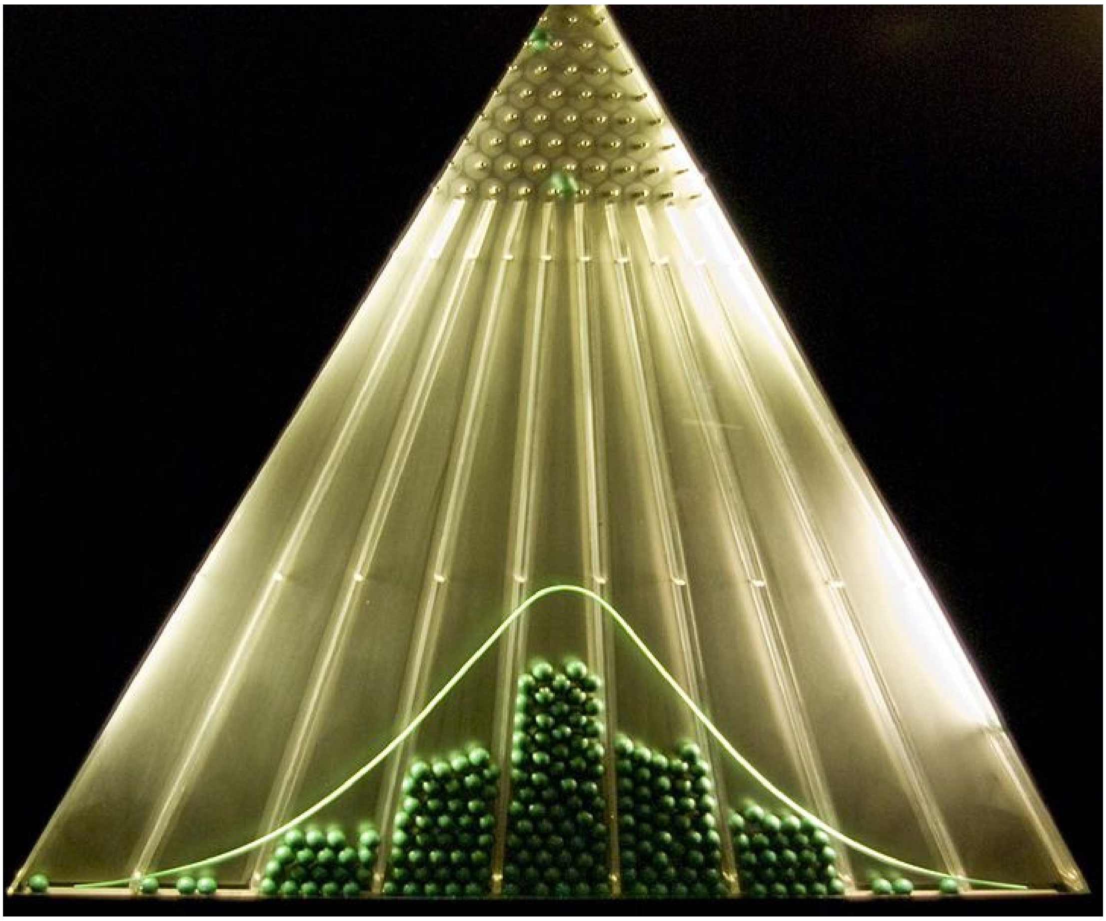
Licence de mathématiques

Calcul des probabilités



Bruno SAUSSEREAU[1]

2012-2013

[1]bruno.saussereau@univ-fcomte.fr

Table des matières.

Présentation du cours

Ce cours correspond à un enseignement dans le cadre du semestre 5 de la licence, suivant la terminologie introduite par la réforme dite "L.M.D.", des cursus à l'université mise en place à la rentrée universitaire 2004.

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

Objectifs du cours

Cet enseignement par correspondance s'adresse en priorité aux étudiants désireux de présenter le concours du CAPES de mathématiques ou à ceux qui se destinent à des études de mathématiques appliquées en vue de devenir ingénieurs-mathématiciens. Il n'est pas conçu pour les étudiants désireux de poursuivre leurs études en vue d'effectuer de la recherche en mathématiques ou de présenter le concours de l'agrégation externe de mathématiques (notamment l'épreuve optionnelle de modélisation probabiliste). Pour ces étudiants d'autres enseignements sur le même sujet sont prévus : "Théorie de la mesure et de l'intégration" (semestre 5) et "Théorie des probabilités" (semestre 6). En effet, dans notre approche, l'usage de la théorie de la mesure sera limité au strict minimum et beaucoup de résultat concernant cette théorie seront simplement admis dans notre cours, alors que la connaissance de leur démonstration est indispensable à tout futur chercheur en mathématiques ou candidat au concours de l'agrégation externe de mathématiques.

C'est dire que le cours qui suit se veut surtout un cours essentiellement tourné à la fois vers la pratique pour l'ingénieur qui utilise les résultats des probabilités et vers la réflexion sur la modélisation pour le professeur de lycée qui, bien qu'enseignant les probabilités à un niveau élémentaire, doit avoir un recul suffisant sur les principes de base de la modélisation probabiliste. Ce cours trouve son prolongement naturel dans un enseignement de la statistique inférentielle au semestre 6.

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 [3] et [4].Certains des exercices proposés dans ce cours sont inspirés de ces deux ouvrages moyennant quelques adaptations de vocabulaire dues au formalisme introduit dans le cours.

Ce cours utilisera bien sûr des connaissances de mathématiques classiques, bien que relativement élémentaires, vues dans les semestres précédents. C'est l'occasion, dès maintenant, de réviser ces notions mathématiques indispensables qui seront supposées connues. Il s'agit 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. le calcul intégral : intégrale au sens de Riemann, intégrale généralisée (ou impropre), changement de variable dans une intégrale, méthode d'intégration par parties,
3. le calcul des sommes de séries : série géométrique, série exponentielle, dérivation des séries entières, ...

Méthode de travail

Le cours proprement dit comprendra des définitions, des propositions, des démonstrations, des exemples et des exercices corrigés.

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.

Les exemples du cours servent à illustrer une définition sur un cas particulier ou montrer une application concrète d'une proposition. Leur rédaction est 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.

l'annexe B explique l'usage de la table de la loi normale centrée réduite (voir page 117).

L'annexe A est un rappel des principales relations mathématiques utilisées dans les calculs de probabilités et des lois de probabilités classiques à connaître. Ce formulaire vous sera fourni en totalité ou en partie lors des épreuves de contrôles ou examens à.

Un fascicule contenant le formulaire de l'annexe A ainsi que la table de la loi normale $\mathcal{N}(0,1)$ qui ce situe à la page 117 accompagne ce cours. Ce document vous sera fourni lors des épreuves s'il est nécessaire.

Remarques sur la bibliographie

Concernant la bibliographie, le cours prendra appui sur l'ouvrage [2] paru aux éditions Ellipses.

Pour une justification du choix du formalisme et de sa signification en tant que modèle de la "réalité", on pourra consulter avec profit le chapitre I de [1] et [2].

Calendrier de travail

Le cours comprend 6 chapitres :

1- Modèles probabilistes
2- Loi d'un vecteur aléatoire
3- Intégrale de Lebesgue
4- Espérance et moments d'un vecteur aléatoire
5- Indépendance de vecteurs aléatoires
6- Convergences et théorèmes-limite.

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

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

Ce calendrier est donné à titre indicatif si vous souhaitez vous organiser sur un semestre. Bien entendu j'accepte de corriger vos copies et répond à vos questions à tout moment de l'année.
Le devoir I portera sur les chapitres 1, 2 et 3. Il peut être envoyé pour le 01/11/2012.
Le devoir II portera principalement sur le chapitre 4 et le paragraphe 1 du chapitre 5 mais pourra bien sûr faire appel à des résultats des chapitres précédents. Il paut être envoyé vers le 01/12/2012.
Le devoir III portera principalement sur les chapitres 4, 5 et 6 mais pourra bien sûr faire appel à des résultats des chapitres précédents. Il doit être envoyé au plus tard pour le 01/01/2013.

Quelques conseils concernant les devoirs.

Un devoir de mathématiques est un devoir de français qui traite de mathématiques, c'est donc avant tout un devoir de français. Il doit donc être rédigé 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 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 exemple de rédaction d'un texte mathématique.

Il est évident que l'appréciation de la copie 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.

Mise en ligne.

Le cours sera mis en ligne sur la plate forme Moodle de l'université de Franche-Comté (Environnement Numérique de Travail).
Les devoirs y seront aussi présent ainsi que les corrections (après la date prévue de retour des copies).
Un forum sera proposé sur Moodle. Ce sera l'occasion de poser vos questions, de faire part de vos remarques et commentaires, de signaler les fautes de frappes et autres imprécisions et aussi de proposer des améliorations du manuscripts. Tout ceci se fera publiquement et pourra ainsi être profitable à tous les utilisateurs de cette ressource.
Vous pouvez aussi poser vos questions par mail à l'adresse qui figure sur la première page à savoir :

bruno.saussereau@univ-fcomte.fr

Notations

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 ^{c}, inclusion (au sens large) $\subseteq$.

Le symbole de Halmos, $\square$, signifie 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.

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

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 1.1. Image-réciproque
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.

L'ensemble image-réciproque peut aussi être noté par $f^{-1}(A)$. Il faut cependant prendre garde car cela ne signifie pas que l'application f est bejective.

Exemple 1.
Si f et g sont deux applications de E dans $\mathbb{R}$ et a un réel,
$\{f = g\} := \{x \in E \ ; \ f(x) = g(x)\}$, $\{f \leq g\} := \{x \in E \ ; \ f(x) \leq g(x)\}$, $\{f = a\} := \{x \in E \ ; \ f(x) = a\}$.

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 $k \in I$, $x \in A_k$. On peut aussi écrire de manière ensembliste

$$\bigcup_{i \in I} A_i = \{x \in F \ ; \ \exists \ i \in I \ , \ x \in A_i\} \ .$$

De même, $\bigcap_{i\in I} A_i$ désigne la partie de F constituée des éléments x de F tels que, pour tout indice $k \in I$, $x \in A_k$. On écrit aussi

$$\bigcap_{i\in I} A_i = \{x \in F \ ; \forall \ i \in I \ , \ x \in A_i\} \ .$$

Voici quelques propriétés classiques de l'image-réciproque :

Proposition 1.1. Propriétés de l'image réciproque
Avec les notations introduites ci-dessus,

1. $\{f \in \emptyset\} := \{x \in E \ / \ f(x) \in \emptyset\} = \emptyset$.
2. *Si* A *et* B *sont des parties de* F *avec* $A \subseteq B$ *alors,* $\{f \in A\} \subseteq \{f \in B\}$.
3. *Si* $(A_i)_{i\in I}$ *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. $\{f \in A\}^c = E \setminus \{f \in A\} = \{f \in F \setminus A\} = \{f \in A^c\}$.

La preuve de cette proposition est l'objet de l'exercice 1, page 11.
On fera attention à l'ambiguité 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 $\{f \in \complement_F A\}$.

Définition 1.2. Fonction indicatrice
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*

$$\mathbb{1}_A(x) := \begin{cases} 0, & \text{si } x \notin A \ ; \\ 1, & \text{si } x \in A. \end{cases}$$

Pour s'entraîner à manipuler ces premières notions, il est conseillé de faire les exercices 1 et 2, page 12.

1.2 Tribu sur un ensemble

Définition 1.3. Tribu, espace mesurable, parties mesurables
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* $(A_n)_{\mathbb{N}}$ *est une suite d'éléments de* $\mathcal{A}$, *alors*
$$\bigcup_{n\in\mathbb{N}} A_n \in \mathcal{A}.$$

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.

Exemple 2.
*$\{\emptyset, E\}$ 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.*

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

Proposition 1.2. Procédé de construction de parties mesurables
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 $i \in \mathbb{N}\backslash I$. 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. □

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". 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.

Définition 1.4. Tribu borélienne
La tribu borélienne (ou tribu de Borel de $\mathbb{R}$) notée $\mathcal{B}(\mathbb{R})$ 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, on définit la tribu borélienne (ou tribu de Borel de $\mathbb{R}^d$,) notée $\mathcal{B}(\mathbb{R}^d)$, comme étant 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** $\overline{\mathbb{R}}$, notée $\mathcal{B}(\overline{\mathbb{R}})$, comme étant 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 $a \in \overline{\mathbb{R}}$.

Définition 1.5. Borélien
Les éléments des tribus $\mathcal{B}(\overline{\mathbb{R}})$, resp. $\mathcal{B}(\mathbb{R}^d)$, sont appelés boréliens de $\overline{\mathbb{R}}$, resp. $\mathbb{R}^d$.

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 aux seuls ensembles qui seront manipulés dans la suite :

Proposition 1.3. Exemples de boréliens

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. 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

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

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.

Dans ce cours nous passerons rapidement sur la démonstration de l'appartenance à la tribu de Borel des ensembles manipulés. Ce type de démonstration ne sera pas exigible dans les épreuves d'évaluation.

Il est conseillé de réfléchir aux exercices 3,...,7.

1.3 Mesures

Définition 1.6. Mesure, espace mesuré
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 $(E,\, \mathcal{A}, \mu)$ s'appelle un espace mesuré.

On notera qu'il ne suffit pas que le deuxième axiome de la définition précédente soit vrai pour les suites finies deux à deux disjointes (simple-additivité) pour qu'il le soit pour les suites infinies deux à deux disjointes. Un contre-exemple est proposé dans l'exercice 8, page 12. Bien entendu, le deuxième axiome de la définition des mesures énoncé pour les suites infinies deux à deux disjointes est aussi vrai pour les suites finies deux à deux disjointes.

Définition 1.7. Mesure de Lebesgue sur $\mathbb{R}^d$
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 $]a_1, b_1]\times]a_2, b_2]\times\cdots\times]a_d, b_d]$ où pour tout entier $1 \le i \le d$, les réels a_i et b_i vérifient $a_i < b_i$,

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

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 dans l'exercice 9, page 13, (faisable dès maintenant) que $\lambda^{(d)}\left(\mathbb{R}^d\right) = +\infty$. 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.

1.4 Probabilités

Définition 1.8. Probabilité, espace probabilisé
Une probabilité sur $(E,\, \mathcal{A})$ est une mesure μ sur $(E,\, \mathcal{A})$ telle que $\mu(E) = 1$. Le triplet $(E,\, \mathcal{A}, \mu)$ s'appelle alors un espace probabilisé. On parle aussi parfois d'espace de probabilité.

Exemple 3.

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 $\left(\mathbb{R}^d,\, \mathcal{B}(\mathbb{R}^d)\right)$ appelée probabilité de Dirac au point a sur $\mathbb{R}^d$.*

2. *La mesure de Lebesgue n'est pas une probabilité sur $\mathbb{R}^d$ d'après le résultat de l'exercice 9, page 13.*

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 de la propriété suivante 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 un cours d'analyse de deuxième année de licence. □

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 $(E, \mathcal{A})$.

Exemple 4.

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 l'ensemble $\{1, 2, \cdots, n\}$) :*
$$\mathcal{U}(n) := \frac{1}{n}\sum_{k=1}^{n} \delta_k.$$

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

Définition 1.9. Probabilité discrète
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 $p_n > 0$.

Exemple 5.
Les probabilités binomiale $\mathcal{B}(n, p)$, de Poisson $\mathcal{P}(\alpha)$, de Dirac δ_a sont discrètes et portées respectivement par les ensembles $\{0, 1, \cdots, n\}$, $\mathbb{N}$, $\{a\}$.
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, 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 la construction précédente.

Définition 1.10. Densité de probabilité sur $\mathbb{R}$
Une densité de probabilité sur $\mathbb{R}$ est une 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.

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 $\int_{-\infty}^{+\infty} \rho(t)dt < +\infty$.
La définition générale d'une densité sur $\mathbb{R}^d$ avec $d \geq 1$ sera donnée au chapitre 5.
Exemple 6.

1. *La probabilité de Gauss-Laplace 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.

L'existence des mesures à densité résulte d'un théorème de prolongement assez technique que nous n'énoncerons pas. En revanche l'unicité dans le cas des probabilités résulte d'un lemme qu'on admettra, dont il est utile de connaître l'énoncé.

Proposition 1.5. Lemme d'unicité de deux probabilités, fonction de répartition

1. *Soient μ et ν deux probabilités sur $\mathbb{R}$. Si pour tout $x \in \mathbb{R}$*
$$\mu(]-\infty, x]) = \nu(]-\infty, x]),$$
alors $\mu = \nu$.
2. *L'application F_μ de $\mathbb{R}$ dans $[0, 1]$, définie, pour tout $x \in \mathbb{R}$, par $F_\mu(x) := \mu(]-\infty, x])$. est la* **fonction de répartition** *de la probabilité μ, en abrégé* **f.r.** .
3. *Deux probabilités sur $\mathbb{R}$ sont identiques si, et seulement si, elles ont la même fonction de répartition.*

Exemple 7.

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 tabulées de la fonction de répartition de la probabilité $\mathcal{N}_1(0, 1)$ se trouvent dans la table de la loi normale centrée réduite de l'annexe B.

Proposition 1.6. Propriétés des probabilités
Soit $(E, \mathcal{A}, \mu)$ désigne un espace de probabilité.

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 $A \in \mathcal{A}$, $\mu(A^c) = 1 - \mu(A)$.*
5. **Inégalité de Bonferroni** *: Pour toute suite $(A_n)_{n\in\mathbb{N}}$ d'éléments de $\mathcal{A}$,*

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

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 $A \subseteq E$.
5) Pour démontrer l'inégalité de Bonferroni nous aurons besoin du résultat ensembliste de l'exercice 17, page 14, dont l'énoncé est le suivant : soit $(A_n)_{n\in\mathbb{N}}$ une suite de parties d'un ensemble E. Posons $B_0 := A_0$ et,pour tout entier $k \geq 1$, $B_k := A_k \setminus (A_0 \cup A_1 \cup \cdots \cup A_{k-1})$. Alors pour tout entier $n \geq 0$, $B_n \subseteq A_n$, la suite $(B_n)_{n\in\mathbb{N}}$ est formée de parties deux à deux disjointes vérifiant

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

On applique la σ-additivité des mesures à la suite $(B_n)_{n\in\mathbb{N}}$ et on utilise le point 1 précédemment démontré. □

Proposition 1.7. Théorème de continuité monotone

1. *Pour toute suite $(A_n)_{n\in\mathbb{N}}$ d'éléments de $\mathcal{A}$, croissante au sens de l'inclusion, $(\mu(A_n))_{n\in\mathbb{N}}$ est une suite réelle croissante convergeant vers $\mu(\bigcup_n A_n)$ c-à-d*

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

2. *Pour toute suite $(A_n)_{n\in\mathbb{N}}$ d'éléments de $\mathcal{A}$, décroissante au sens de l'inclusion, la suite $(\mu(A_n))_{n\in\mathbb{N}}$ est une suite réelle décroissante convergeant vers $\mu(\bigcap_n A_n)$ c-à-d*

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

Démonstration :
1) Soit $(A_n)_{n\in\mathbb{N}}$ suite croissante d'éléments de $\mathcal{A}$. Utilisons la suite construite dans l'exercice I-17. Comme la suite $(A_n)_{n\in\mathbb{N}}$ est croissante, pour tout entier $n \geq 1$, $B_n = A_n \setminus A_{n-1}$ et $B_0 = A_0$. $(B_n)_{n\in\mathbb{N}}$ 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 $(A_n^c)_{n\in\mathbb{N}}$ 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. □

Remarques : Parmi les propriétés énoncées ci-dessus pour les probabilités, certaines tombent en défaut pour les mesures qui ne sont pas des probabilités. La partie 1 de la proposition 1.6 est vraie pour les mesures qui ne sont pas des probabilités. En revanche la partie 2 utilise la propriété que $\mu(A) = 1-\mu(A^c)$, c'est-à-dire que $\mu(A)+\mu(A^c) = 1 = \mu(E)$. Ceci est encore vrai pour les mesures μ telles que $\mu(E) < \infty$ (au lieu de $\mu(E) = 1$) mais elle tombe en défaut si $\mu(E) = +\infty$ comme par exemple pour la mesure de Lebesgue sur $\mathbb{R}$.

Proposition 1.8. Propriétés des fonctions de répartition
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 $\mathbb{R}$,*

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

.

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

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

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,

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

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.
□
On admettra le résultat réciproque de la partie 1 de la proposition 1.8 qui prouve qu'il y a bijection entre l'ensemble des probabilités sur $\mathbb{R}$ et l'ensemble des f.r..

Proposition 1.9. Probabilités sur $\mathbb{R}$ et f.r.
Si F est une application croissante de $\mathbb{R}$ dans $[0,1]$, continue-à-droite sur $\mathbb{R}$ avec

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

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

1.5 Exercices

Exercice 1 Solution p. 77

Démontrer la proposition 1.1.

Exercice 2 Solution p. 77

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

1. Ecrire $\mathbb{1}_{A\cap B}$ 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.* $A \cap B = \varnothing$).
   (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 3 Solution p. 78

Soient n un entier strictement positif et $(A_1, A_2, \cdots, A_n)$ 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 $(A_1, A_2, \cdots, A_n)$, 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 $\{1, 2, \cdots, n\}$. Montrer que la famille $\mathcal{A}$ est une tribu sur E.

Exercice 4 Solution p. 78

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

Exercice 5 Solution p. 79

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

Exercice 6 Solution p. 79

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 $a < b$.

Exercice 7 Solution p. 79

Soient n un entier strictement positif et $(A_1, A_2, \cdots, A_n)$ une partition de E. Montrer que la tribu $\mathcal{A}$ construite dans l'exercice 3, page 12, est la plus petite des tribus sur E contenant la famille $(A_1, A_2, \cdots, A_n)$.

Exercice 8 Solution p. 79

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 $\mu(\varnothing) = 0$. Montrer que

1. μ est **simplement-additive,** *i.e.* pour toute suite finie $A_1, \cdots, A_n$ 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.

Exercice 9 Solution p. 80

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.

Exercice 10 Solution p. 80

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

Exercice 11 Solution p. 80

Vérifier que les probabilités introduites dans l'exemple 4 sont bien des probabilités construites suivant le procédé de la proposition 1.4.

Exercice 12 Solution p. 81

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 $\mathcal{B}(7, \frac{3}{10})(\{0,3,5\})$.

Exercice 13 Solution p. 81

1. Vérifier que les applications définies dans les points 2 et 3 de l'exemple 6 sont bien des densités de probabilités.
2. Dessiner l'allure des courbes représentatives de ces densités.
3. Calculer $\mathcal{U}([0,1])(]-\infty, \frac{4}{3}])$, $\mathcal{E}(2)(]-\infty, 7])$.

Exercice 14 Solution p. 82

Vérifier les exemples de f.r. donnés dans l'exemple 7 et dessiner l'allure des courbes représentatives de ces f.r..

Exercice 15 Solution p. 83

Exprimer les fonctions de répartition des probabilités $\mathcal{E}(\alpha)$ et $\mathcal{U}([a,b])$ et dessiner l'allure de leurs courbes représentatives.

Exercice 16 Solution p. 83

Soit F l'application de $\mathbb{R}$ dans $\mathbb{R}$ définie, pour tout réel x, par

$$F(x) := \begin{cases} 1 - \frac{1}{2}e^{-x}, & \text{si } x \geqslant 0\,; \\ \frac{1}{2}e^{x}, & \text{si } x \leqslant 0. \end{cases}$$

Montrer que F est la fonction de répartition d'une probabilité à densité qu'on déterminera.

Exercice 17 Solution p. 83

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

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

Exercice 18 Solution p. 84

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 19 Solution p. 84

1. Montrer que si μ est une probabilité admettant une densité sur $\mathbb{R}$, alors pour tout réel a, $\mu(\{a\}) = 0$.
2. Avec les notations de la proposition 1.8, montrer que pour tous réels a, b vérifiant $a < b$,
$$\mu(]a,b[) = F(b-) - F(a) \quad \text{et} \quad \mu([a,b[) = F(b-) - F(a-).$$

Exercice 20 Solution p. 84

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

Exercice 21 Solution p. 84

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 f.r. 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 paragraphe est de fournir quelques éléments de réflexion sur la modélisation mathématique de phénomènes aléatoires.
Considérons les deux exemples suivants :

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 $(\Omega, \mathcal{P}(\Omega), \mathbb{P})$ est 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 $\mathbb{R}$.

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 considèrera 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 $\mathcal{C}([0,1],\mathbb{R}^+)$.

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 $\mathcal{F}$ sont alors appelés **événements** de Ω.

2.2 Applications mesurables

Définition 2.1. Application mesurable, application borélienne
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.*
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 un premier exemple d'applications mesurables :

Proposition 2.1. Mesurabilité des fonctions indicatrices
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, $A \in \mathcal{A}$. □
La proposition suivante donne des classes importantes de fonctions boréliennes qui correspondent à la plupart des cas qu'on considèrera par la suite.

Proposition 2.2. Exemples de fonctions boréliennes (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.

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. Dans la suite de ce cours toutes les applications utilisées seront mesurables et on ne s'attardera pas à prouver la mesurablité des applications considérées.
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 2.2. Partie positive, partie négative d'une fonction
Si f est une application d'un ensemble E dans $\overline{\mathbb{R}}$. Les fonctions $f^+ := \sup(f, 0)$ et $f^- := \sup(-f, 0)$ sont appelées respectivement la **partie positive** *et la* **partie négative** *de f.*

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

Exemple 8.
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 :

Proposition 2.3. Opérations classiques et mesurabilité (admis)

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 est une application $\mathcal{A}$-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 2.3. Fonction étagée
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 $\alpha_1, \alpha_2, \cdots, \alpha_n$ 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

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

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.

Proposition 2.4. Fonctions mesurables et fonctions étagées
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.

Démonstration : Pour $n \geqslant 1$ et $i \in \{0, 1, ..., n2^n - 1\}$ on pose $A_n = \{x \in E \ / \ f(x) \geqslant n\}$ et $B_{n,i} = \{x \in E \ / \ i2^{-n} f(x) \leqslant (i+1)2^{-n}\}$. soit ensuite la suite de fonctions étagées positives $(f_n)_{n\geqslant 1}$ définie par

$$f_n = \sum_{i=0}^{n2^n-1} \frac{i}{2^n} 1\!\!1_{B_{n,i}} \ + \ n 1\!\!1_{A_n} \ .$$

On vérifie aisement que pour tout $x \in E$, la suite $(f_n(x))_{n\geqslant 1}$ et croissante et converge vers $f(x)$. □
Ce résultat 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 $1\!\!1_A$, 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 $\sum_{i=1}^{n} \alpha_i 1\!\!1_{A_i}$ 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

Vecteurs 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 $(\Omega, \mathcal{F})$.

Définition 2.4. Vecteur aléatoire, variable aléatoire réelle
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, ...
La proposition suivante est l'énoncé avec un vocabulaire différent du résultat de l'exercice 22, page 22.

Proposition 2.5. Composition d'un vecteur aléatoire par une application borélienne
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. □
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 v.a.r. X.

Loi d'un vecteur aléatoire

Proposition 2.6. Loi de probabilité
Soit X *un vecteur aléatoire de dimension* d*. L'application*

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

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 $\mathbb{P}_X$. □

Définition 2.5. Loi d'un vecteur aléatoire
La probabilité $\mathbb{P}_X$ est appelée la **loi de probabilité relativement à** $\mathbb{P}$ *du vecteur aléatoire X ou plus simplement* **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.*

Remarque. *On admettra le résultat théorique suivant. Si μ est une probabilité sur $\mathbb{R}$, alors il existe un espace de probabilité de base $(\Omega, \mathcal{F}, \mathbb{P})$ et une v.a.r. X sur cet espace telle que $\mathbb{P}_X = \mu$.*

Exemple 9.
Soit X une v.a.r. de loi $\mathcal{N}_1(0,1)$ (on notera qu'une telle affirmation a un sens d'après la remarque précédente). Déterminons la loi de la v.a.r. $Y := X^2$.
Pour cela il suffit d'identifier la fonction de répartition F_Y de la v.a.r. 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. On la note $\gamma(\frac{1}{2}, \frac{1}{2})$.* □

Définition 2.6. Loi gaussienne
Soient $m \in \mathbb{R}$ et $\sigma > 0$. Appelons **loi de Gauss-Laplace de paramètres** m **et** σ^2, *et notons $\mathcal{N}_1(m, \sigma^2)$, la probabilité sur $\mathbb{R}$ admettant pour densité la fonction ρ définie sur $\mathbb{R}$ 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. Il est démontré dans l'exercice 25, page 22.

Proposition 2.7. Procédé de standardisation d'une loi gaussienne
Avec les notations précédentes, une v.a.r. X suit la loi $\mathcal{N}_1(m, \sigma^2)$ si, et seulement si, la v.a.r. $Z := \frac{X-m}{\sigma}$ suit la loi $\mathcal{N}_1(0,1)$.

Définition 2.7. Variable aléatoire discrète, v.a. à densité
Une v.a. X à valeurs dans $\mathbb{R}^d$ est dite **discrète** *si sa loi est discrète.*
Une v.a.r. est dite à densité sur $\mathbb{R}$ si sa loi est à densité sur $\mathbb{R}$.

Exemple 10.

1. *Les v.a.r. 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 de ce cours.
Les v.a.r. discrètes sont les v.a.r. à valeurs "presque-sûrement" dans un ensemble dénombrable. De façon précise :

Proposition 2.8. Caractérisation des v.a. discrètes (admis)
Un vecteur aléatoire X de dimension d est discret si, et seulement si, il existe une partie $D := \{a_k, k \in K \subseteq \mathbb{N}\}$ de $\mathbb{R}^d$ telle que $\mathbb{P}(X \in D) = 1$. Dans ce cas la loi du vecteur aléatoire X s'écrit

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

On dit aussi 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 de $\mathbb{R}^d$.
Ce résultat ramène alors la détermination de la loi d'une v.a.r. 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 v.a.r. 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 v.a.r. à densité. En effet, pour une v.a.r. 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 v.a.r. X admettant un densité, ce qui ne présente plus d'intérêt.
La proposition précédente sera notamment appliquée (en prenant $D = \mathbb{N}$ ou $D = \mathbb{Z}$) dans le cas où les v.a.r. sont **entières** *i.e.* prennent leurs valeurs dans $\mathbb{N}$ ou $\mathbb{Z}$.
Remarque. *Toute v.a.r. X à valeurs dans $\mathbb{N}$, resp. $\mathbb{Z}$, est discrète. Sa loi s'écrit alors*

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

Travail de culture général conseillé mais pas urgent : Étudier dans [3], pages 147 à 171, l'interprétation probabiliste à l'aide de tirages dans une urne, des v.a.r. de lois de Bernoulli, binomiale, géométrique, de Pascal, binomiale-négative, hypergéométrique.

2.4 Exercices

Exercice 22 Solution p. 86

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 $\mathbb{R}^d$.

Exercice 23 Solution p. 86

Soit X une v.a.r. 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 v.a.r. $Y := e^X$ 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** .

Exercice 24 Solution p. 86

On considère une v.a.r. X dont la fonction de répartition est donnée par

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

Déterminer la loi de la v.a.r. $Y = |X|$.

Exercice 25 Solution p. 87

Démontrer la proposition 2.7.

Exercice 26 Solution p. 87

Montrer que si X est une v.a.r. de loi

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

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

Chapitre 3

Intégrale de Lebesgue

3.1 Intégration des applications mesurables positives

On se donne $(E, \mathcal{A}, \mu)$ un espace mesuré quelconque.
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 $\mathbb{E}_\mu$, 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 $\varphi = \sum_{k=1}^n \alpha_k \mathbb{1}_{A_k}$ où $\alpha_1 \geq 0, \alpha_2 \geq 0, \cdots, \alpha_n \geq 0$ sont les valeurs deux à deux distinctes de φ et, pour tout entier k vérifiant $1 \leq k \leq n$, $A_k := \{x \in E \;/\; f(x) = \alpha_k\}$. 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 $f \in \mathcal{M}^+(E, \mathcal{A})$,

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

On remarquera encore que, par sa définition, $\mathbb{E}_\mu(f)$ est un nombre positif éventuellement infini c-à-d $\mathbb{E}_\mu(f) \in [0, +\infty]$.

Exemple 11.
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 $a_0 < a_1 < a_2 < \cdots < a_n$ est un suite strictement croissante de $n+1$ réels, les réels $\alpha_1, \alpha_2, \cdots, \alpha_n$ n'étant pas nécessairement deux à deux distincts. On dit que φ est une **fonction en escalier** *sur $\mathbb{R}$. Alors*

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

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

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 $\mathbb{E}_\mu$ par trois propriétés fondamentales.

Proposition 3.1. Théorème d'existence de l'intégrale d'une fonction mesurable positive (admis)

1. *Si* μ *est une mesure sur* $(E, \mathcal{A})$*, il existe une application notée* $\mathbb{E}_\mu$*, 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* $\mathcal{M}^+(E, \mathcal{A})$,
   $$\lim_{n\to+\infty} \mathbb{E}_\mu(f_n) = \mathbb{E}_\mu\left(\lim_{n\to+\infty} f_n\right).$$
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 $\mathbb{E}_\mu(f)$:

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

$\mathbb{E}_\mu(f)$ 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'expliciter directement le nombre $\mathbb{E}_\mu(f)$ si ce n'est dans des cas simples. Par exemple :

Exemple 12.

1) Si f *est l'application* $f = \sum_{i=1}^n a_i \mathbb{1}_{A_i}$*, où* $(A_1, A_2, \cdots, A_n)$ *est une famille finie d'éléments de* $\mathcal{A}$ *et* $(a_1, a_2, \cdots, a_n)$ *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* $\int_\mathbb{R} f d\mu = \frac{4}{3}\pi + 7$.

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 5.

Proposition 3.2. Intégration des fonctions boréliennes positives par rapport à la mesure de Lebesgue (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

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

On aura par exemple

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

Proposition 3.3. Intégration par rapport à la mesure de Dirac
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

$$\mathbb{E}_\mu(f) = f(a).$$

La preuve du résultat ci-dessus est un des sujets du devoir I.
La proposition qui suit généralise la précédente :

Proposition 3.4. Intégration par rapport à une mesure discrète
On suppose $E := \mathbb{R}^d$, $\mathcal{A} := \mathcal{B}(\mathbb{R}^d)$, $\mu := \sum_{i=0}^{+\infty} \alpha_i\delta_{a_i}$ où $(a_k)_{k\in\mathbb{N}}$ est une suite de vecteurs de $\mathbb{R}^d$ et $(\alpha_k)_{k\in\mathbb{N}}$ une suite de réels positifs ou nuls.
Si f est une application borélienne de $\mathbb{R}^d$ dans $[0, +\infty]$, alors

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

Exemple 13.
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)

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

Exemple 14.
Soit $(u_n)_{n\in\mathbb{N}}$ 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. Intégration par rapport à une mesure 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

$$\mathbb{E}_\mu(f) = \mathbb{E}_\lambda(f\rho).$$

Exemple 15.
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

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

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

3.2 Intégration des fonctions numériques

On se donne $(E, \mathcal{A}, \mu)$ un espace mesuré quelconque.
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. Application intégrable et intégrale d'une fonction numérique
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 $\mathbb{E}_\mu(f) \in \mathbb{R}$.*
De plus f est intégrable si, et seulement si, $\mathbb{E}_\mu(|f|)$ est fini.

On démontre l'équivalence entre l'intégrabilité de f et la finitude de $\mathbb{E}_\mu(|f|)$. tout d'abord $|f| = f^+ + f^-$ est une application $\mathcal{A}$-mesurable positive et, d'après 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.□

Exemple 16.

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,* $f(a) \in \mathbb{R}$.

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 3.6. Intégration des fonctions réelles par rapport à la mesure de Lebesgue sur $\mathbb{R}$ (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*

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

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*

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

3.3 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 $f_1, f_2, \cdots, f_d$, où, pour tout entier $k \leq d$, f_k est une application $\mathcal{A}$-mesurable de E dans $\mathbb{R}$. On écrira $f := (f_1, f_2, \cdots, f_d)$.

Définition 3.2. Fonction vectorielle intégrable, intégrale
On dit que f est **intégrable sur** E **suivant** μ *si toutes les applications-composantes $f_1, f_2, \cdots, f_d$ 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 $\mathbb{E}_\mu(f_1), \mathbb{E}_\mu(f_2), \cdots, \mathbb{E}_\mu(f_d)$, et on note $\mathbb{E}_\mu(f) := (\mathbb{E}_\mu(f_1), \mathbb{E}_\mu(f_2), \cdots, \mathbb{E}_\mu(f_d))$.*
On remarque que $f := (f_1, f_2, \cdots, f_d)$ 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,

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

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 $\mathbb{R}^n$. On aboutit à la définition suivante.

Définition 3.3. Fonction caractéristique
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.*

Exemple 17.
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.

L'intérêt de cette définition résulte dans le théorème admis suivant permettant d'identifier les probabilités sur $\mathbb{R}^n$ comme on le fait, mais seulement sur $\mathbb{R}$, avec les fonctions de répartition.

Proposition 3.7. Théorème d'indentification de deux probabilités sur $\mathbb{R}^n$ (admis)
Deux probabilités sur $\mathbb{R}^n$ sont identiques si, et seulement si, elles ont la même fonction caractéristique.

On trouvera la liste des fonctions caractéristiques des probabilités usuelles sur $\mathbb{R}$ dans le formulaire donné dans l'annexe A de ce cours.

3.4 Propriétés de l'intégrale et théorèmes de convergence

L'intégrale d'une fonction suivant une mesure μ possède toutes les propriétés des intégrales classiques vues en première et deuxième année de Licence.

Proposition 3.8. Propriétés de l'intégrale de Lebesgue (admis)
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* $\mathbb{E}_\mu(f) \leq \mathbb{E}_\mu(g)$.

Les énoncés de théorèmes permettant d'intervertir les symboles $\mathbb{E}_\mu$ et $\lim_{n\to+\infty}$ sont particulièrement simples dans cette théorie de l'intégration. Ce sont ces résultats que je nomme théorèmes de convergence.
Commençons par rappeler :

Proposition 3.9. Théorème de convergence monotone de Beppo-Lévi (admis)
Pour toute suite croissante $(f_n)_{n\in\mathbb{N}}$ d'applications $\mathcal{A}$-mesurables positives,

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

Ce résultat a pour corollaire :

Proposition 3.10. Théorème d'interversion de $\mathbb{E}_\mu$ et $\sum_0^\infty$
Pour toute suite $(f_n)_{n\in\mathbb{N}}$ 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 $(g_n)_{n\in\mathbb{N}}$. □

Notons que ces deux 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. 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 deuxième année de licence.

Proposition 3.11. Théorème de convergence dominée de Lebesgue (admis)
Si $(f_n)_{n\in\mathbb{N}}$ est une suite d'applications $\mathcal{A}$-mesurables vérifiant :

i) la suite $(f_n)_{n\in\mathbb{N}}$ converge simplement vers une application $\mathcal{A}$-mesurable $f = \lim_n f_n$

ii) il existe une application intégrable φ telle que,

$$\text{pour tout } k \in \mathbb{N} \ , \ |f_k| \leq \varphi,$$

alors $f = \lim_n f_n$ est intégrable et

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

Deux exercices d'application de ce théorème sont proposées dans les exercices 34 et 35, page 31. Ces deux exercices sont importants car ils montrent la facilité avce laquelle on peut intervertir limite et intégrale sans se soucier d'une quelconque convergence uniforme comme c'était le cas en deuxième année de licence.

À ce stade de la lecture, vous etes en mesure de faire le devoir I (après avoir fait les exercices proposés ci-après)

3.5 Exercices

Exercice 27 Solution p. 87

Vérifier les affirmations de l'exemple 12.

Exercice 28 Solution p. 87

1. Soient μ et ν deux mesures sur $(E, \mathcal{A})$. A l'aide de la proposition 3.1 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 ?

Exercice 29 Solution p. 88

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 $(\alpha_k)_{k\in\mathbb{N}}$ est une suite de réels positifs telle que la série $\sum_{k=0}^{+\infty} \alpha_k = 1$. 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.

Exercice 30 Solution p. 89

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

Exercice 31 Solution p. 89

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

Exercice 32 Solution p. 89

Soit μ une mesure sur $(E, \mathcal{A})$ et $f \in \mathcal{M}^+(E, \mathcal{A})$. Démontrer que

$$\mu(\{x \in E\ ;\ f(x) = 0\}) = 1 \qquad \Longrightarrow \qquad \int_E f(x) d\mu(x) = 0\ .$$

Exercice 33 Solution p. 89

Soient f et g deux applications de E dans $\mathbb{R}$ intégrables suivant μ. Démontrer que

1. $|\mathbb{E}_\mu(f)| \leq \mathbb{E}_\mu(|f|)$ où $|\cdot|$ désigne la valeur absolue.
2. Si $d = 1$ et $f \leq g$, alors $\mathbb{E}_\mu(f) \leq \mathbb{E}_\mu(g)$.

Exercice 34 Solution p. 90

Calculer $\lim_{n\to\infty} \int_0^n \left(1 + \frac{x}{n}\right)^n e^{-2x} dx$.

Exercice 35 Solution p. 90

Soit f une fonction $\lambda-$intégrable sur l'intervalle $[a, b]$, calculer

$$\lim_{n\to+\infty} \int_{[a,b]} \ln\left(e + \frac{|x|}{n}\right) f(x) d\lambda(x)\ .$$

Chapitre 4

Espérance et moments d'un vecteur aléatoire

4.1 Moments d'une variable aléatoire

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

Définition 4.1. Espérance mathématique d'une variable aléatoire
L'opérateur d'intégration par rapport à la mesure de probabilité $\mathbb{P}$ sera noté $\mathbb{E}$ 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** *. L'espérance mathématique est parfois appelée* ***moyenne de*** X.

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 $\int_\Omega Xd\mathbb{P}$ 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 $h(X) := h \circ X$.

Définition 4.2. Variance, écart-type et moments d'une variable aléatoire
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. $m'_p := \mathbb{E}\left[(X - m_1)^p\right]$. Le moment centré d'ordre 2 s'appelle aussi* **la variance de** X *et se note Var(X). Sa racine carrée positive s'appelle* **l'écart-type de** X *et se note σ_X.*

Définition 4.3. Covariance de deux variables aléatoires
Si X et Y sont deux v.a.r. on appelle **covariance de** X **et** Y, *le réel (s'il est défini) $COV(X,Y) := \mathbb{E}([X - \mathbb{E}(X)][Y - \mathbb{E}(Y)])$.*

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

Proposition 4.1. Existence des moments d'une v.a.r.

1. *Soit X une v.a.r. telle qu'il existe un entier naturel non nul p vérifiant $\mathbb{E}(|X|^p) < +\infty$. Alors, pour tout entier k vérifiant $1 \leq k \leq p$, les moments d'ordre k, $m_k := \mathbb{E}(X^k)$ existent dans $\mathbb{R}$. Il en est de même pour les moments centrés et $m'_k := \mathbb{E}\left[(X - m_1)^k\right]$.*
2. *Si X et Y sont deux v.a.r. vérifiant $\mathbb{E}(X^2) < +\infty$ et $\mathbb{E}(Y^2) < +\infty$, alors la covariance de X et Y, 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 v.a.r. 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 $\mathbb{E}\left[(X - m_1)^k\right]$ 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 v.a.r. XY est donc intégrable ainsi que la v.a.r. $Z := [X - \mathbb{E}(X)][Y - \mathbb{E}(Y)]$, ce qui donne bien un sens à la covariance de X et de Y. □

Par commodité on pose la définition suivante :

Définition 4.4. Variable aléatoire de carré intégrable
Une v.a.r. X est dite **de carré intégrable** *si* $\mathbb{E}(X^2) < +\infty$.

Proposition 4.2. Formules de König-Huygens
Soient X et Y deux v.a.r. de carré intégrable. Alors

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

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 4.3. Espérance de v.a.r. discrète et à densité

1. *Si X est une v.a.r. intégrable discrète de loi $\mathbb{P}_X := \sum_{k=0}^{+\infty} \mathbb{P}(X = a_k)\delta_{a_k}$, alors*

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

2. *Si X est une v.a.r. 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 de ce cours.

La proposition suivante donne un procédé de calcul des moments d'une v.a.r. à l'aide de sa fonction caractéristique :

Proposition 4.4. Moments et fonction caractéristique (admis)
Soient X une v.a.r. 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 Φ_X admet un développement limité en 0 à l'ordre n s'écrivant*

$$\Phi_X(u) = 1 + \sum_{k=1}^{k=n} a_k u^k + o(u^n)$$

alors, pour tout $1 \leq k \leq 2p$, $\mathbb{E}(X^k) = (-i)^k a_k k!$.

Définition 4.5. Espérance d'un vecteur aléatoire
Soit X un vecteur aléatoire de dimension d de composantes $X_1, X_2, ..., X_d$ intégrables suivant $\mathbb{P}$. On appelle **espérance de X suivant $\mathbb{P}$,** *et on note $\mathbb{E}(X)$, le vecteur de $\mathbb{R}^d$,*

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

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 4.6. Variable centrée, variable réduite
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.**

Définition 4.7. Matrice de dispersion, matrice des covariances
Soit X un vecteur aléatoire de dimension d de composantes $X_1, X_2, ..., X_d$ de carré intégrable sur Ω. On appelle **matrice de dispersion** *de X l'espérance de la matrice carrée aléatoire d'ordre d, $[X - \mathbb{E}(X)][X - \mathbb{E}(X)]^t$, où ^{t} désigne l'opération de transposition des matrices. On dit aussi* **matrice des covariances de** *X et on la note D_X.*

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

Proposition 4.5. Propriétés des matrices de dispersion
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 $Cov(X_i, X_j)$ des v.a.r. X_i et X_j. Les éléments diagonaux de D_X sont les variances des composantes de X.*
2. $D_{[X-\mathbb{E}(X)]} = D_X$, $\mathbb{E}(MX) = M\mathbb{E}(X)$, *et* $D_{MX} = MD_X M^t$.
3. *D_X est une matrice symétrique. D_X est* **de type positif** *i.e., pour tout $u \in \mathbb{R}^d$, $u^t D_X u \geq 0$. En particulier, D_X est une matrice diagonalisable sur $\mathbb{R}$ dont les valeurs propres sont des réels positifs ou nuls.*

Exemple 18.
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,

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

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

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

□

4.2 Théorème du transfert

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 4.6. 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

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

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

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

Démonstration (idée) : Cette proposition se démontre à l'aide de la technique des fonctions étagées.

Proposition 4.7. 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

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

Exemple 19.
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 $\mathbb{R}^d$,

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

Exemple 20.
Soit X une v.a.r. 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.

i) 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,*

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

ii) Deuxième méthode : On a vu au chapitre 2 que la v.a.r. $Y := X^2$ *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.

Par application du théorème de transfert il vient aisément :

Proposition 4.8. Expressions tranférée des moyennes, variance et moments
Sous les conditions d'existence des différents moments,

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

4.3 Critères d'indentification des lois

Le théorème du transfert permet d'établir un troisième critère d'identification des lois utilisant les fonctions boréliennes positives. Les deux premiers déjà vus utilisent les fonctions de répartition (proposition 1.5) et les fonctions caractéristiques (proposition 3.7).

Proposition 4.9. Critère des fonctions positives (pour identifier deux 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, 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 avec la notation des opérateurs d'intégration

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

Démonstration :

i) Condition nécessaire : Si $\mathbb{P}_X = \mu$, d'après le théorème du transfert, pour toute application borélienne positive h de $\mathbb{R}^d$ dans $[0, +\infty]$,

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

ii) Condition suffisante : 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 $B \in \mathcal{B}(\mathbb{R}^d)$ la fonction $h := \mathbb{1}_B$ 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 $\mathbb{P}_X = \mu$. □

Exemple 21.
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 v.a.r. $Y := \sup(X_1, X_2)$. Pour cela notons $A := \{(x, y) \in \mathbb{R}^2 \ / \ x < y\}$. Soit h une application borélienne de $\mathbb{R}$ dans $\mathbb{R}^+$. 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 $\mathbb{E}[h(Y)] = \int_{\mathbb{R}} hd\mu$ *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 v.a.r.* Y. □

Le critère des fonctions positives exprime 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 $\mathbb{R}^d$.
Les deux autres critères d'identification de lois utilisant les fonctions de répartition ou les fonctions caractéristiques peuvent aussi s'énoncer sous cette forme. En effet, ces trois critères peuvent se formuler en un seul énoncé :

Proposition 4.10. Critère d'identification de lois (synthèse)
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*

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

lorsque u parcourt $\mathbb{R}^d$.

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

Démonstration :
1) a déja été vu.
2) On remarque que, pour tout $u \in \mathbb{R}^d$,

$$\begin{array}{rcl} \Phi_X(u) & = & \mathbb{E}\left(\exp(i\langle u, X\rangle)\right) = \mathbb{E}\left(h_u(X)\right) \\ \text{et } \Phi_\mu(u) & = & \mathbb{E}_\mu(h_u). \end{array}$$

On conclut par le théorème 3.7.
3) On se limite au cas $d = 1$. On remarque que, pour tout $u \in \mathbb{R}$,

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

On conclut par le lemme d'unicité de la proposition 1.5. □

4.4 Exercices

Exercice 36 Solution p. 91

Vérifier les formules de König-Huygens.

Exercice 37 Solution p. 91

Soit X une v.a.r. suivant la **loi de Rayleigh** de densité définie sur $\mathbb{R}$ par

$$\rho(x) := xe^{-\frac{x^2}{2}}\mathbb{1}_{\mathbb{R}^+}(x).$$

Montrer que, pour tout entier $k \geq 1$,

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

Expliciter l'espérance et la variance de la v.a.r. X.

Exercice 38 Solution p. 91

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

Exercice 39 Solution p. 92

Démontrer la proposition 4.5.

Exercice 40 Solution p. 92

Soient X et Y deux v.a.r. 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 v.a.r. X, Y et XY sont intégrables suivant $\mathbb{P}$.
2. En étudiant le signe de l'expression $\mathbb{E}\left[(X+\alpha Y)^2\right]$ pour tout $\alpha \in \mathbb{R}$, prouver l'**inégalité de Cauchy-Schwarz** :

$$|\mathbb{E}(XY)| \leq \sqrt{\mathbb{E}(X^2)}\sqrt{\mathbb{E}(Y^2)}.$$

Exercice 41 Solution p. 93

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

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

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

Exercice 42 Solution p. 93

Soit X une v.a.r. de loi $\mathcal{N}_1(0,1)$.

1. Calculer de deux façons différentes $\mathbb{E}(e^X)$.
2. Montrer que X^3 est intégrable (suivant la mesure $\mathbb{P}$) et calculer $\mathbb{E}(X^3)$.

Exercice 43 Solution p. 94

Avec les notations de l'exemple 21, montrer que $\mathbb{E}(X) = (2,2)$ et que la loi de la v.a.r. $Z := X_1 + X_2$ est

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

Exercice 44 Solution p. 95

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 45 Solution p. 96

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 46 Solution p. 96

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 47 Solution p. 97

Soit X une variable aléatoire réelle uniforme continue sur $[0, 1]$. Quelle est la loi de la variable aléatoire

$$Y = -\frac{1}{\alpha}\ln(1 - X),$$

où α est un paramètre réel strictement positif ?

Chapitre 5

Indépendance de vecteurs aléatoires

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,

$$A \times B := \left\{(x_1, \cdots, x_{n+p}) \in \mathbb{R}^{n+p} / (x_1, \cdots, x_n) \in A \text{ et } (x_{n+1}, \cdots, x_{n+p}) \in B\right\},$$

c-à-d $A \times B := (A \times \mathbb{R}^p) \cap (\mathbb{R}^n \times B)$. De même, si $a := (a_1, \cdots, a_n) \in \mathbb{R}^n$ et $b := (b_1, \cdots, b_n) \in \mathbb{R}^p$, on notera $(a, b) := (a_1, \cdots, a_n, a_{n+1}, \cdots, a_{n+p})$ considéré comme élément de $\mathbb{R}^{n+p}$. On dit que (a, b) est obtenu par **concaténation** de a et b.

5.1 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 $a := (a_1, \cdots, a_n) \in \mathbb{R}^n$ et $b := (b_1, \cdots, b_n) \in \mathbb{R}^p$. 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 :

Définition 5.1. Mesure produit
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 $\mathbb{R}^p$, $\mu\otimes\nu(A\times B)=\mu(A)\nu(B)$.
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 $\mathbb{R}^{n+p}$.*

Le résultat d'existence de la mesure produit énoncé dans la définition est admis. On notera que toutes les mesures sur $\mathbb{R}^{n+p}$ ne sont pas nécessairement des mesures-produit.

Exemple 22.
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 $\lambda^{(2)}=\lambda\otimes\lambda$.

Le théorème suivant 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 5.1. 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*
$$\int_{\mathbb{R}^{n+p}}fd(\mu\otimes\nu)=\int_{\mathbb{R}^p}\left(\int_{\mathbb{R}^n}f(x,y)d\mu(x)\right)d\nu(y)=\int_{\mathbb{R}^n}\left(\int_{\mathbb{R}^p}f(x,y)d\nu(y)\right)d\mu(x).$$

Le théorème précédent 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 vues au chapitre 4.

Ce théorème 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** dont l'énoncé est le suivant :

Proposition 5.2. Théorème de Fubini (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 de $\mathbb{R}^{n+p}$ intégrable suivant la mesure-produit $\mu \otimes \nu$. En utilisant le théorème de Tonelli, cette intégrabilité s'exprime par :

$$\int_{\mathbb{R}^{n+p}} |f| d(\mu \otimes \nu) = \int_{\mathbb{R}^p} \left(\int_{\mathbb{R}^n} |f(x,y)| d\mu(x) \right) d\nu(y) = \int_{\mathbb{R}^n} \left(\int_{\mathbb{R}^p} |f(x,y)| d\nu(y) \right) d\mu(x).$$

1. *Pour tous $y \in \mathbb{R}^p$ et $x \in \mathbb{R}^n$, les applications partielles $u \in \mathbb{R}^n \mapsto f(u,y)$ et $v \in \mathbb{R}^p \mapsto f(x,v)$ sont boréliennes.*
2. *Les applications*
$$x \in \mathbb{R}^n \mapsto \int_{\mathbb{R}^p} f(x,v) d\nu(v) \quad \text{et} \quad y \in \mathbb{R}^p \mapsto \int_{\mathbb{R}^n} f(u,y) d\mu(u)$$
sont boréliennes $\mu-$intégrable (respectivement $\nu - intgrable$),
3. *On a les égalités*
$$\int_{\mathbb{R}^{n+p}} f d(\mu \otimes \nu) = \int_{\mathbb{R}^p} \left(\int_{\mathbb{R}^n} f(x,y) d\mu(x) \right) d\nu(y) = \int_{\mathbb{R}^n} \left(\int_{\mathbb{R}^p} f(x,y) d\nu(y) \right) d\mu(x).$$

D'autres notations sont utilisées dans les ouvrages. On trouvera indifféremment

$$\begin{aligned} \int_{\mathbb{R}^{n+p}} f d(\mu \otimes \nu) &= \int_{\mathbb{R}^{n+p}} f(x,y) d(\mu \otimes \nu)(x,y) = \int_{\mathbb{R}^{n+p}} f(x,y) d\mu(x) \otimes d\nu(y) \\ &= \int_{\mathbb{R}^{n+p}} f(x,y) d\mu(x) d\nu(y). \end{aligned}$$

Exemple 23.
Considérons l'application f définie sur $\mathbb{R}^2$ par $f(x,y) := 1\!\!1_{]0,+\infty[}(x) 1\!\!1_{]a,b[}(y) e^{-xy}$ 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 $\lambda^{(2)} = \lambda \otimes \lambda$,

$$\begin{aligned} \int_{\mathbb{R}^2} f d\lambda^{(2)} &= \int_{\mathbb{R}} 1\!\!1_{]0,+\infty[}(x) \left(\int_{\mathbb{R}} 1\!\!1_{]a,b[}(y) e^{-xy} d\lambda(y) \right) d\lambda(x) \\ &= \int_{\mathbb{R}} 1\!\!1_{]a,b[}(y) \left(\int_{\mathbb{R}} 1\!\!1_{]0,+\infty[}(x) e^{-xy} d\lambda(x) \right) d\lambda(y). \end{aligned}$$

Ce qui donne en utilisant la règle d'intégration suivant la mesure de Lebesgue sur $\mathbb{R}$ des fonctions boréliennes positives,

$$\begin{aligned} \int_{\mathbb{R}^2} f d\lambda^{(2)} &= \int_0^{+\infty} \frac{e^{-ax} - e^{-bx}}{x} dx \\ &= \int_{\mathbb{R}} 1\!\!1_{]a,b[}(y) \frac{1}{y} d\lambda(y) = \int_a^b \frac{1}{y} dy = \ln\left(\frac{b}{a}\right). \end{aligned}$$

On a ainsi par la même occasion établi la valeur de l'intégrale généralisée au sens de Riemann

$$\int_0^{+\infty} \frac{e^{-ax} - e^{-bx}}{x} dx = \ln\left(\frac{b}{a}\right). \square$$

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 $f_1, \cdots, f_d$ sont les applications-composantes de F, on appelle **jacobien** de F au point $u \in U$ le déterminant de la matrice carré 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 5.3. 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 $x \in \mathbb{R}^d \mapsto \mathbb{1}_{\mathrm{V}}(x) f(T^{-1}(x))|J(x)|$ est une application borélienne positive.*
2. *Si l'application $x \in \mathbb{R}^d \mapsto \mathbb{1}_{\mathrm{U}}(x) f(x)$ est $\lambda^{(d)}$-intégrable sur $\mathbb{R}^d$, alors l'application $x \in \mathbb{R}^d \mapsto \mathbb{1}_{\mathrm{V}}(x) f(T^{-1}(x))|J(x)|$ est $\lambda^{(d)}$-intégrable sur $\mathbb{R}^d$.*

De plus, dans les deux cas ci-dessus

$$\int_{\mathbb{R}^d} \mathbb{1}_{\mathrm{U}}(u) f(u) d\lambda^{(d)}(u) = \int_{\mathbb{R}^d} \mathbb{1}_{\mathrm{V}}(v) f\left(T^{-1}(v)\right) |J(v)|\, d\lambda^{(d)}(v).$$

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.

Exemple 24.
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 $\mathbb{R}$. □

À ce stade de la lecture, vous pouvez commencer à réfléchir au devoir II.

5.2 Indépendance de vecteurs aléatoires

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 $d \geq 1$.

Définition 5.2. Densité de probabilité sur $\mathbb{R}^d$
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 montre que la règle d'intégration suivant une mesure de probabilité à densité est tout à fait analogue à celle déjà vue au chapitre 4.

Proposition 5.4. Intégration par rapport à une densité de probabilité sur $\mathbb{R}^d$ (admis)
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*

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

On remarquera que les définitions de densité et de probabilité à densité introduites au chapitre 1 sont bien des cas particuliers de la définition donnée ci-dessus. Donnons un exemple d'utilisation de cette proposition dans le calcul de lois de probabilité sur $\mathbb{R}^2$.

Exemple 25.
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 v.a.r. X. Soit A un borélien de $\mathbb{R}$. Calculons $\mathbb{P}_X(A)$. Par définition de la notion de loi et comme $\{X \in A\} = \{(X, Y) \in A \times \mathbb{R}\}$,

$$\begin{aligned} \mathbb{P}_X(A) &= \mathbb{P}(X \in A) = \mathbb{P}\left((X, Y) \in A \times \mathbb{R}\right) \\ &= \mathbb{P}_{(X,Y)}(A \times \mathbb{R}) = \int_{\mathbb{R}^2} \mathbb{1}_A(x) \mathbb{1}_\mathbb{R}(y) d\mathbb{P}_{(X,Y)}(x, y). \end{aligned}$$

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 v.a.r. 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 $X = Y$!). □

En fait on vient de montrer, sur un cas particulier, un résultat important qui affirme que si un vecteur aléatoire admet une densité, alors ses composantes sont des v.a.r. à densité. La réciproque est fausse en général comme le montre le contre-exemple proposé dans l'exercice 52, page 59. Plus précisément :

Proposition 5.5. Densité des composantes d'un vecteur aléatoire à densité sur $\mathbb{R}^d$

Si $X := (X_1, \cdots, X_d)$ 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

$$\chi_k(t) := \int_{\mathbb{R}^{d-1}} \rho(x_1, \cdots, x_{k-1}, t, x_{k+1}, \cdots, x_d)d\lambda^{(d-1)}(x_1, \cdots, x_{k-1}, x_{k+1}, \cdots, x_d).$$

La réciproque est fausse (voir exercice 52, page 59).

Exemple 26.

Revenons à l'exemple 25 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,

$$\mathbb{P}_X \otimes \mathbb{P}_Y(G) = \mathbb{P}_X\left(]\frac{1}{2}, 1[\right)\mathbb{P}_Y\left(]\frac{1}{2}, 1[\right).$$

Or

$$\mathbb{P}_X\left(]\frac{1}{2}, 1[\right) = \int_{\mathbb{R}} \mathbb{1}_{]\frac{1}{2},1[}d\mathbb{P}_X = \int_{\mathbb{R}}(1-|x|)\mathbb{1}_{[-1,1]}(x)\mathbb{1}_{]\frac{1}{2},1[}(x)d\lambda(x) = \int_{\frac{1}{2}}^1 (1-x)dx = \frac{1}{8}.$$

Ce qui montre que $0 = \mathbb{P}_{(X,Y)}(G) \neq \mathbb{P}_X \otimes \mathbb{P}_Y(G) = (\frac{1}{8})^2$ 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. $\mathbb{P}_{(X,Y)} \neq \mathbb{P}_X \otimes \mathbb{P}_Y$. □

Les situations pour lesquelles on aura l'égalité seront celles où on dira qu'il y a indépendance des variables suivant la définition :

Définition 5.3. Indépendance d'une suite finie de vecteurs aléatoires
Une suite finie de vecteurs aléatoires $(X_1, \cdots, X_n)$, de dimensions quelconques (éventuellement distinctes), est dite **indépendante (relativement à $\mathbb{P}$)** *si*

$$\mathbb{P}_{(X_1,\cdots,X_n)} = \mathbb{P}_{X_1} \otimes \mathbb{P}_{X_2} \otimes \cdots \otimes \mathbb{P}_{X_n}.$$

On dit aussi, par abus, que les v.a.r. $X_1, \cdots, X_n$ sont indépendantes. Il s'agit là d'un abus car l'indépendance est une propriété de la suite $(X_1, \cdots, X_n)$ et non de chacune des v.a.r. X_k.

Définition 5.4. Loi conjointe, loi marginale
La loi du vecteur aléatoire concaténé $X := (X_1, \cdots, X_n)$ est dite aussi **loi conjointe** *des vecteurs $X_1, \cdots, X_n$. Pour tout entier $k = 1, \cdots, n$, la loi du vecteur aléatoire X_k s'appelle alors la* **loi marginale** *de X de rang k.*

Avec cette terminologie, on peut énoncer le résultat de l'exemple précédent en exprimant que le couple de v.a.r. (X, Y) n'est pas indépendant.

Dans les chapitres sur les convergences de suite de v.a.r. on aura besoin de la définition suivante :

Définition 5.5. Indépendance d'une suite de vecteurs aléatoires
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 v.a.r. définies sur $(\Omega, \mathcal{F}, \mathbb{P})$ telle que, pour tout $k \in I$, μ_k soit la loi de la v.a.r. 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 $(X_0, \cdots, X_n)$ est indépendante.

Les propositions qui suivent ont pour but de donner des critères permettant de reconnaître si une suite finie de vecteurs aléatoires est indépendante.

Proposition 5.6. Critère d'indépendance d'une suite (finie) de vecteurs aléatoires
Soit $(X_1, \cdots, X_n)$ une suite de vecteurs aléatoires de dimensions respectives $d_1, \cdots, d_n$. Les assertions suivantes sont équivalentes :

1. *La suite $(X_1, \cdots, X_n)$ est indépendante.*
2. *Pour tout entier $1 \leq k \leq n$ et tout borélien B_k de $\mathbb{R}^{d_k}$,*

$$\mathbb{P}_{(X_1,\cdots,X_n)}(B_1 \times B_2 \times \cdots \times B_n) = \mathbb{P}_{X_1}(B_1)\mathbb{P}_{X_2}(B_2) \cdots \mathbb{P}_{X_n}(B_n).$$

3. *Pour tout entier $1 \leq k \leq n$ et tout borélien B_k de $\mathbb{R}^{d_k}$,*

$$\mathbb{P}\left[X_1, \cdots, X_n) \in B_1 \times \cdots \times B_n\right] = \mathbb{P}(X_1 \in B_1) \cdots \mathbb{P}(X_n \in B_n).$$

4. *Pour tout entier $1 \leq k \leq n$ et tout borélien B_k de $\mathbb{R}^{d_k}$,*

$$\mathbb{P}\left[\{X_1 \in B_1\} \cap \cdots \cap \{X_n \in B_n\}\right] = \mathbb{P}(X_1 \in B_1) \cdots \mathbb{P}(X_n \in B_n).$$

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 :

$$\{(X_1, \cdots, X_n) \in B_1 \times \cdots \times B_n\} = \{X_1 \in B_1\} \cap \{X_2 \in B_2\} \cap \cdots \cap \{X_n \in B_n\}.$$

- "(4) implique (1)" : admis.□

Dans les cas où on manipule des v.a.r. discrètes la proposition précédente a pour corollaire le critère d'indépendance ci-dessous. Pour simplifier on supposera que les v.a.r. sont portées par $\mathbb{N}$ mais le résultat se généralise aux v.a.r. portées par un ensemble dénombrable quelconque de $\mathbb{R}$, notamment $\mathbb{Z}$.

Proposition 5.7. Critère d'indépendance pour des v.a.r. discrètes
Soit $(X_1, \cdots, X_n)$ une suite de v.a.r. discrètes portées par $\mathbb{N}$, alors la suite $(X_1, \cdots, X_n)$ est indépendante si, et seulement si, pour tout $(k_1, \cdots, k_n) \in \mathbb{N}^n$,

$$\mathbb{P}\left(\{X_1 = k_1\} \cap \cdots \cap \{X_n = k_n\}\right) = \mathbb{P}(X_1 = k_1) \cdots \mathbb{P}(X_n = k_n).$$

Démonstration : La condition nécessaire résulte de l'implication "(1) implique (4)" de la proposition précédente où on a pris $B_i := \{k_i\}$, $i = 1, \cdots, n$, qui sont bien des boréliens de $\mathbb{R}$.
La condition suffisante résulte du fait que $\mathbb{P}_{(X_1,\cdots,X_n)}$ est une probabilité discrète sur $\mathbb{R}^n$, elle est donc entièrement déterminée par la connaissance des nombres

$$\mathbb{P}_{(X_1,\cdots,X_n)}(\{k_1\} \times \{k_2\} \times \cdots \times \{k_n\})$$

qui sont, par hypothèse, égaux à

$$\mathbb{P}(X_1 = k_1)\mathbb{P}(X_2 = k_2) \cdots \mathbb{P}(X_n = k_n).$$

De même on vérifie que $\mathbb{P}_{X_1} \otimes \cdots \otimes \mathbb{P}_{X_n}$ est une probabilités portée par $\mathbb{N}^n$ donc discrète. Elle est donc entièrement déterminée par la connaissance des nombres

$$\mathbb{P}_{X_1} \otimes \cdots \otimes \mathbb{P}_{X_n}(\{k_1\}\times\{k_2\}\times \cdots \times\{k_n\}) = \mathbb{P}(X_1 = k_1)\mathbb{P}(X_2 = k_2)\cdots\mathbb{P}(X_n = k_n).$$

Par suite

$$\mathbb{P}_{(X_1,\cdots,X_n)} = \mathbb{P}_{X_1} \otimes \cdots \otimes \mathbb{P}_{X_n}.$$

Ce qui prouve l'indépendance de la suite $(X_1, \cdots, X_n)$. □

Exemple 27.

Reprenons les notations de l'exemple 18 avec $X := (X_1, X_2)$ 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 v.a.r. (X_1, X_2). □

On vient d'établir un critère d'indépendance pour une grande famille de probabilités, celle des probabilités discrètes, donnons un critère pour une autre grande famille de probabilités, celle des probabilités à densité.

Proposition 5.8. Critère d'indépendance des v.a.r. à densité

1. *Si $(X_1, \cdots, X_n)$ est une suite indépendante de v.a.r. telle que, pour tout $k = 1, \cdots, n$, la v.a.r. X_k soit de densité ρ_k sur $\mathbb{R}$, alors le vecteur aléatoire $X := (X_1, \cdots, X_n)$ de dimension n admet pour densité sur $\mathbb{R}^n$ l'application*

$$\rho \ : \ (x_1, \cdots, x_n) \in \mathbb{R}^n \mapsto \rho(x_1, \cdots, x_n) := \rho_1(x_1)\rho_2(x_2)\cdots\rho_n(x_n) \in [0, +\infty].$$

2. *Réciproquement, si $X := (X_1, \cdots, X_n)$ est un vecteur aléatoire de dimension n admettant pour densité sur $\mathbb{R}^n$ une application ρ définie par une relation de la forme $\rho(x_1, \cdots, x_n) := g_1(x_1)g_2(x_2)\cdots g_n(x_n)$, où, pour tout $k = 1, \cdots, n$, g_k est une application borélienne positive de $\mathbb{R}$ dans $[0, +\infty]$, alors la suite de v.a.r. $(X_1, \cdots, X_n)$ est indépendante et pour tout $k = 1, \cdots, n$, la v.a.r. 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 v.a.r. 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'appli-cation ρ définie par $\rho(u,v,w) := g_1(u)g_2(v)g_3(w)$.
Commençons par étudier la loi de X.
D'après la proposition 5.5, la v.a.r. 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 $\omega \in \Omega$,

$$\mathbb{1}_{\{(X,Y,Z) \in A \times B \times C\}}(\omega) = \mathbb{1}_{\{X \in A\} \cap \{Y \in B\} \cap \{Z \in C\}}(\omega) = \mathbb{1}_{\{X \in A\}}(\omega) \mathbb{1}_{\{Y \in B\}}(\omega) \mathbb{1}_{\{Z \in C\}}(\omega)$$

et $\mathbb{1}_{\{X \in A\}}(\omega) = \mathbb{1}_A(X(\omega))$, en utilisant la définition de l'opérateur d'intégration donnée dans la proposition 3.1, puis en appliquant les théorèmes du transfert et de Tonelli, il vient

$$\begin{aligned}
\mathbb{P}_{(X,Y,Z)}(A \times B \times C) &= \mathbb{E}[\mathbb{1}_{\{(X,Y,Z) \in A \times B \times C\}}] = \mathbb{E}[\mathbb{1}_A(X) \mathbb{1}_B(Y) \mathbb{1}_C(Z)] \\
&= \int_{\mathbb{R}^3} \mathbb{1}_A(x) \mathbb{1}_B(y) \mathbb{1}_C(z) g_1(x) g_2(y) g_3(z) d\lambda^{(3)}(x,y,z,) \\
&= \left(\int_{\mathbb{R}} \mathbb{1}_A g_1 d\lambda \right) \left(\int_{\mathbb{R}} \mathbb{1}_B g_2 d\lambda \right) \left(\int_{\mathbb{R}} \mathbb{1}_C g_3 d\lambda \right).
\end{aligned}$$

De plus, comme les v.a.r. X, Y, et Z ont pour densités respectives χ_1, χ_2, et χ_3,

$$\mathbb{P}_X(A)\mathbb{P}_Y(B)\mathbb{P}_Z(C) = \left(\int_{\mathbb{R}} \mathbb{1}_A \chi_1 d\lambda\right)\left(\int_{\mathbb{R}} \mathbb{1}_B \chi_2 d\lambda\right)\left(\int_{\mathbb{R}} \mathbb{1}_C \chi_3 d\lambda\right)$$
$$= \left(\int_{\mathbb{R}} \mathbb{1}_A(t)\frac{g_1(t)}{\int_{\mathbb{R}} g_1 d\lambda} d\lambda(t)\right)\left(\int_{\mathbb{R}} \mathbb{1}_B(t)\frac{g_2(t)}{\int_{\mathbb{R}} g_2 d\lambda} d\lambda(t)\right)\left(\int_{\mathbb{R}} \mathbb{1}_C(t)\frac{g_3(t)}{\int_{\mathbb{R}} g_3 d\lambda} d\lambda(t)\right)$$
$$= \left(\int_{\mathbb{R}} \mathbb{1}_A g_1 d\lambda\right)\left(\int_{\mathbb{R}} \mathbb{1}_B g_2 d\lambda\right)\left(\int_{\mathbb{R}} \mathbb{1}_C g_3 d\lambda\right),$$

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 v.a.r. (X, Y, Z). □
Donnons un autre énoncé, beaucoup plus utile dans la pratique, du critère d'indépendance des v.a.r. à densité :

Proposition 5.9. Énoncé pratique du critère d'indépendance des v.a.r. à densité
Soit $(X_1, \cdots, X_n)$ une suite de v.a.r. telle que, pour tout $k = 1, \cdots, n$, la v.a.r. X_k soit de densité ρ_k sur $\mathbb{R}$. Alors la suite de v.a.r. $(X_1, \cdots, X_n)$ est indépendante si, et seulement si, le vecteur aléatoire $X := (X_1, \cdots, X_n)$ de dimension n admet pour densité sur $\mathbb{R}^n$ l'application

$$\rho \ : \ (x_1, \cdots, x_n) \in \mathbb{R}^n \mapsto \rho(x_1, \cdots, x_n) := \rho_1(x_1)\rho_2(x_2)\cdots\rho_n(x_n) \in [0, +\infty].$$

Démonstration : La condition nécessaire est l'assertion 1) de la proposition précédente et la condition suffisante résulte de l'assertion 2) avec, pour tout $k = 1, \cdots, n$, $g_k := \rho_k$ en remarquant que, par hypothèse, $\int_{\mathbb{R}} \rho_k d\lambda = 1$. □.

Exemple 28.

1. *Reprenons les notations de l'exercice 53, page 59. 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 v.a.r. 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 v.a.r. (X, Y) en vertu de la proposition précédente.*

2. *Dans l'exemple 25, on vérifie aisément que $\rho(x, y) \neq \chi(x)\chi(y)$, ce qui est une autre façon de prouver que la suite de v.a.r. (X, Y) n'est pas indépendante.* □

Donnons un critère valable pour des vecteurs aléatoires généraux sans hypothèses sur le type de loi qu'ils satisfont.

Proposition 5.10. Critère d'indépendance utilisant des fonctions positives
Soit $(X_1, \cdots, X_n)$ une suite de vecteurs aléatoires de dimensions respectives $d_1, \cdots, d_n$. Alors, la suite $(X_1, \cdots, X_n)$ 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 $[0, +\infty]$,

$$\mathbb{E}\left[f_1(X_1)f_2(X_2)\cdots f_n(X_n)\right] = \mathbb{E}\left[f_1(X_1)\right]\mathbb{E}\left[f_2(X_2)\right]\cdots\mathbb{E}\left[f_n(X_n)\right].$$

Démonstration : Faisons la démonstration pour $n = 2$.
C.N. Par application des théorèmes du transfert et de Tonelli où on utilise l'indépen-dance en écrivant $\mathbb{P}_{(X_1,X_2)} = \mathbb{P}_{X_1} \otimes \mathbb{P}_{X_2}$, il vient

$$\begin{aligned}
\mathbb{E}\left[f_1(X_1)f_2(X_2)\right] &= \int_{\mathbb{R}^{d_1+d_2}} f_1(x)f_2(y)d\mathbb{P}_{(X_1,X_2)}(x,y) \\
&= \int_{\mathbb{R}^{d_1}} f_1(x)\left(\int_{\mathbb{R}^{d_2}} f_2(y)d\mathbb{P}_{X_2}(y)\right)d\mathbb{P}_{X_1}(x) \\
&= \left(\int_{\mathbb{R}^{d_1}} f_1 d\mathbb{P}_{X_1}\right)\left(\int_{\mathbb{R}^{d_2}} f_2 d\mathbb{P}_{X_2}\right) \\
&= \mathbb{E}\left[f_1(X_1)\right]\mathbb{E}\left[f_2(X_2)\right].
\end{aligned}$$

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

$$\mathbb{E}\left[\mathbb{1}_A(X_1)\mathbb{1}_B(X_2)\right] = \mathbb{E}\left[\mathbb{1}_A(X_1)\right]\mathbb{E}\left[\mathbb{1}_B(X_2)\right]$$

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 $\mathbb{P}_{(X_1,X_2)} = \mathbb{P}_{X_1} \otimes \mathbb{P}_{X_2}$. □
A titre d'exemple d'utilisation de cette proposition, donnons un corollaire très utile dans les calculs faisant intervenir des v.a.r. indépendantes et intégrables :

Proposition 5.11. Espérance d'un produit de v.a.r. indépendantes intégrables
Soit $(X_1, \cdots, X_n)$ une suite de v.a.r. intégrables. Si la suite $(X_1, \cdots, X_n)$ est indépendante, alors la v.a.r. $X_1X_2\cdots X_n$ est intégrable et

$$\mathbb{E}[X_1X_2\cdots X_n] = \mathbb{E}(X_1)\mathbb{E}(X_2)\cdots\mathbb{E}(X_n).$$

La réciproque est fausse.

Démonstration : Faisons-la pour deux v.a.r. 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 précédente avec les fonctions positives $x \mapsto |x|$ grâce à l'indépendance de (X, Y). On a donc prouvé que la v.a.r. 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 5.10 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. □
L'exercice 54, page 60, propose un contre-exemple prouvant que la réciproque de la proposition précédente est fausse. Cet exercice est souvent à la base de nombreux contre-exemples concernant l'indépendance des variables aléatoires.
Comme corollaire de la proposition précédente,

Proposition 5.12. Indépendance et matrice de dispersion
Soit $X := (X_1, X_2, ..., X_d)$ un vecteur aléatoire de carré intégrable de dimension d. Si la suite de v.a.r. $(X_1, X_2, ..., X_d)$ est indépendante, alors la matrice de dispersion de X est diagonale (la réciproque est fausse).

Démonstration : D'après la proposition précédente, pour tout couple d'entiers (i, j) avec $i \neq j$, $\mathrm{Cov}(X_i, X_j) = 0$. On conclut en utilisant la proposition 4.5, (3). □
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 5.13. Indépendance de fonctions de v.a.r.
Si $(X_1, \cdots, X_n, Y_1, \cdots, Y_p)$ 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 $(\varphi(X_1, \cdots, X_n), \psi(Y_1, \cdots, Y_p))$ est indépendant.

Démonstration : Considérons les vecteurs aléatoires $X := (X_1, \cdots, X_n)$ et $Y := (Y_1, \cdots, Y_p)$. Commençons par montrer que le couple de vecteurs aléatoires (X, Y) est indépendant. Comme $(X_1, \cdots, X_n, Y_1, \cdots, Y_p)$ est une suite indépendante, pour tous boréliens de $\mathbb{R}$, $A_1, \cdots, A_n$, il vient

$$\begin{aligned}
\mathbb{P}_X[A_1 \times \cdots \times A_n] &= \mathbb{P}_{(X_1,\cdots,X_n)}[A_1 \times \cdots \times A_n] \\
&= \mathbb{P}_{(X_1,\cdots,X_n,Y_1,\cdots,Y_p)}[A_1 \times \cdots \times A_n \times \mathbb{R}^p] \\
&= \mathbb{P}_{X_1}(A_1) \cdots \mathbb{P}_{X_n}(A_n).
\end{aligned}$$

Ce qui prouve que $\mathbb{P}_X = \mathbb{P}_{(X_1,\cdots,X_n)} = \mathbb{P}_{X_1} \otimes \cdots \otimes \mathbb{P}_{X_n}$, et par suite

$$\begin{aligned}
\mathbb{P}_{(X,Y)} &= \mathbb{P}_{(X_1,\cdots,X_n,Y_1,\cdots,Y_p)} \\
&= \mathbb{P}_{X_1} \otimes \cdots \otimes \mathbb{P}_{X_n} \otimes \mathbb{P}_{Y_1} \otimes \cdots \otimes \mathbb{P}_{Y_p} \\
&= \mathbb{P}_X \otimes \mathbb{P}_Y.
\end{aligned}$$

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.□

On pourra vérifier que si un couple de vecteurs aléatoires (X, Y) de dimensions respectives d et k est indépendant avec $X := (X_1, \cdots, X_d)$ et $Y := (Y_1, \cdots, Y_k)$ 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 5.14. Indépendance de fonctions de v.a.r. (cas particulier usuel)
Si $(X_1, \cdots, X_n)$ est une suite indépendante de v.a.r. , alors, pour toute suite d'applications boréliennes $(f_1, \cdots, f_n)$ de $\mathbb{R}$ dans $\mathbb{R}$, la suite de v.a.r. $(f_1(X_1), \cdots, f_n(X_n))$ est indépendante.

Terminons par un critère simple d'application utilisant les fonctions caractéristiques.

Proposition 5.15. Critère d'indépendance par les fonctions caractéristiques
Une suite de v.a.r. $(X_1, \cdots, X_n)$ est indépendante si, et seulement si, pour tout $(u_1, \cdots, u_n) \in \mathbb{R}^n$,

$$\begin{array}{rcl} \mathbb{E}\left[\exp\left(i\sum_{k=1}^n u_k X_k\right)\right] & = & \mathbb{E}\left(e^{iu_1X_1}\right)\cdots\mathbb{E}\left(e^{iu_nX_n}\right) \\ \textit{i.e.}\quad \Phi_{(X_1,\cdots,X_n)}(u_1, \cdots, u_n) & = & \Phi_{X_1}(u_1)\cdots\Phi_{X_n}(u_n). \end{array}$$

5.3 Somme de v.a.r. indépendantes

Démontrons d'abord un important corollaire de l'exercice 41, page 41.

Proposition 5.16. Variance d'une somme de v.a.r. indépendantes
Si $(X_1, X_2, \cdots, X_n)$ est une suite indépendante de v.a.r. de carré intégrable, alors

$$\mathbb{E}\left[\left(\sum_{k=1}^n X_k\right)^2\right] < +\infty \quad \textit{et} \quad \mathit{Var}\left(\sum_{k=1}^n X_k\right) = \sum_{k=1}^n \mathit{Var}(X_k).$$

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 v.a.r.

$$\mathbb{E}\left[\left(\sum_{k=1}^n X_k\right)^2\right] \leq K\mathbb{E}\left(\sum_{k=1}^n |X_k|^2\right) = K\sum_{k=1}^n \mathbb{E}\left(|X_k|^2\right) < +\infty.$$

On a vérifié dans l'exercice 41, page 41, que

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

et on conclut en remarquant que, par indépendance des v.a.r. , $\mathrm{Cov}(X_i, X_j) = 0$ pour tout couple d'entiers (i, j) tel que $i \neq j$. □

On peut généraliser la proposition 5.16 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 v.a.r. indépendantes.

Proposition 5.17. Fonction caractéristique d'une somme de v.a.r. indépendantes

Si $(X_1, X_2, \cdots, X_n)$ est une suite indépendante de v.a.r. , alors, pour tout $u \in \mathbb{R}$,

$$\begin{aligned} \mathbb{E}\left[\exp\left(iu \textstyle\sum_{k=1}^{n} X_k\right)\right] &= \mathbb{E}\left(e^{iuX_1}\right) \cdots \mathbb{E}\left(e^{iuX_n}\right) \\ \text{i.e. } \Phi_{X_1+\cdots+X_n}(u) &= \Phi_{X_1}(u) \cdots \Phi_{X_n}(u). \end{aligned}$$

La réciproque est fausse.

Démonstration : Il suffit de prendre $u := u_1 = u_2 = \cdots = u_n$ dans la condition nécessaire du critère des fonctions caractéristiques. □

Proposition 5.18. Somme de v.a.r. gaussiennes indépendantes

Si $(X_1, X_2, \cdots, X_p)$ est une suite indépendante de v.a.r. de lois respectives $\mathcal{N}_1(m_1, \sigma_1^2), \cdots, \mathcal{N}_1(m_p, \sigma_p^2)$, alors la v.a.r. $S_p := X_1 + \cdots + X_p$ est une loi gaussienne $\mathcal{N}_1(m_1 + \cdots + m_p, \sigma_1^2 + \cdots + \sigma_p^2)$.

Démonstration : On applique le résultat précédent en notant que la fonction ca-ractéristique d'une v.a.r. 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 3.7 en remarquant que la fonction ca-ractéristique de la v.a.r. S_p est celle de la loi $\mathcal{N}_1(m_1 + \cdots + m_p, \sigma_1^2 + \cdots + \sigma_p^2)$. □

Citons, à titre d'exemple, un autre résultat fondamental dont la démonstration est laissée en exercice :

Proposition 5.19. Somme de v.a.r. de Bernoulli indépendantes

Si $(X_1, X_2, \cdots, X_n)$ est une suite indépendante de v.a.r. de Bernoulli de même paramètre $p \in]0, 1[$, alors la v.a.r. $S_n := X_1 + \cdots + X_n$ suit la loi binomiale $\mathcal{B}(n, p)$.

Dans certains cas on peut directement calculer la loi de la v.a.r. "somme". En voici deux exemples énoncés sous forme de propositions.

Proposition 5.20. Loi de la somme de v.a.r. binomiales, de v.a.r. de Poissson

1. *Si (X, Y) est un couple indépendant de v.a.r. de lois binomiales respectives $\mathcal{B}(n, p)$ et $\mathcal{B}(m, p)$ de même paramètre $p \in]0, 1[$, alors la v.a.r. $X + Y$ suit la loi binomiale $\mathcal{B}(n + m, p)$.*
2. *Si (X, Y) est un couple indépendant de v.a.r. de lois de Poisson respectives $\mathcal{P}(\alpha)$ et $\mathcal{P}(\beta)$, où les réels α et β sont strictement positifs, alors la v.a.r. $X + Y$ suit la loi de Poisson $\mathcal{P}(\alpha + \beta)$.*

La démonstration de cette proposition est proposée en exercice.

Voici quelques résultats plus généraux sur les sommes de v.a.r. indépendantes. Plus que de retenir des formules, Il faut surtout être capable de refaire directement les calcul 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 5.21. Densité de la somme de v.a.r. indépendantes
Soit (X, Y) un couple indépendant de v.a.r. admettant pour densités respectives ρ_X et ρ_Y, alors la v.a.r. $X + Y$, 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,

$$\begin{aligned}
\mathbb{E}[h(X+Y)] &= \int_{\mathbb{R}^2} h(u)\rho_X(v)\rho_Y(u-v)d\lambda^{(2)}(u,v) \\
&= \int_{\mathbb{R}} h(u) \left(\int_{\mathbb{R}} \rho_X(v)\rho_Y(u-v)d\lambda(v) \right) d\lambda(u) \\
&= \int_{\mathbb{R}^d} h(u)\rho_{X+Y}(u)d\lambda^{(d)}(u),
\end{aligned}$$

ce qui prouve que ρ_{X+Y} est bien la densité de $X + Y$. □

Exemple 29.
Soit (X, Y) un couple indépendant de v.a.r. . On suppose que la v.a.r. 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 v.a.r. $X + Y$ 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

$$\rho_{X+Y}(t) = \left(1 - e^{-t}\right) \mathbb{1}_{[0,1]}(t) + e^{-t}(e-1)\mathbb{1}_{]1,+\infty[}(t).\square$$

5.4 Exercices

Exercice 48 Solution p. 97

1. Déduire du résultat de l'exemple 24 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

$$\int_{\mathbb{R}^2} e^{-(x^2+2xy+2y^2)} d\lambda^{(2)}(x,y) = \pi.$$

Exercice 49 Solution p. 98

Démontrer la première partie de la proposition 5.4.

Exercice 50 Solution p. 98

Citer un exemple de v.a.r. X et Y distinctes et de même loi.

Exercice 51 Solution p. 98

Démontrer la proposition 5.5 pour $d = 3$ en adaptant la démarche développée dans l'exemple 25 dans lequel on avait $d = 2$.

Exercice 52 Solution p. 99

Soit X une v.a.r. de loi $\mathcal{N}_1(0,1)$. 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 5.5 est fausse.

Exercice 53 Solution p. 99

Soit (U,V) un couple indépendant de v.a.r. 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.
2. Montrer que les v.a.r. X et Y ont toutes les deux pour loi $\mathcal{N}_1(0,1)$.
3. Déduire des questions précédentes que le couple de v.a.r. (X,Y) est indépendant.

Exercice 54 Solution p. 100

Soient X une v.a.r. de loi $\mathcal{N}_1(0,1)$ et ε une v.a.r. indépendante de X de loi $\frac{1}{2}(\delta_{-1}+\delta_1)$. Montrer que la v.a.r. $Y := \varepsilon X$ 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 $\mathbb{E}(X^2Y^2)$ et on utilisera la proposition 5.10).

Exercice 55 Solution p. 101

Soit (X,Y) un couple indépendant de v.a.r. de même loi $\mathcal{N}_1(m,\sigma^2)$, montrer que $\mathbb{E}[(X+Y)^2] = 4m^2 + 2\sigma^2$.

Exercice 56 Solution p. 101

Trouver, parmi les exercices ou exemples déjà proposés, un contre-exemple prouvant que la réciproque de la proposition 5.16 est fausse.

Exercice 57 Solution p. 101

Construire un contre-exemple à la proposition 5.17 en considérant les deux v.a.r. X et $Y := X$ où X est une v.a.r. 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 considèrera le borélien $A := [-\frac{1}{2}, -\frac{1}{2}]$ et on calculera $\mathbb{P}_{(X,Y)}(A \times A^c)$ qu'on comparera au produit $\mathbb{P}_X(A)\mathbb{P}_Y(A^c)$.

Exercice 58 Solution p. 101

Démontrer la proposition 5.19.

Exercice 59 Solution p. 101

Soit $(X_1, \cdots, X_n)$ une suite indépendante de n v.a.r. 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 $X_1(\omega), \cdots, X_n(\omega)$ dans l'ordre décroissant et on note $X_{(k)}(\omega)$ le k$^{\text{ième}}$ de ces nombres ainsi rangés où $k = 1, \cdots, n$.

1. Vérifier que $X_{(1)} = \max(X_1, \cdots, X_n)$ et calculer sa fonction de répartition.
2. Que représente la v.a.r. $X_{(n)}$? Calculer sa fonction de répartition.
3. Soit $t > 0$. Pour tout $k = 1, \cdots, n$, on pose $Y_k := \mathbb{1}_{]t,+\infty[}(X_k)$.
   (a) Prouver que la loi de la v.a.r. Y_k est
   $$\mathbb{P}_{Y_k} = e^{-\alpha t}\delta_1 + (1 - e^{-\alpha t})\delta_0.$$
   Quelle est la loi de la v.a.r. $Y_1 + Y_2 + \cdots + Y_n$?
   (b) Comparer les événements $\{X_{(k)} \leq t\}$ et $\{Y_1 + Y_2 + \cdots + Y_n \leq k - 1\}$.
4. Déterminer, pour tout $k = 1, \cdots, n$, la fonction de répartition de la v.a.r. $X_{(k)}$.

Exercice 60 Solution p. 102

Démontrer la proposition 5.20 en utilisant les fonctions caractéristiques.

Exercice 61 Solution p. 102

Soient X et Y deux v.a.r. indépendantes. On suppose que la v.a.r. X est binomiale de paramètres n, $\frac{1}{2}$ et que la v.a.r. Y est binomiale de paramètres m, $\frac{1}{2}$. On admet la relation $\sum_{k=0}^{n} C_n^k C_m^{m-k} = C_{n+m}^m$.
Calculer la probabilité que $X = Y$.

Exercice 62 Solution p. 103

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 $\omega \in \Omega$,

$$\tau_r(\omega) := \inf\{n \in \mathbb{N}^*/X_1(\omega) + X_2(\omega) + ... + X_n(\omega) = r\}$$

et

$$\theta_r(\omega) := \inf\{n \in \mathbb{N}^*/X_1(\omega) + X_2(\omega) + ... + X_{n+r}(\omega) = r\}$$

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 v.a.r. τ_r est une v.a.r. discrète de loi (dite **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 v.a.r. θ_r est une v.a.r. discrète de loi (dite **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 v.a.r. 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 ?

Chapitre 6

Convergences et théorèmes limites

Dans ce chapitre, toutes les v.a.r. considérées sont supposées définies sur un même espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$.

6.1 Convergence en probabilité d'une suite de v.a.r. , loi faible des grands nombres

Commençons par démontrer deux inégalités souvent utilisées en probabilité :

Proposition 6.1. Inégalité de Markov
Soient X une v.a.r. et φ une application borélienne de $\mathbb{R}$ dans $[0, +\infty]$, alors pour tout $a > 0$,

$$\mathbb{P}\left[\varphi(X) \geq a\right] \leq \frac{\mathbb{E}[\varphi(X)]}{a}.$$

Démonstration : Posons $A := \{\varphi(X) \geq a\}$. Puisque φ est positive, on peut écrire :

$$\mathbb{E}[\varphi(X)] = \mathbb{E}[\varphi(X)\mathbb{1}_A] + \mathbb{E}[\varphi(X)\mathbb{1}_{A^c}] \geq \mathbb{E}[\varphi(X)\mathbb{1}_A] \geq a\mathbb{E}(\mathbb{1}_A) = a\mathbb{P}(A). \ \square$$

Proposition 6.2. Inégalité de Bienaymé-Tchébycheff
Soit X une v.a.r. telle que $\mathbb{E}(X^2) < +\infty$. Alors pour tout $\alpha > 0$,

$$\mathbb{P}\left[|X - \mathbb{E}(X)| \geq \alpha\right] \leq \frac{1}{\alpha^2} Var(X).$$

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. Étudions un exemple d'application de cette inégalité.

Exemple 30.

1. *Si X est une v.a.r. de loi binomiale $\mathcal{B}(10, \frac{1}{2})$ alors $\mathbb{E}(X) = 5$ et $Var(X) = \frac{5}{2}$. Un calcul direct donne l'estimation*

$$\mathbb{P}\left(|X-5| \geq 4\right) = \left(C_{10}^0 + C_{10}^1 + C_{10}^9 + C_{10}^{10}\right)\left(\frac{1}{2}\right)^{10} \approx 0,021,$$

alors que l'inégalité de Bienaymé-Tchébycheff donne la majoration

$$\mathbb{P}\left(|X-5| \geq 4\right) \leq \frac{2,5}{4^2} \approx 0,156.$$

2. *Si X est une v.a.r. de carré intégrable de variance σ^2, l'inégalité de Bienaymé-Tchébycheff devient $\mathbb{P}\left(|X-m| \geq \sigma\right) \leq 1$, alors que dans le cas d'une v.a.r. gaussienne de loi $\mathcal{N}_1(m, \sigma^2)$*

$$\begin{aligned}\mathbb{P}\left(|X-m| \geq \sigma\right) &= 2\int_\sigma^{+\infty} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{x^2}{2\sigma^2}\right) dx \\ &= 2\int_1^{+\infty} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t^2}{2}\right) dt \approx 0,3174,\end{aligned}$$

et par suite $\mathbb{P}\left(m-\sigma < X < m+\sigma\right) \approx 0,6826$. $\square$

Définition 6.1. Suite de v.a.r. identiquement-distribuée
Une suite $(X_n)_{n\in\mathbb{N}}$ de v.a.r. est dite identiquement-distribuée si toutes les v.a.r. de la suite ont la même loi. Une suite **i.i.d.** *de v.a.r. est une suite indépendante et identiquement distribuée de v.a.r. .*

Définition 6.2. Moyenne empirique
Si $(X_n)_{n\in\mathbb{N}}$ est une suite de v.a.r. on note $\overline{X}_n$, pour tout entier $n \geq 1$, la moyenne empirique (d'ordre n) associée à la suite $(X_n)_{n\in\mathbb{N}}$, définie par

$$\overline{X}_n := \frac{X_1 + \cdots + X_n}{n}.$$

La démonstration de la proposition suivante est immédiate et laissée en exercice.

Proposition 6.3. Moyenne et variance de la moyenne empirique
Si $(X_n)_{n\in\mathbb{N}}$ est une suite indépendante et identiquement distribuée de v.a.r. de carré intégrable, de moyenne m et de variance σ^2, alors

$$\mathbb{E}\left(\overline{X}_n\right) = m \quad \text{et } Var\left(\overline{X}_n\right) = \frac{\sigma^2}{n}.$$

Une application de l'inégalité de Bienaymé-Tchébycheff dont le résultat est historiquement célèbre est :

Proposition 6.4. Théorème de Bernoulli
Soit $(X_n)_{n\in\mathbb{N}}$ une suite indépendante et identiquement distribuée de v.a.r. 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 $n \geq 1$,

$$\mathbb{P}\left(\left|\frac{X_1 + \cdots + X_n}{n} - p\right| \geq a\right) \leq \frac{1}{4a^2 n}.$$

Démonstration : On applique l'inégalité de Bienaymé-Tchébycheff à la v.a.r. $\overline{X}_n$ et on utilise la proposition précédente pour obtenir la majoration

$$\mathbb{P}\left(\left|\frac{X_1+\cdots+X_n}{n}-m\right|\geq a\right)\leq\frac{\mathrm{Var}(X_0)}{na^2}.$$

Comme les v.a.r. sont de Bernoulli $\mathbb{E}(X_0)=m=p$ et $\mathrm{Var}(X_0)=p(1-p)$. On conclut en vérifiant que, pour tout $p\in]0,1[$, $p(1-p)\leq\frac{1}{4}$. □

Exemple 31. *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 expérience associe le nombre* 1 *si la pièce tombe sur Pile lors du* $k^{\text{ième}}$ *lancer et* 0 *sinon, la v.a.r.* X_k *est une v.a.r. de Bernoulli de paramètre p inconnu. On modélise cette situation en représentant les lancers successifs indépendants par une suite i.i.d. de v.a.r. 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*

$$\overline{X}_{4040}(\omega_0)=\frac{2049}{4040}\approx 0,507.$$

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,*

$$\mathbb{P}\left(\omega\in\Omega\ /\ \left|\overline{X}_{4040}(\omega)-p\right|\geq 0,0352\right)\leq 0,05.$$

La probabilité que la réalisation ω_0 *de cette expérience satisfasse l'événement*

$$\{\omega\in\Omega\ /\ |\overline{X}_{4040}(\omega)-p|\geq 0,0352\}$$

est inférieure à 0,05. *Autrement dit, la probabilité que le paramètre p vérifie la condition* $|\overline{X}_{4040}(\omega_0)-p|\geq 0,0352$, *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. □

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 de Statistique du semestre suivant.

Définition 6.3. Convergence en probabilité
On dit qu'une suite de v.a.r. $(X_n)_{n\in\mathbb{N}}$ *converge en probabilité vers une v.a.r. Y si, pour tout réel* $a>0$,

$$\lim_{n\to+\infty}\mathbb{P}\left(|X_n-Y|\geq a\right)=0.$$

Proposition 6.5. Loi faible des grands nombres
Si $(X_n)_{n\in\mathbb{N}}$ *est une suite indépendante et identiquement distribuée de v.a.r. de carré intégrable et d'espérance m, alors la suite des moyennes empiriques converge en probabilité vers la v.a.r. constante m.*

Démonstration : On applique l'inégalité de Bienaymé-Tchébycheff comme dans la démonstration du théorème de Bernoulli pour obtenir

$$\mathbb{P}\left(\left|\frac{X_1+\cdots+X_n}{n}-m\right|\geq a\right)\leq\frac{\mathrm{Var}(X_0)}{na^2}.$$

Ce qui donne le résultat par passage à la limite. □

Exemple 32.

Si, pour tout entier $n \geq 1$, X_n est une v.a.r. de loi $\frac{1}{n+1}\delta_n + \frac{n}{n+1}\delta_{\frac{1}{n}}$, alors la suite $(X_n)_{n\in\mathbb{N}}$ converge en probabilité vers la v.a.r. constante 0. En effet, soit $a > 0$ et n un entier tel que $n > a > \frac{1}{n}$. Comme la v.a.r. 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 $+\infty$. □

La limite d'une suite de v.a.r. convergeant en probabilité est "presque-sûrement" unique comme l'énonce de façon précise le résultat suivant :

Proposition 6.6. Unicité de la limite en probabilité

Si $(X_n)_{n\in\mathbb{N}}$ est une suite de v.a.r. convergeant en probabilité vers les v.a.r. X et Y, alors les v.a.r. X et Y sont égales presque-sûrement i.e. $\mathbb{P}(X \neq Y) = 0$.

Démonstration : Avec les notations du théorème on peut écrire,

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

Soit $a > 0$ un réel, on vérifie l'inclusion entre événements

$$\{|X - Y| \geq a\} \subseteq \{|X - X_n| \geq \frac{a}{2}\} \cup \{|X_n - Y| \geq \frac{a}{2}\}.$$

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, $\mathbb{P}(|X - Y| \geq a) = 0$. Or

$$\{X \neq Y\} = \{|X - Y| > 0\} = \bigcup_{n\geq 1} \left\{|X - Y| \geq \frac{1}{n}\right\}.$$

D'où

$$\mathbb{P}(X \neq Y) \leq \sum_{n\geq 1} \mathbb{P}\left(|X - Y| \geq \frac{1}{n}\right) = 0.$$

□

La proposition suivante permet d'effectuer des calculs sur les limites en probabilité.

Proposition 6.7. Calculs sur les limites en probabilité (admis)

1) *Soient f une application continue de $\mathbb{R}^2$ dans $\mathbb{R}$, $(X_n)_{n\in\mathbb{N}}$ et $(Y_n)_{n\in\mathbb{N}}$ deux suites de v.a.r. convergeant en probabilité respectivement vers les v.a.r. X et Y, alors la suite de v.a.r. $(f(X_n, Y_n))_{n\in\mathbb{N}}$ converge en probabilité vers la v.a.r. $f(X, Y)$.*

2) *Si $(X_n)_{n\in\mathbb{N}}$ et $(Y_n)_{n\in\mathbb{N}}$ sont deux suites de v.a.r. convergeant en probabilité respectivement vers les v.a.r. X et Y, alors les suites de v.a.r. $(X_n + Y_n)_{n\in\mathbb{N}}$ et $(X_nY_n)_{n\in\mathbb{N}}$ convergent en probabilité respectivement vers les v.a.r. $X + Y$ et XY.*

6.2 Convergence presque-sûre d'une suite de v.a.r. , loi forte des grands nombres

On va énoncer la loi forte sous sa forme classique en introduisant un autre mode de convergence pour les suites de v.a.r. .
Auparavant remarquons que :

Définition 6.4. Convergence presque-sûre
Si $(X_n)_{n\in\mathbb{N}}$ est une suite de v.a.r. et Y une v.a.r. , alors l'ensemble Δ_Y des $\omega \in \Omega$ tels que la suite réelle $(X_n(\omega))_{n\in\mathbb{N}}$ ne converge pas vers $Y(\omega)$ est un événement de l'espace de probabilité $(\Omega, \mathcal{F}, \mathbb{P})$ i.e. $\Delta_Y \in \mathcal{F}$.
On dira que la suite v.a.r. $(X_n)_{n\in\mathbb{N}}$ **converge presque-sûrement** *vers Y si $\mathbb{P}(\Delta_Y) = 0$.*

Il faut vérifier que $\Delta_Y \in \mathcal{F}$. On remarque 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 $(X_n(\omega))_{n\in\mathbb{N}}$ 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 v.a.r. X_i et Y ainsi que la stabilité des tribus pour les opérations de réunion et d'intersection dénombrables.

Exemple 33.
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 v.a.r. $X_n := \pi + n\mathbb{1}_{[0,\frac{1}{n}]}$. Alors la suite $(X_n)_{n\in\mathbb{N}}$ converge presque-sûrement vers la v.a.r. 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.

La proposition suivante, difficile à démontrer, a une grande importance de par sa signification probabiliste. C'est en partie grâce à ce théorème que le formalisme des probabilités trouve sa cohérence.

Proposition 6.8. Loi forte des grands nombres
Si $(X_n)_{n\in\mathbb{N}}$ est une suite indépendante et identiquement distribuée de v.a.r. intégrables et d'espérance m, alors la suite des moyennes empiriques $(\overline{X}_n)_{n\in\mathbb{N}}$ converge presque-sûrement vers la v.a.r. constante m.

On peut aussi dire que l'ensemble des $\omega \in \Omega$ tels que la suite des moyennes empiriques $(\overline{X}_n(\omega))_{n\in\mathbb{N}}$ 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 v.a.r. X_k est une v.a.r. de Bernoulli de paramètre p inconnu. On modélise cette situation en considérant les essais successifs indépendants par une suite i.i.d. de v.a.r. 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

v.a.r. X_k, c-à-d la fréquence de réalisation de l'événement A, est une bonne approximation de la probabilité de A.
Donnons quelques propriétés de cette convergence. La limite d'une suite de v.a.r. convergeant presque-sûrement est "unique" dans le sens suivant :

Proposition 6.9. Unicité de la limite presque-sûre
Si $(X_n)_{n\in\mathbb{N}}$ est une suite de v.a.r. convergeant presque-sûrement vers les v.a.r. X et Y, alors les v.a.r. 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 $\mathbb{P}(X \neq Y) \leq \mathbb{P}(\Delta_X) + \mathbb{P}(\Delta_Y) = 0$. □
La proposition suivante permet d'effectuer des calculs sur les limites presque-sûres.

Proposition 6.10. Calculs sur les limites presque-sûres
1) *Soit f une application continue de $\mathbb{R}^2$ dans $\mathbb{R}$, $(X_n)_{n\in\mathbb{N}}$ et $(Y_n)_{n\in\mathbb{N}}$ deux suites de v.a.r. convergeant presque-sûrement respectivement vers les v.a.r. X et Y, alors la suite de v.a.r. $(f(X_n, Y_n))_{n\in\mathbb{N}}$ converge presque-sûrement vers la v.a.r. $f(X, Y)$.*
2) *Si $(X_n)_{n\in\mathbb{N}}$ et $(Y_n)_{n\in\mathbb{N}}$ sont deux suites de v.a.r. convergeant presque-sûrement respectivement vers les v.a.r. X et Y, alors les suites de v.a.r. $(X_n + Y_n)_{n\in\mathbb{N}}$ et $(X_nY_n)_{n\in\mathbb{N}}$ convergent presque-sûrement respectivement vers les v.a.r. $X + Y$ et XY.*

La convergence presque-sûre est plus forte que la convergence en probabilité comme le précise la proposition :

Proposition 6.11. Convergence presque-sûre et en probabilité
Si $(X_n)_{n\in\mathbb{N}}$ est une suite de v.a.r. convergeant presque-sûrement vers la v.a.r. Y, alors la suite $(X_n)_{n\in\mathbb{N}}$ converge en probabilité vers la v.a.r. Y. La réciproque est fausse.

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,

$$\mathbb{P}\left(\bigcap_{m\in\mathbb{N}} \bigcup_{n\geq m} \{|X_n - Y| \geq a\}\right) = \lim_{m\to+\infty} \mathbb{P}\left(\bigcup_{n\geq m} \{|X_n - Y| \geq a\}\right).$$

Comme, pour tout entier m,

$$\mathbb{P}\left(\bigcup_{n\geq m} \{|X_n - Y| \geq a\}\right) \geq \mathbb{P}\left(\{|X_m - Y| \geq a\}\right),$$

on obtient

$$0 \geq \mathbb{P}\left(\bigcap_{m\in\mathbb{N}} \bigcup_{n\geq m} \{|X_n - Y| \geq a\}\right) \geq \lim_{m\to+\infty} \mathbb{P}\left(\{|X_m - Y| \geq a\}\right) \geq 0.$$

Ce qui entraîne que $\lim_{m\to+\infty} \mathbb{P}(|X_m - Y| \geq a) = 0$ c-à-d que la suite de v.a.r. $(X_n)_{n\in\mathbb{N}}$ converge en probabilité vers la v.a.r. Y. □

Remarque. *Tenant compte de ce que les v.a.r. de carré intégrable sont des v.a.r. intégrables, la loi faible des grands nombres apparaît ainsi comme une conséquence immédiate de la loi forte.*

6.3 Convergence en loi d'une suite de v.a.r.

L'importance du dernier théorème de convergence qui va être énoncé dans le prochain paragraphe nécessite l'introduction d'une notion de convergence dans l'ensemble 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. On notera $C_b(\mathbb{R}^d)$ l'espace des fonctions continues bornées de $\mathbb{R}^d$ dans $\mathbb{R}$ muni de la norme infinie définie par $\|\varphi\|_\infty = \sup_{x\in\mathbb{R}^d} |\varphi(x)|$.

Définition 6.5. Convergence étroite, convergence en loi

1. *Soient $(\mu_n)_{n\in\mathbb{N}}$ une suite de probabilités sur $\mathbb{R}^d$ et μ une probabilité sur $\mathbb{R}^d$. On dit que la suite de probabilités $(\mu_n)_{n\in\mathbb{N}}$* **converge (étroitement)** *vers la probabilité μ (on note $\mu_n \xrightarrow[n\to\infty]{e} \mu$) si, pour*
$$\forall\varphi \in C_b(\mathbb{R}^d) \quad , \quad \int \varphi d\mu_n \xrightarrow[n\to\infty]{} \int \varphi d\mu \ .$$
2. *Si $(X_n)_{n\in\mathbb{N}}$ est une suite de v.a.r. et X une v.a.r. , on dit parfois par abus que la suite $(X_n)_{n\in\mathbb{N}}$* **converge en loi** *vers la v.a.r. X pour exprimer que la suite $(\mathbb{P}_{X_n})_{n\in\mathbb{N}}$ des lois des v.a.r. X_n converge étroitement vers la loi $\mathbb{P}_X$ de la v.a.r. X. On a $X_n \xrightarrow[n\to\infty]{\text{en loi}}$ équivaut à $\mathbb{E}(\varphi(X_n)) \xrightarrow[n\to\infty]{} \mathbb{E}(\varphi(x))$ pour toute fonction $\varphi \in C_b(\mathbb{R}^d)$.*

On remarquera que les variables aléatoires X_n ne sont pas forcément définies sur le même espace probabilisé. On prendra garde que cette dernière terminologie est très dangereuse car une suite de v.a.r. $(X_n)_{n\in\mathbb{N}}$ peut converger en loi vers des v.a.r. différentes. Tout ce qu'on peut affirmer a priori c'est que toutes ces v.a.r. auront alors la même loi comme conséquence du théorème qui suit :

Proposition 6.12. Unicité de la limite pour la convergence étroite
Si une suite de probabilités sur $\mathbb{R}^d$, $(\mu_n)_{n\in\mathbb{N}}$, converge vers deux probabilités sur $\mathbb{R}^d$, 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. On rappelle qu'on a vu dans le premier chapitre 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 un ensemble dénombrable. 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 $(\mu_n)_{n\in\mathbb{N}}$, 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 $\mu = \nu$. □

Proposition 6.13. Critère des fonctions de répartition pour la convergence en loi (admis)
Soit $(X_n)_{n\in\mathbb{N}}$ une suite de v.a.r. et X une v.a.r. .

1. *Si $X_n \xrightarrow[n\to\infty]{\text{en loi}} X$, alors $F_{X_n}(x) \xrightarrow[n\to\infty]{} F_X(x)$ en tout point de continuité de la fonction de répartition F_X de la v.a.r. X.*
2. *Réciproquement, si*
   - *$F_{X_n}(x) \xrightarrow[n\to\infty]{} F(x)$ en tout point de continuité de la fonction F*
   - *la fonction F est continue à droite*
   - *$\lim_{x\to-\infty} F(x) = 0$*
   - *$\lim_{x\to+\infty} F(x) = 1$*

   alors F est la fonction de répartition d'une variable aléatoire X et $X_n \xrightarrow[n\to\infty]{\text{en loi}} X$.

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 6.14. Critère de convergence pour des 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 $(\mu_n(\{k\}))_{n\in\mathbb{N}}$ 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. □

Donnons comme exemple d'application aux lois classiques du critère des probabilités discrètes le résultat suivant :

Proposition 6.15. Loi binomiale $\mathcal{B}(n;p)$ et loi de Poisson
Soit $\alpha \in]0,+\infty[$. Si $(p_n)_{n\in\mathbb{N}}$ est une suite de réels de $]0,1[$ telle que $\lim_n(np_n)=\alpha$, alors la suite de probabilités $(\mathcal{B}(n,p_n))_{n\in\mathbb{N}^*}$ converge vers $\mathcal{P}(\alpha)$.

Démonstration : Fixons $k \in \mathbb{N}$, par définition de la probabilité binomiale

$$\mathcal{B}(n,p_n)(\{k\}) = C_n^k p_n^k (1-p_n)^{n-k} = \frac{n(n-1)...(n-k+1)}{k!} p_n^k (1-p_n)^{n-k}.$$

Au voisinage de $+\infty$, n^k est un équivalent de $n(n-1)...(n-k+1)$ 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

$$\lim_{n\to+\infty} \frac{(np_n)^k}{k!}(1-p_n)^{n-k} = \frac{\alpha^k}{k!}e^{-\alpha}$$

et par équivalence

$$\lim_{n\to+\infty} \mathcal{B}(n,p_n)(\{k\}) = \frac{\alpha^k}{k!}e^{-\alpha} = \mathcal{P}(\alpha)(\{k\}),$$

pour tout entier k, ce qui donne le résultat cherché. □
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.
Dans les autres cas, par exemple celui des probabilités à densité, on préfère utiliser le critère général de convergence suivant :

Proposition 6.16. Critère des fonctions caractéristiques pour la convergence en loi (admis)
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 $(\mu_n)_{n\in\mathbb{N}}$ converge vers μ si, et seulement si, la suite des fonctions caractéristiques $(\Phi_{\mu_n})_{n\in\mathbb{N}}$ converge simplement sur $\mathbb{R}^d$ vers la fonction caractéristique Φ_μ.

Exemple 34.
Si $(a_n)_{n\in\mathbb{N}}$ et $(\sigma_n)_{n\in\mathbb{N}}$ sont deux suites réelles convergeant respectivement vers les réels a et σ, alors la suite de probabilités $(\mathcal{N}_1(a_n,\sigma_n^2))_{n\in\mathbb{N}}$ converge vers la probabilité $\mathcal{N}_1(a,\sigma^2)$. □
Pour compléter l'étude des liens entre les diverses convergences, signalons le résultat suivant qui prouve que la convergence en probabilité, et a fortiori la convergence presque-sûre, d'une suite de v.a.r. implique la convergence de la suite des lois de ces v.a.r. :

Proposition 6.17. Convergence en probabilité et convergence en loi (admis)
◇ *Si $(X_n)_{n\in\mathbb{N}}$ est une suite de v.a.r. convergeant en probabilité vers la v.a.r. Y, alors la suite des lois $(\mathbb{P}_{X_n})_{n\in\mathbb{N}}$ converge vers la loi $\mathbb{P}_Y$ de la v.a.r. Y.*
◇ *Si $(X_n)_{n\in\mathbb{N}}$ est une suite de v.a.r. 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 v.a.r. $(X_n)_{n\in\mathbb{N}}$ converge en probabilité vers la v.a.r. constante a.*

La convergence des lois d'une suite de v.a.r. n'implique pas nécessairement la convergence en probabilité de la suite de v.a.r. , cependant c'est vrai si la suite des lois converge vers une probabilité de Dirac.

6.4 Théorème-limite central (TLC) et applications

Pour introduire le TLC, commençons par une proposition élémentaire très utilisée en statistique inférentielle.

Proposition 6.18. Loi de $\overline{X}_n$ dans le cas gaussien
Si $(X_n)_{n\in\mathbb{N}}$ est une suite indépendante et identiquement distribuée de v.a.r. gaussiennes d'espérance m et de variance $\sigma^2 > 0$, alors, pour tout entier $n > 0$, la v.a.r. $\frac{X_1+\cdots+X_n}{n} - m$ est une variable aléatoire normale centrée de variance $\frac{\sigma^2}{n}$. On peut aussi dire que la v.a.r. $S_n := X_1 + \cdots + X_n$ est une variable aléatoire normale d'espérance nm et de variance $n\sigma^2$.

Démonstration : Soit $n \in \mathbb{N}^*$ fixé. Comme $(X_1, \cdots, X_n)$ est une suite indépendante de v.a.r. gaussiennes, la somme $X_1 + \cdots + X_n$ est une v.a.r. 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 $\mathcal{N}_1(0, \frac{\sigma^2}{n})$. □

Si, dans les hypothèses de la proposition 6.18, 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, lui aussi très important en statistique inférentielle :

Proposition 6.19. Théorème-limite central
Si $(X_n)_{n\in\mathbb{N}}$ est une suite indépendante et identiquement distribuée de v.a.r. de carré intégrable d'espérance m et de variance $\sigma^2 > 0$, alors,

$$\frac{\frac{1}{n}\sum_{k=1}^{n} X_k - m}{\frac{\sigma}{\sqrt{n}}} \xrightarrow[n\to\infty]{\text{en loi}} \mathcal{N}(0;1) \ .$$

On peut aussi écrire que pour tout $x \in \mathbb{R}$,

$$\lim_{n\to+\infty} \mathbb{P}\left(\frac{\frac{1}{n}\sum_{k=1}^{n} X_k - m}{\frac{\sigma}{\sqrt{n}}} \leqslant x\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt.$$

La loi forte des grands nombres affirme que presque-sûrement, pour n assez grand, la moyenne empirique $\overline{X}_n$ 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 $\overline{X}_n - m$ commise en prenant $\overline{X}_n$ comme estimation de m.

Démonstration : Cette démonstration est proposée dans un des exercices du devoir III. □

Donnons un corollaire du théorème-limite central historiquement important.

Proposition 6.20. Théorème de De Moivre-Laplace
Si $p \in]0,1[$ et, pour tout entier $n \geq 1$, Z_n est une v.a.r. de loi binomiale $\mathcal{B}(n,p)$, alors pour tout réel x

$$\lim_{n\to+\infty} \mathbb{P}\left(\frac{Z_n - np}{\sqrt{np(1-p)}} \leq x\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} dt.$$

Démonstration : D'après la proposition 5.20, la v.a.r.

$$\frac{Z_n - np}{\sqrt{np(1-p)}}$$

a même loi que la v.a.r.

$$Y_n := \frac{X_1 + \cdots + X_n - np}{\sqrt{np(1-p)}}$$

où $(X_n)_{n\in\mathbb{N}}$ est une suite indépendante et identiquement distribuée de v.a.r. de Bernoulli de loi $\mathcal{B}(p)$. De plus avec les notations habituelles, $m = \mathbb{E}(X_0) = p$, $\sigma^2 = \mathrm{Var}(X_0) = p(1-p)$. On applique alors le théorème-limite central à la suite $(X_n)_{n\in\mathbb{N}}$. □

On notera qu'une suite de probabilités discrètes peut converger vers une probabilité non-discrète comme le prouve le théorème de De Moivre-Laplace.

Le théorème de De Moivre-Laplace exprime que dans la pratique, pour n assez grand, une v.a.r. X binomiale de taille n peut être approximée par une v.a.r. 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 v.a.r. 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 il est d'usage de considérer cette approximation satisfaisante si $np \geqslant 5$ et $n(p-1) \geqslant 5$.

À ce stade de la lecture, vous pouvez commencer à réfléchir au devoir III (après avoir fait les exercices proposés ci-après).

6.5 Exercices

Exercice 63 Solution p. 105

Soit X une v.a.r. telle que $\mathbb{E}(X^2) < +\infty$ de variance σ^2 et de moyenne m.

1. Montrer que
$$\mathbb{P}\left(m - 3\sigma < X < m + 3\sigma\right) \geq \frac{8}{9} \text{ et } \mathbb{P}\left(m - 2\sigma < X < m + 2\sigma\right) \geq \frac{3}{4}.$$
2. Comparer les valeurs obtenues dans la question précédente avec celles données par le calcul dans le cas où X est une v.a.r. de loi $\mathcal{N}_1(m, \sigma^2)$.

Exercice 64 Solution p. 105

1. On considère une v.a.r. X de carré intégrable, centrée et de variance σ^2. A l'aide de l'inégalité de Cauchy-Schwarz (voir exercice 40, page 41), montrer que, pour tout $a > 0$,
$$a \leq \mathbb{E}\left[(a - X)\mathbb{1}_{]-\infty,a]}(X)\right] \leq \sqrt{\mathbb{P}(X \leq a)}\sqrt{\sigma^2 + a^2}.$$
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 $\mathrm{Var}(Y) = 400$. 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.

Exercice 65 Solution p. 106

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 ?

Exercice 66 Solution p. 106

Montrer que si, pour tout entier $n \geq 1$, X_n est une v.a.r. de densité $n\mathbb{1}_{[0,\frac{1}{n}]}$ alors la suite $(X_n)_{n\in\mathbb{N}}$ converge en probabilité vers la v.a.r. constante 0.

Exercice 67 Solution p. 106

Trouver, parmi les exercices ou exemples déjà proposés dans ce cours, un contre-exemple prouvant que si la suite de v.a.r. $(X_n)_{n\in\mathbb{N}}$ converge en probabilité vers la v.a.r. X, cela n'entraîne pas nécessairement que la suite des espérances (si elles existent) $(\mathbb{E}(X_n))_{n\in\mathbb{N}}$ converge vers le réel $\mathbb{E}(X)$ (s'il existe).

Exercice 68 Solution p. 107

Démontrer la proposition 6.10.

Exercice 69 Solution p. 107

Soit $(X_n)_{n\in\mathbb{N}}$ une suite i.i.d. de v.a.r. de carré intégrable. On note, pour tout entier $n \geq 2$,

$$S_n^2 := \frac{1}{n-1}\sum_{k=1}^{n}\left(X_k - \overline{X}_n\right)^2,$$

la **variance empirique d'ordre** n de la suite $(X_n)_{n\in\mathbb{N}}$. Montrer que

$$S_n^2 = \frac{1}{n-1}\sum_{k=1}^{n} X_k^2 - \frac{n}{n-1}\left(\overline{X}_n\right)^2$$

et en déduire que la suite de v.a.r. $(S_n^2)_{n\in\mathbb{N}}$ converge presque-sûrement vers $\mathrm{Var}(X_0)$.

Exercice 70 Solution p. 107

En considérant la suite de probabilités de Dirac au point n sur $\mathbb{R}$, $(\delta_n)_{n\in\mathbb{N}}$, 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 $(F_{\mu_n})_{n\in\mathbb{N}}$ sans que la suite $(\mu_n)_{n\in\mathbb{N}}$ converge.

Exercice 71 Solution p. 108

Soit $(X_n)_{n\geq 1}$ une suite indépendante de v.a.r. de même loi de Poisson $\mathcal{P}(1)$. Pour $n \geq 1$, on pose $S_n := \sum_{k=1}^{k=n} X_k$. Préciser la loi de S_n et calculer la limite de la suite

$$\left(e^{-n}\sum_{k=0}^{k=n}\frac{n^k}{k!}\right)_{n\geq 1}.$$

Exercice 72 Solution p. 108

Un fabriquant produit des transistors dont 1 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 ?

Exercice 73 Solution p. 108

Démontrer l'affirmation de l'exemple 34. 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 des probabilités.

Exercice 74 Solution p. 108

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 $p \leq 0,01$?

Chapitre 7

Solutions des exercices

On ne donne pas dans ce chapitre toutes les solutions détaillées des exercices. Cependant, pour chaque exercice, au moins une indication est donnée. Quelques exercices font l'objet d'une attention particulière.

7.1 Solutions des exercices du chapitre 1

Exercice 1 Enoncé p. 11

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 montré 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 équivalence[1]. 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 :

$$x \in \{f \in A\}^c \Leftrightarrow x \notin \{f \in A\} \Leftrightarrow f(x) \notin A \Leftrightarrow f(x) \in A^c \Leftrightarrow x \in \{f \in A^c\} \ .$$

Exercice 2 Enoncé p. 11

[1]Si vous n'êtes pas sûr de vous lors de l'écriture d'une équivalence, vérifiez rapidement les deux implications pour vous en persuader

1. Dans les deux cas (a) et (b) on a $\mathbb{1}_{A\cap B} = \mathbb{1}_A.\mathbb{1}_B$. 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

	$\mathbb{1}_A(x)$	$\mathbb{1}_B(x)$	$\mathbb{1}_A(x).\mathbb{1}_B(x)$	$\mathbb{1}_{A\cup B}(x)$ 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	$\surd$	1

Ainsi dans le cas (a) : $\mathbb{1}_{A\cup B} = \mathbb{1}_A + \mathbb{1}_B$
et dans le cas (b) : $\mathbb{1}_{A\cup B} = \mathbb{1}_A + \mathbb{1}_B - \mathbb{1}_A.\mathbb{1}_B$.

2. On a trivialement $\mathbb{1}_{A^c} = 1 - \mathbb{1}_A$ et $\mathbb{1}_{A\setminus B} = \mathbb{1}_{A\cap B^c} = \mathbb{1}_A.\mathbb{1}_{B^c} = \mathbb{1}_A(1-\mathbb{1}_B)$. On remarque que si $B \subseteq A$, $\mathbb{1}_A.\mathbb{1}_B = \mathbb{1}_B$ et dans ce cas $\mathbb{1}_{A\setminus B} = \mathbb{1}_A - \mathbb{1}_B$.
Enfin $\mathbb{1}_{A\cup B\uplus C} = \mathbb{1}_{(A\cup B)\cup C} = \mathbb{1}_{A\cup B} + \mathbb{1}_C - \mathbb{1}_C\mathbb{1}_{A\cup B}$. On développe de même l'indicatrice de $A\cup B$ et on obtient :

$$\mathbb{1}_{A\cup B\cup C} = \mathbb{1}_A + \mathbb{1}_B + \mathbb{1}_C - (\mathbb{1}_A\mathbb{1}_B + \mathbb{1}_B\mathbb{1}_C + \mathbb{1}_C\mathbb{1}_A) + (\mathbb{1}_A\mathbb{1}_B\mathbb{1}_C)$$

Exercice 3

Enoncé p. 12

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 $E = A_1 \cup \cdots \cup A_n$ par définition d'une partition. Donc E s'écrit bien comme $\cup_{i\in I} A_i$ en choisissant $I = \{1, ..., n\}$.
ii) Soit $B \in \mathcal{A}$. Montrons que $B^c \in \mathcal{A}$. On a :

$$B \in \mathcal{A} \Leftrightarrow \exists I \subset \{1, ..., n\} \; ; \; B = \cup_{i\in I} A_i$$

et comme les $A_1, ..., A_n$ forment une partition de E on a $B^c = \cup_{j\in J} A_j$ où $J = \{1, ..., n\} \setminus I$. 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 $I_k \subset \{1, ..., n\}$ 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 $J = \cup_{k\in\mathbb{N}} I_k \subset \{1, ..., n\}$. Ainsi $\cup_{k\in\mathbb{N}} B_k$ est bien la réunion d'une sous-famille des $A_1, ..., A_n$ et donc $\cup_{k\in\mathbb{N}} B_k \in \mathcal{A}$.

Exercice 4

Enoncé p. 12

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 $A \in \mathcal{B}$, $A^c \in \mathcal{B}$.

- Soit $(A_k)_{k\in\mathbb{N}}$ une suite d'éléments de $\mathcal{B}$. On a $\forall n \in \mathbb{N}$; $\forall i \in I$; $A_k \in \mathcal{A}_i$. 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 $\mathcal{A} = \{\emptyset, A, A^c, E\}$ 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 $A \cup B \in \mathcal{C}$.

Exercice 5 Enoncé p. 12

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 $\mathcal{A}$.

Exercice 6 Enoncé p. 12

On note $\mathcal{F}$ la famille des tribus 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 des tribus de $\mathcal{F}$. C'est un tribu d'après le résultat de l'exercice 4 (voir page 12). On remarque que l'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 $\mathcal{B} \subset \mathcal{A}$ car $\mathcal{B}$ est l'intersection des tribus de la famille $\mathcal{F}$.

Des trois points ci-dessus, on déduit que $\mathcal{B} = \mathcal{B}(\mathbb{R})$ est la tribu de Borel sur $\mathbb{R}$.

Exercice 7 Enoncé p. 12

Il faut montrer que si $\mathcal{B}$ est une tribu contenant les parties $A_1, ..., A_n$, alors $\mathcal{A} \subset \mathcal{B}$. Soit $\mathcal{B}$ une telle tribu. Par définition et propriétés des tribus, elle contient toutes les réunions des sous-familles de $A_1, ..., A_n$ c'est-à-dire que pour tout sous-ensemble I de $\{1, ..., n\}$, $\cup_{i\in I} A_i \in \mathcal{B}$. Donc $\mathcal{B}$ contient tous les éléments de $\mathcal{A}$ et on a donc bien $\mathcal{A} \subset \mathcal{B}$.

Exercice 8 Enoncé p. 12

1. Soit $A_1, ..., A_n$ 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_o} 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

$$\mu(\bigcup_1^n A_i) = \sum_{k\in\bigcup_i A_i} \frac{1}{k^2} = \sum_1^n (\sum_{k\in A_i} \frac{1}{k^2}) = \sum_1^n \mu(A_i).$$

Si $\bigcup_1^n A_i$ est infini, alors il existe un A_{i_o} 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 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 $(\{k\})_{\mathbb{N}^*}$ 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

$$\mu\left(\bigcup_{k\geq 1}\{k\}\right) \neq \sum_{k\geq 1}\mu\left(\{k\}\right).$$

En conséquence, l'application μ n'est pas σ-additive bien qu'elle soit additive.

Exercice 9 Enoncé p. 13

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 $\sigma-$additivité de la mesure de Lebesgue λ sur $\mathbb{R}$, on obtient

$$\lambda(\mathbb{R}) = \lambda\left(\bigcup_{k\in\mathbb{Z}}]k, k+1]\right) = \sum_{k=-\infty}^{k=+\infty}\lambda\big(]k,k+1]\big)\ .$$

Or pour tout $k \in \mathbb{Z}$, $\lambda(]k,k+1]) = 1$ (la longueur de l'intervalle $]k,k+1]$), d'où $\lambda(\mathbb{R}) = \sum_{k=-\infty}^{+\infty} 1 = +\infty$.

Exercice 10 Enoncé p. 13

On peut remarquer tout d'abord que si A est un borélien alors $\delta_a(A) = \mathbb{1}_{\{a\}}(A)$. 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 $a \in \cup_{n\in\mathbb{N}}A_n$ alors il existe un unique n_0 (les A_n sont disjoints deux à deux) tel que $a \in A_{n_0}$ et donc $\delta_a(\cup_{n\in\mathbb{N}}A_n) = 1$ et

$$\sum_{n=0}^{\infty}\delta_a(A_n) = \underbrace{\delta_a(A_0) + ... + \delta_a(A_{n_0-1})}_{=0} + \underbrace{\delta_a(A_{n_0})}_{=1} + \underbrace{\textstyle\sum_{k=n_0+1}^{\infty}\delta_a(A_k)}_{=0},$$

  d'où l'égalité $\delta_a(\cup_{n\in\mathbb{N}}A_n) = \sum_{n\in\mathbb{N}}\delta_a(A_n)$.
  Si maintenant $a \notin \cup_{n\in\mathbb{N}}A_n$, alors $\forall n \in \mathbb{N}$; $a \notin A_n$ et par suite $\delta_a(\cup_{n\in\mathbb{N}}A_n) = 0$ 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.

Exercice 11 Enoncé p. 13

Avec les notations de la proposition 1.4 il suffit de prendre $\mu_k = \delta_k$, pour tout $k \in \mathbb{N}$ et

1. pour la probabilité binomiale,
   - $\alpha_k = C_n^k p^k(1-p)^{n-k}$ pour $0 \leqslant k \leqslant n$

– $\alpha_k = 0$ pour $k \geqslant n+1$.

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 $k \geqslant 1$.
4. pour la probabilité uniforme-discrète,
   – $\alpha_k = \frac{1}{n}$ pour $1 \leqslant k \leqslant n$
   – $\alpha_0 = 0$ et $\alpha_k = 0$ pour $k \geqslant n+1$.

On vérifie aisément que $\sum_{k\geqslant 0} \alpha_k = 1$ dans chacun de ces cas.

Exercice 12 Enoncé p. 13

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 $\mathcal{B}(7;0.3)(\{0,3,5\}) \simeq 0.3343$.

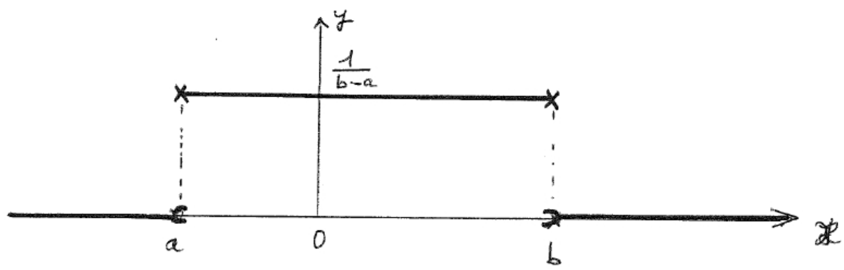
Exercice 13 Enoncé p. 13

1. La fonction $\rho = \frac{\mathbb{1}_{[a,b]}}{b-a}$ est bien positive, continue sauf en deux points de $\mathbb{R}$ et

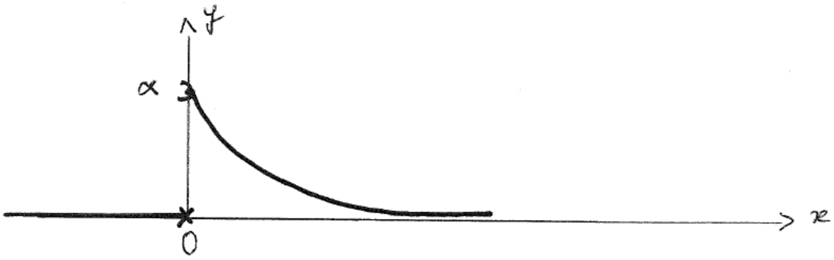
$$\int_{-\infty}^{+\infty} \rho(x)dx = (b-a) \times \frac{1}{b-a} = 1 \ ,$$

   ainsi c'est une densité. Il en est de même pour la deuxième fonction $\rho(t) = \alpha e^{-\alpha t}\mathbb{1}_{\mathbb{R}^+}(t)$.

2. La courbe représentative de $x \mapsto \frac{\mathbb{1}_{[a,b]}(x)}{b-a}$ est



   et celle de $x \mapsto \alpha e^{-\alpha x}\mathbb{1}_{\mathbb{R}^+}(x)$



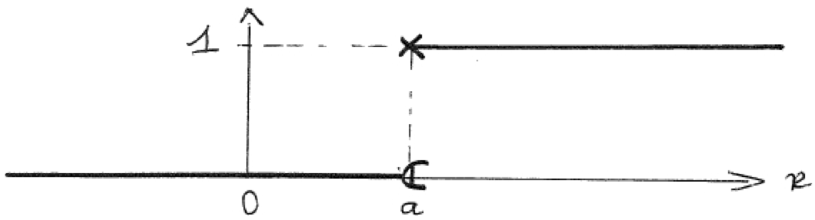
3. Des calculs élémentaires donnent

$$\mathcal{U}\big([0,1]\big)\,\big([-\infty,4/3]\big) = 1 \quad \text{et} \quad \mathcal{E}(2)\,\big([-\infty,7]\big) = 1 - e^{-14}\ .$$

Exercice 14 Enoncé p. 13

Pour chaque valeur de x on calcule $\mu\big(]-\infty,x]\big)$. On trouve les allures suivantes pour chaque courbe.

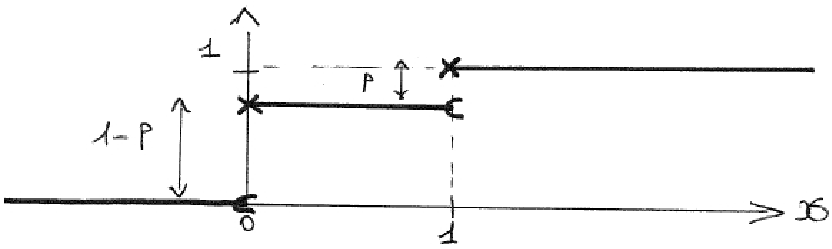
1. La courbe représentative de $F_{\delta_a}(x) = \mathbb{1}_{[a,+\infty[}(x)$ est



2. On a

$$F_{\mathcal{B}(p)}(x) = \begin{cases} 0 & \text{si } x < 0\,; \\ 1-p & \text{si } 0 \leqslant x < 1\,; \\ 1 & \text{si } 1 \leqslant x. \end{cases}$$

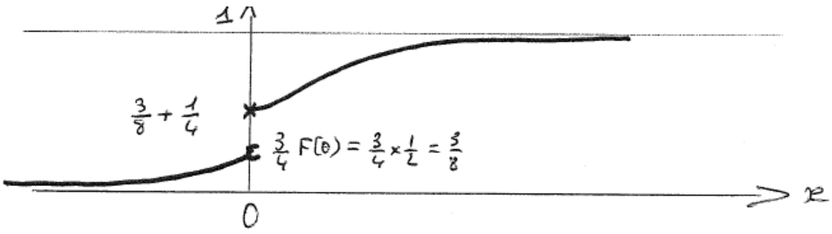
et sa courbe représentative est



3. On a

$$F_{\frac{1}{4}\delta_0+\frac{3}{4}\mathcal{N}(0,1)}(x) = \begin{cases} \frac{3}{4}F(x) & \text{si } x < 0\,; \\ \frac{3}{4}F(x) + \frac{1}{4} & \text{si } x \geqslant 0. \end{cases}$$

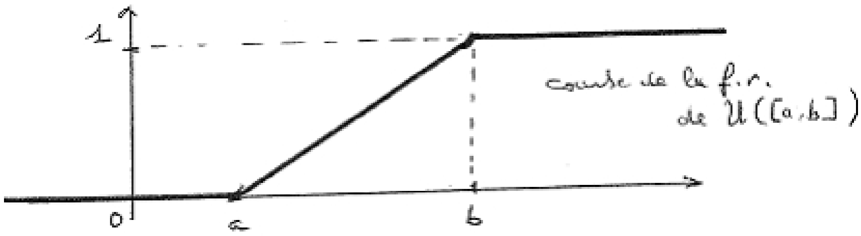
dont la courbe est la suivante :



Exercice 15 Enoncé p. 13

1. On a $\mathcal{E}(\alpha)(]-\infty,x]) = (1-e^{-\alpha x})\mathbb{1}_{]0,+\infty[}(x)$ et
2. Pour la loi uniforme sur $[a,b]$ on a

$$\mathcal{U}([a,b])(]-\infty,x]) = \begin{cases} 0 & \text{si } x < a \\ \frac{x-a}{b-a} & \text{si } a \leqslant x \leqslant b\,; \\ 1 & \text{si } x \geqslant b. \end{cases}$$



Exercice 16 Enoncé p. 13

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

$$f(x) = \begin{cases} \frac{1}{2}e^{+x}, & \text{si } x \leqslant 0\,; \\ \frac{1}{2}e^{-x}, & \text{si } x \geqslant 0. \end{cases}$$

Ainsi F est la fonction de répartition d'une probabilité à densité f définie par $f(x) = \frac{1}{2}e^{-|x|}$.

Exercice 17 Enoncé p. 14

Le fait que $B_n \subset A_n$ 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 $x \in B_m$, $x \notin A_0 \cup A_1 \cup ... \cup A_{m-1}$ donc $x \notin A_n$ car $n \leqslant m-1$. Donc $x \notin B_n$ car $B_n \subset A_n$. 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 $B_n \subset A_n$, $\cup_{n=0}^{\infty} B_n \subset \cup_{n=0}^{\infty} A_n$. Montrons l'inclusion inverse. Soit $x \in \cup_{n=0}^{\infty} A_n$, 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 $x \in \cup_{n=0}^{\infty} B_n$. On vient de montrer l'inclusion $\cup_{n=0}^{\infty} A_n \subset \cup_{n=0}^{\infty} B_n$ et on en déduit donc l'égalité souhaitée.

Exercice 18 Enoncé p. 14

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 1.8, 2.(c) 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. On a alors $\mathcal{N}_1(0;1)(\{\alpha_k\}) = p_k = 0$ pour tout $k \in \mathbb{N}$ 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.

Exercice 19 Enoncé p. 14

1. Cela résulte de la continuité de la fonction de répartition $x \mapsto \int_{-\infty}^{x} \rho(t)dt$ et de la proposition 1.8, 2.(c).
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 $\mu([a,b[) = \mu(]a,b[) + \mu(\{a\}) = (F(b-) - F(a)) + (F(a) - F(a-)) = F(b-) - F(a-)$.

Exercice 20 Enoncé p. 14

La fonction de répartition F de $\mathcal{U}([0,1])$ est continue sur $\mathbb{R}$ et est donnée par

$$F(x) = \begin{cases} 0, & \text{si } x < 0\,; \\ x, & \text{si } 0 \leqslant x \leqslant 1\,; \\ 1, & \text{si } x > 1. \end{cases}$$

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.,

$$\mathcal{U}([0,1])(\mathbb{Q}) = \sum_{q\in\mathbb{Q}} \underbrace{\mathcal{U}([0,1])(\{q\})}_{=0} = 0\ .$$

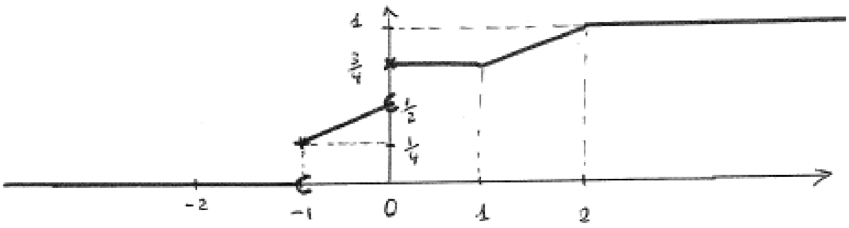
La fonction de répartition de $\mathcal{E}(2)$ est continue et est donnée par

$$F(x) = \begin{cases} 0, & \text{si } x < 0\,; \\ 1 - e^{-2x}, & \text{si } x \geqslant 0. \end{cases}$$

Donc $\mathcal{E}(2)(\{\pi\}) = 0$ et $\mathcal{E}(2)(\{\pi\} \cup [9/2,7]) = \mathcal{E}(2)(\{\pi\}) + \mathcal{E}(2)([9/2,7]) = 0 + F(7) - F(9/2) = e^{-9} - e^{-14}$.

Exercice 21 Enoncé p. 14

La représentation graphique de F est



On peut écrire F de la manière (moins synthétique mais plus lisible) suivante :

$$F(x) = \begin{cases} 0, & \text{si } x < -1\,; \\ \frac{x+2}{4}, & \text{si } -1 \leqslant x < 0 \text{ ou } 1 \leqslant x \leqslant 2\,; \\ \frac{3}{4}, & \text{si } 0 \leqslant x \leqslant 1\,; \\ 1, & \text{si } x \geqslant 2. \end{cases}$$

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

$$F_1(x) = \begin{cases} 0, & \text{si } x < -1\,; \\ \frac{x+1}{4}, & \text{si } -1 \leqslant x < 0\,; \\ \frac{1}{4}, & \text{si } 0 \leqslant x \leqslant 1\,; \\ \frac{x}{4}, & \text{si } 1 \leqslant x \leqslant 2\,; \\ \frac{1}{2}, & \text{si } x \geqslant 2. \end{cases}$$

et

$$F_2(x) = \begin{cases} 0, & \text{si } x < -1\,; \\ \frac{1}{4}, & \text{si } -1 \leqslant x < 0\,; \\ \frac{1}{2}, & \text{si } x \geqslant 0. \end{cases} .$$

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

$$f_1(t) = \begin{cases} \frac{1}{4}, & \text{si } -1 \leqslant t < 0 \text{ ou } 1 \leqslant t \leqslant 2\,; \\ 0, & \text{sinon.} \end{cases} .$$

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.

7.2 Solutions des exercices du chapitre 2

Exercice 22 Enoncé p. 22

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.

Exercice 23 Enoncé p. 22

Soit F_X et F_Y les f.r. de X et Y respectivement. Soit $t \in \mathbb{R}$.

- Si $t \leqslant 0$, comme $\forall \omega \in \Omega$, $Y(\omega) = e^{X(\omega)} > 0$ on en déduit que $F_Y(t) = \mathbb{P}(Y \leqslant t) = \mathbb{P}(\emptyset) = 0$.
- 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

$$F_Y(t) = \mathbb{P}(Y \leqslant t) = \mathbb{P}(e^X \leqslant t) = \mathbb{P}(X \leqslant \ln t) = F_X(\ln t)$$
$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\ln t} e^{-u^2/2} du$$

et en faisant le changement de variable $u = \ln x$ (en remarquant que $-\infty < u \leqslant \ln t \Leftrightarrow 0 < x < t$ 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.

Exercice 24 Enoncé p. 22

On raisonne comme dans l'exercice 23, page 22. Tout d'abord, $F_Y(t) = 0$ pour $t < 0$ car Y est une v.a.r. positive.
Soit $t \geqslant 0$. Comme F_X est continue sur $\mathbb{R}$, on a

$$F_Y(t) = \mathbb{P}(Y \leqslant t) = \mathbb{P}(-t \leqslant X \leqslant t) = F_X(t) - F_X(-t)$$
$$= \frac{1}{2}\left(2 - e^{-t}\right) - \frac{1}{2} e^{-t} = 1 - e^{-t} .$$

Par suite Y est une variable aléatoire à densité $\rho(t) = e^{-t} \mathbb{1}_{]0,\infty[}(t)$ c'est à dire que Y est une v.a.r. de loi $\mathcal{E}(1)$ (voir le formulaire).

Exercice 25 Enoncé p. 22

On écrit les équivalences suivantes :

$$\begin{aligned} Y = \sigma X + m \quad \text{suit une loi } \mathcal{N}_1(m;\sigma^2) &\Leftrightarrow \forall t \in \mathbb{R}\ ;\ \mathbb{P}(Y \leqslant t) = \int_{-\infty}^{t} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(u-m)^2}{2\sigma^2}} du \\ &\Leftrightarrow \forall t \in \mathbb{R}\ ;\ \mathbb{P}\left(X \leqslant \frac{t-m}{\sigma}\right) = \int_{-\infty}^{t} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-m)^2}{2\sigma^2}} du \\ &\Leftrightarrow \forall x \in \mathbb{R}\ ;\ \mathbb{P}\left(X \leqslant x\right) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{v^2}{2}} dv\ , \end{aligned}$$

où on a d'abord posé $x = (t-m)/\sigma$ puis fait le changement de variable $v = (u-m)/\sigma$.

Exercice 26 Enoncé p. 22

Pour n fixé quelconque :

$$\mathbb{P}_X(\{n\}) = \mathbb{P}(X = n) = \sum_{k=0}^{\infty} p_k \delta_k(\{n\}) = p_n$$

car on rappelle que $\delta_k(\{n\}) = 1$ si $k = n$ et 0 sinon.

7.3 Solutions des exercices du chapitre 3

Exercice 27 Enoncé p. 30

1. D'après la conclusion 1.(b) de la proposition 3.1, $\mathbb{E}_\mu(f) = \sum_{i=1}^n a_i \mathbb{E}_\mu(\mathbb{1}_{A_i})$. Or d'après 1.(a) de cette même proposition, $\mathbb{E}_\mu(\mathbb{1}_{A_i}) = \mu(A_i)$ et donc $\mathbb{E}_\mu(f) = \sum_{i=1}^n a_i \mu(\mathbb{1}_{A_i})$.
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 $\mathbb{E}_\mu(f) = \frac{4}{3}\pi + 7$.

Exercice 28 Enoncé p. 30

1. On introduit l'application $\tilde{I}$: : $\mathcal{M}^+ \to [0,\infty]$ 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, c'est-à-dire :
   (a) Pour tout $A \in \mathcal{A}$, $\tilde{I}(\mathbb{1}_A) = (\mu + \nu)(A)$.
   (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 $\mathcal{M}^+(E, \mathcal{A})$,
   $$\lim_{n\to+\infty} \tilde{I}(f_n) = \tilde{I}\left(\lim_{n\to+\infty} f_n\right).$$
   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
   $$\begin{aligned} \mathbb{E}_\mu(f) &= \mathbb{E}_\lambda(\rho f) \quad \text{d'après la proposition 3.5} \\ &= \int_{-\infty}^{+\infty} \rho(x) f(x) dx \quad \text{d'après la proposition 3.2} \\ &= \int_0^1 \alpha e^{-\alpha x} e^{\alpha x} dx = \alpha \ , \\ \text{et} \quad \mathbb{E}_\nu(f) &= e^{-\alpha} f(1) \quad \text{d'après la proposition 3.4} \\ &= e^{-\alpha} e^{\alpha \times 1} = 1 \ . \end{aligned}$$
   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
   $$\mathbb{E}_{\mu+\nu}(\mathbb{1}_\mathbb{R}) = \mathbb{E}_\mu(\mathbb{1}_\mathbb{R}) + \mathbb{E}_\nu(\mathbb{1}_\mathbb{R}) = \int_{-1}^1 e^{-\alpha x} dx + \sum_{k=0}^\infty \frac{\alpha^k}{k!} = \frac{e^\alpha - e^{-\alpha}}{\alpha} + e^\alpha > 1$$
   donc $\mu + \nu$ n'est pas une probabilité. x

Exercice 29 Enoncé p. 30

L'application borélienne f est intégrable suivant μ si, et seulement si, $\mathbb{E}_\mu(|f|) < +\infty$. Or d'après la proposition 3.4,

$$\mathbb{E}_\mu(|f|) = \sum_{i=0}^\infty \alpha_i |f|(a_i) = \sum_{i=0}^\infty \alpha_i |f(a_i)|.$$

Ainsi f sera intégrable par rapport à $\mu = \sum_{=0}^{\infty} \alpha_i \delta_i$ 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 $\sum_{i=0}^{\infty} \alpha_i f(a_i)$.

Exercice 30 Enoncé p. 30

Soit $x = (x_1, ..., x_d) \in \mathbb{R}^d$, $|x| = \sqrt{x_1^2 + ... + x_d^2}$. L'application f sera intégrable pour μ si, et seulement si, $\mathbb{E}_\mu(|f|) < \infty$, c-à-d $\mathbb{E}_\mu(\sqrt{f_1^2 + ... + f_d^2}) < \infty$. Or pour tout $i = 1, ..., d$ on a

$$|f_i| \leqslant \sqrt{f_1^2 + ... + f_d^2} \leqslant |f_1| + ... + |f_d| \;.$$

D'où par la croissance de l'opérateur $\mathbb{E}_\mu$ pour les fonctions positives (proposition 3.1, assertion 2),

$$\mathbb{E}_\mu(|f_i|) \leqslant \mathbb{E}_\mu(|f|) \leqslant \sum_{i=1}^{d} \mathbb{E}_\mu(|f_i|)$$

et par suite $\mathbb{E}_\mu(|f|) < +\infty \Leftrightarrow \forall i = 1, ..., d \;;\; \mathbb{E}_\mu(|f_i|) < +\infty$.

Exercice 31 Enoncé p. 31

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.

$$\Phi_\mu(t) = \int_{\mathbb{R}} e^{itx} \sum_{k=0}^{n} C_n^p p^k (1-p)^{n-k} d\delta_k(x) = \sum_{k=0}^{n} C_n^p (pe^{it})^k (1-p)^{n-k} \;.$$

   Ce qui donne avec la formule du binome de Newton : $\varphi_\mu(t) = (1 - p + pe^{it})^n$.
3. Cas où μ est la probabilité de Poisson de paramètres $\alpha > 0$.

$$\Phi_\mu(t) = \int_{\mathbb{R}} e^{itx} \sum_{k=0}^{+\infty} e^{-\alpha} \frac{(\alpha e^{it})^k}{k!} d\delta_k(x) = e^{\alpha(e^{it}-1)} \;.$$

Exercice 32 Enoncé p. 31

Comme $f \in \mathcal{M}^+$, il existe une suite $(f_n)_{n\geqslant 1}$ de fonctions de $\mathcal{E}^+$ telles que $f = \lim_n \uparrow f_n$ (la flèche vers le haut indique que pour tout $\omega \in \mathcal{E}$, $f_n(\omega) \to f(\omega)$ et la suite $(f_n(\omega))_{n\geqslant 1}$ est croissante). Si on suppose que $\mu(\{x \in E \;;\; f(x) = 0\}) = 1$, alors il en est de même pour tout les f_n. Or comme $f_n \in \mathcal{E}^+$, il existe des nombres positifs $(a_k)_{1\leqslant k\leqslant p}$ et des ensembles mesurables $(A_k)_{1\leqslant k\leqslant p}$ tels que $f_n = \sum_{k=1}^{p} \alpha_k \mathbb{1}_{A_k}$, on en déduit que $\alpha_k = 0$ pour tout k et donc $\int f_n d\mu = \sum_{k=1}^{p} \alpha_k \mu(A_k) = 0$. On en déduit que $\int f d\mu = 0$ par convergence monotone.

Exercice 33 Enoncé p. 31

1. On a $f = f^+ - f^-$ et $|f| = f^+ + f^-$. Ainsi

$$\begin{aligned}\left|\int f d\mu\right| &\leqslant \left|\int f^+ d\mu - \int f^- d\mu\right| \\ &\leqslant \left|\int f^+ d\mu\right| + \left|\int f^- d\mu\right| \\ &\leqslant \int f^+ d\mu + \int f^- d\mu \\ &\leqslant \int |f| d\mu \ .\end{aligned}$$

2. Il suffit d'écrire que

$$\int g d\mu = \int f d\mu + \int (g-f) d\mu \ .$$

Exercice 34 Enoncé p. 31

Nous allons appliquer le Théorème de convergence dominée. On pose pour $x \in \mathbb{R}$:

$$f_n(x) = \left(1 + \frac{x}{n}\right)^n e^{-2x} \mathbb{1}_{[0,n]}(x) \ .$$

On admet que les fonctions f_n sont mesurables. Pour tout $x \in \mathbb{R}$, on a $\lim_{n\to\infty} f_n(x) = e^{-x}\mathbb{1}_{[0,+\infty[}(x) := f(x)$ ainsi $(f_n)_{n\geqslant 1}$ converge simplement vers la fonction f. Il faut prendre garde lorsqu'on veut vérifier l'hypothèse (ii). En effet il faut trouver une fonction φ qui

NE DEPEND SURTOUT PAS DE n et telle que pour tout n : $f_n \leqslant \varphi$.

Pour cela on utilise que $\ln(1 + x/n) \leqslant x/n$ pour $x \geqslant 0$ et ainsi $f_n(x) \leqslant e^{-x}$ pour $x \geqslant 0$. Si on pose $\varphi(x) = e^{-x}\mathbb{1}_{\mathbb{R}^+}(x)$, on obtient bien que pour tout n, $f_n \leqslant \varphi$ et on a bien vérifier notre hypothèse de domination car φ est bien intégrable ($\int_{\mathbb{R}} |\varphi(x)|dx = 1$). On peut donc intervertir limite et intégrale pour obtenir

$$\lim_{n\to\infty} \int_{\mathbb{R}} f_n(x)dx = \int_{\mathbb{R}} \lim_{n\to\infty} f_n(x)dx = \int_{\mathbb{R}} e^{-x}\mathbb{1}_{[0,+\infty[}(x) = \int_0^{\infty} e^{-x}dx = 1 \ .$$

Exercice 35 Enoncé p. 31

Nous allons appliquer le théorème de convergence dominée. On pose pour $x \in \mathbb{R}$:

$$f_n(x) = \ln\left(e + \frac{|x|}{n}\right) f(x)\mathbb{1}_{[a,b]}(x) \ .$$

On admet que les fonctions f_n sont mesurables. Pour tout $x \in \mathbb{R}$, on a $\lim_{n\to\infty} f_n(x) = f(x)\mathbb{1}_{[a,b]}(x)$ ainsi $(f_n)_{n\geqslant 1}$ converge simplement vers la fonction $f\mathbb{1}_{[a,b]}$. Pour $x \in [a,b]$ nous avons la majoration

$$\begin{aligned}|f_n(x)| &\leqslant \ln\left(e + \frac{|x|}{n}\right)|f(x)| \\ &\leqslant \ln\left(e + \max(|a|,|b|)\right)|f(x)| \ .\end{aligned}$$

Ainsi $f_n \leqslant \varphi$ avec $\varphi(x) = \ln(e + \max(|a|, |b|))\, |f(x)| \mathbb{1}_{[a,b]}(x)$. La fonction φ est intégrable car f l'est. On peut donc intervertir limite et intégrale pour obtenir

$$\lim_{n\to\infty} \int_{\mathbb{R}} f_n(x) d\lambda(x) = \int_{\mathbb{R}} \lim_{n\to\infty} f_n(x) d\lambda(x) = \int_{\mathbb{R}} f(x) \mathbb{1}_{[a,b]}(x) d\lambda(x) = \int_a^b f(x) dx \ .$$

7.4 Solutions des exercices du chapitre 4

Exercice 36 Enoncé p. 40

1. On a $Var(X) = \mathbb{E}((X - \mathbb{E}(x)^2) = Cov(X, X)$. En posant $\mathbb{E}(X) = m$:

$$\begin{aligned} Var(X) &= \mathbb{E}(X - m)^2 = \mathbb{E}(X^2 - 2mX + m^2) = \mathbb{E}(X^2) - 2m\mathbb{E}(X) + m^2 \\ &\mathbb{E}(X^2) - 2m^2 + m^2 = \mathbb{E}(X^2) - m^2 = \mathbb{E}(X^2) - (\mathbb{E}(X))^2 \ . \end{aligned}$$

2. On pose $m_X = \mathbb{E}(X)$ et $m_Y = \mathbb{E}(Y)$ et on a :

$$\begin{aligned} Cov(X, Y) &= \mathbb{E}((X - m_X)(Y - m_Y)) = \mathbb{E}(XY) - m_X\mathbb{E}(Y) - m_Y\mathbb{E}(X) + m_X m_Y \\ &= \mathbb{E}(XY) - \mathbb{E}(X)m_Y - m_Y\mathbb{E}(X) + m_X m_Y = \mathbb{E}(XY) - m_X m_Y \\ &= \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y) \ . \end{aligned}$$

Exercice 37 Enoncé p. 40

On peut vérifier que ρ est bien une densité de probabilité sur $\mathbb{R}$. En particulier on a donc $\int_0^{+\infty} xe^{-x^2/2}dx = 1$. 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 :

$$\mathbb{E}(X^r) = \int_{\mathbb{R}} x^r d\mathbb{P}_X(x) = \int_{\mathbb{R}} x^{r+1} e^{-x^2/2} \mathbb{1}_{\mathbb{R}^+}(x) d\lambda(x) = \int_0^\infty x^{r+1} e^{-x^2/2} dx \ .$$

Par des techniques d'intégration par parties successives, on recherche une relation de récurrence sur l'entier $r \geqslant 2$:

$$\int_0^\infty x^{r+1} e^{-x^2/2} dx = r \int_0^\infty x^{r-1} e^{-x^2/2} dx$$

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 $Var(X) = 2 - \pi/2$.

Exercice 38 Enoncé p. 41

Supposons tout d'abord que pout 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 $X_i^2 \leqslant |X|^2$.

Exercice 39 Enoncé p. 41

Notons $X_1, ..., X_d$ les composantes de X dans la base canonique de $\mathbb{R}^d$, M la matrice $[a_{ij}]_{1\leqslant i\leqslant c, 1\leqslant j\leqslant d}$. Le vecteur MX de $\mathbb{R}^c$ a pour composante $((MX)_1, ..., (MX)_c)$ où pour $i = 1, ..., c$, $(MX)_i = \sum_{j=1}^d a_{ij}X_j$.
◇ 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

$$D_Y = \mathbb{E}\left((Y - \mathbb{E}(Y))(Y - \mathbb{E}(Y))^t\right) = \mathbb{E}\left((X - \mathbb{E}(X))(X - \mathbb{E}(X))^t\right) = D_X$$

d'où une première assertion du point 2) de la proposition 4.5.
◇ Le vecteur $\mathbb{E}(MX)$ a pour comosantes $\sum_{j=1}^d a_{ij}\mathbb{E}(X_j)$ pour $i = 1, ..., c$. 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 $\mathbb{E}[(X_i - \mathbb{E}(X_i))(X_j - \mathbb{E}(X_j))] = Cov(X_i,X_j)$, ce qui prouve aussi que D_X est symétrique car $Cov(X_i,X_j) = Cov(x_j,X_i)$. Par ailleur, les éléments diagonaux de D_X sont $Cov(X_i,X_i) = Var(X_i)$ pour $i = 1, ..., d$.
◇ Montrons que $D_{MX} = MD_XM^t$. Soit $i, j = 1, ..., c$, le coefficient (i, j) de D_{MX} est

$$\begin{aligned}
Cov\big((MX)_i,(MX)_j\big) &= \mathbb{E}[(MX)_i(MX)_j] - \mathbb{E}[(MX)_i]\mathbb{E}[(MX)_j] \\
&= \mathbb{E}\left[\sum_{k=1}^d a_{ik}X_k \sum_{l=1}^d a_{jl}X_l\right] - \mathbb{E}\left[\sum_{k=1}^d a_{ik}X_k\right]\mathbb{E}\left[\sum_{l=1}^d a_{jl}X_l\right] \\
&= \sum_{k=1}^d\sum_{l=1}^d a_{ik}a_{jl}\Big[\underbrace{\mathbb{E}(X_kX_l) - \mathbb{E}(X_k)\mathbb{E}(X_l)}_{=Cov(X_k,X_l)}\Big] \\
&= \sum_{k=1}^d\sum_{l=1}^d a_{ik}Cov(X_k,X_l)a_{jl} \\
&= \text{ coefficient } (i, j) \text{ de la matrice } MD_XM^t \text{ (matrice carrée d'ordre } c)
\end{aligned}$$

d'où l'égalité cherchée.
◇ Montrons que d_x est positive. Soit $U = (u_1, ..., u_d)^t \in \mathbb{R}^d$ (U est aussi une matrice colonne). On a $D_{U^tX} = U^tD_X(U^t)^t = U^tD_XU$. Or $U^tX = \sum_{j=1}^d u_jX_j$ et D_{U^tX} représente la variance de cette variable aléatoire. Ainsi $U^tD_XU \geqslant 0$ 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.

Exercice 40 Enoncé p. 41

1. Les équivalences suivantes donnent :

$$\big(|X| - |Y|\big)^2 \geqslant 0 \Leftrightarrow X^2 + Y^2 - 2|XY| \geqslant 0 \Leftrightarrow X^2 + Y^2 \geqslant 2|XY| \geqslant |XY|\ .$$

On en déduit que $\mathbb{E}(|XY|) \leqslant \mathbb{E}(X^2)+\mathbb{E}(Y^2) < \infty$, ainsi la v.a.r. XY est intégrable. En choisissant $Y = \mathbb{1}_\Omega$ (c'est la variable aléatoire constante égale à 1), on a $|XY| = |X| \leqslant X^2 + \mathbb{1}_\Omega$ d'où

$$\mathbb{E}(|X|) \leqslant \mathbb{E}(X^2) + \mathbb{E}(\mathbb{1}_\Omega) = \mathbb{E}(X^2) + 1 < \infty \ .$$

On procède de même pour Y.

2. Pour tout $\alpha \in \mathbb{R}$,

$$0 \leqslant \mathbb{E}((X+\alpha Y)^2) = \alpha^2\mathbb{E}(Y^2) + 2\alpha\mathbb{E}(XY) + \mathbb{E}(X^2) \ .$$

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 $\big(\mathbb{E}(XY)\big)^2 - \mathbb{E}(Y^2)\mathbb{E}(X^2) \leqslant 0$ d'où le résultat.

Exercice 41 Enoncé p. 41

1. Pour tout $i = 1, ..., n$, X_i est de carrè intégrable car $\mathbb{E}(|X|^2) < \infty$. On a

$$\begin{aligned} Y^2 &= \sum_{i=1}^n X_i^2 + \sum_{1\leqslant i\neq j\leqslant n} X_i Xj \\ &\leqslant \sum_{i=1}^n X_i^2 + \sum_{1\leqslant i\neq j\leqslant n} |X_i Xj| \quad \text{ce qui entraîne} \\ \mathbb{E}(Y^2) &\leqslant \underbrace{\sum_{i=1}^n \mathbb{E}(X_i^2)}_{<\infty \text{ car } X_i \text{ intégrables}} + \underbrace{\sum_{1\leqslant i\neq j\leqslant n} \mathbb{E}(|X_i Xj|)}_{<\infty \text{ d'après l'exercice 40, page 41}} \ , \end{aligned}$$

d'où $\mathbb{E}(Y^2) < \infty$ ce qui prouve que $\sum_{k=1}^n X_i$ est de carré intégrable.

2. On a :

$$\begin{aligned} Var(Y) = \mathbb{E}[(Y-\mathbb{E}(Y))^2] &= \mathbb{E}\left[\left(\sum_{i=1}^n (X_i - \mathbb{E}(X_i))\right)^2\right] \\ &= \sum_{i=1}^n \underbrace{\mathbb{E}[(X_i-\mathbb{E}(X_i))^2]}_{=Var(X_i)} + \sum_{1\leqslant i\neq j\leqslant n} \underbrace{\mathbb{E}[(X_i-\mathbb{E}(X_i))(X_j-\mathbb{E}(X_j))]}_{=Cov(X_i,X_j)} \ , \end{aligned}$$

d'où le résultat en remarquant que

$$\sum_{1\leqslant i\neq j\leqslant n} \mathbb{E}[(X_i-\mathbb{E}(X_i))(X_j-\mathbb{E}(X_j))] = 2\sum_{1\leqslant i<j\leqslant n} \mathbb{E}[(X_i-\mathbb{E}(X_i))(X_j-\mathbb{E}(X_j))]$$

car $Cov(X_i, X_j) = Cov(X_j, X_i)$.

Exercice 42 Enoncé p. 41

1. Méthode 1) : on applique le théorème de transfert 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 2, exercice 23, 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

$$\begin{aligned}\mathbb{E}(e^X) &= \int_{\mathbb{R}} y\rho(y) d\lambda(y) = \int_{-\infty}^{+\infty} \frac{y}{y\sqrt{2\pi}} \exp\left(-\frac{1}{2}(\ln y)^2\right) \mathbb{1}_{]0,+\infty[}(y) \\ &= \frac{1}{\sqrt{2\pi}} \int_0^{\infty} e^{-\frac{1}{2}(\ln y)^2} dy = \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-u^2/2} e^u du\end{aligned}$$

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 $\mathbb{R}$.

Exercice 43 Enoncé p. 41

• 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 $\mathbb{R}^+ \times \mathbb{R}^+ \to \mathbb{R}^+$ définie par $h(x,y) = x$. D'où

$$\mathbb{E}(X_1) = \int_{\mathbb{R}^2} h(x,y)d\left(\textstyle\sum_{k,l=1}^{\infty} \frac{1}{2^{k+l}}\delta_{(k,l)}\right)(x,y) = \sum_{k=1}^{\infty}\sum_{l=1}^{\infty} \frac{1}{2^{k+l}}\underbrace{h(k,l)}_{=k} = \sum_{k=1}^{\infty}\sum_{l=1}^{\infty} \frac{k}{2^{k+l}} .$$

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 $\mathbb{E}(X_1) = \frac{1/2}{1-1/2} = 2$. 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'indentification des lois par les fonctions boréliennes positives. Soit h de $\mathbb{R} \to [0,\infty]$ 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,

$$\begin{aligned}\mathbb{E}(h(Z)) &= \int_{\mathbb{R}^2} \varphi(x,y)d\mathbb{P}_X(x,y) = \int_{\mathbb{R}^2} h(x+y)d\left(\textstyle\sum_{k,l=1}^{\infty} \frac{1}{2^{k+l}}\delta_{(k,l)}\right)(x,y)\\ &= \sum_{k=1}^{\infty}\sum_{l=1}^{\infty} \frac{1}{2^{k+l}}h(k+l) = \sum_{i=2}^{\infty}\sum_{(k,l)/k+l=i} \frac{1}{2^i}h(i) .\end{aligned}$$

Or il existe $i-1$ couples (k,l) tels que $k+l = i$ avec $1 \leqslant k$ et $1 \leqslant l$ 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 $i = 1$.

Exercice 44 Enoncé p. 41

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 $x \leqslant -e$, $F(x) = 0$.
   - ◇ 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 $-1 \leqslant x < 0$, $F(x) = \int_{-e}^{-1} -1/(4t)dt + \int_{-1}^{x}(t+5/4)dt = x^2/2 + 5x/4 + 1$.
   - ◇ Si $x \geqslant 0$, $F(x) = \int_{-\infty}^{x} f(t)dt = \int_{-e}^{0} f(t)dt = 1$.

D'où F est définie par

$$F(x)=\begin{cases}0 & \text{si } x\leqslant -e\,;\\ -(1/4)\ln(-x)+1/4, & \text{si } -e<x\leqslant -1\,;\\ x^2/2+5x/4+1, & \text{si } -1<x\leqslant 0\,;\\ 1, & \text{si } x\geqslant 0.\end{cases}$$

2. Pour calculer $\mathbb{E}(X)$ on écrit que

$$\mathbb{E}(X)=\int_{-\infty}^{+\infty}xf(x)dx=\int_{-e}^{-1}-\frac{x}{4x}dx+\int_{-1}^{0}(x^2+5x/4)dx=-(6e+1)/24\ .$$

Exercice 45 Enoncé p. 42

1. Écrivons la fonction de répartition de $Y=X^3$. Soit $y\in\mathbb{R}$,

$$\begin{aligned}F_Y(y)=\mathbb{P}(Y\leqslant y)=\mathbb{P}(X^3\leqslant y)=\mathbb{P}(X\leqslant y^{1/3})&=\int_{-\infty}^{y^{1/3}}\frac{1}{\sqrt{2\pi}}e^{-t^2/2}dt\\ =\int_{-\infty}^{y}\underbrace{\frac{1}{\sqrt{2\pi}}e^{-z^{2/3}/2}\frac{1}{3}z^{-2/3}}_{:=f(z)}dz\ ,\end{aligned}$$

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 $y\in]0,1[$

$$\mathbb{P}(Y\leqslant y)=\mathbb{P}(F(X)\leqslant y)=\mathbb{P}(X\leqslant F^{-1}(y))=F(F^{-1}(y))=y\ .$$

Comme F prend ses valeurs dans l'intervalle $]0,1[$, l'ensemble $\{F(X)\leqslant y\}=\emptyset$ dès que $y\leqslant 0$ et donc $F_Y(y)=\mathbb{P}(F(X)\leqslant y)=0$ quand $y\leqslant 0$.
Pour la même raison, $\{F(X)\leqslant y\}=\Omega$ dès que $y\geqslant 1$ et dans ce cas $F_Y(y)=\mathbb{P}(\Omega)=1$. En conclusion on a

$$F_Y(y)=\begin{cases}0, & \text{si } y\leqslant 0\,;\\ y, & \text{si } 0<y<1\,;\\ 1, & \text{si } y\geqslant 1.\end{cases}$$

donc Y est une v.a.r. uniforme sur $[0,1]$.

Exercice 46 Enoncé p. 42

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 45. Soit $y \in \mathbb{R}^+$,

$$F_Y(y) = \mathbb{P}(Y \leqslant y) = \mathbb{P}(|X| \leqslant y) = \mathbb{P}(-y \leqslant X \leqslant y) = \int_{-y}^{y} f(t)dt$$
$$= \int_0^y \alpha e^{-\alpha t} dt = 1 - e^{-\alpha y} \ .$$

Si maintenant $y < 0$, $\{T \leqslant y\} = \{|X| \leqslant y\} = \emptyset$ 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, on sait que $\mathbb{E}(Y) = 1/\alpha$ et $Var(Y) = 1/\alpha^2$. Par la formule de König, $\mathbb{E}(Y^2) = Var(Y) + (\mathbb{E}(Y))^2 = 2/\alpha^2$. 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 $Var(x) = \mathbb{E}(X^2) = 2/\alpha^2$.

Exercice 47 Enoncé p. 42

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 $y >^0$, $F_Y(y) = \mathbb{P}(Y \leqslant y) = \mathbb{P}(X \leqslant 1 - e^{-\alpha y}) = 1 - e^{-\alpha y}$ car $0 \leqslant 1 - e^{-\alpha y} \leqslant 1$ et x suit une loi uniforme sur [0,1]. Comme pour $y \leqslant 0$, $\mathbb{P}(Y \leqslant y) = 0$ on reconnait que Y suit la loi exponentielle de paramètre $\alpha > 0$;

7.5 Solutions des exercices du chapitre 5

Exercice 48 Enoncé p. 59

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 24 on a

$$\frac{\pi}{4} = \int_{\mathbb{R}^2} f(x,y)d\lambda^{(2)}(x,y) = \int_{\mathbb{R}} g(y)\left(\int_{\mathbb{R}} h(x)d\lambda(x)\right)d\lambda(y)$$
$$= \left[\int_{\mathbb{R}} g(y)d\lambda(y)\right] \times \left[\int_{\mathbb{R}} h(x)d\lambda(x)\right]$$
$$= \left[\int_0^\infty e^{-y^2}dy\right] \times \left[\int_0^\infty e^{-x^2}dx\right].$$

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 $T : (x,y) \in \mathbb{R}^2 \to T(x,y) = (u,v) \in \mathbb{R}^2$ 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 :

$$\begin{aligned}\int_{\mathbb{R}^2} e^{-(x+y)^2+y^2} d\lambda^{(2)}(x,y) &= \int_{\mathbb{R}^2} e^{-(u^2+v^2)} d\lambda^{(2)}(u,v)\\ &= \left[\int_{\mathbb{R}} e^{-u^2} d\lambda(u)\right] \times \left[\int_{\mathbb{R}} e^{-v^2} d\lambda(v)\right]\\ &= \left(\int_{-\infty}^{+\infty} e^{-x^2} dx\right)^2 = \left(2\int_2^{+\infty} e^{-x^2} dx\right)^2 = 4\left(\frac{\sqrt{\pi}}{2}\right)^2 = \pi \ ,\end{aligned}$$

où l'on a utilisé le théorème de Tonelli et le résultat de la première question.

Exercice 49 Enoncé p. 59

Vérifions qur 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 le fonction nulle sur $\mathbb{R}^d$ donc $\nu(\emptyset) = \int_{\mathbb{R}^d} \mathbb{1}_\emptyset \rho d\lambda^{(d)} = 0$.
- Par définitionde 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 $\sigma-$additivité. Soit $(A_k)_{k\in\mathbb{N}}$ une suite d'éléments de $\mathcal{B}$ deux à deux disjoints. Alors si on note $A = \cup_{k=1}^\infty A_k$, $\mathbb{1}_A = \sum_{k=0}^\infty \mathbb{1}_{A_k}$ car les A_k sont deux à deux disjoints. On a alors en utilisant le théorème de Beppo-Levi

$$\nu(A) = \int_{\mathbb{R}^d} \mathbb{1}_A \rho d\lambda^{(d)} = \int_{\mathbb{R}^d} \rho \sum_{k=0}^\infty d\lambda^{(d)} = \sum_{k=0}^\infty \int_{\mathbb{R}^d} \mathbb{1}_{A_k}\rho d\lambda^{(d)} = \sum_{k=0}^\infty \nu(A_k) \ ,$$

ce qui donne la $\sigma-$additivité.

On a donc ν qui est bien une mesure de probabilité.

Exercice 50 Enoncé p. 59

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 $k = 1,...,6$, $\mathbb{P}(X = k) = \mathbb{P}(Y = k) = 1/6$, 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.

Exercice 51 Enoncé p. 59

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

$$\begin{aligned}\mathbb{P}_Y(A) = \mathbb{P}(Y \in A) &= \mathbb{P}\big((X,Y,Z) \in \mathbb{R} \times A \times \mathbb{R}\big) = \mathbb{P}_{(X,Y,Z)}(\mathbb{R} \times A \times \mathbb{R})\\ &= \int_{\mathbb{R}^3} \mathbb{1}_{\mathbb{R}}(x)\mathbb{1}_A(y)\mathbb{1}_{\mathbb{R}}(z) d\mathbb{P}_{(X,Y,Z)}(x,y,z)\end{aligned}$$

et par application des régles d'intégration des mesures à densité (proposition 5.4) et du thèorème de Tonelli, on obtient comme dans l'exemple 25 que

$$\mathbb{P}_Y(A) = \int_{\mathbb{R}} 1\!\!1_A(y) \underbrace{\left(\int_{\mathbb{R}^2} 1\!\!1_{\mathbb{R}}(x) 1\!\!1_{\mathbb{R}}(z) \rho(x,y,z) d\lambda^{(2)}(x,z)\right)}_{:=\chi(y)} d\lambda(x) = \int_{\mathbb{R}} 1\!\!1_A(y)\chi(y)d\lambda(y)$$

ce qui prouve que la v.a.r. Y admet pour densité la fonction χ définie ci-dessus.

$\boxed{\textbf{Exercice 52}}$ Enoncé p. 59

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 :

$$\begin{aligned}\mathbb{P}((X,X) \in \Delta) = \mathbb{P}_{(X,X)}(\Delta) &= \int_{\mathbb{R}^2} 1\!\!1_\Delta(x,y)\rho(x,y)d\lambda^{(2)}(x,y)\\ &= \int_{\mathbb{R}^2} 1\!\!1_\Delta(x,y)\rho(x,x)d\lambda^{(2)}(x,y)\\ &= \int_{\mathbb{R}} \rho(x,x)\left(\int_{\mathbb{R}} 1\!\!1_\Delta(x,y)d\lambda(y)\right)d\lambda(x)\ .\end{aligned}$$

Or $\int_{\mathbb{R}} 1\!\!1_\Delta(x,y)d\lambda(y) = \int_{\mathbb{R}} 1\!\!1_{\{x\}}(y)d\lambda(y) = \lambda(\{x\}) = 0$ 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 à la proposition 5.5.

$\boxed{\textbf{Exercice 53}}$ Enoncé p. 59

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 :

$$\begin{aligned}\mathbb{E}(h(X,Y)) &= \mathbb{E}\left(h\big(\sqrt{-2\ln U}\cos(2\pi V), \sqrt{-2\ln U}\sin(2\pi V)\big)\right)\\ &= \int_{\mathbb{R}^2} h\big(\sqrt{-2\ln u}\cos(2\pi v), \sqrt{-2\ln u}\sin(2\pi v)\big)d\mathbb{P}_{(U,V)}(u,v)\\ &= \int_{\mathbb{R}^2} h\big(\sqrt{-2\ln u}\cos(2\pi v), \sqrt{-2\ln u}\sin(2\pi v)\big)d\mathbb{P}_U \otimes d\mathbb{P}_V(u,v)\\ &= \int_{\mathbb{R}^2} h\big(\sqrt{-2\ln u}\cos(2\pi v), \sqrt{-2\ln u}\sin(2\pi v)\big)1\!\!1_{]0,1[}(u)1\!\!1_{]0,1[}(v)d\lambda^{(2)}(u,v)\ .\end{aligned}$$

Effectuons le changement de variable $T\ :]0,1[^2 \to \mathbb{R}^2$ défini pour $(u,v) \in \mathbb{R}^2$ par

$$\left\{\begin{array}{l} x = \sqrt{-2\ln u}\cos(2\pi v)\\ y = \sqrt{-2\ln u}\sin(2\pi v)\end{array}\right. \iff \left\{\begin{array}{l} u = \exp(-(x^2+y^2)/2)\\ v = \frac{1}{2\pi}\arctan(y/x)\end{array}\right. .$$

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 :

$$\mathbb{E}\big(h(X,Y)\big) = \int_{\mathbb{R}^2} h(x,y)\frac{1}{2\pi}e^{-\frac{x^2+y^2}{2}}d\lambda^{(2)}(x,y) \ ,$$

et donc la v.a.r. $(X, Y$ admet bien la densité ρ annoncée.

2. Pour trouver la loi de la v.a.r. X (idem pour Y), on applique la proposition 5.5 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 $\mathbb{R}$

$$\begin{aligned}\mathbb{P}_{(X,Y)}(A\times B) &= \int_{\mathbb{R}^2} \mathbb{1}_A(x)\mathbb{1}_B(y)\frac{1}{2\pi}e^{-\frac{x^2+y^2}{2}}d\lambda^{(2)}(x,y) \\ &= \left(\int_{\mathbb{R}} \mathbb{1}_A(x)\frac{1}{\sqrt{2\pi}}e^{-x^2/2}d\lambda(x)\right) \times \left(\int_{\mathbb{R}} \mathbb{1}_B(y)\frac{1}{\sqrt{2\pi}}e^{-y^2/2}d\lambda(y)\right) \\ &= \mathbb{P}_X(A)\times\mathbb{P}_Y(B)\end{aligned}$$

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 indépendnat.

Exercice 54 Enoncé p. 60

Étudions la fonction de répartition de la v.a.r. Y. Pour tout $y \in \mathbb{R}$, $F_Y(y) = \mathbb{P}(Y \leqslant y)$. On peut exprimer l'événement $\{Y \leqslant y\}$ en fonction des v.a.r. ε et X :

$$\{Y \leqslant y\} = [\{\varepsilon = 1\}\cap\{X \leqslant y\}]\cup[\{\varepsilon = -1\}\cap\{X \geqslant -y\}]$$

Comme (ε, X) indépendants, les événements $\{\varepsilon = 1\}$ et $\{X \leqslant y\}$ sont indépendants (il en est de même pour $\{\varepsilon = -1\}$ et $\{X \geqslant -1\}$). Les événements entre crochets étant disjoints, on obtient :

$$\begin{aligned}F_Y(y) &= \mathbb{P}(\varepsilon = 1)\times\mathbb{P}(X \leqslant y) + \mathbb{P}(\varepsilon = -1)\times\mathbb{P}(X \geqslant -y) \\ &= \frac{1}{2}\int_{-\infty}^{y}\frac{1}{\sqrt{2\pi}}e^{-x^2/2}dx + \frac{1}{2}\int_{-y}^{+\infty}\frac{1}{\sqrt{2\pi}}e^{-x^2/2}dx = \int_{-\infty}^{y}\frac{1}{\sqrt{2\pi}}e^{-x^2/2}dx\end{aligned}$$

et donc Y est une v.a.r. de loi $\mathcal{N}(0;1)$.
On remarque que $\mathbb{E}(XY) = \mathbb{E}(\varepsilon X^2) = \mathbb{E}(\varepsilon)\times\mathbb{E}(X^2)$ car ε et X sont indépendantes. Ainsi $\mathbb{E}(XY) = 0\times 1 = 0$.
En remarquant que $\varepsilon^2 = 1$ on obtient $\mathbb{E}(X^2Y^2) = \mathbb{E}(\varepsilon^2X^4) = \mathbb{E}(X^4) = 3$ alors que $\mathbb{E}(X^2)\times\mathbb{E}(Y^2) = \mathbb{E}(X^2)\times\mathbb{E}(X^2) = 1\times 1 = 1 \neq 3$ et donc X et Y ne sont pas indépendantes.

Exercice 55 Enoncé p. 60

D'après les formules de König-Huygens (Proposition 4.2), $Var(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2$. D'où $\mathbb{E}(X^2) = Var(X) + (\mathbb{E}(X))^2 = \sigma^2 + m^2$ (idem pour Y). Par indépendance de (X, Y), $\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y)$ et donc on a

$$\begin{aligned}\mathbb{E}((X+Y)^2) &= \mathbb{E}(X^2 + 2XY + Y^2) = \mathbb{E}(X^2) + 2\mathbb{E}(XY) + \mathbb{E}(Y^2)\\ &== \mathbb{E}(X^2) + 2\mathbb{E}(X)\mathbb{E}(Y) + \mathbb{E}(Y^2) = 4m^2 + 2\sigma^2\ .\end{aligned}$$

Exercice 56 Enoncé p. 60

On reprend les notations de l'exercice 54, page 60. On a

$$\begin{aligned}Var(X+Y) &= Var(X) + 2\underbrace{Cov(X,Y)}_{=0} + Var(Y)\\ &= Var(X) + Var(Y)\end{aligned}$$

bien que (X, Y) ne soit pas indépendant.

Exercice 57 Enoncé p. 60

- 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) $A = [-1/2, 1/2]$ et on calcule

$$\mathbb{P}_X(A) = \int_{-1/2}^{1/2} \frac{1}{\pi}\frac{1}{1+x^2}dx = \frac{1}{\pi}\left[\arctan(x)\right]_{-1/2}^{1/2} = \frac{1}{\pi}\left(\frac{\pi}{4} - \left(\frac{\pi}{4}\right)\right) = \frac{1}{2}\ .$$

  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.

Exercice 58 Enoncé p. 60

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 $X_1, ..., X_n$ on a

$$\Phi_{S_n}(t) = \Phi_{X_1}(t) \times ... \times \Phi_{X_n}(t) = \left[pe^{it} + (1-p)\right]^n$$

et on reconnait la fonction caractéristique d'une loi $\mathcal{B}(n; p)$. Donc S_n est une v.a.r. $\mathcal{B}(n; p)$.

Exercice 59 Enoncé p. 60

1. Pour tout $\omega \in \Omega$, $X_{(1)}(\omega) = \max(X_1(\omega), ..., X_n(\omega))$ car c'est le premier nombre dans la suite $X_1(\omega), ..., X_n(\omega)$ réordonnés de façon décroissante. Soit $x \in \mathbb{R}$, $F_{X_{(1)}}(x) = \mathbb{P}(X_{(1)} \leqslant x)$ or $\{\max(X_1, ..., X_n) \leqslant x\} = \cap_{k=1}^n \{X_k \leqslant x\}$ et par indépendance de $X_1, ..., X_n$ on déduit que

$$F_{x_{(1)}}(x) = \mathbb{P}\left(\cap_{k=1}^n \{X_k \leqslant x\}\right) = \prod_{k=1}^n \mathbb{P}(X_k \leqslant x) = \left(F(x)\right)^n$$

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 $X_{(n)} = \min(X_1, ..., X_n)$. Soit $x \in \mathbb{R}$, $\{X_{(n)} \leqslant x\} = \{X_{(n)} > x\}^c = \{\cap_{k=1}^n \{X_k > x\}\}^c$. Par suite

$$\begin{aligned} F_{X_{(n)}}(x) &= 1 - \mathbb{P}(X_{(n)} > x) = 1 - \mathbb{P}\left(\cap_{k=1}^n \{X_k > x\}\right) = 1 - \prod_{k=1}^n \mathbb{P}(X_k > x) \\ &= 1 - \prod_{k=1}^n (1 - \mathbb{P}(X_k \leqslant x)) = 1 - \prod_{k=1}^n [1 - F(x)] = 1 - [1 - F(x)]^n \\ &= (1 - e^{-n\alpha x})\mathbb{1}_{]0,\infty[}(x) \ . \end{aligned}$$

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

$$p = \mathbb{P}(Y_k = 1) = \mathbb{P}(X_k > t) = 1 - \mathbb{P}(X_k \leqslant t) = 1 - F(t) = 1 - (1 - e^{-\alpha t})$$

d'où $p = e^{-\alpha t}$ et donc $\mathbb{P}_{Y_k} = e^{-\alpha t}\delta_1 + (1 - e^{-\alpha t})\delta_0$, pour $k = 1, ..., n$. Par suite $Y_1 + ... + Y_n$ 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 $f_1(X_1), ..., f_n(X_n)$ avec $f_k = \mathbb{1}_{]t,\infty[}$ (on utilise la proposition 5.14). La v.a.r. $Y_1 + ... + Y_n$ est donc une v.a.r. $\mathcal{B}(n; p)$.

(b) On remarque que $Y_k = 1$ signifie que X_k est strictement supérieur à t. Ainsi $Y_1 + ... + Y_k$ est égal au nombre de v.a.r. X_i qui sont strictement supérieure à t. Ceci entraîne que $\{Y_1 + ... + Y_n \leqslant k - 1\} = \{X_{(k)} \leqslant t\}$.

4. Soit $t \in \mathbb{R}$, d'après ce qui précède on a :

$$\begin{aligned} F_{X_{(k)}}(t) &= \mathbb{P}(X_{(k)} \leqslant t) = \mathbb{P}(Y_1 + ... + Y_n \leqslant k - 1) = \sum_{j=0}^{k-1} C_n^j p^j (1-p)^{n-j} \\ &= \sum_{j=0}^{k-1} C_n^j e^{-j\alpha t}(1 - e^{-\alpha t})^{n-j} \ . \end{aligned}$$

Exercice 60 Enoncé p. 61

On fait la preuve du deuxième point (l'autre preuve est identique). 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 $\alpha + \beta$.

Exercice 61 Enoncé p. 61

Sans restreindre la généralité, on peut supposer que $n \leqslant m$. 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

$$\begin{aligned}
\mathbb{P}(X=Y) &= \sum_{k=0}^{n} \mathbb{P}\big(\{X=k\}\cap\{Y=k\}\big) \\
&= \sum_{k=0}^{n} \mathbb{P}(X=k)\times\mathbb{P}(Y=k) \qquad \text{par indépendance de } X \text{ et } Y \\
&= \sum_{k=0}^{n} C_n^k \left(\frac{1}{2}\right)^n C_m^k \left(\frac{1}{2}\right)^m \\
&= \left(\frac{1}{2}\right)^{n+m} \left(\sum_{k=0}^{n} C_n^k C_m^k\right) = \left(\frac{1}{2}\right)^{n+m} \left(\sum_{k=0}^{n} C_n^k C_m^{m-k}\right)
\end{aligned}$$

et en utilisant l'indication de l'énoncé, on obtient $\mathbb{P}(X=Y)=\frac{1}{2^{n+m}}C_{n+m}^m$.

Exercice 62 Enonce p. 61

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 $\{n\in\mathbb{N}^*/X_1(\omega)+X_2(\omega)+\ldots+X_n(\omega)=r\}$ 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. $(X_i)_{i\in\mathbb{N}}$,

$$\begin{aligned}
\mathbb{P}\{\tau_r=k\} &= \mathbb{P}\left(\left\{\sum_{i=1}^{k-1} X_i = r-1\right\} \cap \{X_k=1\}\right) \\
&= \mathbb{P}\left(\left\{\sum_{i=1}^{k-1} X_i = r-1\right\}\right) \mathbb{P}(\{X_k=1\}).
\end{aligned}$$

   La v.a.r. $\sum_{i=1}^{k-1} X_i$ 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 v.a.r. 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. $(X_k)_{k \in \mathbb{N}}$ représente les valeurs successives obtenues dans un jeu de Pile-ou-Face, en notant $X_k = 1$ quand on a obtenu "Face" lors du $\mathrm{k}^{\text{ième}}$-lancer et par conséquent $X_k = 0$ quand on a obtenu "Pile" lors du $\mathrm{k}^{\text{ième}}$-lancer.
La v.a.r. $X_1 + X_2 + ... + X_n$ 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 $\mathrm{r}^{\text{ième}}$-"succès" et la v.a.r. θ_r représente le nombre d'"échecs" avant le $\mathrm{r}^{\text{ième}}$-"succès" dans une infinité de lancers d'une pièce de monnaie.
Si on prend $r = 1$, la v.a.r. 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 v.a.r. géométrique de paramètre p. Une v.a.r. géométrique de paramètre p est une v.a.r. 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 v.a.r. définie, pour tout $\omega \in \Omega$, par $X_n(\omega) = 0$ si, au $\mathrm{n}^{\text{ième}}$ coup, on cherche l'allumette dans la poche droite et $X_n(\omega) = 1$ si, au $\mathrm{n}^{\text{ième}}$ 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 v.a.r. X_n suit une loi de Bernoulli de paramètre $\frac{1}{2}$. La v.a.r. $S_n := X_1 + X_2 + ... + X_n$ représente le nombre de fois où le fumeur a puisé dans sa poche gauche. On considère que le fumeur s'aperçevra que la boîte gauche est vide lorsqu'il cherchera une allumette dans cette poche pour la $(\text{N+1})^{\text{ième}}$-fois. Cela signifie que l'événement G_k est réalisé lorsqu'on extrait $N - k$ allumettes de la boîte droite avant la $(\text{N+1})^{\text{ième}}$-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

$$\mathbb{P}(A_k) = 2\mathbb{P}(G_k) = 2C_N^{2N-k}(\frac{1}{2})^{2N-k+1} = C_N^{2N-k}(\frac{1}{2})^{2N-k}.$$

7.6 Solutions des exercices du chapitre 6

Exercice 63 Enoncé p. 74

1. D'après l'inégalité de Bienaymé-Tchébicheff, $\mathbb{P}(|X - m| \geqslant 3\sigma) \leqslant \frac{\sigma^2}{9\sigma^2} = \frac{1}{9}$ et donc

$$\mathbb{P}(m - 3\sigma < X < m + 3\sigma) = 1 - \mathbb{P}(|X - m| \geqslant \sigma) \geqslant \frac{8}{9} \simeq 0.88 \ .$$

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 l'annexe B) 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 $\mathbb{P}(m - 2\sigma < X < m + 2\sigma) \simeq 0.9544$.

Exercice 64 Enoncé p. 74

1. En utilisant le fait que X est d'espérance nulle on a :

$$\begin{aligned} a \ &= \ \mathbb{E}[a - X] \ = \ \mathbb{E}[(a - X)\mathbb{1}_{(X \leq a)} + (a - X)\mathbb{1}_{(X > a)}] \\ &\leq \ \mathbb{E}[(a - X)\mathbb{1}_{(X \leq a)}] \leq (\mathbb{P}(X \leq a))^{\frac{1}{2}}(\mathbb{E}[(a - X)^2])^{\frac{1}{2}} \end{aligned}$$

où la dernière inégalité est obtenue en utilisant l'inégalité de Cauchy-Schwarz (voir exercice 40 page 41). De plus $\mathbb{E}[(a - X)^2] = a^2 - 2a\mathbb{E}[X] + \mathbb{E}[X^2] = a^2 + \sigma^2$ ce qui donne :

$$a \leq \mathbb{E}[(a - X)\mathbb{1}_{(X \leq a)}] \leq (\mathbb{P}(X \leq a))^{\frac{1}{2}}\sqrt{a^2 + \sigma^2}.$$

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

$$\mathrm{Var}(X) = \mathbb{E}[(Y-100)^2] = \mathbb{E}[(Y-\mathbb{E}[Y])^2] = \mathrm{Var}(Y) = 400.$$

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.

Exercice 65 Enoncé p. 74

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. $X_1, ..., X_N$ sont indépendantes et de même loi. On note $\overline{X}_N = \frac{1}{N}\sum_{k=1}^{N} X_k$ la moyenne empirique des valeurs observées. Par l'inégalité de Bienaymé-Tchébicheff, $\mathbb{P}(|\overline{X}_N - m| \geqslant 0.05) \leqslant \frac{Var(\overline{X}_N)}{(0.05)^2}$. Or

$$Var(\overline{X}_N) = \frac{Var(X_1)}{N} = \frac{0.25}{N}$$

et donc $\mathbb{P}(|\overline{X}_N - m| \geqslant 0.05) \leqslant \frac{0.25}{N(0.05)^2}$. On cherche donc N tel que $\frac{0.25}{N(0.05)^2} \leqslant 0.01$ d'où $N \geqslant \frac{0.25}{(0.05)^2 \times 0.01} = 10000$.

Exercice 66 Enoncé p. 74

Soit $a > 0$, prenons n suffisamment grand pour que $0 < 1/n < a$. On veut étudier la limite quand $n \to \infty$ de $\mathbb{P}(|X_n - 0| \geqslant a)$. Pour n tel que $1/n < a$, on a

$$\begin{aligned}\mathbb{P}(|X_n - 0| \geqslant a) &= \mathbb{P}(|X_n| \geqslant a) = \mathbb{P}(X_n \leqslant -a) + \mathbb{P}(X_n \geqslant a)\\ &= \int_{-\infty}^{-a} n \underbrace{\mathbb{1}_{[0,\frac{1}{n}]}(x)}_{=0 \text{ sur }]-\infty,-a[} dx + \int_{a}^{+\infty} n \underbrace{\mathbb{1}_{[0,\frac{1}{n}]}(x)}_{=0 \text{ sur } [a,+\infty[} dx\end{aligned}$$

d'où pour tout $a > 0$, $\lim_{n\to\infty} \mathbb{P}(|X_n - 0| \geqslant a) = 0$ ce qui prouve que $(X_n)_{n\in\mathbb{N}}$ converge en probabilité vers 0.

Exercice 67 Enoncé p. 74

On considère l'exemple 32 du cours. La suite $(X_n)_{n\in\mathbb{N}}$ converge en probabilité vers le v.a.r.

0. Mais pour tout entier $n \geqslant 1$, $\mathbb{E}(X_n) = \frac{n}{n+1} + \frac{1}{n}\frac{n}{n+1} = 1$. Donc la suite des expérances $(\mathbb{E}(X_n))_{n\in\mathbb{N}^*}$ ne converge pas vers l'espérance de la limite qui est $0 = \mathbb{E}(0)$.

Exercice 68 Enoncé p. 75

1) Comme $X_n \xrightarrow[n\to\infty]{\text{p.s}} X$, il existe Δ_X tel que $\mathbb{P}(\Delta_X) = 0$ et pour tout $\omega \in \Delta_X^c$, $\lim_{n\to\infty} X_n(\omega) = X(\omega)$. De même il existe Δ_Y de probabilité nulle tel que pour tout $\omega \in \Delta_Y^c$, $\lim_{n\to\infty} Y_n(\omega) = Y(\omega)$. Posons $\Delta = \Delta_X \cup \Delta_Y$, alors pour tout $\omega \in \Delta^c$, $\lim_{n\to\infty} X_n(\omega) = X(\omega)$ et $\lim_{n\to\infty} Y_n(\omega) = Y(\omega)$. 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, $0 \leqslant \mathbb{P}(\Delta) \leqslant \mathbb{P}(\Delta_X) + \mathbb{P}(\Delta_Y)$ 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 $f : (x, y) \to x + y$ et $g : (x, y) \to xy$ qui sont des applications continues sur $\mathbb{R}^2$.

Exercice 69 Enoncé p. 75

On commence par remarquer que $\sum_{k=1}^n X_k = n\overline{X}_n$ et on a

$$\begin{aligned}
\sum_{k=1}^n (X_k - \overline{X}_n)^2 &= \sum_{k=1}^n (X_k^2 - 2X_k\overline{X}_n + \overline{X}_n^2) \\
&= \left(\sum_{k=1}^n X_k^2\right) - 2\overline{X}_n \left(\sum_{k=1}^n X_k\right) + n\overline{X}_n^2 \\
&= \left(\sum_{k=1}^n X_k^2\right) - 2n\overline{X}_n^2 + n\overline{X}_n^2 = \left(\sum_{k=1}^n X_k^2\right) - n\overline{X}_n^2
\end{aligned}$$

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 $\sigma^2 = Var(X_0)$. On a

$$\overline{X}_n \xrightarrow[n\to\infty]{\text{p.s.}} m \quad \text{et} \quad \overline{X}_n^2 \xrightarrow[n\to\infty]{\text{p.s.}} m^2 .$$

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

$$\frac{1}{n}\sum_{k=1}^n X_k^2 \xrightarrow[n\to\infty]{\text{p.s.}} \mathbb{E}(X_0^2) .$$

Ainsi

$$S_n^2 = \underbrace{\frac{n}{n-1}}_{\to 1} \times \underbrace{\frac{1}{n}\sum_{k=1}^n X_k^2}_{\to \mathbb{E}(X_0^2)} - \underbrace{\frac{n}{n-1}}_{\to 1} \times \underbrace{\overline{X}_n^2}_{\to m^2} \xrightarrow[n\to\infty]{\text{p.s.}} \mathbb{E}(X_0^2) - m^2 = \sigma^2 .$$

Exercice 70 Enoncé p. 75

La fonction de répartition de δ_n est $\mathbb{1}_{[n,\infty[}$ pour tout $n \in \mathbb{N}$. De plus $\lim_{n\to\infty} \mathbb{1}_{[n,\infty[}(x) = 0$ donc la limite de la suite de fonction 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 $\lim_{x\to\infty} F(x) = 1$.
On a donc la convergence simple de la suite de fonction de répartitions $(F_{\delta_n})_{n\in\mathbb{N}}$ mais la suite $(\delta_n)_{n\in\mathbb{N}}$ ne converge pas étroitement vers une limite μ, sinon on aurait $\lim_{n\to\infty} F_{\delta_n}(x) = F_\mu(x) = 0$ pour tout point x de continuité de F_μ et donc $F_\mu = 0$ ce qui est impossible.

Exercice 71 Enoncé p. 75

On sait par le théorème de stabilité de la somme de variables de Poisson indépendantes (proposition 5.20), que la loi de S_n est $\mathcal{P}(n)$. Remarquons alors que

$$e^{-n}\sum_{k=0}^{k=n}\frac{n^k}{k!} = \mathbb{P}(S_n \le n) \text{ et } \{S_n \le n\} = \{\frac{1}{\sqrt{n}}(S_n - n) \le 0\}.$$

Par suite,

$$e^{-n}\sum_{k=0}^{k=n}\frac{n^k}{k!} = \mathbb{P}\left(\frac{1}{\sqrt{n}}(S_n - n) \le 0\right).$$

Le théorème-limite central permet d'affirmer que la suite $\left(\frac{1}{\sqrt{n}}(S_n - n)\right)_{n\ge 1}$ converge en loi vers une variable Y de loi $\mathcal{N}_1(0,1)$. En particulier $\mathbb{P}(\frac{1}{\sqrt{n}}(S_n - n) \le 0)$ tend vers $\mathbb{P}(Y \le 0) = \frac{1}{2}$ quand n tend vers $+\infty$. On conclut donc que la suite $\left(e^{-n}\sum_{k=0}^{k=n}\frac{n^k}{k!}\right)_{n\ge 1}$ converge vers $\frac{1}{2}$.

Exercice 72 Enoncé p. 75

Soit X le nombre de transistor défectueux dans un sachet de 100, X suit une loi $\mathcal{B}(100;0.01)$. L'événement "la garantie tombe en defaut" se modélise par $\{X \geqslant 3\}$. Comme le proposent les commentaires qui suivent la proposition 6.15, on approxime la loi binomiale par la loi de Poisson de paramètre 1 car on a $n = 100 \geqslant 30$, $np = 100 \times 0.01 < 10$ et $p \leqslant 0.1$. D'où

$$\mathbb{P}(X \geqslant 3) = 1 - \mathbb{P}(X \leqslant 2) \simeq 1 - \mathbb{P}(\mathcal{P}(1) \leqslant 2) \simeq 1 - 0.92 \simeq 8\% .$$

Exercice 73 Enoncé p. 75

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 $(\Phi_{\mu_n}(t))_{n\in\mathbb{N}} = \left(\exp\left(ia_nt - \frac{1}{2}t^2\sigma_n^2\right)\right)_{n\in\mathbb{N}}$ converge vers $\exp\left(iat - \frac{1}{2}t^2\sigma^2\right)$. La suite des fonctions caractéristiques $\left(\Phi_{\mathcal{N}_1(a_n,\sigma_n^2)}\right)_{n\in\mathbb{N}}$ 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 de la proposition 6.16.

Exercice 74 Enoncé p. 75

1) Considérons pour $k = 1, ..., N$ 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, $X_1, ..., X_N$ sont des v.a.r. indépendantes. Le nombre de spectateurs qui désirent aller dans la salle 1 est donc $S_N = X_1 + ... + X_N$ 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 que dans ce cas il y a au moins un spectateur qui ne vera pas son film. Dans ce cas, $p = 1$. Étudions le cas où $N \leqslant 2n$ c'est-à-dire $N - n \leqslant n$. Dans ce cas, $\{S > n\} \cap \{S < N - n\} = \emptyset$ et donc $p = \mathbb{P}(S > n) + \mathbb{P}(S \leqslant N - n)$. D'après le théorème central limite (N est implicitement supposé grand), on a

$$\lim_{N \to \infty} \mathbb{P}\left(\frac{\frac{S}{N} - \frac{1}{2}}{\frac{1}{2N}} \leqslant x \right) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-u^2/2} du \ .$$

Si on note $T = \frac{\frac{S}{N} - \frac{1}{2}}{\frac{1}{2N}}$, c'est une v.a.r. asymptotiquement de loi $\mathcal{N}(0,1)$ et donc

$$\begin{aligned} p &= \mathbb{P}\left(T > \frac{n - \frac{N}{2}}{\sqrt{N/4}} \right) + \mathbb{P}\left(T < \frac{N - n - \frac{N}{2}}{\sqrt{N/4}} \right) \\ &= \mathbb{P}\left(|T| \geqslant \frac{n - \frac{N}{2}}{\sqrt{N/4}} \right) = 2\Big(1 - \Phi\big[(2n - N)/\sqrt{N}\big]\Big) \ , \end{aligned}$$

où Φ désigne la fonction de répartition de la loi $\mathcal{N}(0,1)$.

2) Si $N = 1000$ et si on veut $p \leqslant 0.01$, il faut choisir n pour que $\Phi[(2n - 1000)/\sqrt{1000}] \geqslant 0.95$. La lecture inverse de la table de $\mathcal{N}(0; 1)$ donne $\Phi(2.58) \simeq 0.95$ d'où $(2n - 1000)/\sqrt{1000} \geqslant 2.58$ et par suite il faut prendre $n \geqslant 541$.

Chapitre 8

Bibliographie.

[1] Brémaud P., Introduction aux probabilités : modélisation des phénomènes aléatoires, Springer-Verlag, 1988. ix

[2] Ducel Y., Introduction à la théorie mathématique des probabilités, Ellipses, 1998. ix

[3] Leboeuf C.- Roque J.L.- Guegand J., Cours de probabilités et de statistiques, Ellipses, 2ème édition 1983. viii, 21

[4] Leboeuf C.- Roque J.L.- Guegand J., Exercices corrigés de probabilités, Ellipses, 1987. viii

Annexes

Annexe A

Formulaire

Ce petit formulaire est autorisé pendant les examens.

A.1 Quelques lois discrètes

- Loi de Bernoulli $\mathcal{B}(1, p)$, $0 \leqslant p \leqslant 1$:

$$\mathbb{P}_X = p\delta_1 + (1-p)\delta_0 \ .$$

- Loi binômiale $\mathcal{B}(n, p)$, $(n, p) \in \mathbb{N}^* \times [0, 1]$:

$$\mathbb{P}_X = \sum_{k=0}^{n} C_n^k p^k (1-p)^{n-k} \delta_k \ .$$

- Loi géométrique $\mathcal{G}(p)$, $0 < p \leqslant 1$:

$$\mathbb{P}_X = \sum_{k=1}^{\infty} p(1-p)^{k-1} \delta_k \ .$$

- Loi de Poisson $\mathcal{P}(\theta)$, $\theta \in]0, \infty[$:

$$\mathbb{P}_X = \sum_{k=0}^{\infty} e^{-\theta} \frac{\theta^k}{k!} \delta_k \ .$$

A.2 Quelques lois à densité (par rapport à la mesure de Lebesgue)

- Loi uniforme $\mathcal{U}_{[a,b]}$ sur l'intervalle $[a, b]$:

$$f_X(x) = \frac{1}{b-a} \mathbb{1}_{[a,b]}(x) \ .$$

- Loi gaussienne ou normale $\mathcal{N}(m, \sigma^2)$ de moyenne m et de variance $\sigma^2 > 0$:

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-(x-m)^2/2\sigma^2} \ .$$

- Loi exponentielle $\mathcal{E}(\lambda)$, $\lambda > 0$:

$$f_X(x) = \lambda e^{-\lambda x} \mathbb{1}_{]0,+\infty[}(x) \ .$$

- Loi gamma $\Gamma(\lambda, \alpha)$, $\lambda > 0$ et $\alpha > 0$:

$$f_X(x) = \frac{1}{\Gamma(\alpha)} \lambda^\alpha x^{\alpha-1} e^{-\lambda x} \mathbb{1}_{]0,+\infty[}(x) \ .$$

- Loi du khi 2 à d degré de liberté $\chi^2(d)$, $d \in \mathbb{N}^*$: c'est la loi $\Gamma(1/2, d/2)$.
- Loi béta $\beta(a, b)$, $a > 0$, $b > 0$:

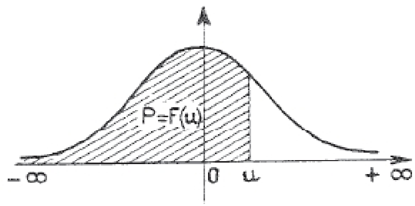
$$f_X(x) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} x^{a-1} (1_x)^{b-1} \mathbb{1}_{[0,1]}(x) \ .$$

A.3 Quelques propriétés des lois usuelles

Loi	$\mathbb{E}(X)$	$\mathrm{Var}(X)$	$\Phi_X(t)$
$\mathcal{B}(n, p)$	np	$np(1-p)$	$(1 - p + pe^{it})^n$
$\mathcal{G}(p)$	$1/p$	$(1-p)/p^2$	$pe^{it}/\big(1 - (1-p)e^{it}\big)$
$\mathcal{P}(\theta)$	θ	θ	$e^{-\theta(1-e^{it})}$
$\mathcal{U}_{[a,b]}$	$(a+b)/2$	$(b-a)^2/12$	$(e^{itb} - e^{ita})/\big(it(b-a)\big)$
$\mathcal{N}(m, \sigma^2)$	m	σ^2	$\exp\big(imt - \sigma^2 t^2/2\big)$
$\mathcal{E}(\lambda)$	$1/\lambda$	$1/\lambda^2$	$\lambda/(\lambda - it)$
$\Gamma(\lambda, \alpha)$	α/λ	α/λ^2	$\big(\lambda/(\lambda - it)\big)^\alpha$
$\chi^2(d)$	d	$2d$	
$\beta(a, b)$	$a/(a+b)$	$ab/\big((a+b)^2(a+b+1)\big)$	

Fonction de répartition de la loi normale centrée-réduite

(Probabilité de réaliser une valeur inférieure à **u**)



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 B

Utilisation de la table de la loi normale centrée-réduite

Calculs avec des v.a.r. normales centrées-réduites

Voici quelques exemples d'utilisation de la table du fascicule joint à ce cours et présente page 117.

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 donne les valeurs, connaissant le réel a positif, des expressions $\mathbb{P}(X < a)$. On peut toujours se ramener à ces cas moyennant les relations

1. Si a est une réel positif, $\mathbb{P}(X < a)$ est donné par la table.
2. Si a est une réel positif, $\mathbb{P}(X > a) = 1 - \mathbb{P}(X < a)$.
3. Si a est une réel strictement négatif, $\mathbb{P}(X < a) = 1 - \mathbb{P}(X < -a)$.
4. Si a est une réel strictement négatif, $\mathbb{P}(X > a) = \mathbb{P}(X < -a)$.

Pour lire dans la table la valeur de $\mathbb{P}(X < u)$ pour u positif, par exemple pour $u = 2,37$, 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 $\mathbb{P}(X < u)$ que pour $0 < u < 3$. Cela est dû au fait que pour les valeurs supérieures à 3, $\mathbb{P}(X < u) \approx 1$ et par suite $\mathbb{P}(X > u) \approx 0$. Toutefois la table donne les valeurs de $\mathbb{P}(X < u)$ pour u prenant des valeurs entre 3 et $4,5$ avec cinq décimales (tables des grandes valeurs pour u située au bas de la page).

Calculs avec des v.a.r. normales quelconques

Pour ce qui est du calcul de probabilité dans le cas de v.a.r. normales quelconques, on rappelle la proposition 2.7. 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 $a < b$,

$$\{a < X < b\} = \{\frac{a - m}{\sigma} < Z < \frac{b - m}{\sigma}\},$$

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.

Annexe C

Liste des définitions.

Annexe D

Liste des propositions.