

Table des matières

Probabilités

Barbara SCHAPIRA

29 janvier 2009

1 Probabilités et variables aléatoires	4
1.1 Espaces de probabilité	4
1.1.1 Le paradoxe de Banach-Tarski	4
1.1.2 Tribus	4
1.1.3 Mesures de probabilité	5
1.1.4 Le vocabulaire des probabilités	5
1.1.5 Rappels à connaître du S4	6
1.2 Retour sur les classes monotones	7
1.3 Variables aléatoires	8
1.3.1 Variable aléatoire, loi d'une variable aléatoire	8
1.3.2 Fonction de répartition d'une variable aléatoire à valeurs dans \mathbb{R} (ou $\mathbb{N}, \mathbb{Z}, \dots$)	10
1.4 Lois et variables aléatoires classiques	11
1.4.1 Lois discrètes	12
1.4.2 Lois continues	13
1.5 Variables aléatoires à densité	15
1.5.1 Définitions et premières propriétés	15
1.5.2 Changement de variables	17
2 Les grands outils probabilistes	19
2.1 Espérance, variance, écart-type	19
2.1.1 Définition	19
2.1.2 Exemples	20
2.1.3 Variance, écart-type	20
2.1.4 Corrélation de deux variables aléatoires	21
2.1.5 Moments	21
2.1.6 Inégalités classiques	21
2.2 Fonctions génératrices	22
2.3 Fonctions génératrices	22
2.3.1 Calcul des moments	23
2.4 Fonctions caractéristiques	23
2.4.1 Fonction caractéristique d'une variable aléatoire réelle	23
2.4.2 Fonction caractéristique d'une variable aléatoire vectorielle	25
3 Indépendance	26
3.1 Produit de lois	26
3.2 Couple de variables aléatoires	26
3.2.1 Couple de variables aléatoires, densité marginale	26
3.3 Variables aléatoires indépendantes	27
3.3.1 Définitions	27
3.3.2 Variables aléatoires indépendantes et produit de lois	28
3.4 Le lemme de Borel-Cantelli	29
3.5 Espérance et indépendance	30
3.6 Indépendance et fonctions caractéristiques	31
3.6.1 Exemple fondamental : somme de variables aléatoires indépendantes.	31

4	Convergences et théorèmes limites	33
4.1	Convergences de variables aléatoires	33
4.1.1	Convergence presque sûre	33
4.1.2	Convergence en probabilité	34
4.1.3	Convergence de mesures, Convergence en loi	34
4.1.4	Convergence dans L^p	36
4.2	Loi faible et forte des grands nombres	37
4.2.1	Loi faible des grands nombres	37
4.2.2	Loi forte des grands nombres	38
4.3	Théorème limite central	40

L'alphabet grec

Beaucoup d'entre vous n'ont jamais fait de grec ancien. Or les mathématicien-ne-s usent et abusent des lettres grecques.

Les voici donc, dans l'ordre, avec minuscule, majuscule, et nom.

α , A , Alpha
β , B , Bêta
γ , Γ , Gamma
δ , Δ , Delta
ε (ou parfois ϵ), E , epsilon
ζ , Z , zêta (prononcer "dzeta")
η , H , êta
θ Θ Thêta
ι I iota
κ K Kappa
λ Λ lambda
μ M mu
ν V nu
ξ Ξ xi
o O omicron (prononcer "omicrone")
π Π pi
ρ P rho
σ Σ Sigma
τ T tau
υ U upsilon
φ (ou parfois ϕ) Φ Phi
χ X Chi (prononcer "ki")
ψ Ψ psi
ω Ω omega.

Chapitre 1

Probabilités et variables aléatoires

1.1 Espaces de probabilité

1.1.1 Le paradoxe de Banach-Tarski

Soit X un ensemble. Idéalement, on voudrait dire a priori qu'une mesure μ sur X , ce serait une application $\mu : \mathcal{P}(X) \rightarrow \overline{\mathbb{R}}_+$ définie sur l'ensemble de toutes les parties de X , additive ($\mu(A \cup B) = \mu(A) + \mu(B)$ si A et B sont disjoints, avec $\mu(\emptyset) = 0$). (L'intuition de ce qu'est une mesure peut être guidée par la notion de volume et/ou de probabilité.)

Cela existe. Par exemple, sur \mathbb{R} , pour toute partie $A \subset \mathbb{R}$, on peut définir $\mu_n(A) = \frac{1}{n} \# \{0, \dots, n-1\} \cap A$. On peut passer à la limite et obtenir ce qu'on appelle une "moyenne" μ qui est additive, positive et vérifie $\mu(\mathbb{R}) = 1$. Toutefois, elle vérifie aussi $\mu(A) = 0$ pour toute partie A bornée. Ennuyeux... On aimerait dire quand même que $\mu(\mathbb{R}) = \sum_{n \in \mathbb{Z}} \mu([n, n+1[)$.

C'est ce qu'on appelle la σ -additivité. Le problème est alors différent. On ne peut alors pas supposer que toutes les parties d'un ensemble sont mesurables. En effet, c'est le paradoxe suivant, dit de *Banach-Tarski*. On peut prendre une sphère disons de rayon 1, la couper en un nombre fini de morceaux, et les recoller pour faire DEUX sphères de rayon 1 ! Autrement dit, quelque chose d'absolument anti-intuitif. Ces morceaux ne sont pas mesurables, c'est ce qui empêche de dire que les deux sphères ont même volume...

En effet, imaginons qu'on puisse calculer la mesure de tout ensemble de la sphère (i.e. ici l'aire). Dans ce cas, à l'aide du paradoxe ci-dessus, on obtiendrait que la surface d'une sphère de rayon 1 est égale à celle d'une sphère de rayon 2. Contradiction ! Autrement dit, il n'est pas possible que les ensembles intervenant dans le découpage de la sphère ci-dessus soient mesurables.

Notons que la démonstration de ce résultat fait appel à l'axiome du choix. Cet axiome, refusé par certains mathématiciens, dit en gros qu'étant donnée une collection non dénombrable d'ensembles, on peut choisir un élément de chaque ensemble. (Ceci est toujours possible pour des collections finies ou dénombrables en procédant par récurrence.) Plus précisément, si $(E_i)_{i \in I}$ est une collection d'ensembles, il existe un élément $x = (x_i)_{i \in I}$ dans le produit infini $\prod_{i \in I} E_i$.

Remarque En TD d'intégration (exo ? feuille ?) nous avons vu un exo montrant qu'à l'aide de l'axiome du choix, on peut construire un ensemble non mesurable.

1.1.2 Tribus

Définition 1.1.1 Soit X un ensemble. Une tribu, ou σ -algèbre est une collection $\mathcal{A} \subset \mathcal{P}(X)$ de parties de X qui contient \emptyset et X , est stable par union dénombrable et passage au complémentaire.

Remarquons qu'une tribu est stable par intersection dénombrable (exo).

Une *algèbre* est une collection de parties qui contient l'ensemble vide et est stable par union finie et passage au complémentaire.

Les exemples classiques et les plus utiles en pratique sont les suivants. Si X est (au plus) dénombrable, $\mathcal{P}(X)$ est une tribu. Sur \mathbb{R}^d , $d \geq 1$ (et en particulier sur \mathbb{R}), on considère la *tribu borélienne* engendrée par les ouverts. C'est la plus petite tribu qui contient tous les ouverts, mais donc aussi tous les fermés...

Autres tribus : sur X la famille $\{\emptyset, X\}$ est toujours une tribu. Si $A \subset \mathcal{P}(X)$ est un ensemble, la tribu engendrée par A est $\{\emptyset, X, A, X \setminus A\}$.

En Master, vous aurez l'occasion de rencontrer d'autres tribus, et même des familles de tribus, utiles. Cette année, les tribus $\mathcal{P}(\mathbb{N})$ et $\mathcal{B}(\mathbb{R})$ sont les plus importantes.

Exercice 1.1.2 Trouver toutes les tribus de $X = \{1, 2\}$, puis de $X = \{1, 2, 3\}$.

1.1.3 Mesures de probabilité

Rappelons que si \mathcal{A} est une tribu sur X , un ensemble *mesurable* est simplement un ensemble $A \in \mathcal{P}(X)$ qui est dans la tribu \mathcal{A} . Autrement dit, $A \in \mathcal{A} \subset \mathcal{P}(X)$.

Définition 1.1.3 Soit (X, \mathcal{A}) un espace mesurable. Une (mesure de) probabilité est une application $\mathbf{P} : \mathcal{A} \rightarrow [0, 1]$ qui vérifie $\mathbf{P}(X) = 1$ et si $(A_n)_{n \in \mathbb{N}}$ est une suite de sous-ensembles mesurables de X deux à deux disjoints, alors $\mathbf{P}(\cup_{n \in \mathbb{N}} A_n) = \sum_{n \in \mathbb{N}} \mathbf{P}(A_n)$.

Comparez avec la définition d'une mesure.

Voici quelques propriétés utiles en pratique, à savoir démontrer (certaines ne sont que des reformulations de la définition).

1. $\mathbf{P}(\emptyset) = 0$
2. Si $(A, B) \in \mathcal{A}^2$ et $A \subset B$ alors $\mathbf{P}(A) \leq \mathbf{P}(B)$.
3. Si $(A, B) \in \mathcal{A}^2$ et $A \cap B = \emptyset$ alors $\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B)$
4. $\mathbf{P}(A^c) = 1 - \mathbf{P}(A)$
5. $\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B) - \mathbf{P}(A \cap B)$
6. $\mathbf{P}(A \cap B) \leq \min(\mathbf{P}(A), \mathbf{P}(B))$
7. $\mathbf{P}(A \cup B) \geq \max(\mathbf{P}(A), \mathbf{P}(B))$
8. Si $(A_n)_{n \in \mathbb{N}} \in \mathcal{A}^{\mathbb{N}}$ est une suite croissante pour l'inclusion (i.e. $A_n \subset A_{n+1}$ pour tout $n \in \mathbb{N}$) alors la suite $(\mathbf{P}(A_n))_{n \in \mathbb{N}}$ est croissante et $\lim_{n \rightarrow \infty} \mathbf{P}(A_n) = \mathbf{P}(\cup_{n \in \mathbb{N}} A_n)$
9. Si $(A_n)_{n \in \mathbb{N}} \in \mathcal{A}^{\mathbb{N}}$ est une suite décroissante pour l'inclusion alors la suite $(\mathbf{P}(A_n))_{n \in \mathbb{N}}$ est décroissante et $\lim_{n \rightarrow \infty} \mathbf{P}(A_n) = \mathbf{P}(\cap_{n \in \mathbb{N}} A_n)$.

Exercice 1.1.4 Démontrer ces propriétés. Indication : Considérer $B_n = A_n \setminus A_{n-1}$.

Définition 1.1.5 Un espace de probabilité, ou espace probabilisé, $(X, \mathcal{A}, \mathbb{R})$ est la donnée d'un ensemble X muni d'une tribu \mathcal{A} et d'une mesure de probabilité \mathbf{P} .

En pratique, les espaces sur lesquels on travaillera le plus sont $(\mathbb{N}, \mathcal{P}(\mathbb{N}))$ muni d'une des mesures de probabilité introduites au paragraphe 1.4, ou bien $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ et plus généralement $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$, $d \geq 1$, muni d'une mesure de probabilité de 1.4.

1.1.4 Le vocabulaire des probabilités

La théorie des probabilités permet de fournir des modèles mathématiques pertinents à des phénomènes faisant intervenir du hasard, de l'aléa (lancer de dé, cours d'une action en bourse, durée de vie d'une ampoule, temps d'attente d'un bus, fluctuations de certaines quantités : variations des différentes mesures d'une grandeur physique autour de la valeur théorique attendue, variation de la température autour de la valeur moyenne saisonnière attendue, etc

On modélise un ensemble de situations possibles par un espace de probabilité $(\Omega, \mathcal{A}, \mathbf{P})$. L'ensemble Ω est l'espace des états. Par exemple, si on tire une pièce à pile ou face, on peut considérer $\Omega = \{\text{pile, face}\}$. Pour un lancer de dé, on utilisera volontiers $\Omega = \{1, 2, 3, 4, 5, 6\}$. Si on tire à pile ou face trois fois de suite, on utilisera sans doute $\Omega = \{\text{pile, face}\}^3 = \{ppp, ppf, pfp, pff, fpp, fpf, ffp, fff\} \dots$

Rappelons que quand Ω est fini ou dénombrable, on utilise $\mathcal{A} = \mathcal{P}(\Omega)$. Un événement est un ensemble $A \in \mathcal{A}$. Par exemple, si on lance un dé, l'événement « obtenir un nombre pair » s'écrit encore $\{2, 4, 6\} \subset \Omega$.

En lançant deux dés, avec $\Omega = \{1, \dots, 6\}^2$, l'événement « obtenir un 7 » s'écrit $A = \{(1, 6), (2, 5), (3, 4), (4, 3), (5, 2), (6, 1)\}$.

On appelle $\mathbf{P}(A)$ la probabilité que l'événement A se réalise. Si $\mathbf{P}(A) = 1$ on dit que l'événement A est certain, ou presque sûr (p.s.).

Intuitivement (historiquement) si on répète une expérience un grand nombre n de fois, et qu'on mesure la fréquence $f_n(A)$ d'un événement A , ($f_n(A) = \frac{1}{n} \times \text{nb de réalisations de } A$), on souhaite définir la probabilité de A comme étant la limite de $f_n(A)$ quand le nombre n d'expériences tend vers $+\infty$. Dans la théorie moderne des probabilités, on se donne a priori une mesure de probabilité \mathbf{P} et on démontre (voir le dernier chapitre de ce cours, lois des grands nombres) que les fréquences $f_n(A)$ convergent vers $\mathbf{P}(A)$.

Une *variable aléatoire* est une fonction mesurable de l'espace mesurable (Ω, \mathcal{A}) dans un autre espace mesurable (E, \mathcal{E}) . En pratique, on verra qu'on ne connaît que peu de choses sur l'espace Ω , qu'on ne définit même pas forcément précisément lorsqu'on modélise un phénomène. Et toute l'étude porte sur l'espace d'arrivée (E, \mathcal{E}) qui est souvent fini ou $(\mathbb{N}, \mathcal{P}(\mathbb{N}))$ ou $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$.

Par exemple, lorsque je lance un dé, je peux considérer Ω l'ensemble de toutes les trajectoires possibles du dé de ma main au sol, et X la fonction définie sur Ω à valeurs dans $\{1, 2, 3, 4, 5, 6\}$ qui à une trajectoire associe le numéro sur la face du dessus à l'arrivée du dé sur le sol. Pour une modélisation correcte, on supposera que X est mesurable (ce qui revient à choisir sur Ω la tribu adaptée pour cela). On pourrait aussi choisir $\Omega = \{1, 2, 3, 4, 5, 6\}$.

Autre exemple : lancer de deux dés, $\Omega = \{1, \dots, 6\}^2$ et $X : (i, j) \mapsto i + j$.

1.1.5 Rappels à connaître du S4

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

Définition 1.1.6 Soit $B \in \mathcal{A}$ un ensemble mesurable t.q. $\mathbf{P}(B) > 0$. La fonction $\mathbf{P}(\cdot|B)$ définie par $A \in \mathcal{A} \mapsto \mathbf{P}(A|B) = \frac{\mathbf{P}(A \cap B)}{\mathbf{P}(B)}$ est une probabilité sur (Ω, \mathcal{A}) , et la quantité $\mathbf{P}(A|B)$ est appelée la probabilité conditionnelle de A sachant B .

Exercice 1.1.7 Vérifier que $\mathbf{P}(\cdot|B)$ est bien une probabilité.

$\mathbf{P}(A|B)$ représente la probabilité que A se réalise sachant que B se réalise.

Définition 1.1.8 Deux événements A et B sont indépendants si $\mathbf{P}(A \cap B) = \mathbf{P}(A)\mathbf{P}(B)$.

Une famille $(A_i)_{i \in I}$ finie ou dénombrable est indépendante si pour tout sous-ensemble fini d'indices $J \subset I$ $\mathbf{P}(\cap_{j \in J} A_j) = \prod_{j \in J} \mathbf{P}(A_j)$.

Les événements $(A_i)_{i \in I}$ sont deux à deux indépendants si pour tout couple (i, j) d'indices avec $i \neq j$, A_i et A_j sont indépendants.

Bien sûr, si la famille (A_i) est indépendante dans son ensemble, alors les (A_i) sont deux à deux indépendants.

Vocabulaire : en l'absence de précision, « des événements indépendants » signifie des événements indépendants dans leur ensemble.

Un exemple : on joue deux fois de suite à pile ou face. A est l'événement « obtenir pile au premier lancer », B « obtenir pile au deuxième lancer » et C « obtenir deux lancers identiques ». Ces événements sont-ils deux à deux indépendants ? Indépendants dans leur ensemble ?

Exercice 1.1.9 On lance un dé, et on note i le chiffre obtenu. (Ici $\Omega = \{1, \dots, 6\}$). Donner un exemple de deux événements A et B indépendants.

Remarque : Si A et B sont indépendants alors $\mathbf{P}(A|B) = \mathbf{P}(A)$.

Exercice 1.1.10 Si A et B sont indépendants, alors A et B^c aussi, A^c et B aussi, A^c et B^c aussi.

Proposition 1.1.11 (Probabilités composées) Soit $(A_1, \dots, A_n) \in \mathcal{A}^n$ tel que $\mathbf{P}(A_1 \cap \dots \cap A_n) > 0$. Alors $\mathbf{P}(A_1 \cap A_2 \cap \dots \cap A_n) = \mathbf{P}(A_1)\mathbf{P}(A_2|A_1)\mathbf{P}(A_3|A_1 \cap A_2) \dots \mathbf{P}(A_n|A_1 \cap \dots \cap A_{n-1})$

Démonstration : Exercice très facile à savoir faire! □

Définition 1.1.12 Une partition mesurable de X est une famille d'ensembles $(E_i)_{i \in I}$ mesurables deux à deux disjoints tels que $\cup_{i \in I} E_i = \Omega$. On suppose souvent implicitement que $\mathbf{P}(E_i) > 0$ pour tout $i \in I$. (Pourquoi est-ce possible sans difficulté ?)

Proposition 1.1.13 (Probabilités totales) Soit $(E_i)_{i \in I} \subset \mathcal{A}^I$ une partition mesurable finie ou dénombrable de Ω . Alors pour tout $A \in \mathcal{A}$, on a $\mathbf{P}(A) = \sum_{i \in I} \mathbf{P}(A \cap E_i) = \sum_{i \in I} \mathbf{P}(A|E_i)\mathbf{P}(E_i)$

Démonstration : Exercice très facile à savoir faire! \square

Mise en garde Cette formule élémentaire provoque un nombre incalculable d'erreurs scandaleuses. Débrouillez-vous mais apprenez la, sachez la démontrer et l'utiliser sans écrire des choses du style $\mathbf{P}(A) \dots \sum_i \mathbf{P}(A|E_i)$
Exemples : voir TD et compléter.

Proposition 1.1.14 (Formule de Bayes) Soit $(E_i)_{i \in I} \subset \mathcal{A}^I$ une partition mesurable finie ou dénombrable de Ω . Si $A \in \mathcal{A}$ vérifie $\mathbf{P}(A) > 0$ alors

$$\mathbf{P}(E_n|A) = \frac{\mathbf{P}(A|E_n)\mathbf{P}(E_n)}{\sum_{m \in I} \mathbf{P}(A|E_m)\mathbf{P}(E_m)}$$

Démonstration : Exercice très facile à savoir faire! \square

Remarque : ce résultat complètement élémentaire ne mérite pas le nom de proposition par sa difficulté, mais par son utilité pratique. (Voir TD).

1.2 Retour sur les classes monotones

Ce paragraphe est tiré de [R1].
Rappelons la définition

Définition 1.2.1 Une famille \mathcal{M} de parties d'un ensemble E est appelée classe monotone si

- $E \in \mathcal{M}$,
- si $A \in \mathcal{M}$, $B \in \mathcal{M}$, et $A \subset B$, alors $B \setminus A \in \mathcal{M}$,
- Si $(A_n)_{n \in \mathbb{N}}$ est une suite croissante d'éléments de \mathcal{M} , alors $\cup_{n \in \mathbb{N}} A_n$ est dans \mathcal{M} .

La Classe monotone $\mathcal{M}(\mathcal{C})$ engendrée par une famille \mathcal{C} de parties de E est la plus petite classe monotone contenant \mathcal{C} , ou encore l'intersection de toutes les classes monotones contenant \mathcal{C} .

Une tribu est une classe monotone. La classe monotone $\mathcal{M}(\mathcal{C})$ engendrée par une famille \mathcal{C} de parties est donc incluse dans la tribu $\sigma(\mathcal{C})$ engendrée par \mathcal{C} .

Remarque 1.2.2 Une classe monotone stable par intersection finie est une tribu. En effet, les deux premières propriétés garantissent alors que \emptyset est dans la classe monotone, et que la classe est stable par union finie. La dernière propriété assure enfin que la classe est stable par union dénombrable.

La notion de classe monotone prend toute son utilité dans le théorème d'apparence technique suivant, dont on verra l'intérêt dans le corollaire qui suit.

Théorème 1.2.3 (Théorème de la classe monotone) Si \mathcal{C} est une famille stable par intersection finie, alors $\mathcal{M}(\mathcal{C}) = \sigma(\mathcal{C})$.

Démonstration : Comme $\mathcal{M}(\mathcal{C}) \subset \sigma(\mathcal{C})$, il suffit de montrer que $\mathcal{M}(\mathcal{C})$ est une tribu. Pour cela, vue la remarque ci-dessus, il suffit de montrer qu'elle est stable par intersection finie.

• Posons $\mathcal{M}_1 = \{A \in \mathcal{M}(\mathcal{C}), \forall C \in \mathcal{C}, A \cap C \in \mathcal{M}(\mathcal{C})\}$. Alors $\mathcal{C} \subset \mathcal{M}_1$ (par hypothèse sur \mathcal{C}). Et \mathcal{M}_1 est une classe monotone contenant \mathcal{C} et incluse dans $\mathcal{M}(\mathcal{C})$, donc $\mathcal{M}_1 = \mathcal{M}(\mathcal{C})$. En effet, $E \in \mathcal{M}_1$ bien sûr. Si A et B sont dans \mathcal{M}_1 et $A \subset B$, et $C \in \mathcal{C}$, alors $(B \setminus A) \cap C = (B \cap C) \setminus (A \cap C)$; $B \cap C$ et $A \cap C$ sont dans $\mathcal{M}(\mathcal{C})$ par définition de \mathcal{M}_1 , et leur différence aussi car $\mathcal{M}(\mathcal{C})$ est une classe monotone. La stabilité de \mathcal{M}_1 par union dénombrable croissante découle du fait que $(\cup_{n \in \mathbb{N}} A_n) \cap C = \cup_{n \in \mathbb{N}} (A_n \cap C)$.

• Posons $\mathcal{M}_2 = \{A \in \mathcal{M}(\mathcal{C}), \forall C \in \mathcal{M}(\mathcal{C}), A \cap C \in \mathcal{M}(\mathcal{C})\}$. De la même façon, on vérifie que \mathcal{M}_2 est une classe monotone, incluse dans $\mathcal{M}(\mathcal{C})$ et contenant \mathcal{C} , donc égale à $\mathcal{M}(\mathcal{C})$. Donc $\mathcal{M}(\mathcal{C})$ est stable par intersection finie. \square

Corollaire 1.2.4 (Unicité des mesures) Si \mathcal{C} est une classe d'ensembles stable par intersection finie et engendrant une tribu \mathcal{A} sur E , alors deux mesures de probabilité sur \mathcal{A} qui sont égales sur \mathcal{C} sont égales sur \mathcal{A} .

Ce corollaire nous sera utile par la suite.

Démonstration : Soient μ et ν deux mesures de probabilité égales sur \mathcal{C} ; notons

$$\mathcal{M} = \{A \in \mathcal{A}, \mu(A) = \nu(A)\}.$$

On vérifie aisément que \mathcal{M} est une classe monotone qui contient \mathcal{C} . Le théorème précédent permet de conclure. \square

Fin cours 1 (2h)

1.3 Variables aléatoires

1.3.1 Variable aléatoire, loi d'une variable aléatoire

Définition 1.3.1 Une variable aléatoire est une application mesurable $X : (\Omega, \mathcal{A}) \rightarrow (E, \mathcal{E})$. (Souvent, $E = \mathbb{N}$ ou $E = \mathbb{R}^d$). Lorsque $E = \mathbb{R}$, ou $E \subset \mathbb{R}$ on parle de variable aléatoire réelle (v.a.r.) Lorsque $E = \mathbb{R}^d$ avec $d \geq 2$ on parle de vecteur aléatoire.

Lorsque E est un ensemble discret, fini ou dénombrable, on parle de variable aléatoire discrète. Lorsque $E \subset \mathbb{N}$ ou \mathbb{Z} , on parle de variable aléatoire entière.

Bien évidemment, une v.a. à valeurs dans \mathbb{N} est une v.a.r. mais on parle de variable aléatoire discrète.

Remarque 1.3.2 « Rappelons » qu'un ensemble discret est un espace topologique sur lequel les singletons sont ouverts.

Définition 1.3.3 La tribu engendrée par une v.a. $X : (\Omega, \mathcal{A}) \rightarrow (E, \mathcal{E})$ est la plus petite tribu $\sigma(X) \subset \mathcal{A}$ rendant X mesurable. Elle peut encore être définie comme $\sigma(X) = \{X^{-1}(B), B \in \mathcal{E}\}$.

Démonstration : On commence par remarquer que $\sigma(X)$ contient $X^{-1}(\mathcal{E})$, puis que $X^{-1}(\mathcal{E})$ est une tribu, enfin que $\sigma(X)$ est par définition la plus petite tribu rendant X mesurable, i.e. la plus petite tribu contenant $X^{-1}(\mathcal{E})$. \square

Définition 1.3.4 Soit \mathbf{P} une probabilité sur (Ω, \mathcal{A}) . La loi image ou mesure image de \mathbf{P} par X est la mesure \mathbf{P}_X sur (E, \mathcal{E}) définie par $\mathbf{P}_X(A) = \mathbf{P}(X^{-1}(A))$ pour tout $A \in \mathcal{E}$. La mesure de probabilité \mathbf{P}_X est appelée la loi de X .

Notation : on écrit souvent (abus de notation utile pour l'intuition) $\mathbf{P}(X \in A)$ à la place de $\mathbf{P}_X(A) = \mathbf{P}(X^{-1}(A)) = \mathbf{P}(\{\omega \in \Omega, X(\omega) \in A\})$.

Lorsqu'on modélise une situation réelle avec des outils probabilistes, en général, on ne connaît pas $(\Omega, \mathcal{A}, \mathbf{P})$. Lors d'une expérience aléatoire $\omega \in \Omega$ on ne fait qu'observer la valeur $X(\omega) \in \mathbb{N}$ ou \mathbb{R} . On est alors intéressé par la probabilité que $X(\omega)$ prenne telle ou telle valeur. Autrement dit, on veut savoir étudier \mathbf{P}_X sur E et pas du tout \mathbf{P} sur Ω . L'espace de probabilité $(\Omega, \mathcal{A}, \mathbf{P})$ est important du point de vue théorique, mais n'est pas explicite en général.

Vous rencontrerez donc fréquemment des phrases du type « $X : \Omega \rightarrow \{0, 1\}$ suit une loi de Bernoulli de paramètre p », alors qu'on n'a même pas précisé Ω . On dit même X suit une loi de Bernoulli $\mathcal{B}(p)$ sur $\{0, 1\}$ sans évoquer l'espace Ω quand seule la loi \mathbf{P}_X de X est utile.

Un exemple de loi image : $\Omega = \{(i, j), 1 \leq i, j \leq 6\}$; $\mathbf{P}(i, j) = \frac{1}{36}$ et $X : \Omega \rightarrow \mathbb{N}$ définie par $X(i, j) = i + j$. Pourquoi X est-elle mesurable ? Sur \mathbb{N} la loi de X est définie par $\mathbf{P}_X(\{n\}) = \mathbf{P}(X^{-1}(n)) = \mathbf{P}(\{\omega \in \Omega, X(\omega) = n\})$. Et on a $\mathbf{P}_X(0) = 0 = \mathbf{P}_X(n)$ pour $n \geq 13$. $\mathbf{P}_X(2) = \frac{1}{36}$, $\mathbf{P}_X(3) = \frac{1}{18}$, ...

Autre exemple : $\Omega = [0, 1]$ muni de sa tribu borélienne et de la mesure de Lebesgue. Soit X la variable aléatoire définie par $X(\omega) = 1$ si $\omega \in [0, 1/2]$ et 0 sinon ($X = \mathbf{1}_{[0, 1/2]}$). Alors \mathbf{P}_X vérifie $\mathbf{P}_X(0) = \mathbf{P}_X(1) = \frac{1}{2}$ et $\mathbf{P}_X(\mathbb{R} \setminus \{0, 1\}) = 0$. Autrement dit, $\mathbf{P}_X = \frac{1}{2}(\delta_0 + \delta_1)$ où δ_x désigne la masse de Dirac en x .

Exercice 1.3.5 On considère $(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathbf{P})$ où $\mathbf{P} = \frac{1}{3}\delta_{-1} + \frac{1}{2}\delta_0 + \frac{1}{6}\delta_1$. C'est une probabilité sur \mathbb{R} . Soit $X : \omega \mapsto |\omega|$. Montrer que $\mathbf{P}_X = \frac{1}{2}(\delta_0 + \delta_1)$.

Remarque 1.3.6 Une mesure de probabilité \mathbf{P} sur un espace mesurable (Ω, \mathcal{A}) est toujours la loi d'une variable aléatoire. Considérer $X = Id : (\Omega, \mathcal{A}) \rightarrow (\Omega, \mathcal{A})$. Si au départ Ω est muni de la loi de probabilité \mathbf{P} , alors à l'arrivée la loi image de \mathbf{P} par X , i.e. la loi \mathbf{P}_X de X est égale à \mathbf{P} .

Remarque 1.3.7 (Important !) Il existe des variables aléatoires différentes de même loi. Par exemple sur $[0, 1]$ muni de la tribu borélienne et de la mesure de Lebesgue, considérer $X = \mathbf{1}_{[0,2/3]}$ et $Y = \mathbf{1}_{[1/4,7/12]} + \mathbf{1}_{[2/3,1]}$. Considérer même $Z : \{1, \dots, 6\} \rightarrow \{0, 1\}$ défini par $Z(\omega) = 1$ si ω n'est pas divisible par trois. Alors Z a la même loi que X et Y (exo) alors qu'il n'est pas défini sur le même espace!

Théorème 1.3.8 Soit $X : (\Omega, \mathcal{A}, \mathbf{P}) \rightarrow (E, \mathcal{E})$ une variable aléatoire. Alors pour toute fonction $h : E \rightarrow \mathbb{R}$ mesurable bornée ou mesurable positive, on a

$$\int_{\Omega} h \circ X(\omega) d\mathbf{P}(\omega) = \int_E h(x) d\mathbf{P}_X(x)$$

Plus généralement, h est \mathbf{P}_X -intégrable ssi $h \circ X$ est \mathbf{P} -intégrable et la formule ci-dessus est encore vraie dans ce cas.

Remarque 1.3.9 (Cas particuliers importants) Lorsque $E = \mathbb{N}$ le terme de droite se réécrit $\sum_{n \in \mathbb{N}} h(n) \mathbf{P}(X = n)$.

Lorsque Ω est discret (souvent $\Omega = \mathbb{N}$ ou Ω fini en pratique), alors $\int_{\Omega} h \circ X(\omega) d\mathbf{P}(\omega) = \sum_{\omega \in \Omega} p(\omega) h \circ X(\omega)$.

Cette formule dont la démonstration est élémentaire est fondamentale. En effet, répétons encore une fois qu'en pratique, on ne connaît pas Ω ni \mathbf{P} . C'est donc sur E qu'on effectuera tous les calculs. Le terme de gauche est important théoriquement, le terme de droite sert en pratique. (Ce résultat est appelé principe de transfert dans certains ouvrages.)

Remarque 1.3.10 Pourquoi ces hypothèses sur h ? Rappelons que l'intégrale $\int_E h(x) d\mathbf{P}_X(x)$ est définie pour les fonctions mesurables positives ou les fonctions \mathbf{P}_X -intégrables. Or en théorie des probabilités, les variables aléatoires bornées sont des fonctions intégrables (Si $|f| \leq C$, alors $\int_E f d\mathbf{P}_X \leq C \mathbf{P}_X(E) = C < \infty$).

Cela dit, le bon cadre de ce théorème est f mesurable positive ou intégrable.

Attention : C'est faux en analyse! Sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ muni de la mesure de Lebesgue, les fonctions bornées ne sont pas toutes intégrables! Exemple, les fonctions constantes...

Démonstration : Principe de preuve très classique, à connaître. On passe progressivement des indicatrices de boréliens à toutes les fonctions mesurables. On reverra ce type de preuve à plusieurs reprises.

- Si $A \in \mathcal{E}$ et $h = \mathbf{1}_A$ il s'agit simplement de la définition de $\mathbf{P}_X : \mathbf{1}_A \circ X = \mathbf{1}_{X^{-1}(A)}$ et $\mathbf{P}(X^{-1}(A)) = \mathbf{P}_X(A)$.
- Si $h = \sum_i \alpha_i \mathbf{1}_{A_i}$ est une fonction simple (somme finie d'indicatrices pondérées par les α_i), alors le résultat souhaité est vrai par linéarité de l'intégrale.
- Si h est mesurable positive, alors il existe une suite (h_n) croissante de fonctions simples positives telles que $h_n \rightarrow h$, et donc aussi $h_n \circ X \rightarrow h \circ X$ presque sûrement. Le théorème de convergence monotone (Beppo-Levi) donne le résultat souhaité.
- Si h est mesurable bornée, elle est limite de fonctions simples uniformément bornées; le théorème de convergence dominée de Lebesgue s'applique.
- Si h est \mathbf{P}_X -intégrable (plus faible que mesurable bornée car \mathbf{P}_X est une probabilité), alors il existe h_+ et h_- mesurables positives telles que $h = h_+ - h_-$. On a $|h| = h_+ + h_-$, et $\int_{\mathbb{R}} |h| d\mathbf{P}_X < \infty$ implique $\int_{\mathbb{R}} h_{\pm} d\mathbf{P}_X < \infty$. Encore une fois, la linéarité de l'intégrale va mener à l'égalité voulue, et donc en particulier au fait que $h \circ X$ est \mathbf{P} -intégrable. Le même raisonnement en partant de cette dernière hypothèse donnerait $h \mathbf{P}_X$ intégrable. \square

La proposition suivante est un cas particulier très utile.

Proposition 1.3.11 La v.a.r. X est intégrable ssi $\int_{\mathbb{R}} |x| d\mathbf{P}_X(x) < \infty$.

Démonstration : C'est un corollaire du théorème 1.3.8 avec $h(x) = |x|$. \square

1.3.2 Fonction de répartition d'une variable aléatoire à valeurs dans \mathbb{R} (ou $\mathbb{N}, \mathbb{Z}, \dots$)

Dans ce paragraphe, sauf mention du contraire, $X : (\Omega, \mathcal{A}) \rightarrow (E, \mathcal{E})$ désigne une variable aléatoire à valeurs dans un sous-ensemble de \mathbb{R} .

Définition 1.3.12 Soit \mathbf{P} une mesure de probabilité sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ ou $(\mathbb{N}, \mathcal{P}(\mathbb{N}))$. La fonction de répartition de \mathbf{P} est la fonction $F : \mathbb{R} \rightarrow [0, 1]$ définie par $F(x) = \mathbf{P}(\cdot - \infty, x]$.

Exemples : $\delta_2, \frac{\delta_1 + \delta_2}{2}, \dots$

Exercice 1.3.13 Tracer le graphe de la fonction de répartition de δ_0 , de $\frac{1}{3}\delta_{-2} + \frac{1}{2}\delta_1 + \frac{1}{6}\delta_{13}$, de la mesure uniforme sur $[-1, 1] : \mathbf{P} = \frac{1}{2}\mathbf{1}_{[-1,1]}\lambda$, où λ est la mesure de Lebesgue...

Définition 1.3.14 La fonction de répartition de X , notée F_X , est la fonction de répartition de sa loi \mathbf{P}_X . Autrement dit,

$$F_X(x) = \mathbf{P}_X(\cdot - \infty, x] = \mathbf{P}(X \leq x) = \mathbf{P}(\{\omega \in \Omega, X(\omega) \leq x\})$$

Proposition 1.3.15 Soit F_X la fonction de répartition de \mathbf{P}_X . Alors on a

1. F_X est croissante.
2. F_X est à valeurs dans $[0, 1]$
3. F_X est continue à droite.
4. F_X a des limites à gauche en tout point, et $\lim_{x \rightarrow a^-} F_X(x) = F_X(a^-) = \mathbf{P}(X < a) \leq \mathbf{P}(X \leq a) = F_X(a)$.
5. $F_X(x) \rightarrow 0$ quand $x \rightarrow -\infty$ et $F_X(x) \rightarrow 1$ quand $x \rightarrow +\infty$.
6. $F_X(b) - F_X(a) = \mathbf{P}(a < X \leq b)$
7. $\mathbf{P}(X = a) = F_X(a) - \lim_{x \rightarrow a^-} F_X(x) = F_X(a) - F_X(a^-)$

Remarque 1.3.16 On verra plus loin des cas agréables dans lesquels F_X est continue sur \mathbb{R} .

Démonstration : 1. Si $x \leq y,]-\infty, x] \subset]-\infty, y]$ d'où $F_X(x) = \mathbf{P}_X(\cdot - \infty, x] \leq \mathbf{P}_X(\cdot - \infty, y] = F_X(y)$.

2. Pour tout $x \in \mathbb{R}, F_X(x)$ est la probabilité d'un ensemble.

3. Soit $x \in \mathbb{R}$. On veut montrer que $\lim_{y \rightarrow x^+} F_X(y) = F_X(x)$. Comme F_X est croissante, il suffit de montrer que $\lim_{n \rightarrow \infty} F_X(x + \frac{1}{n}) = F_X(x)$. (En effet, supposons que $F_X(x + \frac{1}{n}) \rightarrow F_X(x)$ quand $n \rightarrow \infty$. Alors pour tout $\varepsilon > 0$, il existe $N > 0$ tq pour tout $n \geq N, F_X(x) \leq F_X(x + \frac{1}{n}) \leq F_X(x) + \varepsilon$. Comme F_X est croissante, pour tout $x \leq y \leq x + \frac{1}{n}, F_X(x) \leq F_X(y) \leq F_X(x + \frac{1}{n}) \leq F_X(x) + \varepsilon$. Donc $\lim_{y \rightarrow x^+} F_X(y) = F_X(x)$.)

Remarquons que $F_X(x + \frac{1}{n}) = \mathbf{P}_X(\cdot - \infty, x + \frac{1}{n}]$. La suite d'intervalles $]-\infty, x + \frac{1}{n}]$ est décroissante, la probabilité \mathbf{P}_X vérifie donc $\lim_{n \rightarrow \infty} \mathbf{P}_X(\cdot - \infty, x + \frac{1}{n}] = \mathbf{P}_X(\cap_{n \in \mathbb{N}}]-\infty, x + \frac{1}{n}] = \mathbf{P}_X(\cdot - \infty, x] = F_X(x)$.

4. Considérons cette fois la suite d'ensembles $]-\infty, x - \frac{1}{n}]$. C'est une suite croissante d'intervalles. On a donc $\lim_{n \rightarrow \infty} F_X(x - \frac{1}{n}) = \mathbf{P}_X(\cdot - \infty, x]$. (Alors que $F_X(x) = \mathbf{P}_X(\cdot - \infty, x]$.) Comme F_X est croissante, on en déduit que F_X admet $\mathbf{P}_X(\cdot - \infty, x]$ comme limite à gauche en x . (Remarquons que l'existence de la limite découle immédiatement du fait que si $y < x$ l'application $y \mapsto F_X(y)$ est croissante et majorée par $F_X(x)$.)

5. On a $\emptyset = \cap_{n \in \mathbb{N}}]-\infty, -n]$. Le même raisonnement que ci-dessus donne $\lim_{n \rightarrow \infty} F_X(-n) = 0$. Comme F_X est croissante, ceci implique (vérifier!) que $\lim_{x \rightarrow -\infty} F_X(x) = 0$. Pour la limite en $+\infty$ considérer les intervalles $]-\infty, n]$.

6 et 7. Exo immédiat. \square

Proposition 1.3.17 Soient P_1 et P_2 deux mesures de probabilité sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Si elles ont même fonction de répartition $F_1 = F_2 = F$, alors elles sont égales.

Démonstration : Pour tout intervalle $]a, b]$, on a $\mathbf{P}_1(\cdot - \infty, b] - \mathbf{P}_1(\cdot - \infty, a] = F(b) - F(a) = \mathbf{P}_2(\cdot - \infty, b] - \mathbf{P}_2(\cdot - \infty, a])$. La classe des intervalles du type $]a, b]$, $a \in \mathbb{R}, b \in \mathbb{R}$ est stable par intersection finie et engendre la tribu borélienne (voir les TD du module intégration. Par intersection/union dénombrable, il est clair que la tribu engendrée contient tous les intervalles, et donc tous les boréliens.)

Le corollaire 1.2.4 permet de conclure que $\mathbf{P}_1 = \mathbf{P}_2$. \square

Proposition 1.3.18 Soient X, Y de variables aléatoires définies respectivement sur $(\Omega, \mathcal{A}, \mathbf{P})$ et $(\Omega', \mathcal{A}', \mathbf{P}')$, à valeurs réelles ou entières. Si X et Y ont même fonction de répartition, alors X et Y ont même loi.

Démonstration : Reformulation immédiate de la proposition précédente. □

Proposition 1.3.19 Si F est une fonction de \mathbb{R} dans $[0, 1]$ qui est croissante, continue à droite, et admet pour limite 0 en $-\infty$ et 1 en $+\infty$, alors il existe une mesure de probabilité \mathbf{P} sur \mathbb{R} dont F est la fonction de répartition.

Démonstration : L'idée de la preuve est élémentaire. On définit \mathbf{P} sur tout intervalle $]a, b]$ par $\mathbf{P}(]a, b]) = F(b) - F(a)$. On a bien $\mathbf{P}(\emptyset) = 0$ et $\mathbf{P}(\mathbb{R}) = 1$. La famille des intervalles $]a, b]$ est une algèbre : stable par intersection et union finie, passage au complémentaire, et contient \emptyset et \mathbb{R} . Cette algèbre engendre la tribu borélienne. Pour conclure la preuve, il suffit d'utiliser le théorème ci-dessous 1.3.20. □

Théorème 1.3.20 (Théorème de prolongement de Carathéodory) Si m est une mesure bornée définie sur une algèbre \mathcal{B} de parties de E , il existe une unique mesure prolongeant m sur la tribu $\sigma(\mathcal{B})$ engendrée par \mathcal{B} .

Démonstration : L'unicité découle du corollaire 1.2.4. L'existence sera démontrée en Master I. □

Proposition 1.3.21 Une fonction de répartition a au plus un nombre dénombrable de discontinuités.

Démonstration : Voir TD □

Proposition 1.3.22 Soit F la fonction de répartition de la probabilité μ sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. On définit un « pseudo-inverse » de F par $F^{-1}(t) = \inf\{x \in \mathbb{R}, F(x) \geq t\}$. Soit U une variable aléatoire de loi uniforme sur $[0, 1]$. Alors $Y = F^{-1}(U)$ est une variable aléatoire de loi μ .

Ceci sert en simulation : si on sait simuler une v.a.r. de loi uniforme, alors on sait simuler n'importe quelle loi sur \mathbb{R} .

Démonstration : Voir TD □

Définition 1.3.23 Soit $X = (X_1, \dots, X_d)$ un vecteur aléatoire à valeurs dans \mathbb{R}^d . Sa fonction de répartition est définie par

$$F_X(x_1, \dots, x_d) = \mathbf{P}_X([-\infty, x_1] \times \dots \times [-\infty, x_d]) = \mathbf{P}(X_1 \leq x_1 \text{ et } \dots X_d \leq x_d).$$

Proposition 1.3.24 Soient X_1 et X_2 deux vecteurs aléatoires réels, à valeurs dans \mathbb{R}^d . S'ils ont même fonction de répartition, alors ils ont même loi.

Démonstration : Ceci se fait exactement comme la proposition 1.3.18. □

1.4 Lois et variables aléatoires classiques

En pratique, on modélise souvent les phénomènes aléatoires de la vie réelle par quelques lois qu'on connaît bien, et que nous allons revoir ici. Les lois discrètes ont été vues au S4, et doivent être connues.

1.4.1 Lois discrètes

Le pile ou face, loi uniforme

On lance une pièce et lorsqu'elle retombe, on regarde si la face visible est "face" (la face de Marianne, sur les francs d'autrefois) ou "pile" (la semeuse, de l'autre côté des francs). On suppose que la pièce a autant de chances de tomber sur pile que sur face. On introduit alors $\Omega = \{\text{pile}, \text{face}\}$, la tribu est $\mathcal{P}(\Omega)$ et la probabilité \mathbf{P} définie par $\mathbf{P}(\text{pile}) = \mathbf{P}(\text{face}) = \frac{1}{2}$.

Une variante mathématique est de regarder l'espace $\{0, 1\}$ muni de la tribu $\mathcal{P}(\{0, 1\})$ et de la probabilité $\mathbf{P} = \frac{1}{2}\delta_0 + \frac{1}{2}\delta_1$. Si une variable aléatoire X est à valeurs dans $\{0, 1\}$ muni de \mathbf{P} on dit que X suit une loi uniforme sur $\{0, 1\}$.

Sa fonction de répartition vérifie $F_X(x) = 0$ si $x < 0$, $F_X(x) = \frac{1}{2}$ si $x \in [0, 1[$ et $F_X(x) = 1$ si $x \geq 1$.

Plus généralement, si X est une v.a. à valeurs dans un ensemble fini E , on dit qu'elle suit une loi uniforme sur E si pour tout $\omega \in E$, on a $P_X(\{\omega\}) = \frac{1}{\#E}$. On dit alors que chacun des événements élémentaires $\{\omega\}$ est équiprobable.

Un autre exemple classique est celui du lancer de dé équilibré. L'ensemble à considérer est $\{1, 2, 3, 4, 5, 6\}$ muni de la probabilité uniforme qui assigne à chaque événement élémentaire la probabilité $\frac{1}{6}$.

Exercice 1.4.1 Tracer le graphe de la fonction de répartition d'une loi de pile ou face sur $\{0, 1\}$ puis d'une loi uniforme sur $\{1, \dots, 6\}$.

Marches aléatoires

Imaginons un-e ivrogne qui se promène le long d'une rue (orientée est ouest), et tous les 10 mètres, il tire à pile ou face pour savoir s'il va à l'ouest (pile) ou à l'est (face). Cette marche dite aléatoire est modélisée de la façon suivante : on imagine que l'ivrogne se promène sur \mathbb{Z} ; à l'instant $n \geq 0$, il se trouve à la position $S_n \in \mathbb{Z}$ (avec $S_0 = 0$ pour simplifier). Entre l'instant n et l'instant $n + 1$ il tire à pile ou face, on note X_{n+1} la variable aléatoire qui vaut $+1$ s'il tire face (il va à droite) et -1 sinon. Sa position à l'instant $n + 1$ est alors $S_{n+1} = S_n + X_{n+1}$. Autrement dit, on modélise cette marche par une suite $(X_n)_{n \geq 1}$ de variables aléatoires à valeurs dans $\{+1, -1\}$, de loi uniforme sur $\{+1, -1\}$. La position à l'instant n est donnée par $S_n = \sum_{k=1}^n X_k$. Le dernier chapitre nous donnera des outils pour étudier le comportement de S_n quand n grand (retour en 0? divergence en $\pm\infty$...)

Plus généralement, on peut effectuer des marches aléatoires sur \mathbb{Z}^d , $d \geq 2$. Alors X_n est une variable aléatoire à valeurs dans $\{\pm 1\}^d$, et S_n est une v.a. à valeurs dans \mathbb{Z}^d .

Remarque 1.4.2 Les marches aléatoires sur \mathbb{Z}^d ont un comportement bien connu depuis longtemps, mais les chercheurs étudient encore très activement les marches aléatoires sur des groupes G non commutatifs, comme des groupes de matrices par exemple. On parle alors de produits de matrices aléatoires. À chaque instant n , au lieu d'ajouter à S_{n-1} un élément X_n qui fait partie des générateurs de \mathbb{Z}^d et de leurs inverses, on multiplie (à gauche ou à droite, ce n'est pas pareil, il faut choisir !) S_{n-1} par un générateur X_n du groupe non commutatif G (noté multiplicativement au lieu d'additivement pour \mathbb{Z}^d). On se promène alors aléatoirement dans G , et on voudrait bien savoir si on va revenir à l'origine ou pas (ce qui dépend de la nature et de la taille du groupe !)

Cette théorie des matrices aléatoires intéresse les mathématiciens pour sa beauté intrinsèque d'une part, mais aussi pour ses intérêts en modélisation. Plusieurs chercheurs-euses du LAMFA travaillent par exemple à modéliser le développement du Prunus en forêt de Compiègne à l'aide de produits aléatoires de matrices.

Fin cours 2 (2h, le 28 janvier 2009)

Loi de Bernoulli

La loi de Bernoulli de paramètre p , notée parfois $\mathcal{B}(p)$ est la loi sur $\{0, 1\}$ (ou autre ensemble à deux éléments) définie par $\mathbf{P}(1) = p$ et $\mathbf{P}(0) = 1 - p$, p étant un paramètre fixé dans $[0, 1]$. Autrement dit $\mathbf{P} = p\delta_1 + (1 - p)\delta_0$.

Par exemple, on tire une boule dans une urne contenant R boules rouges et N noires, soit une proportion $p = \frac{R}{R+N}$ de boules rouges, et $1 - p$ de boules noires. On considère $\Omega = \{1, \dots, R + N\}$ l'ensemble des boules numérotées, et X la variable aléatoire définie sur Ω qui vaut 1 si la boule est rouge et 0 sinon. En supposant que chaque boule a une probabilité identique d'être attrapée, la v.a. X suit une loi de Bernoulli de paramètre p .

Exercice 1.4.3 Tracer la fonction de répartition d'une loi de Bernoulli de paramètre $\frac{1}{3}$.

Loi binomiale

Une urne contient R boules rouges, N boules noires, soit encore une proportion $p = \frac{R}{R+N}$ de boules rouges. Si on tire n boules avec remise (on remet la boule dans l'urne à chaque fois), l'espace des états est alors $\Omega = \{1, \dots, R+N\}^n$; soit X la v.a. qui à un tirage aléatoire $\omega \in \Omega$ de n boules avec remise associe le nombre de boules rouges obtenues. Un calcul élémentaire montre que $\mathbf{P}_X(k) = \mathbf{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$ pour tout $0 \leq k \leq n$. On dit que X suit une loi binomiale de paramètres (n, p) parfois encore notée $\mathcal{B}(n, p)$.

Exercice 1.4.4 Vérifier que \mathbf{P}_X est une probabilité sur $\{0, 1, \dots, n\}$.

Exercice 1.4.5 Tracer le graphe de la fonction de répartition d'une loi binomiale de paramètres $n = 4$ et $p = \frac{1}{3}$.

Loi hypergéométrique

C'est la loi de la variable aléatoire X comptant le nombre de boules rouges obtenues lors d'un tirage de n boules sans remise dans une urne contenant R rouges et N noires. On la note parfois $\mathcal{H}(n, N, R)$, et on vérifie que $\mathbf{P}_X(k) = \mathbf{P}(X = k) = \frac{C_R^k C_N^{n-k}}{C_{R+N}^n}$.

Loi de Poisson

Notée parfois $\mathcal{P}(\lambda)$. Loi de Poisson de paramètre $\lambda \in \mathbb{R}$, c'est la mesure de probabilité sur \mathbb{N} définie par $\mathbf{P}(n) = \frac{\lambda^n}{n!} e^{-\lambda}$.

Exercice 1.4.6 Vérifier que \mathbf{P} est une probabilité sur \mathbb{N}

Exercice 1.4.7 Tracer le graphe de la fonction de répartition d'une loi de Poisson de paramètre 1.

La loi de Poisson est la loi des événements rares.

Par exemple, on sait que le temps moyen d'attente à un guichet (en nombre de minutes) vaut $\lambda > 0$. Alors la probabilité d'attendre k minutes est bien modélisée par une loi de Poisson de paramètre λ .

De même, si on sait qu'une ampoule dure en moyenne λ jours, alors on peut modéliser le temps entre deux pannes par une loi de Poisson de paramètre λ .

Autre exemple pratique : on prélève (sans remise donc) n individus dans une population qui contient des individus de deux types A et B (malades/sains, bactéries mutantes/ordinaires, etc). Le type A est présent en proportion p très faible, le type B est en proportion $1-p$ proche de 1. Le nombre moyen d'individus A prélevés vaut $\lambda = np$. Si n est grand, p petit et $1 \leq \lambda = np \leq 10$, la v.a. représentant le nombre d'individus de type A obtenus suit théoriquement une loi hypergéométrique, et en pratique on peut supposer que c'est une loi binomiale (quand n est très grand, un tirage avec ou sans remise importe peu). En fait, on peut même considérer que c'est une loi de Poisson de paramètre $\lambda = np$. Ceci se démontre en utilisant la formule de Stirling $n! \sim \sqrt{2\pi n} (n/e)^n$ quand $n \rightarrow \infty$.

Loi géométrique

La loi géométrique de paramètre $\alpha \in]0, 1[$ est la mesure de probabilité sur \mathbb{N} définie par $\mathbf{P}(n) = (1-\alpha)\alpha^n$. C'est la probabilité d'obtenir une boule noire après n boules rouges lorsque la probabilité d'obtenir une boule rouge est α . (NB : la suite $(\mathbf{P}(n))_{n \in \mathbb{N}}$ est une suite géométrique).

Il arrive (dans certains livres) que les paramètres α et $1-\alpha$ soient échangés, ou bien même la loi soit définie sur $\mathbb{N}^* = \mathbb{N} \setminus \{0\}$ par $\mathbf{P}(n) = \alpha(1-\alpha)^{n-1}$, $n \geq 1$. Par conséquent, on précise toujours les notations lorsqu'on parle de loi géométrique.

Exercice 1.4.8 Vérifier que \mathbf{P} est une probabilité sur \mathbb{N} (ou \mathbb{N}^*) pour chacune des conventions de notation ci-dessus.

Exercice 1.4.9 Tracer la fonction de répartition d'une loi géométrique de paramètre $\frac{1}{2}$.

1.4.2 Lois continues

Définition 1.4.10 Une v.a.r. a une loi continue si sa fonction de répartition est continue.

Les v.a. ci-dessous sont continues. Vérifiez-le sur chaque exemple.

Loi uniforme

La loi uniforme sur l'intervalle $[a, b] \subset \mathbb{R}$ muni de sa tribu borélienne) est la mesure de probabilité définie pour tout $A \in \mathcal{B}([a, b])$ par

$$\mathbf{P}(A) = \frac{1}{b-a} \int_A dx = \frac{1}{b-a} \lambda(A)$$

où λ est la mesure de Lebesgue sur \mathbb{R} .

Sa fonction de répartition vérifie $F(x) = \frac{1}{b-a} \int_{-\infty, x] \cap [a, b]} dx$. D'où $F(x) = 0$ si $x \leq a$, $F(x) = \frac{x-a}{b-a}$ si $a \leq x \leq b$, $F(x) = 1$ après. Dessinez le graphe de F . Qu'observez-vous ?

Notez que rien ne change si on considère l'intervalle (a, b) ouvert en a ou en b .

Plus généralement, on peut considérer la loi uniforme sur un ensemble E borélien quelconque de \mathbb{R} , définie pour tout $A \in \mathcal{B}(E) \subset \mathcal{B}(\mathbb{R})$ par $\mathbf{P}(A) = \frac{\lambda(A)}{\lambda(E)}$.

La même définition est valide dans \mathbb{R}^d , $d \geq 2$, en remplaçant la mesure de Lebesgue sur \mathbb{R} par la mesure de Lebesgue sur \mathbb{R}^d .

Loi gaussienne, ou loi normale

D'une certaine façon, c'est la loi de probabilité la plus importante qui existe. Vous comprendrez pourquoi au dernier chapitre, dans le théorème limite central (TLC). Cette loi « universelle » modélise les écarts à la moyenne. Typiquement lorsqu'on mesure une grandeur physique qui a une valeur théorique attendue, il y a toujours une erreur, dont les causes peuvent être multiples et dues à des sommes de nombreuses causes microscopiques ou macroscopiques (instrument de mesure, précision de l'expérience, etc) et qu'on modélise souvent par une variable aléatoire de loi normale. En d'autres termes, la loi normale modélise des erreurs macroscopiques dues à une somme d'erreurs microscopiques. Vous comprendrez pourquoi lorsqu'on étudiera le TLC.

La loi normale centrée réduite notée $\mathcal{N}(0, 1)$ sur \mathbb{R} donne à tout borélien A de \mathbb{R} la probabilité

$$\mathbf{P}(A) = \frac{1}{\sqrt{2\pi}} \int_A e^{-x^2/2} dx$$

Exercice 1.4.11 Vérifier que \mathbf{P} est une probabilité sur \mathbb{R} .

La loi normale de moyenne m , d'écart-type σ (variance σ^2) est notée $\mathcal{N}(m, \sigma)$. Elle est définie par

$$\mathbf{P}(A) = \frac{1}{\sqrt{2\pi}\sigma} \int_A e^{-\frac{(x-m)^2}{2\sigma^2}} dx.$$

Loi exponentielle

La loi exponentielle de paramètre a , notée parfois $\mathcal{E}(a)$, vérifie pour tout $A \in \mathcal{B}(\mathbb{R}_+)$

$$\mathbf{P}(A) = \int_A a e^{-ax} \mathbf{1}_{\mathbb{R}_+}(x) dx.$$

Exercice 1.4.12 Vérifier que \mathbf{P} est une probabilité sur \mathbb{R} .

Exercice 1.4.13 Tracer le graphe de la fonction de répartition d'une loi exponentielle de paramètre 1.

Cette loi permet de modéliser les processus sans mémoire, qui ne vieillissent pas, les durées de vie. Par exemple, c'est la loi de la variable aléatoire X qui décrit le temps que met une particule instable à se désintégrer. Elle vérifie la propriété suivante.

Proposition 1.4.14 Soit X une v.a. réelle. Alors X suit une loi exponentielle si et seulement si pour tous $x > 0$ et $y > 0$,

$$\mathbf{P}(X > x + y | X > y) = \mathbf{P}(X > x).$$

Démonstration : EXO. Voir TD □

La durée de vie d'un néon suit une loi exponentielle. Par conséquent, inutile de changer un néon qui fonctionne encore...

Loi de Cauchy

Elle est définie pour tout $A \in \mathcal{B}(\mathbb{R}_+)$ par

$$\mathbf{P}(A) = \frac{1}{\pi} \int_A \frac{1}{1+x^2} dx.$$

Exercice 1.4.15 Vérifier que \mathbf{P} est une probabilité sur \mathbb{R} .

Nous nous en servons souvent comme contre-exemple.

Si U suit une loi uniforme sur $] -\frac{\pi}{2}, \frac{\pi}{2}[$ alors $\tan U$ suit une loi de Cauchy.

1.5 Variables aléatoires à densité

1.5.1 Définitions et premières propriétés

Définition 1.5.1 Soit μ une probabilité définie sur $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. Elle est à densité, ou absolument continue par rapport à la mesure de Lebesgue λ s'il existe une fonction $f : \mathbb{R} \rightarrow \mathbb{R}_+$ Lebesgue-intégrable et positive, telle que pour tout borélien A de \mathbb{R} ,

$$\mu(A) = \int_A f d\lambda.$$

On dit que f est une densité de μ par rapport à λ .

Remarquons qu'une densité f vérifie

$$\int_{\mathbb{R}} f d\lambda = 1.$$

Remarque 1.5.2 Une mesure de probabilité peut avoir plusieurs densités. Il n'y a donc pas unicité. Cela dit, si f et g sont deux densités de μ alors $f = g$ μ -presque sûrement. Par exemple, la loi uniforme sur $[a, b]$ a pour densités (entre autres) $\frac{1}{b-a} \mathbf{1}_{[a,b]}$ mais aussi $\frac{1}{b-a} \mathbf{1}_{]a,b]}$, ou $\frac{1}{b-a} \mathbf{1}_{[a,b[}$, ou ... Cela dit, il y a unicité de la densité dans $L^1(\mathbb{R}, \mathcal{B}(\mathbb{R}), \lambda)$.

En fait, cette définition a parfaitement un sens pour des mesures positives μ quelconques, et pas nécessairement des mesures de probabilité. Dans ce cas, la seule chose qui change est que la densité f est positive mesurable mais non nécessairement Lebesgue-intégrable.

Remarque 1.5.3 La définition est inchangée dans \mathbb{R}^d , $d \geq 2$. Relisez tout le paragraphe en vérifiant ce qui s'applique dans \mathbb{R}^d , $d \geq 2$.

Exercice 1.5.4 Vérifier que toutes les lois du paragraphe 1.4.2 sont à densité, et donner leur densité.

Proposition 1.5.5 Si \mathbf{P} est une mesure de probabilité à densité f sur \mathbb{R} , alors sa fonction de répartition F est continue sur \mathbb{R} et vérifie pour tous $a \leq b$ réels

$$F(b) - F(a) = \mathbf{P}(]a, b]) = \int_{]a, b]} f d\lambda$$

Démonstration : Par définition, $F(b) - F(a) = \mathbf{P}(] - \infty, b]) - \mathbf{P}(] - \infty, a])$. L'additivité de \mathbf{P} donne $F(b) - F(a) = \mathbf{P}(]a, b]) = \int_{]a, b]} f d\lambda$ (la dernière égalité résultant du fait que \mathbf{P} est à densité).

Cette expression permet de déduire la continuité de F . En effet, $F(x) = F(a) + \int_{]a, x]} f d\lambda$ pour n'importe quel a fixé. Et votre cours d'intégration vous permet d'affirmer qu'une fonction définie par une intégrale est continue. (Exo : vérifier cela dans votre cours d'intégration). \square

Proposition 1.5.6 Si X a une fonction de répartition F qui est C^1 (respectivement C^1 par morceaux, i.e. C^1 en dehors d'un nombre fini de discontinuités), alors X a pour densité $f = F'$, et f est continue (resp. continue par morceaux).

Démonstration : Soit $f = F'$. Cette fonction est définie partout sauf au plus en un nombre fini de points, elle est continue par morceaux, positive (car F est croissante), et $\int_{\mathbb{R}} f d\lambda = F(\infty) - F(-\infty) = 1$. Soit donc μ la mesure de probabilité de densité $f = F'$. Cette mesure coïncide avec \mathbf{P}_X sur tous les intervalles $]a, b]$. Le corollaire 1.2.4 donne $\mu = \mathbf{P}_X$. \square

Exercice 1.5.7 Donner l'allure du graphe des densités et des fonctions de répartition de tous les exemples du paragraphe 1.4.2.

Remarque 1.5.8 Si μ est à densité, f est mesurable mais pas nécessairement continue. C'est F qui est continue.

Remarque 1.5.9 La réciproque de la proposition 1.5.5 est fautive : il existe des lois μ dont la fonction de répartition est continue mais qui n'ont pas de densité. Autrement dit, il existe des lois continues mais pas absolument continues.

Une façon d'obtenir de telles lois est d'utiliser la construction ci-dessous. On construit une fonction F continue croissante sur $[0, 1]$, avec $F(0) = 0$ et $F(1) = 1$, mais qui est constante presque partout ! Si la loi correspondante était à densité, sa densité serait nulle, donc la loi serait nulle, contradiction avec le fait que c'est une probabilité.

L'un de ces monstres, appelé *escalier du diable*, est construit de la manière suivante.

À l'étape 0, on définit $F_0(x) = 0$ si $x \leq 0$, $F_0(x) = 1$ si $x \geq 1$ et $F_0(x) = x$ sur $[0, 1]$. F_0 est continue, croissante, et a pour limites 0 en $-\infty$ et 1 en $+\infty$.

À l'étape 1, on découpe $[0, 1]$ en trois intervalles égaux de longueur $\frac{1}{3}$. Sur l'intervalle du milieu $[1/3, 2/3]$ on pose $F_1(x) = \frac{1}{3}$. Sur les deux intervalles $[0, 1/3]$ et $[2/3, 1]$, on prolonge F_1 en une fonction affine par morceaux et continue : $F_1(x) = \frac{3}{2}x$ sur $[0, 1/3]$ et $F_1(x) = \frac{3}{2}x - \frac{1}{2}$ sur $[2/3, 1]$.

À l'étape 2, on recommence sur chacun des intervalles $[0, 1/3]$ et $[2/3, 1]$. (En dehors, sur $] - \infty, 0] \cup [1/3, 2/3] \cup [1, +\infty[$, on garde $F_2(x) = F_1(x)$.) Sur l'intervalle $[1/9, 2/9]$ on pose $F_2(x) = \frac{1}{4}$, sur l'intervalle $[7/9, 8/9]$ $F_2(x) = \frac{3}{4}$ et sur les quatre intervalles $[0, \frac{1}{9}]$, $[\frac{2}{9}, \frac{1}{3}]$, $[\frac{2}{9}, \frac{7}{9}]$, $[\frac{8}{9}, 1]$, on prolonge F_2 en une fonction affine par morceaux continue croissante.

Et on continue la construction... Par récurrence, on construit une suite (F_n) de fonctions continues sur \mathbb{R} , croissantes, valant 0 avant 0 et 1 après 1, et qui sont constantes par morceaux sur une union d'intervalles I_n de longueur de plus en plus grande.

À la limite, (F_n) converge. En effet, il est élémentaire de vérifier que pour tout $n \geq 1$, on a $\|F_{n+1} - F_n\| \leq \frac{1}{2^n}$. On en déduit aisément que pour tout $p \geq 1$, on a $\|F_{n+p} - F_n\| \leq \frac{1}{2^{n-p}}$. En particulier, $(F_n)_{n \in \mathbb{N}}$ est une suite de Cauchy pour la topologie de la norme infinie sur les fonctions continues sur $[0, 1]$. Elle converge donc et sa limite, notée F est continue. On l'appelle *l'escalier du diable* : elle croît sur $[0, 1]$ de 0 à 1, de façon continue, en étant constante presque partout ! C'est la fonction de répartition d'une mesure de probabilité \mathbf{P} qui n'admet pas de densité.

Voici une proposition peu difficile dont l'intérêt apparaîtra au paragraphe 1.5.2.

Proposition 1.5.10 La variable aléatoire réelle X a pour densité la fonction mesurable positive f si et seulement si pour toute fonction $h : \mathbb{R} \rightarrow \mathbb{R}$ mesurable bornée, on a

$$\int_{\mathbb{R}} h(x) d\mathbf{P}_X(x) = \int_{\mathbb{R}} h(x) f(x) d\lambda(x)$$

Démonstration : Si pour toute $h : \mathbb{R} \rightarrow \mathbb{R}$ mesurable bornée, $\int_{\mathbb{R}} h(x) d\mathbf{P}_X(x) = \int_{\mathbb{R}} h(x) f(x) d\lambda(x)$, alors pour $h = \mathbf{1}_A$, A borélien quelconque, on obtient $\mathbf{P}_X(A) = \int_A f(x) d\lambda(x)$. Donc \mathbf{P}_X est une loi à densité f .

Réciproquement, si \mathbf{P}_X a pour densité f , l'égalité cherchée est vraie pour toutes les fonctions $h = \mathbf{1}_A$ indicatrices de boréliens A (par définition d'une loi à densité). Par linéarité, l'égalité est encore vraie pour toutes les fonctions simples $h = \sum_i c_i \mathbf{1}_{A_i}$ (somme finie). (Par convergence monotone, l'égalité est vraie pour toutes les fonctions h mesurables positives. Inutile ici, mais vrai.) Par convergence dominée, l'égalité est vraie pour toutes les fonctions intégrables, et en particulier toutes les fonctions mesurables bornées. \square

Exercice 1.5.11 Soit $X : \Omega = [0, 1] \rightarrow \mathbb{R}$ définie par $X(\omega) = 2\omega + 1$. Donner la loi (et la densité) de X .

On calcule $\int_{\mathbb{R}} h d\mathbf{P}_X = \int_{[0,1]} h(2\omega + 1) d\omega = \int_{[0,3]} h(w) \frac{dw}{2} = \int_{\mathbb{R}} h(x) \frac{\mathbf{1}_{[0,3]}(x)}{2} dx$. La densité de X est donc $\frac{\mathbf{1}_{[0,3]}}{2}$.

Finissons ce paragraphe par une question que se posent souvent les mathématicien-nes : qu'est-ce qui caractérise une densité ? Autrement dit, quelles fonctions sont des densités de lois de probabilité ?

Proposition 1.5.12 Si $f : (\mathbb{R}, \mathcal{B}(\mathbb{R})) \rightarrow (\mathbb{R}_+, \mathcal{B}(\mathbb{R}_+))$ est une fonction mesurable positive avec $\int_{\mathbb{R}} f d\lambda = 1$ alors il existe une mesure de probabilité μ sur \mathbb{R} qui admet f pour densité.

Démonstration : Fait dans le module intégration. L'idée est complètement élémentaire : si A est un borélien de \mathbb{R} , on définit $\mu(A) = \int_A f d\lambda$. Vérifiez en exercice que cela définit bien une mesure de probabilité (à densité évidemment) sur \mathbb{R} . \square

1.5.2 Changement de variables

Dans ce paragraphe, nous partons d'une variable aléatoire X à densité, et nous cherchons à savoir si la variable aléatoire $Y = \varphi(X)$ est à densité, et si oui, quelle est sa densité?

Nous utiliserons la proposition 1.5.10. La stratégie est la suivante. On sait que X a pour densité f_X . On applique cette proposition 1.5.10. Soit h une fonction mesurable bornée ; alors $h \circ \varphi$ est encore mesurable bornée. On déduit de 1.5.10 que pour toute fonction mesurable bornée du type $h \circ \varphi$, on a

$$\int_{\Omega} h \circ \varphi \circ X(\omega) d\mathbf{P}(\omega) = \int_{\mathbb{R}} h \circ \varphi(x) d\mathbf{P}_X(x) = \int_{\mathbb{R}} h \circ \varphi(x) f_X(x) d\lambda(x).$$

Remarquons que le premier terme vaut encore $\int_{\Omega} h \circ Y(\omega) d\mathbf{P}(\omega) = \int_{\mathbb{R}} h(y) d\mathbf{P}_Y(y)$.

On essaie alors d'appliquer le théorème du changement de variables (en posant $y = \varphi(x)$) au terme $\int_{\mathbb{R}} h \circ \varphi(x) f_X(x) d\lambda(x)$, en espérant le mettre sous la forme $\int_{\mathbb{R}} h(y) f_Y(y) d\lambda(y)$, avec f_Y une fonction mesurable positive. La proposition 1.5.10 permet alors de conclure que f_Y est la densité de \mathbf{P}_Y .

Voyons cela sur un exemple très détaillé. Cherchons la densité de la loi de $Y = X + 1$, si X est une v.a. à densité f_X . On veut utiliser la proposition 1.5.10. On étudie donc la quantité $\int_{\mathbb{R}} h \circ Y(y) d\mathbf{P}_Y(y)$ qui vaut $\int_{\Omega} h \circ Y(\omega) d\mathbf{P}(\omega)$ (théorème 1.3.8), puis $\int_{\Omega} h \circ \varphi \circ X(\omega) d\mathbf{P}(\omega)$ avec $\varphi(x) = x + 1$. Cette intégrale vaut encore (théorème 1.3.8 appliqué à X) $\int_{\mathbb{R}} h \circ \varphi(x) d\mathbf{P}_X(x)$, puis $\int_{\mathbb{R}} h(x+1) f_X(x) dx$ (car X est à densité). Le théorème de changement de variables donne $\int_{\mathbb{R}} h(x+1) f_X(x) dx = \int_{\mathbb{R}} h(x) f_X(x-1) dx$. Finalement, pour toute fonction h mesurable bornée, on a montré que $\int_{\mathbb{R}} h(y) d\mathbf{P}_Y(y) = \int_{\mathbb{R}} h(y) f_X(y-1) dy$. La proposition 1.5.10 nous permet de conclure que $Y = X + 1$ est à densité et que $f_Y(y) = f_X(y-1)$.

Rappelons le théorème du changement de variables :

Théorème 1.5.13 (Changement de variable) Soit $\varphi : U \subset \mathbb{R}^d \rightarrow V \subset \mathbb{R}^d$ un difféomorphisme de classe C^1 entre deux ouverts, et $g : V \rightarrow \mathbb{R}$ Lebesgue intégrable. Alors

$$\int_V g(y) dy = \int_U g(\varphi(x)) |\text{Jac } \varphi(x)| dx.$$

Pour la démonstration, voir votre cours d'intégration.

Un exemple fondamental : le passage en coordonnées polaires

Soit $U =]0, 2\pi[\times \mathbb{R}_+^*$ et $V = \mathbb{R}^2 \setminus \{(x, 0), x \geq 0\}$. Et $\varphi(r, \theta) = (r \cos \theta, r \sin \theta)$. Bien sûr, $|\text{Jac } \varphi(r, \theta)| = r$. Si $g : V \rightarrow \mathbb{R}$ est intégrable ou positive, alors

$$\int_V g(x, y) dx dy = \int_U g(r \cos \theta, r \sin \theta) dr r d\theta.$$

Application : calcul de $\int_{\mathbb{R}^2} e^{-x^2/2} dx$. Indication : Passer en coordonnées polaires dans l'intégrale $\int_{\mathbb{R}^2} e^{-\frac{x^2+y^2}{2}} dx dy$.

Remarque très importante en pratique : ci-dessus, l'application $\varphi : \mathbb{R}^2 \rightarrow \mathbb{R}^2$ n'est pas un difféomorphisme (car pas bijective) mais quitte à enlever des ensembles de mesure nulle, l'application $\varphi|_U$ est un difféomorphisme.

Reformulation de ce qui précède

On peut donner un énoncé de changement de variable probabiliste, comme ci-dessous. Néanmoins, il est sans doute plus utile de retenir et savoir refaire sa preuve que de retenir son énoncé.

Proposition 1.5.14 Soit φ un difféomorphisme d'un ouvert U de \mathbb{R}^d dans un ouvert V de \mathbb{R}^d . Si X est une v.a. qui prend p.s. ses valeurs dans U et qui a pour densité p , alors la v.a. $Y = \varphi \circ X$ a pour densité $p \circ \varphi^{-1} \times |J_{\varphi^{-1}}|$.

Démonstration : Soit h une fonction mesurable bornée définie sur V . A l'aide du théorème de transfert et du théorème de changement de variable, on montre que

$$\begin{aligned} \int_V h(y) d\mathbf{P}_Y &= \int_{\Omega} h \circ Y(\omega) d\mathbf{P}(\omega) = \int_{\Omega} h \circ \varphi \circ X(\omega) d\mathbf{P}(\omega) = \int_U h \circ \varphi(x) d\mathbf{P}_X(x) \\ &= \int_U h \circ \varphi(x) p(x) dx = \int_V h(y) p \circ \varphi^{-1}(y) |J_{\varphi^{-1}}(y)| dy \end{aligned}$$

La proposition 1.5.10 permet de conclure. \square

Exercice 1.5.15 Soit X une v.a.r. de loi exponentielle et $Y = X^2$. Donner la densité de Y .

Soit $\varphi : x \mapsto x^2$. C'est un difféo de \mathbb{R}_+^* dans \mathbb{R}_+^* , et de \mathbb{R}_-^* dans \mathbb{R}_+^* . Et on a

$$\begin{aligned} \int_{\mathbb{R}} h(y) d\mathbf{P}_Y(y) &= \int_{\Omega} h \circ Y d\mathbf{P} = \int_{\Omega} h \circ \varphi \circ X d\mathbf{P} = \int_{\mathbb{R}} h \circ \varphi(x) f_X(x) dx \\ &= \int_{\mathbb{R}_+^*} + \int_{\mathbb{R}_-^*} \dots = \int_{\mathbb{R}_+^*} h(y) \frac{1}{2\sqrt{y}} (f_X(\sqrt{y}) + f_X(-\sqrt{y})) dy \end{aligned}$$

Exercice 1.5.16 Exemple à détailler. Soit X une variable aléatoire réelle de densité $f_X(x) = e^{-x} 1_{\mathbb{R}^+}(x)$, $Y = |\log X|$, et $\varphi(x) = |\log x|$. On pose $\varphi_1(x) = \log x$ de $U_1 =]1, +\infty[$ dans \mathbb{R}_+^* , et $\varphi_2(x) = -\log x$ de $U_2 =]0, 1[$ dans \mathbb{R}_+^* . On a bien sûr $\varphi_1^{-1}(y) = e^y$ et $\varphi_2^{-1}(y) = e^{-y}$. La proposition ci-dessus donne

$$f_Y(y) = \left(e^{-e^y} e^y + e^{-e^{-y}} e^{-y} \right) 1_{\mathbb{R}_+^*}(y).$$