k-1} + U_k$, pour tout entier $k \geq 2$, et $X_1 = U_1$.

Montrer que, pour tout $n \in \mathbb{N}^*$, $X := (X_1, X_2, \dots, X_n)$ est un vecteur gaussien.

L'espérance et la matrice de dispersion déterminent complètement la fonction caractéristique d'un vecteur gaussien comme le montre le résultat suivant :

Proposition 5.3.

Soit X un vecteur aléatoire de dimension d admettant une espérance $m := (m_1, \dots, m_d) \in \mathbb{R}^d$ et une matrice de dispersion D . Alors X est un vecteur gaussien si, et seulement si, sa fonction caractéristique Φ_X est donnée, pour tout $u \in \mathbb{R}^d$, par

$$\Phi_X(u) = \exp \left(i \langle u, m \rangle - \frac{1}{2} \langle u, Du \rangle \right) \text{ ou } \Phi_X(u) = \exp \left(i u^* m - \frac{1}{2} u^* D u \right).$$

Démonstration : • Montrons la condition nécessaire. Posons $X := (X_1, \dots, X_d)$, $u := (u_1, \dots, u_d)$ et $Y := u_1 X_1 + \dots + u_d X_d$. Comme X est un vecteur gaussien, la variable aléatoire réelle Y est de loi gaussienne i.e. $\mathbb{P}_Y = \mathcal{N}_1(m_Y, \sigma_Y^2)$. De plus

$$m_Y := \mathbb{E}(Y) = u_1 \mathbb{E}(X_1) + \dots + u_d \mathbb{E}(X_d) = u_1 m_1 + \dots + u_d m_d = \langle u, m \rangle = u^* m,$$

et

$$\begin{aligned} \sigma_Y^2 &= \mathbb{E}[(Y - m_Y)^2] = \mathbb{E} \left[(u_1(X_1 - m_1) + \dots + u_d(X_d - m_d))^2 \right] \\ &= \sum_{1 \leq i, j \leq d} u_i u_j \mathbb{E}[(X_i - m_i)(X_j - m_j)] \\ &= \sum_{1 \leq i, j \leq d} u_i u_j \text{Cov}(X_i, X_j) = \langle u, Du \rangle = u^* D u. \end{aligned}$$

Comme pour tout $u \in \mathbb{R}^d$,

$$\Phi_X(u) = \mathbb{E}(e^{i(u_1 X_1 + \dots + u_d X_d)}) = \mathbb{E}(e^{iY}) = \Phi_Y(1)$$

et que $\Phi_Y(1) = \exp(i m_Y - \frac{1}{2} \sigma_Y^2)$, on obtient $\Phi_X(u) = \exp(i \langle u, m \rangle - \frac{1}{2} \langle u, Du \rangle)$.

• Montrons la condition suffisante. Soit $X := (X_1, \dots, X_d)$ un vecteur aléatoire quelconque de fonction caractéristique définie sur \mathbb{R}^d par

$$\Phi_X(u) = \exp \left(i \langle u, m \rangle - \frac{1}{2} \langle u, Du \rangle \right).$$

Soit $Y := a_1 X_1 + \dots + a_d X_d$ une combinaison linéaire des composantes de X . En considérant la fonction caractéristique de Y , il vient, pour tout réel t ,

$$\begin{aligned} \Phi_Y(t) &= \mathbb{E}(e^{iYt}) = \mathbb{E}(e^{i(t a_1 X_1 + \dots + t a_d X_d)}) = \Phi_X(a_1 t, \dots, a_d t) \\ &= \exp \left(i t \langle a, m \rangle - \frac{1}{2} t^2 \langle a, Da \rangle \right) \end{aligned}$$

où on a posé $a := (a_1, \dots, a_d)$.

Ce qui prouve que, pour tout n -uplet de réels (a_1, \dots, a_d) , la variable aléatoire réelle $a_1 X_1 + \dots + a_d X_d$ est une variable aléatoire réelle gaussienne de loi $\mathcal{N}_1(\langle a, m \rangle, \langle a, Da \rangle)$. X est bien un vecteur gaussien. \square

5.2 Loi d'un vecteur gaussien

La dernière proposition 5.1 a pour conséquence que la loi d'un vecteur gaussien X de dimension d est entièrement déterminée par la connaissance des deux paramètres que sont l'espérance m et la matrice de dispersion D du vecteur gaussien X .

L'existence de vecteurs gaussiens d'espérance m et de matrice de dispersion D données a priori est affirmée par le résultat suivant que nous admettrons (pour la démonstration consulter [3] exercice VII-11) :

Proposition 5.4.

Si $m \in \mathbb{R}^d$ et D est une matrice carrée d'ordre d à coefficients réels, symétrique et de type positif, il existe un espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$ et un vecteur gaussien de dimension d sur $(\Omega, \mathcal{F}, \mathbb{P})$ d'espérance m et de matrice de dispersion D .

Les résultats précédents autorisent la définition suivante :

Définition 5.3.

On appelle loi de Gauss-Laplace ou loi normale sur \mathbb{R}^d de paramètres m et D la loi de probabilité d'un vecteur gaussien de dimension d d'espérance m et de matrice de dispersion D . Dans ce cas on note $\mathcal{N}_d(m, D)$ cette probabilité.

La proposition ci-dessous sera souvent utilisée pour prouver que certains vecteurs sont gaussiens :

Proposition 5.5.

Si X est un vecteur gaussien de dimension d , A une matrice rectangulaire $k \times d$ à coefficients réels et b un vecteur de dimension k , alors le vecteur aléatoire $Y := AX + b$ est un vecteur gaussien de dimension k . De plus si $\mathcal{N}_d(m, D)$ est la loi de X , la loi de Y est $\mathcal{N}_k(Am + b, ADA^)$,*

Démonstration : Il suffit de remarquer que les composantes de AX sont des combinaisons linéaires des composantes de X . Donc toute combinaison linéaire des composantes de Y est une combinaison linéaire des composantes de X (qui est une v.a.r. gaussienne car X est supposé gaussien) à laquelle on ajoute une constante, dont la somme est encore une v.a.r. gaussienne. Y suit donc une loi normale de dimension k . Il reste à en préciser les paramètres : espérance et matrice de dispersion. On applique alors l'item 2 de la proposition 3.24, page 48, pour préciser les paramètres de la loi normale de dimension k . \square

Exercice 5.2. (Corrigé de l'exercice : page 183)

En reprenant les hypothèses et notations de l'exercice 5.1,

1. Montrer que le vecteur gaussien X admet une densité sur \mathbb{R}^n qu'on explicitera.

2. Montrer que le vecteur gaussien X est centré de matrice de dispersion

$$D_X = [d_{i,j}]_{1 \leq i,j \leq n} \text{ où}$$

$$\begin{aligned} d_{1,1} &= \sigma^2 \\ d_{i,i} &= \sigma^2(\theta^2 + 1) && \text{pour } i = 2, \dots, n, \\ d_{j+1,j} &= d_{j,j+1} = \theta\sigma^2 && \text{pour } j = 1, \dots, n-1, \\ d_{i,j} &= 0 && \text{dans les autres cas.} \end{aligned}$$

La forme de la fonction caractéristique d'un vecteur gaussien permet d'établir un critère important d'indépendance des composantes d'un vecteur gaussien.

Proposition 5.6.

Soit $X := (X_1, X_2, \dots, X_d)$ un vecteur gaussien de dimension d . Alors la suite de variables aléatoires réelles (X_1, X_2, \dots, X_d) est indépendante si, et seulement si, la matrice de dispersion de X est diagonale.

Démonstration : • La condition nécessaire résulte de la proposition 4.13, page 75.

• Pour la condition suffisante, en vertu du critère d'indépendance 4.16, page 76, utilisant les fonctions caractéristiques, il suffit de montrer que si la matrice de dispersion est diagonale, alors pour tout $(u_1, \dots, u_d) \in \mathbb{R}^d$,

$$\phi_X(u_1, \dots, u_n) = \phi_{X_1}(u_1) \cdots \phi_{X_d}(u_d).$$

Or, en utilisant la proposition 5.3 pour obtenir la première des deux égalités suivantes,

$$\phi_X(u_1, \dots, u_n) = \exp \left(i \sum_{k=1}^d u_k \mathbb{E}(X_k) - \frac{1}{2} \sum_{k=1}^d u_k^2 \text{Var}(X_k) \right) = \phi_{X_1}(u_1) \cdots \phi_{X_d}(u_d),$$

car, pour tout entier k , la variable aléatoire réelle X_k est gaussienne et

$$\phi_{X_k}(u_k) = \exp \left(i u_k \mathbb{E}(X_k) - \frac{1}{2} u_k^2 \text{Var}(X_k) \right). \square$$

On peut donner un énoncé plus général de cette proposition. Pour cela introduisons une définition.

Définition 5.4.

Deux vecteurs aléatoires X et Y de dimensions quelconques seront dits **non-corrélés** si la matrice d'intercorrélation

$$I_{X,Y} := \mathbb{E}([X - \mathbb{E}(X)][Y - \mathbb{E}(Y)]^*)$$

est égale à la matrice-nulle.

On peut vérifier facilement, en explicitant les coefficients des matrices, que :

Proposition 5.7.

Si X et Y sont deux vecteurs aléatoires de dimensions respectives d et k , D_Z la matrice de dispersion du vecteur concaténé $Z := (X, Y)$ de dimension $d + k$, D_X et D_Y les matrices de dispersion de X et Y , alors :

1. La matrice d'intercorrélation $I_{X,Y}$ est une matrice rectangulaire à d lignes et k colonnes dont le coefficient général d'indice (i, j) , où $1 \leq i \leq d$ et $1 \leq j \leq k$, est $\text{Cov}(X_i, Y_j)$.
2. $I_{Y,X} = I_{X,Y}^*$, $D_X = I_{X,X}$ et D_Z est la matrice par blocs

$$D_Z = \begin{pmatrix} D_X & I_{X,Y} \\ I_{Y,X} & D_Y \end{pmatrix}.$$

D'où une version plus générale de la proposition 5.6 :

Proposition 5.8.

Soient $X := (X_1, X_2, \dots, X_d)$ un vecteur aléatoire de dimension d et $Y := (Y_1, Y_2, \dots, Y_k)$ un vecteur aléatoire de dimension k .

Supposons que le vecteur concaténé $Z := (X_1, X_2, \dots, X_d, Y_1, Y_2, \dots, Y_k)$ soit gaussien de dimension $d + k$. Alors le couple de vecteurs aléatoires (X, Y) est indépendant si, et seulement si, les vecteurs X et Y sont non-corrélés.

Démonstration : • La condition nécessaire résulte du fait que, si le couple de vecteurs aléatoires (X, Y) est indépendant, alors, pour tout (i, j) avec $1 \leq i \leq d$ et $1 \leq j \leq k$, le couple de variable aléatoire réelle (X_i, Y_j) est indépendant et par suite le coefficient général de la matrice d'intercorrélation $\text{Cov}(X_i, Y_j) = 0$.

• Réciproquement, si $I_{X,Y} = 0$, alors, d'après la proposition 5.7, $I_{Y,X} = 0$, et $D_Z = \begin{pmatrix} D_X & 0 \\ 0 & D_Y \end{pmatrix}$ où Z est le vecteur concaténé (X, Y) . On vérifie alors que, pour tout vecteur $w := (u_1, \dots, u_d, v_1, \dots, v_k) \in \mathbb{R}^{d+k}$, si on considère les vecteurs $u := (u_1, \dots, u_d)$, $v := (v_1, \dots, v_k)$, $m_X := \mathbb{E}(X)$ et $m_Y := \mathbb{E}(Y)$,

$$w^* D_Z w = u^* D_X u + v^* D_Y v \text{ et } w^* \mathbb{E}(Z) = u^* m_X + v^* m_Y.$$

La fonction caractéristique du vecteur Z est alors définie sur \mathbb{R}^{d+k} par

$$\begin{aligned} \Phi_Z(u_1, \dots, u_d, v_1, \dots, v_k) &= \exp(iu^* m_X + iv^* m_Y) \exp\left(-\frac{1}{2}u^* D_X u - \frac{1}{2}v^* D_Y v\right) \\ &= \Phi_X(u_1, \dots, u_d) \Phi_Y(v_1, \dots, v_k). \end{aligned}$$

On conclut en appliquant le critère d'indépendance utilisant les fonctions caractéristiques. \square

On notera la nécessité pour le vecteur $Z = (X_1, X_2, \dots, X_d, Y_1, Y_2, \dots, Y_k)$ d'être gaussien a priori. Cette hypothèse implique que les vecteurs X et Y sont gaussiens, mais on prendra garde qu'il ne suffit pas que les vecteurs X et Y soient gaussiens pour que le vecteur concaténé $(X_1, X_2, \dots, X_d, Y_1, Y_2, \dots, Y_k)$ soit gaussien.

L'exercice 4.6, page 74, fournit encore un contre-exemple prouvant que la proposition devient fausse si on ne suppose plus a priori que le vecteur $(X_1, X_2, \dots, X_d, Y_1, Y_2, \dots, Y_k)$ est gaussien. On vérifie facilement que la proposition 5.6 est bien un cas particulier de la proposition 5.8.

Exercice 5.3. (Corrigé de l'exercice : page 184)

On reprend les notations, hypothèses et résultats de l'exercice 3.15, page 59.

1. Existe-t-il un réel strictement positif a tel que (X, X_a) soit non-corrélé ?
2. Existe-t-il un réel strictement positif a tel que (X, X_a) soit un vecteur gaussien ?
3. Existe-t-il un réel strictement positif a tel que (X, X_a) soit un couple indépendant ?

Terminons par un résultat précisant la forme de la loi normale dans le cas où D est une matrice inversible, ce qui a pour conséquence que D est une matrice définie-positive. La loi de Gauss s'explique alors facilement :

Proposition 5.9.

Soient $m \in \mathbb{R}^d$ et D une matrice carrée d'ordre d à coefficients réels, symétrique et de type positif. Si D est inversible, alors la probabilité $\mathcal{N}_d(m, D)$ admet la densité ρ sur \mathbb{R}^d

$$\rho : x \in \mathbb{R}^d \mapsto \rho(x) := \frac{1}{\sqrt{(2\pi)^d \det(D)}} \exp \left(-\frac{1}{2} (x - m)^* D^{-1} (x - m) \right).$$

Démonstration : Soit X un vecteur gaussien de loi $\mathcal{N}_d(m, D)$ où $m \in \mathbb{R}^d$ et D une matrice carrée d'ordre d à coefficients réels, symétrique, positive et inversible. Comme D est symétrique réelle, il existe alors une matrice A d'ordre d orthogonale telle que $ADA^* = \Delta$ où Δ est une matrice diagonale. Les éléments diagonaux de Δ sont strictement positifs car D est positive et inversible. Notons $\sigma_1^2, \dots, \sigma_d^2$ les éléments diagonaux de Δ . Considérons le vecteur aléatoire $Z := A(X - m)$. D'après la proposition 5.5, Z est un vecteur gaussien de dimension d de loi $\mathcal{N}_d(0, \Delta)$. Soient (Z_1, \dots, Z_d) les composantes de Z . Comme la matrice de dispersion de Z , Δ , est diagonale, la suite de variable aléatoire réelle (Z_1, \dots, Z_d) est indépendante et, pour tout entier $1 \leq k \leq d$, la loi de Z_k est $\mathcal{N}_1(0, \sigma_k^2)$ d'après la proposition 4.5, page 66. La variable aléatoire réelle Z_k a donc pour densité l'application f_k définie sur \mathbb{R} par

$$f_k(t) := \frac{1}{\sqrt{2\pi\sigma_k^2}} \exp \left(-\frac{t^2}{2\sigma_k^2} \right).$$

Par suite, d'après la proposition 4.9, page 72, Z admet pour densité l'application définie sur \mathbb{R}^d par

$$f(t_1, \dots, t_d) = f_1(t_1) \cdots f_d(t_d) = \left(\frac{1}{\sqrt{2\pi}} \right)^d \frac{1}{\sqrt{\sigma_1^2 \cdots \sigma_d^2}} \exp \left(-\frac{1}{2} \sum_{k=1}^d \frac{t_k^2}{\sigma_k^2} \right).$$

Comme $\sigma_1^2 \cdots \sigma_d^2 = \det(D)$ et $\sum_{k=1}^d \frac{1}{\sigma_k^2} t_k^2 = t^* \Delta^{-1} t$, où $t := (t_1, \dots, t_d)$, on obtient

$$f(t) = f(t_1, \dots, t_d) = \left(\frac{1}{\sqrt{2\pi}} \right)^d \frac{1}{\sqrt{\det(D)}} \exp \left(-\frac{1}{2} t^* \Delta^{-1} t \right).$$

Montrons que la loi de X admet une densité. Soit h une application borélienne positive de \mathbb{R}^d dans $[0, +\infty]$. Comme $X = A^* Z + m$, en appliquant le théorème du transfert à Z ,

$$\mathbb{E}[h(X)] = \int_{\mathbb{R}^d} h(A^* z + m) d\mathbb{P}_Z(z) = \int_{\mathbb{R}^d} h(A^* z + m) f(z) d\lambda^{(d)}(z).$$

Effectuons le changement de variable dans \mathbb{R}^d , $z \in \mathbb{R}^d \mapsto x := A^*z + m$ dont le jacobien de la transformation $x \in \mathbb{R}^d \mapsto z = A(x - m)$ est $\det(A)$. Remarquons que, A étant orthogonale, $|\det(A)| = 1$ et appliquons le théorème de changement de variable. Il vient

$$\mathbb{E}[h(X)] = \int_{\mathbb{R}^d} h(x) f(A(x - m)) d\lambda^{(d)}(x).$$

Le vecteur X admet donc la densité définie sur \mathbb{R}^d par $\rho(x) := f(A(x - m))$. Par suite,

$$\begin{aligned} \rho(x) &= \left(\frac{1}{\sqrt{2\pi}} \right)^d \frac{1}{\sqrt{\det(D)}} \exp \left(-\frac{1}{2} (x - m)^* A^* \Delta^{-1} A (x - m) \right) \\ &= \left(\frac{1}{\sqrt{2\pi}} \right)^d \frac{1}{\sqrt{\det(D)}} \exp \left(-\frac{1}{2} (x - m)^* D^{-1} (x - m) \right), \end{aligned}$$

d'où le résultat cherché. \square

On admettra que si la matrice D n'est pas inversible, la loi de Gauss n'a pas de densité par rapport à la mesure de Lebesgue sur \mathbb{R}^d car on montre qu'elle est portée par un sous-espace affine de \mathbb{R}^d de dimension strictement inférieure à d .

Définition 5.5.

Si la matrice D est inversible, on dira que la loi de Gauss (ou le vecteur gaussien) $\mathcal{N}_d(m, D)$ est **non-dégénérée**. Dans le cas contraire on dira qu'elle est **dégénérée**.

Exemples 5.3.

1) Une variable aléatoire réelle gaussienne est non-dégénérée si, et seulement si, sa loi est une loi normale de variance non nulle. Donc une variable aléatoire réelle gaussienne dégénérée est une variable de loi δ_m où m est un réel quelconque.

2) Soit (X, Y) un couple de variables aléatoires réelles admettant pour densité l'application définie sur \mathbb{R}^2 par $\rho(x, y) := \frac{\sqrt{3}}{4\pi} \exp \left(-\frac{1}{2} (x^2 - xy + y^2) \right)$. On vérifie alors que

$$x^2 - xy + y^2 = \begin{pmatrix} x & y \end{pmatrix} \begin{pmatrix} 1 & -1/2 \\ -1/2 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} = \begin{pmatrix} x & y \end{pmatrix} D^{-1} \begin{pmatrix} x \\ y \end{pmatrix}$$

où $D := \begin{pmatrix} 4/3 & 2/3 \\ 2/3 & 4/3 \end{pmatrix}$ est la matrice de dispersion du vecteur (X, Y) . On en déduit que (X, Y) est un vecteur gaussien de loi $\mathcal{N}_2(0, D)$. On peut aussi affirmer que X et Y suivent la loi $\mathcal{N}_1(0, \frac{4}{3})$ et, puisque D n'est pas diagonale, que le couple de variables aléatoires réelles (X, Y) n'est pas indépendant. \square

Exercice 5.4. (Corrigé de l'exercice : page 185)

Soit X un vecteur aléatoire de dimension 3. On suppose que la loi de X est $\mathcal{N}_3(0, \Gamma)$ où

$$\Gamma := \begin{pmatrix} 3 & -1 & 0 \\ -1 & 3 & 0 \\ 0 & 0 & 2 \end{pmatrix}. \text{ Trouver une matrice } A \text{ carrée d'ordre 3 telle que les composantes}$$

du vecteur AX soient indépendantes et non dégénérées.

5.3 Exercices de révision sur les chapitres I à V

Exercice 5.5. (Corrigé de l'exercice : page 185)

Soient X un vecteur aléatoire de dimension 2, de loi $\mathcal{N}_2(0, I)$ où 0 désigne le vecteur nul de \mathbb{R}^2 et I la matrice-identité d'ordre 2. Soit A une matrice orthogonale d'ordre 2 à coefficients réels. Déterminer la loi du vecteur aléatoire de dimension 2 défini par $U := AX$.

Exercice 5.6. (Corrigé de l'exercice : page 185)

Soit (X, Y) un couple de variables aléatoires réelles de loi $\mathbb{P}_{(X,Y)} = \left(\alpha e^{-\frac{1}{2}(x^2 - xy + y^2)} \right) \cdot \lambda^{(2)}$ où $\lambda^{(2)}$ est la mesure de Lebesgue sur \mathbb{R}^2 . Déterminer la constante α et la matrice de dispersion du couple (X, Y) . Préciser les lois respectives des variables aléatoires réelles X et Y . Le couple de variables aléatoires réelles (X, Y) est-il indépendant ?

Exercice 5.7. (Corrigé de l'exercice : page 186)

Soit X une variable aléatoire réelle de loi $\mathcal{N}_1(0, 1)$. On pose

$$Y = X \mathbb{1}_{[0,\pi]}(|X|) - X \mathbb{1}_{]\pi,+\infty[}(|X|).$$

1. Vérifier que, pour toute application h de \mathbb{R} dans \mathbb{R} ,

$$h(Y) = h(X) \mathbb{1}_{[0,\pi]}(|X|) + h(-X) \mathbb{1}_{]\pi,+\infty[}(|X|).$$

2. Montrer que la variable aléatoire réelle Y suit la loi normale réduite centrée.
3. Le vecteur aléatoire (X, Y) est-il gaussien ?
4. Le couple de variable aléatoire réelle (X, Y) est-il indépendant ?

Exercice 5.8. (Corrigé de l'exercice : page 187)

Soient X, Y, Z trois variables aléatoires réelles indépendantes de même loi $\mathcal{N}_1(0, 1)$.

1. Déterminer les lois respectives des variables aléatoires réelles $U := X + Y + Z$, $V := 2X - Y - Z$ et $W := Y - Z$.
2. Montrer que les variables aléatoires réelles $X - Y$, $Y - Z$ et $Z - X$ sont chacune indépendantes de la v.a.r. U .
3. Le vecteur aléatoire de dimension 3, (U, V, W) , est-il gaussien ? Préciser sa loi.
4. Le triplet de variables aléatoires réelles (U, V, W) est-il indépendant ?
5. On note φ la fonction caractéristique de X^2 [on admettra que, pour tout réel x , $\varphi(x) = (1 - 2ix)^{-\frac{1}{2}}$]. On pose $T := (X - Y)^2 + (Y - Z)^2 + (Z - X)^2$. Vérifier, pour tout $(x, y, z) \in \mathbb{R}^3$, l'égalité

$$(x - y)^2 + (y - z)^2 + (z - x)^2 = 2 \left(x - \frac{1}{2}(y + z) \right)^2 + \frac{3}{2}(y - z)^2,$$

et exprimer la fonction caractéristique $\Phi_{(U,T)}$ du vecteur aléatoire (U, T) à l'aide de φ . En déduire l'expression de $\Phi_{(U,T)}(u, t)$ en fonction de u et t .

Exercice 5.9. (Corrigé de l'exercice : page 188)

Théorème de Fisher-Cochran

Soit $n \in \mathbb{N}^*$ et (X_1, \dots, X_n) une suite indépendante de v.a.r. toutes de même loi $\mathcal{N}_1(0, 1)$.

On définit respectivement les v.a.r. **moyenne empirique** et **variance empirique** par

$$\bar{X} := \frac{X_1 + \dots + X_n}{n} \quad \text{et} \quad S^2 := \frac{1}{n-1} \sum_{k=1}^n (X_k - \bar{X})^2.$$

1. Montrer que la v.a.r. X_1^2 suit la loi $\gamma(\frac{1}{2}, \frac{1}{2})$ aussi appelée loi du **Khi-deux à 1 degré de liberté** et notée $\chi^2(1)$.
2. En utilisant la fonction caractéristique des lois Gamma, en déduire que la loi de la v.a.r. $\sum_{k=1}^n X_k^2$ est $\gamma(\frac{1}{2}, \frac{n}{2})$ aussi appelée loi du **Khi-deux à n degrés de liberté** notée $\chi^2(n)$.
3. Montrer qu'il existe une matrice orthogonale C de la forme

$$C = \begin{pmatrix} c_{1,1} & c_{1,2} & \cdots & c_{1,n} \\ c_{2,1} & c_{2,2} & \cdots & c_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ c_{n-1,1} & c_{n-1,2} & \cdots & c_{n-1,n} \\ \frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}} & \cdots & \frac{1}{\sqrt{n}} \end{pmatrix}.$$

4. Déterminer la loi du vecteur aléatoire $Y := CX$.
5. Calculer Y_n et $\sum_{k=1}^n Y_k^2$ à l'aide de X_1, \dots, X_n . En déduire que $\bar{X} = \frac{1}{\sqrt{n}} Y_n$ et $S^2 = \frac{1}{n-1} \sum_{k=1}^{n-1} Y_k^2$.
6. Démontrer le **théorème de Fisher-Cochran** : Soit (X_1, \dots, X_n) une suite indépendante de v.a.r. de même loi $\mathcal{N}_1(0, 1)$. Alors (\bar{X}, S^2) est indépendant, \bar{X} suit la loi $\mathcal{N}_1(0, \frac{1}{n})$ et $(n-1)S^2$ suit la loi $\chi^2(n-1)$.

Exercice 5.10. (Corrigé de l'exercice : page 189)

Soit $(\varepsilon_i)_{i \geq 1}$ une suite indépendante de v.a.r. de même loi $\mathcal{N}_1(0, 1)$ et X_0 une v.a.r. indépendante de la suite $(\varepsilon_i)_{i \geq 1}$ et de loi $\mathbb{P}_{X_0} = \mathcal{N}_1(m, \sigma^2)$. On définit la suite de v.a.r. $(X_n)_{n \geq 1}$ de la façon suivante : $X_n := l_n(X_0, \dots, X_{n-1}) + b_n \varepsilon_n$ où $(b_n)_{n \geq 1}$ est une suite de réels et $(l_n)_{n \geq 1}$ une suite de formes linéaires sur \mathbb{R}^n . Montrer que, pour tout $n \geq 1$, il existe une forme linéaire L_n sur \mathbb{R}^{n+1} telle que $X_n = L_n(X_0, \varepsilon_1, \dots, \varepsilon_n)$ et en déduire que le vecteur (X_0, \dots, X_n) est gaussien.

Chapitre 6

Lois des grands nombres et convergences de v.a.r.

Dans ce chapitre, sauf indication contraire, toutes les variables aléatoires considérées seront réelles et définies sur un même espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$.

La plupart des résultats et définitions de ce chapitre sont une traduction en langage probabiliste de ceux vus en théorie de la mesure (cf. [2] chapitre V). Ils peuvent s'étendre aux v.a. à valeurs dans \mathbb{R}^d en prenant pour $|\cdot|$ la norme euclidienne de \mathbb{R}^d .

Les espaces $\mathcal{L}^p(\Omega, \mathcal{F}, \mathbb{P})$ et $L^p(\Omega, \mathcal{F}, \mathbb{P})$ ont été introduits en théorie de la mesure pour $p \in [1, +\infty]$. On rappelle que, pour tout $1 < p < q < +\infty$,

$$L^\infty \subseteq L^q \subseteq L^p \subseteq L^1 \text{ et } \|\cdot\|_{L^p} \leq \|\cdot\|_{L^q}.$$

De plus $\bigcap_{p \geq 1} L^p \neq L^\infty$ et $\bigcup_{p \geq 1} L^p \neq L^1$ (cf. [2] p. 67 ex IV-8 et IV-11).

On note $L^0(\Omega, \mathcal{F}, \mathbb{P})$ ou plus simplement L^0 , l'ensemble des classes d'équivalence, pour l'égalité \mathbb{P} -presque-sûre, des v.a. à valeurs dans $\overline{\mathbb{R}}$ sur $(\Omega, \mathcal{F}, \mathbb{P})$, définies et finies \mathbb{P} -presque-sûrement. On notera en cas de nécessité \widehat{X} la classe d'équivalence \mathbb{P} -presque-sûre de la v.a. X . On a, pour tout $p \in [1, +\infty]$, $L^p \subseteq L^0$. $(\widehat{X}_n)_{n \in \mathbb{N}}$ désignera une suite d'éléments de L^0 et \widehat{X} un élément de L^0 .

Dans ce chapitre, nous allons passer en revue ces différents modes de convergence initialement étudiés dans le cours de théorie de la mesure en les appliquant au cas particulier où la mesure est une probabilité.

6.1 Convergence en probabilité d'une suite de v.a.r.

6.1.1 Loi faible des grands nombres

Commençons par démontrer deux inégalités souvent utilisées en probabilité :

Proposition 6.1.**Inégalité de Markov**

Soient X une variable aléatoire réelle et φ une application borélienne de \mathbb{R} dans $[0, +\infty]$, alors pour tout réel $a > 0$,

$$\mathbb{P}[\varphi(X) \geq a] \leq \frac{\mathbb{E}[\varphi(X)]}{a}.$$

Démonstration : Posons $A := \{\varphi(X) \geq a\}$. Puisque φ est positive, on peut écrire :

$$\mathbb{E}[\varphi(X)] = \mathbb{E}[\varphi(X)\mathbf{1}_A] + \mathbb{E}[\varphi(X)\mathbf{1}_{A^c}] \geq \mathbb{E}[\varphi(X)\mathbf{1}_A] \geq a\mathbb{E}(\mathbf{1}_A) = a\mathbb{P}(A). \quad \square$$

Proposition 6.2.**Inégalité de Bienaymé-Tchébycheff**

Soit X une v.a.r. telle que $\mathbb{E}(X^2) < +\infty$. Alors pour tout réel $\alpha > 0$,

$$\mathbb{P}[|X - \mathbb{E}(X)| \geq \alpha] \leq \frac{1}{\alpha^2} \text{Var}(X).$$

Démonstration : On applique l'inégalité de Markov avec $\varphi(t) := [t - \mathbb{E}(X)]^2$ et $a := \alpha^2$. \square

Très utile en théorie, l'inégalité de Bienaymé-Tchébycheff peut être parfois trop grossière pour apporter des informations utiles dans la pratique. Dans ce cas on établit d'autres majorations plus appropriées à la situation étudiée. A titre d'exemple, on pourra étudier l'exercice [6.2](#).

Exemples 6.1.

1) Si X est une variable aléatoire réelle de loi binomiale $\mathcal{B}(10, \frac{1}{2})$ alors $\mathbb{E}(X) = 5$ et $\text{Var}(X) = \frac{5}{2}$. Un calcul direct donne l'estimation

$$\mathbb{P}(|X - 5| \geq 4) = (C_{10}^0 + C_{10}^1 + C_{10}^9 + C_{10}^{10}) \left(\frac{1}{2}\right)^{10} \approx 0,021,$$

alors que l'inégalité de Bienaymé-Tchébycheff donne la majoration

$$\mathbb{P}(|X - 5| \geq 4) \leq \frac{2,5}{4^2} \approx 0,156.$$

2) Si X est une variable aléatoire réelle de carré intégrable de variance σ^2 , l'inégalité de Bienaymé-Tchébycheff devient $\mathbb{P}(|X - m| \geq \sigma) \leq 1$, alors que dans le cas d'une variable aléatoire réelle gaussienne de loi $\mathcal{N}_1(m, \sigma^2)$

$$\begin{aligned} \mathbb{P}(|X - m| \geq \sigma) &= 2 \int_{\sigma}^{+\infty} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{x^2}{2\sigma^2}\right) dx \\ &= 2 \int_1^{+\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t^2}{2}\right) dt \approx 0,3174, \end{aligned}$$

et par suite $\mathbb{P}(m - \sigma < X < m + \sigma) \approx 0,6826$. \square

Exercice 6.1. (Corrigé de l'exercice : page 190)

Règle des trois "sigmas"

Soit X une v.a.r. telle que $\mathbb{E}(X^2) < +\infty$ de variance σ^2 et de moyenne m .

a) Montrer que

$$\mathbb{P}(m - 3\sigma < X < m + 3\sigma) \geq \frac{8}{9} \text{ et } \mathbb{P}(m - 2\sigma < X < m + 2\sigma) \geq \frac{3}{4}.$$

b) Comparer les valeurs obtenues dans la question précédente avec celles données par le calcul dans le cas où X est une variable aléatoire réelle de loi $\mathcal{N}_1(m, \sigma^2)$.

Exercice 6.2. (Corrigé de l'exercice : page 190)

1. On considère une variable aléatoire réelle X de carré intégrable, centrée et de variance σ^2 . A l'aide de l'inégalité de Cauchy-Schwarz de l'exercice 3.13, page 48, montrer que, pour tout réel $a > 0$,

$$a \leq \mathbb{E}[(a - X)\mathbb{1}_{]-\infty, a]}(X)] \leq \sqrt{\mathbb{P}(X \leq a)}\sqrt{\sigma^2 + a^2}.$$

En déduire que

$$\mathbb{P}(X > a) \leq \frac{\sigma^2}{\sigma^2 + a^2}.$$

2. Une usine fabrique chaque semaine un nombre aléatoire Y d'objets. On suppose $\mathbb{E}[Y] = 100$ et $\text{Var}(Y) = 400$. Trouver à l'aide de la question précédente un majorant de la probabilité que la production hebdomadaire dépasse 120. Comparer ce résultat avec celui obtenu par application de l'inégalité de Bienaymé-Tchébycheff.

Définition 6.1.

Une suite $(X_n)_{\mathbb{N}}$ de variables aléatoires réelles est dite **identiquement-distribuée**, en abrégé **i.d.**, si toutes les variables aléatoires réelles de la suite ont la même loi. Une suite **i.i.d.** de variables aléatoires réelles est une suite indépendante et identiquement distribuée de v.a.r..

Définition 6.2.

Si $(X_n)_{\mathbb{N}}$ est une suite de variables aléatoires réelles pour tout entier $n \geq 1$, on appelle **moyenne empirique d'ordre n** associée à la suite $(X_n)_{\mathbb{N}}$, et on note $\bar{X}_{(n)}$ ou plus simplement \bar{X} , la v.a.r. définie par

$$\bar{X}_{(n)} := \frac{X_1 + \cdots + X_n}{n}.$$

La démonstration de la proposition suivante est immédiate et laissée en exercice.

Proposition 6.3.

Si $(X_n)_{\mathbb{N}}$ est une suite indépendante et identiquement distribuée de variables aléatoires réelles de carré intégrable, d'espérance m et de variance σ^2 , alors

$$\mathbb{E}\left(\frac{X_1 + \cdots + X_n}{n}\right) = m \text{ et } \text{Var}\left(\frac{X_1 + \cdots + X_n}{n}\right) = \frac{\sigma^2}{n}.$$

Une application de l'inégalité de Bienaymé-Tchébycheff dont le résultat est historiquement célèbre est :

Proposition 6.4.

Théorème de Bernoulli

Soit $(X_n)_{\mathbb{N}}$ une suite indépendante et identiquement distribuée de variables aléatoires réelles de Bernoulli de paramètre $p \in]0, 1[$ i.e. de loi $\mathcal{B}(1, p)$. Alors, pour tout réel $a > 0$ et tout entier $n \geq 1$,

$$\mathbb{P} \left(\left| \frac{X_1 + \dots + X_n}{n} - p \right| \geq a \right) \leq \frac{1}{4a^2n}.$$

Démonstration : On applique l'inégalité de Bienaymé-Tchébycheff à la variable aléatoire réelle $\bar{X}_{(n)}$ et on utilise la proposition précédente pour obtenir la majoration

$$\mathbb{P} \left(\left| \frac{X_1 + \dots + X_n}{n} - m \right| \geq a \right) \leq \frac{\text{Var}(X_0)}{na^2}.$$

Comme les variable aléatoire réelle sont de Bernoulli $\mathbb{E}(X_0) = m = p$ et $\text{Var}(X_0) = p(1 - p)$. On conclut en vérifiant que, pour tout $p \in]0, 1[$, $p(1 - p) \leq \frac{1}{4}$. \square

Exemples 6.2.

Dans son *Essai d'arithmétique morale* paru en 1777, Buffon relate l'expérience qui consiste à lancer 4040 fois une pièce de monnaie. Dans la réalisation ω_0 de cette expérience Buffon obtient 2049 fois "Pile". Si on note X_k l'application qui à chaque réalisation associe le nombre 1 si la pièce tombe sur "Pile" lors du $k^{\text{ième}}$ lancer et 0 sinon, la variable aléatoire réelle X_k est une variable aléatoire réelle de Bernoulli de paramètre p inconnu. On modélise cette situation en représentant les lancers successifs indépendants par une suite i.i.d. de variable aléatoire réelle de Bernoulli. On est donc dans les conditions du théorème de Bernoulli. Avec les notations introduites précédemment, pour l'observation ω_0 de Buffon, on peut écrire

$$\bar{X}_{(4040)}(\omega_0) = \frac{2049}{4040} \approx 0,507.$$

Choisissons a tel que $\frac{1}{4a^2 \cdot 4040} \approx 0,05$. Par exemple prenons $a = 0,0352$. D'après le théorème de Bernoulli,

$$\mathbb{P} (\omega \in \Omega / |\bar{X}_{(4040)}(\omega) - p| \geq 0,0352) \leq 0,05.$$

La probabilité que la réalisation ω_0 de cette expérience satisfasse l'événement

$$\{\omega \in \Omega / |\bar{X}_{(4040)}(\omega) - p| \geq 0,0352\}$$

est inférieure à 0,05. Autrement dit, la probabilité que le paramètre p vérifie la condition $|\bar{X}_{(4040)}(\omega_0) - p| \geq 0,0352$, c-à-d approximativement $|0,507 - p| \geq 0,0352$, est inférieure à 0,05. Par suite, on peut affirmer, avec une probabilité supérieure à 0,95, que l'encadrement de p obtenu lors de l'observation ω_0 , i.e. $0,4718 \leq p \leq 0,5422$, est correct.

L'intervalle $[0,4718; 0,5422]$ est dit **intervalle de confiance pour p de niveau de confiance 0,95**. \square

La technique de détermination des intervalles de confiance pour des paramètres à estimer relève du domaine de la statistique inférentielle et sera développée dans l'unité d'enseignement de Statistique.

Exercice 6.3. (Corrigé de l'exercice : page 191)

On cherche à mesurer une grandeur physique en faisant N mesures indépendantes. Toutes ces mesures sont entachées d'erreur et donnent donc des résultats aléatoires dont l'espérance commune m est la "vraie" valeur de la grandeur à mesurer. Ces mesures sont supposées identiquement distribuées. Sachant que la variance de chacune d'elles est 0,25, on veut obtenir un résultat qui soit éloigné de moins de 0,05 de la "vraie" valeur m avec une probabilité de 0,99. Quelle valeur choisir pour N ?

En fait la majoration du théorème de Bernoulli peut être améliorée pour montrer que la convergence vers 0 du second membre est de type exponentiel. C'est ce résultat, donné à titre d'information, qu'énonce le cas particulier suivant du théorème des grandes déviations (admis et hors programme, pour la démonstration voir [3] exercice IV-13) :

Proposition 6.5.

Théorème des grandes déviations pour les v.a.r. de Bernoulli (Hors programme)

Si $(X_n)_{\mathbb{N}}$ est une suite indépendante de v.a.r. de Bernoulli de même paramètre $p \in]0, 1[$, alors pour tout $\varepsilon > 0$, il existe une constante $C_\varepsilon > 0$ telle que pour tout $n \in \mathbb{N}^*$,

$$\mathbb{P} \left(\left| \frac{X_1 + X_2 + \dots + X_n}{n} - p \right| \geq \varepsilon \right) \leq 2 \exp(-nC_\varepsilon).$$

Énonçons une conséquence plus faible de l'inégalité de Bienaymé-Tchébycheff :

Proposition 6.6.

Loi faible des grands nombres (1^{er} énoncé)

Si $(X_n)_{\mathbb{N}}$ est une suite indépendante et identiquement distribuée de variables aléatoires réelles de carré intégrable et d'espérance m , alors, pour tout réel $a > 0$,

$$\lim_{n \rightarrow +\infty} \mathbb{P} \left(\left| \frac{X_1 + \dots + X_n}{n} - m \right| \geq a \right) = 0.$$

Démonstration : On applique l'inégalité de Bienaymé-Tchébycheff comme dans la démonstration du théorème de Bernoulli pour obtenir

$$\mathbb{P} \left(\left| \frac{X_1 + \dots + X_n}{n} - m \right| \geq a \right) \leq \frac{\text{Var}(X_0)}{na^2}.$$

Ce qui donne le résultat par passage à la limite. \square

6.1.2 Convergence en probabilité

L'énoncé de la loi faible des grands nombres suggère alors la définition suivante :

Définition 6.3.

On dit qu'une suite de variables aléatoires réelles $(X_n)_\mathbb{N}$ **converge en probabilité** vers une variable aléatoire réelle Y si, pour tout réel $a > 0$,

$$\lim_{n \rightarrow +\infty} \mathbb{P}(|X_n - Y| \geq a) = 0.$$

On retrouve ici la traduction de la "convergence en mesure" dans le cas où la mesure est une probabilité.

Avec cette nouvelle définition, la loi faible s'énonce :

Proposition 6.7.**Loi faible des grands nombres (2^{ième} énoncé)**

Si $(X_n)_\mathbb{N}$ est une suite indépendante et identiquement distribuée de variables aléatoires réelles de carré intégrable et d'espérance m , alors la suite des moyennes empiriques $\left(\frac{X_1 + X_2 + \dots + X_n}{n}\right)_{\mathbb{N}^*}$ converge en probabilité vers la variable aléatoire réelle constante m .

Exemples 6.3.

Si, pour tout entier $n \geq 1$, X_n est une variable aléatoire réelle de loi $\frac{1}{n+1}\delta_n + \frac{n}{n+1}\delta_{\frac{1}{n}}$, alors la suite $(X_n)_\mathbb{N}$ converge en probabilité vers la variable aléatoire réelle constante 0. En effet, soit $a > 0$ et n un entier tel que $n > \frac{1}{a} > \frac{1}{n}$. Comme la variable aléatoire réelle X_n est discrète portée par l'ensemble $\{n, \frac{1}{n}\}$,

$$\mathbb{P}(|X_n| \geq a) = \mathbb{P}(X_n = n) = \frac{1}{n+1}.$$

Ce qui prouve le résultat en faisant tendre n vers $+\infty$. \square

Exercice 6.4. (Corrigé de l'exercice : page 191)

Montrer que si, pour tout entier $n \geq 1$, X_n est une variable aléatoire réelle de densité $n\mathbb{1}_{[0, \frac{1}{n}]}$ alors la suite $(X_n)_\mathbb{N}$ converge en probabilité vers la variable aléatoire réelle constante 0.

La limite d'une suite de variables aléatoires réelles convergeant en probabilité est "presque-sûrement" unique comme l'énonce de façon précise le résultat suivant :

Proposition 6.8.

Si $(X_n)_\mathbb{N}$ est une suite de variables aléatoires réelles convergeant en probabilité vers les variables aléatoires réelles X et Y , alors les variables aléatoires réelles X et Y sont égales presque-sûrement, i.e. $\mathbb{P}(X \neq Y) = 0$.

Démonstration : Avec les notations du théorème on peut écrire, pour tout entier naturel n , $|X - Y| \leq |X - X_n| + |X_n - Y|$. Soit $a > 0$ un réel, on vérifie l'inclusion entre événements

$$\{|X - Y| \geq a\} \subseteq \left\{|X - X_n| \geq \frac{a}{2}\right\} \cup \left\{|X_n - Y| \geq \frac{a}{2}\right\}.$$

Par suite, pour tout $a > 0$,

$$\mathbb{P}(|X - Y| \geq a) \leq \mathbb{P}\left(|X - X_n| \geq \frac{a}{2}\right) + \mathbb{P}\left(|X_n - Y| \geq \frac{a}{2}\right)$$

et en passant à la limite dans le second membre de l'inégalité précédente, on trouve par définition de la convergence en probabilité, que, pour tout réel $a > 0$, $\mathbb{P}(|X - Y| \geq a) = 0$. Or

$$\{X \neq Y\} = \{|X - Y| > 0\} = \bigcup_{n \geq 1} \left\{|X - Y| \geq \frac{1}{n}\right\}.$$

D'où

$$\mathbb{P}(X \neq Y) \leq \sum_{n \geq 1} \mathbb{P}\left(|X - Y| \geq \frac{1}{n}\right) = 0. \square$$

La proposition précédente entraîne que la convergence en probabilité induit une convergence dans l'espace L^0 définie de la façon suivante :

Définition 6.4.

On dit que $(\widehat{X}_n)_\mathbb{N}$ converge en probabilité vers \widehat{X} si, pour tout $\delta > 0$, $\lim_{n \rightarrow +\infty} \mathbb{P}(|X_n - X| \geq \delta) = 0$.

On note alors $(\widehat{X}_n)_\mathbb{N} \xrightarrow{\mathbb{P}} \widehat{X}$ ou $\widehat{X} = \mathbb{P}\text{-}\lim_n(\widehat{X}_n)$.

On montre (cf. [2] p. 86 ex V-9), et nous l'admettrons, le résultat suivant :

Proposition 6.9.

(Hors programme)

L'application

$$\Delta : (\widehat{X}, \widehat{Y}) \in L^0 \times L^0 \mapsto \Delta(\widehat{X}, \widehat{Y}) := \inf\{\delta > 0, \mathbb{P}(|X - Y| \geq \delta) \leq \delta\}$$

définit une métrique sur L^0 appelée **métrique de Ky-Fan**. De plus (L^0, Δ) est complet et $(\widehat{X}_n)_\mathbb{N}$ converge en probabilité vers \widehat{X} si, et seulement si, $\lim_n \Delta(\widehat{X}_n, \widehat{X}) = 0$.

La proposition suivante permet d'effectuer des calculs sur les limites en probabilité.

Proposition 6.10.

Soient f une application continue de \mathbb{R}^2 dans \mathbb{R} , $(X_n)_\mathbb{N}$ et $(Y_n)_\mathbb{N}$ deux suites de variables aléatoires réelles convergeant en probabilité respectivement vers les variables aléatoires réelles X et Y , alors la suite de variables aléatoires réelles $(f(X_n, Y_n))_\mathbb{N}$ converge en probabilité vers la variable aléatoire réelle $f(X, Y)$.

Démonstration : Nous allons démontrer cette proposition dans le cadre plus restrictif des applications f uniformément continues sur \mathbb{R}^2 . On admettra le résultat pour les fonctions seulement continues.

Comme f est uniformément continue sur \mathbb{R}^2 , pour tout réel $\varepsilon > 0$, il existe un réel $\eta > 0$ (dépendant de ε), tel que, pour tout $(x, y) \in \mathbb{R}^2$ et $(x', y') \in \mathbb{R}^2$, $|x - x'| + |y - y'| < \eta$ implique $|f(x, y) - f(x', y')| < \varepsilon$.

Soit $\varepsilon > 0$ fixé. Pour tout entier naturel n , on a alors

$$\begin{aligned} \{|f(X_n, Y_n) - f(X, Y)| \geq \varepsilon\} &\subseteq \{|X_n - X| + |Y_n - Y| \geq \eta\} \\ &\subseteq \left\{|X_n - X| \geq \frac{\eta}{2}\right\} \cup \left\{|Y_n - Y| \geq \frac{\eta}{2}\right\}. \end{aligned}$$

D'où, en faisant tendre n vers $+\infty$, $\lim_{n \rightarrow +\infty} \mathbb{P}(|f(X_n, Y_n) - f(X, Y)| \geq \varepsilon) = 0$, ce qui prouve que la suite de variables aléatoires réelles $(f(X_n, Y_n))_{\mathbb{N}}$ converge en probabilité vers la variable aléatoire réelle $f(X, Y)$. \square

On admettra également que cette proposition devient fausse si l'application f n'est plus supposée continue.

Un corollaire immédiat, et très utilisé en pratique, de la proposition précédente est :

Proposition 6.11.

Si $(X_n)_{\mathbb{N}}$ et $(Y_n)_{\mathbb{N}}$ sont deux suites de variables aléatoires réelles convergeant en probabilité respectivement vers les variables aléatoires réelles X et Y , alors les suites de variables aléatoires réelles $(X_n + Y_n)_{\mathbb{N}}$ et $(X_n Y_n)_{\mathbb{N}}$ convergent en probabilité respectivement vers les v.a.r. $X + Y$ et XY .

Exercice 6.5. (Corrigé de l'exercice : page 191)

Trouver, parmi les exercices ou exemples déjà proposés dans ce cours, un contre-exemple prouvant que si la suite de variables aléatoires réelles $(X_n)_{\mathbb{N}}$ converge en probabilité vers la variable aléatoire réelle X , cela n'entraîne pas nécessairement que la suite des espérances (si elles existent) $(\mathbb{E}(X_n))_{\mathbb{N}}$ converge vers le réel $\mathbb{E}(X)$ (s'il existe).

6.2 Convergence presque-sûre d'une suite de v.a.r.

6.2.1 Loi forte des grands nombres

La proposition suivante, difficile à démontrer, a une grande importance de par sa signification probabiliste. Nous en donnerons une démonstration plus loin (cf. proposition 6.19, page 117) dans le cas des variables aléatoires de carré intégrable. C'est en partie grâce à ce théorème que le formalisme des probabilités trouve sa cohérence.

Proposition 6.12.

Loi forte des grands nombres de Kolmogorov (1^{er} énoncé) (admis)

Si $(X_n)_{\mathbb{N}}$ est une suite indépendante et identiquement distribuée de variables aléatoires réelles intégrables et d'espérance m , alors l'ensemble des $\omega \in \Omega$ tels que la suite $\left(\frac{X_1(\omega) + X_2(\omega) + \dots + X_n(\omega)}{n} \right)_{\mathbb{N}^}$ ne converge pas vers m est un événement de probabilité nulle.*

Donnons une interprétation intuitive de la loi forte des grands nombres. Considérons un événement A de probabilité inconnue p relatif à une situation aléatoire. Répétons un grand nombre de fois l'expérience relative à cette situation aléatoire et notons X_k l'application qui à chaque essai associe le nombre 1 si l'événement A est réalisé lors de la $k^{\text{ième}}$ expérience et 0 sinon. La variable aléatoire réelle X_k est une variable aléatoire réelle de Bernoulli de paramètre p inconnu. On modélise cette situation en considérant les essais successifs indépendants par une suite i.i.d. de variables aléatoires réelles de Bernoulli. On est donc dans les conditions d'application de la loi forte de Kolmogorov. Cette loi signifie que presque-sûrement, pour n assez grand, la moyenne empirique des variables aléatoires réelles X_k , c'est-à-dire la fréquence de réalisation de l'événement A , est une bonne approximation de la probabilité de A .

6.2.2 Convergence presque-sûre

On peut énoncer la loi forte sous sa forme classique en introduisant un autre mode de convergence pour les suites de variables aléatoires.

Auparavant remarquons que :

Proposition 6.13.

Si $(X_n)_\mathbb{N}$ est une suite de variables aléatoires réelles et Y une v.a.r., alors l'ensemble Δ_Y des $\omega \in \Omega$ tels que la suite réelle $(X_n(\omega))_\mathbb{N}$ ne converge pas vers $Y(\omega)$ est un événement de l'espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$, i.e. $\Delta_Y \in \mathcal{F}$.

Démonstration : Vérifions au préalable que

$$\Delta_Y = \bigcup_{n=1}^{+\infty} \bigcap_{k=0}^{+\infty} \bigcup_{l=k}^{+\infty} \left\{ |X_l - Y| > \frac{1}{n} \right\}.$$

En effet en prenant la contraposée de la définition de la convergence de la suite réelle $(X_n(\omega))_\mathbb{N}$ vers $Y(\omega)$: $\omega \in \Delta_Y$ si, et seulement si, il existe un entier $n \geq 1$ tel que, pour tout entier naturel k , il existe un entier $l \geq k$ avec $|X_l(\omega) - Y(\omega)| > \frac{1}{n}$.

On conclut alors en utilisant la mesurabilité des variables aléatoires réelles X_i et Y ainsi que la stabilité des tribus pour les opérations de réunion et d'intersection dénombrables (cf. Chapitre I). \square

On peut alors donner la définition suivante :

Définition 6.5.

On dit qu'une suite de variables aléatoires réelles $(X_n)_\mathbb{N}$ converge presque-sûrement vers une variable aléatoire réelle Y si $\mathbb{P}(\Delta_Y) = 0$.

Cette définition est la traduction en langage probabiliste de la définition de la "convergence presque-partout" définie en théorie de la mesure.

Exemples 6.4.

Soient $\Omega := \mathbb{R}$, $\mathcal{F} := \mathcal{B}(\mathbb{R})$ et \mathbb{P} la probabilité sur \mathbb{R} de densité $\mathbb{1}_{[0,1]}$. Pour tout entier $n \geq 1$, considérons la variable aléatoire réelle $X_n := \pi + n\mathbb{1}_{[0, \frac{1}{n}]}$. Alors la suite $(X_n)_\mathbb{N}$ converge presque-sûrement vers la variable aléatoire réelle constante $Y := \pi$.

En effet, on vérifie aisément que $\Delta_Y = \{0\}$ car $\lim_n X_n(0) = +\infty$ et, pour tout $\omega \in \Omega \setminus \{0\}$, $\lim_n X_n(\omega) = \pi$. De plus $\mathbb{P}(\{0\}) = 0$, d'où le résultat. \square

Donnons quelques propriétés de la convergence presque-sûre. La limite d'une suite de variables aléatoires réelles convergeant presque-sûrement est "unique" dans le sens suivant :

Proposition 6.14.

Si $(X_n)_\mathbb{N}$ est une suite de variable aléatoire réelle convergeant presque-sûrement vers les variables aléatoires réelles X et Y , alors les variables aléatoires réelles X et Y sont égales presque-sûrement, i.e. $\mathbb{P}(X \neq Y) = 0$.

Démonstration : Avec les notations introduites ci-dessus on peut écrire

$$\{X \neq Y\} \subseteq \Delta_X \cup \Delta_Y,$$

et par suite $\mathbb{P}(X \neq Y) \leq \mathbb{P}(\Delta_X) + \mathbb{P}(\Delta_Y) = 0$. \square

La proposition précédente entraîne que la convergence presque-sûre induit une convergence dans l'espace L^0 définie de la façon suivante :

Définition 6.6.

On dit que $(\widehat{X}_n)_\mathbb{N}$ converge \mathbb{P} -presque-sûrement (ou plus simplement presque-sûrement, ou en abrégé p.s.,) vers \widehat{X} , si l'ensemble des $\omega \in \Omega$ tels que $(X_n(\omega))_\mathbb{N}$ ne converge pas dans $\overline{\mathbb{R}}$ vers $X(\omega)$ est \mathbb{P} -négligeable.

On note alors $(\widehat{X}_n)_\mathbb{N} \xrightarrow{\mathbb{P}\text{-p.s.}} \widehat{X}$ ou $\widehat{X} = p.s.\text{-}\lim_n(\widehat{X}_n)$.

En général il n'existe pas de métrique d sur L^0 telle que, pour toute suite $(\widehat{X}_n)_\mathbb{N}$ d'éléments de L^0 et $\widehat{X} \in L^0$, on ait : $\lim_{n \rightarrow +\infty} d(\widehat{X}_n, \widehat{X}) = 0$ si, et seulement si, $(\widehat{X}_n)_\mathbb{N}$ converge \mathbb{P} -presque-sûrement vers \widehat{X} . On peut montrer qu'il existe une topologie associée à la convergence presque-sûre si, et seulement si, la probabilité \mathbb{P} est discrète (cf. [2] p. 86 ex V-10).

La proposition suivante permet d'effectuer des calculs sur les limites presque-sûres.

Proposition 6.15.

Soit f une application continue de \mathbb{R}^2 dans \mathbb{R} , $(X_n)_\mathbb{N}$ et $(Y_n)_\mathbb{N}$ deux suites de variables aléatoires réelles convergeant presque-sûrement respectivement vers les variables aléatoires réelles X et Y , alors la suite de variables aléatoires réelles $(f(X_n, Y_n))_\mathbb{N}$ converge presque-sûrement vers la variable aléatoire réelle $f(X, Y)$.

Cette proposition devient fausse si l'application f n'est plus supposée continue (cf. [14] ex 14-11).

Un corollaire immédiat de la proposition précédente est :

Proposition 6.16.

Si $(X_n)_\mathbb{N}$ et $(Y_n)_\mathbb{N}$ sont deux suites de variables aléatoires réelles convergeant presque-sûrement respectivement vers les variables aléatoires réelles X et Y , alors les suites de variables aléatoires réelles $(X_n + Y_n)_\mathbb{N}$ et $(X_n Y_n)_\mathbb{N}$ convergent presque-sûrement respectivement vers les variables aléatoires réelles $X + Y$ et XY .

Exercice 6.6. (Corrigé de l'exercice : page 192)

Démontrer les propositions 6.15 et 6.16.

L'inégalité de Tchébycheff est utile dans l'étude des questions de convergence en probabilité. Le lemme de Borel-Cantelli, quant à lui, est souvent utilisé dans les questions de convergence p.s. grâce à la proposition suivante :

Proposition 6.17.

1. La suite $(X_n)_{\mathbb{N}}$ converge p.s. vers la v.a.r. X si, et seulement si, pour tout $\varepsilon > 0$, $\mathbb{P}(\limsup_n \{|X_n - X| \geq \varepsilon\}) = 0$.
2. Si pour tout $\varepsilon > 0$, la série $\sum_n \mathbb{P}(\{|X_n - X| \geq \varepsilon\})$ converge dans \mathbb{R} , alors la suite $(X_n)_{\mathbb{N}}$ converge p.s. vers la v.a.r. X .
3. Si la suite d'événements $(\{|X_n - X| \geq \varepsilon\})_{\mathbb{N}}$ est indépendante, alors la suite $(X_n)_{\mathbb{N}}$ converge p.s. vers la v.a.r. X si, et seulement si, pour tout $\varepsilon > 0$, la série $\sum_n \mathbb{P}(\{|X_n - X| \geq \varepsilon\})$ converge dans \mathbb{R} .

Démonstration de 6.17 :

1. • $(X_n(\omega))_{\mathbb{N}}$ converge vers $X(\omega)$ si, et seulement si, pour tout $\varepsilon > 0$, il existe un entier naturel N tel que, pour tout $n \in \mathbb{N}$, $n > N$ implique $|X_n(\omega) - X(\omega)| \leq \varepsilon$. En prenant la négation, $(X_n(\omega))_{\mathbb{N}}$ ne converge pas vers $X(\omega)$ si, et seulement si, il existe $\varepsilon > 0$, tel que, pour tout entier naturel N , il existe $n \in \mathbb{N}$, vérifiant $n > N$ et $|X_n(\omega) - X(\omega)| > \varepsilon$. La dernière partie de la phrase peut aussi s'énoncer : il existe $\varepsilon > 0$, tel que ω appartienne à une infinité d'événements de la suite $(\{|X_n - X| > \varepsilon\})_{\mathbb{N}}$; ou encore par définition de la limite-supérieure d'une suite d'événements : il existe $\varepsilon > 0$, tel que $\omega \in \limsup_n \{|X_n - X| > \varepsilon\}$. Par suite,

$$\{\omega \in \Omega / (X_n(\omega))_{\mathbb{N}} \text{ ne converge pas vers } X(\omega)\} = \bigcup_{\varepsilon \in]0, \infty[} \limsup_n \{|X_n - X| > \varepsilon\}.$$

En remarquant que $(X_n(\omega))_{\mathbb{N}}$ converge vers $X(\omega)$ si, et seulement si, pour tout $p \in \mathbb{N}^*$, il existe un entier naturel N tel que, pour tout $n \in \mathbb{N}$, $n > N$ implique $|X_n(\omega) - X(\omega)| \leq \frac{1}{p}$, et en raisonnant comme précédemment on obtient l'égalité

$$\{\omega \in \Omega / (X_n(\omega))_{\mathbb{N}} \text{ ne converge pas vers } X(\omega)\} = \bigcup_{p \in \mathbb{N}^*} \limsup_n \{|X_n - X| > \frac{1}{p}\}.$$

- Si la suite $(X_n)_{\mathbb{N}}$ converge presque-sûrement vers X , alors

$$\mathbb{P}(\{(X_n)_{\mathbb{N}} \text{ ne converge pas vers } X\}) = 0.$$

Comme, pour tout $\varepsilon > 0$,

$$\limsup_n \{|X_n - X| > \varepsilon\} \subseteq \{\omega \in \Omega / (X_n(\omega))_{\mathbb{N}} \text{ ne converge pas vers } X(\omega)\}$$

on en déduit que, pour tout $\varepsilon > 0$, $\mathbb{P}(\limsup_n \{|X_n - X| > \varepsilon\}) = 0$.

Réciproquement, supposons que, pour tout $\varepsilon > 0$, $\mathbb{P}(\limsup_n \{|X_n - X| > \varepsilon\}) = 0$, alors

$$\mathbb{P}\left(\bigcup_{p \in \mathbb{N}^*} \limsup_n \{|X_n - X| > \frac{1}{p}\}\right) \leq \sum_{p \in \mathbb{N}^*} \mathbb{P}\left(\limsup_n \{|X_n - X| > \frac{1}{p}\}\right) = 0.$$

D'après la première partie on obtient

$$\mathbb{P}(\{(X_n)_{\mathbb{N}} \text{ ne converge pas vers } X\}) = 0$$

c'est-à-dire que $(X_n)_{\mathbb{N}}$ converge presque-sûrement vers X .

2. Compte tenu de l'item 1) qu'on vient de démontrer, le résultat des item 2) et 3) de la proposition est une conséquence immédiate du lemme de Borel-Cantelli. \square

On remarquera que dans le troisième item l'indépendance est une hypothèse essentielle ; on trouvera un contre-exemple dans ([14] ex 14-4). Le premier item souligne le lien existant entre les convergences p.s. et en probabilité en s'énonçant :

Proposition 6.18.

La suite $(X_n)_{\mathbb{N}}$ converge presque-sûrement vers X si, et seulement si, la suite $(M_n)_{\mathbb{N}}$, définie pour tout $n \in \mathbb{N}$ par $M_n := \sup_{k \geq n} |X_k - X|$, converge en probabilité vers 0.

Démonstration : Soit $\varepsilon > 0$, $\limsup \{|X_n - X| \geq \varepsilon\} = \bigcap_{n=0}^{\infty} \left(\bigcup_{k=n}^{\infty} \{|X_k - X| \geq \varepsilon\} \right)$. En utilisant le théorème de convergence monotone des probabilités, on obtient

$$\mathbb{P}(\limsup \{|X_n - X| \geq \varepsilon\}) = \lim_{n \rightarrow +\infty} \mathbb{P} \left(\bigcup_{k=n}^{\infty} \{|X_k - X| \geq \varepsilon\} \right).$$

De plus, $\bigcup_{k=n}^{\infty} \{|X_k - X| \geq \varepsilon\} = \{\sup_{k \geq n} |X_k - X| \geq \varepsilon\}$. On conclut alors en vertu du premier item de la proposition 6.17 : $(X_n)_{\mathbb{N}}$ converge p.s. vers la v.a.r. X si, et seulement si, pour tout $\varepsilon > 0$, $\mathbb{P}(\limsup_n \{|X_n - X| \geq \varepsilon\}) = 0$, ou encore $(X_n)_{\mathbb{N}}$ converge p.s. vers la v.a.r. X si, et seulement si, pour tout $\varepsilon > 0$, $\lim_{n \rightarrow +\infty} \mathbb{P} \left(\sup_{k \geq n} |X_k - X| \geq \varepsilon \right) = 0$, ou encore $(X_n)_{\mathbb{N}}$ converge p.s. vers la v.a.r. X si, et seulement si, pour tout $\varepsilon > 0$, $\lim_{n \rightarrow +\infty} \mathbb{P}(M_n \geq \varepsilon) = 0$, où on a posé $M_n = \sup_{k \geq n} |X_k - X|$. Ce qui traduit la convergence en probabilité vers 0 de la suite $(M_n)_{\mathbb{N}}$. \square

Exercice 6.7. (Corrigé de l'exercice : page 192)

Soit $(X_n)_{\mathbb{N}}$ une suite i.i.d. de variables aléatoires réelles de carré intégrable, de variance σ^2 . On note, pour tout entier $n \geq 2$,

$$S_n^2 := \frac{1}{n-1} \sum_{k=1}^n (X_k - \bar{X}_{(n)})^2,$$

la **variance empirique d'ordre n** de la suite $(X_n)_{\mathbb{N}}$. Montrer que

$$S_n^2 = \frac{1}{n-1} \sum_{k=1}^n X_k^2 - \frac{n}{n-1} (\bar{X}_{(n)})^2 \quad (\text{Formule de Steiner})$$

et en déduire que la suite de variables aléatoires réelles $(S_n^2)_{\mathbb{N}}$ converge presque-sûrement vers σ^2 .

Avec la notion de convergence presque-sûre, la loi forte de Kolmogorov s'énonce :

Proposition 6.19.**Loi forte des grands nombres de Kolmogorov (2^{ième} énoncé)**

Si $(X_n)_{\mathbb{N}}$ est une suite indépendante et identiquement distribuée de variables aléatoires réelles intégrables et d'espérance m , alors la suite des moyennes empiriques $\left(\frac{X_1 + X_2 + \dots + X_n}{n}\right)_{n \in \mathbb{N}^*}$ converge presque-sûrement vers la variable aléatoire réelle constante m .

Démonstration : Conformément au programme de l'unité, on ne fera la démonstration de ce théorème que dans le cadre plus restreint des variables de carré intégrable seulement. Remarquons qu'en posant, pour tout entier naturel n , $Y_n = X_n - m$, on peut se ramener, sans perte de généralité, à ne faire la démonstration que dans le cas d'une suite de variables aléatoires centrées. On notera σ^2 la variance commune des variables X_n .

Soit $(X_n)_{\mathbb{N}}$ une suite indépendante et identiquement distribuée de variables aléatoires réelles centrées de carré intégrable. Pour tout entier naturel non nul n , posons $S_n = X_1 + X_2 + \dots + X_n$.

- Montrons que la suite $\left(\frac{S_{n^2}}{n^2}\right)_{n \in \mathbb{N}^*}$ converge presque-sûrement vers 0.

En effet, pour tout $\varepsilon > 0$, nous avons, par application de l'inégalité de Bienaymé-Tchebychev, en tenant compte de l'égalité $\text{Var}\left(\frac{S_{n^2}}{n^2}\right) = \frac{\sigma^2}{n^2}$ (vérifier le calcul),

$$\mathbb{P}\left(\left|\frac{S_{n^2}}{n^2}\right| \geq \varepsilon\right) \leq \frac{\text{Var}\left(\frac{S_{n^2}}{n^2}\right)}{\varepsilon^2} = \frac{\sigma^2}{n^2 \varepsilon^2}.$$

Ce qui prouve que la série numérique à termes positifs $\sum_n \mathbb{P}\left(\left|\frac{S_{n^2}}{n^2}\right| \geq \varepsilon\right)$ est convergente, car son terme général est majorée par le terme général d'une série de Riemann convergente. On conclut alors, en appliquant l'item 2) de la proposition 6.17, que la suite $\left(\frac{S_{n^2}}{n^2}\right)_{n \in \mathbb{N}^*}$ converge presque-sûrement vers 0.

- Pour tout entier naturel non nul n , notons p_n la partie entière de \sqrt{n} , i.e. p_n est l'unique entier naturel vérifiant la double inégalité $p_n \leq \sqrt{n} < p_n + 1$. Montrons que la suite $\left(\frac{S_{p_n^2}}{n}\right)_{n \in \mathbb{N}^*}$ converge presque-sûrement vers 0.

De la double inégalité, on obtient $\sqrt{n} - 1 < p_n \leq \sqrt{n}$, puis, en élevant au carré, $n - 2\sqrt{n} + 1 < p_n^2 \leq n$, ce qui implique $\lim_{n \rightarrow +\infty} \frac{p_n^2}{n} = 1$ et $n - p_n^2 \leq 2\sqrt{n}$. Comme, d'après ce qui vient d'être

vu, la suite $\left(\frac{S_{n^2}}{n^2}\right)_{n \in \mathbb{N}^*}$ converge presque-sûrement vers 0, la suite $\left(\frac{S_{p_n^2}}{p_n^2}\right)_{n \in \mathbb{N}^*}$ converge aussi

presque-sûrement vers 0. De la relation, pour tout entier naturel non nul n , $\frac{S_{p_n^2}}{n} = \frac{S_{p_n^2}}{p_n^2} \cdot \frac{p_n^2}{n}$, on

peut conclure, en passant à la limite presque-sûre, que $\left(\frac{S_{p_n^2}}{n}\right)_{n \in \mathbb{N}^*}$ converge presque-sûrement vers 0.

- Pour conclure la démonstration du théorème, utilisons à nouveau l'inégalité de Bienaymé-Tchebychev. Pour tout $\varepsilon > 0$,

$$\mathbb{P}\left(\left|\frac{S_n}{n} - \frac{S_{p_n^2}}{n}\right| \geq \varepsilon\right) = \mathbb{P}\left(\left|\frac{1}{n} X_{p_n^2+1} + X_{p_n^2+2} + \dots + X_n\right| \geq \varepsilon\right) \leq \frac{(n - p_n^2)\sigma^2}{n^2 \varepsilon^2} \leq \frac{2\sqrt{n}\sigma^2}{n^2 \varepsilon^2},$$

car on a vu que $n - p_n^2 \leq 2\sqrt{n}$. Ce qui prouve que la série numérique à termes positifs $\sum_n \mathbb{P}\left(\left|\frac{S_n}{n} - \frac{S_{p_n^2}}{n}\right| \geq \varepsilon\right)$ est convergente, car son terme général est majorée par le terme général d'une série de Riemann convergente. Toujours en appliquant l'item 2) de la proposition 6.17, on en conclut que $\left(\frac{S_n}{n} - \frac{S_{p_n^2}}{n}\right)_{n \in \mathbb{N}^2}$ converge presque-sûrement vers 0. Donc en tenant compte des résultats démontrés dans les deux points précédents, et de l'inégalité triangulaire $\left|\frac{S_n}{n}\right| \leq \left|\frac{S_n}{n} - \frac{S_{p_n^2}}{n}\right| + \left|\frac{S_{p_n^2}}{n}\right|$, vraie pour tout entier naturel non nul n , on en conclut que $\left(\frac{S_n}{n}\right)_{n \in \mathbb{N}^*}$ converge aussi presque-sûrement vers 0. \square

En fait la réciproque de 6.19 est aussi vraie. Pour cela on se reportera à l'exercice 6.12, page 122, qui montre que si la v.a.r. X n'est pas intégrable, alors la suite $\left(\frac{X_1 + X_2 + \dots + X_n}{n}\right)_{n \in \mathbb{N}^*}$ ne peut pas converger presque-sûrement dans \mathbb{R} . Par contraposée, on obtient justement la réciproque de la proposition 6.19.

On notera que des conditions suffisantes d'existence de loi forte peuvent être démontrées sans l'hypothèse de l'identité des lois ou de l'indépendance mutuelle de la suite de v.a.r..

Par exemple on montre, et on les admettra, les deux propositions 6.20 et 6.21 (hors programme) suivantes :

Proposition 6.20.

Loi forte des grands nombres pour des v.a.r. uniformément bornées (Hors programme)

Si $(X_n)_{n \in \mathbb{N}}$ est une suite indépendante de v.a.r., centrées et telles qu'il existe $M > 0$ vérifiant, pour tous $n \in \mathbb{N}$ et $\omega \in \Omega$, $|X_n(\omega)| \leq M$, alors la suite des moyennes empiriques $\left(\frac{X_1 + X_2 + \dots + X_n}{n}\right)_{n \in \mathbb{N}^*}$ converge presque-sûrement vers 0.

Proposition 6.21.

Loi forte des grands nombres pour des v.a.r. deux à deux indépendantes (Hors programme)

Si $(X_n)_{n \in \mathbb{N}}$ est une suite de v.a.r. intégrables, deux à deux indépendantes, de même loi et d'espérance m , alors la suite des moyennes empiriques $\left(\frac{X_1 + X_2 + \dots + X_n}{n}\right)_{n \in \mathbb{N}^*}$ converge presque-sûrement vers m .

6.3 Convergence dans $L^p(\Omega, \mathcal{F}, \mathbb{P})$ où $p \in [1, +\infty]$

La topologie d'espace vectoriel normé de L^p permet de définir un autre mode de convergence pour les suites d'éléments de L^0 qui sont dans L^p .

Définition 6.7.

Si $p \in [1, +\infty]$, on dit que la suite de v.a.r. $(X_n)_{\mathbb{N}}$ converge en moyenne d'ordre p vers la v.a.r. X ou plus simplement converge dans L^p vers la v.a.r. X si $\widehat{X} \in L^p$ et pour tout $n \in \mathbb{N}$, $\widehat{X}_n \in L^p$ avec $\lim_n \|\widehat{X}_n - \widehat{X}\|_{L^p} = 0$. Si $p = 2$ on parle aussi de convergence en moyenne quadratique.

On note alors $(X_n)_{\mathbb{N}} \xrightarrow{L^p} X$ ou $X = L^p\text{-}\lim_n(X_n)$.

On ne confondra pas la convergence dans L^p avec la convergence de la suite des espérances mathématiques. Plus précisément :

Proposition 6.22.

Si la suite $(X_n)_{\mathbb{N}}$ converge dans L^p vers la v.a.r. X , alors la suite des espérances $(\mathbb{E}(X_n))_{\mathbb{N}}$ converge dans \mathbb{R} vers $\mathbb{E}(X)$.

La réciproque est fautive.

Concernant la réciproque on a cependant le résultat suivant :

Proposition 6.23.

Si la suite $(X_n)_{\mathbb{N}}$ converge p.s. vers la v.a.r. X et si, pour tout entier naturel n , $X_n \geq 0$, alors $(X_n)_{\mathbb{N}}$ converge dans L^1 vers la v.a.r. X si, et seulement si, $(\mathbb{E}(X_n))_{\mathbb{N}}$ converge dans \mathbb{R} vers $\mathbb{E}(X)$.

6.4 Comparaison des convergences dans $L^0(\Omega, \mathcal{F}, \mathbb{P})$

Il existe des relations entre les différents modes de convergence de suites de variables aléatoires. Ces relations sont la traduction en langage probabiliste des relations entre les modes de convergence vus en théorie de la mesure. Rappelons-les pour mémoire :

La convergence presque-sûre est plus forte que la convergence en probabilité comme le précise la proposition :

Proposition 6.24.

Si $(X_n)_{\mathbb{N}}$ est une suite de variables aléatoires réelles convergeant presque-sûrement vers la variable aléatoire réelle Y , alors la suite $(X_n)_{\mathbb{N}}$ converge en probabilité vers la variable aléatoire réelle Y .

La réciproque est fautive.

Démonstration : En partant de la définition de la convergence des suites, on vérifie d'abord que, pour tout réel $a > 0$,

$$\bigcap_{m \in \mathbb{N}} \bigcup_{n \geq m} \{|X_n - Y| \geq a\} \subseteq \Delta_Y.$$

On observe de plus que $\mathbb{P}(\Delta_Y) = 0$ et que, par le théorème de convergence monotone de Beppo-Lévi,

$$\mathbb{P}\left(\bigcap_{m \in \mathbb{N}} \bigcup_{n \geq m} \{|X_n - Y| \geq a\}\right) = \lim_{m \rightarrow +\infty} \mathbb{P}\left(\bigcup_{n \geq m} \{|X_n - Y| \geq a\}\right).$$

Comme, pour tout entier m ,

$$\mathbb{P}\left(\bigcup_{n \geq m} \{|X_n - Y| \geq a\}\right) \geq \mathbb{P}(\{|X_m - Y| \geq a\}),$$

on obtient

$$0 \geq \mathbb{P}(\Delta_Y) \geq \mathbb{P}\left(\bigcap_{m \in \mathbb{N}} \bigcup_{n \geq m} \{|X_n - Y| \geq a\}\right) \geq \lim_{m \rightarrow +\infty} \mathbb{P}(\{|X_m - Y| \geq a\}) \geq 0.$$

Ce qui entraîne que $\lim_{m \rightarrow +\infty} \mathbb{P}(|X_m - Y| \geq a) = 0$ c-à-d que la suite de variables aléatoires réelles $(X_n)_{\mathbb{N}}$ converge en probabilité vers la variable aléatoire réelle Y . \square

Ce dernier résultat n'est plus vrai si la mesure n'est pas finie ([2] p. 85 ex V-8). Quant à la réciproque, elle n'est pas vraie, la convergence en probabilité n'entraîne pas la convergence presque-sûre (cf. [14] ex 14-3). Cependant on a les résultats suivants :

Proposition 6.25.

Si $(X_n)_{\mathbb{N}}$ converge en probabilité vers X , alors il existe une sous-suite convergeant presque-sûrement vers X .

Le résultat suivant montre à quelle condition les notions de convergences en probabilité et presque-sûre coïncident (cf. [2] p. 86 ex V-10) :

Proposition 6.26.

Les notions de convergences en probabilité et presque-sûre coïncident si, et seulement si, la probabilité \mathbb{P} est discrète.

Tenant compte de ce que les variables aléatoires réelles de carré intégrable sont intégrables d'après la proposition 3.20, page 45, la loi faible des grands nombres apparaît ainsi comme une conséquence immédiate de la loi forte de Kolmogorov.

Proposition 6.27.

Si $1 < p < q < +\infty$ et si la suite $(X_n)_{\mathbb{N}}$ converge dans L^q vers la v.a.r. X , alors la suite $(X_n)_{\mathbb{N}}$ converge dans L^p vers la v.a.r. X .

La réciproque est fautive (cf. [14] ex 14-5). La topologie de L^q contient la topologie induite sur L^q par celle de L^p , mais on n'a pas l'égalité des topologies.

Concernant le lien entre les convergences presque-sûre et dans L^p , en général la convergence presque-sûre d'une suite de v.a. de L^p n'entraîne pas la convergence de cette suite dans L^p (cf. [14] ex 14-8). Cependant on a les deux résultats suivants :

Proposition 6.28.

Si $(X_n)_{\mathbb{N}}$ est dominée dans L^p , $1 \leq p < +\infty$, et converge presque-sûrement vers X , alors elle converge dans L^p vers X .

La réciproque de cette proposition est fautive, la convergence dans L^p n'entraîne pas la convergence presque-sûre (cf. [14] ex 14-7, [2] p. 83 ex V-1-(3)), mais :

Proposition 6.29.

Si $(X_n)_{n \in \mathbb{N}}$ converge dans L^p vers X , $1 \leq p < +\infty$, alors il existe une sous-suite convergeant presque-sûrement vers X .

Concernant le lien entre les convergences dans L^p et en probabilité, on a :

Proposition 6.30.

Si $(X_n)_{n \in \mathbb{N}}$ converge dans L^p vers X , $1 \leq p < +\infty$, alors elle converge en probabilité vers X .

La réciproque est fautive, la convergence en probabilité n'entraîne pas la convergence dans L^p (cf. [14] ex 14-6, [2] p. 83 ex V-1-(5)).

6.5 Exercices de révision sur les chapitres I à VI

Exercice 6.8. (Corrigé de l'exercice : page 192)

Soient $(X_n)_{n \geq 1}$ une suite de v.a.r. de carré intégrable non corrélées. On suppose qu'il existe un réel μ et un réel positif C tels que, pour tout $n \geq 1$, $\mathbb{E}[X_n] = \mu$ et $\text{Var}(X_n) \leq C$. Montrer que la suite

$$\left(\frac{X_1 + \cdots + X_n}{n} \right)_{n \geq 1}$$

converge vers μ dans L^2 et en probabilité.

Exercice 6.9. (Corrigé de l'exercice : page 193)**Théorème de Monte-Carlo**

Soit f une application de $[0, 1]$ dans \mathbb{R} de carré intégrable au sens de Lebesgue sur $[0, 1]$. On considère une suite indépendante $(U_n)_{n \in \mathbb{N}}$ de v.a.r. de loi uniforme sur $[0, 1]$. Démontrer directement et sans utiliser la loi des grands nombres que la suite des moyennes empiriques associée à la suite de v.a.r. $(f(U_n))_{n \in \mathbb{N}}$ converge en probabilité vers l'intégrale au sens de Lebesgue $\int_{[0,1]} f d\lambda$.

Exercice 6.10. (Corrigé de l'exercice : page 193)

A l'aide de la loi forte des grands nombres et en considérant une suite indépendante de v.a.r. $(X_i)_{i \geq 1}$ de même loi uniforme $\mathcal{U}([0, 1])$, calculer

$$\lim_{n \rightarrow +\infty} \int_{[0,1]^n} f\left(\frac{x_1 + \cdots + x_n}{n}\right) d\lambda^{(n)}(x_1, \dots, x_n),$$

où $\lambda^{(n)}$ est la mesure de Lebesgue dans \mathbb{R}^n et f une application continue bornée de \mathbb{R} dans \mathbb{R} .

Exercice 6.11. (Corrigé de l'exercice : page 194)**Théorème de Weierstrass**

Soient f une application continue de $[0, 1]$ dans \mathbb{R} et $x \in [0, 1]$. Pour tout $n \in \mathbb{N}^*$, notons S_n une v.a.r. binomiale de loi $\mathcal{B}(n, x)$.

1. Montrer que $p_n(x) := \mathbb{E}[f(\frac{1}{n}S_n)]$ est un polynôme en x appelé **polynôme de Bernstein de f** .
2. En utilisant l'uniforme continuité de f sur $[0, 1]$ montrer que, pour tout $\varepsilon > 0$, il existe $\delta > 0$ tel que, pour tout $n \in \mathbb{N}^*$ et tout $x \in [0, 1]$,

$$\begin{aligned} |p_n(x) - f(x)| &\leq \mathbb{E}[|f(\frac{1}{n}S_n) - f(x)|] \\ &\leq \varepsilon \mathbb{P}\left(\left|\frac{1}{n}S_n - x\right| < \delta\right) + 2\mathbb{P}\left(\left|\frac{1}{n}S_n - x\right| \geq \delta\right) \sup_{0 \leq x \leq 1} |f(x)|. \end{aligned}$$

En déduire que, pour tout $\varepsilon > 0$, il existe $\delta > 0$ tel que, pour tout $n \in \mathbb{N}^*$ et tout $x \in [0, 1]$,

$$|p_n(x) - f(x)| \leq \varepsilon + 2 \frac{x(1-x)}{n\delta^2} \sup_{0 \leq x \leq 1} |f(x)|.$$

3. Démontrer le **théorème de Weierstrass** : *Toute application continue de $[0, 1]$ dans \mathbb{R} est limite uniforme sur $[0, 1]$ d'une suite de polynômes.*

Exercice 6.12. (Corrigé de l'exercice : page 195)

Pour la première question de cet exercice, on pourra utiliser le résultat de l'exercice 4.17, page 92.

1. Soit X une v.a.r., montrer que $\mathbb{E}(|X|) \leq \sum_{n \geq 0} \mathbb{P}(|X| \geq n)$.
2. Soit $(X_n)_{\mathbb{N}}$ une suite indépendante de v.a.r. de même loi. On pose, pour tout $n \in \mathbb{N}^*$, $S_n := X_1 + X_2 + \dots + X_n$. En utilisant la relation

$$\frac{S_n}{n} - \frac{S_{n+1}}{n+1} = \frac{S_n}{n(n+1)} - \frac{X_{n+1}}{n+1},$$

Montrer que

$$\left\{ \left(\frac{1}{n} S_n \right)_{\mathbb{N}^*} \text{ converge dans } \mathbb{R} \right\} \subseteq \left[\limsup_{n \rightarrow +\infty} \{ |X_n| \geq n \} \right]^c.$$

3. On suppose de plus que X_1 n'est pas intégrable. A l'aide du lemme de Borel-Cantelli, montrer que la suite de v.a.r. $(\frac{1}{n} S_n)_{\mathbb{N}^*}$ ne converge pas presque-sûrement dans \mathbb{R} .

Exercice 6.13. (Corrigé de l'exercice : page 196)

Soit $(X_n)_{\mathbb{N}}$ une suite indépendante de v.a.r. intégrables, de même loi et d'espérance $m \neq 0$. On note, pour tout entier $n \geq 1$, $S_n := X_1 + X_2 + \dots + X_n$. Soit I un intervalle borné de \mathbb{R} .

1. Montrer l'inclusion

$$\left\{ \left(\frac{1}{n} S_n \right)_{\mathbb{N}^*} \text{ converge vers } m \right\} \subseteq \left(\limsup_n \{ S_n \in I \} \right)^c.$$

2. En déduire que presque-sûrement seulement un nombre fini des événements $\{S_n \in I\}$ sont réalisés.

Chapitre 7

Théorème-limite central et convergence de lois

Dans ce chapitre, sauf indication contraire, toutes les variables aléatoires considérées seront réelles et définies sur un même espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$. On notera $M_1(\mathbb{R})$, ou plus simplement M_1 , l'ensemble des mesures de probabilité sur \mathbb{R} . On se propose de munir cet ensemble d'une structure topologique appelée **topologie de la convergence étroite**. La notation $C_b(\mathbb{R}^d)$ désignera l'espace fonctionnel des applications, appelées **fonctions-test**, continues et bornées de \mathbb{R}^d dans \mathbb{R} .

7.1 Théorème-limite central (TLC)

7.1.1 Énoncé du théorème-limite central (TLC)

Pour introduire l'énoncé du TLC, commençons par une proposition élémentaire très utilisée en statistique inférentielle.

Proposition 7.1.

Si $(X_n)_{n \in \mathbb{N}}$ est une suite indépendante et identiquement distribuée (i.i.d.) de variables aléatoires réelles gaussiennes d'espérance m et de variance $\sigma^2 > 0$, alors, pour tout entier $n > 0$, la variable aléatoire réelle

$$\frac{X_1 + \cdots + X_n}{n} - m$$

est une variable aléatoire normale centrée de variance $\frac{\sigma^2}{n}$.

Démonstration : Soit $n \in \mathbb{N}^*$ fixé. Comme (X_1, \dots, X_n) est une suite indépendante de variables aléatoires réelles gaussiennes, la somme $X_1 + \cdots + X_n$ est une variable aléatoire réelle gaussienne de loi $\mathcal{N}_1(nm, n\sigma^2)$. Par suite la v.a.r.

$$\frac{X_1 + \cdots + X_n}{n} - m$$

est de loi $\mathcal{N}_1(0, \frac{\sigma^2}{n})$. \square

En particulier, pour tout $x \in \mathbb{R}$, on peut écrire avec les notations de la proposition précédente,

$$\mathbb{P} \left(\frac{X_1 + \cdots + X_n}{n} - m \leq x \sqrt{\frac{\sigma^2}{n}} \right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{u^2}{2}} du.$$

On peut aussi énoncer la proposition précédente sous la forme :

Proposition 7.2.

Si $(X_n)_{\mathbb{N}}$ est une suite indépendante et identiquement distribuée de variables aléatoires réelles gaussiennes d'espérance m et de variance $\sigma^2 > 0$, alors, pour tout entier $n > 0$, la variable aléatoire réelle $S_n := X_1 + \cdots + X_n$ est une variable aléatoire normale d'espérance nm et de variance $n\sigma^2$.

Si, dans les hypothèses de la proposition 7.1, on supprime la connaissance a priori de la loi commune des v.a.r., le résultat précédent devient seulement "asymptotiquement" vrai au sens précisé dans l'énoncé du théorème suivant, connu sous le nom de **Théorème-limite central** ou en abrégé **TLC**, très important en statistique inférentielle. Vu son importance, nous donnerons dans ce chapitre plusieurs énoncés équivalents de ce résultat que nous démontrerons à la page 138 après avoir étudié la notion de convergence étroite d'une suite de probabilités.

Proposition 7.3.

Théorème-limite central (version "moyenne empirique")

Si $(X_n)_{\mathbb{N}}$ est une suite indépendante et identiquement distribuée de variables aléatoires réelles de carré intégrable d'espérance m et de variance $\sigma^2 > 0$, alors, pour tout $x \in \mathbb{R}$,

$$\lim_{n \rightarrow +\infty} \mathbb{P} \left(\frac{X_1 + \cdots + X_n}{n} - m \leq x \sqrt{\frac{\sigma^2}{n}} \right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt.$$

La loi forte des grands nombres affirme que presque-sûrement, pour n assez grand, la moyenne empirique $\frac{X_1 + \cdots + X_n}{n}$ est proche de m . Le théorème-limite central quant à lui, donne des renseignements, pour n "assez grand", sur la loi approximative de l'erreur $\frac{X_1 + \cdots + X_n}{n} - m$ commise en prenant $\frac{X_1 + \cdots + X_n}{n}$ comme estimation de m .

D'un point de vue pratique, si $(X_n)_{\mathbb{N}}$ est une suite indépendante et identiquement distribuée de variables aléatoires réelles de carré intégrable d'espérance m et de variance $\sigma^2 > 0$, alors, pour tout entier naturel n assez grand et pour tout $x \in \mathbb{R}$,

$$\mathbb{P} \left(\frac{X_1 + \cdots + X_n}{n} - m \leq x \right) \approx \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x\sqrt{\frac{n}{\sigma^2}}} e^{-\frac{t^2}{2}} dt$$

c-à-d avec le changement de variable $u := t\sqrt{\frac{\sigma^2}{n}}$,

$$\mathbb{P} \left(\frac{X_1 + \cdots + X_n}{n} - m \leq x \right) \approx \frac{1}{\sqrt{2\pi\frac{\sigma^2}{n}}} \int_{-\infty}^x e^{-\frac{u^2}{2\frac{\sigma^2}{n}}} du.$$

On peut dire qu'**asymptotiquement**, i.e. pour tout entier n assez grand, l'erreur $\frac{X_1 + \dots + X_n}{n} - m$ suit la loi normale $\mathcal{N}_1(0, \frac{\sigma^2}{n})$ ou encore que la moyenne empirique $\frac{X_1 + \dots + X_n}{n}$ suit approximativement la loi $\mathcal{N}_1(m, \frac{\sigma^2}{n})$. Cela signifie qu'asymptotiquement la v.a.r. $S_n = X_1 + \dots + X_n$ peut être approximée par une variable aléatoire normale d'espérance nm et de variance $n\sigma^2$.

On peut alors donner une autre forme équivalente du TLC qui porte sur la somme de n v.a.r. i.i.d. :

Proposition 7.4.

Théorème-limite central (version "somme de n v.a.r.")

Si $(X_n)_{\mathbb{N}}$ est une suite indépendante et identiquement distribuée de variables aléatoires réelles de carré intégrable, d'espérance m et de variance $\sigma^2 > 0$, alors, pour tout réel x ,

$$\lim_{n \rightarrow +\infty} \mathbb{P} \left(\frac{S_n - nm}{\sqrt{n}} \leq x \right) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2\sigma^2}} dt.$$

On peut aussi exprimer le TCL sous une forme plus intuitive, calquée sur l'énoncé de la proposition 7.2 mais cette fois sans l'hypothèse de normalité des variables, ce qui conduit à un résultat analogue vrai seulement asymptotiquement sur n , au lieu de l'être pour tout n . C'est d'ailleurs souvent sous cette dernière forme que le TLC est énoncé et utilisé dans les applications pratiques en statistique.

Proposition 7.5.

Approximation d'une somme de v.a.r. par la loi normale

Si $(X_n)_{\mathbb{N}}$ est une suite indépendante et identiquement distribuée de variables aléatoires réelles de carré intégrable, d'espérance m et de variance $\sigma^2 > 0$, alors, pour tout entier naturel $n \geq 1$ suffisamment grand (en pratique $n \geq 30$), la variable aléatoire réelle $S_n := X_1 + \dots + X_n$ se comporte approximativement comme une variable aléatoire normale d'espérance nm et de variance $n\sigma^2$.

Exemples 7.1.

Considérons une suite i.i.d. $(X_n)_{\mathbb{N}}$ de variables aléatoires réelles de loi uniforme sur l'intervalle $[-\frac{1}{2}, \frac{1}{2}]$. Posons $S := \sum_{k=1}^{365} X_k$. Alors $\mathbb{P}(|S| \leq 15) \approx 0,99$.

En effet, en gardant les mêmes notations que plus haut, un calcul d'espérance et de variance pour la variable aléatoire réelle X_1 de loi uniforme sur l'intervalle $[-\frac{1}{2}, \frac{1}{2}]$ on obtient, $m = \mathbb{E}(X_1) = 0$ et $\sigma^2 = \text{Var}(X_1) = \frac{1}{12}$. Par suite $\mathbb{E}(S) = 0$ et, en vertu de l'identité des lois et de l'indépendance de la suite $(X_n)_{\mathbb{N}}$,

$$\text{Var}(S) = \text{Var} \left(\sum_{k=1}^{365} X_k \right) = \sum_{k=1}^{365} \text{Var}(X_k) = \frac{365}{12}.$$

Comme d'après ce qui a été écrit plus haut, S est pratiquement une variable aléatoire réelle normale car n peut être considéré comme grand, on en déduit que la loi de S est pratiquement $\mathcal{N}_1(0, \frac{365}{12})$. Cela signifie aussi que la variable aléatoire réelle $\sqrt{\frac{12}{365}}S$ est

pratiquement de loi $\mathcal{N}_1(0, 1)$. Par suite comme $15\sqrt{\frac{12}{365}} \approx 2,72$,

$$\mathbb{P}(|S| \leq 15) = \mathbb{P}\left(\left|\sqrt{\frac{12}{365}}S\right| \leq 15\sqrt{\frac{12}{365}}\right) = \int_{-2,72}^{2,72} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt.$$

En utilisant la table de la loi normale centrée réduite de l'annexe B, page 211, on trouve $\mathbb{P}(|S| \leq 15) \approx 0,99$.

Comparons ce résultat à ce que donnerait l'inégalité de Bienaymé-Tchébychev. Par application de cette inégalité,

$$\mathbb{P}(|S| > 15) = \mathbb{P}(|S - \mathbb{E}(S)| > 15) \leq \frac{\text{Var}(S)}{15^2} \approx 0,135,$$

c-à-d $\mathbb{P}(|S| \leq 15) \geq 0,865$. \square

On voit qu'un des intérêts du TLC est de permettre d'approximer la loi d'une variable aléatoire réelle dans des situations où le calcul exact de cette loi serait pratiquement très compliqué, voire impossible.

7.1.2 Cas particuliers du théorème-limite central (TLC)

Appliquons le TLC à des cas particuliers. Voici un corollaire du théorème-limite central historiquement important.

Proposition 7.6.

Approximation d'une loi binomiale par une loi normale (Théorème de De Moivre-Laplace)

Si $p \in]0, 1[$ et, pour tout entier $n \geq 1$, Z_n est une variable aléatoire réelle de loi binomiale $\mathcal{B}(n, p)$, alors pour tout réel x

$$\lim_{n \rightarrow +\infty} \mathbb{P}\left(\frac{Z_n - np}{\sqrt{np(1-p)}} \leq x\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt.$$

Démonstration : D'après la proposition 4.33, page 86, la variable aléatoire réelle

$$\frac{Z_n - np}{\sqrt{np(1-p)}}$$

a même loi que la variable aléatoire réelle

$$Y_n := \frac{X_1 + \cdots + X_n - np}{\sqrt{np(1-p)}}$$

où $(X_n)_{\mathbb{N}}$ est une suite indépendante et identiquement distribuée de variables aléatoires réelles de Bernoulli de loi $\mathcal{B}(p)$. De plus avec les notations habituelles, $m = \mathbb{E}(X_0) = p$, $\sigma^2 = \text{Var}(X_0) = p(1-p)$.

On applique alors le théorème-limite central (version "moyenne empirique", cf. proposition 7.3)

à la suite $(X_n)_\mathbb{N}$. \square

Le théorème de De Moivre-Laplace exprime que dans la pratique, pour n assez grand, une variable aléatoire réelle X binomiale de taille n peut être approximée par une variable aléatoire réelle normale X' . Plus précisément, si a et b sont des réels avec $a < b$, alors

$$\mathbb{P}(a < X \leq b) = \mathbb{P}\left(\frac{a - np}{\sqrt{np(1-p)}} < \frac{X - np}{\sqrt{np(1-p)}} \leq \frac{b - np}{\sqrt{np(1-p)}}\right).$$

D'où pour un entier n assez grand,

$$\mathbb{P}(a < X \leq b) \approx \frac{1}{\sqrt{2\pi}} \int_{\frac{a-np}{\sqrt{np(1-p)}}}^{\frac{b-np}{\sqrt{np(1-p)}}} e^{-\frac{t^2}{2}} dt$$

c-à-d avec le changement de variable $t = \frac{u - np}{\sqrt{np(1-p)}}$,

$$\mathbb{P}(a < X \leq b) \approx \frac{1}{\sqrt{2\pi np(1-p)}} \int_a^b \exp\left(-\frac{(u - np)^2}{2np(1-p)}\right) du = \mathbb{P}(a < X' \leq b)$$

où X' est une variable aléatoire réelle normale d'espérance np et de variance $np(1-p)$.

En conclusion, on peut dire qu'**asymptotiquement**, i.e. pour n assez grand, la v.a.r. X binomiale $\mathcal{B}(n, p)$, peut être approximée par une variable aléatoire normale X' d'espérance np et de variance $np(1-p)$. Dans la pratique cette approximation est considérée satisfaisante si $n \geq 20$ et $p \approx \frac{1}{2}$.

Exercice 7.1. (Corrigé de l'exercice : page 196)

Un cinéma comporte deux salles contenant chacune n places. N personnes se présentent à l'entrée de ce cinéma. On admet que les choix des spectateurs sont indépendants les uns des autres et qu'un spectateur quelconque a une chance sur deux d'aller dans la première salle.

1. Quelle est la probabilité P que tous les spectateurs ne puissent pas voir le film qu'ils ont choisi ?
2. Comment le constructeur aurait-il dû choisir n si on sait que $N = 1000$ et si on veut que $P \leq 0,01$?

On peut faire, pour les variables aléatoires réelles de Poisson, un raisonnement analogue à celui fait pour les variables binomiales pour obtenir la proposition suivante :

Proposition 7.7.

Approximation d'une loi de Poisson par une loi normale

Soient un réel $\alpha > 0$ et, pour tout entier $n \geq 1$, Z_n une variable aléatoire réelle de loi de Poisson $\mathcal{P}(n\alpha)$, alors pour tout réel x ,

$$\lim_{n \rightarrow +\infty} \mathbb{P}\left(\frac{Z_n - n\alpha}{\sqrt{n\alpha}} \leq x\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt.$$

Démonstration : La démonstration de cette proposition se calque facilement sur celle du théorème de De Moivre-Laplace. \square

Une conséquence de la proposition 7.7 est que, pour un réel β assez grand, une variable aléatoire réelle X de Poisson $\mathcal{P}(\beta)$ peut être approximée par une variable aléatoire réelle X' normale $\mathcal{N}_1(\beta, \beta)$. Dans la pratique cette approximation est considérée satisfaisante si $\beta \geq 10$.

7.1.3 Correction de continuité

Prenons le cas de l'approximation d'une variable aléatoire réelle X binomiale $\mathcal{B}(n, p)$, par une variable aléatoire réelle X' normale $\mathcal{N}_1[np, np(1-p)]$. Si on remplace "brutalement" X par X' dans les calculs, on aura, pour tout entier $0 \leq k \leq n$, $\mathbb{P}(X = k) \approx \mathbb{P}(X' = k) = 0$ car X' étant une variable aléatoire réelle à densité, $\mathbb{P}(X' = x) = 0$ pour tout réel x . Ce qui n'a aucun intérêt.

En général pour éviter cet inconvénient, quand on veut approximer dans un calcul pratique une variable aléatoire réelle X discrète portée par \mathbb{N} par une variable aléatoire réelle X' admettant une densité ρ sur \mathbb{R} de fonction de répartition F , on effectue une correction appelée **correction de continuité**.

Les corrections de continuité sont données par :

1. Pour tout entier $0 \leq k \leq n$, on approxime $\mathbb{P}(X = k)$ de la façon suivante :

$$\mathbb{P}(X = k) = \mathbb{P}\left(k - \frac{1}{2} \leq X \leq k + \frac{1}{2}\right) \approx \mathbb{P}\left(k - \frac{1}{2} \leq X' \leq k + \frac{1}{2}\right)$$

c'est-à-dire

$$\mathbb{P}(X = k) \approx \int_{k-\frac{1}{2}}^{k+\frac{1}{2}} \rho(t) d\lambda(t) = F\left(k + \frac{1}{2}\right) - F\left(k - \frac{1}{2}\right).$$

2. Plus généralement, si a et b sont des réels avec $a < b$, on approxime $\mathbb{P}(a < X < b)$ de la façon suivante (mêmes écritures avec les inégalités larges) :

$$\mathbb{P}(a < X < b) = \mathbb{P}\left(a - \frac{1}{2} \leq X \leq b + \frac{1}{2}\right) \approx \mathbb{P}\left(a - \frac{1}{2} \leq X' \leq b + \frac{1}{2}\right)$$

c'est-à-dire

$$\mathbb{P}(a < X < b) \approx \int_{a-\frac{1}{2}}^{b+\frac{1}{2}} \rho(t) d\lambda(t) = F\left(b + \frac{1}{2}\right) - F\left(a - \frac{1}{2}\right).$$

3. On écrit, pour tout réel b , (mêmes écritures avec les inégalités larges) :

$$\mathbb{P}(X < b) = \mathbb{P}\left(X \leq b + \frac{1}{2}\right) \approx \mathbb{P}\left(X' \leq b + \frac{1}{2}\right) = \int_{-\infty}^{b+\frac{1}{2}} \rho(t) d\lambda(t) = F\left(b + \frac{1}{2}\right).$$

4. On écrit, pour tout réel a , (mêmes écritures avec les inégalités larges) :

$$\mathbb{P}(a < X) = \mathbb{P}\left(a - \frac{1}{2} \leq X\right) \approx \mathbb{P}\left(a - \frac{1}{2} \leq X'\right) = \int_{a-\frac{1}{2}}^{+\infty} \rho(t) d\lambda(t) = 1 - F\left(a - \frac{1}{2}\right).$$

7.2 Convergence d'une suite de probabilités, convergence en loi

L'importance du théorème-limite central conduit à introduire une notion de convergence dans l'ensemble $M_1(\mathbb{R})$ des probabilités sur \mathbb{R} . Auparavant disons qu'un réel x est **point de continuité**, resp. **point de discontinuité**, d'une application f de \mathbb{R} dans \mathbb{R} si l'application f est continue en x , resp. discontinue en x .

Définition 7.1.

Soient $(\mu_n)_{\mathbb{N}}$ une suite de probabilités sur \mathbb{R} et μ une probabilité sur \mathbb{R} . On dit que la suite de probabilités $(\mu_n)_{\mathbb{N}}$ **converge étroitement** vers la probabilité μ , si, pour tout point de continuité x de la fonction de répartition F_μ de μ , la suite réelle $(F_{\mu_n}(x))_{\mathbb{N}}$ converge vers le réel $F_\mu(x)$.

On remarquera qu'on n'exige pas que la suite de réels $(F_{\mu_n}(x))_{\mathbb{N}}$ converge vers $F_\mu(x)$ aux points $x \in \mathbb{R}$ où la fonction de répartition F_μ de μ n'est pas continue. Il se peut qu'il n'y ait pas convergence en certains points de discontinuité.

On notera le résultat suivant d'analyse, conséquence de la monotonie de la fonction de répartition :

Proposition 7.8.

L'ensemble des points de discontinuité d'une fonction de répartition, est dénombrable.

Démonstration : On rappelle qu'on a vu dans le premier chapitre, proposition 1.14, page 15, que la fonction de répartition de μ n'est pas continue en un point x si, et seulement si, $\mu(\{x\}) > 0$. On en déduit donc que l'ensemble des discontinuités de F_μ est égal à

$$\bigcup_{k=1}^{+\infty} \left\{ x \in \mathbb{R} / \mu(\{x\}) > \frac{1}{k} \right\}.$$

Mais, pour tout entier $k \geq 1$, $\{x \in \mathbb{R} / \mu(\{x\}) > \frac{1}{k}\}$ contient au plus k éléments. Par suite l'ensemble des discontinuités de la fonction de répartition d'une probabilité est une réunion dénombrable d'ensembles dénombrables, donc est un ensemble dénombrable. \square

On prendra garde que, si une suite de fonctions de répartition de probabilités converge, sa limite n'est pas nécessairement la fonction de répartition d'une probabilité, comme le montre l'exercice 7.2 ci-dessous. En revanche, on admettra que :

Proposition 7.9.

La limite, si elle existe, d'une suite de fonctions de répartition est une application de \mathbb{R} dans $[0, 1]$ continue à droite et croissante.

Exercice 7.2. (Corrigé de l'exercice : page 197)

En considérant la suite de probabilités de Dirac au point n sur \mathbb{R} , $(\delta_n)_{\mathbb{N}}$, montrer que la limite, si elle existe, d'une suite de fonctions de répartition n'est pas nécessairement une fonction de répartition et qu'on peut avoir la convergence simple de la suite $(F_{\mu_n})_{\mathbb{N}}$ sans

que la suite $(\mu_n)_\mathbb{N}$ converge.

Définition 7.2.

Si $(X_n)_\mathbb{N}$ est une suite de variables aléatoires réelles et X une v.a.r., on dit que la suite $(X_n)_\mathbb{N}$ **converge en loi** vers la variable aléatoire réelle X pour exprimer que la suite $(\mathbb{P}_{X_n})_\mathbb{N}$ des lois des variables aléatoires réelles X_n converge vers la loi \mathbb{P}_X de la variable aléatoire réelle X .

On prendra garde que cette dernière terminologie est très dangereuse car une suite de variables aléatoires réelles $(X_n)_\mathbb{N}$ peut converger en loi vers des variables aléatoires réelles différentes. Tout ce qu'on peut affirmer a priori c'est que toutes ces variables aléatoires réelles auront alors la même loi, comme le précise le résultat qui suit :

Proposition 7.10.

Si une suite de probabilités sur \mathbb{R} , $(\mu_n)_\mathbb{N}$, converge étroitement vers deux probabilités sur \mathbb{R} , μ et ν , alors $\mu = \nu$.

Démonstration : Notons $C(F_\mu)$, resp. $C(F_\nu)$, l'ensemble des points de continuité de la fonction de répartition de μ , resp. ν . Comme l'ensemble des discontinuités de la fonction de répartition d'une probabilité est un ensemble dénombrable (cf. proposition 7.8), l'ensemble des points de continuité communs à F_μ et F_ν , i.e. $C(F_\mu) \cap C(F_\nu)$, est dense dans \mathbb{R} comme complémentaire d'un ensemble dénombrable. Pour tout $x \in C(F_\mu) \cap C(F_\nu)$, $F_\mu(x) = \lim_n F_{\mu_n}(x) = F_\nu(x)$ car la suite $(\mu_n)_\mathbb{N}$ converge à la fois vers les deux probabilités μ et ν . F_μ et F_ν étant continues à droite sur \mathbb{R} , on en conclut que $F_\mu = F_\nu$ et par suite $\mu = \nu$. \square

On admettra qu'on peut munir l'ensemble $M_1(\mathbb{R})$ d'une structure d'espace métrique dont la topologie associée est celle de la convergence étroite (cf. [3], exercice V-17). Précisément :

Proposition 7.11.

(Hors programme)

1. L'application d de $M_1 \times M_1$ dans \mathbb{R}^+ définie par

$$d(\mu, \nu) := \inf \{ \varepsilon > 0; \forall x \in \mathbb{R}, F_\mu(x - \varepsilon) - \varepsilon \leq F_\nu(x) \leq F_\mu(x + \varepsilon) + \varepsilon \},$$

où F_μ et F_ν désignent les fonctions de répartition respectivement de μ et ν , définit une métrique, dite **métrique de Lévy**, sur $M_1(\mathbb{R})$.

2. La suite $(\mu_n)_\mathbb{N}$ de probabilités converge étroitement vers la probabilité μ si, et seulement si, la suite réelle $(d(\mu_n, \mu))_\mathbb{N}$ converge vers 0.
3. De plus, pour tout couple de v.a.r. (X, Y) , on a $d(\mathbb{P}_X, \mathbb{P}_Y) \leq \Delta(\hat{X}, \hat{Y})$, où Δ désigne la métrique de Ky-Fan (cf. proposition 6.9, page 111).

En utilisant la terminologie (abusiv) de la convergence en loi et la version "somme de n v.a.r." du TLC, on vérifie aisément qu'on peut énoncer le TLC en termes de convergence en loi. Cet énoncé englobe le cas où $\sigma^2 = 0$ et se généralise au cas où les variables aléatoires sont vectorielles (cf. proposition 7.25, page 138) :

Proposition 7.12.**Théorème-limite central (version réelle)**

Soient $(X_n)_{n \in \mathbb{N}}$ une suite indépendante et identiquement distribuée de variables aléatoires réelles de carré intégrable (i.e. $\mathbb{E}(|X_0|^2) < +\infty$), d'espérance $m \in \mathbb{R}$ et de variance σ^2 . On pose, pour tout entier naturel non nul n , $S_n = X_1 + X_2 + \cdots + X_n$. Alors, la suite de variables aléatoires réelles $\left(\frac{S_n - nm}{\sqrt{n}}\right)_{n \in \mathbb{N}^*}$ converge en loi vers une variable aléatoire gaussienne centrée de variance σ^2 .

Exercice 7.3. (Corrigé de l'exercice : page 197)

Soit $(X_n)_{n \geq 1}$ une suite indépendante de variables aléatoires réelles de même loi de Poisson

$\mathcal{P}(1)$. Pour $n \geq 1$, on pose $S_n := \sum_{k=1}^n X_k$. Préciser la loi de S_n et calculer la limite de la suite réelle

$$\left(e^{-n} \sum_{k=0}^n \frac{n^k}{k!}\right)_{n \geq 1}.$$

Dans le cas où la suite de probabilités est composée de probabilités discrètes portées par \mathbb{N} , on utilise le critère de convergence suivant :

Proposition 7.13.**Critère de convergence pour les probabilités discrètes**

Soient, pour tout $n \in \mathbb{N}$, μ_n et μ des probabilités discrètes portées par \mathbb{N} . Alors la suite $(\mu_n)_{n \in \mathbb{N}}$ converge vers μ si, et seulement si, pour tout $k \in \mathbb{N}$, la suite de réels $(\mu_n(\{k\}))_{n \in \mathbb{N}}$ converge vers $\mu(\{k\})$.

Démonstration : Montrons la condition nécessaire. Comme μ est une probabilité discrète portée par \mathbb{N} , sa fonction de répartition est définie sur \mathbb{R} par

$$F_\mu(x) = \sum_{k=0}^{+\infty} \mu(\{k\}) \mathbb{1}_{[k, +\infty[}(x).$$

On a une écriture analogue pour les fonctions de répartition des probabilités μ_n . L'ensemble des points de discontinuité de F_μ est inclus dans \mathbb{N} . Par suite, pour tout entier k , $k + \frac{1}{2}$ et $k - \frac{1}{2}$ sont des points de continuité de F_μ et F_{μ_n} . Comme $\mu(\{k\}) = F_\mu(k + \frac{1}{2}) - F_\mu(k - \frac{1}{2})$, il vient en utilisant le fait que, pour tout point de continuité x de F_μ , $F_\mu(x) = \lim_n F_{\mu_n}(x)$,

$$\begin{aligned} \mu(\{k\}) &= \lim_n F_{\mu_n}\left(k + \frac{1}{2}\right) - \lim_n F_{\mu_n}\left(k - \frac{1}{2}\right) \\ &= \lim_n \left[F_{\mu_n}\left(k + \frac{1}{2}\right) - F_{\mu_n}\left(k - \frac{1}{2}\right) \right] \\ &= \lim_n \mu_n(\{k\}). \end{aligned}$$

La condition suffisante est immédiate d'après l'écriture des fonctions de répartition de μ et μ_n . \square

On notera que le critère précédent devient faux si les probabilités sont portées par une partie dénombrable D de \mathbb{R} dont les points ne sont pas tous topologiquement isolés. On rappelle qu'un

point $x \in D$ est dit **topologiquement isolé** s'il existe un intervalle de la forme $]x - \varepsilon, x + \varepsilon[$, avec $\varepsilon > 0$, ne contenant pas d'autre point de D que x lui-même. On peut construire un contre-exemple en considérant, pour tout entier $n \geq 1$, les probabilités $\mu_n := \delta_{\frac{1}{n}}$ et prendre $\mu := \delta_0$.

Donnons comme exemple d'application aux lois classiques du critère des probabilités discrètes le résultat suivant :

Proposition 7.14.

Soit $\alpha \in]0, +\infty[$. Si $(p_n)_{\mathbb{N}}$ est une suite de réels de $]0, 1[$ telle que $\lim_n(np_n) = \alpha$, alors la suite de probabilités $(\mathcal{B}(n, p_n))_{\mathbb{N}^*}$ converge vers $\mathcal{P}(\alpha)$.

Démonstration : Fixons $k \in \mathbb{N}$, par définition de la probabilité binomiale

$$\mathcal{B}(n, p_n)(\{k\}) = C_n^k p_n^k (1 - p_n)^{n-k} = \frac{n(n-1)\dots(n-k+1)}{k!} p_n^k (1 - p_n)^{n-k}.$$

Au voisinage de $+\infty$, n^k est un équivalent de $n(n-1)\dots(n-k+1)$ et $\mathcal{B}(n, p_n)(\{k\})$ admet pour équivalent

$$\frac{(np_n)^k}{k!} (1 - p_n)^{n-k}.$$

De plus, toujours au voisinage de $+\infty$,

$$\ln(1 - p_n)^{n-k} = (n-k) \ln(1 - p_n) \sim (n-k)(-p_n) \sim \frac{n-k}{n}(-\alpha) \sim -\alpha,$$

par suite $\lim_n (1 - p_n)^{n-k} = e^{-\alpha}$. Par passage à la limite on obtient donc

$$\lim_{n \rightarrow +\infty} \frac{(np_n)^k}{k!} (1 - p_n)^{n-k} = \frac{\alpha^k}{k!} e^{-\alpha}$$

et par équivalence

$$\lim_{n \rightarrow +\infty} \mathcal{B}(n, p_n)(\{k\}) = \frac{\alpha^k}{k!} e^{-\alpha} = \mathcal{P}(\alpha)(\{k\}),$$

pour tout entier k , ce qui donne le résultat cherché. \square

Dans les calculs pratiques ce résultat est utilisé de la façon suivante :

si $n \geq 30$, $p \leq 0,1$ $np \leq 10$, on assimile une variable binomiale de loi $\mathcal{B}(n, p)$ à une variable de Poisson de paramètre np .

Exercice 7.4. (Corrigé de l'exercice : page 198)

Un fabricant produit des transistors dont un pour cent sont défectueux. Il les ensache par paquets de 100 et les garantit à 98 pour cent. Quelle est la probabilité que cette garantie tombe en défaut ?

Donnons, toujours à titre d'exemple, une approximation de la loi hypergéométrique.

Définition 7.3.

On appelle **loi hypergéométrique** de paramètres N, n, p , la probabilité sur \mathbb{R} définie par

$$\mathcal{H}(N, n, p) := \sum_{k=0}^n \frac{C_{Np}^k C_{Nq}^{n-k}}{C_N^n} \delta_k,$$

où $n \in \mathbb{N}^*$, $N \in \mathbb{N}^*$, $p \in]0, 1[$ tels que $Np \in \mathbb{N}$, et $q := 1 - p$.

On trouvera la valeur de l'espérance et de la variance de la loi hypergéométrique dans le formulaire de l'annexe A, page 205.

Proposition 7.15.

Soient $n \in \mathbb{N}^*$ et $p \in]0, 1[$. Notons $S := \{N \in \mathbb{N} / Np \in \mathbb{N}\}$.

Alors la suite de probabilités $(\mathcal{H}(N, n, p))_{N \in S}$ converge étroitement vers $\mathcal{B}(n, p)$.

Exercice 7.5. (Corrigé de l'exercice : page 198)

Démontrer la proposition 7.15 précédente.

Dans les calculs pratiques ce résultat est utilisé de la façon suivante :

si $N > 10n$, on assimile une variable hypergéométrique de loi $\mathcal{H}(N, n, p)$ à une variable binomiale de loi $\mathcal{B}(n, p)$.

On notera aussi qu'une suite de probabilités discrètes peut converger vers une probabilité non-discrète comme le prouve le théorème de De Moivre-Laplace.

Dans les cas généraux, on dispose de critères de convergence étroite utilisant des familles de fonctions-test, par exemple le critère suivant (pour une démonstration, on pourra se reporter à [4], page 178) :

Proposition 7.16.**Critère de convergence étroite par les fonctions continues bornées**

Soient $(\mu_n)_{\mathbb{N}}$ une suite de probabilités sur \mathbb{R} et μ une probabilité sur \mathbb{R} . La suite de probabilités $(\mu_n)_{\mathbb{N}}$ converge étroitement vers la probabilité μ si, et seulement si, pour tout $f \in C_b(\mathbb{R})$, la suite de réels $(\int_{\mathbb{R}} f d\mu_n)_{\mathbb{N}}$ converge dans \mathbb{R} vers $\int_{\mathbb{R}} f d\mu$.

Énoncé en terme de convergence en loi de v.a.r. cette dernière proposition devient, par application du théorème du transfert :

Proposition 7.17.**Critère de convergence en loi par les fonctions continues bornées**

Soient $(X_n)_{\mathbb{N}}$ une suite de v.a.r. et X une v.a.r.. La suite de v.a.r. $(X_n)_{\mathbb{N}}$ converge en loi vers la v.a.r. X si, et seulement si, pour tout $f \in C_b(\mathbb{R})$, la suite de réels $(\mathbb{E}(f(X_n)))_{\mathbb{N}}$ converge dans \mathbb{R} vers $\mathbb{E}(f(X))$.

A titre d'application de ce résultat, prouvons la proposition suivante :

Proposition 7.18.

Soit $(X_n)_{n \in \mathbb{N}}$ une suite de v.a.r. et f une application continue de \mathbb{R} dans \mathbb{R} . Si la suite de v.a.r. $(X_n)_{n \in \mathbb{N}}$ converge en loi vers une v.a.r. X , alors la suite de v.a.r. $(f(X_n))_{n \in \mathbb{N}}$ converge en loi vers $f(X)$.

Démonstration : On remarque tout d'abord que, pour tout $n \in \mathbb{N}$, l'application $f(X_n)$ de Ω dans \mathbb{R} est bien une variable aléatoire puisqu'elle est la composée de l'application X_n qui est $(\mathcal{F}, \mathcal{B}(\mathbb{R}))$ -mesurable et de l'application f qui est continue donc $(\mathcal{B}(\mathbb{R}), \mathcal{B}(\mathbb{R}))$ -mesurable. Si φ est une application de \mathbb{R} dans \mathbb{R} continue bornée alors $\varphi \circ f$ est également continue bornée. L'hypothèse de convergence en loi de la suite $(X_n)_{n \in \mathbb{N}}$ vers X entraîne, compte tenu de la proposition 7.16 appliquée à $\varphi \circ f$ et du théorème de transfert, $\lim_n \mathbb{E}[\varphi \circ f(X_n)] = \lim_n \int_{\mathbb{R}} \varphi \circ f(x) d\mathbb{P}_{X_n} = \int_{\mathbb{R}} \varphi \circ f(x) d\mathbb{P}_X = \mathbb{E}[\varphi \circ f(X)]$ que l'on peut écrire $\lim_n \mathbb{E}[\varphi(f(X_n))] = \mathbb{E}[\varphi(f(X))]$. Ceci étant valable pour toute application continue bornée φ , on conclut que la suite $(f(X_n))_{n \in \mathbb{N}}$ converge en loi vers $f(X)$. \square

Il peut être utile d'avoir des critères utilisant d'autres familles de fonctions-test, comme par exemple celle des fonctions continues à support compact ou celle des fonctions continues nulles à l'infini :

Proposition 7.19.**Critère de convergence étroite par les fonctions continues à support compact**

Soient $(\mu_n)_{n \in \mathbb{N}}$ une suite de probabilités sur \mathbb{R}^d et μ une probabilité sur \mathbb{R}^d . La suite de probabilités $(\mu_n)_{n \in \mathbb{N}}$ converge étroitement vers la probabilité μ si, et seulement si, pour toute fonction f de \mathbb{R}^d dans \mathbb{R} , continue et à support compact sur \mathbb{R}^d , la suite de réels $(\int_{\mathbb{R}^d} f d\mu_n)_{n \in \mathbb{N}}$ converge dans \mathbb{R} vers $\int_{\mathbb{R}^d} f d\mu$.

Démonstration : Faisons la démonstration dans le cas où $d = 1$. On admettra le résultat pour $d > 1$.

• C.N. - Si la suite $(\mu_n)_{n \in \mathbb{N}}$ de probabilités converge étroitement vers la probabilité μ , alors, d'après la proposition 7.16, pour tout $f \in C_b(\mathbb{R})$, la suite de réels $(\int_{\mathbb{R}} f d\mu_n)_{n \in \mathbb{N}}$ converge dans \mathbb{R} vers $\int_{\mathbb{R}} f d\mu$. Comme les fonctions continues à support compact sont bornées, on a bien que, pour toute fonction f continue et à support compact sur \mathbb{R} , la suite de réels $(\int_{\mathbb{R}} f d\mu_n)_{n \in \mathbb{N}}$ converge dans \mathbb{R} vers $\int_{\mathbb{R}} f d\mu$.

• C.S. - On suppose que, pour toute fonction f continue et à support compact sur \mathbb{R} , la suite de réels $(\int_{\mathbb{R}} f d\mu_n)_{n \in \mathbb{N}}$ converge dans \mathbb{R} vers $\int_{\mathbb{R}} f d\mu$.

◇ Considérons, pour tout entier naturel non nul k , la fonction φ_k , définie, pour tout réel x , par

$$\varphi_k(x) = \begin{cases} 0, & \text{si } x < -(k+1) \text{ ou } x > k+1; \\ 1, & \text{si } -k \leq x \leq k; \\ -x + k + 1, & \text{si } k \leq x \leq k+1; \\ x + k + 1, & \text{si } -(k+1) \leq x \leq -k. \end{cases}$$

Pour tout entier naturel non nul k , la fonction φ_k , est une fonction continue à support compact telle que $0 \leq \varphi_k \leq 1$ et la suite de fonctions $(\varphi_k)_{\mathbb{N}^*}$ converge simplement sur \mathbb{R} vers la fonction constante égale à 1. Donc, la suite $(1 - \varphi_k)_{\mathbb{N}^*}$ converge simplement sur \mathbb{R} vers la fonction nulle et cette suite de fonctions est dominée par la fonction μ -intégrable $\mathbf{1}_{\mathbb{R}}$, car μ est par hypothèse une probabilité. Donc par le théorème de convergence dominée, on obtient

$$\lim_{k \rightarrow \infty} \int_{\mathbb{R}} (1 - \varphi_k) d\mu = \int_{\mathbb{R}} \lim_{k \rightarrow \infty} (1 - \varphi_k) d\mu = 0.$$

◇ Soit h une fonction continue, bornée sur \mathbb{R} par M . On a, pour tout entier naturel non nul k , et tout entier naturel n ,

$$\begin{aligned} \left| \int_{\mathbb{R}} h d\mu_n - \int_{\mathbb{R}} h d\mu \right| &\leq \left| \int_{\mathbb{R}} (h - h\varphi_k) d\mu_n \right| + \left| \int_{\mathbb{R}} h\varphi_k d\mu_n - \int_{\mathbb{R}} h\varphi_k d\mu \right| + \left| \int_{\mathbb{R}} (h\varphi_k - h) d\mu \right| \\ &\leq M \int_{\mathbb{R}} (1 - \varphi_k) d\mu_n + \left| \int_{\mathbb{R}} h\varphi_k d\mu_n - \int_{\mathbb{R}} h\varphi_k d\mu \right| + M \int_{\mathbb{R}} (1 - \varphi_k) d\mu, \\ &\leq M \int_{\mathbb{R}} (1 - \varphi_k) d\mu + M \left| \int_{\mathbb{R}} (1 - \varphi_k) d\mu_n - \int_{\mathbb{R}} (1 - \varphi_k) d\mu \right| + \\ &\quad + \left| \int_{\mathbb{R}} h\varphi_k d\mu_n - \int_{\mathbb{R}} h\varphi_k d\mu \right| + M \int_{\mathbb{R}} (1 - \varphi_k) d\mu, \\ &\leq M \int_{\mathbb{R}} (1 - \varphi_k) d\mu + M \left| \int_{\mathbb{R}} \varphi_k d\mu_n - \int_{\mathbb{R}} \varphi_k d\mu \right| + \\ &\quad + \left| \int_{\mathbb{R}} h\varphi_k d\mu_n - \int_{\mathbb{R}} h\varphi_k d\mu \right| + M \int_{\mathbb{R}} (1 - \varphi_k) d\mu, \end{aligned}$$

car $\int_{\mathbb{R}} (1 - \varphi_k) d\mu_n - \int_{\mathbb{R}} (1 - \varphi_k) d\mu = \int_{\mathbb{R}} \varphi_k d\mu - \int_{\mathbb{R}} \varphi_k d\mu_n$, puisque, μ et μ_n étant des probabilités, $\int_{\mathbb{R}} \mathbf{1} d\mu = \int_{\mathbb{R}} \mathbf{1} d\mu_n = 1$. Or les fonctions φ_k et $h\varphi_k$ sont continues à support compact. Comme, par hypothèse, pour toute fonction f continue et à support compact sur \mathbb{R} , la suite de réels $\left(\int_{\mathbb{R}} f d\mu_n \right)_{\mathbb{N}}$ converge dans \mathbb{R} vers $\int_{\mathbb{R}} f d\mu$, on a, pour tout entier k fixé,

$$\lim_{n \rightarrow \infty} \int_{\mathbb{R}} \varphi_k d\mu_n = \int_{\mathbb{R}} \varphi_k d\mu, \text{ ou encore } \lim_{n \rightarrow \infty} \left| \int_{\mathbb{R}} \varphi_k d\mu_n - \int_{\mathbb{R}} \varphi_k d\mu \right| = 0. \text{ Pour la même raison,}$$

$$\text{pour tout entier } k \text{ fixé, } \lim_{n \rightarrow \infty} \left| \int_{\mathbb{R}} (h\varphi_k) d\mu_n - \int_{\mathbb{R}} h\varphi_k d\mu \right| = 0.$$

◇ Soit $\varepsilon > 0$ donné. Alors, à condition de choisir et fixer k suffisamment grand, puis de prendre n suffisamment grand, on pourra avoir $\left| \int_{\mathbb{R}} h d\mu_n - \int_{\mathbb{R}} h d\mu \right| \leq \varepsilon$. Ce qui prouve que, pour tout $h \in C_b(\mathbb{R})$, la suite de réels $\left(\int_{\mathbb{R}} h d\mu_n \right)_{\mathbb{N}}$ converge dans \mathbb{R} vers $\int_{\mathbb{R}} h d\mu$, et donc que la suite de probabilités $(\mu_n)_{\mathbb{N}}$ converge étroitement vers la probabilité μ . □

On peut remplacer dans la proposition précédente les fonctions continues à support compact par les fonctions continues "nulles à l'infini".

Définition 7.4.

Une application f de \mathbb{R}^d dans \mathbb{R} est dite **nulle à l'infini** si, pour tout $\varepsilon > 0$, il existe un compact $K \subseteq \mathbb{R}^d$ tel que, pour tout $x \in K^c$, $|f(x)| \leq \varepsilon$.

On note souvent $C_0(\mathbb{R}^d)$ l'ensemble des applications de \mathbb{R}^d dans \mathbb{R} continues et nulles à l'infini.

On obtient alors le critère suivant :

Proposition 7.20.

Critère de convergence étroite par les fonctions continues nulles à l'infini

Soient $(\mu_n)_\mathbb{N}$ une suite de probabilités sur \mathbb{R}^d et μ une probabilité sur \mathbb{R}^d . La suite de probabilités $(\mu_n)_\mathbb{N}$ converge étroitement vers la probabilité μ si, et seulement si, pour toute fonction f de \mathbb{R}^d dans \mathbb{R} , continue sur \mathbb{R}^d et nulle à l'infini, la suite de réels $(\int_{\mathbb{R}^d} f d\mu_n)_\mathbb{N}$ converge dans \mathbb{R} vers $\int_{\mathbb{R}^d} f d\mu$.

Démonstration : Raisonnons toujours en dimension $d = 1$. On peut reprendre mot pour mot la démonstration de la proposition 7.19 en remplaçant l'expression "à support compact" par "nulle à l'infini", car on vérifie en reprenant ses notations que, pour tout k , les fonctions φ_k et $h\varphi_k$ sont continues et nulles à l'infini. \square

Dans le cas des probabilités à densité on dispose de la condition suffisante (mais non nécessaire) de convergence étroite suivante :

Proposition 7.21.

Théorème de Scheffé

Soient, pour tout $n \in \mathbb{N}$, μ_n et μ des probabilités absolument continues sur \mathbb{R} de densités respectives f_n et f par rapport à la mesure de Lebesgue λ . Si la suite des densités $(f_n)_\mathbb{N}$ converge λ -presque-partout vers la densité f , alors la suite de probabilités $(\mu_n)_\mathbb{N}$ converge étroitement vers μ .

La réciproque est fausse.

Démonstration : Comme f et f_n sont des densités, pour tout $n \in \mathbb{N}$, $\int_{\mathbb{R}} (f - f_n) d\lambda = 0$. Donc

$$\int_{\mathbb{R}} (f - f_n)^+ d\lambda = \int_{\mathbb{R}} (f - f_n)^- d\lambda = \frac{1}{2} \int_{\mathbb{R}} |f - f_n| d\lambda$$

où on utilise que, pour tout $x \in \mathbb{R}$, $x = x^+ - x^-$ et $|x| = x^+ + x^-$. La suite $((f - f_n)^+)_\mathbb{N}$ est dominée par f et converge λ -presque-partout vers 0 car l'application $x \mapsto x^+$ est continue.

D'après le théorème de convergence dominée de Lebesgue, on en déduit que $\left(\int_{\mathbb{R}} (f - f_n)^+ d\lambda\right)_\mathbb{N}$

converge vers 0 et par suite $\left(\int_{\mathbb{R}} |f - f_n| d\lambda\right)_\mathbb{N}$ converge également vers 0. Enfin comme, pour

tout $t \in \mathbb{R}$, $\left|\int_{]-\infty, t]} (f - f_n) d\lambda\right| \leq \int_{\mathbb{R}} |f - f_n| d\lambda$, la suite $\left(\int_{]-\infty, t]} (f - f_n) d\lambda\right)_\mathbb{N}$ tend vers 0

quand n tend vers l'infini. On a montré que, pour tout $t \in \mathbb{R}$, la suite $(F_n(t))_\mathbb{N}$ converge vers $F(t)$ où F_n est la fonction de répartition de μ_n et F celle de μ . D'où la convergence étroite de $(\mu_n)_\mathbb{N}$ vers μ . \square

La réciproque est fausse comme le prouve l'exercice suivant :

Exercice 7.6. (Corrigé de l'exercice : page 198)

En considérant, pour tout entier $n \geq 1$, les applications définies, pour tout réel x , par

$f_n(x) := [1 - \cos(2\pi nx)]\mathbb{1}_{[0,1]}(x)$, montrer qu'il existe une suite de probabilités qui converge étroitement vers une probabilité μ sans que la suite des densités associées converge λ -presque-partout vers la densité de μ .

Les fonctions caractéristiques (f.c.) sont aussi un outil extrêmement commode dans l'étude des convergences étroites, grâce au résultat suivant qu'on admettra (on pourra en trouver une démonstration dans [4], pages 179-181) :

Proposition 7.22.

Théorème de continuité de Paul Lévy

Soit $(\mu_n)_{\mathbb{N}}$ une suite de probabilités sur \mathbb{R}^d . La suite des f.c. $(\Phi_{\mu_n})_{\mathbb{N}}$ converge simplement sur \mathbb{R}^d vers une application φ de \mathbb{R}^d dans \mathbb{C} continue en 0 si, et seulement si, il existe une probabilité μ sur \mathbb{R}^d , de fonction caractéristique $\Phi_\mu = \varphi$, telle que la suite $(\mu_n)_{\mathbb{N}}$ converge étroitement vers μ .

L'hypothèse de la continuité de φ en 0 est essentielle. Dans la proposition 7.22, le fait que $(\mu_n)_{\mathbb{N}}$ converge vers une limite μ qui est une probabilité fait partie du résultat de la proposition, contrairement à la proposition 7.23 qui suit, où le fait que μ soit une probabilité fait partie des hypothèses de la proposition.

Le théorème de continuité de Lévy a pour corollaire un critère pratique de convergence étroite utilisant les fonctions caractéristiques.

Proposition 7.23.

Critère de convergence étroite par les f.c.

Soient $(\mu_n)_{\mathbb{N}}$ une suite de probabilités sur \mathbb{R}^d et μ une probabilité sur \mathbb{R}^d . La suite de probabilités $(\mu_n)_{\mathbb{N}}$ converge étroitement vers la probabilité μ si, et seulement si, la suite des f.c. $(\Phi_{\mu_n})_{\mathbb{N}}$ converge simplement sur \mathbb{R}^d vers la f.c. Φ_μ .

ou encore, énoncé avec les v.a.r. :

Proposition 7.24.

Critère de convergence en loi par les f.c.

Soient $(X_n)_{\mathbb{N}}$ une suite de v.a.r. et X une v.a.r.. La suite $(X_n)_{\mathbb{N}}$ converge en loi vers X si, et seulement si, la suite des f.c. $(\Phi_{X_n})_{\mathbb{N}}$ converge simplement sur \mathbb{R} vers la f.c. Φ_X .

Démonstration : • C.N. - Supposons que $(\mu_n)_{\mathbb{N}}$ converge étroitement vers la probabilité μ . Les fonctions $e_u : x \in \mathbb{R}^d \mapsto e^{i\langle x, u \rangle} \in \mathbb{C}$, où $u \in \mathbb{R}^d$, sont continues et bornées sur \mathbb{R}^d . Par application du critère de convergence étroite par les fonctions continues et bornées (proposition 7.16) appliqué aux parties réelles et imaginaires des fonctions e_u , on en conclut que, pour tout $u \in \mathbb{R}^d$, la suite de nombres complexes $\left(\int_{\mathbb{R}^d} e^{i\langle x, u \rangle} d\mu_n(x) \right)_{\mathbb{N}}$ converge dans \mathbb{C} vers

$\int_{\mathbb{R}^d} e^{i\langle x, u \rangle} d\mu(x)$, donc la suite des f.c. $(\Phi_{\mu_n})_{\mathbb{N}}$ converge simplement sur \mathbb{R}^d vers la f.c. Φ_μ .

• C.S. - Supposons que la suite des f.c. $(\Phi_{\mu_n})_{\mathbb{N}}$ converge simplement sur \mathbb{R}^d vers la f.c. Φ_μ . Prenons, avec les notations du théorème de continuité de Lévy, $\varphi = \Phi_\mu$. Alors, la suite des f.c. $(\Phi_{\mu_n})_{\mathbb{N}}$ converge simplement sur \mathbb{R}^d vers une application φ de \mathbb{R}^d dans \mathbb{C} continue en 0. Par le théorème de continuité de Lévy, on en conclut qu'il existe une probabilité ν sur \mathbb{R}^d , de fonction caractéristique $\Phi_\nu = \varphi$, et que la suite $(\mu_n)_{\mathbb{N}}$ converge étroitement vers ν . Comme

$\Phi_\nu = \varphi = \Phi_\mu$, on conclut par le théorème d'injectivité des f.c. que $\mu = \nu$, et que la suite $(\mu_n)_\mathbb{N}$ converge bien étroitement vers μ . \square

Exemples 7.2.

Si $(a_n)_\mathbb{N}$ et $(\sigma_n)_\mathbb{N}$ sont deux suites réelles convergeant respectivement vers les réels a et σ , alors la suite de probabilités $(\mathcal{N}_1(a_n, \sigma_n^2))_\mathbb{N}$ converge étroitement vers la probabilité $\mathcal{N}_1(a, \sigma^2)$. \square

Exercice 7.7. (Corrigé de l'exercice : page 199)

Démontrer l'affirmation de l'exemple 7.2 précédent. Pour cela utiliser la fonction caractéristique de la loi $\mathcal{N}_1(a_n, \sigma_n^2)$, faire tendre n vers l'infini et conclure en appliquant le critère des fonctions caractéristiques pour la convergence étroite des probabilités.

Utilisons le critère de convergence étroite par les f.c. pour donner une démonstration du théorème-limite central que nous énonçons maintenant dans le cadre vectoriel qui généralise l'énoncé donné dans la proposition 7.12 en dimension $d = 1$:

Proposition 7.25.

Théorème-limite central (version vectorielle)

Soient $(X_n)_\mathbb{N}$ une suite indépendante et identiquement distribuée de vecteurs aléatoires réels de dimension d , de carré intégrable (i.e. $\mathbb{E}(|X_0|^2) < +\infty$), d'espérance $m \in \mathbb{R}^d$ et de matrice de dispersion D . On pose, pour tout entier naturel non nul n , $S_n = X_1 + X_2 + \dots + X_n$.

Alors, la suite des lois des vecteurs aléatoires $\left(\frac{S_n - nm}{\sqrt{n}}\right)_{\mathbb{N}^*}$ converge étroitement vers la loi gaussienne de dimension d , $\mathcal{N}_d(0, D)$.

Démonstration : Nous allons démontrer le TLC dans le cas $d = 1$.

Soit $(X_n)_\mathbb{N}$ une suite indépendante et identiquement distribuée de variables aléatoires réelles de carré intégrable, d'espérance m et de variance σ^2 .

• Considérons d'abord le cas $\sigma^2 = 0$. Les v.a.r. sont alors déterministes et égales à la constante m . Dans ce cas, pour tout entier non nul n , $S_n = nm$ et $\frac{S_n - nm}{\sqrt{n}} = 0$. La suite des v.a.r.

$\left(\frac{S_n - nm}{\sqrt{n}}\right)_{\mathbb{N}^*}$ est la suite stationnaire nulle. Elle converge étroitement vers la loi gaussienne dégénérée de dimension $\mathcal{N}_1(0, 0) = \delta_0$. Ce qui prouve le théorème dans le cas $\sigma^2 = 0$.

• Plaçons-nous maintenant dans le cas $\sigma^2 > 0$.

◇ Soit Φ la fonction caractéristique de la v.a.r. $X_1 - m$. D'après la proposition 3.33, page 57, comme les variables sont de carré intégrable, la f.c. Φ est de classe C^2 . De plus, on a $\Phi(0) = 1$, $\Phi'(0) = \mathbb{E}(X_1 - m) = 0$ et $\Phi''(0) = i^2 \mathbb{E}[(X_1 - m)^2] = -\text{Var}(X_1) = -\sigma^2$. La fonction Φ admet un développement limité en 0 à l'ordre 2 donné, pour tout réel t , par $\Phi(t) = 1 - \frac{\sigma^2}{2} t^2 + t^2 \varepsilon(t)$ avec $\lim_{t \rightarrow 0} \varepsilon(t) = 0$. Donc, la fonction caractéristique de la v.a.r. $\frac{S_n - nm}{\sqrt{n}}$ est l'application Φ_n définie, pour tout réel t , par

$$\Phi_n(t) = \mathbb{E} \left[e^{i \langle \frac{S_n - nm}{\sqrt{n}}, t \rangle} \right] = \mathbb{E} \left[e^{i \langle S_n - nm, \frac{t}{\sqrt{n}} \rangle} \right] = \mathbb{E} \left[e^{i \langle \sum_{k=1}^n X_k, \frac{t}{\sqrt{n}} \rangle} \right].$$

D'où $\phi_n(t) = \mathbb{E} \left[e^{i \langle X_1, \frac{1}{\sqrt{n}} \rangle} \right]^n = \left[\phi \left(\frac{t}{\sqrt{n}} \right) \right]^n$, ou encore $\phi_n(t) = \left[1 - \frac{\sigma^2}{2n} t^2 + \frac{t^2}{n} \varepsilon \left(\frac{t}{\sqrt{n}} \right) \right]^n$.

◇ Il ne reste plus qu'à montrer que, pour tout réel t , $\lim_{n \rightarrow +\infty} \phi_n(t) = e^{-\frac{\sigma^2 t^2}{2}}$.

En effet, d'après la formule du binôme de Newton,

$$\left[1 - \frac{\sigma^2}{2n} t^2 + \frac{t^2}{n} \varepsilon \left(\frac{t}{\sqrt{n}} \right) \right]^n = \sum_{k=0}^{k=n} \frac{C_n^k t^{2k}}{n^k} \left[-\frac{\sigma^2}{2} + \varepsilon \left(\frac{t}{\sqrt{n}} \right) \right]^k.$$

Considérons la série numérique $\sum_k u_k(n)$ de terme général : $u_k(n) = 0$ si $k > n$, et

$$u_k(n) = \frac{C_n^k t^{2k}}{n^k} \left[-\frac{\sigma^2}{2} + \varepsilon \left(\frac{t}{\sqrt{n}} \right) \right]^k \text{ si } 0 \leq k \leq n. \text{ Pour tout réel } t \text{ fixé, il existe une}$$

constante M telle que, pour tout entier naturel non nul n , $|u_k(n)| \leq \frac{M^k}{k!}$. Donc la série

$\sum_k u_k(n)$ converge normalement. On peut donc intervertir les symboles $\lim_{n \rightarrow +\infty}$ et $\sum_{k=0}^{+\infty}$, on ob-

tient $\lim_{n \rightarrow +\infty} \sum_{k=0}^{+\infty} u_k(n) = \sum_{k=0}^{+\infty} \lim_{n \rightarrow +\infty} u_k(n) = \sum_{k=0}^{+\infty} \frac{-\sigma^2 t^{2k}}{2(k!)} = e^{-\frac{\sigma^2 t^2}{2}}$. Donc la suite des fonctions

caractéristiques des v.a.r. $\frac{S_n - nm}{\sqrt{n}}$ converge vers la fonction caractéristique de la loi normale centrée de variance σ^2 . Ce qui prouve, en vertu du critère de convergence en loi par les fonctions caractéristiques (proposition 7.24, page 137) que la suite des v.a.r. $\frac{S_n - nm}{\sqrt{n}}$ converge en loi vers une v.a.r. normale centrée de variance σ^2 . □

Pour compléter l'étude des liens, commencée au chapitre précédent, entre les divers modes de convergences, signalons le résultat suivant qui prouve que la convergence en probabilité, et a fortiori la convergence presque-sûre, d'une suite de variables aléatoires réelles implique la convergence de la suite des lois de ces v.a.r. :

Proposition 7.26.

Si $(X_n)_{\mathbb{N}}$ est une suite de variables aléatoires réelles convergeant en probabilité vers la variable aléatoire réelle Y , alors la suite des lois $(\mathbb{P}_{X_n})_{\mathbb{N}}$ converge vers la loi \mathbb{P}_Y de la variable aléatoire réelle Y .

La réciproque est fausse.

Démonstration : Soit h une fonction numérique positive définie sur \mathbb{R}^d , continue et à support compact. La fonction h est donc uniformément continue sur \mathbb{R}^d . Fixons $\varepsilon > 0$. Il existe donc un réel $\eta_\varepsilon > 0$, tel que, pour tout $x, y \in \mathbb{R}^d$, $|x - y| \leq \eta_\varepsilon$ implique $|h(x) - h(y)| \leq \varepsilon$.

Comme la suite $(X_n)_{\mathbb{N}}$ converge en probabilité vers la variable aléatoire réelle Y , il existe un entier naturel N_ε tel que, pour tout entier $n \geq N_\varepsilon$, $\mathbb{P}(|X_n - Y| > \eta_\varepsilon) \leq \varepsilon$. Alors, pour tout entier naturel $n \geq N_\varepsilon$, on a

$$\begin{aligned}
|\mathbb{E}[h(X_n) - h(Y)]| &\leq \mathbb{E}[|h(X_n) - h(Y)|] \\
&= \mathbb{E}[|h(X_n) - h(Y)| \mathbf{1}_{\{|X_n - Y| \leq \eta_\varepsilon\}}] + \mathbb{E}[|h(X_n) - h(Y)| \mathbf{1}_{\{|X_n - Y| > \eta_\varepsilon\}}] \\
&\leq \varepsilon \mathbb{E}[\mathbf{1}_{\{|X_n - Y| \leq \eta_\varepsilon\}}] + 2\|h\|_\infty \mathbb{E}[\mathbf{1}_{\{|X_n - Y| > \eta_\varepsilon\}}] \\
&\leq \varepsilon \mathbb{E}[\mathbf{1}_{\{|X_n - Y| \leq \eta_\varepsilon\}}] + 2\|h\|_\infty \mathbb{P}(|X_n - Y| > \eta_\varepsilon) \\
&\leq \varepsilon + 2\|h\|_\infty \varepsilon = (1 + 2\|h\|_\infty)\varepsilon.
\end{aligned}$$

Ce qui prouve que $\lim_{n \rightarrow \infty} \mathbb{E}[h(X_n)] = \mathbb{E}[h(Y)]$, pour toute fonction numérique h positive définie sur \mathbb{R}^d , continue et à support compact. On conclut alors en vertu du critère de convergence étroite par les fonctions continues à support compact. \square

La convergence des lois d'une suite de variables aléatoires réelles n'implique pas nécessairement la convergence en probabilité de la suite de v.a.r., cependant cette conclusion devient vraie si la suite des lois converge vers une probabilité de Dirac :

Proposition 7.27.

Si $(X_n)_{n \in \mathbb{N}}$ est une suite de variables aléatoires réelles sur un même espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$ telle que la suite des lois $(\mathbb{P}_{X_n})_{n \in \mathbb{N}}$ converge vers la probabilité de Dirac δ_a où a est un réel, alors la suite de variables aléatoires réelles $(X_n)_{n \in \mathbb{N}}$ converge en probabilité vers la variable aléatoire réelle constante a .

Résultat qu'on peut aussi énoncer :

Proposition 7.28.

Une suite de v.a.r. $(X_n)_{n \in \mathbb{N}}$ converge en loi vers une constante $a \in \mathbb{R}$ si, et seulement si, elle converge en probabilité vers a .

Démonstration : Soit $\varepsilon > 0$ et f_ε l'application de \mathbb{R} dans \mathbb{R} définie par

$$f_\varepsilon(x) := \mathbf{1}_{]a-\varepsilon, a+\varepsilon[}(x) + \frac{1}{\varepsilon}|x - a|\mathbf{1}_{]a-\varepsilon, a+\varepsilon[}(x).$$

f_ε est continue bornée donc la suite $(\mathbb{E}[f_\varepsilon(X_n)])_{n \geq 1}$ converge vers $\mathbb{E}[f_\varepsilon(a)] = 0$, d'après la proposition 7.17. Comme

$$\mathbb{P}(|X_n - a| \geq \varepsilon) = \mathbb{E}[\mathbf{1}_{]a-\varepsilon, a+\varepsilon[}(X_n)] \leq \mathbb{E}[f_\varepsilon(X_n)],$$

on conclut que la suite $(X_n)_{n \geq 1}$ converge en probabilité vers a . On remarque que, lorsque la limite est une v.a.r. presque-sûrement constante, il y a équivalence entre la convergence en loi et la convergence en probabilité. \square

Notons que, dans le cas général, on ne peut pas effectuer d'opérations élémentaires sur les limites-en-loi. Cependant on admettra (pour une démonstration de la première assertion on pourra se reporter à l'exercice 7.10, page 142) :

Proposition 7.29.

Théorème de Slutsky

Soient $(X_n)_{n \in \mathbb{N}}$ une suite de v.a.r. convergeant en loi vers une v.a.r. X et $(Y_n)_{n \in \mathbb{N}}$ une suite de v.a.r. convergeant en loi vers 0, alors

1. la suite de v.a.r. $(X_n + Y_n)_{n \in \mathbb{N}}$ converge en loi vers X ,
2. la suite de v.a.r. $(X_n Y_n)_{n \in \mathbb{N}}$ converge en loi (et aussi en probabilité) vers 0.

Pour terminer citons pour information deux théorèmes-limites particulièrement importants portant sur la convergence de lois, (que nous admettrons et qui sont hors programme) :

Proposition 7.30.

Loi des événements rares (Hors programme)

Soient $(X_k^{(n)})_{(k,n) \in \mathbb{N}^2}$ une suite de v.a.r. indépendantes à valeurs dans \mathbb{N} . On pose $p_k^{(n)} := \mathbb{P}(X_k^{(n)} = 1)$ et $\epsilon_k^{(n)} := \mathbb{P}(X_k^{(n)} \geq 2)$. On suppose de plus que

$$1. \lim_n (p_1^{(n)} + p_2^{(n)} + \dots + p_n^{(n)}) = \alpha \in]0, +\infty[,$$

$$2. \lim_n (\max(p_1^{(n)}, p_2^{(n)}, \dots, p_n^{(n)})) = 0,$$

$$3. \lim_n (\epsilon_1^{(n)} + \epsilon_2^{(n)} + \dots + \epsilon_n^{(n)}) = 0.$$

Alors la suite de v.a.r. $(X_1^{(n)} + X_2^{(n)} + \dots + X_n^{(n)})_{n \in \mathbb{N}^*}$ converge en loi vers une v.a.r. de loi $\mathcal{P}(\alpha)$.

En remarquant qu'une v.a.r. de loi binomiale $\mathcal{B}(n, p)$ a même loi que la somme de n v.a.r. indépendantes de loi $\mathcal{B}(p)$, la proposition 7.14 est un cas particulier de la loi des événements rares (cf. [3], exercice V-16).

Notons que les théorèmes-limites jouent un rôle théorique important en statistique dans la vérification des modèles probabilistes de phénomènes aléatoires. En particulier celui-ci (cf. [3], problème V-1) :

Proposition 7.31.

Théorème fondamental de la statistique (Hors programme)

Soit $(X_n)_{n \in \mathbb{N}}$ une suite de v.a.r. indépendantes et de même loi. Alors, pour \mathbb{P} -presque-tout $\omega \in \Omega$, la suite des lois empiriques de (X_1, X_2, \dots, X_n)

$$\left(\frac{1}{n} \sum_{k=1}^{k=n} \delta_{X_k(\omega)} \right)_{n \in \mathbb{N}^*}$$

converge étroitement vers la loi de X_1 .

7.3 Exercices de révision sur les chapitres I à VII

Exercice 7.8. (Corrigé de l'exercice : page 199)

Étudier la convergence étroite de la suite de probabilités $(\mu_n)_{n \geq 1}$ de densités respectives $(f_n)_{n \geq 1}$ où, pour tout $n \geq 1$, f_n est définie par $f_n(x) := nx^{n-1} \mathbb{1}_{[0,1]}(x)$.

Exercice 7.9. (Corrigé de l'exercice : page 199)

Soit $(X_n)_{n \geq 1}$ une suite indépendante de v.a.r. de même loi de Cauchy $\mathcal{C}(1)$ (Pour la définition, cf. formulaire de l'annexe A, page 205). Pour tout $n \geq 1$, on pose $S_n := \sum_{k=1}^{k=n} X_k$.

Étudier les convergences en probabilité et en loi des suites de v.a.r. $\left(\frac{1}{\sqrt{n}} S_n \right)_{n \geq 1}$, $\left(\frac{1}{n} S_n \right)_{n \geq 1}$

et $\left(\frac{1}{n^2}S_n\right)_{n \geq 1}$.

Exercice 7.10. (Corrigé de l'exercice : page 200)

Le but de cet exercice est de prouver l'item 1 du **Théorème de Slutsky** (cf. proposition 7.29, page 140).

Soient X une v.a.r., $(X_n)_\mathbb{N}$ et $(Y_n)_\mathbb{N}$ deux suites de v.a.r..

1. Montrer que, pour tous $t \in \mathbb{R}$, $\alpha > 0$, et $n \in \mathbb{N}$,

$$|\phi_{X_n+Y_n}(t) - \phi_{X_n}(t)| \leq 2\mathbb{P}(|Y_n| > \alpha) + \mathbb{E}[\mathbb{1}_{]-\infty, \alpha]}(|Y_n|)|e^{itY_n} - 1|]$$

où ϕ_Z désigne la fonction caractéristique de la v.a.r. Z .

2. Démontrer le **théorème de Slutsky** : Si $(X_n)_\mathbb{N}$ converge en loi vers X et $(Y_n)_\mathbb{N}$ converge en loi vers 0, alors la suite $(X_n + Y_n)_\mathbb{N}$ converge en loi vers X .
3. A l'aide d'un exemple, montrer que l'on a pas nécessairement $(X_n - X)_\mathbb{N}$ qui converge en loi vers 0.

Exercice 7.11. (Corrigé de l'exercice : page 200)

Soit $(U_k)_\mathbb{N}$ une suite indépendante de v.a.r. de loi normale centrée et de variance $\sigma^2 > 0$. Pour tout $\theta \in \mathbb{R}$, on définit la suite $(X_k)_\mathbb{N}$ par la relation de récurrence $X_n = \theta X_{n-1} + U_n$, pour tout $n \geq 1$, avec $X_0 = 0$.

1. Déterminer, pour tout $n \in \mathbb{N}$, la loi de la v.a.r. X_n .
2. Étudier la convergence en loi de la suite de v.a.r. $(X_k)_\mathbb{N}$.

Exercice 7.12. (Corrigé de l'exercice : page 201)

Soient X une v.a.r. et $(X_n)_\mathbb{N}$ une suite de v.a.r.. On suppose que, pour tout $n \in \mathbb{N}$, $\mathbb{P}_{X_n} := \delta_{x_n}$ où $x_n \in \mathbb{R}$.

1. Si $\mathbb{P}_X = \delta_x$, où $x \in \mathbb{R}$, montrer que la suite $(X_n)_\mathbb{N}$ converge en loi vers X si et seulement si la suite de réels $(x_n)_\mathbb{N}$ converge vers x .
2. Montrer que si la suite $(X_n)_\mathbb{N}$ converge en loi vers X , alors il existe $x \in \mathbb{R}$ tel que $\mathbb{P}_X = \delta_x$. On pourra utiliser le résultat de l'exercice 3.23, page 60.

Exercice 7.13. (Corrigé de l'exercice : page 201)

Dans cet exercice on se propose de démontrer la réciproque du résultat de l'exercice 7.7, page 138.

Soit $(\mu_n)_\mathbb{N}$ une suite de probabilités de Gauss où, pour tout $n \in \mathbb{N}$, $\mu_n := \mathcal{N}_1(a_n, \sigma_n^2)$. On suppose que cette suite converge étroitement vers une probabilité μ . On se propose de montrer qu'alors les suites réelles $(a_n)_\mathbb{N}$ et $(\sigma_n^2)_\mathbb{N}$ convergent respectivement vers les réels a et σ^2 et que $\mu = \mathcal{N}_1(a, \sigma^2)$ (avec la convention $\mathcal{N}_1(a, 0) = \delta_a$).

1. En utilisant les fonctions caractéristiques, montrer que la suite $(\sigma_n^2)_\mathbb{N}$ est bornée et qu'elle admet une seule valeur d'adhérence dans \mathbb{R} . En déduire qu'elle converge dans \mathbb{R} , on notera σ^2 sa limite.

-
2. A l'aide du théorème de continuité de Lévy, montrer que la suite $(\delta_{a_n})_{\mathbb{N}}$ converge étroitement vers la probabilité de Dirac en un point a . En déduire que la suite $(a_n)_{\mathbb{N}}$ converge vers a et que $\mu = \mathcal{N}_1(a, \sigma^2)$.

Chapitre 8

Corrigés des exercices

8.1 Corrigés des exercices du chapitre I

Corrigé de l'exercice 1.1, page 2

1. S'il existe $x \in \{f \in \emptyset\}$, alors $f(x)$ appartient à \emptyset ce qui est absurde. Donc $\{f \in \emptyset\} = \emptyset$.
2. Soit $x \in \{f \in A\}$, alors $f(x) \in A$ et comme $A \subset B$, $f(x) \in B$ et donc $x \in \{f \in B\}$.
On a donc montrer que tout élément x de l'ensemble $\{f \in A\}$ est aussi dans $\{f \in B\}$ ce qui signifie que $\{f \in A\} \subset \{f \in B\}$.
3. On va raisonner par équivalence¹. On a

$$\begin{aligned} x \in \{f \in \cup_{i \in I} A_i\} &\Leftrightarrow f(x) \in \cup_{i \in I} A_i \Leftrightarrow \exists i_0 ; f(x) \in A_{i_0} \Leftrightarrow \exists i_0 ; x \in \{f \in A_{i_0}\} \\ &\Leftrightarrow x \in \cup_{i \in I} \{f \in A_i\} . \end{aligned}$$

D'où l'égalité. De même on écrit

$$\begin{aligned} x \in \{f \in \cap_{i \in I} A_i\} &\Leftrightarrow f(x) \in \cap_{i \in I} A_i \Leftrightarrow \forall i ; f(x) \in A_i \Leftrightarrow \forall i ; x \in \{f \in A_i\} \\ &\Leftrightarrow x \in \cap_{i \in I} \{f \in A_i\} . \end{aligned}$$

4. On procède encore par équivalence :

$$x \in \{f \in A\}^c \Leftrightarrow x \notin \{f \in A\} \Leftrightarrow f(x) \notin A \Leftrightarrow f(x) \in A^c \Leftrightarrow x \in \{f \in A^c\} .$$

Corrigé de l'exercice 1.2, page 2

1. Dans les deux cas (a) et (b) on a $\mathbb{1}_{A \cap B} = \mathbb{1}_A \cdot \mathbb{1}_B$. Par contre $\mathbb{1}_{A \cup B}$ s'exprime différemment suivant que A et B sont disjoints ou non. On peut vérifier cela en étudiant toutes les valeurs prises par ces fonctions

	$\mathbb{1}_A(x)$	$\mathbb{1}_B(x)$	$\mathbb{1}_A(x) \cdot \mathbb{1}_B(x)$	$\mathbb{1}_{A \cup B}(x)$ cas (a)	$\mathbb{1}_{A \cup B}$ cas (b)
$x \in (A \cup B)^c$	0	0	0	0	0
$x \in A \cap B^c$	1	0	0	1	1
$x \in A^c \cap B$	0	1	0	1	1
$x \in A \cap B$	1	1	1	1	0

¹Si vous n'êtes pas sûr de vous lors de l'écriture d'une équivalence, vérifiez rapidement les deux implications pour vous en persuader

Ainsi dans le cas (a) : $\mathbb{1}_{A \cup B} = \mathbb{1}_A + \mathbb{1}_B$
 et dans le cas (b) : $\mathbb{1}^{A \cup B} = \mathbb{1}_A + \mathbb{1}_B - \mathbb{1}_A \cdot \mathbb{1}_B$.

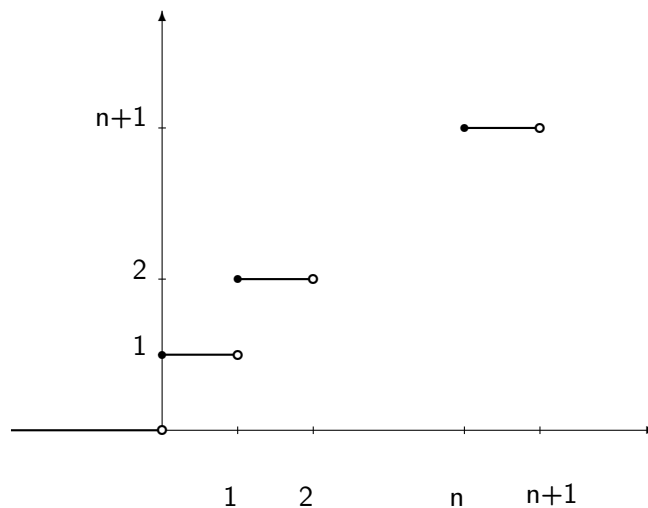
2. On a trivialement $\mathbb{1}_{A^c} = 1 - \mathbb{1}_A$ et $\mathbb{1}_{A \setminus B} = \mathbb{1}_{A \cap B^c} = \mathbb{1}_A \cdot \mathbb{1}_{B^c} = \mathbb{1}_A(1 - \mathbb{1}_B)$. On remarque que si $B \subseteq A$, $\mathbb{1}_A \cdot \mathbb{1}_B = \mathbb{1}_B$ et dans ce cas $\mathbb{1}_{A \setminus B} = \mathbb{1}_A - \mathbb{1}_B$.

Enfin $\mathbb{1}_{A \cup B \cup C} = \mathbb{1}_{(A \cup B) \cup C} = \mathbb{1}_{A \cup B} + \mathbb{1}_C - \mathbb{1}_C \mathbb{1}_{A \cup B}$. On développe de même l'indicatrice de $A \cup B$ et on obtient :

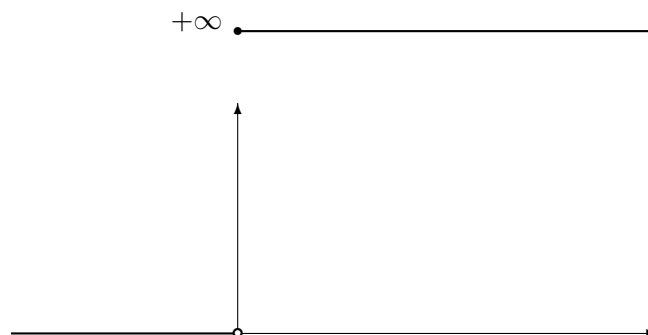
$$\mathbb{1}_{A \cup B \cup C} = \mathbb{1}_A + \mathbb{1}_B + \mathbb{1}_C - (\mathbb{1}_A \mathbb{1}_B + \mathbb{1}_B \mathbb{1}_C + \mathbb{1}_C \mathbb{1}_A) + (\mathbb{1}_A \mathbb{1}_B \mathbb{1}_C)$$

Corrigé de l'exercice 1.3, page 2

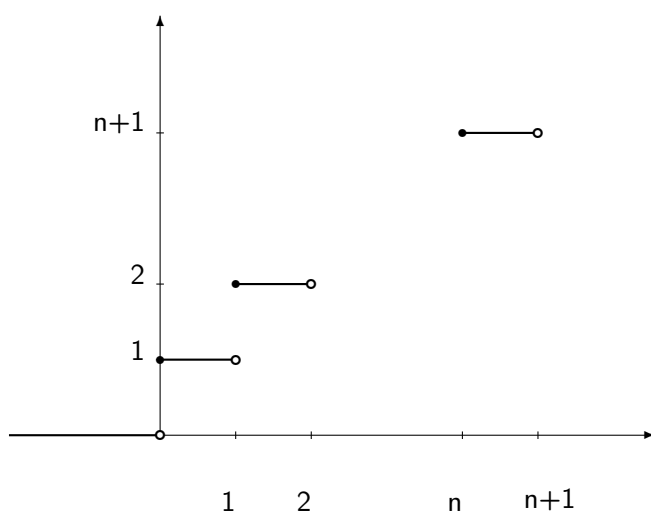
1. Représentation graphique de la fonction $\sum_{n \geq 0} \mathbb{1}_{[n, +\infty[}$:



2. Représentation graphique de la fonction $\sum_{n \geq 0} \mathbb{1}_{[0, n]}$:



3. Représentation graphique de la fonction $\sum_{n \geq 0} (n+1) \mathbb{1}_{[n, n+1]}$:



Corrigé de l'exercice 1.4, page 3

Cet exercice n'est pas difficile mais demande de la rigueur lors de sa rédaction. Il faut démontrer les trois axiomes qui feront de \mathcal{A} une tribu.

i) Tout d'abord $E \in \mathcal{A}$ car $E = A_1 \cup \dots \cup A_n$ par définition d'une partition. Donc E s'écrit bien comme $\cup_{i \in I} A_i$ en choisissant $I = \{1, \dots, n\}$.

ii) Soit $B \in \mathcal{A}$. Montrons que $B^c \in \mathcal{A}$. On a :

$$B \in \mathcal{A} \Leftrightarrow \exists I \subset \{1, \dots, n\} ; B = \cup_{i \in I} A_i$$

et comme les A_1, \dots, A_n forment une partition de E on a $B^c = \cup_{j \in J} A_j$ où J est le complémentaire de I dans $\{1, \dots, n\}$. Donc $B^c \in \mathcal{A}$.

iii) Soit maintenant $(B_k)_{k \in \mathbb{N}}$ une suite d'éléments de \mathcal{A} . Pour tout $k \in \mathbb{N}$, il existe un sous-ensemble $I_k \subset \{1, \dots, n\}$ tel que $B_k = \cup_{i \in I_k} A_i$. Par suite

$$\cup_{k \in \mathbb{N}} B_k = \cup_{k \in \mathbb{N}} (\cup_{i \in I_k} A_i) = \cup_{j \in J} A_j$$

où $J = \cup_{k \in \mathbb{N}} I_k \subset \{1, \dots, n\}$. Ainsi $\cup_{k \in \mathbb{N}} B_k$ est bien la réunion d'une sous-famille des A_1, \dots, A_n et donc $\cup_{k \in \mathbb{N}} B_k \in \mathcal{A}$.

Corrigé de l'exercice 1.5, page 3

1) On commence par considérer une famille quelconque (finie ou infinie) de tribu : $(\mathcal{A}_i)_{i \in I}$. On considère \mathcal{B} la famille des parties de E communes à toutes les tribus \mathcal{A}_i pour $i \in I$. On dit que \mathcal{B} est l'intersection des tribus \mathcal{A}_i et on écrit $\mathcal{B} = \cap_{i \in I} \mathcal{A}_i$. On va montrer que \mathcal{B} est elle-même une tribu.

- $E \in \mathcal{A}_i$ pour tout i car les \mathcal{A}_i sont des tribus et donc $E \in \mathcal{B}$.
- Soit $A \in \mathcal{B}$. Alors $A \in \mathcal{A}_i$ pour tout i et donc $A^c \in \mathcal{A}_i$ pour tout i . Donc pour tout $A \in \mathcal{B}$.
- Soit $(A_k)_{k \in \mathbb{N}}$ une suite d'éléments de \mathcal{B} . On a $\forall k \in \mathbb{N}, \forall i \in I, A_k \in \mathcal{A}_i$. Donc comme \mathcal{A}_i est une tribu, pour tout $i \in I, \cup_{k \in \mathbb{N}} A_k \in \mathcal{A}_i$. Par suite $\cup_{k \in \mathbb{N}} A_k \in \mathcal{B}$.

Ainsi \mathcal{B} est une tribu sur E .

2) Prouvons par un contre-exemple que la réunion d'une famille de tribu sur E n'est pas nécessairement une tribu sur E . En effet prenons $E = \mathbb{R}$, $A = \{1\}$ et $B = \{2\}$ deux parties de E . On vérifie facilement que la famille à quatre éléments $\mathcal{I} = \{\emptyset, A, A^c, E\}$ est une tribu sur E . Il

en est de même pour la famille $\mathcal{B} = \{\emptyset, B, B^c, E\}$. Considérons la réunion $\mathcal{C} = \mathcal{A} \cup \mathcal{B}$. On a $\mathcal{C} = \{\emptyset, A, A^c, B, B^c, E\}$. Ce n'est pas une tribu car on n'a pas $A \cup B \in \mathcal{C}$.

Corrigé de l'exercice 1.6, page 4

A et B mesurables relativement à la tribu \mathcal{A} signifie juste que $A \in \mathcal{A}$ et $B \in \mathcal{A}$. Comme \mathcal{A} est une tribu, $B^c \in \mathcal{A}$ et donc $A \setminus B = A \cap B^c$ est l'intersection de deux éléments de la tribu \mathcal{A} et donc c'est aussi un élément de \mathcal{A} . Donc $A \setminus B$ est mesurable par rapport à \mathcal{A} .

Corrigé de l'exercice 1.7, page 5

On note \mathcal{F} la famille des tribus sur \mathbb{R} contenant tous les intervalles de la forme $]a, b]$ où a et b sont des réels tels que $a < b$. Notons \mathcal{B} la tribu obtenue par intersection de toutes les tribus de la famille \mathcal{F} . \mathcal{B} est une tribu d'après le résultat de l'exercice 1.5. On remarque qu'on a :

- \mathcal{B} est une tribu,
- \mathcal{B} contient tous les intervalles de la forme $]a, b]$ où a et b sont des réels tels que $a < b$. \mathcal{B} appartient donc à la famille \mathcal{F} définie plus haut.
- \mathcal{B} est la plus petite, au sens de l'inclusion, des tribus de la famille \mathcal{F} . Cela signifie que si \mathcal{A} est une tribu sur \mathbb{R} appartenant à la famille \mathcal{F} , alors $\mathcal{B} \subseteq \mathcal{A}$, car \mathcal{B} , qui est l'intersection des tribus de la famille \mathcal{F} , est contenue dans toutes les tribus de la famille \mathcal{F} et en particulier dans la tribu \mathcal{A} .

Des trois points ci-dessus, on déduit que $\mathcal{B} = \mathcal{B}(\mathbb{R})$ est la tribu de Borel sur \mathbb{R} .

Corrigé de l'exercice 1.8, page 5

Il faut montrer que si \mathcal{B} est une tribu contenant les parties A_1, \dots, A_n , alors $\mathcal{A} \subseteq \mathcal{B}$. Soit \mathcal{B} une telle tribu. Par définition et propriétés des tribus, elle contient toutes les réunions des sous-familles de A_1, \dots, A_n c'est-à-dire que pour tout sous-ensemble I de $\{1, \dots, n\}$, $\bigcup_{i \in I} A_i \in \mathcal{B}$. Donc \mathcal{B} contient tous les éléments de \mathcal{A} et on a donc bien $\mathcal{A} \subseteq \mathcal{B}$.

Corrigé de l'exercice 1.9, page 6

1. Soit A_1, \dots, A_n une suite finie de parties de \mathbb{N} deux à deux disjointes. Deux cas se présentent :

- Premier cas : $0 \in \bigcup_1^n A_i$. Alors $\mu(\bigcup_1^n A_i) = +\infty$ et $\sum_1^n \mu(A_i) = +\infty$ car il existe un A_{i_0} contenant 0 et donc de mesure infinie. Donc $\mu(\bigcup_1^n A_i) = \sum_1^n \mu(A_i)$.
- Second cas : $0 \notin \bigcup_1^n A_i$. Alors si $\bigcup_1^n A_i$ est fini, tous les A_i sont finis et

$$\mu\left(\bigcup_1^n A_i\right) = \sum_{k \in \bigcup_1^n A_i} \frac{1}{k^2} = \sum_1^n \left(\sum_{k \in A_i} \frac{1}{k^2}\right) = \sum_1^n \mu(A_i).$$

Si $\bigcup_1^n A_i$ est infini, alors il existe un A_{i_0} infini et on a $\mu(\bigcup_1^n A_i) = +\infty = \sum_1^n \mu(A_i)$.
L'application μ est donc additive.

2. Cependant que μ n'est pas σ -additive (et donc ce n'est pas une mesure). En effet, considérons \mathbb{N}^* , $\mu(\mathbb{N}^*) = +\infty$ et $\mathbb{N}^* = \bigcup_1^{+\infty} \{k\}$. Comme la suite $(\{k\})_{k \in \mathbb{N}^*}$ est formée de parties deux à deux disjointes on a $\sum_{k \geq 1} \mu(\{k\}) = \sum_{k \geq 1} \frac{1}{k^2} < +\infty$. Ainsi

$$\mu\left(\bigcup_{k \geq 1} \{k\}\right) \neq \sum_{k \geq 1} \mu(\{k\}).$$

En conséquence, l'application μ n'est pas σ -additive bien qu'elle soit additive.

Corrigé de l'exercice 1.10, page 7

On vérifie aisément que $\mathbb{R} = \bigcup_{k \in \mathbb{Z}}]k, k+1]$. Les intervalles $]k, k+1]$ sont des boréliens disjoints deux à deux. Par σ -additivité de la mesure de Lebesgue λ sur \mathbb{R} , on obtient

$$\lambda(\mathbb{R}) = \lambda\left(\bigcup_{k \in \mathbb{Z}}]k, k+1]\right) = \sum_{k=-\infty}^{k=+\infty} \lambda(]k, k+1]) .$$

Or pour tout $k \in \mathbb{Z}$, $\lambda(]k, k+1]) = 1$ (la longueur de l'intervalle $]k, k+1]$), d'où $\lambda(\mathbb{R}) = \sum_{k=-\infty}^{+\infty} 1 = +\infty$.

Corrigé de l'exercice 1.11, page 8

On peut remarquer tout d'abord que si A est un borélien alors $\delta_a(A) = \mathbf{1}_{\{a\}}(A)$. Vérifions que δ_a est une mesure sur $\mathbb{R}, \mathcal{B}(\mathbb{R})$.

- C'est bien une application positive de $\mathcal{B}(\mathbb{R})$ dans $[0, +\infty]$.
- On a $\delta_a(\emptyset) = 0$.
- Soit $(A_n)_{n \in \mathbb{N}}$ une suite de boréliens deux à deux disjoints. Deux cas se présentent.
Tout d'abord si $a \in \bigcup_{n \in \mathbb{N}} A_n$ alors il existe un unique n_0 (les A_n sont disjoints deux à deux) tel que $a \in A_{n_0}$ et donc $\delta_a(\bigcup_{n \in \mathbb{N}} A_n) = 1$ et

$$\sum_{n=0}^{\infty} \delta_a(A_n) = \underbrace{\delta_a(A_0) + \dots + \delta_a(A_{n_0-1})}_{=0} + \underbrace{\delta_a(A_{n_0})}_{=1} + \underbrace{\sum_{k=n_0+1}^{\infty} \delta_a(A_k)}_{=0},$$

d'où l'égalité $\delta_a(\bigcup_{n \in \mathbb{N}} A_n) = \sum_{n \in \mathbb{N}} \delta_a(A_n)$.

Si maintenant $a \notin \bigcup_{n \in \mathbb{N}} A_n$, alors $\forall n \in \mathbb{N}$; $a \notin A_n$ et par suite $\delta_a(\bigcup_{n \in \mathbb{N}} A_n) = 0$ et $\sum_{n \in \mathbb{N}} \delta_a(A_n) = 0$ donc on a encore l'égalité.

Ainsi δ_a est une mesure et comme $\delta_a(\mathbb{R}) = 1$, c'est aussi une probabilité.

Corrigé de l'exercice 1.12, page 9

Avec les notations de la proposition 1.4, page 8, il suffit de prendre $\mu_k = \delta_k$, pour tout $k \in \mathbb{N}$ et, suivant le cas,

1. pour la probabilité binomiale :
 - $\alpha_k = C_n^k p^k (1-p)^{n-k}$ pour $0 \leq k \leq n$
 - $\alpha_k = 0$ pour $k \geq n+1$.
2. pour la probabilité de Poisson : $\alpha_k = e^{-\alpha} \frac{\alpha^k}{k!}$, pour tout $k \in \mathbb{N}$.
3. pour la probabilité géométrique : $\alpha_0 = 0$, et $\alpha_k = p(1-p)^{k-1}$ pour $k \geq 1$.
4. pour la probabilité uniforme-discrète :
 - $\alpha_k = \frac{1}{n}$ pour $1 \leq k \leq n$
 - $\alpha_0 = 0$ et $\alpha_k = 0$ pour $k \geq n+1$.

Corrigé de l'exercice 1.13, page 9

1) $\mathcal{B}(n; p)(\{i\}) = \sum_{k=0}^n C_n^k p^k (1-p)^{n-k} \delta_k(\{i\})$ où $\delta_k(\{i\}) = 1$ si $i = k$ et 0 sinon. On a donc $\mathcal{B}(n; p)(\{i\}) = C_n^i p^i (1-p)^{n-i}$.

De même pour la loi de Poisson on trouve $\mathcal{P}(\alpha)(\{i\}) = e^{-\alpha} \frac{\alpha^i}{i!}$.

2) On a par l'additivité des probabilités pour les ensembles deux à deux disjoints :

$$\mathcal{P}(1/10)(\{1, 3, 5, 7\}) = \mathcal{P}(1/10)(\{1\}) + \mathcal{P}(1/10)(\{3\}) + \mathcal{P}(1/10)(\{5\}) + \mathcal{P}(1/10)(\{7\}).$$

On trouve donc

$$\begin{aligned} \mathcal{P}(1/10)(\{1, 3, 5, 7\}) &= e^{-0,1} \left[\frac{(0,1)^1}{1!} + \frac{(0,1)^3}{3!} + \frac{(0,1)^5}{5!} + \frac{(0,1)^7}{7!} \right] \\ &\simeq 0,0905 + 0,0002 + 0 + 0 \simeq 0,0907. \end{aligned}$$

De même on trouve $\mathcal{B}(7; 0.3)(\{0, 3, 5\}) \simeq 0.3343$.

Corrigé de l'exercice 1.14, page 11

Le fait que, pour tout entier naturel n , $B_n \subseteq A_n$ est évident. Pour montrer que les B_n sont disjoints deux à deux, supposons que $B_n \cap B_m \neq \emptyset$ pour $n \neq m$. On peut considérer que $n < m$. Soit $x \in B_n \cap B_m$. Comme $x \in B_m$, $x \notin A_0 \cup A_1 \cup \dots \cup A_{m-1}$ donc $x \notin A_n$ car $n \leq m-1$. Donc $x \notin B_n$ car $B_n \subseteq A_n$. Ainsi $x \notin B_n \cap B_m$ et il y a donc contradiction. Donc par l'absurde, $B_n \cap B_m = \emptyset$.

Comme $B_n \subseteq A_n$, $\bigcup_{n=0}^{\infty} B_n \subseteq \bigcup_{n=0}^{\infty} A_n$. Montrons l'inclusion inverse. Soit $x \in \bigcup_{n=0}^{\infty} A_n$, il existe n_0 tel que $x \in A_{n_0}$ et pour tout $k < n_0$, $x \notin A_k$ (cet indice n_0 peut être 0), donc $x \in B_{n_0}$ et ainsi $x \in \bigcup_{n=0}^{\infty} B_n$. On vient de montrer l'inclusion $\bigcup_{n=0}^{\infty} A_n \subseteq \bigcup_{n=0}^{\infty} B_n$ et on en déduit donc l'égalité souhaitée.

Soit K un entier naturel. Comme $B_n \subseteq A_n$, $\bigcup_{n=0}^{n=K} B_n \subseteq \bigcup_{n=0}^{n=K} A_n$. Montrons l'inclusion inverse. Soit $x \in \bigcup_{n=0}^{n=K} A_n$, il existe n_0 tel que $x \in A_{n_0}$ et pour tout $k < n_0$, $x \notin A_k$ (cet indice n_0 peut être 0), donc $x \in B_{n_0}$ et ainsi $x \in \bigcup_{n=0}^{n=K} B_n$. On vient de montrer l'inclusion $\bigcup_{n=0}^{n=K} A_n \subseteq \bigcup_{n=0}^{n=K} B_n$ et on en déduit donc l'égalité souhaitée.

Corrigé de l'exercice 1.15, page 15

On doit trouver une fonction f de \mathbb{R} dans \mathbb{R} intégrable telle que $F(x) = \int_{-\infty}^x f(t)dt$. Comme F est dérivable à dérivée continue sur \mathbb{R} , il suffit de prendre $f = F'$ c'est-à-dire

$$f(x) = \begin{cases} \frac{1}{2}e^{+x}, & \text{si } x \leq 0; \\ \frac{1}{2}e^{-x}, & \text{si } x \geq 0. \end{cases}$$

Ainsi F est la fonction de répartition d'une probabilité à densité f définie par $f(x) = \frac{1}{2}e^{-|x|}$.

Corrigé de l'exercice 1.16, page 16

La f.r. de $\mathcal{N}_1(0; 1)$ est l'application Φ de \mathbb{R} dans $[0,1]$ définie par

$$\Phi(x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$$

c'est donc une application continue sur \mathbb{R} . Par suite d'après la proposition 1.14 2.(c), page 15, on en déduit que pour tout réel x , $\mathcal{N}_1(0; 1)(\{x\}) = 0$.

Supposons maintenant que $\mathcal{N}_1(0; 1) = \sum_{k=0}^{\infty} p_k \delta_{\alpha_k}$ où $(p_k)_{k \in \mathbb{N}}$ est une suite de réels positifs telle que $\sum_{k=0}^{\infty} p_k = 1$ et $(\alpha_k)_{k \in \mathbb{N}}$ est une suite de nombres réels, deux à deux distincts. On a alors, pour tout entier naturel k , $\mathcal{N}_1(0; 1)(\{\alpha_k\}) = p_k = 0$; ce qui contredit le fait que $\sum_{k=0}^{\infty} p_k = 1$. Donc $\mathcal{N}_1(0; 1)$ n'est pas une probabilité discrète.

Corrigé de l'exercice 1.17, page 16

1. Cela résulte de la continuité de la fonction de répartition $x \mapsto \int_{-\infty}^x \rho(t) dt$ et de la proposition 1.14, 2.(c), page 15.
2. On a $\mu(]a, b]) = \mu(]a, b]) - \mu(\{b\}) = (F(b) - F(a)) - (F(b) - F(b-)) = F(b-) - F(a)$.
De même $\mu([a, b[) = \mu(]a, b]) + \mu(\{a\}) = (F(b-) - F(a)) + (F(a) - F(a-)) = F(b-) - F(a-)$.

Corrigé de l'exercice 1.18, page 17

La fonction de répartition F de $\mathcal{U}([0, 1])$ est continue sur \mathbb{R} et est donnée par

$$F(x) = \begin{cases} 0, & \text{si } x < 0; \\ x, & \text{si } 0 \leq x \leq 1; \\ 1, & \text{si } x > 1. \end{cases}$$

On a donc $\mathcal{U}([0, 1])([1/6, 4/3]) = F(4/3) - F(1/6-) = F(4/3) - F(1/6) = 1 - 1/6 = 5/6$.
Comme $\mathbb{Q} = \cup_{q \in \mathbb{Q}} \{q\}$ (union de singletons disjoints deux à deux), on a par continuité de la f.r.,

$$\mathcal{U}([0, 1])(\mathbb{Q}) = \sum_{q \in \mathbb{Q}} \underbrace{\mathcal{U}([0, 1])(\{q\})}_{=0} = 0.$$

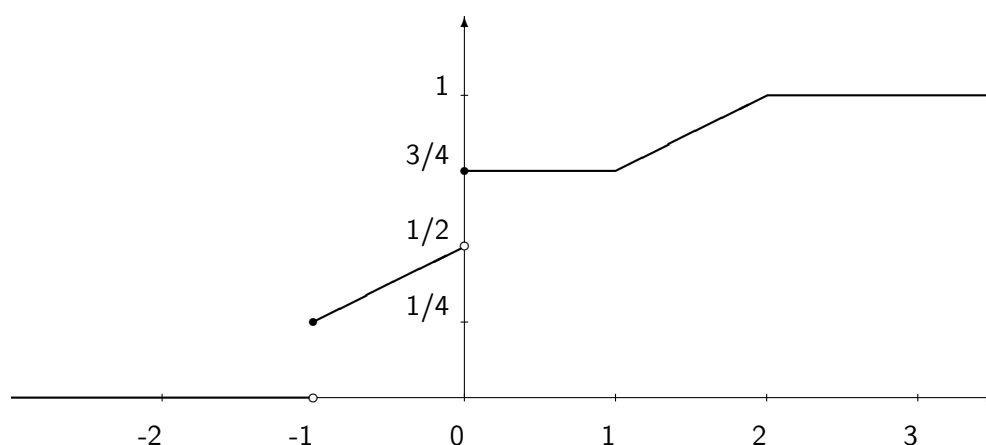
La fonction de répartition de $\mathcal{E}(2)$ est continue et est donnée par

$$F(x) = \begin{cases} 0, & \text{si } x < 0; \\ 1 - e^{-2x}, & \text{si } x \geq 0. \end{cases}$$

Donc $\mathcal{E}(2)(\{\pi\}) = 0$ et $\mathcal{E}(2)(\{\pi\} \cup [9/2, 7]) = \mathcal{E}(2)(\{\pi\}) + \mathcal{E}(2)([9/2, 7]) = 0 + F(7) - F(9/2) = e^{-9} - e^{-14}$.

Corrigé de l'exercice 1.19, page 17

La représentation graphique de F est :



On peut écrire F de la manière (moins synthétique mais plus lisible) suivante :

$$F(x) = \begin{cases} 0, & \text{si } x < -1; \\ \frac{x+2}{4}, & \text{si } -1 \leq x < 0 \text{ ou } 1 \leq x \leq 2; \\ \frac{3}{4}, & \text{si } 0 \leq x \leq 1; \\ 1, & \text{si } x \geq 2. \end{cases}$$

La fonction F présente des sauts ce qui est révélateur de la présence de Dirac dans l'expression de la probabilité.

Par ailleurs, F est bien une fonction de répartition car elle est croissante, continue à droite et $\lim_{x \rightarrow -\infty} F(x) = 0$ et $\lim_{x \rightarrow +\infty} F(x) = 1$.

La mesure μ sera la somme d'une mesure à densité et d'une variable discrète. A priori on ne dispose pas de résultat dans le cours pour conjecturer ce fait mais en pratique (en refaisant d'autres exercices de ce type) la méthode décrite ci-après permet de conclure. On peut considérer que pour tout $x \in \mathbb{R}$, $F(x) = F_1(x) + F_2(x)$ où

$$F_1(x) = \begin{cases} 0, & \text{si } x < -1; \\ \frac{x+1}{4}, & \text{si } -1 \leq x < 0; \\ \frac{1}{4}, & \text{si } 0 \leq x \leq 1; \\ \frac{x}{4}, & \text{si } 1 \leq x \leq 2; \\ \frac{1}{2}, & \text{si } x \geq 2. \end{cases}$$

et

$$F_2(x) = \begin{cases} 0, & \text{si } x < -1; \\ \frac{1}{4}, & \text{si } -1 \leq x < 0; \\ \frac{1}{2}, & \text{si } x \geq 0. \end{cases}$$

On remarque que F_1 est continue et F_2 permet de prendre en compte les sauts. On peut écrire $F_1(x) = \int_{-\infty}^x f_1(t) dt$ où

$$f_1(t) = \begin{cases} \frac{1}{4}, & \text{si } -1 \leq t < 0 \text{ ou } 1 \leq t \leq 2; \\ 0, & \text{sinon.} \end{cases}$$

Ainsi $F_1(x) = \mu_1([-\infty, x])$ où μ_1 est la mesure de densité f_1 . Pour F_2 on écrit

$$\begin{aligned} F_2(x) &= \frac{1}{4} \mathbb{1}_{[-1, +\infty[}(x) + \frac{1}{4} \mathbb{1}_{[0, +\infty[}(x) \\ &= \frac{1}{4} \delta_{-1}([-\infty, x]) + \frac{1}{4} \delta_0([-\infty, x]) \\ &= \mu_2([-\infty, x]) \end{aligned}$$

où $\mu_2 = \frac{1}{4} \delta_{-1} + \frac{1}{4} \delta_0$. Finalement F est la fonction de répartition de la probabilité $\mu = \mu_1 + \mu_2$. On remarquera que μ_1 et μ_2 ne sont pas elles-mêmes des probabilités.

8.2 Corrigés des exercices du chapitre II

Corrigé de l'exercice 2.1, page 21

Il suffit de vérifier que $\forall B \in \mathcal{B}(\mathbb{R}^d)$, $(f \circ \varphi)^{-1}(B) \in \mathcal{A}$. Or $(f \circ \varphi)^{-1}(B) = \varphi^{-1}(f^{-1}(B))$, comme f est borélienne, $f^{-1}(B) \in \mathcal{B}(\mathbb{R}^k)$ et comme φ est \mathcal{A} -mesurable, $\varphi^{-1}(f^{-1}(B)) \in \mathcal{A}$ d'où le résultat.

Corrigé de l'exercice 2.2, page 25

Soit F_X et F_Y les f.r. de X et Y respectivement. Soit $t \in \mathbb{R}$.

- Si $t \leq 0$, comme $\forall \omega \in \Omega$, $Y(\omega) = e^{X(\omega)} > 0$ on en déduit que $F_Y(t) = \mathbb{P}(Y \leq t) = \mathbb{P}(\emptyset) = 0$.

- Supposons maintenant que $t > 0$. On rappelle que X étant une v.a.r. de loi $\mathcal{N}_1(0; 1)$, $F_X(x) = (1/\sqrt{2\pi}) \int_{-\infty}^x e^{-u^2/2} du$. On a donc

$$\begin{aligned} F_Y(t) &= \mathbb{P}(Y \leq t) = \mathbb{P}(e^X \leq t) = \mathbb{P}(X \leq \ln t) = F_X(\ln t) \\ &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\ln t} e^{-u^2/2} du \end{aligned}$$

et en faisant le changement de variable $u = \ln x$ (en remarquant que $-\infty < u \leq \ln t \Leftrightarrow 0 < x \leq t$) on obtient

$$F_Y(t) = \frac{1}{\sqrt{2\pi}} \int_0^t \frac{e^{-(\ln x)^2/2}}{x} dx.$$

Il reste à faire apparaître une densité ce qui revient à écrire $F_Y(t)$ sous forme d'une intégrale entre $-\infty$ et t :

$$F_Y(t) = \int_{-\infty}^t \rho(x) dx$$

où on a posé

$$\rho(x) = \frac{1}{\sqrt{2\pi x}} e^{-\frac{1}{2}(\ln x)^2} \mathbb{1}_{]0, +\infty[}(x)$$

d'où le résultat cherché.

Corrigé de l'exercice 2.3, page 25

Déterminons la fonction de répartition de la v.a.r. Y .

$$F_Y(t) = \mathbb{P}(Y \leq t) = \begin{cases} 0 & \text{si } t < 0 \\ \mathbb{P}(-t \leq X \leq t) & \text{si } t \geq 0 \end{cases}$$

Puisque F_X est continue, on peut écrire $\mathbb{P}(-t \leq X \leq t) = F_X(t) - F_X(-t) = \frac{1}{2}(2 - e^{-t}) - \frac{1}{2}e^{-t} = 1 - e^{-t}$. On reconnaît la fonction de répartition d'une v.a. de loi exponentielle $\mathcal{E}(1)$.

Corrigé de l'exercice 2.4, page 26

On écrit les équivalences suivantes :

$$\begin{aligned} Y = \sigma X + m \text{ suit une loi } \mathcal{N}_1(m; \sigma^2) &\Leftrightarrow \forall t \in \mathbb{R} ; \mathbb{P}(Y \leq t) = \int_{-\infty}^t \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-m)^2}{2\sigma^2}} dx \\ &\Leftrightarrow \forall t \in \mathbb{R} ; \mathbb{P}\left(X \leq \frac{t-m}{\sigma}\right) = \int_{-\infty}^{\frac{t-m}{\sigma}} \frac{1}{\sqrt{2\pi}} e^{-\frac{v^2}{2}} dv \\ &\Leftrightarrow \forall x \in \mathbb{R} ; \mathbb{P}(X \leq x) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-\frac{v^2}{2}} dv, \end{aligned}$$

où on a d'abord posé $x = (t - m)/\sigma$ puis fait le changement de variable $v = (u - m)/\sigma$.

Corrigé de l'exercice 2.5, page 26

Soit X une variable aléatoire réelle de loi

$$\mathbb{P}_X := \sum_{k \in \mathbb{N}} p_k \delta_k.$$

Fixons $n \in \mathbb{N}$, il vient

$$\mathbb{P}_X(\{n\}) := \sum_{k \in \mathbb{N}} p_k \delta_k(\{n\}).$$

Or $\delta_k(\{n\}) = 0$, si $k \neq n$ et $\delta_n(\{n\}) = 1$. Par suite $\sum_{k \in \mathbb{N}} p_k \delta_k(\{n\}) = p_n$ et $\mathbb{P}_X(\{n\}) = p_n$. D'où le résultat cherché.

Corrigé de l'exercice 2.6, page 28

1. Le graphe de F est une fonction en escalier, croissante et continue à droite, présentant des sauts de discontinuité de première espèce en tout point d'abscisse $n \geq 1$, la hauteur du saut étant égale à $\mathbb{P}(X = n)$.
2. F est une fonction de répartition car elle est définie sur \mathbb{R} à valeurs dans $[0, 1]$, croissante, continue à droite en tout point, avec $\lim_{x \rightarrow -\infty} F(x) = 0$ et $\lim_{x \rightarrow +\infty} F(x) = 1$.
3. Soit $n > 1$ un entier naturel. D'après le cours, pour tout réel x , $\mathbb{P}(X = x)$ est égal à la valeur du saut de F au point d'abscisse x . C'est-à-dire

$$\mathbb{P}(X = n) = F(n) - F(n-) = \left(1 - \frac{1}{n(n+1)}\right) - \left(1 - \frac{1}{n(n-1)}\right) = \frac{2}{n(n^2-1)}.$$

On a de même $\mathbb{P}(X = 0) = 0$ et $\mathbb{P}(X = 1) = \frac{1}{2}$.

4. Par définition, comme X est une variable discrète à valeurs dans \mathbb{N} ,

$$\begin{aligned} \mathbb{E}(X) &:= \sum_{n=1}^{\infty} n \mathbb{P}(X = n) \\ \mathbb{E}(X) &= \frac{1}{2} + \sum_{n=2}^{\infty} \frac{2}{(n-1)(n+1)} \\ \mathbb{E}(X) &= \frac{1}{2} + \sum_{n=2}^{\infty} \left(\frac{1}{n-1} - \frac{1}{n+1} \right) \\ \mathbb{E}(X) &= \frac{1}{2} + \left(1 - \frac{1}{3}\right) + \left(\frac{1}{2} - \frac{1}{3}\right) + \left(\frac{1}{3} - \frac{1}{4}\right) + \cdots + \left(\frac{1}{n-1} - \frac{1}{n+1}\right) + \cdots \\ \mathbb{E}(X) &= \frac{1}{2} + 1 + \frac{1}{2} = 2. \end{aligned}$$

5. Calculons la variance de X . Vérifions auparavant que la variable X est de carré intégrable, c'est-à-dire que $\mathbb{E}(X^2)$ est un réel fini. Or, par définition des moments d'ordre 2, et suivant un calcul analogue au précédent, il vient :

$$\mathbb{E}(X^2) = \sum_{n=1}^{\infty} n^2 \mathbb{P}(X = n) = \frac{1}{2} + \sum_{n=2}^{\infty} \frac{2n}{(n-1)(n+1)}.$$

La série à terme général réel positif $\sum_{n=2}^{\infty} \frac{2n}{(n-1)(n+1)}$ ne converge pas. En effet, son terme général est équivalent à $\frac{1}{n}$, terme général d'une série divergente (série harmonique).

D'après le critère de l'équivalence pour les séries à terme général positif, la série $\sum_{n=2}^{\infty} \frac{2n}{(n-1)(n+1)}$ diverge. Donc la variable aléatoire X n'admet pas de variance.

8.3 Corrigés des exercices du chapitre III

Corrigé de l'exercice 3.1, page 31

1. D'après la conclusion 1.(b) de la proposition 3.1, page 30, $\mathbb{E}_{\mu}(f) = \sum_{i=1}^n a_i \mathbb{E}_{\mu}(\mathbf{1}_{A_i})$. Or d'après 1.(a) de cette même proposition, $\mathbb{E}_{\mu}(\mathbf{1}_{A_i}) = \mu(A_i)$ et donc $\mathbb{E}_{\mu}(f) = \sum_{i=1}^n a_i \mu(\mathbf{1}_{A_i})$.
2. D'après ce qui vient d'être vu, $\mathbb{E}_{\mu}(f) = \pi \mu([0, 1/3]) + \mu([6, 10]) + 3\mu(\{5\})$. Mais

$$\mu([0, 1/3]) = \underbrace{\delta_0([0, 1/3])}_{=1} + \underbrace{\delta_5([0, 1/3])}_{=0} + \underbrace{\lambda([0, 1/3])}_{=1/3-0} = 4/3.$$

De même

$$\begin{aligned} \mu([6, 10]) &= \underbrace{\delta_0([6, 10])}_{=0} + \underbrace{\delta_5([6, 10])}_{=0} + \underbrace{\lambda([6, 10])}_{=10-6} = 4 \text{ et} \\ \mu(\{5\}) &= \underbrace{\delta_0(\{5\})}_{=0} + \underbrace{\delta_5(\{5\})}_{=1} + \underbrace{\lambda(\{5\})}_{=0} = 1. \end{aligned}$$

Par suite $\mathbb{E}_{\mu}(f) = \frac{4}{3}\pi + 7$.

Corrigé de l'exercice 3.2, page 33

1. On introduit l'application $\tilde{I} : \mathcal{M}^+ \rightarrow [0, \infty]$ définie pour $f \in \mathcal{M}^+(E, \mathcal{A})$ par $\tilde{I}(f) = \mathbb{E}_{\mu}(f) + \mathbb{E}_{\nu}(f)$. On montre aisément que \tilde{I} vérifie les conditions 1.(a), 1.(b) et 1.(c) de la proposition 3.1, page 30, c'est-à-dire :

- (a) Pour tout $A \in \mathcal{A}$, $\tilde{I}(\mathbf{1}_A) = (\mu + \nu)(A)$.
- (b) Pour tous f et g appartenant à $\mathcal{M}^+(E, \mathcal{A})$ et tout réel $\alpha \geq 0$.

$$\tilde{I}(f + g) = \tilde{I}(f) + \tilde{I}(g) \text{ et } \tilde{I}(\alpha f) = \alpha \tilde{I}(f).$$

- (c) Pour toute suite croissante $(f_n)_{n \in \mathbb{N}}$ d'éléments de $\mathcal{M}^+(E, \mathcal{A})$,

$$\lim_{n \rightarrow +\infty} \tilde{I}(f_n) = \tilde{I}\left(\lim_{n \rightarrow +\infty} f_n\right).$$

On conclut alors par l'unicité de l'opérateur que $\tilde{I} = \mathbb{E}_{\mu+\nu}$, c'est-à-dire que pour tout $f \in \mathcal{M}^+(E, \mathcal{A})$, $\tilde{I}(f) = \mathbb{E}_{\mu+\nu}(f)$ ou encore $\mathbb{E}_{\mu}(f) + \mathbb{E}_{\nu}(f) = \mathbb{E}_{\mu+\nu}(f)$.

2. D'après la question précédente, $\mathbb{E}_{\mu}(f) + \mathbb{E}_{\nu}(f) = \mathbb{E}_{\mu+\nu}(f)$ or

$$\mathbb{E}_{\mu}(f) = \mathbb{E}_{\lambda}(\rho f) \text{ d'après la proposition 3.1, page 30,}$$

$$= \int_{-\infty}^{+\infty} \rho(x) f(x) dx \text{ d'après la proposition 3.2, page 31,}$$

$$= \int_0^1 \alpha e^{-\alpha x} e^{\alpha x} dx = \alpha,$$

$$\text{et } \mathbb{E}_{\nu}(f) = e^{-\alpha} f(1) \text{ d'après la proposition 3.4, page 32,}$$

$$= e^{-\alpha} e^{\alpha \times 1} = 1.$$

Finalement $\mathbb{E}_{\mu+\nu}(f) = \alpha + 1$.

3. Un raisonnement analogue conduit à :

$$\begin{aligned}\mathbb{E}_{\mu}(f) &= \int_{-1}^1 e^{-\alpha x} e^{\alpha x} dx = 2 \\ \mathbb{E}_{\nu}(f) &= \sum_{k=0}^{\infty} \frac{\alpha^k}{k!} f(k) = \sum_{k=0}^{\infty} \frac{\alpha^k}{k!} e^{\alpha k} = \sum_{k=0}^{\infty} \frac{(\alpha e^{\alpha})^k}{k!} = \exp(\alpha e^{\alpha}).\end{aligned}$$

Par suite $\mathbb{E}_{\mu+\nu}(f) = 2 + \exp(\alpha e^{\alpha})$. De même

$$\mathbb{E}_{\mu+\nu}(\mathbf{1}_{\mathbb{R}}) = \mathbb{E}_{\mu}(\mathbf{1}_{\mathbb{R}}) + \mathbb{E}_{\nu}(\mathbf{1}_{\mathbb{R}}) = \int_{-1}^1 e^{-\alpha x} dx + \sum_{k=0}^{\infty} \frac{\alpha^k}{k!} = \frac{e^{\alpha} - e^{-\alpha}}{\alpha} + e^{\alpha} > 1$$

donc $\mu + \nu$ n'est pas une probabilité.

Corrigé de l'exercice 3.3, page 34

L'application borélienne f est intégrable suivant μ si, et seulement si, $\mathbb{E}_{\mu}(|f|) < +\infty$. Or d'après la proposition 3.4, page 32,

$$\mathbb{E}_{\mu}(|f|) = \sum_{i=0}^{\infty} \alpha_i |f(a_i)| = \sum_{i=0}^{\infty} \alpha_i |f(a_i)|.$$

Ainsi f sera intégrable par rapport à $\mu = \sum_{i=0}^{\infty} \alpha_i \delta_i$ si, et seulement si, la série numérique à termes positifs $\sum_{i=0}^{\infty} \alpha_i |f(a_i)|$ est convergente ce qui est équivalent à l'absolue convergence de la série numérique $\sum_{i=0}^{\infty} \alpha_i f(a_i)$.

Corrigé de l'exercice 3.5, page 36

Soit $x = (x_1, \dots, x_d) \in \mathbb{R}^d$, $|x| = \sqrt{x_1^2 + \dots + x_d^2}$. L'application f sera intégrable pour μ si, et seulement si, $\mathbb{E}_{\mu}(|f|) < \infty$, c-à-d $\mathbb{E}_{\mu}(\sqrt{f_1^2 + \dots + f_d^2}) < \infty$. Or pour tout $i = 1, \dots, d$ on a

$$|f_i| \leq \sqrt{f_1^2 + \dots + f_d^2} \leq |f_1| + \dots + |f_d|.$$

D'où par la croissance de l'opérateur \mathbb{E}_{μ} pour les fonctions positives (proposition 3.1, assertion 2, page 30),

$$\mathbb{E}_{\mu}(|f_i|) \leq \mathbb{E}_{\mu}(|f|) \leq \sum_{i=1}^d \mathbb{E}_{\mu}(|f_i|)$$

et par suite $\mathbb{E}_{\mu}(|f|) < +\infty \Leftrightarrow \forall i = 1, \dots, d ; \mathbb{E}_{\mu}(|f_i|) < +\infty$.

Corrigé de l'exercice 3.6, page 37

1. Cas où μ est la probabilité de Dirac sur \mathbb{R} au point a .

$$\Phi_{\mu}(t) = \int_{\mathbb{R}} e^{itx} d(\delta_a)(x) = e^{iat}.$$

2. Cas où μ est la probabilité binomiale de paramètres n et p .

$$\Phi_{\mu}(t) = \int_{\mathbb{R}} e^{itx} \sum_{k=0}^n C_n^p p^k (1-p)^{n-k} d\delta_k(x) = \sum_{k=0}^n C_n^p (pe^{it})^k (1-p)^{n-k}. \text{ Ce qui donne avec la formule du binôme de Newton : } \phi_{\mu}(t) = (1 - p + pe^{it})^n.$$

3. Cas où μ est la probabilité de Poisson de paramètres $\alpha > 0$.

$$\phi_{\mu}(t) = \int_{\mathbb{R}} e^{itx} \sum_{k=0}^{+\infty} e^{-\alpha} \frac{(\alpha e^{it})^k}{k!} d\delta_k(x) = e^{-\alpha(e^{it}-1)}.$$

Corrigé de l'exercice 3.7, page 42

1. Méthode 1) : on applique le théorème de transfert (cas positif) à la fonction h définie par $h(x) = e^x$. En notant $\mu = \mathbb{P}_X = \mathcal{N}(0; 1)$ on a (d'après les règles d'intégration)

$$\mathbb{E}(e^X) = \mathbb{E}_{\mu}(h) = \int_{\mathbb{R}} e^x d\mathbb{P}_X(x) = \int_{\mathbb{R}} e^x \frac{1}{\sqrt{2\pi}} e^{-x^2/2} d\lambda(x) = \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{x-x^2/2} dx.$$

Or $x - x^2/2 = -(x-1)^2/2 + 1/2$ d'où

$$\mathbb{E}(e^X) = \frac{\sqrt{e}}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{\frac{1}{2}(x-1)^2} dx = \sqrt{e}.$$

Méthode 2) : On a vu au chapitre II, exercice 2.3, page 22, que la v.a.r. $Y = e^X$ est une v.a.r. à densité ρ donnée par

$$\rho(x) := \frac{1}{x\sqrt{2\pi}} \exp\left(-\frac{1}{2}(\ln x)^2\right) \mathbb{1}_{]0, +\infty[}(x).$$

On a alors $\mathbb{E}(e^X) = \mathbb{E}(Y) = \mathbb{E}_{\mu}(f)$ où $f(y) = y$ et μ est la probabilité de densité ρ . Ainsi

$$\begin{aligned} \mathbb{E}(e^X) &= \int_{\mathbb{R}} y \rho(y) d\lambda(y) = \int_{-\infty}^{+\infty} \frac{y}{y\sqrt{2\pi}} \exp\left(-\frac{1}{2}(\ln y)^2\right) \mathbb{1}_{]0, +\infty[}(y) dy \\ &= \frac{1}{\sqrt{2\pi}} \int_0^{\infty} e^{-\frac{1}{2}(\ln y)^2} dy = \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-u^2/2} e^u du \end{aligned}$$

où on a effectué le changement de variables $u = \ln y$. D'où $\mathbb{E}(e^X) = \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{u-u^2/2} du$ et on est ramené à la même intégrale que dans la première méthode.

2. En effectuant le changement de variable $u = x^2$ et en faisant une intégration par parties on écrit

$$\begin{aligned} \mathbb{E}(|X|^3) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} |x|^3 e^{-x^2/2} dx = \frac{2}{\sqrt{2\pi}} \int_0^{+\infty} x^3 e^{-x^2/2} dx \\ &= \frac{2}{\sqrt{2\pi}} \int_0^{+\infty} u \sqrt{u} e^{-u/2} \frac{du}{2\sqrt{u}} \\ &= \underbrace{\left[\frac{1}{\sqrt{2\pi}} (-2u) e^{-u/2} \right]_0^{\infty}}_{=0} - \underbrace{\frac{1}{\sqrt{2\pi}} \int_0^{+\infty} (-2) e^{-u/2} du}_{=4/\sqrt{2\pi}}, \end{aligned}$$

d'où $\mathbb{E}(|X|^3) = 4/\sqrt{2\pi} < \infty$ donc X^3 est intégrable.

Calculons $\mathbb{E}(X^3)$:

$$\mathbb{E}(X^3) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} x^3 e^{-x^2/2} dx = 0$$

car la fonction $x \mapsto x^3 e^{-x^2/2}$ est impaire sur \mathbb{R} .

Corrigé de l'exercice 3.8, page 44

• On montre que $\mathbb{E}(X) = (2; 2)$. On sait que si $X = (X_1, X_2)$, $\mathbb{E}(X) = (\mathbb{E}(X_1), \mathbb{E}(X_2))$. Or $\mathbb{E}(X_1) = \mathbb{E}(h(X)) = \int_{\mathbb{R}^2} h(x, y) d\mathbb{P}_X(x, y)$ où on a posé h fonction de $\mathbb{R}^+ \times \mathbb{R}^+ \rightarrow \mathbb{R}^+$ définie par $h(x, y) = x$. D'où

$$\mathbb{E}(X_1) = \int_{\mathbb{R}^2} h(x, y) d\left(\sum_{k,l=1}^{\infty} \frac{1}{2^{k+l}} \delta_{(k,l)}\right)(x, y) = \sum_{k=1}^{\infty} \sum_{l=1}^{\infty} \frac{1}{2^{k+l}} \underbrace{h(k, l)}_{=k} = \sum_{k=1}^{\infty} \sum_{l=1}^{\infty} \frac{k}{2^{k+l}}.$$

Par suite

$$\mathbb{E}(X_1) = \sum_{k=1}^{\infty} \frac{k}{2^k} \underbrace{\left(\sum_{l=1}^{\infty} \frac{1}{2^l}\right)}_{=1} = \sum_{k=1}^{\infty} k \left(\frac{1}{2}\right)^k$$

et comme en dérivant terme à terme une série géométrique on montre facilement que $x \sum_{k=1}^{\infty} kx^{k-1} = \frac{x}{(x-1)^2}$ pour $0 < x < 1$, on déduit que $\mathbb{E}(X_1) = \frac{1/2}{(1-1/2)^2} = 2$. On raisonne de même pour montrer que $\mathbb{E}(X_2) = 2$.

• On détermine la loi de la v.a.r. $Z = X_1 + X_2$. On applique le critère d'identification des lois par les fonctions boréliennes positives. Soit h de $\mathbb{R} \rightarrow [0, \infty]$ borélienne. En posant $\varphi(x, y) = h(x + y)$ (qui est borélienne positive car h l'est), on a $\mathbb{E}(h(Z)) = \mathbb{E}(h(X_1 + X_2)) = \mathbb{E}(\varphi(Z))$. Par le théorème du transfert,

$$\begin{aligned} \mathbb{E}(h(Z)) &= \int_{\mathbb{R}^2} \varphi(x, y) d\mathbb{P}_X(x, y) = \int_{\mathbb{R}^2} h(x + y) d\left(\sum_{k,l=1}^{\infty} \frac{1}{2^{k+l}} \delta_{(k,l)}\right)(x, y) \\ &= \sum_{k=1}^{\infty} \sum_{l=1}^{\infty} \frac{1}{2^{k+l}} h(k + l) = \sum_{i=2}^{\infty} \sum_{(k,l)/k+l=i} \frac{1}{2^i} h(i). \end{aligned}$$

Or il existe $i - 1$ couples (k, l) tels que $k + l = i$ avec $1 \leq k$ et $1 \leq l$ donc $\sum_{(k,l)/k+l=i} \frac{1}{2^i} h(i) = (i - 1) \times h(i) / 2^i$. D'où

$$\mathbb{E}(h(Z)) = \sum_{i=2}^{\infty} \frac{i-1}{2^i} h(i) = \int_{\mathbb{R}} h(x) d\mu(x)$$

avec $\mu = \sum_{i=2}^{\infty} \frac{i-1}{2^i} \delta_i = \sum_{i=1}^{\infty} \frac{i-1}{2^i} \delta_i$ car $i - 1 = 0$ pour $i = 1$.

Corrigé de l'exercice 3.9, page 46

1. On a $\text{Var}(X) = \mathbb{E}((X - \mathbb{E}(X))^2) = \text{Cov}(X, X)$. En posant $\mathbb{E}(X) = m$:

$$\begin{aligned} \text{Var}(X) &= \mathbb{E}(X - m)^2 = \mathbb{E}(X^2 - 2mX + m^2) = \mathbb{E}(X^2) - 2m\mathbb{E}(X) + m^2 \\ &= \mathbb{E}(X^2) - 2m^2 + m^2 = \mathbb{E}(X^2) - m^2 = \mathbb{E}(X^2) - (\mathbb{E}(X))^2. \end{aligned}$$

2. On pose $m_X = \mathbb{E}(X)$ et $m_Y = \mathbb{E}(Y)$ et on a :

$$\begin{aligned} \text{Cov}(X, Y) &= \mathbb{E}((X - m_X)(Y - m_Y)) = \mathbb{E}(XY) - m_X \mathbb{E}(Y) - m_Y \mathbb{E}(X) + m_X m_Y \\ &= \mathbb{E}(XY) - \mathbb{E}(X)m_Y - m_Y \mathbb{E}(X) + m_X m_Y = \mathbb{E}(XY) - m_X m_Y \\ &= \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y). \end{aligned}$$

Corrigé de l'exercice 3.10, page 47

On peut vérifier que ρ est bien une densité de probabilité sur \mathbb{R} . En particulier on a donc $\int_0^{+\infty} x e^{-x^2/2} dx = 1$. On remarque que X est une v.a.r. positive (presque-sûrement) c'est-à-dire $\mathbb{P}(X < 0) = 0$ ou encore $\mathbb{P}(X = 0) = 1$. On peut donc faire les calculs comme si X est à valeurs dans $[0, +\infty]$.

Soit r un entier naturel strictement positif, par le théorème de transfert :

$$\mathbb{E}(X^r) = \int_{\mathbb{R}} x^r d\mathbb{P}_X(x) = \int_{\mathbb{R}} x^{r+1} e^{-x^2/2} \mathbf{1}_{\mathbb{R}^+}(x) d\lambda(x) = \int_0^{\infty} x^{r+1} e^{-x^2/2} dx.$$

Par des techniques d'intégration par parties successives, on recherche une relation de récurrence sur l'entier $r \geq 2$:

$$\int_0^{\infty} x^{r+1} e^{-x^2/2} dx = r \int_0^{\infty} x^{r-1} e^{-x^2/2} dx$$

c'est-à-dire $\mathbb{E}(X^r) = r\mathbb{E}(X^{r-2})$. En distinguant les cas où $r = 2k$ et $r = 2k - 1$, et en tenant compte que $\int_0^{\infty} e^{-x^2/2} dx = \sqrt{\pi/2}$, on trouve les relations demandées. En particulier $\mathbb{E}(X) = \sqrt{\pi/2}$ et $\mathbb{E}(X^2) = 2$ d'où $\text{Var}(X) = 2 - \pi/2$.

Corrigé de l'exercice 3.11, page 47

Supposons tout d'abord que pour tout i , $\mathbb{E}(X_i^2) < \infty$. Alors comme $|X|^2 = \sum_{i=1}^d X_i^2$, $\mathbb{E}(|X|^2) = \sum_{i=1}^d \mathbb{E}(X_i^2) < \infty$.

Réciproquement, si $\mathbb{E}(|X|^2) < \infty$ alors $\mathbb{E}(X_i^2) < \infty$ car $X_i^2 \leq |X|^2$.

Corrigé de l'exercice 3.12, page 48

Notons X_1, \dots, X_d les composantes de X dans la base canonique de \mathbb{R}^d , M la matrice $[a_{ij}]_{1 \leq i \leq c, 1 \leq j \leq d}$. Le vecteur MX de \mathbb{R}^c a pour composante $((MX)_1, \dots, (MX)_c)$ où pour $i = 1, \dots, c$, $(MX)_i = \sum_{j=1}^d a_{ij} X_j$.

• Le vecteur $Y = X - \mathbb{E}(X) \in \mathbb{R}^d$ a pour composante $Y_i = X_i - \mathbb{E}(X_i)$ d'où $\mathbb{E}(Y_i) = 0$. Ainsi $Y = Y - \mathbb{E}(Y) = X - \mathbb{E}(X)$ et par suite

$$D_Y = \mathbb{E}((Y - \mathbb{E}(Y))(Y - \mathbb{E}(Y))^t) = \mathbb{E}((X - \mathbb{E}(X))(X - \mathbb{E}(X))^t) = D_X$$

d'où une première assertion du point 2) de la proposition 3.24, page 48.

• Le vecteur $\mathbb{E}(MX)$ a pour composantes $\sum_{j=1}^d a_{ij} \mathbb{E}(X_j)$ pour $i = 1, \dots, c$. Ce sont aussi les composantes de $M \times \mathbb{E}(X)$. D'où l'égalité $\mathbb{E}(MX) = M\mathbb{E}(X)$.

• Le terme (i, j) de la matrice D_X est $\mathbb{E}[(X_i - \mathbb{E}(X_i))(X_j - \mathbb{E}(X_j))] = \text{Cov}(X_i, X_j)$, ce qui prouve aussi que D_X est symétrique car $\text{Cov}(X_i, X_j) = \text{Cov}(X_j, X_i)$. Par ailleurs, les éléments diagonaux de D_X sont $\text{Cov}(X_i, X_i) = \text{Var}(X_i)$ pour $i = 1, \dots, d$.

- Montrons que $D_{MX} = MD_X M^t$. Soit $i, j = 1, \dots, c$, le coefficient (i, j) de D_{MX} est

$$\begin{aligned}
 \text{Cov}((MX)_i, (MX)_j) &= \mathbb{E}[(MX)_i (MX)_j] - \mathbb{E}[(MX)_i] \mathbb{E}[(MX)_j] \\
 &= \mathbb{E} \left[\sum_{k=1}^d a_{ik} X_k \sum_{l=1}^d a_{jl} X_l \right] - \mathbb{E} \left[\sum_{k=1}^d a_{ik} X_k \right] \mathbb{E} \left[\sum_{l=1}^d a_{jl} X_l \right] \\
 &= \sum_{k=1}^d \sum_{l=1}^d a_{ik} a_{jl} \left[\underbrace{\mathbb{E}(X_k X_l) - \mathbb{E}(X_k) \mathbb{E}(X_l)}_{= \text{Cov}(X_k, X_l)} \right] \\
 &= \sum_{k=1}^d \sum_{l=1}^d a_{ik} \text{Cov}(X_k, X_l) a_{jl} \\
 &= \text{coefficient } (i, j) \text{ de la matrice } MD_X M^t \text{ (matrice carrée d'ordre } c)
 \end{aligned}$$

d'où l'égalité cherchée.

- Montrons que d_X est positive. Soit $U = (u_1, \dots, u_d)^t \in \mathbb{R}^d$ (U est aussi une matrice colonne). On a $D_{U^t X} = U^t D_X (U^t)^t = U^t D_X U$. Or $U^t X = \sum_{j=1}^d u_j X_j$ et $D_{U^t X}$ représente la variance de cette v.a.r. . Ainsi $U^t D_X U \geq 0$ pour tout vecteur U (ou matrice unicolonne).

Le reste des propriétés de la matrice D_X se déduisent de la théorie de la diagonalisation des matrices symétriques, de type positif.

Corrigé de l'exercice 3.13, page 48

1. Les équivalences suivantes donnent :

$$(|X| - |Y|)^2 \geq 0 \Leftrightarrow X^2 + Y^2 - 2|XY| \geq 0 \Leftrightarrow X^2 + Y^2 \geq 2|XY| \geq |XY|.$$

On en déduit que $\mathbb{E}(|XY|) \leq \mathbb{E}(X^2) + \mathbb{E}(Y^2) < \infty$, ainsi la v.a.r. XY est intégrable. En choisissant $Y = \mathbf{1}_\Omega$ (c'est la variable aléatoire constante égale à 1), on a $|XY| = |X| \leq X^2 + \mathbf{1}_\Omega$ d'où

$$\mathbb{E}(|X|) \leq \mathbb{E}(X^2) + \mathbb{E}(\mathbf{1}_\Omega) = \mathbb{E}(X^2) + 1 < \infty.$$

On procède de même pour Y .

2. Pour tout $\alpha \in \mathbb{R}$,

$$0 \leq \mathbb{E}((X + \alpha Y)^2) = \alpha^2 \mathbb{E}(Y^2) + 2\alpha \mathbb{E}(XY) + \mathbb{E}(X^2).$$

Nous avons un polynôme en α de degré deux qui est positif. Il ne peut donc pas avoir de racines réelles distinctes et son discriminant est donc négatif ou nul, c'est-à-dire $(\mathbb{E}(XY))^2 - \mathbb{E}(Y^2)\mathbb{E}(X^2) \leq 0$ d'où le résultat.

Corrigé de l'exercice 3.14, page 49

1. Pour tout $i = 1, \dots, n$, X_i est de carré intégrable car $\mathbb{E}(|X|^2) < \infty$. On a

$$\begin{aligned} Y^2 &= \sum_{i=1}^n X_i^2 + \sum_{1 \leq i \neq j \leq n} X_i X_j \\ &\leq \sum_{i=1}^n X_i^2 + \sum_{1 \leq i \neq j \leq n} |X_i X_j| \quad \text{ce qui entraîne} \\ \mathbb{E}(Y^2) &\leq \underbrace{\sum_{i=1}^n \mathbb{E}(X_i^2)}_{< \infty \text{ car } X_i \text{ intégrables}} + \underbrace{\sum_{1 \leq i \neq j \leq n} \mathbb{E}(|X_i X_j|)}_{< \infty \text{ d'après l'exercice 3.13, page 48}}, \end{aligned}$$

d'où $\mathbb{E}(Y^2) < \infty$ ce qui prouve que $\sum_{k=1}^n X_k$ est de carré intégrable.

2. On a :

$$\begin{aligned} \text{Var}(Y) &= \mathbb{E}[(Y - \mathbb{E}(Y))^2] = \mathbb{E} \left[\left(\sum_{i=1}^n (X_i - \mathbb{E}(X_i)) \right)^2 \right] \\ &= \sum_{i=1}^n \underbrace{\mathbb{E}[(X_i - \mathbb{E}(X_i))^2]}_{=\text{Var}(X_i)} + \sum_{1 \leq i \neq j \leq n} \underbrace{\mathbb{E}[(X_i - \mathbb{E}(X_i))(X_j - \mathbb{E}(X_j))]}_{=\text{Cov}(X_i, X_j)}, \end{aligned}$$

d'où le résultat en remarquant que

$$\sum_{1 \leq i \neq j \leq n} \mathbb{E}[(X_i - \mathbb{E}(X_i))(X_j - \mathbb{E}(X_j))] = 2 \sum_{1 \leq i < j \leq n} \mathbb{E}[(X_i - \mathbb{E}(X_i))(X_j - \mathbb{E}(X_j))]$$

car $\text{Cov}(X_i, X_j) = \text{Cov}(X_j, X_i)$.

Corrigé de l'exercice 3.15, page 59

1. Pour vérifier la relation, on étudie la valeur de chacun des deux membres de l'égalité pour ω tel que $|X(\omega)| > a$, puis pour ω tel que $|X(\omega)| \leq a$. On constate alors que les deux membres coïncident dans chaque cas.
2. Remarquons que X et $-X$ ont la même loi. En effet l'étude des fonctions caractéristiques donne $\phi_{-X}(t) = \mathbb{E}[e^{-itX}] = \phi_X(-t) = e^{-\frac{1}{2}t^2} = \phi_X(t)$. Soient h une application borélienne positive de \mathbb{R} dans \mathbb{R} , a un réel strictement positif, de la relation,

$$h(X_a) = h(X)\mathbb{1}_{[0,a]}(|X|) + h(-X)\mathbb{1}_{]a,+\infty[}(|X|)$$

on obtient, en passant à l'espérance,

$$\mathbb{E}[h(X_a)] = \mathbb{E}[h(X)\mathbb{1}_{[0,a]}(|X|)] + \mathbb{E}[h(-X)\mathbb{1}_{]a,+\infty[}(|X|)].$$

En remarquant que X et $-X$ ont la même loi et en utilisant le théorème du transfert, il vient

$$\mathbb{E}[h(-X)\mathbb{1}_{]a,+\infty[}(|X|)] = \mathbb{E}[h(X)\mathbb{1}_{]a,+\infty[}(|X|)].$$

Ce qui, en reportant dans le deuxième membre de l'égalité précédente, donne $\mathbb{E}[h(X_a)] = \mathbb{E}[h(X)]$, pour toute application borélienne positive de \mathbb{R} dans \mathbb{R} . Ce qui prouve que X_a suit la même loi que X c'est-à-dire $\mathcal{N}_1(0, 1)$.

Corrigé de l'exercice 3.16, page 59

1. Le réel a est déterminée par la contrainte $\int_{-\infty}^{+\infty} f(x)dx = 1$ qui s'exprime ici par

$$\int_{-e}^{-1} \frac{a}{x} dx + \int_{-1}^0 (x+1-a) dx = 0$$

et un rapide calcul donne $a = -1/4$. La fonction de répartition de X est définie pour $x \in \mathbb{R}$ par $F(x) = \int_{-\infty}^x f(t)dt$. On a :

- ◇ Si $x \leq -e$, $F(x) = 0$.
- ◇ Si $-e < x < -1$, $F(x) = \int_{-e}^x f(t)dt = \int_{-e}^x -1/(4t)dt = -(1/4) \ln(-x) + 1/4$.
- ◇ Si $-1 \leq x < 0$, $F(x) = \int_{-e}^{-1} -1/(4t)dt + \int_{-1}^x (t+5/4)dt = x^2/2 + 5x/4 + 1$.
- ◇ Si $x \geq 0$, $F(x) = \int_{-\infty}^x f(t)dt = \int_{-e}^0 f(t)dt = 1$.

D'où F est définie par

$$F(x) = \begin{cases} 0 & \text{si } x \leq -e ; \\ -(1/4) \ln(-x) + 1/4, & \text{si } -e < x \leq -1 ; \\ x^2/2 + 5x/4 + 1, & \text{si } -1 < x \leq 0 ; \\ 1, & \text{si } x \geq 0. \end{cases}$$

2. Pour calculer $\mathbb{E}(X)$ on écrit que

$$\mathbb{E}(X) = \int_{-\infty}^{+\infty} xf(x)dx = \int_{-e}^{-1} -\frac{x}{4x} dx + \int_{-1}^0 (x^2 + 5x/4) dx = -(6e + 1)/24 .$$

Corrigé de l'exercice 3.17, page 59

1. Écrivons la fonction de répartition de $Y = X^3$. Soit $y \in \mathbb{R}$,

$$\begin{aligned} F_Y(y) &= \mathbb{P}(Y \leq y) = \mathbb{P}(X^3 \leq y) = \mathbb{P}(X \leq y^{1/3}) = \int_{-\infty}^{y^{1/3}} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} dt \\ &= \int_{-\infty}^y \underbrace{\frac{1}{\sqrt{2\pi}} e^{-z^2/2} \frac{1}{3} z^{-2/3}}_{:=f(z)} dz , \end{aligned}$$

où on a effectué le changement de variables $t = z^{1/3}$. Par suite $F_Y(y) = \int_{-\infty}^y f(z)dz$ et X^3 est une v.a.r. à de densité f donnée ci-dessus.

2. La fonction de répartition de la loi normale centrée réduite est une application continue, strictement croissante donc bijective de \mathbb{R} sur $]0,1[$. Ainsi F^{-1} existe bien et on peut écrire pour tout $y \in]0,1[$

$$\mathbb{P}(Y \leq y) = \mathbb{P}(F(X) \leq y) = \mathbb{P}(X \leq F^{-1}(y)) = F(F^{-1}(y)) = y .$$

Comme F prend ses valeurs dans l'intervalle $]0,1[$, l'ensemble $\{F(X) \leq y\} = \emptyset$ dès que $y \leq 0$ et donc $F_Y(y) = \mathbb{P}(F(X) \leq y) = 0$ quand $y \leq 0$.

Pour la même raison, $\{F(X) \leq y\} = \Omega$ dès que $y \geq 1$ et dans ce cas $F_Y(y) = \mathbb{P}(\Omega) = 1$.

En conclusion on a

$$F_Y(y) = \begin{cases} 0, & \text{si } y \leq 0 ; \\ y, & \text{si } 0 < y < 1 ; \\ 1, & \text{si } y \geq 1. \end{cases}$$

donc Y est une v.a.r. uniforme sur $[0,1]$.

Corrigé de l'exercice 3.18, page 60

1. L'application f est bien positive, continue sur \mathbb{R} et on a $\int_{-\infty}^{+\infty} f(t)dt = \int_0^{\infty} \alpha e^{-\alpha u} du = 1$.
2. La v.a.r. $Y = |X|$ ne prend que des valeurs positives. On procède comme dans l'exercice 3.17, page 59. Soit $y \in \mathbb{R}^+$,

$$\begin{aligned} F_Y(y) &= \mathbb{P}(Y \leq y) = \mathbb{P}(|X| \leq y) = \mathbb{P}(-y \leq X \leq y) = \int_{-y}^y f(t)dt \\ &= \int_0^y \alpha e^{-\alpha t} dt = 1 - e^{-\alpha y}. \end{aligned}$$

Si maintenant $y < 0$, $\{T \leq y\} = \{|X| \leq y\} = \emptyset$ et donc $F_Y(y) = \mathbb{P}(\emptyset) = 0$. Finalement, $F_Y(y) = (1 - e^{-\alpha y})\mathbb{1}_{\mathbb{R}^+}(y)$ et donc Y est une v.a.r. exponentielle de paramètre $\alpha > 0$.

D'après le formulaire de l'annexe A, page 205, on sait que $\mathbb{E}(Y) = 1/\alpha$ et $\text{Var}(Y) = 1/\alpha^2$. Par la formule de König (proposition 3.22, page 46), $\mathbb{E}(Y^2) = \text{Var}(Y) + (\mathbb{E}(Y))^2 = 2/\alpha^2$. Or $Y^2 = X^2$ et donc $\mathbb{E}(X^2) = 2/\alpha^2$. On vérifie que $\mathbb{E}(X) = 0$ et on conclut que $\text{Var}(X) = \mathbb{E}(X^2) = 2/\alpha^2$.

Corrigé de l'exercice 3.19, page 60

Remarquons que si $\omega \in \Omega$ est tel que $X(\omega) = 1$, alors $Y(\omega)$ n'est pas définie. Mais comme $\mathbb{P}(\{X = 1\}) = 0$, la probabilité que Y soit définie est 1. On remarque que Y prend ses valeurs dans \mathbb{R}^+ . Soit $y > 0$, $F_Y(y) = \mathbb{P}(Y \leq y) = \mathbb{P}(X \leq 1 - e^{-\alpha y}) = 1 - e^{-\alpha y}$ car $0 \leq 1 - e^{-\alpha y} \leq 1$ et x suit une loi uniforme sur $[0, 1]$. Comme pour $y \leq 0$, $\mathbb{P}(Y \leq y) = 0$ on reconnaît que Y suit la loi exponentielle de paramètre $\alpha > 0$.

Corrigé de l'exercice 3.20, page 60

Si on considère la v.a. $Y := \frac{X - m}{\sigma}$ alors Y suit la loi normale $\mathcal{N}_1(0, 1)$, d'après le théorème de standardisation. D'où, pour tout réel u , $\phi_X(u) = \mathbb{E}[e^{iu(\sigma Y + m)}] = e^{ium} \phi(\sigma u)$ avec

$$\phi(t) = \mathbb{E}(e^{iYt}) = \int_{\mathbb{R}} e^{ixt} d\mathbb{P}_Y(x) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-\frac{x^2}{2}} e^{ixt} d\lambda(x).$$

Pour calculer cette intégrale, on va montrer que ϕ vérifie une équation différentielle.

$$\begin{aligned} \phi(t) &= \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-\frac{x^2}{2}} (\cos(tx) + i \sin(tx)) d\lambda(x) \\ &= \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-\frac{x^2}{2}} \cos(tx) d\lambda(x). \end{aligned}$$

Comme $|-xe^{-\frac{x^2}{2}} \cos(tx)| \leq |x|e^{-\frac{x^2}{2}}$ et que cette dernière fonction est intégrable sur \mathbb{R} pour la mesure de Lebesgue, par le théorème de dérivation sous le signe \int de la théorie de Lebesgue, on en déduit que ϕ est dérivable et que

$$\begin{aligned} \phi'(t) &= -\frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} x e^{-\frac{x^2}{2}} \sin(tx) d\lambda(x) \\ &= \frac{1}{\sqrt{2\pi}} \left[e^{-\frac{x^2}{2}} \sin(tx) \right]_{-\infty}^{+\infty} - \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} t e^{-\frac{x^2}{2}} \cos(tx) d\lambda(x) \\ &= -t\phi(t). \end{aligned}$$

Car $\int_{\mathbb{R}} e^{-\frac{x^2}{2}} \sin(tx) d\lambda(x) = 0$ comme intégrale par rapport à la mesure de Lebesgue, sur un intervalle centré en 0 d'une fonction impaire. Comme Φ est une f.c., elle prend nécessairement la valeur 1 au point 0. Donc Φ est la solution particulière de l'équation différentielle $y' + ty = 0$ telle que $y(0) = 1$. La solution particulière de cette équation différentielle qui prend la valeur 1 en 0 est $\Phi(t) = e^{-\frac{t^2}{2}}$. Par suite, en revenant au début du corrigé $\Phi_X(u) = \mathbb{E}[e^{iu(\sigma Y + m)}] = e^{ium} \Phi(\sigma u) = e^{ium} e^{-\frac{\sigma^2 u^2}{2}} = \exp\left(ium - \frac{\sigma^2 u^2}{2}\right)$.

Corrigé de l'exercice 3.21, page 60

1. Soit A un événement. On a $\mathbb{P}(X \in A) = \mathbb{E}(\mathbb{1}_A(X))$ et $\mathbb{P}(Y \in A) = \mathbb{E}(\mathbb{1}_A(Y))$. Comme $X = Y$ presque-sûrement, on a $\mathbb{1}_A(X) = \mathbb{1}_A(Y)$ presque-sûrement. Par suite $\mathbb{P}(X \in A) = \mathbb{P}(Y \in A)$, c-à-d $\mathbb{P}_X = \mathbb{P}_Y$.

La réciproque est fausse. Pour montrer cela donnons deux contre-exemples, un dans le cas discret, un dans le cas non discret.

Pour le cas discret, considérons le jeu de Pile-ou-Face modélisé de la façon suivante : $\Omega := \{P, F\}$, $\mathcal{F} := \mathcal{P}(\{P, F\})$, \mathbb{P} est définie sur \mathcal{F} à partir de $\mathbb{P}(\{F\}) = \mathbb{P}(\{P\}) = \frac{1}{2}$, considérons les v.a.r. $X := \mathbb{1}_A$ et $Y := \mathbb{1}_{A^c}$. On vérifie aisément que ces deux v.a.r. ont pour loi $\frac{1}{2}(\delta_0 + \delta_1)$ (il s'agit d'une loi de Bernoulli), cependant ces v.a.r. ne sont pas presque-sûrement égales.

Pour le cas non discret, considérons une v.a.r. X de loi $\mathcal{N}_1(0, 1)$ et posons $Y := -X$. On a alors, en utilisant le théorème du transfert,

$$\begin{aligned} \mathbb{P}(Y \in A) &= \mathbb{E}(\mathbb{1}_A(-X)) = \int_{\Omega} \mathbb{1}_A(-X) d\mathbb{P} = \int_{\mathbb{R}} \mathbb{1}_A(-x) d\mathbb{P}_X(x) \\ &= \int_{-\infty}^{+\infty} \mathbb{1}_A(-x) \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) dx = \int_{-\infty}^{+\infty} \mathbb{1}_A(t) \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t^2}{2}\right) dt \\ &= \mathbb{P}(X \in A). \end{aligned}$$

X et Y ont donc la même loi, bien qu'elles ne soient pas presque-sûrement-égales car $\mathbb{P}(X = Y) = \mathbb{P}(X = -X) = \mathbb{P}(X = 0) = 0$.

2. Soit h une application positive borélienne de \mathbb{R} dans \mathbb{R} .

$$\begin{aligned} \mathbb{E}[h(g(X))] &= \int_{\Omega} h(g(X)) d\mathbb{P} = \int_{\mathbb{R}} h(g(x)) d\mathbb{P}_X(x) \\ &= \int_{\mathbb{R}} h(g(x)) d\mathbb{P}_Y(x) = \int_{\Omega} h(g(Y)) d\mathbb{P} = \mathbb{E}[h(g(Y))], \end{aligned}$$

d'où le résultat cherché. On peut aussi remarquer que $h \circ g$ est une application positive borélienne de \mathbb{R} dans \mathbb{R} . On applique alors l'équivalence $\mathbb{P}_X = \mathbb{P}_Y$ si, et seulement si, pour toute application k positive borélienne de \mathbb{R} dans \mathbb{R} , $\mathbb{E}(k(X)) = \mathbb{E}(k(Y))$ avec $k = h \circ g$.

Montrons par un contre-exemple qu'on peut avoir $\mathbb{P}_X = \mathbb{P}_Y$ et $\mathbb{P}_{XZ} \neq \mathbb{P}_{YZ}$. Considérons le jeu de Pile-ou-Face précédent et avec les notations précédentes posons $Z = X$. On a alors $XZ = X^2 = X$ et $YZ = YX = 0$. Par suite $\mathbb{P}_X = \mathbb{P}_{XZ} \neq \mathbb{P}_{YZ} = \delta_0$ bien que $\mathbb{P}_X = \mathbb{P}_Y$.

Corrigé de l'exercice 3.22, page 60

D'une façon générale, par le théorème du transfert, on a $\Phi_X(t) = \mathbb{E}[e^{itX}] = \int_{\mathbb{R}} e^{itx} d\mathbb{P}_X(x)$.

1. $\Phi_X(t) = \int_{\mathbb{R}} e^{itx} \sum_{k=1}^{+\infty} pq^{k-1} d\delta_k(x) = \sum_{k=1}^{+\infty} pq^{k-1} e^{itk} = \frac{pe^{it}}{1 - qe^{it}}.$
2. $\Phi_Y(t) = \frac{1}{b-a} \int_{\mathbb{R}} e^{itx} \mathbb{1}_{[a,b]}(x) d\lambda(x) = \frac{1}{b-a} \frac{e^{ita} - e^{itb}}{it}.$
3. $\Phi_Z(t) = \int_{\mathbb{R}} e^{itx} \alpha e^{-\alpha x} \mathbb{1}_{[0,+\infty[}(x) d\alpha^{(1)}(x) = \left[\frac{\alpha}{it - \alpha} e^{(it-\alpha)x} \right]_0^{+\infty} = \frac{\alpha}{\alpha - it}.$

Corrigé de l'exercice 3.23, page 60

Montrons que, pour tout $t \in \mathbb{R}$, $F_{\mu}(t) \in \{0, 1\}$. En effet, par densité de D dans \mathbb{R} si $t \in \mathbb{R}$, il existe une suite décroissante $(t_n)_{\mathbb{N}}$ de réels de D convergeant vers t . Par continuité à droite de la fonction de répartition F_{μ} , la suite $(F_{\mu}(t_n))_{\mathbb{N}}$ converge vers $F_{\mu}(t)$. Comme pour tout $n \in \mathbb{N}$ $F_{\mu}(t_n) \in \{0, 1\}$, par passage à la limite $F_{\mu}(t) \in \{0, 1\}$.

Considérons $a := \inf\{t \in \mathbb{R} / F_{\mu}(t) = 1\}$. On ne peut pas avoir $a = +\infty$ sinon, pour tout $t \in \mathbb{R}$, $F_{\mu}(t) = 0$, ce qui contredirait le fait que $\lim_{t \rightarrow +\infty} F_{\mu}(t) = 1$. De même on ne peut pas avoir $a = -\infty$ sinon, pour tout $t \in \mathbb{R}$, $F_{\mu}(t) = 1$, ce qui contredirait le fait que $\lim_{t \rightarrow -\infty} F_{\mu}(t) = 0$.

a est donc un réel et la continuité à droite de F_{μ} implique que $F_{\mu}(a) = 1$. Enfin la propriété de croissance des fonctions de répartition implique que $F_{\mu} = \mathbb{1}_{[a, +\infty[}$, c'est-à-dire que μ est la mesure de Dirac au point a .

Ce résultat s'applique en particulier à l'ensemble D des points où la fonction de répartition est continue, puisque son complémentaire dans \mathbb{R} est un ensemble dénombrable.

8.4 Corrigés des exercices du chapitre IV

Corrigé de l'exercice 4.1, page 65

1. En posant $f(x, y) = \mathbb{1}_{]0, \infty[}(x) \mathbb{1}_{]0, \infty[}(y) e^{-x^2} e^{-y^2} = h(x) \times g(y)$ avec $h(x) = \mathbb{1}_{]0, \infty[}(x) e^{-x^2}$ et $g(y) = \mathbb{1}_{]0, \infty[}(y) e^{-y^2}$, et en utilisant l'exemple 4.3, page 64, on a

$$\begin{aligned} \frac{\pi}{4} &= \int_{\mathbb{R}^2} f(x, y) d\lambda^{(2)}(x, y) = \int_{\mathbb{R}} g(y) \left(\int_{\mathbb{R}} h(x) d\lambda(x) \right) d\lambda(y) \\ &= \left[\int_{\mathbb{R}} g(y) d\lambda(y) \right] \times \left[\int_{\mathbb{R}} h(x) d\lambda(x) \right] \\ &= \left[\int_0^{\infty} e^{-y^2} dy \right] \times \left[\int_0^{\infty} e^{-x^2} dx \right]. \end{aligned}$$

Comme les deux intégrales ci-dessus sont égales, on obtient $\int_0^{\infty} e^{-x^2} dx = \sqrt{\pi}/2$. L'autre égalité est immédiate après un changement de variable.

2. D'après la remarque de l'énoncé,

$$\int_{\mathbb{R}^2} e^{-(x^2+2xy+2y^2)} d\lambda^{(2)}(x, y) = \int_{\mathbb{R}^2} e^{-(x+y)^2+y^2} d\lambda^{(2)}(x, y)$$

Effectuons le changement de variables T de classe \mathcal{C}^1 défini par $T : (x, y) \in \mathbb{R}^2 \rightarrow T(x, y) = (u, v) \in \mathbb{R}^2$ avec $u = x + y$ et $v = y$. Le jacobien de T^{-1} au point (u, v)

vaut 1. Le théorème de changement de variable donne :

$$\begin{aligned} \int_{\mathbb{R}^2} e^{-(x+y)^2+y^2} d\lambda^{(2)}(x, y) &= \int_{\mathbb{R}^2} e^{-(u^2+v^2)} d\lambda^{(2)}(u, v) \\ &= \left[\int_{\mathbb{R}} e^{-u^2} d\lambda(u) \right] \times \left[\int_{\mathbb{R}} e^{-v^2} d\lambda(v) \right] \\ &= \left(\int_{-\infty}^{+\infty} e^{-x^2} dx \right)^2 = \left(2 \int_0^{+\infty} e^{-x^2} dx \right)^2 = 4 \left(\frac{\sqrt{\pi}}{2} \right)^2 = \pi, \end{aligned}$$

où l'on a utilisé le théorème de Tonelli et le résultat de la première question.

Corrigé de l'exercice 3.4, page 36

Vérifions que ν est une probabilité sur $\mathcal{B}(\mathbb{R}^d)$ que l'on notera \mathcal{B} dans cette correction.

- L'application $\mathbb{1}_A \rho$ est borélienne et positive pour tout $A \in \mathcal{B}$ donc $\nu(A) \in [0, \infty]$.
- La fonction $\mathbb{1}_\emptyset$ est la fonction-nulle sur \mathbb{R}^d donc $\nu(\emptyset) = \int_{\mathbb{R}^d} \mathbb{1}_\emptyset \rho d\lambda^{(d)} = 0$.
- Par définition de la densité ρ , $\nu(\mathbb{R}^d) = \int_{\mathbb{R}^d} \mathbb{1}_{\mathbb{R}^d} \rho d\lambda^{(d)} = \int_{\mathbb{R}^d} \rho d\lambda^{(d)} = 1$.
- Il reste à montrer la σ -additivité. Soit $(A_k)_{k \in \mathbb{N}}$ une suite d'éléments de \mathcal{B} deux à deux disjoints. Alors si on note $A = \bigcup_{k=1}^{\infty} A_k$, $\mathbb{1}_A = \sum_{k=0}^{\infty} \mathbb{1}_{A_k}$ car les A_k sont deux à deux disjoints. On a alors en utilisant le théorème de Beppo-Levi

$$\nu(A) = \int_{\mathbb{R}^d} \mathbb{1}_A \rho d\lambda^{(d)} = \int_{\mathbb{R}^d} \rho \sum_{k=0}^{\infty} \mathbb{1}_{A_k} d\lambda^{(d)} = \sum_{k=0}^{\infty} \int_{\mathbb{R}^d} \mathbb{1}_{A_k} \rho d\lambda^{(d)} = \sum_{k=0}^{\infty} \nu(A_k),$$

ce qui donne la σ -additivité.

ν est donc bien une mesure de probabilité.

Corrigé de l'exercice 4.2, page 66

Considérons un dé équilibré à 6 faces. On lance le dé deux fois de suite. Soit X le résultat obtenu au premier lancer et Y celui obtenu au second.

- Les v.a.r. X et Y ont la même loi. En effet elles sont à valeurs dans $\{1, 2, 3, 4, 5, 6\}$ et pour $k = 1, \dots, 6$, $\mathbb{P}(X = k) = \mathbb{P}(Y = k) = 1/6$, c'est-à-dire que $\mathbb{P}_X = \mathbb{P}_Y$.
- En revanche, $X \neq Y$ car l'égalité signifierait que, à chaque double lancer, le résultat du second lancer est toujours le même que celui du premier lancer ce qui est faux.

Corrigé de l'exercice 4.3, page 66

Soit (X, Y, Z) un vecteur aléatoire de densité ρ sur \mathbb{R}^3 . Cherchons la loi de Y (ce serait la même chose pour X et Z). Soit A un borélien de \mathbb{R} , $\{Y \in A\} = \{(X, Y, Z) \in \mathbb{R} \times A \times \mathbb{R}\}$ d'où

$$\begin{aligned} \mathbb{P}_Y(A) &= \mathbb{P}(Y \in A) = \mathbb{P}((X, Y, Z) \in \mathbb{R} \times A \times \mathbb{R}) = \mathbb{P}_{(X, Y, Z)}(\mathbb{R} \times A \times \mathbb{R}) \\ &= \int_{\mathbb{R}^3} \mathbb{1}_{\mathbb{R}}(x) \mathbb{1}_A(y) \mathbb{1}_{\mathbb{R}}(z) d\mathbb{P}_{(X, Y, Z)}(x, y, z) \end{aligned}$$

et par application des règles d'intégration des mesures à densité (proposition 3.8, page 35,) et du théorème de Tonelli, on obtient comme dans l'exemple 4.4, page 65, que

$$\mathbb{P}_Y(A) = \int_{\mathbb{R}} \mathbb{1}_A(y) \underbrace{\left(\int_{\mathbb{R}^2} \mathbb{1}_{\mathbb{R}}(x) \mathbb{1}_{\mathbb{R}}(z) \rho(x, y, z) d\lambda^{(2)}(x, z) \right)}_{:=\chi(y)} d\lambda(x) = \int_{\mathbb{R}} \mathbb{1}_A(y) \chi(y) d\lambda(y)$$

ce qui prouve que la v.a.r. Y admet pour densité la fonction χ définie ci-dessus.

Corrigé de l'exercice 4.1, page 66

Tout d'abord, comme l'événement $\{(X, X) \in \Delta\}$ est certain, on a $\mathbb{P}((X, X) \in \Delta) = 1$. Supposons que ρ soit une densité pour le vecteur (X, X) . Un autre calcul donnerait alors :

$$\begin{aligned}\mathbb{P}((X, X) \in \Delta) &= \mathbb{P}_{(X, X)}(\Delta) = \int_{\mathbb{R}^2} \mathbb{1}_{\Delta}(x, y) \rho(x, y) d\lambda^{(2)}(x, y) \\ &= \int_{\mathbb{R}^2} \mathbb{1}_{\Delta}(x, y) \rho(x, x) d\lambda^{(2)}(x, y) \\ &= \int_{\mathbb{R}} \rho(x, x) \left(\int_{\mathbb{R}} \mathbb{1}_{\Delta}(x, y) d\lambda(y) \right) d\lambda(x) .\end{aligned}$$

Or $\int_{\mathbb{R}} \mathbb{1}_{\Delta}(x, y) d\lambda(y) = \int_{\mathbb{R}} \mathbb{1}_{\{x\}}(y) d\lambda(y) = \lambda(\{x\}) = 0$ pour tout $x \in \mathbb{R}$. Par suite $\mathbb{P}((X, X) \in \Delta) = 0$ ce qui contredit le premier calcul et fournit ainsi un contre-exemple pour la réciproque de la proposition 4.5, page 66.

Corrigé de l'exercice 4.5, page 68

1. Soit h une application borélienne positive de \mathbb{R}^2 dans \mathbb{R}^+ . En utilisant successivement le théorème du transfert, l'indépendance de U et V , les règles d'intégration par rapport à des mesures à densité, on obtient :

$$\begin{aligned}\mathbb{E}(h(X, Y)) &= \mathbb{E}\left(h(\sqrt{-2 \ln U} \cos(2\pi V), \sqrt{-2 \ln U} \sin(2\pi V))\right) \\ &= \int_{\mathbb{R}^2} h(\sqrt{-2 \ln u} \cos(2\pi v), \sqrt{-2 \ln u} \sin(2\pi v)) d\mathbb{P}_{(U, V)}(u, v) \\ &= \int_{\mathbb{R}^2} h(\sqrt{-2 \ln u} \cos(2\pi v), \sqrt{-2 \ln u} \sin(2\pi v)) d\mathbb{P}_U \otimes d\mathbb{P}_V(u, v) \\ &= \int_{\mathbb{R}^2} h(\sqrt{-2 \ln u} \cos(2\pi v), \sqrt{-2 \ln u} \sin(2\pi v)) \mathbb{1}_{]0,1[}(u) \mathbb{1}_{]0,1[}(v) d\lambda^{(2)}(u, v) .\end{aligned}$$

Effectuons le changement de variable $T :]0,1[^2 \rightarrow \mathbb{R}^2$ défini pour $(u, v) \in \mathbb{R}^2$ par

$$\begin{cases} x = \sqrt{-2 \ln u} \cos(2\pi v) \\ y = \sqrt{-2 \ln u} \sin(2\pi v) \end{cases} \iff \begin{cases} u = \exp(-(x^2 + y^2)/2) \\ v = \frac{1}{2\pi} \arctan(y/x) \end{cases} .$$

Le jacobien de T^{-1} est

$$J_{T^{-1}} = \det \begin{pmatrix} -xe^{-\frac{x^2+y^2}{2}} & -ye^{-\frac{x^2+y^2}{2}} \\ -\frac{1}{2\pi} \frac{y}{x^2+y^2} & \frac{1}{2\pi} \frac{x}{x^2+y^2} \end{pmatrix} = -\frac{1}{2\pi} e^{-\frac{x^2+y^2}{2}} .$$

On obtient par le théorème de changement de variable :

$$\mathbb{E}(h(X, Y)) = \int_{\mathbb{R}^2} h(x, y) \frac{1}{2\pi} e^{-\frac{x^2+y^2}{2}} d\lambda^{(2)}(x, y) ,$$

et donc la v.a. (X, Y) admet bien la densité ρ annoncée.

2. Pour trouver la loi de la v.a.r. X (idem pour Y), on applique la proposition 4.5, page 66, qui nous permet d'affirmer que X est une v.a.r. de densité ρ_X définie par

$$\begin{aligned}\rho_X(x) &= \int_{\mathbb{R}} \rho(x, y) d\lambda(y) = \int_{\mathbb{R}} \frac{1}{2\pi} e^{-x^2/2} e^{-y^2/2} dy \\ &= \frac{1}{2\pi} e^{-x^2/2} \underbrace{\int_{\mathbb{R}} e^{-y^2/2} dy}_{=\sqrt{2\pi}} = \frac{1}{\sqrt{2\pi}} e^{-x^2/2},\end{aligned}$$

donc X est une v.a.r. normale centrée réduite (idem pour Y).

3. Pour tout A, B boréliens de \mathbb{R}

$$\begin{aligned}\mathbb{P}_{(X,Y)}(A \times B) &= \int_{\mathbb{R}^2} \mathbf{1}_A(x) \mathbf{1}_B(y) \frac{1}{2\pi} e^{-\frac{x^2+y^2}{2}} d\lambda^{(2)}(x, y) \\ &= \left(\int_{\mathbb{R}} \mathbf{1}_A(x) \frac{1}{\sqrt{2\pi}} e^{-x^2/2} d\lambda(x) \right) \times \left(\int_{\mathbb{R}} \mathbf{1}_B(y) \frac{1}{\sqrt{2\pi}} e^{-y^2/2} d\lambda(y) \right) \\ &= \mathbb{P}_X(A) \times \mathbb{P}_Y(B)\end{aligned}$$

et ceci pour tout boréliens A et B ce qui prouve que $\mathbb{P}_{(X,Y)} = \mathbb{P}_X \times \mathbb{P}_Y$ et donc (X, Y) est un couple indépendant.

Corrigé de l'exercice 4.6, page 74

Étudions la fonction de répartition de la v.a.r. Y . Pour tout $y \in \mathbb{R}$, $F_Y(y) = \mathbb{P}(Y \leq y)$. On peut exprimer l'événement $\{Y \leq y\}$ en fonction des v.a.r. ε et X :

$$\{Y \leq y\} = [\{\varepsilon = 1\} \cap \{X \leq y\}] \cup [\{\varepsilon = -1\} \cap \{X \geq -y\}]$$

Comme (ε, X) indépendants, les événements $\{\varepsilon = 1\}$ et $\{X \leq y\}$ sont indépendants (il en est de même pour $\{\varepsilon = -1\}$ et $\{X \geq -y\}$). Les événements entre crochets étant disjoints, on obtient :

$$\begin{aligned}F_Y(y) &= \mathbb{P}(\varepsilon = 1) \times \mathbb{P}(X \leq y) + \mathbb{P}(\varepsilon = -1) \times \mathbb{P}(X \geq -y) \\ &= \frac{1}{2} \int_{-\infty}^y \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx + \frac{1}{2} \int_{-y}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx = \int_{-\infty}^y \frac{1}{\sqrt{2\pi}} e^{-x^2/2} dx\end{aligned}$$

et donc Y est une v.a.r. de loi $\mathcal{N}(0; 1)$.

Pour montrer que X et Y sont non-corrélés, il suffit de calculer

$$\mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])] = \mathbb{E}[\varepsilon X^2] = \mathbb{E}[\varepsilon] \mathbb{E}[X^2] = 0$$

du fait que (X, ε) est indépendant et que ε est centrée. Si (X, Y) était indépendant, alors (X^2, Y^2) le serait aussi et on aurait

$$\mathbb{E}[X^2 Y^2] = \mathbb{E}[X^2] \mathbb{E}[Y^2].$$

Cependant, du fait que $Y^2 = X^2$,

$$\mathbb{E}[X^2 Y^2] = \mathbb{E}[X^2] \mathbb{E}[Y^2] = (\mathbb{E}[X^2])^2 = 1.$$

Mais, par ailleurs,

$$\mathbb{E}[X^2 Y^2] = \mathbb{E}[X^4] = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} x^4 e^{-\frac{1}{2}x^2} d\lambda^{(1)}(x) = 3,$$

où la dernière intégrale est calculée par une intégration par parties ; ce qui montre une contradiction. Le couple de v.a.r. (X, Y) n'est donc pas indépendant.

Corrigé de l'exercice 4.7, page 74

D'après les formules de König-Huygens (Proposition 3.22, page 46), $\text{Var}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2$. D'où $\mathbb{E}(X^2) = \text{Var}(X) + (\mathbb{E}(X))^2 = \sigma^2 + m^2$ (idem pour Y). Par indépendance de (X, Y) , $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$ et donc on a

$$\begin{aligned} \mathbb{E}((X + Y)^2) &= \mathbb{E}(X^2 + 2XY + Y^2) = \mathbb{E}(X^2) + 2\mathbb{E}(XY) + \mathbb{E}(Y^2) \\ &= \mathbb{E}(X^2) + 2\mathbb{E}(X)\mathbb{E}(Y) + \mathbb{E}(Y^2) = 4m^2 + 2\sigma^2. \end{aligned}$$

Corrigé de l'exercice 4.8, page 85

On reprend les notations de l'exercice 4.6, page 74. On a

$$\begin{aligned} \text{Var}(X + Y) &= \text{Var}(X) + 2 \underbrace{\text{Cov}(X, Y)}_{=0} + \text{Var}(Y) \\ &= \text{Var}(X) + \text{Var}(Y) \end{aligned}$$

bien que (X, Y) ne soit pas indépendant.

Corrigé de l'exercice 4.9, page 86

- On a $\phi_{X+Y}(t) = \phi_{2X}(t) = \mathbb{E}[e^{2itX}] = e^{-2|t|} = (e^{-|t|})^2 = (\phi_X(t))^2 = \phi_X(t) \times \phi_Y(t)$.
- Pour montrer que (X, Y) est non indépendant, on prend (comme le recommande l'énoncé) $A = [-1, 1]$ et on calcule

$$\mathbb{P}_X(A) = \int_{-1}^1 \frac{1}{\pi} \frac{1}{1+x^2} dx = \frac{1}{\pi} [\arctan(x)]_{-1}^1 = \frac{1}{\pi} \left(\frac{\pi}{4} + \frac{\pi}{4} \right) = \frac{1}{2}.$$

On a alors $\mathbb{P}_{(X,Y)}(A \times A^c) = \mathbb{P}((X, X) \in A \times A^c) = 0$ et $\mathbb{P}(A) \times \mathbb{P}(A^c) = \mathbb{P}(A)(1 - \mathbb{P}(A)) = 1/4 \neq 0$. Donc X et Y ne sont pas indépendants.

Corrigé de l'exercice 4.10, page 86

Si X_k est une v.a.r. de Bernoulli de paramètre $p \in]0, 1[$, sa fonction caractéristique est $\phi_{X_k}(t) = pe^{it} + (1-p)$. Par indépendance de X_1, \dots, X_n on a

$$\phi_{S_n}(t) = \phi_{X_1}(t) \times \dots \times \phi_{X_n}(t) = [pe^{it} + (1-p)]^n$$

et on reconnaît la fonction caractéristique d'une loi $\mathcal{B}(n; p)$. Donc S_n est une v.a.r. $\mathcal{B}(n; p)$.

Corrigé de l'exercice 4.11, page 86

1. Pour tout $\omega \in \Omega$, $X_{(1)}(\omega) = \max(X_1(\omega), \dots, X_n(\omega))$ car c'est le premier nombre dans la suite $X_1(\omega), \dots, X_n(\omega)$ réordonnés de façon croissante. Soit $x \in \mathbb{R}$, $F_{X_{(1)}}(x) = \mathbb{P}(X_{(1)} \leq x)$ or $\{\max(X_1, \dots, X_n) \leq x\} = \cap_{k=1}^n \{X_k \leq x\}$ et par indépendance de X_1, \dots, X_n on déduit que

$$F_{X_{(1)}}(x) = \mathbb{P}(\cap_{k=1}^n \{X_k \leq x\}) = \prod_{k=1}^n \mathbb{P}(X_k \leq x) = (F(x))^n$$

où F est la fonction de répartition des v.a.r. X_i , c'est à dire que $F(x) = (1 - e^{-\alpha x})\mathbb{1}_{]0, \infty[}(x)$. D'où $F_{X_{(1)}}(x) = (1 - e^{-\alpha x})^n \mathbb{1}_{]0, \infty[}(x)$.

2. On vérifie de même que $X_{(n)} = \min(X_1, \dots, X_n)$. Soit $x \in \mathbb{R}$, $\{X_{(n)} \leq x\} = \{X_{(n)} > x\}^c = \cap_{k=1}^n \{X_k > x\}$. Par suite

$$\begin{aligned} F_{X_{(n)}}(x) &= 1 - \mathbb{P}(X_{(n)} > x) = 1 - \mathbb{P}(\cap_{k=1}^n \{X_k > x\}) = 1 - \prod_{k=1}^n \mathbb{P}(X_k > x) \\ &= 1 - \prod_{k=1}^n (1 - \mathbb{P}(X_k \leq x)) = 1 - \prod_{k=1}^n [1 - F(x)] = 1 - [1 - F(x)]^n \\ &= (1 - e^{-n\alpha x})\mathbb{1}_{]0, \infty[}(x). \end{aligned}$$

3. (a) La v.a.r. Y_k ne prend que deux valeurs, 0 et 1. C'est donc une v.a.r. de Bernoulli de paramètre

$$p = \mathbb{P}(Y_k = 1) = \mathbb{P}(X_k > t) = 1 - \mathbb{P}(X_k \leq t) = 1 - F(t) = 1 - (1 - e^{-\alpha t})$$

d'où $p = e^{-\alpha t}$ et donc $\mathbb{P}_{Y_k} = e^{-\alpha t}\delta_1 + (1 - e^{-\alpha t})\delta_0$, pour $k = 1, \dots, n$. Par suite $Y_1 + \dots + Y_n$ est une somme de v.a.r. de Bernoulli de même paramètre p . Ces v.a.r. sont indépendantes car elles sont de la formes $f_1(X_1), \dots, f_n(X_n)$ avec $f_k = \mathbb{1}_{]t, \infty[}$ (on utilise la proposition 4.15, page 76). La v.a.r. $Y_1 + \dots + Y_n$ est donc une v.a.r. de loi $\mathcal{B}(n; p)$.

- (b) On remarque que $Y_k = 1$ signifie que X_k est strictement supérieur à t . Ainsi $Y_1 + \dots + Y_k$ est égal au nombre de v.a.r. X_i qui sont strictement supérieure à t . Ceci entraîne que $\{Y_1 + \dots + Y_n \leq k - 1\} = \{X_{(k)} \leq t\}$.

4. Soit $t \in \mathbb{R}$, d'après ce qui précède on a :

$$\begin{aligned} F_{X_{(k)}}(t) &= \mathbb{P}(X_{(k)} \leq t) = \mathbb{P}(Y_1 + \dots + Y_n \leq k - 1) = \sum_{j=0}^{k-1} C_n^j p^j (1 - p)^{n-j} \\ &= \sum_{j=0}^{k-1} C_n^j e^{-j\alpha t} (1 - e^{-\alpha t})^{n-j}. \end{aligned}$$

Corrigé de l'exercice 4.12, page 88

On fait la preuve du deuxième point (l'autre preuve est identique, à quelques aménagements évidents près, à celle utilisée dans la démonstration de la proposition 4.34, page 87). On a $\Phi_X(t) = \exp[\alpha(e^{it} - 1)]$ et $\Phi_Y(t) = [\beta(e^{it} - 1)]$. Comme (X, Y) indépendant, la fonction caractéristique de $X + Y$ est donnée pour tout $t \in \mathbb{R}$ par

$$\Phi_{X+Y}(t) = \Phi_X(t) \times \Phi_Y(t) = \exp[(\alpha + \beta)(e^{it} - 1)]$$

et ainsi ϕ_{X+Y} est la fonction caractéristique d'une loi de Poisson de paramètre $\alpha + \beta$.

Corrigé de l'exercice 4.13, page 90

Pour $\omega \in \Omega$ donné, le polynôme $x^2 - 2A(\omega)x + B(\omega)$ admet :

1. deux racines réelles distinctes si, et seulement si, $A^2(\omega) - B(\omega) > 0$.

En utilisant le théorème de transfert, l'indépendance du couple de v.a.r. (A, B) , le théorème de Tonelli, ainsi que le passage de l'intégrale de Lebesgue à celle de Riemann, il vient

$$\begin{aligned} \mathbb{P}(A^2 - B > 0) &= \mathbb{E}[\mathbf{1}_{]0, +\infty[}(A^2 - B)] \\ &= \int_{\mathbb{R}^2} \mathbf{1}_{]0, +\infty[}(x^2 - y) d\mathbb{P}_{(A, B)}(x, y) \\ &= \int_{\mathbb{R}} \left(\int_{\mathbb{R}} \mathbf{1}_{]0, +\infty[}(x^2 - y) \mathbf{1}_{[0, 1]}(x) \mathbf{1}_{[0, 1]}(y) d\lambda(y) \right) d\lambda(x) \\ &= \int_{\mathbb{R}} \mathbf{1}_{[0, 1]}(x) \left(\int_{\mathbb{R}} \mathbf{1}_{]0, +\infty[}(x^2 - y) \mathbf{1}_{[0, 1]}(y) d\lambda(y) \right) d\lambda(x) \\ &= \int_0^1 \left(\int_0^{x^2} dy \right) dx = \int_0^1 x^2 dx = \frac{1}{3}. \end{aligned}$$

2. deux racines complexes et non réelles distinctes si, et seulement si, $A^2(\omega) - B(\omega) < 0$.

En menant le calcul de façon analogue au cas précédent, il vient

$$\begin{aligned} \mathbb{P}(A^2 - B < 0) &= \mathbb{E}[\mathbf{1}_{]-\infty, 0[}(A^2 - B)] \\ &= \int_{\mathbb{R}^2} \mathbf{1}_{]-\infty, 0[}(x^2 - y) d\mathbb{P}_{(A, B)}(x, y) \\ &= \int_{\mathbb{R}} \left(\int_{\mathbb{R}} \mathbf{1}_{]-\infty, 0[}(x^2 - y) \mathbf{1}_{[0, 1]}(x) \mathbf{1}_{[0, 1]}(y) d\lambda(y) \right) d\lambda(x) \\ &= \int_{\mathbb{R}} \mathbf{1}_{[0, 1]}(x) \left(\int_{\mathbb{R}} \mathbf{1}_{]-\infty, 0[}(x^2 - y) \mathbf{1}_{[0, 1]}(y) d\lambda(y) \right) d\lambda(x) \\ &= \int_0^1 \left(\int_{x^2}^1 dy \right) dx = \int_0^1 (1 - x^2) dx = \frac{2}{3}. \end{aligned}$$

3. une racine double si, et seulement si, $A^2(\omega) - B(\omega) = 0$.

En remarquant que $\{A^2 - B = 0\}^c = \{A^2 - B > 0\} \cup \{A^2 - B < 0\}$, il vient

$$\mathbb{P}(A^2 - B = 0) = 1 - \mathbb{P}(A^2 - B < 0) - \mathbb{P}(A^2 - B > 0) = 0.$$

4. Déterminons la loi de la v.a.r. $\Delta := A^2 - B$. Soit h une application borélienne positive de \mathbb{R} dans \mathbb{R} .

$$\mathbb{E}[h(\Delta)] = \int_{[0, 1]^2} h(x^2 - y) d\lambda^{(2)}(x, y) = \int_{[0, 1]^2} h(x^2 - y) d\lambda^{(2)}(x, y).$$

Par le changement de variable de classe \mathcal{C}^1 $y = t^2 - z$ et $x = t$ l'ouvert $]0, 1]^2$ a pour image réciproque l'ouvert

$$\begin{aligned} U := & \{(t, z) \in \mathbb{R}^2 / z \in [0, 1[, t \in]\sqrt{z}, 1]\} \\ & \cup \{(t, z) \in \mathbb{R}^2 / z \in]-1, 0[, t \in]0, \sqrt{1+z}]\}. \end{aligned}$$

Par le théorème de changement de variable (cf. 4.4, page 64) puisque la valeur absolue du jacobien est égale à 1, il vient

$$\int_{]0,1[^2} h(x^2 - y) d\lambda^{(2)}(x, y) = \int_U h(z) d\lambda^{(2)}(t, z).$$

Par suite, en utilisant le théorème de Tonelli, et en remarquant que $\mathbb{1}_U(t, z) = \mathbb{1}_{]0,1[}(z) \mathbb{1}_{] \sqrt{z}, 1[}(t) + \mathbb{1}_{]-1,0[}(z) \mathbb{1}_{]0, \sqrt{1+z}[}(t)$,

$$\begin{aligned} \mathbb{E}[h(\Delta)] &= \\ &= \int_{\mathbb{R}} h(z) \left(\mathbb{1}_{]0,1[}(z) \int_{\mathbb{R}} \mathbb{1}_{] \sqrt{z}, 1[}(t) d\lambda(t) + \mathbb{1}_{]-1,0[}(z) \int_{\mathbb{R}} \mathbb{1}_{]0, \sqrt{1+z}[}(t) d\lambda(t) \right) d\lambda(z) \\ &= \int_{\mathbb{R}} h(z) \left(\mathbb{1}_{]0,1[}(z) (1 - \sqrt{z}) + \mathbb{1}_{]-1,0[}(z) \sqrt{1+z} \right) d\lambda(z). \end{aligned}$$

La loi de Δ est la mesure de probabilité sur \mathbb{R} admettant la fonction

$$\mathbb{1}_{]0,1[}(z)(1 - \sqrt{z}) + \mathbb{1}_{]-1,0[}(z)\sqrt{1+z}$$

pour densité par rapport à la mesure de Lebesgue. On retrouve les résultats des questions précédentes en calculant

$$\begin{aligned} \mathbb{P}(\Delta > 0) &= \int_{\Omega} \mathbb{1}_{\{\Delta > 0\}}(\omega) d\mathbb{P}(\omega) \\ &= \int_{\mathbb{R}} \mathbb{1}_{]0,+\infty[}(z) (\mathbb{1}_{]0,1[}(z)(1 - \sqrt{z}) + \mathbb{1}_{]-1,0[}(z)\sqrt{1+z}) d\lambda(z) \\ &= \int_0^1 1 - \sqrt{z} dz = 1 - \frac{2}{3} = \frac{1}{3}, \end{aligned}$$

$$\begin{aligned} \mathbb{P}(\Delta < 0) &= \int_{\Omega} \mathbb{1}_{\{\Delta < 0\}}(\omega) d\mathbb{P}(\omega) \\ &= \int_{\mathbb{R}} \mathbb{1}_{]-\infty,0[}(z) (\mathbb{1}_{]0,1[}(z)(1 - \sqrt{z}) + \mathbb{1}_{]-1,0[}(z)\sqrt{1+z}) d\lambda(z) \\ &= \int_{-1}^0 \sqrt{1+z} dz = \left[\frac{2}{3}(1+z)^{3/2} \right]_{-1}^0 = \frac{2}{3}, \end{aligned}$$

$$\mathbb{P}(\Delta = 0) = 1 - \mathbb{P}(\Delta > 0) - \mathbb{P}(\Delta < 0) = 0.$$

Corrigé de l'exercice 4.14, page 91

1. La fonction f étant borélienne positive, on peut lui appliquer le théorème de Tonelli et en se ramenant à des intégrales de Riemann on obtient

$$\begin{aligned} \mathbb{P}_{(X,Y)}(\mathbb{R}^2) &= \int_{\mathbb{R}^2} f(x, y) d\lambda^{(2)}(x, y) \\ &= \alpha \int_0^1 (1 - x^2) dx \int_0^{+\infty} y e^{-3y} dy \\ &= \alpha \left[x - \frac{x^3}{3} \right]_0^1 \left(\left[-\frac{1}{3} y e^{-3y} \right]_0^{+\infty} + \left[-\frac{1}{9} e^{-3y} \right]_0^{+\infty} \right) = \alpha \frac{2}{27}. \end{aligned}$$

Comme $\mathbb{P}_{(X,Y)}$ est une mesure de probabilité sur \mathbb{R}^2 , on a $\mathbb{P}_{(X,Y)}(\mathbb{R}^2) = 1$ d'où on déduit $\alpha = \frac{27}{2}$.

2. En utilisant le résultat de l'exercice I-9, on sait que X et Y admettent des densités f_X et f_Y déterminées par :

$$\begin{aligned} f_X(x) &:= \int_{\mathbb{R}} f(x, y) d\lambda(y) = \alpha(1 - x^2) \mathbb{1}_{[0,1]}(x) \frac{1}{9} = \frac{3}{2}(1 - x^2) \mathbb{1}_{[0,1]}(x) \\ f_Y(y) &:= \int_{\mathbb{R}} f(x, y) d\lambda(x) = 9ye^{-3y} \mathbb{1}_{[0,+\infty[}(y). \end{aligned}$$

3. La relation ensembliste $\{0 < X \leq 2\} \cap \{Y \geq 1\} = \{(X, Y) \in]0, 2] \times [1, +\infty[\}$ permet d'écrire

$$\begin{aligned} \mathbb{P}(0 < X \leq 2, Y \geq 1) &= \mathbb{P}((X, Y) \in]0, 2] \times [1, +\infty[) \\ &= \mathbb{P}_{(X,Y)}(]0, 2] \times [1, +\infty[) \\ &= \int_{]0,2] \times [1,+\infty[} f(x, y) d\lambda^{(2)}(x, y) \\ &= \alpha \left[x - \frac{x^3}{3} \right]_0^1 \left(\left[\frac{-1}{3} ye^{-3y} \right]_1^{+\infty} + \left[\frac{-1}{9} e^{-3y} \right]_1^{+\infty} \right) \\ &= 9 \left(\frac{1}{3} e^{-3} + \frac{1}{9} e^{-3} \right) = 4e^{-3}. \end{aligned}$$

4. Il s'agit de calculer $\text{Var}(X)$, $\text{Var}(Y)$ et $\text{Cov}(X, Y)$.

$$\begin{aligned} \mathbb{E}[X] &= \int_{\mathbb{R}} x f_X(x) d\lambda(x) = \int_0^1 \frac{3}{2}(x - x^3) dx = \frac{3}{8} \\ \mathbb{E}[X^2] &= \int_{\mathbb{R}} x^2 f_X(x) d\lambda(x) = \int_0^1 \frac{3}{2}(x^2 - x^4) dx = \frac{1}{5} \\ \mathbb{E}[Y] &= \int_{\mathbb{R}} y f_Y(y) d\lambda(y) = \int_0^{+\infty} 9y^2 e^{-3y} dy = \frac{2}{3} \\ \mathbb{E}[Y^2] &= \int_{\mathbb{R}} y^2 f_Y(y) d\lambda(y) = \int_0^{+\infty} 9y^3 e^{-3y} dy = \frac{2}{3}. \end{aligned}$$

Par suite

$$\begin{aligned} \text{Var}(X) &= \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \frac{19}{320} \\ \text{Var}(Y) &= \mathbb{E}[Y^2] - (\mathbb{E}[Y])^2 = \frac{2}{9} \\ \text{Cov}(X, Y) &= \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] = 0 \end{aligned}$$

où la dernière égalité est une conséquence directe du théorème de Fubini en remarquant que $xyf(x, y) = (xf_X(x))(yf_Y(y))$ et par suite que $\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y]$. La matrice de dispersion est alors donnée par :

$$D = \begin{pmatrix} \frac{11}{64} & 0 \\ 0 & \frac{2}{9} \end{pmatrix}.$$

Corrigé de l'exercice 4.15, page 91

Sans restreindre la généralité, on peut supposer que $n \leq m$. On a

$$\{X = Y\} = \cup_{k=0}^n [\{X = k\} \cap \{Y = k\}]$$

et la réunion est formée de sous-ensemble de Ω deux à deux disjoints. On a alors

$$\begin{aligned} \mathbb{P}(X = Y) &= \sum_{k=0}^n \mathbb{P}(\{X = k\} \cap \{Y = k\}) \\ &= \sum_{k=0}^n \mathbb{P}(X = k) \times \mathbb{P}(Y = k) \quad \text{par indépendance de } X \text{ et } Y \\ &= \sum_{k=0}^n C_n^k \left(\frac{1}{2}\right)^n C_m^k \left(\frac{1}{2}\right)^m \\ &= \left(\frac{1}{2}\right)^{n+m} \left(\sum_{k=0}^n C_n^k C_m^k\right) = \left(\frac{1}{2}\right)^{n+m} \left(\sum_{k=0}^n C_n^k C_m^{m-k}\right) \end{aligned}$$

et en utilisant l'indication de l'énoncé, on obtient $\mathbb{P}(X = Y) = \frac{1}{2^{n+m}} C_{n+m}^m$.

Corrigé de l'exercice 4.16, page 91

1. Cette relation est obtenue par $(r - 1)$ dérivations successives à partir de la somme de la série entière $\sum_{k \in \mathbb{N}} x^k = \frac{1}{1-x}$.

2. τ_r est une v.a.r. à valeurs dans $\mathbb{N} \cup \{+\infty\}$. En effet, a priori il se peut très bien qu'on ait un ω tel que $\{n \in \mathbb{N}^* / X_1(\omega) + X_2(\omega) + \dots + X_n(\omega) = r\}$ soit vide, dans ce cas par convention, $\tau_r(\omega) := +\infty$.

On sait alors que la loi de τ_r sera de la forme

$$\mathbb{P}_{\tau_r} = \sum_{k \geq 0} \mathbb{P}\{\tau_r = k\} \delta_k + \mathbb{P}(\tau_r = +\infty) \delta_{+\infty}$$

Fixons un entier $k \in \mathbb{N}$ et considérons l'événement $\{\tau_r = k\}$.

Si $k < r$, $\{\tau_r = k\} = \emptyset$, et si $k \geq r$, on peut écrire,

$$\{\tau_r = k\} = \left\{ \sum_{i=1}^{k-1} X_i = r-1 \right\} \cap \{X_k = 1\}.$$

De plus en vertu de l'indépendance de la suite de v.a.r. $(X_i)_{\mathbb{N}}$,

$$\begin{aligned} \mathbb{P}\{\tau_r = k\} &= \mathbb{P}\left(\left\{ \sum_{i=1}^{k-1} X_i = r-1 \right\} \cap \{X_k = 1\}\right) \\ &= \mathbb{P}\left(\left\{ \sum_{i=1}^{k-1} X_i = r-1 \right\}\right) \mathbb{P}(\{X_k = 1\}). \end{aligned}$$

La variable aléatoire réelle $\sum_{i=1}^{k-1} X_i$ est la somme de $(k-1)$ v.a. de Bernoulli indépendantes et de paramètre p , elle suit donc la loi binomiale $\mathcal{B}(k-1, p)$. On trouve alors, en posant pour simplifier les écritures $q := 1 - p$:

$$\mathbb{P}\{\tau_r = k\} = C_{k-1}^{r-1} p^{r-1} q^{(k-1)-(r-1)} p = C_{k-1}^{r-1} p^r q^{k-r}$$

Remarquons que

$$\sum_{k=0}^{+\infty} \mathbb{P}\{\tau_r = k\} = \sum_{k \geq r} C_{k-1}^{r-1} p^r q^{k-r} = \frac{1}{(1-q)^r} p^r = 1$$

où la dernière égalité a été obtenue par application de la relation de la première question. Ce qui donne la loi cherchée pour la variable aléatoire réelle τ_r car $\mathbb{P}(\tau_r = +\infty) = 0$. Il suffit de remarquer pour cela que $\mathbb{P}(\tau_r = +\infty) = 1 - \mathbb{P}\{\tau_r \in \mathbb{N}\}$, et

$$\mathbb{P}\{\tau_r \in \mathbb{N}\} = \sum_{k=0}^{+\infty} \mathbb{P}\{\tau_r = k\} = 1.$$

Il vient alors que $\mathbb{P}\{\tau_r = +\infty\} = 0$. On dit que τ_r est presque-sûrement finie.

3. Remarquons d'abord que $\theta_r = \tau_r - r \geq 0$. Cela prouve que θ_r est une v.a. à valeurs $\mathbb{N} \cup \{+\infty\}$, qu'elle est presque-sûrement finie et que sa loi est donnée par

$$\mathbb{P}_{\theta_r} = \sum_{k \geq 0} \mathbb{P}\{\theta_r = k\} \delta_k$$

où

$$\mathbb{P}\{\theta_r = k\} = \mathbb{P}\{\tau_r = k + r\} = C_{k+r-1}^{r-1} p^r q^k,$$

ce qui donne le résultat cherché.

4. Interprétons ce qui précède de la façon suivante : Appelons **jeu de Pile-ou-Face** l'expérience aléatoire qui consiste à lancer, une infinité de fois, la même pièce de monnaie truquée. On note p la probabilité d'avoir "Face" (ou "succès") et $q = 1 - p$ la probabilité d'avoir "Pile" (ou "échec").

La suite des v.a.r. $(X_k)_{\mathbb{N}}$ représente les valeurs successives obtenues dans un jeu de Pile-ou-Face, en notant $X_k = 1$ quand on a obtenu "Face" lors du $k^{\text{ième}}$ -lancer et par conséquent $X_k = 0$ quand on a obtenu "Pile" lors du $k^{\text{ième}}$ -lancer.

La v.a.r. $X_1 + X_2 + \dots + X_n$ représente alors le nombre de "succès" en n lancers dans ce jeu.

La v.a.r. τ_r représente le rang d'arrivée (ou encore temps d'attente) du $r^{\text{ième}}$ -succès et la v.a.r. θ_r représente le nombre d'"échecs" avant le $r^{\text{ième}}$ -succès dans une infinité de lancers d'une pièce de monnaie.

Si on prend $r = 1$, la variable aléatoire réelle τ_1 est le temps d'attente du premier "succès" dans une infinité de lancers d'une pièce de monnaie, c'est-à-dire une variable aléatoire réelle géométrique de paramètre p . Une variable aléatoire réelle géométrique de paramètre p est une variable aléatoire réelle de Pascal de paramètres $r = 1$ et p .

5. Soit $(\Omega, \mathcal{F}, \mathbb{P})$ un espace probabilisé. Notons A_k l'événement "le fumeur se rendant compte pour la première fois qu'une boîte est vide, l'autre boîte contient k allumettes", G_k l'événement "le fumeur se rendant compte pour la première fois que la boîte Gauche est vide, la boîte droite contient k allumettes", D_k l'événement "le fumeur se rendant compte pour la première fois que la boîte Droite est vide, la boîte gauche contient k allumettes". Nous avons la réunion disjointe $A_k = G_k \cup D_k$ et, pour des raisons de symétrie, $\mathbb{P}(D_k) = \mathbb{P}(G_k)$ d'où $\mathbb{P}(A_k) = 2\mathbb{P}(G_k)$.

Notons X_n la variable aléatoire réelle définie, pour tout $\omega \in \Omega$, par $X_n(\omega) = 0$ si, au $n^{\text{ième}}$ -coup, on cherche l'allumette dans la poche droite et $X_n(\omega) = 1$ si, au $n^{\text{ième}}$ -coup, on cherche l'allumette dans la poche gauche. On suppose l'équiprobabilité de tirer dans la poche gauche ou dans celle de droite, ce que l'on exprime en disant que la variable aléatoire réelle X_n suit une loi de Bernoulli de paramètre $\frac{1}{2}$. La variable aléatoire réelle $S_n := X_1 + X_2 + \dots + X_n$ représente le nombre de fois où le fumeur a puisé dans sa poche gauche. On considérera que le fumeur s'apercevra que la boîte gauche est vide lorsqu'il cherchera une allumette dans cette poche pour la $(N+1)^{\text{ième}}$ -fois. Cela signifie que l'événement G_k est réalisé lorsqu'on extrait $N-k$ allumettes de la boîte droite avant la $(N+1)^{\text{ième}}$ -fois où on puise dans la poche gauche. Par suite avec les notations de la question 2, $G_k = \{\theta_{N+1} = N-k\}$. Il vient alors

$$\mathbb{P}(A_k) = 2\mathbb{P}(G_k) = 2C_N^{2N-k} \left(\frac{1}{2}\right)^{2N-k+1} = C_N^{2N-k} \left(\frac{1}{2}\right)^{2N-k}.$$

Corrigé de l'exercice 4.17, page 92

1. Considérons sur l'espace mesuré $(\Omega \times \mathbb{R}^+, \mathcal{F} \otimes \mathcal{B}(\mathbb{R}^+), \mathbb{P} \otimes \lambda)$, où λ est la mesure de Lebesgue, la fonction à valeurs réelles définie par $H(\omega, t) := nt^{n-1} \mathbb{1}_{[t, +\infty[}(X(\omega))$. H est positive et $(\mathcal{F} \otimes \mathcal{B}(\mathbb{R}^+), \mathcal{B}(\mathbb{R}))$ -mesurable, on peut appliquer le théorème de Tonelli,

$$\begin{aligned} \int_{\Omega \times \mathbb{R}^+} Hd(\mathbb{P} \otimes \lambda) &= \int_{\Omega} \left\{ \int_{\mathbb{R}^+} nt^{n-1} \mathbb{1}_{[0, X(\omega)]}(t) d\lambda(t) \right\} d\mathbb{P}(\omega) \\ &= \int_{\Omega} (X(\omega))^n d\mathbb{P}(\omega) \\ &= \int_{\Omega} X^n d\mathbb{P} = \mathbb{E}[X^n] \end{aligned}$$

et

$$\begin{aligned} \int_{\Omega \times \mathbb{R}^+} Hd(\mathbb{P} \otimes \lambda) &= \int_{\mathbb{R}^+} \left\{ \int_{\Omega} nt^{n-1} \mathbb{1}_{\{X > t\}}(\omega) d\mathbb{P}(\omega) \right\} d\lambda(t) \\ &= \int_{\mathbb{R}^+} nt^{n-1} \mathbb{P}(X > t) d\lambda(t). \end{aligned}$$

Comme l'application $t \mapsto \mathbb{P}(X > t)$ est monotone, son ensemble des points de discontinuités est dénombrable donc de mesure de Lebesgue nulle. Par suite l'application $t \mapsto nt^{n-1} \mathbb{P}(X > t)$ est intégrable au sens de Riemann sur tout compact de \mathbb{R}^+ (cf [2] proposition 0-6 p.9). On obtient le résultat en remarquant que $\mathbb{P}(X > t) = 1 - F_X(t)$.

2. Considérons maintenant une v.a. dont la loi est donnée par $\mathbb{P}_X := \frac{1}{2}(\delta_{-1} + \delta_1)$. Sa fonction de répartition est $F_X(t) = \frac{1}{2}(\mathbb{1}_{[-1, +\infty[}(t) + \mathbb{1}_{[1, +\infty[}(t))$ et

$$\int_0^{+\infty} nt^{n-1}(1 - F(t))dt = \frac{1}{2} \int_0^1 nt^{n-1}dt = \frac{1}{2}$$

alors que $\mathbb{E}[X^n] = \frac{1}{2}((-1)^n + 1)$. La formule n'est vérifiée pour aucune valeur de $n \geq 1$. La condition X positive est bien nécessaire. On peut évidemment calculer les espérances $\mathbb{E}[|X|^n]$ par la formule démontrée ici mais en prenant la fonction de répartition de $|X|$.

3. Comme les trois variables sont positives presque-sûrement, on peut utiliser la relation de la question 1) avec $n = 1$ pour l'espérance et $n = 2$ pour $\mathbb{E}[X^2]$, puis on calcule $\sigma_X^2 = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$.

(a) Pour la variable X on obtient :

$$\begin{aligned}\mathbb{E}[X] &= \int_0^{+\infty} (1 - F_X(t)) dt = \int_0^1 (1 - t) dt + \int_1^{+\infty} 0 dt = \frac{1}{2} \\ \mathbb{E}[X^2] &= \int_0^{+\infty} 2t(1 - F_X(t)) dt = \int_0^1 2t(1 - t) dt = [t^2 - \frac{2}{3}t^3]_0^1 = \frac{1}{3} \\ \sigma_X^2 &= \mathbb{E}[X^2] - (\mathbb{E}[X])^2 = \frac{1}{3} - \frac{1}{4} = \frac{1}{12}.\end{aligned}$$

(b) Pour la variable Y on obtient :

$$\begin{aligned}\mathbb{E}[Y] &= \int_0^{+\infty} (1 - F_Y(t)) dt = \int_0^{+\infty} e^{-\lambda t} dt = \frac{1}{\lambda} \\ \mathbb{E}[Y^2] &= \int_0^{+\infty} 2t(1 - F_Y(t)) dt = \int_0^{+\infty} 2te^{-\lambda t} dt \\ &= \left[-\frac{2te^{-\lambda t}}{\lambda} \right]_0^{+\infty} + \int_0^{+\infty} \frac{2}{\lambda} e^{-\lambda t} dt = \frac{2}{\lambda^2} \\ \sigma_Y^2 &= \mathbb{E}[Y^2] - (\mathbb{E}[Y])^2 = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \frac{1}{\lambda^2}.\end{aligned}$$

- (c) On a $1 - F_Z(t) = \sum_{k=0}^{+\infty} e^{-\lambda} \frac{\lambda^k}{k!} \mathbb{1}_{]-\infty, k[}(t)$. Comme $1 - F_Z(t)$ est la somme d'une série de fonctions positives mesurables, on peut intervertir \int et \sum .

$$\begin{aligned}\mathbb{E}[Z] &= \int_0^{+\infty} (1 - F_Z(t)) dt = \sum_{k=0}^{+\infty} \int_0^k e^{-\lambda} \frac{\lambda^k}{k!} dt \\ &= \sum_{k=1}^{+\infty} e^{-\lambda} \frac{\lambda^k}{(k-1)!} = \lambda \\ \mathbb{E}[Z^2] &= \int_0^{+\infty} 2t(1 - F_Z(t)) dt = \sum_{k=0}^{+\infty} \int_0^k e^{-\lambda} \frac{\lambda^k}{k!} 2t dt = \sum_{k=1}^{+\infty} e^{-\lambda} \frac{\lambda^k}{(k-1)!} k \\ &= \lambda e^{-\lambda} \left[\lambda \sum_{k=2}^{+\infty} \frac{\lambda^{k-2}}{(k-2)!} + \sum_{k=1}^{+\infty} \frac{\lambda^{k-1}}{(k-1)!} \right] = \lambda^2 + \lambda \\ \sigma_Z^2 &= \mathbb{E}[Z^2] - (\mathbb{E}[Z])^2 = \lambda.\end{aligned}$$

Corrigé de l'exercice 4.18, page 92

Pour tout entier $k \geq 1$ on a $\{T = k\} = \{X_k = 1\} \cap \bigcap_{i=1}^{k-1} \{X_i = 0\} \in \mathcal{F}$ et

$\{T = +\infty\} = \bigcap_{i=1}^{+\infty} \{X_i = 0\} \in \mathcal{F}$. T est bien une variable aléatoire à valeurs dans $\overline{\mathbb{N}}^*$. Calculons la loi de T . Pour $k \geq 1$ on a, en utilisant l'indépendance des variables (X_1, \dots, X_k) ,

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

Comme $\mathbb{P}(T < +\infty) = \sum_{k=1}^{+\infty} \mathbb{P}(T = k) = \sum_{k=1}^{+\infty} p(1-p)^{k-1} = 1$, on en déduit que $\mathbb{P}(T = +\infty) = 0$ c-à-d que T est finie p.s. . $\mathbb{P}_T = \sum_{k=1}^{+\infty} p(1-p)^{k-1} \delta_k$ est la loi géométrique $\mathcal{G}(p)$.

Calculons maintenant l'espérance de T . Il vient en utilisant le théorème du transfert

$$\mathbb{E}[T] = \int_{\mathbb{R}} t d\mathbb{P}_T(t) = \sum_{k=1}^{+\infty} p(1-p)^{k-1} \int_{\mathbb{R}} t d\delta_k(t) = \sum_{k=1}^{+\infty} p(1-p)^{k-1} k.$$

On reconnaît la dérivée terme à terme de la série $\sum_{k=1}^{+\infty} px^k$ calculée pour $x = p-1$. De plus $\sum_{k=1}^{+\infty} x^k = \frac{1}{1-x}$ et sa dérivée est $\frac{1}{(1-x)^2}$. En conclusion $\mathbb{E}[T] = \frac{1}{p}$.

Corrigé de l'exercice 4.19, page 92

1. On a

$$\limsup_k (p_k \mathbb{N}^*) = \bigcap_k \bigcup_{n \geq k} p_n \mathbb{N}^* \subseteq \bigcap_k [p_k, +\infty[= \emptyset,$$

donc $\mathbb{P}(\limsup_k (p_k \mathbb{N}^*)) = 0$.

2. Montrons que la suite $(p_k \mathbb{N}^*)_{k \in \mathbb{N}^*}$ est indépendante. Soit $\{p_{k_1}, \dots, p_{k_n}\}$ des entiers premiers, comme $p_{k_1} \mathbb{N}^* \cap \dots \cap p_{k_n} \mathbb{N}^* = (p_{k_1} \dots p_{k_n}) \mathbb{N}^*$, il vient

$$\begin{aligned} \mathbb{P}(p_{k_1} \mathbb{N}^* \cap \dots \cap p_{k_n} \mathbb{N}^*) &= \mathbb{P}((p_{k_1} \dots p_{k_n}) \mathbb{N}^*) \\ &= \frac{1}{p_{k_1} \dots p_{k_n}} \\ &= \mathbb{P}(p_{k_1} \mathbb{N}^*) \dots \mathbb{P}(p_{k_n} \mathbb{N}^*). \end{aligned}$$

Ce qui prouve l'indépendance de la suite $(p_k \mathbb{N}^*)_{k \in \mathbb{N}^*}$. D'après un résultat d'arithmétique la série $\sum_{k=1}^{+\infty} \frac{1}{p_k} = +\infty$, et comme les $p_k \mathbb{N}^*$ sont indépendants par rapport à \mathbb{P} , en appliquant le théorème de Borel-Cantelli on déduit que $\mathbb{P}(\limsup_k (p_k \mathbb{N}^*)) = 1$. La contradiction prouve que la probabilité \mathbb{P} n'existe pas.

Corrigé de l'exercice 4.20, page 93

1. Considérons la famille de toutes les sous-tribus \mathcal{A} de \mathcal{F} telles que, pour tout $i \in I$, X_i est \mathcal{A} -mesurable. Cette famille est non vide car elle contient la tribu \mathcal{F} . On vérifie alors que $\sigma(X_i, i \in I)$ est l'intersection de toutes les tribus composant cette famille.
2. Un élément de \mathcal{C} est donc de la forme

$$\{X_{i_1} \in B_1\} \cap \{X_{i_2} \in B_2\} \cap \dots \cap \{X_{i_n} \in B_n\}$$

où $n \in \mathbb{N}^*$, $i_1, i_2, \dots, i_n \in I$ sont distincts deux à deux et $B_1, B_2, \dots, B_n \in \mathcal{B}(\mathbb{R})$.

Comme, pour tout $k \in I$, X_k est $\sigma(X_i, i \in I)$ -mesurable, $\{X_k \in B\} \in \sigma(X_i, i \in I)$ pour tout borélien B . Par suite la famille \mathcal{C} est incluse dans la tribu $\sigma(X_i, i \in I)$. Il en est de même pour la tribu engendrée par \mathcal{C} .

Réciproquement, pour tous $k \in I$ et borélien B de \mathbb{R} , $\{X_k \in B\} \in \mathcal{C}$. Donc, pour tout $k \in I$, l'application X_k est mesurable par rapport à la tribu engendrée par \mathcal{C} . Par suite la tribu engendrée par \mathcal{C} contient la tribu $\sigma(X_i, i \in I)$ qui est la plus petite à posséder cette propriété.

3. Notons \mathcal{C}_J (resp. \mathcal{C}_K) la famille définie comme dans la question précédente et engendrant la tribu $\sigma(X_i, i \in J)$ (resp. $\sigma(X_i, i \in K)$).

Comme les familles \mathcal{C}_J et \mathcal{C}_K sont stables par intersections finies, pour montrer l'indépendance des tribus $\sigma(X_i, i \in J)$ et $\sigma(X_i, i \in K)$, il suffit de montrer que, pour tous $C \in \mathcal{C}_J$ et $D \in \mathcal{C}_K$, on a $\mathbb{P}(C \cap D) = \mathbb{P}(C)\mathbb{P}(D)$.

$C \in \mathcal{C}_J$ est donc de la forme $\{X_{j_1} \in A_1\} \cap \{X_{j_2} \in A_2\} \cap \dots \cap \{X_{j_n} \in A_n\}$ où $n \in \mathbb{N}^*$, $j_1, j_2, \dots, j_n \in J$ sont distincts deux à deux et $A_1, A_2, \dots, A_n \in \mathcal{B}(\mathbb{R})$, et, de même, $D \in \mathcal{C}_K$ est de la forme $\{X_{k_1} \in B_1\} \cap \{X_{k_2} \in B_2\} \cap \dots \cap \{X_{k_m} \in B_m\}$ où $m \in \mathbb{N}^*$, $k_1, k_2, \dots, k_m \in K$ sont distincts deux à deux et $B_1, B_2, \dots, B_m \in \mathcal{B}(\mathbb{R})$.

Par suite, en vertu de l'indépendance des v.a.r. $(X_i)_I$, il vient

$$\begin{aligned} \mathbb{P}(C \cap D) &= \\ &= \mathbb{P}(\{X_{j_1} \in A_1\} \cap \dots \cap \{X_{j_n} \in A_n\} \cap \{X_{k_1} \in B_1\} \cap \dots \cap \{X_{k_m} \in B_m\}) \\ &= \mathbb{P}(X_{j_1} \in A_1) \dots \mathbb{P}(X_{j_n} \in A_n) \mathbb{P}(X_{k_1} \in B_1) \dots \mathbb{P}(X_{k_m} \in B_m) \\ &= \mathbb{P}(\{X_{j_1} \in A_1\} \cap \dots \cap \{X_{j_n} \in A_n\}) \mathbb{P}(\{X_{k_1} \in B_1\} \cap \dots \cap \{X_{k_m} \in B_m\}) \\ &= \mathbb{P}(C)\mathbb{P}(D). \end{aligned}$$

D'où la relation cherchée.

Corrigé de l'exercice 4.21, page 93

Nous utiliserons le fait que la fonction caractéristique d'une somme finie de variables aléatoires indépendantes est égale au produit des fonctions caractéristiques des variables aléatoires de la somme (Pour obtenir les fonctions caractéristiques des lois classiques on pourra se reporter au formulaire de l'annexe A, page 205).

1. Comme les fonctions caractéristiques de X_1 et X_2 sont données par $\phi_{X_1}(t) = (1 - it\alpha)^{-a}$ et $\phi_{X_2}(t) = (1 - it\alpha)^{-b}$, on a $\phi_Y(t) = (1 - it\alpha)^{-(a+b)}$ et Y est une v.a.r. de loi gamma $\gamma(a+b, \alpha)$.
2. On sait que, pour tout $k = 1, 2, \dots, n$, $\phi_{X_k}(t) = (1 - it\alpha)^{-1}$, on a donc $\phi_Z(t) = (1 - it\alpha)^{-n}$ et Z est une v.a.r. de loi gamma $\gamma(n, \alpha)$.

Corrigé de l'exercice 4.22, page 93

Par définition

$$\phi_Z(t) = \mathbb{E}[e^{itZ}] = \mathbb{E}\left[e^{it\sum_{j=1}^Y X_j}\right] = \mathbb{E}\left[\sum_{n=1}^{+\infty} e^{it\sum_{j=1}^n X_j} \mathbf{1}_{(Y=n)}\right].$$

Or la suite de v.a.r. $(\sum_{n=1}^k e^{it\sum_{j=1}^n X_j} \mathbf{1}_{(Y=n)})_{k \geq 1}$ est bornée en module par 1, une seule des indicatrices $\mathbf{1}_{(Y=n)}$ est éventuellement non nulle. On obtient donc en utilisant le théorème

de convergence dominée puis l'indépendance des v.a.r.

$$\begin{aligned}\phi_Z(t) &= \sum_{n=1}^{+\infty} \mathbb{E} \left[e^{it \sum_{j=1}^n X_j} \mathbf{1}_{(Y=n)} \right] = \sum_{n=1}^{+\infty} \mathbb{E}[\mathbf{1}_{(Y=n)}] \prod_{j=1}^n \mathbb{E}[e^{itX_j}] \\ &= \sum_{n=1}^{+\infty} \mathbb{P}(Y = n) (\phi(t))^n = \psi \circ \phi(t).\end{aligned}$$

L'application ψ qui est aussi définie par $\psi(t) = \mathbb{E}[t^Y]$ est appelée **fonction génératrice** de la variable discrète Y .

Corrigé de l'exercice 4.23, page 93

D est une v.a.r. à valeurs \mathbb{Z} et M à valeurs \mathbb{N}^* . De plus, les v.a.r. étant discrètes, le couple (D, M) est indépendant si, et seulement si, pour tout $(i, j) \in \mathbb{Z} \times \mathbb{N}^*$, $\mathbb{P}(\{D = i\} \cap \{M = j\}) = \mathbb{P}(D = i)\mathbb{P}(M = j)$.

1. Supposons que X et Y suivent la loi géométrique de paramètre $p \in]0, 1[$.

On vérifie aisément que, pour $j \in \mathbb{N}^*$ fixé,

$$\{D = i\} \cap \{M = j\} = \begin{cases} \{X = i + j\} \cap \{Y = j\} & \text{si } i \geq 0, \\ \{X = j\} \cap \{Y = j - i\} & \text{si } i < 0. \end{cases}$$

Par suite en vertu de l'indépendance du couple (X, Y) il vient

$$\begin{aligned}\text{si } i \geq 0, \quad \mathbb{P}(\{D = i\} \cap \{M = j\}) &= \mathbb{P}(\{X = i + j\} \cap \{Y = j\}) \\ &= \mathbb{P}(X = i + j)\mathbb{P}(Y = j) = p^2 q^{i+j-1} q^{j-1} = p^2 q^{i+2j-2}\end{aligned}$$

$$\begin{aligned}\text{si } i < 0, \quad \mathbb{P}(\{D = i\} \cap \{M = j\}) &= \mathbb{P}(\{X = j\} \cap \{Y = j - i\}) \\ &= \mathbb{P}(X = j)\mathbb{P}(Y = j - i) = p^2 q^{j-i-1} q^{j-1} = p^2 q^{2j-i-2}\end{aligned}$$

On peut rassembler ces deux écritures sous la forme $\mathbb{P}(\{D = i\} \cap \{M = j\}) = \frac{p^2}{q^2} q^{2j+|i|}$.

Calculons, pour tout $i \in \mathbb{Z}$ fixé, $\mathbb{P}(D = i)$. Comme $\{D = i\} = \bigcup_{j \in \mathbb{N}^*} (\{D = i\} \cap \{M = j\})$

où l'union est mutuellement disjointe et par un calcul simple de somme de séries géométriques, il vient

$$\begin{aligned}\mathbb{P}(D = i) &= \sum_{j \in \mathbb{N}^*} \mathbb{P}(\{D = i\} \cap \{M = j\}) \\ &= \sum_{j \in \mathbb{N}^*} \frac{p^2}{q^2} q^{2j+|i|} = \frac{p^2}{1 - q^2} q^{|i|}.\end{aligned}$$

Calculons maintenant, pour tout $j \in \mathbb{N}^*$ fixé, $\mathbb{P}(M = j)$. De la même façon que précédemment $\{M = j\} = \bigcup_{i \in \mathbb{Z}} (\{D = i\} \cap \{M = j\})$ où l'union est mutuellement

disjointe. Par un calcul simple de somme de séries géométriques, il vient

$$\begin{aligned}\mathbb{P}(M = j) &= \sum_{i \in \mathbb{Z}} \mathbb{P}(\{D = i\} \cap \{M = j\}) \\ &= \sum_{i \in \mathbb{Z}} \frac{p^2}{q^2} q^{2j+|i|} = \frac{p}{q^2} q^{2j} (q + 1).\end{aligned}$$

Comparant les trois valeurs trouvées, on vérifie aisément que $\mathbb{P}(\{D = i\} \cap \{M = j\}) = \mathbb{P}(D = i)\mathbb{P}(M = j)$, ce qui prouve l'indépendance du couple (D, M) .

2. Réciproquement, supposons les v.a.r. D et M indépendantes. Soit $n \in \mathbb{N}^*$, fixé. Par l'indépendance du couple (X, Y) , il vient

$$\frac{\mathbb{P}((X = n+1) \cap (Y = n))}{\mathbb{P}((X = n) \cap (Y = n))} = \frac{\mathbb{P}(X = n+1)\mathbb{P}(Y = n)}{\mathbb{P}(Y = n)\mathbb{P}(X = n)} = \frac{\mathbb{P}(X = n+1)}{\mathbb{P}(X = n)},$$

ce qui donne la première égalité.

De plus $(X = n+1) \cap (Y = n) = (D = 1) \cap (M = n)$ et $(X = n) \cap (Y = n) = (D = 0) \cap (M = n)$. Par l'indépendance du couple (M, D) , il vient

$$\frac{\mathbb{P}((X = n+1) \cap (Y = n))}{\mathbb{P}((X = n) \cap (Y = n))} = \frac{\mathbb{P}(D = 1)\mathbb{P}(M = n)}{\mathbb{P}(D = 0)\mathbb{P}(M = n)} = \frac{\mathbb{P}(D = 1)}{\mathbb{P}(D = 0)}$$

d'où la deuxième égalité.

On remarque que le rapport ne dépend pas de l'entier n . Soit α sa valeur. On a la relation de récurrence, pour tout $n \in \mathbb{N}^*$, $\mathbb{P}(X = n+1) = \alpha\mathbb{P}(X = n)$.

Par suite, pour tout $n \in \mathbb{N}^*$, $\mathbb{P}(X = n) = \alpha^{n-1}\mathbb{P}(X = 1)$. De la relation $\sum_{k \geq 1} \mathbb{P}(X = k) = 1$ on déduit que $\sum_{k \geq 1} \alpha^{k-1}\mathbb{P}(X = 1) = \frac{1}{1-\alpha}\mathbb{P}(X = 1) = 1$, et par suite que

$$\mathbb{P}(X = 1) = 1 - \alpha.$$

La loi de X et Y s'écrit alors : $\mathbb{P}_X = \mathbb{P}_Y = \sum_{k \geq 1} \alpha^{k-1}(1-\alpha)\delta_k$. Les v.a.r. X et Y suivent donc la loi géométrique de paramètre $1 - \alpha$.

Corrigé de l'exercice 4.24, page 94

La probabilité que les variables aléatoires X et Y prennent leurs valeurs dans le complémentaire de l'intervalle $]0, \alpha[$ est nulle. On peut donc considérer, dans la suite, que les variables X et Y sont à valeurs dans l'intervalle $]0, \alpha[$.

Par ailleurs, en vertu de l'indépendance du couple (X, Y) , le vecteur aléatoire (X, Y) de dimension 2 admet pour densité ρ la fonction définie sur \mathbb{R}^2 par : pour tout $(x, y) \in \mathbb{R}^2$, $\rho(x, y) = \frac{1}{\alpha^2}$ si $(x, y) \in]0, \alpha[^2$, et $\rho(x, y) = 0$ si $(x, y) \notin]0, \alpha[^2$. C'est-à-dire que la vecteur aléatoire (X, Y) de dimension 2 suit la loi uniforme de dimension 2 sur $]0, \alpha[^2$.

En conséquence, pour tout borélien A de \mathbb{R}^2 , $\mathbb{P}[(X, Y) \in A] = \frac{a(\Delta \cap]0, \alpha[^2)}{\alpha^2}$, où $a(\Delta \cap]0, \alpha[^2)$ désigne la mesure de l'aire de l'intersection du borélien Δ avec le carré ouvert $]0, \alpha[^2$.

Pour déterminer les lois des variables Z et T , nous pouvons étudier leurs fonctions de répartition, que nous noterons respectivement F et G .

Loi de Z .

Remarquons que la variable Z peut aussi s'écrire $Z = \max(X, Y) - \min(X, Y) = |X - Y|$. C'est une variable aléatoire positive qui prend ses valeurs dans l'intervalle $[0, \alpha]$.

Étudions la fonction de répartition F de Z . Soit z réel fixé. Cela revient à étudier la probabilité $F(z) = \mathbb{P}(Z \leq z)$ de l'événement $\{Z \leq z\}$.

- ◊ Si $z < 0$, comme Z est positive, l'événement $\{Z \leq z\} = \emptyset$ (événement impossible). Par suite $F(z) = 0$.

- ◇ Si $z \geq \alpha$, comme Z prend ses valeurs dans $[0, \alpha]$, l'événement $\{Z \leq z\} = \Omega$ (événement certain). Par suite $F(z) = 1$.
- ◇ Il reste à étudier le cas $0 \leq z < \alpha$. Supposons $0 \leq z < \alpha$ fixé. On peut écrire $\{Z \leq z\} = \{(X, Y) \in \Delta \cap]0, \alpha]^2\}$ où $\Delta = \{(x, y) \in \mathbb{R}^2 / |x - y| \leq z\}$ désigne la bande définie par les points de \mathbb{R}^2 contenus entre les deux droites d'équations respectives $y = x + z$ et $y = x - z$ dans un système d'axes orthonormé. Un raisonnement géométrique élémentaire montre que la mesure de l'aire de l'intersection $\Delta \cap]0, \alpha]^2$ est : $a(\Delta \cap]0, \alpha]^2) = \alpha^2 - (\alpha - z)^2$.
En conséquence, d'après la remarque préliminaire sur la loi du couple (X, Y) ,

$$F(z) = \mathbb{P}(Z \leq z) = \mathbb{P}[(X, Y) \in \Delta \cap]0, \alpha]^2] = \frac{\alpha^2 - (\alpha - z)^2}{\alpha^2} = 1 - \left(1 - \frac{z}{\alpha}\right)^2.$$

En résumé, la fonction de répartition de la variable aléatoire Z est donnée, pour tout réel z , par :

$$F(z) = \begin{cases} 0 & \text{si } z < 0, \\ 1 - \left(1 - \frac{z}{\alpha}\right)^2 & \text{si } 0 \leq z < \alpha, \\ 1 & \text{si } z \geq \alpha. \end{cases}$$

Comme la fonction de répartition est de classe C^1 sur les trois intervalles $] -\infty, 0[$, $]0, \alpha[$ et $]\alpha, +\infty[$, une densité f vérifiera l'équation $F' = f$ sur chacun de ces intervalles. Les valeurs de f aux bornes, $f(0)$ et $f(\alpha)$, peuvent être choisies arbitrairement (on rappelle qu'une densité n'est définie que presque-partout pour la mesure de Lebesgue).

En conclusion, la variable aléatoire Z admet pour densité, la fonction f définie pour tout réel z , par :

$$f(z) = \begin{cases} 0 & \text{si } z < 0 \text{ ou si } z > \alpha, \\ \frac{2}{\alpha} \left(1 - \frac{z}{\alpha}\right) & \text{si } 0 \leq z \leq \alpha, \end{cases}$$

Loi de T

Remarquons que la variable T est une variable aléatoire positive qui prend ses valeurs dans l'intervalle $[0, 1]$.

Étudions la fonction de répartition G de T . Soit z réel fixé. Cela revient à étudier la probabilité $G(z) = \mathbb{P}(T \leq z)$ de l'événement $\{T \leq z\}$.

- ◇ Si $z < 0$, comme T est positive, l'événement $\{T \leq z\} = \emptyset$ (événement impossible). Par suite $G(z) = 0$.
- ◇ Si $z \geq 1$, comme T prend ses valeurs dans $[0, 1]$, l'événement $\{T \leq z\} = \Omega$ (événement certain). Par suite $G(z) = 1$.
- ◇ Il reste à étudier le cas $0 \leq z < 1$. Supposons $0 \leq z < 1$ fixé. On peut écrire $\{T \leq z\} = \{(X, Y) \in \Delta' \cap]0, \alpha]^2\}$ où $\Delta' = \Delta_1 \cup \Delta_2$ avec $\Delta_1 = \left\{ (x, y) \in \mathbb{R}^2 / 0 < x < y \text{ et } \frac{x}{y} \leq z \right\}$ et $\Delta_2 = \left\{ (x, y) \in \mathbb{R}^2 / 0 < y < x \text{ et } \frac{y}{x} \leq z \right\}$. Δ' est le complémentaire (dans $]0, +\infty[^2$) du secteur angulaire défini par les points de $]0, +\infty[^2$ contenus entre les deux droites d'équations respectives $y = \frac{x}{z}$ et $y = zx$ dans un système d'axes orthonormé. Un raisonnement géométrique élémentaire montre que la mesure de l'aire de l'intersection $\Delta' \cap]0, \alpha]^2$ est : $a(\Delta' \cap]0, \alpha]^2) = z\alpha^2$.

En conséquence, d'après la remarque préliminaire sur la loi du couple (X, Y) ,

$$G(z) = \mathbb{P}(T \leq z) = \mathbb{P}[(X, Y) \in \Delta' \cap]0, \alpha[^2] = \frac{z\alpha^2}{\alpha^2} = z.$$

En résumé, la fonction de répartition de la variable aléatoire T est donnée, pour tout réel z , par :

$$G(z) = \begin{cases} 0 & \text{si } z < 0, \\ z & \text{si } 0 \leq z < 1, \\ 1 & \text{si } z \geq 1. \end{cases}$$

On reconnaît la fonction de répartition de la loi uniforme sur l'intervalle $[0, 1]$. La variable aléatoire T suit donc la loi uniforme sur l'intervalle $[0, 1]$.

8.5 Corrigés des exercices du chapitre V

Corrigé de l'exercice 5.1, page 97

Le vecteur aléatoire X est l'image du vecteur U par l'application linéaire A dont la matrice $[a_{i,j}]_{1 \leq i,j \leq n}$ dans la base canonique de \mathbb{R}^n a pour coefficients

$$\begin{aligned} a_{i,i} &= 1 && \text{pour } i = 1, \dots, n, \\ a_{j+1,j} &= \theta && \text{pour } j = 1, \dots, n-1, \\ a_{i,j} &= 0 && \text{dans les autres cas.} \end{aligned}$$

Par suite toute combinaison linéaire des variables aléatoires X_1, X_2, \dots, X_n est une combinaison linéaire des variables aléatoires U_1, U_2, \dots, U_n . Comme (U_1, U_2, \dots, U_n) est une suite indépendante de variable aléatoire réelle gaussiennes, le vecteur U est gaussien. Par suite toute combinaison linéaire des variables aléatoires U_1, U_2, \dots, U_n est une variable aléatoire gaussienne. Il en est donc de même de toute combinaison linéaire des variables aléatoires X_1, X_2, \dots, X_n . Ce qui prouve, par définition des vecteurs gaussiens, que le vecteur $X = AU$ est lui-même un vecteur gaussien.

Corrigé de l'exercice 5.2, page 98

1. Déterminons la loi de X par le critère d'identification des lois par les fonctions positives. Soit h une application borélienne positive de \mathbb{R}^d dans $[0, +\infty[$. Reprenons les notations du corrigé 8.5, page 183, de l'exercice 5.1, page 97. Par application du théorème du transfert au vecteur U , par indépendance de la suite (U_1, U_2, \dots, U_n) en notant $f(t)$ la densité de $\mathcal{N}_1(0, \sigma^2)$ et $x = (x_1, x_2, \dots, x_n)$, il vient

$$\begin{aligned} \mathbb{E}(h(X)) &= \mathbb{E}(h(AU)) = \int_{\mathbb{R}^n} h(Ax) d\mathbb{P}_U(x) \\ &= \int_{\mathbb{R}^n} h(Ax) f(x_1) f(x_2) \cdots f(x_n) d\lambda^{(n)}(x). \end{aligned}$$

Faisons le changement de variable à l'aide du C^1 -difféomorphisme A i.e. $y := Ax$, pour tout $x \in \mathbb{R}^n$. L'application réciproque est définie, pour tout entier $1 \leq k \leq n$, par

$x_k = \sum_{j=0}^{k-1} (-\theta)^j y_{k-j}$ et son jacobien en un point y est $\det(A^{-1}) = 1$. On obtient

$$\begin{aligned}\mathbb{E}(h(X)) &= \int_{\mathbb{R}^n} h(y) f(y_1) f(y_2 - \theta y_1) \cdots f\left(\sum_{j=0}^{n-1} (-\theta)^j y_{n-j}\right) d\lambda^{(n)}(y) \\ &= \int_{\mathbb{R}^n} h(y) \left(\frac{1}{\sqrt{2\pi}}\right)^n \exp\left[-\frac{1}{2\sigma^2} \sum_{k=1}^n \left(\sum_{j=0}^{k-1} (-\theta)^j y_{k-j}\right)^2\right] d\lambda^{(n)}(y).\end{aligned}$$

Ce qui prouve que X est un vecteur aléatoire de densité définie, pour tout $x = (x_1, x_2, \dots, x_n) \in \mathbb{R}^n$, par

$$f_X(x) = \left(\frac{1}{\sqrt{2\pi}}\right)^n \exp\left[-\frac{1}{2\sigma^2} \sum_{k=1}^n \left(\sum_{j=0}^{k-1} (-\theta)^j x_{k-j}\right)^2\right].$$

2. L'espérance de X est $\mathbb{E}[X] = \mathbb{E}[AU] = A\mathbb{E}[U] = 0$ (vecteur nul de \mathbb{R}^n).
La matrice de dispersion de X est donnée par

$$D_X := \mathbb{E}(XX^*) = \mathbb{E}(AU(AU)^*) = AD_U A^* = \sigma^2 AA^*$$

car $D_U = \sigma^2 I_n$, où I_n désigne la matrice-identité d'ordre n . On peut calculer D_X en effectuant le produit matriciel AA^* . On trouve que $D_X = [d_{i,j}]_{1 \leq i,j \leq n}$ où

$$\begin{aligned}d_{1,1} &= \sigma^2 \\ d_{i,i} &= \sigma^2(\theta^2 + 1) && \text{pour } i = 2, \dots, n, \\ d_{j+1,j} &= d_{j,j+1} = \theta\sigma^2 && \text{pour } j = 1, \dots, n-1, \\ d_{i,j} &= 0 && \text{dans les autres cas.}\end{aligned}$$

Corrigé de l'exercice 5.3, page 101

1. On remarque que $\mathbb{E}[XX_a] = \mathbb{E}[X^2 \mathbf{1}_{\{|X| \leq a\}}] - \mathbb{E}[X^2 \mathbf{1}_{\{|X| > a\}}]$. Le théorème du transfert permet d'écrire

$$\mathbb{E}[X^2 \mathbf{1}_{\{|X| \leq a\}}] = \frac{1}{\sqrt{2\pi}} \int_{-a}^a x^2 e^{-\frac{1}{2}x^2} dx, \quad \mathbb{E}[X^2 \mathbf{1}_{\{|X| > a\}}] = \frac{2}{\sqrt{2\pi}} \int_a^{+\infty} x^2 e^{-\frac{1}{2}x^2} dx.$$

$\mathbb{E}[XX_a]$ est donc égale à la différence de deux fonctions de la variable a réelle positive, continues sur \mathbb{R}^+ , dont la première est strictement croissante de 0 à $\mathbb{E}[X^2] = 1$ et la seconde strictement décroissante de $\mathbb{E}[X^2] = 1$ à 0. Il existe donc une unique valeur de a_0 pour laquelle $\mathbb{E}[XX_{a_0}] = 0$ c'est-à-dire (X, X_{a_0}) est non-corrélé.

2. Comme, pour tout réel $a > 0$, $X + X_a = 2X \mathbf{1}_{\{|X| \leq a\}}$ n'est pas une v.a.r. gaussienne car $X + X_a$ est une variable aléatoire bornée par $2a$, le vecteur (X, X_a) n'est gaussien pour aucune valeur de a .
3. Si le couple (X, X_a) était indépendant, le vecteur (X, X_a) serait gaussien car ses composantes seraient des v.a.r. gaussiennes indépendantes. D'après la question précédente, il y aurait contradiction. Pour tout réel $a > 0$, le couple (X, X_a) n'est donc pas indépendant.

Corrigé de l'exercice 5.4, page 102

Si on considère X comme un vecteur colonne, alors le vecteur aléatoire $Y := AX$ suit une loi gaussienne $\mathcal{N}_3(A0, A\Gamma A^*) = \mathcal{N}_3(0, A\Gamma A^*)$. Comme Y est un vecteur gaussien il suffit, pour que ses composantes soient indépendantes, que la matrice $A\Gamma A^*$ soit diagonale. Si de plus on veut que les composantes de AX soient non-dégénérées, alors il faut que la matrice $A\Gamma A^*$ soit inversible. Comme Γ est réelle symétrique, elle est diagonalisable dans \mathbb{R} et il existe une matrice orthogonale U telle que la matrice $\Delta := U^{-1}\Gamma U$ soit diagonale. Les valeurs propres de Γ sont 2 et 4. La matrice Δ est donc inversible ainsi que la matrice Γ . Les vecteurs $e_1 := \frac{1}{\sqrt{2}}(1, 1, 0)$ et $e_2 := (0, 0, 1)$ sont des vecteurs propres linéairement-indépendants associés à la valeur propre double 2 et $e_3 := \frac{1}{\sqrt{2}}(1, -1, 0)$ est un vecteur propre associé à la valeur propre 4. La base de vecteurs propres (e_1, e_2, e_3) est orthonormée et la matrice

$$U := \begin{pmatrix} 1/\sqrt{2} & 0 & 1/\sqrt{2} \\ 1/\sqrt{2} & 0 & -1/\sqrt{2} \\ 0 & 1 & 0 \end{pmatrix}$$

est orthogonale. Posons $A = U^* = U^{-1}$. On vérifie que la matrice A ainsi définie répond à la question posée.

Corrigé de l'exercice 5.5, page 103

Avec les notations du cours concernant les opérations matricielles, l'espérance de U est donnée par $A0 = 0$ et sa matrice de dispersion est donnée par $A/A^* = AA^* = I$. En effet A est une matrice orthogonale donc $A^* = A^{-1}$. L'image du vecteur gaussien X par l'endomorphisme de \mathbb{R}^2 de matrice A dans la base canonique de \mathbb{R}^2 est encore un vecteur gaussien. Par suite la loi de U est $\mathcal{N}_2(0, I)$.

Corrigé de l'exercice 5.6, page 103

1. Comme $\mathbb{P}_{(X,Y)}$ est une probabilité sur \mathbb{R}^2 , α doit être tel que $\mathbb{P}_{(X,Y)}(\mathbb{R}^2) = 1$. Le théorème de Tonelli permet d'écrire

$$\begin{aligned} \mathbb{P}_{(X,Y)}(\mathbb{R}^2) &= \alpha \int_{\mathbb{R}^2} e^{-\frac{1}{2}(x^2 - xy + y^2)} d\lambda^{(2)}(x, y) \\ &= \alpha \int_{\mathbb{R}} e^{-\frac{3}{8}y^2} \int_{\mathbb{R}} e^{-\frac{1}{2}(x - \frac{y}{2})^2} d\lambda^{(1)}(x) d\lambda^{(1)}(y) \\ &= \alpha \int_{\mathbb{R}} e^{-\frac{3}{8}y^2} \int_{\mathbb{R}} e^{-\frac{1}{2}x^2} d\lambda^{(1)}(x) d\lambda^{(1)}(y) \\ &= \alpha \int_{\mathbb{R}} e^{-\frac{3}{8}y^2} \sqrt{2\pi} d\lambda^{(1)}(y) = \alpha \sqrt{2\pi} \int_{\mathbb{R}} e^{-\frac{3}{8}y^2} d\lambda^{(1)}(y) = \alpha \frac{4\pi}{\sqrt{3}}. \end{aligned}$$

Où on a utilisé la valeur de l'intégrale de Gauss $\int_{\mathbb{R}} e^{-\frac{(x-m)^2}{2\sigma^2}} d\lambda^{(1)}(x) = \sigma\sqrt{2\pi}$. Par suite $\alpha = \frac{\sqrt{3}}{4\pi}$.

2. On rappelle que la loi $\mathcal{N}_2(m, D)$ admet une densité par rapport à la mesure de Lebesgue $\lambda^{(2)}$ si et seulement si la matrice D est inversible (cf cours). Dans ce cas la densité s'écrit, pour tout $t \in \mathbb{R}^2$,

$$\frac{1}{2\pi\sqrt{\det(D)}} \exp\left(-\frac{1}{2}(t-m)^* D^{-1}(t-m)\right).$$

Le terme constant $m^*D^{-1}m$, obtenu par $t = 0$ dans l'expression $(t - m)^*D^{-1}(t - m)$, représente la norme du vecteur m relativement à la forme quadratique définie positive définie par D^{-1} . Il est nul si et seulement si $m = 0$. Comme

$$x^2 - xy + y^2 = \begin{pmatrix} x & y \end{pmatrix} \begin{pmatrix} 1 & -1/2 \\ -1/2 & 1 \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix},$$

on reconnaît en $\mathbb{P}_{(X,Y)}$ la loi gaussienne $\mathcal{N}_2(m, D)$ où $m = 0$ puisqu'il n'y a pas de terme constant dans l'exponentielle et $D^{-1} = \begin{pmatrix} 1 & -1/2 \\ -1/2 & 1 \end{pmatrix}$. Donc $D := \begin{pmatrix} 4/3 & 2/3 \\ 2/3 & 4/3 \end{pmatrix}$ est la matrice de dispersion du vecteur (X, Y) .

On en déduit que X et Y suivent la loi $\mathcal{N}_1(0, \frac{4}{3})$ et, puisque D n'est pas diagonale, le couple de variable aléatoire réelle (X, Y) n'est pas indépendant.

On peut aussi déterminer directement $\text{Var}(X)$, $\text{Var}(Y)$ et $\text{Cov}(X, Y)$ par un calcul d'intégrales à partir de la densité de la loi du couple (X, Y) .

Corrigé de l'exercice 5.7, page 103

Remarquons que X et $-X$ ont la même loi. En effet l'étude des fonctions caractéristiques donne $\Phi_{-X}(t) = \mathbb{E}[e^{-itX}] = \Phi_X(-t) = e^{-\frac{1}{2}t^2} = \Phi_X(t)$ pour tout réel t .

1. La relation

$$h(Y) = h(X)\mathbb{1}_{[0,\pi]}(|X|) + h(-X)\mathbb{1}_{\pi,+\infty[}(|X|)$$

est facile à vérifier. Il suffit pour cela d'étudier séparément le cas où ω satisfait $|X(\omega)| \in [0, \pi]$ et le cas où ω satisfait $|X(\omega)| \in \pi, +\infty[$. Pour observer que dans le premier cas, $h(Y(\omega)) = h(X(\omega))$ et dans le second cas $h(Y(\omega)) = h(-X(\omega))$. La relation avec les indicatrices sert à résumer ces deux cas en une seule expression.

2. Soient h une application borélienne positive de \mathbb{R} dans \mathbb{R} , de la relation

$$h(Y) = h(X)\mathbb{1}_{[0,\pi]}(|X|) + h(-X)\mathbb{1}_{\pi,+\infty[}(|X|).$$

on obtient, en passant à l'espérance,

$$\mathbb{E}[h(Y)] = \mathbb{E}[h(X)\mathbb{1}_{[0,\pi]}(|X|)] + \mathbb{E}[h(-X)\mathbb{1}_{\pi,+\infty[}(|X|)].$$

En remarquant que X et $-X$ ont la même loi et en utilisant le théorème du transfert, il vient

$$\mathbb{E}[h(-X)\mathbb{1}_{\pi,+\infty[}(|X|)] = \mathbb{E}[h(X)\mathbb{1}_{\pi,+\infty[}(|X|)].$$

Ce qui, en reportant dans le deuxième membre de l'égalité précédente, donne $\mathbb{E}[h(Y)] = \mathbb{E}[h(X)]$, pour toute application borélienne positive de \mathbb{R} dans \mathbb{R} . Ce qui prouve que Y suit la même loi que X c'est-à-dire $\mathcal{N}_1(0, 1)$.

3. La variable aléatoire réelle $X + Y = 2X\mathbb{1}_{[0,\pi]}(|X|)$ n'est pas une gaussienne car $X + Y$ est une variable aléatoire bornée par 2π . Le vecteur (X, Y) n'est donc pas gaussien.

4. Si le couple (X, Y) était indépendant, le vecteur aléatoire (X, Y) serait gaussien car ses composantes seraient des variable aléatoire réelle gaussiennes indépendantes. D'après la question précédente, il y aurait contradiction. En conclusion, le couple (X, Y) n'est donc pas indépendant.

Corrigé de l'exercice 5.8, page 103

1. Du fait de l'indépendance des variables aléatoires réelles X , Y et Z , la fonction caractéristique de U est égale au produit des fonctions caractéristiques de X , Y et Z . Cette fonction caractéristique est donc donnée, pour tout $t \in \mathbb{R}$, par $\Phi_U(t) = e^{-\frac{3}{2}t^2}$ et par suite U a pour loi $\mathcal{N}_1(0, 3)$.

De la même façon, on montre que la variable aléatoire réelle $V := 2X - Y - Z$ suit la loi $\mathcal{N}_1(0, 6)$ et que la variable aléatoire réelle $W := Y - Z$ suit la loi $\mathcal{N}_1(0, 2)$.

2. On remarque que le couple $(U, X - Y)$ est l'image de (X, Y, Z) par l'application linéaire de \mathbb{R}^3 dans \mathbb{R}^2 de matrice, dans les bases canoniques respectives de \mathbb{R}^3 et \mathbb{R}^2 , $A := \begin{pmatrix} 1 & 1 & 1 \\ 1 & -1 & 0 \end{pmatrix}$. Comme (X, Y, Z) est un vecteur gaussien de loi $\mathcal{N}_3(0, I)$ où 0 est le vecteur nul de \mathbb{R}^3 et I la matrice identité d'ordre 3, le vecteur $(U, X - Y)$ est aussi gaussien de loi $\mathcal{N}_2(0, AIA^*)$. La matrice de dispersion de $(U, X - Y)$ est donc $AA^* = \begin{pmatrix} 3 & 0 \\ 0 & 2 \end{pmatrix}$. Cette matrice étant diagonale, le couple de variables aléatoires réelles $(U, X - Y)$ est indépendant.

On procède de façon analogue pour les couples $(U, Y - Z)$ et $(U, Z - X)$.

3. On remarque que le vecteur $(X + Y + Z, 2X - Y - Z, Y - Z)$ est l'image du vecteur gaussien (X, Y, Z) par l'endomorphisme de \mathbb{R}^3 de matrice (dans la base canonique de \mathbb{R}^3)

$$B := \begin{pmatrix} 1 & 1 & 1 \\ 2 & -1 & -1 \\ 0 & 1 & -1 \end{pmatrix}. \text{ Le vecteur aléatoire } (U, V, W) \text{ est donc un vecteur gaussien de}$$

$$\text{loi } \mathcal{N}_3(0, BIB^*) = \mathcal{N}_3(0, BB^*) \text{ où } BB^* = \begin{pmatrix} 3 & 0 & 0 \\ 0 & 6 & 0 \\ 0 & 0 & 2 \end{pmatrix}.$$

4. La matrice des covariances du vecteur (U, V, W) , BB^* , est diagonale et le vecteur (U, V, W) gaussien, le triplet des variables aléatoires réelles (U, V, W) est donc indépendant.

5. La fonction caractéristique du couple (U, T) est définie, pour tout $(u, t) \in \mathbb{R}^2$, par

$$\begin{aligned} \Phi_{(U, T)}(u, t) &= \mathbb{E}[\exp(i(uU + vT))] \\ &= \mathbb{E}\left[\exp\left(iu(X + Y + Z) + \frac{1}{2}it(2X - Y - Z)^2 + \frac{3}{2}it(Y - Z)^2\right)\right], \end{aligned}$$

où on a utilisé l'identité à vérifier. Comme la famille des v.a.r. $(U, 2X - Y - Z, Y - Z)$

est indépendante,

$$\begin{aligned}\Phi_{(U,T)}(u, t) &= \mathbb{E}[e^{iuU}] \mathbb{E}\left[\exp\left(\frac{1}{2}it(2X - Y - Z)^2\right)\right] \mathbb{E}\left[\exp\left(\frac{3}{2}it(Y - Z)^2\right)\right] \\ &= e^{-\frac{3}{2}u^2} \Phi_{\frac{1}{6}(2X-Y-Z)^2}(3t) \Phi_{\frac{1}{2}(Y-Z)^2}(3t).\end{aligned}$$

Comme $(2X - Y - Z)$ suit la loi $\mathcal{N}_1(0, 6)$ et que $(Y - Z)$ suit la loi $\mathcal{N}_1(0, 2)$, $\frac{1}{\sqrt{6}}(2X - Y - Z)$ et $\frac{1}{\sqrt{2}}(Y - Z)$ suivent la loi $\mathcal{N}_1(0, 1)$. On conclut à l'égalité des fonctions caractéristiques $\Phi_{\frac{1}{\sqrt{6}}(2X-Y-Z)^2} = \Phi_{\frac{1}{\sqrt{2}}(Y-Z)^2} = \varphi$. Il vient alors

$$\Phi_{(U,T)}(u, t) = e^{-\frac{3}{2}u^2} [\varphi(3t)]^2 = e^{-\frac{3}{2}u^2} \frac{1}{1 - 6it}.$$

Corrigé de l'exercice 5.9, page 104

1. En utilisant la technique des fonctions boréliennes positives, si h est une telle fonction, on a

$$\begin{aligned}\mathbb{E}[h(X_1^2)] &= \int_{\mathbb{R}} h(x^2) d\mathbb{P}_{X_1}(x) = \int_{\mathbb{R}} h(x^2) \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} d\lambda(x) \\ &= 2 \int_{\mathbb{R}^+} h(x^2) \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} d\lambda(x),\end{aligned}$$

puisque la fonction intégrée par rapport à la mesure de Lebesgue est paire. On utilise ensuite le théorème de changement de variable dans une intégrale relative à la mesure de Lebesgue, en posant $t = x^2$.

$$\mathbb{E}[h(X_1^2)] = \int_{\mathbb{R}} h(t) \frac{1}{\sqrt{2\pi}} e^{-\frac{t}{2}} t^{\frac{1}{2}} \mathbf{1}_{\mathbb{R}^+}(t) d\lambda(t) = \int_{\mathbb{R}} h(t) d\gamma\left(\frac{1}{2}, \frac{1}{2}\right)(t).$$

On a montré $\mathbb{P}_{X_1^2} = \gamma(\frac{1}{2}, \frac{1}{2}) = \chi^2(1)$.

2. En utilisant les fonctions caractéristiques des lois Gamma (voir le formulaire de l'annexe A, page 205) et la proposition 4.31, page 86, la somme de deux v.a.r. indépendantes de loi respective $\gamma(a, b)$ et $\gamma(a, b')$ est une v.a.r. de loi $\gamma(a, b + b')$. Par récurrence immédiate sur n , on montre que la v.a.r. $\sum_{k=1}^n X_k^2$ suit une loi $\gamma(\frac{1}{2}, \frac{n}{2}) = \chi^2(n)$.

3. On considère l'espace vectoriel euclidien \mathbb{R}^n et on pose $V_n := \frac{1}{\sqrt{n}} \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$. D'après les

théorèmes d'algèbre linéaire, il est possible de construire une base \mathcal{B} orthomormée qui complète la famille libre (V_n) , $\mathcal{B} = (V_1, \dots, V_n)$. La matrice de passage de la base canonique à \mathcal{B} est la matrice orthogonale (i.e. sa transposée est égale à son inverse) formée des vecteurs colonnes de \mathcal{B} . Sa transposée, que l'on note C , est aussi orthogonale et de la forme

$$C = \begin{pmatrix} c_{1,1} & c_{1,2} & \cdots & c_{1,n} \\ c_{2,1} & c_{2,2} & \cdots & c_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ c_{n-1,1} & c_{n-1,2} & \cdots & c_{n-1,n} \\ \frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}} & \cdots & \frac{1}{\sqrt{n}} \end{pmatrix}.$$

4. En utilisant la proposition 5.5, page 98, comme X suit la loi $\mathcal{N}_n(0, I_n)$, $Y = CX$ suit la loi $\mathcal{N}_n(0, CI_nC^*) = \mathcal{N}_n(0, I_n)$, où C^* désigne la transposée de C , et $C^* = C^{-1}$ puisque C est orthogonale.

5. De $Y = CX$, on déduit facilement que $Y_n = \frac{1}{\sqrt{n}} \sum_{k=1}^n X_k = \sqrt{n} \bar{X}$. Avec les règles du calcul matriciel, on remarque que

$$\sum_{k=1}^n Y_k^2 = Y^*Y = (CX)^*(CX) = X^*(C^*C)X = X^*X = \sum_{k=1}^n X_k^2.$$

De plus,

$$\begin{aligned} (n-1)S^2 &= \sum_{k=1}^n (X_k - \bar{X})^2 = \sum_{k=1}^n (X_k^2 - 2\bar{X}X_k + \bar{X}^2) \\ &= \sum_{k=1}^n X_k^2 - 2\bar{X} \sum_{k=1}^n X_k + n\bar{X}^2 = \sum_{k=1}^n X_k^2 + n(-2\bar{X}^2 + \bar{X}^2) \\ &= \sum_{k=1}^n X_k^2 - n\bar{X}^2 = \sum_{k=1}^n Y_k^2 - Y_n^2 = \sum_{k=1}^{n-1} Y_k^2. \end{aligned}$$

Ainsi $\bar{X} = \frac{1}{\sqrt{n}} Y_n$ et $S^2 = \frac{1}{n-1} \sum_{k=1}^{n-1} Y_k^2$. Le vecteur Y est gaussien et sa matrice de dispersion est diagonale donc (Y_1, \dots, Y_n) est indépendant. Par suite \bar{X} et S^2 sont indépendantes en vertu du théorème d'indépendance des fonctions de v.a.r. (cf proposition 4.14, page 75).

6. La fonction caractéristique de \bar{X} est

$$\phi_{\bar{X}}(t) = \mathbb{E}[e^{it\bar{X}}] = \mathbb{E}[e^{i\frac{t}{\sqrt{n}}Y_n}] = \phi_{Y_n}\left(\frac{t}{\sqrt{n}}\right) = e^{-\frac{1}{2n}t^2}.$$

Donc $\mathbb{P}_{\bar{X}} = \mathcal{N}_1(0, \frac{1}{n})$. Enfin, d'après la deuxième question, le vecteur aléatoire (Y_1, \dots, Y_{n-1}) suit la loi $\mathcal{N}_{n-1}(0, I_{n-1})$, la v.a.r. $(n-1)S^2$ suit la loi $\chi^2(n-1)$.

Corrigé de l'exercice 5.10, page 104

Montrons que, pour tout entier $n \geq 1$, il existe une forme linéaire L_n sur \mathbb{R}^{n+1} telle que $X_n = L_n(X_0, \varepsilon_1, \dots, \varepsilon_n)$. Pour $n = 1$, posons $L_1(x_0, x_1) := l_1(x_0) + b_1x_1$ et remarquons que $X_1 = L_1(X_0, \varepsilon_1)$. Supposons construites les formes linéaires L_i pour $i = 1, \dots, n$ et construisons L_{n+1} . On a

$$\begin{aligned} X_{n+1} &= l_{n+1}(X_0, X_1, \dots, X_n) + b_{n+1}\varepsilon_{n+1} \\ &= l_{n+1}(X_0, L_1(X_0, \varepsilon_1), \dots, L_n(X_0, \varepsilon_1, \dots, \varepsilon_n)) + b_{n+1}\varepsilon_{n+1}. \end{aligned}$$

Posons

$$L_{n+1}(x_0, \dots, x_{n+1}) := l_{n+1}(x_0, L_1(x_0, x_1), \dots, L_n(x_0, x_1, \dots, x_n)) + b_{n+1}x_{n+1}.$$

L_{n+1} est bien une forme linéaire sur \mathbb{R}^{n+2} et $X_{n+1} = L_{n+1}(X_0, \varepsilon_1, \dots, \varepsilon_{n+1})$, ce qui prouve l'existence de la suite $(L_n)_{n \geq 1}$.

Maintenant, pour tout $n \geq 0$, construisons l'endomorphisme A_n de \mathbb{R}^{n+1} en posant, pour tout $(x_0, x_1, \dots, x_n) \in \mathbb{R}^{n+1}$,

$$A_{n+1}(x_0, \dots, x_n) := (x_0, L_1(x_0, x_1), \dots, L_n(x_0, \dots, x_n)).$$

Avec ces notations,

$$(X_0, X_1, \dots, X_n) = A_n(X_0, \varepsilon_1, \dots, \varepsilon_n).$$

Comme les composantes de $(X_0, \varepsilon_1, \dots, \varepsilon_n)$ sont gaussiennes et mutuellement indépendantes, ce vecteur est gaussien. Par suite le vecteur (X_0, X_1, \dots, X_n) est également gaussien comme image d'un vecteur gaussien par une application linéaire.

8.6 Corrigés des exercices du chapitre VI

Dans ce paragraphe, toutes les variables aléatoires considérées sont supposées définies sur un même espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$.

Corrigé de l'exercice 6.1, page 107

1. D'après l'inégalité de Bienaymé-Tchébycheff, $\mathbb{P}(|X - m| \geq 3\sigma) \leq \frac{\sigma^2}{9\sigma^2} = \frac{1}{9}$ et donc

$$\mathbb{P}(m - 3\sigma < X < m + 3\sigma) = 1 - \mathbb{P}(|X - m| \geq \sigma) \geq \frac{8}{9} \simeq 0.88.$$

Un raisonnement analogue donne l'autre partie de la question.

2. Comme X est de loi $\mathcal{N}(m; \sigma^2)$, on utilise les tables numériques (voir mode d'emploi dans l'annexe B, page 211) et on obtient en notant $T = (X - m)/\sigma$ qui est de loi gaussienne centrée réduite

$$\begin{aligned} \mathbb{P}(m - 3\sigma < X < m + 3\sigma) &= \mathbb{P}\left(-3 < \frac{X - m}{\sigma} < 3\right) \\ &= \mathbb{P}(-3 < T < 3) = \mathbb{P}(T < 3) - \mathbb{P}(T < -3) \\ &= \mathbb{P}(T < 3) - [1 - \mathbb{P}(T < 3)] = 2\mathbb{P}(T < 3) - 1 \\ &\simeq 2 \times 0.99865 - 1 \simeq 0.9973, \end{aligned}$$

où l'on a lu $\mathbb{P}(T < 3)$ dans la partie basse de la table correspondant aux grandes valeurs de u . On trouverait de même $\mathbb{P}(m - 2\sigma < X < m + 2\sigma) \simeq 0.9544$.

Corrigé de l'exercice 6.2, page 107

1. En utilisant le fait que X est d'espérance nulle on a :

$$\begin{aligned} a &= \mathbb{E}[a - X] = \mathbb{E}[(a - X)\mathbf{1}_{(X \leq a)} + (a - X)\mathbf{1}_{(X > a)}] \\ &\leq \mathbb{E}[(a - X)\mathbf{1}_{(X \leq a)}] \leq (\mathbb{P}(X \leq a))^{\frac{1}{2}} (\mathbb{E}[(a - X)^2])^{\frac{1}{2}} \end{aligned}$$

où la dernière inégalité est obtenue en utilisant l'inégalité de Cauchy-Schwarz. De plus $\mathbb{E}[(a - X)^2] = a^2 - 2a\mathbb{E}[X] + \mathbb{E}[X^2] = a^2 + \sigma^2$ ce qui donne :

$$a \leq \mathbb{E}[(a - X)\mathbf{1}_{(X \leq a)}] \leq (\mathbb{P}(X \leq a))^{\frac{1}{2}} \sqrt{a^2 + \sigma^2}.$$

En élevant au carré l'inégalité précédente on obtient :

$$\mathbb{P}(X > a) = 1 - \mathbb{P}(X \leq a) \leq 1 - \frac{a^2}{a^2 + \sigma^2} = \frac{\sigma^2}{a^2 + \sigma^2}.$$

2. Si on pose $X = Y - 100$ alors X est une v.a. centrée de variance

$$\text{Var}(X) = \mathbb{E}[(Y - 100)^2] = \mathbb{E}[(Y - \mathbb{E}[Y])^2] = \text{Var}(Y) = 400.$$

L'inégalité de la question 1) nous conduit à

$$\mathbb{P}(Y > 120) = \mathbb{P}(X > 20) \leq \frac{400}{20^2 + 400} = \frac{1}{2}.$$

Avec l'inégalité de Bienaymé-Tchebycheff on obtient :

$$\mathbb{P}(Y > 120) = \mathbb{P}(X > 20) \leq \mathbb{P}(|X| > 20) \leq \frac{400}{20^2} = 1$$

qui ne donne pas de renseignement.

Corrigé de l'exercice 6.3, page 109

Notons X_k la $k^{\text{ième}}$ mesure effectuée sur les N . On peut considérer que X_k est une v.a.r. d'espérance m , que les v.a.r. X_1, \dots, X_N sont indépendantes et de même loi. On note

$\bar{X}_N = \frac{1}{N} \sum_{k=1}^N X_k$ la moyenne empirique des valeurs observées. Par l'inégalité de Bienaymé-

Tchébycheff, $\mathbb{P}(|\bar{X}_N - m| \geq 0.05) \leq \frac{\text{Var}(\bar{X}_N)}{(0.05)^2}$. Or $\text{Var}(\bar{X}_N) = \frac{\text{Var}(X_1)}{N} = \frac{0.25}{N}$ et donc

$\mathbb{P}(|\bar{X}_N - m| \geq 0.05) \leq \frac{0.25}{N(0.05)^2}$. On cherche donc N tel que $\frac{0.25}{N(0.05)^2} \leq 0.01$ d'où

$$N \geq \frac{0.25}{(0.05)^2 \times 0.01} = 10000.$$

Corrigé de l'exercice 6.4, page 110

Soit $a > 0$, prenons un entier n suffisamment grand pour que $0 < 1/n < a$. On veut étudier la limite quand $n \rightarrow \infty$ de $\mathbb{P}(|X_n - 0| \geq a)$. Pour n tel que $1/n < a$, on a

$$\begin{aligned} \mathbb{P}(|X_n - 0| \geq a) &= \mathbb{P}(|X_n| \geq a) = \mathbb{P}(X_n \leq -a) + \mathbb{P}(X_n \geq a) \\ &= \int_{-\infty}^{-a} n \underbrace{\mathbf{1}_{[0, \frac{1}{n}]}(x)}_{=0 \text{ sur }]-\infty, -a[} dx + \int_a^{+\infty} n \underbrace{\mathbf{1}_{[0, \frac{1}{n}]}(x)}_{=0 \text{ sur } [a, +\infty[} dx \end{aligned}$$

d'où pour tout $a > 0$, $\lim_{n \rightarrow \infty} \mathbb{P}(|X_n - 0| \geq a) = 0$ ce qui prouve que $(X_n)_{n \in \mathbb{N}}$ converge en probabilité vers 0.

Corrigé de l'exercice 6.5, page 112

On considère l'exemple 6.3, page 110, du cours. La suite $(X_n)_{n \in \mathbb{N}}$ converge en probabilité vers la v.a.r. 0. Mais pour tout entier $n \geq 1$, $\mathbb{E}(X_n) = \frac{n}{n+1} + \frac{1}{n} \frac{n}{n+1} = 1$. Donc la suite des espérances $(\mathbb{E}(X_n))_{n \in \mathbb{N}^*}$ ne converge pas vers l'espérance de la limite qui est $0 = \mathbb{E}(0)$.

Corrigé de l'exercice 6.6, page 114

1) Comme $X_n \xrightarrow[n \rightarrow \infty]{\text{p.s.}} X$, il existe Δ_X tel que $\mathbb{P}(\Delta_X) = 0$ et pour tout $\omega \in \Delta_X^c$, $\lim_{n \rightarrow \infty} X_n(\omega) = X(\omega)$. De même il existe Δ_Y de probabilité nulle tel que pour tout $\omega \in \Delta_Y^c$, $\lim_{n \rightarrow \infty} Y_n(\omega) = Y(\omega)$. Posons $\Delta = \Delta_X \cup \Delta_Y$, alors pour tout $\omega \in \Delta^c$, $\lim_{n \rightarrow \infty} X_n(\omega) = X(\omega)$ et $\lim_{n \rightarrow \infty} Y_n(\omega) = Y(\omega)$. Comme la fonction f est continue sur \mathbb{R}^2 , la suite $(f(X_n(\omega), Y_n(\omega)))_{n \in \mathbb{N}}$ converge vers $f(X(\omega), Y(\omega))$. De plus, $0 \leq \mathbb{P}(\Delta) \leq \mathbb{P}(\Delta_X) + \mathbb{P}(\Delta_Y)$ et donc $\mathbb{P}(\Delta) = 0$ ce qui prouve la convergence presque-sûre de $(f(X_n, Y_n))_{n \in \mathbb{N}}$ vers $f(X, Y)$.

2) Il suffit d'appliquer le premier point avec $f : (x, y) \rightarrow x + y$ et $g : (x, y) \rightarrow xy$ qui sont des applications continues sur \mathbb{R}^2 .

Corrigé de l'exercice 6.7, page 116

On commence par remarquer que $\sum_{k=1}^n X_k = n\bar{X}_n$ et on a

$$\begin{aligned} \sum_{k=1}^n (X_k - \bar{X}_n)^2 &= \sum_{k=1}^n (X_k^2 - 2X_k\bar{X}_n + \bar{X}_n^2) \\ &= \left(\sum_{k=1}^n X_k^2 \right) - 2\bar{X}_n \left(\sum_{k=1}^n X_k \right) + n\bar{X}_n^2 \\ &= \left(\sum_{k=1}^n X_k^2 \right) - 2n\bar{X}_n^2 + n\bar{X}_n^2 = \left(\sum_{k=1}^n X_k^2 \right) - n\bar{X}_n^2 \end{aligned}$$

d'où la relation pour S_n^2 .

D'après la loi forte des grands nombres appliquée à la suite $(X_n)_{n \in \mathbb{N}}$ où la moyenne commune à tous les X_k est $m = \mathbb{E}(X_0)$ et la variance commune est $\sigma^2 = \text{Var}(X_0)$. On a

$$\bar{X}_n \xrightarrow[n \rightarrow \infty]{\text{p.s.}} m \quad \text{et} \quad \bar{X}_n^2 \xrightarrow[n \rightarrow \infty]{\text{p.s.}} m^2.$$

De même la suite $(X_k^2)_{k \in \mathbb{N}}$ est une suite i.i.d. et on peut donc lui appliquer la loi forte des grands nombres qui prouve que

$$\frac{1}{n} \sum_{k=1}^n X_k^2 \xrightarrow[n \rightarrow \infty]{\text{p.s.}} \mathbb{E}(X_0^2).$$

Ainsi

$$S_n^2 = \underbrace{\frac{n}{n-1}}_{\rightarrow 1} \times \underbrace{\frac{1}{n} \sum_{k=1}^n X_k^2}_{\rightarrow \mathbb{E}(X_0^2)} - \underbrace{\frac{n}{n-1}}_{\rightarrow 1} \times \underbrace{\bar{X}_n^2}_{\rightarrow m^2} \xrightarrow[n \rightarrow \infty]{\text{p.s.}} \mathbb{E}(X_0^2) - m^2 = \sigma^2.$$

Corrigé de l'exercice 6.8, page 121

Posons, pour tout $n \geq 1$, $S_n := X_1 + \cdots + X_n$. En utilisant le fait que $\mathbb{E}(S_n) = n\mu$ et que la variance d'une somme de variables non corrélées est égale à la somme des variances, on obtient

$$\mathbb{E} \left[\left(\frac{S_n}{n} - \mu \right)^2 \right] = \text{Var} \left(\frac{S_n}{n} \right) = \frac{1}{n^2} \sum_{i=1}^n \text{Var}(X_i) \leq \frac{C}{n}.$$

On a donc la convergence vers μ dans L^2 , et par conséquent en probabilité, de la suite $(\frac{S_n}{n})_{n \geq 1}$.

Corrigé de l'exercice 6.9, page 121

Posons, pour tout entier $n \geq 1$, $I_n := \sum_{k=1}^n \frac{1}{n} f(U_k)$ et remarquons que

$$\mathbb{E}(I_n) = \mathbb{E}(f(U_1)), \quad \text{Var}(I_n) = \frac{1}{n^2} \text{Var} \left(\sum_{k=1}^n f(U_k) \right) = \frac{1}{n} \text{Var}(f(U_1)).$$

Pour obtenir ces deux résultats, on utilise l'indépendance et l'identité des lois des v.a.r. $f(U_k)$. De plus comme f est une application de carré intégrable, la constante $C := \text{Var}(f(U_1))$ est finie.

Soit $\delta > 0$, appliquons l'inégalité de Bienaymé-Tchebycheff à la v.a.r. $I_n \in L^2$. Il vient, pour tout entier $n \geq 1$,

$$\mathbb{P}(|I_n - \mathbb{E}(I_n)| \geq \delta) \leq \frac{\text{Var}(I_n)}{\delta^2} \text{ d'où } \mathbb{P}(|I_n - \mathbb{E}(f(U_1))| \geq \delta) \leq \frac{C}{n\delta^2}$$

en reportant les expressions trouvées ci-dessus. Cela prouve que, pour tout $\delta > 0$, $\lim_n \mathbb{P}(|I_n - \mathbb{E}(I_n)| \geq \delta) = 0$, c'est-à-dire que la suite de v.a.r. $(I_n)_{n \in \mathbb{N}^*}$ converge en probabilité vers $\mathbb{E}(f(U_1))$.

Calculons $\mathbb{E}(f(U_1))$. Par le théorème de transfert il vient

$$\begin{aligned} \mathbb{E}(f(U_1)) &= \int_{\Omega} f(U_1) d\mathbb{P} = \int_{\mathbb{R}} f(x) d\mathbb{P}_{U_1}(x) \\ &= \int_{\mathbb{R}} f(x) \mathbb{1}_{[0,1]} d\lambda(x) = \int_{[0,1]} f d\lambda \end{aligned}$$

ce qui achève la démonstration.

Corrigé de l'exercice 6.10, page 121

Pour tout $i \geq 1$, X_i est une v.a.r. intégrable. La loi forte des grands nombres implique

$$\frac{1}{n} \sum_{i=1}^n X_i \xrightarrow{p.s.} \mathbb{E}[X_1] = \frac{1}{2},$$

d'où, comme f est continue,

$$f\left(\frac{1}{n} \sum_{i=1}^n X_i\right) \xrightarrow{p.s.} f\left(\frac{1}{2}\right).$$

Comme f est bornée, il existe une constante M telle que, pour tout $n \geq 1$, $|f(\frac{1}{n} \sum_{i=1}^n X_i)| \leq M$.

Par application du théorème de convergence dominée de Lebesgue

$$\mathbb{E} \left[f\left(\frac{1}{n} \sum_{i=1}^n X_i\right) \right] \longrightarrow \mathbb{E} \left[f\left(\frac{1}{2}\right) \right] = f\left(\frac{1}{2}\right).$$

On applique maintenant le théorème du transfert à la variable aléatoire vectorielle (X_1, \dots, X_n) , il vient

$$\begin{aligned} \mathbb{E} \left[f\left(\frac{1}{n} \sum_{i=1}^n X_i\right) \right] &= \int_{\mathbb{R}^n} f\left(\frac{x_1 + \dots + x_n}{n}\right) d\mathbb{P}_{(X_1, \dots, X_n)}(x_1, \dots, x_n) \\ &= \int_{\mathbb{R}^n} f\left(\frac{x_1 + \dots + x_n}{n}\right) \prod_{i=1}^n \mathbb{1}_{[0,1]}(x_i) d\lambda^{(n)}(x_1, \dots, x_n) \\ &= \int_{[0,1]^n} f\left(\frac{x_1 + \dots + x_n}{n}\right) d\lambda^{(n)}(x_1, \dots, x_n). \end{aligned}$$

en conclusion,

$$\lim_{n \rightarrow +\infty} \int_{[0,1]^n} f\left(\frac{x_1 + \dots + x_n}{n}\right) d\lambda^{(n)}(x_1, \dots, x_n) = f\left(\frac{1}{2}\right).$$

Corrigé de l'exercice 6.11, page 121

1. Comme f est continue sur le compact $[0, 1]$ elle est bornée et en particulier $\sup_{0 \leq x \leq 1} |f(x)|$ est fini, de plus la v.a.r. $f(\frac{1}{n} S_n)$ est intégrable. Par le théorème de transfert il vient

$$\begin{aligned} p_n(x) &:= \mathbb{E}[f(\frac{1}{n} S_n)] = \int_{\mathbb{R}} f\left(\frac{t}{n}\right) d\left(\sum_{k=0}^n C_n^k x^k (1-x)^{n-k} \delta_k\right)(t) \\ &= \sum_{k=0}^n f\left(\frac{k}{n}\right) C_n^k x^k (1-x)^{n-k}. \end{aligned}$$

2. Comme f est continue sur le compact $[0, 1]$ elle est uniformément continue sur $[0, 1]$. Fixons $\varepsilon > 0$, il existe alors $\delta > 0$ tel que, pour tout $(x, y) \in [0, 1]^2$, $|x - y| < \delta$ implique $|f(x) - f(y)| < \varepsilon$. De plus

$$|p_n(x) - f(x)| = |\mathbb{E}[f(\frac{1}{n} S_n) - f(x)]| \leq \mathbb{E} \left[|f(\frac{1}{n} S_n) - f(x)| \right].$$

Considérons l'événement $A_n := \{|\frac{1}{n} S_n - x| < \delta\}$, il vient

$$\begin{aligned} \mathbb{E} \left[|f(\frac{1}{n} S_n) - f(x)| \right] &= \mathbb{E} \left[\mathbb{1}_{A_n} |f(\frac{1}{n} S_n) - f(x)| \right] + \mathbb{E} \left[\mathbb{1}_{A_n^c} |f(\frac{1}{n} S_n) - f(x)| \right] \\ &\leq \varepsilon \mathbb{E}(\mathbb{1}_{A_n}) + 2\mathbb{E}(\mathbb{1}_{A_n^c}) \sup_{0 \leq x \leq 1} |f(x)| \\ &\leq \varepsilon \mathbb{P}(A_n) + 2\mathbb{P}(A_n^c) \sup_{0 \leq x \leq 1} |f(x)| \\ &\leq \varepsilon + 2\mathbb{P} \left(\left| \frac{1}{n} S_n - x \right| \geq \delta \right) \sup_{0 \leq x \leq 1} |f(x)| \end{aligned}$$

ce qui prouve la deuxième inégalité.

Utilisons maintenant l'inégalité de Bienaymé-Tchebycheff pour majorer le deuxième terme du second membre de l'inégalité précédente.

Comme $\{|\frac{1}{n}S_n - x| \geq \delta\} = \{|S_n - \mathbb{E}(S_n)| \geq n\delta\}$ il vient

$$\mathbb{P}(|\frac{1}{n}S_n - x| \geq \delta) = \mathbb{P}(|S_n - \mathbb{E}(S_n)| \geq n\delta) \leq \frac{1}{(n\delta)^2} \text{Var}(S_n) = \frac{x(1-x)}{n\delta^2}$$

ce qui en revenant à la deuxième inégalité donne, pour tout $n \in \mathbb{N}^*$ et tout $x \in [0, 1]$,

$$|p_n(x) - f(x)| \leq \varepsilon + 2 \frac{x(1-x)}{n\delta^2} \sup_{0 \leq x \leq 1} |f(x)|.$$

3. D'après la question précédente, pour tout $n \in \mathbb{N}^*$, on obtient

$$\sup_{0 \leq x \leq 1} |p_n(x) - f(x)| \leq \varepsilon + \frac{1}{2n\delta^2} \sup_{0 \leq x \leq 1} |f(x)|$$

où on a utilisé $\sup_{0 \leq x \leq 1} (x(1-x)) = \frac{1}{4}$ (Étudier la fonction numérique $h(x) := x(1-x)$ sur $[0, 1]$).

Par suite, pour tout $\varepsilon > 0$, il existe $N \in \mathbb{N}$ tel que, pour tout $n \in \mathbb{N}^*$, $n \geq N$ implique $\sup_{0 \leq x \leq 1} |p_n(x) - f(x)| < 2\varepsilon$, ce qui prouve que la suite de polynômes $(p_n)_{\mathbb{N}}$ converge uniformément sur $[0, 1]$ vers la fonction f .

Corrigé de l'exercice 6.12, page 122

1. Appliquons le résultat de l'exercice 4.17, page 92, à la v.a.r. positive $|X|$ pour $n = 1$. Cela donne $\mathbb{E}(|X|) = \int_0^{+\infty} \mathbb{P}(|X| > t) dt$. On peut alors écrire

$$\begin{aligned} \mathbb{E}(|X|) &= \int_0^{+\infty} \sum_{n \geq 0} \mathbb{1}_{[n, n+1[}(t) \mathbb{P}(|X| > t) dt \\ &= \sum_{n \geq 0} \int_0^{+\infty} \mathbb{1}_{[n, n+1[}(t) \mathbb{P}(|X| > t) dt \leq \sum_{n \geq 0} \mathbb{P}(|X| \geq n) \end{aligned}$$

où pour obtenir la dernière inégalité on a utilisé la majoration

$$\mathbb{1}_{[n, n+1[}(t) \mathbb{P}(|X| > t) \leq \mathbb{1}_{[n, n+1[}(t) \mathbb{P}(|X| \geq n).$$

2. Soit $\omega \in \{(\frac{1}{n}S_n)_{\mathbb{N}^*} \text{ converge dans } \mathbb{R}\}$. Comme

$$\frac{S_n(\omega)}{n} - \frac{S_{n+1}(\omega)}{n+1} = \frac{S_n(\omega)}{n(n+1)} - \frac{X_{n+1}(\omega)}{n+1}$$

et que

$$\lim_{n \rightarrow +\infty} \left| \frac{S_n(\omega)}{n} - \frac{S_{n+1}(\omega)}{n+1} \right| = 0 = \lim_{n \rightarrow +\infty} \left| \frac{S_n(\omega)}{n(n+1)} \right|,$$

la suite $\left(\frac{1}{n+1}X_{n+1}(\omega)\right)_n$ converge vers 0. Par suite seuls un nombre fini d'entiers n vérifient $\frac{1}{n}X_n(\omega) \geq 1$ ou encore $|X_n(\omega)| \geq n$. ω ne peut donc pas appartenir à l'ensemble $\limsup_n \{|X_n| \geq n\}$, ce qui prouve l'inclusion cherchée.

3. Comme les v.a.r. sont toutes de même loi, pour tout $n \in \mathbb{N}$, nous avons $\mathbb{P}(|X_1| \geq n) = \mathbb{P}(|X_n| \geq n)$ et par suite, en revenant à l'inégalité de la première question appliquée à la v.a.r. X_1 ,

$$\mathbb{E}(|X_1|) \leq \sum_{n \geq 0} \mathbb{P}(|X_1| \geq n) = \sum_{n \geq 0} \mathbb{P}(|X_n| \geq n).$$

Comme X_1 n'est pas intégrable, $\mathbb{E}(|X_1|) = +\infty$ et par suite en vertu de l'inégalité précédente, la série $\sum_n \mathbb{P}(|X_n| \geq n)$ diverge. Comme la suite $(X_n)_n$ est indépendante, les événements $\{|X_n| \geq n\}$ sont indépendants. Par application du lemme de Borel-Cantelli on en déduit que $\mathbb{P}(\limsup_n \{|X_n| \geq n\}) = 1$ ou encore $\mathbb{P}\left(\left[\limsup_n \{|X_n| \geq n\}\right]^c\right) = 0$. Mais de l'inclusion de la deuxième question on déduit que

$$\mathbb{P}\left(\left(\frac{1}{n}S_n\right)_{n \in \mathbb{N}^*} \text{ converge dans } \mathbb{R}\right) = 0.$$

Ce qui signifie que la suite de v.a.r. $\left(\frac{1}{n}S_n\right)_{n \in \mathbb{N}^*}$ ne converge pas presque-sûrement dans \mathbb{R} .

Corrigé de l'exercice 6.13, page 122

1. Considérons l'événement $A := \left\{\left(\frac{1}{n}S_n\right)_{n \in \mathbb{N}^*} \text{ converge vers } m\right\}$. Soit $\omega \in A$, alors la suite $(S_n(\omega))_{n \in \mathbb{N}^*}$ converge vers $+\infty$ si $m > 0$ ou $-\infty$ si $m < 0$. Par suite l'ensemble $\{n \in \mathbb{N} / S_n(\omega) \in I\}$ est fini car l'intervalle I est supposé borné; ce qui signifie que $\omega \in \left(\limsup_n \{S_n \in I\}\right)^c$. On a donc l'inclusion d'événements $A \subseteq \left(\limsup_n \{S_n \in I\}\right)^c$.
2. Il s'agit de montrer que $\mathbb{P}\left[\left(\limsup_n \{S_n \in I\}\right)^c\right] = 1$. D'après la loi forte des grands nombres, l'événement $A = \left\{\left(\frac{1}{n}S_n\right)_{n \in \mathbb{N}^*} \text{ converge vers } m\right\}$ est de probabilité égale à 1, c'est-à-dire $\mathbb{P}(A) = 1$. Ce qui donne, compte-tenu de l'inclusion démontrée en 1), le résultat cherché.

8.7 Corrigés des exercices du chapitre VII

Dans ce paragraphe, toutes les variables aléatoires considérées sont supposées définies sur un même espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$.

Corrigé de l'exercice 7.1, page 127

- 1) Considérons pour $k = 1, \dots, N$ la v.a.r. $X_k = 1$ si la $k^{\text{ième}}$ va dans la salle 1 et $X_k = 0$ sinon (elle va alors dans la salle 2). La v.a.r. X_k est de loi de Bernoulli de paramètre $1/2$. Comme le choix des spectateurs est supposé indépendant, X_1, \dots, X_N sont des v.a.r. indépendantes. Le nombre de spectateurs qui désirent aller dans la salle 1 est donc $S = S_N = X_1 + \dots + X_N$ qui

est une v.a.r. $\mathcal{B}(N; \frac{1}{2})$. L'événement "tous les spectateurs ne peuvent pas voir le film qu'ils ont choisi" se modélise par :

$$\{S > n\} \cup \{N - S > n\} = \{S > n\} \cup \{S < N - n\}.$$

On remarque que si $N > 2n$, on est sûr qu'il y a au moins un spectateur qui ne verra pas son film. Dans ce cas, $P = 1$. De même, si $N < n$, on est sûr que tous les spectateurs pourront voir leur film. Dans ce cas, $P = 0$.

Étudions le cas où $n \leq N \leq 2n$ c'est-à-dire $0 \leq N - n \leq n$. Dans ce cas, $\{S > n\} \cap \{S < N - n\} = \emptyset$ et donc $P = \mathbb{P}(S > n) + \mathbb{P}(S \leq N - n)$. D'après le théorème central limite (N est implicitement supposé grand), on a

$$\lim_{N \rightarrow \infty} \mathbb{P}\left(\frac{\frac{S}{N} - \frac{1}{2}}{\frac{1}{2\sqrt{N}}} \leq x\right) = \int_{-\infty}^x \frac{1}{\sqrt{2\pi}} e^{-u^2/2} du.$$

Si on note $T = \frac{\frac{S}{N} - \frac{1}{2}}{\frac{1}{2\sqrt{N}}}$, c'est une v.a.r. asymptotiquement de loi $\mathcal{N}(0,1)$ et donc

$$\begin{aligned} P &= \mathbb{P}\left(T > \frac{n - \frac{N}{2}}{\sqrt{N/4}}\right) + \mathbb{P}\left(T < \frac{N - n - \frac{N}{2}}{\sqrt{N/4}}\right) \\ &= \mathbb{P}\left(|T| \geq \frac{n - \frac{N}{2}}{\sqrt{N/4}}\right) = 2\left(1 - \Phi[(2n - N)/\sqrt{N}]\right), \end{aligned}$$

où Φ désigne la fonction de répartition de la loi $\mathcal{N}(0,1)$.

2) Si $N = 1000$ et si on veut $P \leq 0.01$, il faut choisir n pour que $\Phi[(2n - 1000)/\sqrt{1000}] \geq 0.95$. La lecture inverse de la table de $\mathcal{N}(0,1)$ donne $\Phi(2.58) \simeq 0.95$ d'où $(2n - 1000)/\sqrt{1000} \geq 2.58$ et par suite il faut prendre $n \geq 541$.

Corrigé de l'exercice 7.2, page 129

La fonction de répartition de δ_n est $\mathbb{1}_{[n, \infty[}$ pour tout $n \in \mathbb{N}$. De plus $\lim_{n \rightarrow \infty} \mathbb{1}_{[n, \infty[}(x) = 0$ donc la limite de la suite de fonctions de répartition $(\mathbb{1}_{[n, \infty[})_{n \in \mathbb{N}}$ est la fonction-nulle qui n'est pas une fonction de répartition car elle ne vérifie pas $\lim_{x \rightarrow \infty} F(x) = 1$.

On a donc la convergence simple de la suite de fonction de répartition $(F_{\delta_n})_{n \in \mathbb{N}}$ mais la suite $(\delta_n)_{n \in \mathbb{N}}$ ne converge pas étroitement vers une limite μ , sinon on aurait $\lim_{n \rightarrow \infty} F_{\delta_n}(x) = F_{\mu}(x) = 0$ pour tout point x de continuité de F_{μ} et donc $F_{\mu} = 0$, ce qui est impossible.

Corrigé de l'exercice 7.3, page 131

On sait par le théorème de stabilité de la somme de variables de Poisson indépendantes, que la loi de S_n est $\mathcal{P}(n)$. Remarquons alors que

$$e^{-n} \sum_{k=0}^{k=n} \frac{n^k}{k!} = \mathbb{P}(S_n \leq n) \text{ et } \{S_n \leq n\} = \left\{\frac{1}{\sqrt{n}}(S_n - n) \leq 0\right\}.$$

Par suite,

$$e^{-n} \sum_{k=0}^{k=n} \frac{n^k}{k!} = \mathbb{P}\left(\frac{1}{\sqrt{n}}(S_n - n) \leq 0\right).$$

Le théorème-limite central permet d'affirmer que la suite $\left(\frac{1}{\sqrt{n}}(S_n - n)\right)_{n \geq 1}$ converge en loi vers une variable Y de loi $\mathcal{N}_1(0, 1)$. En particulier $\mathbb{P}\left(\frac{1}{\sqrt{n}}(S_n - n) \leq 0\right)$ tend vers $\mathbb{P}(Y \leq 0) = \frac{1}{2}$ quand n tend vers $+\infty$. On conclut donc que la suite $\left(e^{-n} \sum_{k=0}^n \frac{n^k}{k!}\right)_{n \geq 1}$ converge vers $\frac{1}{2}$.

Corrigé de l'exercice 7.4, page 132

Soit X le nombre de transistors défectueux dans un sachet de 100, X suit une loi $\mathcal{B}(100; 0.01)$. L'événement "la garantie tombe en défaut" se modélise par $\{X \geq 3\}$. Comme le proposent les commentaires qui suivent la proposition 7.14, page 132, on approxime la loi binomiale par la loi de Poisson de paramètre 1 car on a $n = 100 \geq 30$, $np = 100 \times 0.01 < 10$ et $p \leq 0.1$. D'où

$$\mathbb{P}(X \geq 3) = 1 - \mathbb{P}(X \leq 2) \simeq 1 - \mathbb{P}(\mathcal{P}(1) \leq 2) \simeq 1 - 0.92 \simeq 8\% .$$

Corrigé de l'exercice 7.5, page 133

Pour tout $N \in S$, posons $\mu_N := \mathcal{H}(N, n, p)$. μ_N est une probabilité discrète portée par l'ensemble des entiers compris entre 0 et n . D'après le critère de convergence des probabilités discrètes, il suffit de montrer que, pour tout entier k compris entre 0 et n , la suite $(\mu_N(\{k\}))_{N \in S}$ converge vers $\mathcal{B}(n, p)(\{k\}) = C_n^k p^k q^{n-k}$.

Soit k un entier compris entre 0 et n , après explicitation des coefficients binomiaux, on peut écrire

$$\mu_N(\{k\}) = C_n^k \frac{[(Np)(Np-1)\dots(Np-k+1)][(Nq)(Nq-1)\dots(Nq-n+k+1)]}{[N(N-1)\dots(N-n+1)]}.$$

Mais, pour N voisin de l'infini,

$$[(Np)(Np-1)\dots(Np-k+1)], \quad [(Nq)(Nq-1)\dots(Nq-n+k+1)]$$

et

$$[N(N-1)\dots(N-n+1)]$$

sont respectivement équivalents à $(Np)^k$, $(Nq)^{n-k}$ et N^n . Par suite, pour N voisin de l'infini, $\mu_N(\{k\})$ est équivalent à $C_n^k p^k q^{n-k}$, ce qui termine la démonstration. Ce résultat exprime qu'une loi hypergéométrique peut être approximée par une loi binomiale.

Corrigé de l'exercice 7.6, page 136

Pour tout $n \geq 1$, on vérifie facilement que l'application définie par

$$f_n(x) := \begin{cases} 1 - \cos(2\pi nx) & \text{si } x \in [0, 1] \\ 0 & \text{si } x \in [0, 1]^c \end{cases}$$

est une densité de probabilité. La fonction de répartition F_n associée à la probabilité de densité f_n est donnée par

$$F_n(t) = \int_{]-\infty, t]} f_n(x) d\lambda(x) = \begin{cases} 0 & \text{si } t \in]-\infty, 0], \\ t - \frac{1}{2\pi n} \sin(2\pi nt) & \text{si } t \in]0, 1], \\ 1 & \text{si } t \in]1, +\infty[. \end{cases}$$

La suite $(F_n)_{n \geq 1}$ converge simplement vers la fonction F définie par $F(t) := t\mathbb{1}_{[0,1]}(t) + \mathbb{1}_{[1,+\infty]}(t)$. F est la fonction de répartition de la loi uniforme $\mathcal{U}([0, 1])$ de densité $f = \mathbb{1}_{[0,1]}$. On a donc une convergence étroite des probabilités de densités f_n vers la loi uniforme, mais la suite des densités n'a de limite en aucun point de $]0, 1[$.

Corrigé de l'exercice 7.7, page 138

Si $(a_n)_{n \in \mathbb{N}}$ et $(\sigma_n)_{n \in \mathbb{N}}$ convergent respectivement vers les réels a et σ , alors $(a_n)_{n \in \mathbb{N}}$ et $(\sigma_n^2)_{n \in \mathbb{N}}$ convergent respectivement vers les réels a et σ^2 , et pour tout $t \in \mathbb{R}$, la suite $(\phi_{\mu_n}(t))_{n \in \mathbb{N}} = \left(\exp \left(ia_n t - \frac{1}{2} t^2 \sigma_n^2 \right) \right)_{n \in \mathbb{N}}$ converge vers $\exp \left(iat - \frac{1}{2} t^2 \sigma^2 \right)$. La suite des fonctions caractéristiques $(\phi_{\mathcal{N}_1(a_n, \sigma_n^2)})_{n \in \mathbb{N}}$ converge simplement vers la fonction caractéristique de la probabilité $\mathcal{N}_1(a, \sigma^2)$. On conclut alors en utilisant le critère des fonctions caractéristiques.

Corrigé de l'exercice 7.8, page 141

On obtient facilement, pour tout $n \geq 1$, la fonction de répartition F_n de μ_n : $F_n(t) = t^n \mathbb{1}_{[0,1]}(t) + \mathbb{1}_{[1,+\infty]}(t)$. Cette suite de fonctions de répartition converge simplement vers $\mathbb{1}_{[1,+\infty]}$ qui est la fonction de répartition de δ_1 . La suite $(\mu_n)_{n \geq 1}$ converge donc étroitement vers δ_1 . On remarquera qu'une suite de probabilités absolument continues peut converger étroitement vers une probabilité discrète.

Corrigé de l'exercice 7.9, page 141

On sait que la fonction caractéristique de la loi de Cauchy est définie par $\phi(t) = e^{-|t|}$.

- Pour tout $n \geq 1$, la fonction caractéristique de $\frac{1}{\sqrt{n}} S_n$ est donnée par

$$\phi_{\frac{1}{\sqrt{n}} S_n}(t) = \mathbb{E}[e^{it \frac{1}{\sqrt{n}} S_n}] = \left(\mathbb{E}[e^{it \frac{1}{\sqrt{n}} X_1}] \right)^n = e^{-|t| \sqrt{n}}$$

du fait de l'indépendance de la suite $(X_n)_{n \geq 1}$. Comme la suite $(e^{-|t| \sqrt{n}})_{n \geq 1}$ tend vers $\mathbb{1}_{\{0\}}(t)$ qui définit une application non continue en 0, c-à-d que la limite des fonctions caractéristiques n'est pas une fonction caractéristique, la suite $\left(\frac{1}{\sqrt{n}} S_n \right)_{n \geq 1}$ ne converge pas en loi et donc ne peut pas converger en probabilité.

- Pour tout $n \geq 1$, la fonction caractéristique de $\frac{1}{n} S_n$ est donnée par $\phi_{\frac{1}{n} S_n}(t) = e^{-|t|}$. La suite $\left(\frac{1}{n} S_n \right)_{n \geq 1}$ converge donc en loi vers une v.a.r. de loi de Cauchy. Si cette suite convergeait en probabilité, alors la suite $\left(\frac{S_n}{n} - \frac{S_{2n}}{2n} \right)_{n \geq 1}$ convergerait en probabilité vers 0 et donc convergerait en loi vers 0. Or la fonction caractéristique de $\frac{S_n}{n} - \frac{S_{2n}}{2n}$ est donnée par

$$\phi_{\frac{S_n}{n} - \frac{S_{2n}}{2n}}(t) = \mathbb{E}[e^{it \frac{1}{2n} (X_1 + \dots + X_n - (X_{n+1} + \dots + X_{2n}))}] = e^{-|t|}$$

qui ne converge pas vers $1_{\mathbb{R}}$, fonction caractéristique de la v.a.r. constante 0, lorsque n tend vers $+\infty$. Donc la suite $\left(\frac{1}{n} S_n \right)_{n \geq 1}$ ne converge pas en probabilité.

- Pour tout $n \geq 1$, la fonction caractéristique de $\frac{1}{n^2}S_n$ est donnée par $\phi_{\frac{1}{n^2}S_n}(t) = e^{-\frac{|t|}{n}}$ qui converge en $+\infty$ vers 1, fonction caractéristique de la v.a.r. 0. La suite $\left(\frac{1}{n^2}S_n\right)_{n \geq 1}$ converge donc en loi vers 0 et d'après la première question, elle converge aussi en probabilité vers 0.

Corrigé de l'exercice 7.10, page 142

1. Pour tous $t \in \mathbb{R}$, $\alpha > 0$, et $n \in \mathbb{N}$,

$$\begin{aligned} |\phi_{X_n+Y_n}(t) - \phi_{X_n}(t)| &= |\mathbb{E}[e^{itX_n}(e^{itY_n} - 1)]| \leq \mathbb{E}[|e^{itY_n} - 1|] \\ &= \int_{\{|Y_n| > \alpha\}} |e^{itY_n} - 1| d\mathbb{P} + \int_{\{|Y_n| \leq \alpha\}} |e^{itY_n} - 1| d\mathbb{P} \\ &\leq 2\mathbb{P}(|Y_n| > \alpha) + \int_{\{|Y_n| \leq \alpha\}} |e^{itY_n} - 1| d\mathbb{P}. \end{aligned}$$

2. Remarquons que $(Y_n)_{\mathbb{N}}$ converge en loi vers la v.a.r. 0 si et seulement si $(Y_n)_{\mathbb{N}}$ converge en probabilité vers la v.a.r. 0 (cf. exercice V-6). Soit $\varepsilon > 0$, il existe $\alpha_0 > 0$ tel que pour $|y| \leq \alpha_0$ on ait $|e^{ity} - 1| \leq \varepsilon$. De plus la convergence de la suite $(Y_n)_{\mathbb{N}}$ en probabilité vers 0 conduit à l'existence de $n_0 \in \mathbb{N}$ tel que pour $n \geq n_0$, $\mathbb{P}(|Y_n| > \alpha_0) \leq \varepsilon$ et par suite $|\phi_{X_n+Y_n}(t) - \phi_{X_n}(t)| \leq 3\varepsilon$. La convergence en loi vers X de la suite $(X_n)_{\mathbb{N}}$ entraîne l'existence d'un entier n_1 , que l'on peut choisir plus grand que n_0 , tel que, pour tout $n \geq n_1$, $|\phi_{X_n}(t) - \phi_X(t)| \leq \varepsilon$. On a montré que pour tout $n \geq n_1$,

$$|\phi_{X_n+Y_n}(t) - \phi_X(t)| \leq |\phi_{X_n+Y_n}(t) - \phi_{X_n}(t)| + |\phi_{X_n}(t) - \phi_X(t)| \leq 4\varepsilon,$$

ce qui prouve la convergence en loi vers X de la suite $(X_n + Y_n)_{\mathbb{N}}$.

3. Supposons que X soit une v.a.r. de loi **symétrique**, i.e. X a même loi que $-X$ (par exemple $\mathbb{P}_X := \frac{1}{2}(\delta_{-1} + \delta_1)$ ou $\mathbb{P}_X := \mathcal{N}_1(0, 1)$), et posons, pour tout $n \in \mathbb{N}$, $X_n := -X$. Il est clair que la suite $(X_n)_{\mathbb{N}}$ converge en loi vers X et que la suite $(X_n - X)_{\mathbb{N}}$ converge en loi vers $-2X \neq 0$.

Corrigé de l'exercice 7.11, page 142

1. Un raisonnement par récurrence montre facilement que, pour tout $n \in \mathbb{N}^*$,

$$X_n = \sum_{k=1}^n \theta^{n-k} U_k.$$

La suite des v.a.r. $(\theta^{n-k} U_k)_{k=1, \dots, n}$ est indépendante, car $(U_k)_{\mathbb{N}}$ l'est. De plus on vérifie aisément que la loi de $\theta^{n-k} U_k$ est $\mathcal{N}_1(0, \theta^{2n-2k} \sigma^2)$.

X_n est la somme de v.a.r. normales indépendantes centrées, sa loi est donc la loi normale centrée de variance

$$\begin{aligned} \text{si } |\theta| \neq 1, \quad \text{Var}(X_n) &= \sum_{k=1}^n \theta^{2n-2k} \sigma^2 = \frac{1-\theta^{2n}}{1-\theta^2} \sigma^2 \\ \text{si } |\theta| = 1, \quad \text{Var}(X_n) &= n\sigma^2. \end{aligned}$$

2. Considérons la fonction caractéristique de la v.a.r. X_n . Pour tout $t \in \mathbb{R}$,

$$\begin{aligned} \text{si } |\theta| \neq 1, \quad \phi_{X_n}(t) &= \exp\left(-\frac{1-\theta^{2n}}{1-\theta^2}\sigma^2 t^2\right) \\ \text{si } |\theta| = 1, \quad \phi_{X_n}(t) &= \exp(-n\sigma^2 t^2). \end{aligned}$$

Si $|\theta| < 1$, la suite $(\phi_{X_n}(t))_{\mathbb{N}}$ converge, pour tout $t \in \mathbb{R}$, vers $\exp\left(-\frac{1}{1-\theta^2}\sigma^2 t^2\right)$ qui est la valeur en t de la fonction caractéristique de la loi $\mathcal{N}_1\left(0, \frac{1}{1-\theta^2}\sigma^2\right)$. Dans ce cas la suite $(X_k)_{\mathbb{N}}$ converge en loi vers une v.a.r. de loi $\mathcal{N}_1\left(0, \frac{1}{1-\theta^2}\sigma^2\right)$.

Si $|\theta| = 1$, la suite $(\phi_{X_n}(t))_{\mathbb{N}}$ converge, pour tout $t \neq 0$, vers 0 et pour $t = 0$, vers 1. La fonction limite n'étant pas continue en 0, elle ne peut pas être la fonction caractéristique d'une probabilité. Dans ce cas la suite $(X_k)_{\mathbb{N}}$ ne converge pas en loi.

Si $|\theta| > 1$, la suite $(\phi_{X_n}(t))_{\mathbb{N}}$ converge, pour tout $t \in \mathbb{R}$, vers 0. La fonction limite ne prenant pas la valeur 1 en 0, elle ne peut pas être la fonction caractéristique d'une probabilité. Dans ce cas la suite $(X_k)_{\mathbb{N}}$ ne converge pas en loi.

Corrigé de l'exercice 7.12, page 142

1. Supposons que la suite $(X_n)_{\mathbb{N}}$ converge en loi vers X et considérons l'application continue bornée f_ε , $\varepsilon > 0$, définie par

$$f_\varepsilon(t) := (x - \varepsilon)\mathbb{1}_{]-\infty, x-\varepsilon[}(t) + t\mathbb{1}_{[x-\varepsilon, x+\varepsilon[}(t) + (x + \varepsilon)\mathbb{1}_{[x+\varepsilon, +\infty[}(t).$$

Comme $\mathbb{E}[f_\varepsilon(X_n)] = f_\varepsilon(x_n)$, pour tout $n \in \mathbb{N}$, la suite $(f_\varepsilon(x_n))_{\mathbb{N}}$ converge vers $f_\varepsilon(x) = x$. Il existe un entier n_0 tel que, pour tout entier $n \geq n_0$, $|f_\varepsilon(x_n) - x| < \varepsilon$, par suite $f_\varepsilon(x_n) = x_n$ et $|x_n - x| < \varepsilon$. La suite $(x_n)_{\mathbb{N}}$ converge vers x .

Réciproquement si la suite $(x_n)_{\mathbb{N}}$ converge vers x alors, pour toute application f de \mathbb{R} dans \mathbb{R} continue bornée et tout $n \in \mathbb{N}$, $\mathbb{E}[f(X_n)] = f(x_n)$. Comme $\mathbb{E}[f(X)] = f(x)$, on en déduit que la suite $(\mathbb{E}[f(X_n)])_{\mathbb{N}}$ converge vers $\mathbb{E}[f(X)]$, d'où la convergence en loi de la suite $(X_n)_{\mathbb{N}}$ vers X .

2. Soit F_X la fonction de répartition de la v.a.r. X et F_{X_n} celle de X_n , $n \in \mathbb{N}$. Par hypothèse, en dehors de l'ensemble \mathcal{D} des points de discontinuité de F , la suite $(F_n)_{\mathbb{N}}$ converge simplement vers F_X . Comme, pour tous $n \in \mathbb{N}$ et $t \notin \mathcal{D}$, $F_{X_n}(t) = \mathbb{1}_{[x_n, +\infty[}(t)$, il vient $F_X(t) \in \{0, 1\}$. On peut alors utiliser le résultat de l'exercice 3.23, page 60, puisque $D := \mathcal{D}^c$ est partout dense comme complémentaire d'un ensemble dénombrable dans \mathbb{R} (cf. proposition 7.8, page 129). Donc il existe $x \in \mathbb{R}$ tel que $\mathbb{P}_X = \delta_x$ et d'après la question précédente on peut ajouter que la suite $(x_n)_{\mathbb{N}}$ converge vers x .

Corrigé de l'exercice 7.12, page 142

On rappelle que la f.c. de la probabilité de Gauss $\mathcal{N}_1(a, \sigma^2)$ est l'application définie sur \mathbb{R} par $\phi_\mu(t) = \exp\left(iat - \frac{1}{2}t^2\sigma^2\right)$.

1. (a) Montrons que la suite $(\sigma_n^2)_{n \in \mathbb{N}}$ est bornée.

Raisonnons par l'absurde et supposons que la suite de réels positifs $(\sigma_n^2)_{n \in \mathbb{N}}$ n'est pas bornée. Il existe alors une suite $(\sigma_{n_k}^2)_{k \in \mathbb{N}}$ extraite de la précédente convergeant vers $+\infty$. Comme la suite $(\mu_n)_{n \in \mathbb{N}}$ converge étroitement vers μ , d'après le critère des fonctions caractéristiques, la f.c. Φ_μ de μ est la limite simple sur \mathbb{R} de la suite de f.c. $(\Phi_{\mu_n})_{n \in \mathbb{N}}$. En considérant la suite des modules de ces f.c. on obtient, pour tout $t \in \mathbb{R}$,

$$\begin{aligned} |\Phi_\mu(t)| &= \lim_n |\exp(ia_n t)| \exp\left(-\frac{1}{2}t^2 \sigma_n^2\right) \\ &= \lim_n \exp\left(-\frac{1}{2}t^2 \sigma_n^2\right) = \lim_k \exp\left(-\frac{1}{2}t^2 \sigma_{n_k}^2\right). \end{aligned}$$

Par suite, $|\Phi_\mu(0)| = 1$ et si $t \neq 0$, $|\Phi_\mu(t)| = 0$. La f.c. Φ_μ n'est donc pas continue en 0, ce qui contredit la propriété de continuité sur \mathbb{R} des f.c. et prouve ainsi que la suite $(\sigma_n^2)_{n \in \mathbb{N}}$ est bornée.

Montrons que la suite $(\sigma_n^2)_{n \in \mathbb{N}}$ est convergente.

La suite $(\sigma_n^2)_{n \in \mathbb{N}}$ est bornée, donc il existe $M > 0$ tel que, pour tout $n \in \mathbb{N}$, $\sigma_n^2 \in [0, M]$. Comme $[0, M]$ est un compact, la suite $(\sigma_n^2)_{n \in \mathbb{N}}$ admet au moins une valeur d'adhérence σ^2 dans $[0, M]$. Si $(\sigma_n^2)_{n \in \mathbb{N}}$ admet une autre valeur d'adhérence τ^2 dans $[0, M]$, alors τ^2 (resp. σ^2) est limite de la suite $(\sigma_{m_k}^2)_{k \in \mathbb{N}}$ (resp. $(\sigma_{n_k}^2)_{k \in \mathbb{N}}$) extraite de $(\sigma_n^2)_{n \in \mathbb{N}}$. Par un raisonnement déjà effectué, il vient, pour tout $t \in \mathbb{R}$,

$$\begin{aligned} |\Phi_\mu(t)| &= \lim_k \exp\left(-\frac{1}{2}t^2 \sigma_{n_k}^2\right) = \exp\left(-\frac{1}{2}t^2 \sigma^2\right) \\ &= \lim_k \exp\left(-\frac{1}{2}t^2 \sigma_{m_k}^2\right) = \exp\left(-\frac{1}{2}t^2 \tau^2\right), \end{aligned}$$

ce qui implique que $\tau^2 = \sigma^2$. Cela montre que σ^2 est la seule valeur d'adhérence de la suite $(\sigma_n^2)_{n \in \mathbb{N}}$ dans le compact $[0, M]$, d'où

$$\liminf_n (\sigma_n^2) = \limsup_n (\sigma_n^2) = \lim_n (\sigma_n^2) = \sigma^2.$$

- (b) Considérons, pour tout $n \in \mathbb{N}$, la f.c. Φ_n de δ_{a_n} . On peut écrire, pour tout $t \in \mathbb{R}$,

$$\Phi_n(t) = \exp(ia_n t) = \Phi_{\mu_n}(t) \exp\left(\frac{1}{2}t^2 \sigma_n^2\right)$$

et par passage à la limite $\Psi(t) := \lim_n \Phi_n(t) = \Phi_\mu(t) \exp\left(\frac{1}{2}t^2 \sigma^2\right)$. L'application Ψ ainsi définie est continue en 0 et limite simple d'une suite de f.c.. D'après le théorème de continuité de Lévy, il existe une probabilité ν sur \mathbb{R} dont Ψ est la f.c. et la suite $(\delta_{a_n})_{n \in \mathbb{N}}$ converge étroitement vers ν . Par suite, si on note C l'ensemble des réels où la fonction de répartition de ν est continue, pour tout $t \in C$, $F_\nu(t) = \lim_n F_{\delta_{a_n}}(t)$.

On en déduit que, pour tout $t \in C$, $F_\nu(t) \in \{0, 1\}$, mais C est une partie partout dense de \mathbb{R} , ce qui implique que ν est la mesure de Dirac en un point $a \in \mathbb{R}$. On conclut alors que la suite $(a_n)_{n \in \mathbb{N}}$ converge dans \mathbb{R} vers le réel a .

Revenant à la fonction caractéristique de μ , on peut écrire, pour tout $t \in \mathbb{R}$,

$$\Phi_\mu(t) = \lim_n \exp\left(ia_n t - \frac{1}{2}t^2 \sigma_n^2\right) = \exp\left(iat - \frac{1}{2}t^2 \sigma^2\right),$$

ce qui, avec la convention adoptée, prouve que $\mu = \mathcal{N}_1(a, \sigma^2)$.

Annexe A

Formulaire

Ce formulaire sera fourni avec les sujets lors des épreuves terminales.

A.1 Rappels de notations

On rappelle les conventions et notations suivantes :

1. Pour tous entiers relatifs k et n , $C_n^k := \frac{n!}{k!(n-k)!}$ si $0 \leq k \leq n$, et $C_n^k := 0$ sinon.
2. L'écriture $\mu := \rho \cdot \lambda$ signifie que ρ est une densité de probabilité définie sur \mathbb{R} par $\rho(x)$ et que μ est la probabilité sur \mathbb{R} définie par cette densité ρ .
3. λ désigne la mesure de Lebesgue sur \mathbb{R} .
4. Pour tout réel $a > 0$, $\Gamma(a) := \int_0^{+\infty} e^{-x} x^{a-1} dx$. En particulier, $\Gamma(1/2) = \sqrt{\pi}$ et, pour tout entier $n \geq 1$, $\Gamma(n) = (n-1)!$.

Pour chaque probabilité μ de la liste, on trouvera :

1. son nom, sa notation, sa définition ;
2. sa fonction de répartition F définie sur \mathbb{R} par $F(x) := \mu([-\infty, x])$;
3. sa fonction caractéristique ϕ définie sur \mathbb{R} par $\phi(t) := \int_{\mathbb{R}} e^{itx} d\mu(x)$ (sauf pour la probabilité hypergéométrique) ;
4. son espérance (moment d'ordre 1) $m := \int_{\mathbb{R}} x d\mu(x)$ et sa variance (moment centré d'ordre 2) $\sigma^2 := \int_{\mathbb{R}} (x - m)^2 d\mu(x)$, si ces moments existent.

A.2 Quelques relations à connaître en probabilités

Somme d'une série géométrique et ses dérivées terme à terme

Pour tout réel $0 < x < 1$,

$$\sum_{k=0}^{+\infty} x^k = \frac{1}{1-x}$$

$$\sum_{k=1}^{+\infty} kx^{k-1} = \frac{1}{(1-x)^2}$$

$$\sum_{k=2}^{+\infty} k(k-1)x^{k-2} = \frac{2}{(1-x)^3}$$

et de façon plus générale, pour tout entier naturel $p \geq 1$,

$$\sum_{k=p}^{+\infty} k(k-1)(k-2)\cdots(k-p+1)x^{k-p} = \frac{p!}{(1-x)^{p+1}}.$$

Somme de la série exponentielle népérienne

Pour tout réel x ,

$$\sum_{k=0}^{+\infty} \frac{x^k}{k!} = e^x$$

Intégrale de Gauss

$$\int_{-\infty}^{+\infty} e^{-\frac{1}{2}t^2} dt = \sqrt{2\pi}$$

Formule du binôme de Newton

Pour tous réels a, b , et tout entier naturel n ,

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

Relation de Vandermonde

Pour tous entiers naturels n, m , et N tels que $0 \leq N \leq n+m$,

$$\sum_{k=0}^N C_n^k C_m^{N-k} = C_{n+m}^N$$

Relation de Pascal

Pour tous entiers naturels n, k tels que $0 \leq k \leq n$,

$$C_{n+1}^k = C_n^k + C_n^{k-1}$$

Deux autres relations utiles

Pour tous entiers naturels n, p tels que $0 \leq p \leq n$,

$$pC_n^p = nC_{n-1}^{p-1}$$

$$\sum_{k=n}^{k=n+p} C_k^n = C_{n+p+1}^{n+1}$$

A.3 Probabilités usuelles discrètes

1. **Dirac** : $\delta_a, a \in \mathbb{R}$

$$F(x) = \mathbb{1}_{[a, +\infty[}(x) \text{ et } \Phi(t) = e^{ita}$$

$$m = a \text{ et } \sigma^2 = 0$$

2. **Bernoulli** : $\mathcal{B}(p) := p\delta_1 + (1-p)\delta_0, p \in]0, 1[$

$$F(x) = (1-p)\mathbb{1}_{[0,1[}(x) + \mathbb{1}_{[1,+\infty[}(x) \text{ et } \Phi(t) = 1-p+pe^{it}$$

$$m = p \text{ et } \sigma^2 = p(1-p)$$

3. **Bernoulli-symétrique** : $\mathcal{B}_s(p) := (1-p)\delta_{-1} + p\delta_1, p \in]0, 1[$

$$F(x) = (1-p)\mathbb{1}_{[-1,1[}(x) + \mathbb{1}_{[1,+\infty[}(x) \text{ et } \Phi(t) = (1-p)e^{-it} + pe^{it}$$

$$m = 2p-1 \text{ et } \sigma^2 = 4p(1-p)$$

4. **Binomiale** : $\mathcal{B}(n, p) := \sum_{k=0}^n C_n^k p^k (1-p)^{n-k} \delta_k, p \in]0, 1[\text{ et } n \in \mathbb{N}^*$

$$F(x) = \sum_{k=0}^n C_n^k p^k (1-p)^{n-k} \mathbb{1}_{[k,+\infty[}(x) \text{ et } \Phi(t) = (1-p+pe^{it})^n$$

$$m = np \text{ et } \sigma^2 = np(1-p)$$

5. **Binomiale-négative** : $\mathcal{I}(r, p) := \sum_{k=0}^{r-1} C_{k+r-1}^{r-1} p^r (1-p)^k \delta_k, p \in]0, 1[\text{ et } r \in \mathbb{N}^*$

$$F(x) = \sum_{k=0}^{+\infty} C_{k+r-1}^{r-1} p^r (1-p)^k \mathbb{1}_{[k,+\infty[}(x) \text{ et } \Phi(t) = \left(\frac{p}{1-(1-p)e^{it}} \right)^r$$

$$m = \frac{r(1-p)}{p} \text{ et } \sigma^2 = \frac{r(1-p)}{p^2}$$

6. **Géométrique** : $\mathcal{G}(p) := \sum_{k=1}^{+\infty} p(1-p)^{k-1} \delta_k, p \in]0, 1[$

$$F(x) = \sum_{k=1}^{+\infty} p(1-p)^{k-1} \mathbb{1}_{[k,+\infty[}(x) \text{ et } \Phi(t) = \frac{pe^{it}}{1-(1-p)e^{it}}$$

$$m = \frac{1}{p} \text{ et } \sigma^2 = \frac{1-p}{p^2}$$

7. **Hypergéométrique** : $\mathcal{H}(n_1, n_2, n) := \sum_{k=0}^n \frac{C_{n_1}^k C_{n_2}^{n-k}}{C_{n_1+n_2}^n} \delta_k, n \in \mathbb{N}^*, n_1 \in \mathbb{N}^* \text{ et } n_2 \in \mathbb{N}^*$

avec $n \leq n_1 + n_2$

$$F(x) = \sum_{k=0}^n \frac{C_{n_1}^k C_{n_2}^{n-k}}{C_{n_1+n_2}^n} \mathbb{1}_{[k,+\infty[}(x)$$

$$m = \frac{nn_1}{n_1 + n_2} \text{ et } \sigma^2 = n \frac{n_1 n_2 (n_1 + n_2 - n)}{(n_1 + n_2)^2 (n_1 + n_2 - 1)}$$

$$8. \text{ Poisson : } \mathcal{P}(\alpha) := \sum_{k=0}^{+\infty} e^{-\alpha} \frac{\alpha^k}{k!} \delta_k, \alpha > 0$$

$$F(x) = \sum_{k=0}^{+\infty} e^{-\alpha} \frac{\alpha^k}{k!} \mathbb{1}_{[k, +\infty[}(x) \text{ et } \Phi(t) = \exp(\alpha(e^{it} - 1))$$

$$m = \alpha \text{ et } \sigma^2 = \alpha$$

$$9. \text{ Uniforme-discrète : } \mathcal{U}(n) := \frac{1}{n} \sum_{k=1}^{k=n} \delta_k, n \in \mathbb{N}^*$$

$$F(x) = \frac{1}{n} \sum_{k=1}^{k=n} \mathbb{1}_{[k, +\infty[}(x) \text{ et } \Phi(t) = \frac{1}{n} \sum_{k=1}^{k=n} e^{itk}$$

$$m = \frac{n+1}{2} \text{ et } \sigma^2 = \frac{n^2 - 1}{12}$$

A.4 Probabilités usuelles à densité

$$1. \text{ Uniforme-continue : } \mathcal{U}([a, b]) := \rho \cdot \lambda, a, b \in \mathbb{R} \text{ avec } a < b \text{ et}$$

$$\rho(x) := \frac{1}{b-a} \mathbb{1}_{[a, b]}(x)$$

$$F(x) = \frac{x-a}{b-a} \mathbb{1}_{[a, b]}(x) + \mathbb{1}_{[b, +\infty[}(x) \text{ et } \Phi(t) = \frac{e^{itb} - e^{ita}}{it(b-a)}$$

$$m = \frac{a+b}{2} \text{ et } \sigma^2 = \frac{(b-a)^2}{12}$$

$$2. \text{ Gamma : } \gamma(a, \theta) := \rho \cdot \lambda, a > 0 \text{ et } \theta > 0 \text{ avec } \rho(x) := \frac{\theta^a x^{a-1}}{\Gamma(a)} e^{-\theta x} \mathbb{1}_{[0, +\infty[}(x)$$

$$F(x) = \int_{-\infty}^x \frac{\theta^a}{\Gamma(a)} e^{-\theta t} t^{a-1} \mathbb{1}_{[0, +\infty[}(t) dt \text{ et } \Phi(t) = \left(\frac{\theta}{\theta - it} \right)^a$$

$$m = \frac{a}{\theta} \text{ et } \sigma^2 = \frac{a}{\theta^2}$$

$$3. \text{ Exponentielle : } \mathcal{E}(\alpha) := \gamma(1, \alpha) = \rho \cdot \lambda, \alpha > 0 \text{ avec } \rho(x) := \alpha e^{-\alpha x} \mathbb{1}_{[0, +\infty[}(x)$$

$$F(x) = (1 - e^{-\alpha x}) \mathbb{1}_{[0, +\infty[}(x) \text{ et } \Phi(t) = \frac{\alpha}{\alpha - it}$$

$$m = \frac{1}{\alpha} \text{ et } \sigma^2 = \frac{1}{\alpha^2}$$

$$4. \text{ Khi-deux : } \chi^2(n) := \gamma\left(\frac{n}{2}, \frac{1}{2}\right) = \rho \cdot \lambda, n \in \mathbb{N}^* \text{ avec } \rho(x) := \frac{x^{\frac{n}{2}-1}}{\Gamma\left(\frac{n}{2}\right) 2^{\frac{n}{2}}} e^{-\frac{x}{2}} \mathbb{1}_{[0, +\infty[}(x)$$

$$F(x) = \int_{-\infty}^x \rho(t) dt \text{ et } \Phi(t) = \left(\frac{1}{1 - 2it} \right)^{\frac{n}{2}}$$

$$m = n \text{ et } \sigma^2 = 2n$$

5. **Cauchy** : $\mathcal{C}(a) := \rho \cdot \lambda$, $a > 0$ avec $\rho(x) := \frac{a}{\pi(a^2 + x^2)}$

$$F(x) = \frac{1}{\pi} \left[\frac{\pi}{2} + \arctan \left(\frac{x}{a} \right) \right] \text{ et } \Phi(t) = e^{-a|t|}$$

Les moments m et σ^2 n'existent pas

6. **Normale ou Gauss-Laplace** : $\mathcal{N}_1(a, b) := \rho \cdot \lambda$, $a \in \mathbb{R}$ et $b > 0$ avec $\rho(x) := \frac{1}{\sqrt{2b\pi}} e^{-\frac{(x-a)^2}{2b}}$

$$F(x) = \frac{1}{\sqrt{2b\pi}} \int_{-\infty}^x e^{-\frac{(u-a)^2}{2b}} du \text{ et } \Phi(t) = \exp \left(iat - \frac{bt^2}{2} \right)$$

$$m = a \text{ et } \sigma^2 = b$$

7. **Normale d -dimensionnelle** : $\mathcal{N}_d(m, D) := \rho \cdot \lambda^d$, où $d \in \mathbb{N}^*$, $m \in \mathbb{R}^d$, D matrice carrée d'ordre d à coefficients réels, symétrique, inversible, de type positif, et $\rho(x) := \frac{1}{\sqrt{(2\pi)^d \det(D)}} \exp \left(-\frac{1}{2}(x - m)^* D^{-1}(x - m) \right)$, où $x \in \mathbb{R}^d$;

$$\Phi_X(u) = \exp \left(iu^* m - \frac{1}{2}u^* D u \right), \text{ où } u \in \mathbb{R}^d \text{ et l'opération } * \text{ désigne la transposition ;}$$

m est le vecteur-espérance et D la matrice de dispersion de $\mathcal{N}_d(m, D)$

Annexe B

Table de la loi normale standard

Dans ce cours nous n'utiliserons que la table de la fonction de répartition de la loi normale standard appelée aussi **loi normale centrée-réduite** ou **loi de Gauss-Laplace standard**. Cette table, qui est reproduite à la fin de l'annexe, sera fournie, sans les explications d'utilisation, avec les sujets lors des épreuves terminales.

B.1 Calculs avec des v.a.r. normales centrées-réduites

Voici quelques exemples d'utilisation de la table de la loi normale standard.

Tout calcul numérique de probabilité avec une variable aléatoire X normale standard se ramène à déterminer la valeur d'expressions de la forme $\mathbb{P}(a < X < b)$ ou $\mathbb{P}(X < b)$ ou $\mathbb{P}(a < X)$. Les inégalités pouvant être strictes ou larges, cela ne change rien aux calculs car la fonction de répartition de la loi normale standard est continue sur \mathbb{R} .

La table de la loi normale standard reproduite donne les valeurs, connaissant le réel t positif, des expressions $\mathbb{P}(X < t)$. On peut toujours se ramener à ces cas moyennant les relations

1. Si t est un réel positif, $\mathbb{P}(X < t)$ est donné par la table.
2. Si t est un réel positif, $\mathbb{P}(X > t) = 1 - \mathbb{P}(X < t)$.
3. Si t est un réel strictement négatif, $\mathbb{P}(X < t) = 1 - \mathbb{P}(X < -t)$.
4. Si t est un réel strictement négatif, $\mathbb{P}(X > t) = \mathbb{P}(X < -t)$.

Pour lire dans la table la valeur de $\mathbb{P}(X < t)$ pour t positif, par exemple pour $t = 2,37$, on procède de la façon suivante. On remarque que $2,37 = 2,3 + 0,07$. La valeur de $\mathbb{P}(X < 2,37)$ est lue à l'intersection de la ligne horizontale 2,3 (valeur lue dans la première colonne de la table) et de la colonne verticale 0,07 (valeur lue dans la première ligne de la table). On trouve $\mathbb{P}(X < 2,37) = 0,9911$.

On peut remarquer que la table ne donne des valeurs de $\mathbb{P}(X < t)$ que pour $0 < t < 3$. Cela est dû au fait que pour les valeurs supérieures à 3, $\mathbb{P}(X < t) \approx 1$ et par suite $\mathbb{P}(X > t) \approx 0$. Toutefois la table donne les valeurs de $\mathbb{P}(X < t)$ pour t prenant des valeurs entre 3 et 4,5 avec cinq décimales (tables des grandes valeurs pour t située au bas de la page).

B.2 Calculs avec des v.a.r. normales de paramètres quelconques

Pour ce qui est du calcul de probabilité dans le cas de v.a.r. normales quelconques, on rappelle la proposition suivante, qui est une réécriture avec le langage des v.a.r. de la proposition 2.9, page 26 :

Proposition B.1.

Procédé de standardisation

Une v.a.r. X est normale d'espérance m et de variance $\sigma^2 > 0$ si, et seulement, si la v.a.r.

$Z := \frac{X - m}{\sigma}$ est une v.a.r. normale centrée-réduite.

Comme, pour tout réel a et b avec $a < b$,

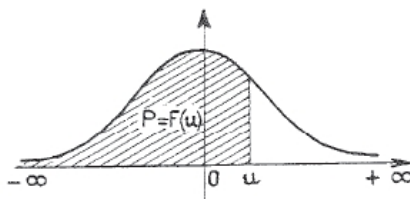
$$\{a < X < b\} = \left\{ \frac{a - m}{\sigma} < Z < \frac{b - m}{\sigma} \right\},$$

on a

$$\mathbb{P}(a < X < b) = \mathbb{P}\left(\frac{a - m}{\sigma} < Z < \frac{b - m}{\sigma}\right).$$

Ainsi tout événement faisant intervenir dans sa formulation une v.a.r. X normale d'espérance m et de variance $\sigma^2 > 0$ peut donc être exprimé avec la v.a.r. $Z := \frac{X - m}{\sigma}$ de loi normale centrée-réduite. Le procédé de standardisation permet de ramener tout calcul de probabilité relatif à une loi normale quelconque à un calcul de probabilité relatif à la loi normale centrée-réduite, et donc à l'utilisation uniquement de la table statistique de la loi normale centrée-réduite.

FONCTION DE RÉPARTITION DE LA LOI NORMALE CENTRÉE-RÉDUITE OU STANDARD

(Pour tout $u > 0$, la table donne la probabilité que la v.a. prenne une valeur inférieure à $u > 0$)

u	0	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5438	0,5478	0,5517	0,5557	0,5596	0,5636	0,5675	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,7580	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,8023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8212	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,8770	0,8790	0,8810	0,8830
1,2	0,8849	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9265	0,9279	0,9292	0,9306	0,9319
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441
1,6	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545
1,7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633
1,8	0,9641	0,9649	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9761	0,9767
2,0	0,9772	0,9778	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817
2,1	0,9821	0,9826	0,9830	0,9834	0,9838	0,9842	0,9846	0,9850	0,9854	0,9857
2,2	0,9861	0,9864	0,9868	0,9871	0,9875	0,9878	0,9881	0,9884	0,9887	0,9890
2,3	0,9893	0,9896	0,9898	0,9901	0,9904	0,9906	0,9909	0,9911	0,9913	0,9916
2,4	0,9918	0,9920	0,9922	0,9925	0,9927	0,9929	0,9931	0,9932	0,9934	0,9936
2,5	0,9938	0,9940	0,9941	0,9943	0,9945	0,9946	0,9948	0,9949	0,9951	0,9952
2,6	0,9953	0,9955	0,9956	0,9957	0,9959	0,9960	0,9961	0,9962	0,9963	0,9964
2,7	0,9965	0,9966	0,9967	0,9968	0,9969	0,9970	0,9971	0,9972	0,9973	0,9974
2,8	0,9974	0,9975	0,9976	0,9977	0,9977	0,9978	0,9979	0,9979	0,9980	0,9981
2,9	0,9981	0,9982	0,9982	0,9983	0,9984	0,9984	0,9985	0,9985	0,9986	0,9986

TABLE POUR LES GRANDES VALEURS DE u

u	3	3,1	3,2	3,3	3,4	3,5	4	4,5
F(u)	0,99865	0,99903	0,99931	0,99952	0,99966	0,99977	0,99997	0,999997

Annexe C

Devoirs à envoyer à la correction

Les trois devoirs ci-dessous sont à renvoyer pour leur correction, au plus tard à la date indiquée, à l'adresse suivante :

Bruno Saussereau,
Laboratoire de Mathématiques de Besançon,
UFR Sciences et Techniques,
16, route de Gray,
25030 Besançon cedex, FRANCE

Le but premier d'un devoir est de montrer au correcteur que vous avez compris le cours, que vous connaissez les résultats vus en cours et les hypothèses qui les commandent, et que vous savez les mobiliser pour répondre à une question ou démontrer un résultat nouveau. Il est donc recommander de tout mettre en oeuvre pour atteindre cet objectif.

En particulier :

Un devoir de mathématiques est un devoir de français qui traite de mathématiques, c'est donc avant tout un texte de français. Il doit donc être rédigée de façon correcte en français. Les hypothèses spécifiques justifiant l'utilisation de chaque théorème doivent être correctement explicitées et le résultat du cours utilisé doit être clairement identifié voire explicitement énoncé. Les résultats intermédiaires et les conclusions obtenues doivent être mis en évidence. Les notations utilisées ou introduites, surtout si elles sont nouvelles par rapport au cours, doivent être clairement annoncées. La rédaction du cours peut être considérée comme un guide de rédaction d'un texte mathématique.

C.1 Devoir 1 à renvoyer le 21 février 2014 au plus tard

La rédaction et la présentation de la copie, la justification des affirmations par référence aux résultats du cours seront des éléments d'appréciation essentiels dans la notation.

Exercice I

Soit X une variable aléatoire normale centrée réduite. Préciser, dans chacun des cas ci-dessous, la loi de probabilité de la variable aléatoire Y définie en fonction de X

1. $Y = X^3$.
2. $Y = F(X)$ où F est la fonction de répartition de la variable X .

Exercice II

Soit X une v.a.r. normale de loi $\mathcal{N}(m, \sigma^2)$, où m et σ sont des réels avec $\sigma > 0$.

1. Montrer que la fonction caractéristique de X peut s'exprimer à l'aide de la fonction caractéristique Φ de la loi de Gauss-Laplace standard $\mathcal{N}(0, 1)$.
2. En utilisant le théorème de dérivation sous le signe \int , montrer que Φ est une solution particulière de l'équation différentielle du premier ordre $y'(t) + ty(t) = 0$. En déduire l'expression analytique de la fonction Φ , puis celle de la fonction caractéristique de la variable X .

Exercice III

Soit $(X_k)_{k \in \mathbb{N}}$ une suite indépendante de v.a.r. de Bernoulli toutes de même paramètre $0 < p < 1$.

1. Soit un entier $r \geq 1$, on définit deux nouvelles v.a.r., en posant pour tout $\omega \in \Omega$,

$$\tau_r(\omega) := \inf\{n \in \mathbb{N}^* / X_1(\omega) + X_2(\omega) + \dots + X_n(\omega) = r\}$$

et

$$\theta_r(\omega) := \inf\{n \in \mathbb{N}^* / X_1(\omega) + X_2(\omega) + \dots + X_{n+r}(\omega) = r\}$$

avec la convention $\inf \emptyset := +\infty$.

1. Montrer, pour tout $x \in]0, 1[$, la relation

$$\sum_{k=r-1}^{+\infty} C_k^{r-1} x^{k-r+1} = \frac{1}{(1-x)^r}.$$

2. Montrer que la variable aléatoire réelle τ_r est une variable aléatoire réelle discrète de loi (dite loi de Pascal de paramètres r et p)

$$\mathcal{P}(r, p) := \sum_{k=r}^{+\infty} C_{k-1}^{r-1} p^r (1-p)^{k-r} \delta_k.$$

Vérifier que $\mathbb{P}(\tau_r = +\infty) = 0$.

3. Montrer que la variable aléatoire réelle θ_r est une variable aléatoire réelle discrète de loi (dite **loi binomiale-négative de paramètres r et p**)

$$\mathcal{I}(r, p) := \sum_{k=0}^{+\infty} C_{k+r-1}^{r-1} p^r (1-p)^k \delta_k.$$

Vérifier que $\mathbb{P}(\theta_r = +\infty) = 0$.

4. Donner une interprétation des variables aléatoires réelles τ_r et θ_r en terme de jeu de Pile-ou-Face.
5. Montrer qu'un des deux modèles précédents permet de formaliser le problème dit **des boîtes d'allumettes de Stephan Banach** :

Un fumeur a dans chacune de ses deux poches une boîte contenant au départ N allumettes. Chaque fois qu'il désire fumer une cigarette, il choisit une poche au hasard. Quelle est la probabilité que, le fumeur se rendant compte pour la première fois qu'une boîte est vide, l'autre boîte contienne k allumettes où k est un entier naturel inférieur ou égal à N ?

Exercice IV

Le but de cet exercice est de montrer qu'il n'existe pas de probabilité \mathbb{P} sur l'espace $(\mathbb{N}^*, \mathcal{P}(\mathbb{N}^*))$ telle que, pour tout $n \geq 1$, $\mathbb{P}(n\mathbb{N}^*) = \frac{1}{n}$ où $n\mathbb{N}^* = \{nk, k \in \mathbb{N}^*\}$.

Supposons qu'une telle probabilité existe. Soit $(p_k)_{\mathbb{N}}$ la suite des nombres entiers premiers rangés en ordre croissant.

1. Par un raisonnement simple montrer que $\mathbb{P}(\limsup_k (p_k \mathbb{N}^*)) = 0$.
2. Montrer que la suite $(p_k \mathbb{N}^*)_{\mathbb{N}}$ est indépendante. En déduire, en utilisant le fait que la série $\sum_k \frac{1}{p_k} = +\infty$, une autre valeur de $\mathbb{P}(\limsup_k (p_k \mathbb{N}^*))$. Conclure que la probabilité \mathbb{P} n'existe pas.

C.2 Devoir 2 à renvoyer le 28 mars 2014 au plus tard

La rédaction et la présentation de la copie, la justification des affirmations par référence aux résultats du cours seront des éléments d'appréciation essentiels dans la notation.

Exercice I

Soit (X, Y) un couple de variables aléatoires réelles de loi $\mathbb{P}_{(X,Y)} = \left(\alpha e^{-\frac{1}{2}(x^2 - xy + y^2)} \right) \cdot \lambda^{(2)}$ où $\lambda^{(2)}$ est la mesure de Lebesgue sur \mathbb{R}^2 . Déterminer la constante α et la matrice de dispersion du couple (X, Y) . Préciser les lois respectives des variables aléatoires réelles X et Y . Le couple de variables aléatoires réelles (X, Y) est-il indépendant ?

Exercice II

Théorème de Fisher-Cochran

Soit $n \in \mathbb{N}^*$ et (X_1, \dots, X_n) une suite indépendante de v.a.r. toutes de même loi $\mathcal{N}_1(0, 1)$. On définit respectivement les v.a.r. **moyenne empirique** et **variance empirique** par

$$\bar{X} := \frac{X_1 + \dots + X_n}{n} \quad \text{et} \quad S^2 := \frac{1}{n-1} \sum_{k=1}^n (X_k - \bar{X})^2.$$

1. Montrer que la v.a.r. X_1^2 suit la loi $\gamma(\frac{1}{2}, \frac{1}{2})$ aussi appelée loi du **Khi-deux à 1 degré de liberté** et notée $\chi^2(1)$.
2. En utilisant la fonction caractéristique des lois Gamma, en déduire que la loi de la v.a.r. $\sum_{k=1}^n X_k^2$ est $\gamma(\frac{1}{2}, \frac{n}{2})$ aussi appelée loi du **Khi-deux à n degrés de liberté** notée $\chi^2(n)$.
3. Montrer qu'il existe une matrice orthogonale C de la forme

$$C = \begin{pmatrix} c_{1,1} & c_{1,2} & \cdots & c_{1,n} \\ c_{2,1} & c_{2,2} & \cdots & c_{2,n} \\ \vdots & \vdots & \ddots & \vdots \\ c_{n-1,1} & c_{n-1,2} & \cdots & c_{n-1,n} \\ \frac{1}{\sqrt{n}} & \frac{1}{\sqrt{n}} & \cdots & \frac{1}{\sqrt{n}} \end{pmatrix}.$$

4. Déterminer la loi du vecteur aléatoire $Y := CX$.
5. Calculer Y_n et $\sum_{k=1}^n Y_k^2$ à l'aide de X_1, \dots, X_n . En déduire que $\bar{X} = \frac{1}{\sqrt{n}} Y_n$ et $S^2 = \frac{1}{n-1} \sum_{k=1}^{n-1} Y_k^2$.
6. Démontrer le **théorème de Fisher-Cochran** : Soit (X_1, \dots, X_n) une suite indépendante de v.a.r. de même loi $\mathcal{N}_1(0, 1)$. Alors (\bar{X}, S^2) est indépendant, \bar{X} suit la loi $\mathcal{N}_1(0, \frac{1}{n})$ et $(n-1)S^2$ suit la loi $\chi^2(n-1)$.

Exercice III

Soit $(\varepsilon_i)_{i \geq 1}$ une suite indépendante de v.a.r. de même loi $\mathcal{N}_1(0, 1)$ et X_0 une v.a.r. indépendante de la suite $(\varepsilon_i)_{i \geq 1}$ et de loi $\mathbb{P}_{X_0} = \mathcal{N}_1(m, \sigma^2)$. On définit la suite de v.a.r. $(X_n)_{n \geq 1}$ de la façon suivante : $X_n := l_n(X_0, \dots, X_{n-1}) + b_n \varepsilon_n$ où $(b_n)_{n \geq 1}$ est une suite de réels et $(l_n)_{n \geq 1}$ une suite de formes linéaires sur \mathbb{R}^n . Montrer que, pour tout $n \geq 1$, il existe une forme linéaire L_n sur \mathbb{R}^{n+1} telle que $X_n = L_n(X_0, \varepsilon_1, \dots, \varepsilon_n)$ et en déduire que le vecteur (X_0, \dots, X_n) est gaussien.

C.3 Devoir 3 à renvoyer le 18 avril 2014 au plus tard

La rédaction et la présentation de la copie, la justification des affirmations par référence aux résultats du cours seront des éléments d'appréciation essentiels dans la notation.

Exercice I

Théorème de Weierstrass

Soient f une application continue de $[0, 1]$ dans \mathbb{R} et $x \in [0, 1]$. Pour tout $n \in \mathbb{N}^*$, notons S_n une v.a.r. binomiale de loi $\mathcal{B}(n, x)$.

1. Montrer que $p_n(x) := \mathbb{E}[f(\frac{1}{n}S_n)]$ est un polynôme en x appelé **polynôme de Bernstein de f** .
2. En utilisant l'uniforme continuité de f sur $[0, 1]$ montrer que, pour tout $\varepsilon > 0$, il existe $\delta > 0$ tel que, pour tout $n \in \mathbb{N}^*$ et tout $x \in [0, 1]$,

$$\begin{aligned} |p_n(x) - f(x)| &\leq \mathbb{E}[|f(\frac{1}{n}S_n) - f(x)|] \\ &\leq \varepsilon \mathbb{P}\left(|\frac{1}{n}S_n - x| < \delta\right) + 2\mathbb{P}\left(|\frac{1}{n}S_n - x| \geq \delta\right) \sup_{0 \leq x \leq 1} |f(x)|. \end{aligned}$$

En déduire que, pour tout $\varepsilon > 0$, il existe $\delta > 0$ tel que, pour tout $n \in \mathbb{N}^*$ et tout $x \in [0, 1]$,

$$|p_n(x) - f(x)| \leq \varepsilon + 2 \frac{x(1-x)}{n\delta^2} \sup_{0 \leq x \leq 1} |f(x)|.$$

3. Démontrer le **théorème de Weierstrass** : *Toute application continue de $[0, 1]$ dans \mathbb{R} est limite uniforme sur $[0, 1]$ d'une suite de polynômes.*

Exercice II

Soit $(X_n)_{n \geq 1}$ une suite indépendante de v.a.r. de même loi de Cauchy $\mathcal{C}(1)$ (Pour la définition, cf. formulaire de l'annexe A, page 205). Pour tout $n \geq 1$, on pose $S_n := \sum_{k=1}^n X_k$. Étudier les con-

vergences en probabilité et en loi des suites de v.a.r. $\left(\frac{1}{\sqrt{n}}S_n\right)_{n \geq 1}$, $\left(\frac{1}{n}S_n\right)_{n \geq 1}$ et $\left(\frac{1}{n^2}S_n\right)_{n \geq 1}$.

Exercice III

Soit $(U_k)_{k \in \mathbb{N}}$ une suite indépendante de v.a.r. de loi normale centrée et de variance $\sigma^2 > 0$. Pour tout $\theta \in \mathbb{R}$, on définit la suite $(X_k)_{k \in \mathbb{N}}$ par la relation de récurrence $X_n = \theta X_{n-1} + U_n$, pour tout $n \geq 1$, avec $X_0 = 0$.

1. Déterminer, pour tout $n \in \mathbb{N}$, la loi de la v.a.r. X_n .
2. Étudier la convergence en loi de la suite de v.a.r. $(X_k)_{k \in \mathbb{N}}$.

Bibliographie.

- [1] Lelong-Ferrand J. - Arnaudière J.M., *Cours de mathématiques*, Dunod, 1977.
- [2] Ansel J.P.- Duce Y., *Exercices corrigés en théorie de la mesure et de l'intégration*, Ellipses, 1995.
- [3] Ansel J.P.- Duce Y., *Exercices corrigés en théorie des probabilités*, Ellipses, 1996.
- [4] Bouleau N., *Probabilités de l'ingénieur : variables aléatoires et simulation*, Hermann, 1986.
- [5] Brémaud P., *Introduction aux probabilités : modélisation des phénomènes aléatoires*, Springer-Verlag, 1988.
- [6] Commission inter-IREM "Statistique et Probabilités" (coordination M. Henry), *Autour de la modélisation en probabilités*, Presses universitaires de Franche-Comté, collection "Didactiques", Besançon, 2001
- [7] Duce Y., *Introduction à la théorie mathématique des probabilités*, Ellipses, 1998.
- [8] Gramain A., *Intégration*, Hermann, coll. Méthodes, 1994.
- [9] Guinot M., *Le paradoxe de Banach-Tarski*, Aléas, 1991.
- [10] Hennequin P.L., *Pourquoi des tribus ?*, Bulletin APMEP n° 303, pp 183-195.
- [11] Leboeuf C.- Roque J.L.- Guegan J., *Cours de probabilités et de statistiques*, Ellipses, 2ème édition 1983.
- [12] Leboeuf C.- Roque J.L.- Guegan J., *Exercices corrigés de probabilités*, Ellipses, 1987.
- [13] Revuz D., *Mesure et intégration*, Hermann, coll. Méthodes, 1994.
- [14] Stoyanov J., *Counterexamples in probability*, John Wiley and Sons, 1989.