Probabilités et variables aléatoires
Préparation à l'agrégation interne

Frédérique Bienvenüe-Duheille
frederique.bienvenue@univ-lyon1.fr

1 Probabilité

1.1 Définitions

On se place sur un ensemble Ω appelé **espace de probabilité** ou univers.

Une **tribu** sur Ω est un sous-ensemble Σ de l'ensemble des parties de Ω tel que $\emptyset \in \Sigma$, Σ est stable par passage au complémentaire (i.e. si $A \in \Sigma$, alors $A^c \in \Sigma$) et par réunion dénombrable (i.e. si (A_n) est une famille dénombrable de parties de Ω telle que pour tout n, $A_n \in \Sigma$, alors $\cup_n A_n \in \Sigma$).

On peut alors vérifier qu'une tribu est également stage par intersection dénombrable .

Le plus souvent, si Ω est dénombrable, la tribu utilisée sera $\mathcal{P}(\Omega)$.

Dans le vocabulaire probabiliste,
- Un élément ω de Ω est appelé une **épreuve**
- Un sous-ensemble A de Ω qui appartient à Σ est un **événement**.
- Un **événement élémentaire** est un singleton de Ω.
- L'**événement certain** est Ω.
- L'**événement impossible** est l'ensemble vide.
- Deux événements disjoints sont dits **incompatibles.**

Définition 1.1 *Une* **mesure de probabilité** $(\mathbf{P}, \Sigma)$ *est une fonction définie sur Σ et à valeurs dans $\mathbb{R}$ vérifiant les propriétés suivantes :*

1. *Pour tout événement A de Ω, $\mathbf{P}(A) \geq 0$.*
2. *$\mathbf{P}(\Omega) = 1$.*
3. *Si $(A_n)_{n\geq 1}$ est une famille dénombrables de sous-ensembles de Ω deux à deux* **disjoints***, on a*
$$\mathbf{P}\big(\bigcup_{n\geq 1} A_n\big) = \sum_n \mathbf{P}(A_n).$$

On déduit la proposition suivante de la définition d'une mesure de probabilité :

Proposition 1.2
1. *$\mathbf{P}(\emptyset) = 0$,*
2. *Si A est un événement, $\mathbf{P}(\Omega\backslash A) = 1 - \mathbf{P}(A)$,*
3. *Si $A \subset B$ sont deux événements, $\mathbf{P}(A) \leq \mathbf{P}(B)$,*
4. *Pour tout événement A, $\mathbf{P}(A) \leq 1$,*
5. *Si A et B sont deux événements, $\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B) - \mathbf{P}(A \cap B)$.*

En toute rigueur, $\mathbf{P}$, en tant que fonction, est une mesure de probabilité. Le terme de probabilité se rapporte à la probabilité $\mathbf{P}(A)$ de l'événement A. Par abus de langage, on utilisera le mot « probabilité » dans les deux cas.

Remarque : Le troisième point de la définition signifie qu'une probabilité est une fonction croissante : c'est une façon de calculer la « taille » des événements.

Pour la suite de ce cours, on se placera sur un espace Ω muni d'une mesure de probabilité $\mathbf{P}$.

1.2 Probabilités discrètes

La mesure de probabilité $\mathbf{P}$ est dite **discrète** dès que l'espace Ω est fini ou dénombrable ou plus généralement, dès qu'il existe un sous-ensemble Ω_0 de Ω fini ou dénombrable et tel que $\mathbf{P}(\Omega_0) = 1$. Une probabilité sur un ensemble dénombrable sera toujours discrète.

On se placera dans la suite de ce paragraphe dans le cas où Ω **est fini ou dénombrable.**

Proposition 1.3 *Une probabilité sur un ensemble* **dénombrable** *est complètement déterminée par les $\mathbf{P}(\{\omega\})$ pour tout $\omega \in \Omega$. En effet, pour $A \subset \Omega$, on a*

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

Remarques :

- Les poids d'une probabilité discrète $\mathbf{P}$ vérifient $\sum_{\omega\in\Omega} \mathbf{P}(\{\omega\}) = 1$.
- Une mesure de probabilité ne permet d'évaluer a priori que la taille de sous-ensembles de Ω.

Des exemples

• *Lancer d'une pièce équilibrée* : on souhaite modéliser le résultat du lancer d'une pièce sans tricherie. Pour cela, on choisit $\Omega_1 = \{\text{pile}, \text{face}\}$, et donc $\text{card}\Omega_1 = 2$. L'ensemble des parties de Ω_1 comporte quatre éléments et on définit la mesure de probabilité $\mathbf{P}$ par $\mathbf{P}\{\text{pile}\} = \mathbf{P}\{\text{face}\} = 1/2$ puisque les deux événements sont équiprobables (c'est-à-dire de même probabilité).

Remarque : On aurait très bien pu choisir $\Omega_1 = \{\text{pile}, \text{face}, \text{rouge}, \text{vert}\}$, et comme mesure de probabilité $\mathbf{P}\{\text{pile}\} = \mathbf{P}\{\text{face}\} = 1/2$ et $\mathbf{P}\{\text{rouge}\} = \mathbf{P}\{\text{vert}\} = 0$, mais tant qu'à faire, on choisit le plus simple...

• *Lancer de k pièces*, $k \geq 2$: on prend cette fois-ci $\Omega_k = (\Omega_1)^k$, c'est-à-dire l'ensemble des k-uplets de pile ou face. On a $\text{card}\Omega_k = 2^k$ et $\text{card}\mathcal{P}(\Omega_k) = 2^{2^k}$. Les différents k-uplets sont tous équiprobables donc $\mathbf{P}(\{\omega\}) = 2^{-k}$, pour tout $\omega \in \Omega_k$.

• *Probabilité uniforme discrète* : sur un ensemble **fini** $\Omega = \{\omega_1, \dots, \omega_n\}$, avec $n = \text{card}(\Omega)$, on définit la probabilité uniforme par $\mathbf{P}(\{\omega_i\}) = 1/n$ pour tout i entre 1 et n. Dans ce cas, tous les ω_i ont la même probabilité de se produire (i.e. sont **équiprobables**), et pour une partie A de Ω, on a

$$\mathbf{P}(A) = \frac{\text{card}A}{n} = \frac{\text{nb cas favorables}}{\text{nb cas possibles}}.$$

Par exemple, lors du lancer d'un dé régulier à six faces, chaque face est obtenue avec la même probabilité 1/6.

Remarque : Il ne peut bien sûr pas y avoir de loi uniforme sur $\mathbb{N}$.

• *Exemple de mesure de probabilité sur* $\mathbb{N}^*$. On lance un dé de façon répétée jusqu'à obtenir un 6, et on note le numéro du tirage du premier 6. On a évidemment $\mathbf{P}(\{1\}) = 1/6$.

On a également

$$\begin{aligned}\mathbf{P}(\{2\}) &= \mathbf{P}(\text{au premier tirage, on n'a pas eu de 6 ; au deuxième tirage, on a eu un 6}) \\ &= \frac{5}{36}\end{aligned}$$

car sur les 36 tirages possibles équiprobables, seuls 5 permettent d'obtenir le premier 6 au deuxième tirage.

De même, pour tout $k \geq 2$,

$$\mathbf{P}(\{k\}) = \mathbf{P}(k-1 \text{ échecs puis une réussite}) = \frac{5^{k-1}}{6^k} = \left(\frac{5}{6}\right)^{k-1}\frac{1}{6}.$$

Cela constitue bien une mesure de probabilité discrète sur $\mathbb{N}^*$ puisque $\sum_{k\geq 1}\mathbf{P}(k) = 1$.

Attention : Ne pas confondre cette probabilité avec la probabilité de tirer un 6 exactement parmi les k premiers lancers.

Remarque : On pourrait chercher à écrire un univers U permettant de décrire l'intégralité des résultats des tirages successifs. Le plus simple est de choisir $U = \{1, \ldots, 6\}^{\mathbb{N}}$. Cet ensemble n'est pas dénombrable. Une tribu raisonnable dont on peut le munir est la tribu cylindrique : c'est la tribu qui est « engendrée » par tous les événements de la forme $(x_1, \ldots, x_n) \times \{0,1\}^{\mathbb{N}}$, avec $n \in \mathbb{N}^*$ et $(x_i)_{i\leq n}) \in \{0,1\}^n$ (on fixe les n premières composantes et on laisse les autres libres).

1.3 Probabilité à densité

On se place sur $\mathbb{R}$ et on note dx l'élément d'intégration de la mesure de Lebesgue. Soit $f : \mathbb{R} \to \mathbb{R}$ une fonction positive et d'intégrale sur $\mathbb{R}$ égale à 1. On supposera que f est continue par morceaux. Il est facile de vérifier que l'on définit une mesure de probabilité μ en posant, pour tout $I \subset \mathbb{R}$:

$$\mu(I) = \int_{\mathbb{R}} \mathbf{1}_I(x) f(x)\ dx.$$

Une telle mesure est dite à densité (par rapport à la mesure de Lebesgue sur $\mathbb{R}$). On dit également que c'est une probabilité continue.

Des exemples

• La mesure uniforme sur l'intervalle $[a,b]$, où $a < b$: On définit

$$\mu(A) = \int_{\mathbb{R}} \mathbf{1}_{A\cap[a,b]}(x)\frac{dx}{b-a} = \int_A \mathbf{1}_{[a,b]}(x)\frac{dx}{b-a}.$$

• La mesure de Gauss sur $\mathbb{R}$. On utilise ici la fonction

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

où $m \in \mathbb{R}$ et $\sigma \in \mathbb{R}^{+*}$ sont deux paramètres fixés. Un joli exercice consiste à prouver (au moins dans le cas $m = 0$ et $\sigma = 1$ que l'intégrale de la fonction f sur $\mathbb{R}$ est égale à 1.

2 Probabilité conditionnelle, indépendance

Définition 2.1 *On se donne deux événements A et B de Ω, avec $\mathbf{P}(B) > 0$. On définit la* **probabilité conditionnelle de A sachant B**, *notée $\mathbf{P}(A|B)$ ou $\mathbf{P}_B(A)$ par*

$$\mathbf{P}_B(A) = \mathbf{P}(A|B) = \mathbf{P}(A \cap B)/\mathbf{P}(B).$$

Un cas typique où interviennent des probabilités conditionnelles est les expériences aléatoires obtenues par des tirages successifs de boules dans des urnes (mais il y a bien sûr d'autres cadres où elles apparaissent naturellement !) On résout ces question le plus souvent en dressant un arbre de probabilité : les données figurant sur les arêtes de l'arbre ont des probabilités conditionnelles.

La probabilité conditionnelle vérifie les mêmes propriétés qu'une probabilité : on a ainsi $\mathbf{P}_B(\Omega) = 1$, $\mathbf{P}_B(\emptyset) = 0$, si A_1 et A_2 sont disjoints, $\mathbf{P}_B(A_1 \cup A_2) = \mathbf{P}_B(A_1) + \mathbf{P}_B(A_2)$, $\mathbf{P}_B(\Omega \backslash A) = 1 - \mathbf{P}_B(A)$...

On peut donc énoncer la proposition suivante :

Proposition 2.2 *Soit B un événement de probabilité strictement positive. On note $\mathbf{P}_B$ la probabilité conditionnelle sachant l'événement B. Alors $\mathbf{P}_B$ est une probabilité sur Ω, c'est-à-dire que*
- *Pour tout $A \in \Omega$, $\mathbf{P}_B(A) \geq 0$.*
- $\mathbf{P}_B(\Omega) = 1$
- *Si A_1 et A_2 sont incompatibles, $\mathbf{P}_B(A_1 \cup A_2) = \mathbf{P}_B(A_1) + \mathbf{P}_B(A_2)$.*

Les probabilités conditionnelles permettent de décomposer un événement suivant des sous-ensembles de Ω sur lesquels on maîtrise mieux ce qui se passe. Pour cela, introduisons la notion de système complet d'événements :

Définition 2.3 *Un* **système complet d'événements** *est une famille dénombrable ou finie (B_n) d'événements deux à deux disjoints et vérifiant $\cup_n B_n = \Omega$.*

Remarque : Plusieurs définitions d'un système complet d'événements cohabitent : suivant l'une d'elle par exemple, un système complet d'événements est une partition de Ω, d'autres définitions imposent que les B_n soient tous de probabilité strictement positive ; on peut aussi ne pas imposer que la réunion des B_n soit égale à Ω, mais plutôt qu'elle soit de probabilité 1... Le point commun à ces définitions est que les B_n sont en nombre dénombrables, deux à deux disjoints et que leur réunion est « presque » Ω. La définition indiquée ici n'implique en particulier pas que les B_n soient non vides.

Remarque : Si l'ensemble Ω est fini, tout système complet ne comporte qu'un nombre fini d'événements non vides.

Proposition 2.4 (Formule des probabilités totales) *Soit (B_n) un système complet d'événements tel que, pour tout $n \geq 1$, $\mathbf{P}(B_n) > 0$, et A un événement quelconque. On a*

$$\mathbf{P}(A) = \sum_n \mathbf{P}(A \cap B_n) = \sum_n \mathbf{P}_{B_n}(A)\mathbf{P}(B_n).$$

Remarque : Si par exemple $\mathbf{P}(B_1) = 0$, on pourrait poser $\mathbf{P}_{B_1}(A) = 0$, ou 1, ou 1/2 pour tout A, cela n'interviendrait pas dans la formule ci-dessus. Néanmoins il est plus pédagogique d'imposer que les B_n soient tous de probabilité strictement positive, pour que la formule ci-dessus soit rigoureuse.

Preuve : Par définition, pour tout n, $\mathbf{P}_{B_n}(A)\mathbf{P}(B_n) = \mathbf{P}(A \cap B_n)$ et les événements $A \cap B_n$ sont deux à deux disjoints car les B_n le sont. On en déduit donc que

$$\begin{aligned}\sum_n \mathbf{P}_{B_n}(A)\mathbf{P}(B_n) &= \mathbf{P}(\cup_n (A \cap B_n)) \\ &= \mathbf{P}(A \cap (\cup_n B_n)) = \mathbf{P}(A).\end{aligned}$$

□

Un problème courant est de déterminer $\mathbf{P}_B(A)$ à partir de $\mathbf{P}_A(B)$. La seule donnée de $\mathbf{P}_A(B)$ n'y suffit pas. Il faut par exemple connaître aussi $\mathbf{P}(A)$ et $\mathbf{P}(B)$: on a alors

$$\mathbf{P}_B(A) = \mathbf{P}_A(B)\mathbf{P}(A)/\mathbf{P}(B).$$

Une autre possibilité est de connaître $\mathbf{P}(A)$ et $\mathbf{P}_{\bar{A}}(B)$ où $\bar{A}$ est le complémentaire de A :
Formule de Bayes :

- Soient A et B deux événements de probabilité strictement positive, vérifiant également $\mathbf{P}(\bar{A}) > 0$. On vérifie que

$$\mathbf{P}_B(A) = \frac{\mathbf{P}_A(B)\mathbf{P}(A)}{\mathbf{P}_A(B)\mathbf{P}(A) + \mathbf{P}_{\bar{A}}(B)\mathbf{P}(\overline{A})}.$$

- Soient (A_n) un système complet d'événements tel que, pour tout n, $\mathbf{P}(A_n) > 0$ et B un événement tel que $\mathbf{P}(B) > 0$. On a pour tout i :

$$\mathbf{P}_B(A_i) = \frac{\mathbf{P}_{A_i}(B)\mathbf{P}(A_i)}{\sum_n \mathbf{P}_{A_n}(B)\mathbf{P}(A_n)}.$$

Preuve : Le dénominateur du membre de droite vaut en fait $\mathbf{P}(B)$, alors que le numérateur vaut $\mathbf{P}(A \cap B)$, d'où le résultat. □

Définition 2.5 *Deux événements A et B sont dits* **indépendants** *si $\mathbf{P}(A \cap B) = \mathbf{P}(A)\mathbf{P}(B)$. On a alors $\mathbf{P}(A|B) = \mathbf{P}(A)$ et $\mathbf{P}(B|A) = \mathbf{P}(B)$ si $\mathbf{P}(A) > 0$ et $\mathbf{P}(B) > 0$.*

Exercice 1
1. *Montrer qu'un événement de probabilité nulle est indépendant de tout événement.*
2. *Montrer que si A et B sont indépendants, alors $\Omega \backslash A$ et B le sont.*
3. *Montrer qu'un événement de probabilité 1 est indépendant de tout événement.*

Exemples :

• *Lors d'un lancer de pile ou face, les événements « tomber sur pile au premier tirage » et « tomber sur pile au deuxième tirage » sont généralement indépendants (sauf en cas de tricherie...)*

• Tirage avec remise. *On dispose d'une urne contenant N boutons noirs et J boutons jaunes. À chaque tirage, on prend un bouton au hasard, on note la couleur du bouton obtenu et on le remet dans l'urne. Les événements $A = \{$tirer un bouton noir au premier tirage$\}$ et $B = \{$tirer un bouton jaune au deuxième tirage$\}$ sont-ils indépendants ?*

• Urne de Polya. *On dispose toujours d'une urne contenant N boutons noirs et J boutons jaunes. À chaque tirage, on note la couleur du bouton obtenu et on le remet dans l'urne accompagné d'un bouton de la même couleur. Même question que précédemment.*

Définition 2.6 *Soit n un entier supérieur ou égal à 2. n événements $A_1, \dots, A_n$ sont* **(mutuellement ou n à n) indépendants** *si pour tout choix d'indices $i_1, \dots, i_k$ deux à deux distincts, on a*

$$\mathbf{P}(A_{i_1} \cap \ldots \cap A_{i_k}) = \mathbf{P}(A_{i_1}) \times \cdots \times \mathbf{P}(A_{i_k}).$$

Des événements n à n indépendants le sont bien évidemment 2 à 2 mais la réciproque est fausse.

Exercice 2 • *On choisit $\Omega = \{1, 2, 3, 4\}$ et on le munit de la probabilité uniforme. Trouver trois événements deux à deux indépendants mais pas trois à trois.*

• *Sur $\Omega = \{1, \dots, 8\}$ muni de la probabilité uniforme, trouver trois événements A, B et C tels que $\mathbf{P}(A \cap B \cap C) = \mathbf{P}(A)\mathbf{P}(B)\mathbf{P}(C)$ mais tels que A, B et C ne sont pas indépendants.*

3 Variables aléatoires réelles

3.1 La loi

3.1.1 Définition

Une **variable aléatoire** *X sur Ω est une fonction $X : (\Omega, \Sigma) \to \mathbb{R}$ telle que pour tout intervalle I de $\mathbf{P}$, l'image-réciproque de I par X appartienne à Σ.*

Notation : on notera $\{X \in I\} = \{\omega \in \Omega, X(\omega) \in I\} = X^{-1}(I)$.

Notation : Pour tout intervalle I et pour tout $x \in \mathbb{R}$, on note

$$\{X \in I\} = \{\omega \in \Omega, X(\omega) \in I\} = X^{-1}(I),$$

et pour tout $x \in \mathbb{R}$,

$$\{X = x\} = \{\omega \in \Omega, X(\omega) = x\} = X^{-1}(\{x\}).$$

Les ensembles $\{X \in I\}$ et $\{X = x\}$ sont des sous-ensembles de Ω. On pourra donc étudier par exemple $\mathbf{P}\{X \in I\}$ pour tout intervalle I de $\mathbb{R}$, mais pas $\mathbf{P}(I)$.

Énonçons la propriété fondamentale de μ :

Proposition 3.1 *La fonction μ ainsi définie est une probabilité sur $\mathbb{R}$ (ou sur l'ensemble $X(\Omega)$).*

Définition 3.2 *La probabilité μ est appelée la* **mesure image** *de* $\mathbf{P}$ *par X, ou la* **loi** *de X.*

Sa loi est ainsi complètement déterminée par la donnée de l'ensemble $X(\Omega)$ ainsi que par les quantités $\mu(B) = \mathbf{P}(X^{-1}(I))$ pour tout intervalle I de $\mathbb{R}$. On note parfois $\mu = X(\mathbf{P})$ ou $\mu = \mathbf{P}_X$ (attention dans ce dernier cas à ne pas faire de confusion avec la probabilité conditionnelle).

La loi est la principale information dont on disposera sur une variable aléatoire : souvent l'ensemble Ω sera inconnu ou implicite, on n'aura donc pas d'information sur $X(\omega)$.

Définition 3.3 – *La variable aléatoire X sera* **discrète** *si elle prend ses valeurs dans un ensemble discret (et sa mesure-image est alors une mesure discrète). Sa loi sera caractérisée par l'ensemble $X(\Omega)$ (ou par un ensemble dénombrable contenant $X(\Omega)$) et par les probabilités $\mathbf{P}(X = x)$ pour tout $x \in X(\Omega)$.*
– *X sera* **à densité** *(on dit aussi que X est continue) si sa mesure image admet une densité, c'est-à-dire s'il existe une fonction $f : \mathbb{R} \to \mathbb{R}^+$, continue par morceaux, d'intégrale sur $\mathbb{R}$ égale à 1, telle que pour tous réels a et b vérifiant $a \leq b$,*

$$\mathbf{P}(X \in [a, b]) = \int_a^b f(x)\, dx.$$

En particulier en prenant $a = b$ dans l'égalité ci-dessus, on remarque $\mathbf{P}(X = a) = 0$ pour tout $a \in \mathbb{R}$.

Remarque : *Si Ω est un ensemble fini ou dénombrable, toute variable aléatoire définie sur Ω sera discrète.*

Attention : *Deux variables aléatoires peuvent suivre la même loi sans être égales : par exemple deux tirages successifs de pile ou face.*

Nous allons maintenant étudier quelques exemples de variables aléatoires discrètes ou à densité, mais il faut garder à l'esprit que cela ne recouvre pas tous les types de variables aléatoires.

3.2 Exemples de variables aléatoires discrètes

Définition 3.4 *La* **loi** *d'une variable aléatoire discrète est donnée par*
– *l'ensemble (dénombrable) $X(\Omega)$,*
– *pour tout $x \in X(\Omega)$, la quantité*

$$\mathbf{P}(\{\omega \in \Omega \text{ tels que } X(\omega) = x\}) = \mathbf{P}(X^{-1}\{x\}) = \mathbf{P}(X = x)$$

Remarque : *On doit avoir $\sum_x \mathbf{P}(X = x) = 1$, où la somme est prise sur $x \in X(\Omega)$.*

Pour construire une variable aléatoire discrète à valeurs dans $\mathbb{N}$, on peut aussi commencer par définir une mesure de probabilité sur $\mathbb{N}$ en se donnant le poids p_n de chaque entier n (avec $\sum p_n = 1$) puis considérer une variable aléatoire X d'un certain espace Ω dans $\mathbb{N}$ dont la loi est donnée par $\mathbf{P}(X = n) = p_n$.

Exercice 3 *On se donne une variable aléatoire $X : \Omega \to \mathbb{N}$. Montrer que la famille $A_n = \{\omega, X(\omega) = n\}$ pour tout $n \geq 0$ forme un système complet d'événements.*

Des exemples

• *Pour un événement $A \subset \Omega$, on note $\mathbf{1}_A$ la fonction suivante :* $\mathbf{1}_A(\omega) = 1$ *si* $\omega \in A$ *et* $\mathbf{1}_A(\omega) = 0$ *sinon. Cette fonction, appelée* **l'indicatrice de l'événement** *A, est une variable aléatoire discrète très utile.*

• *Le nombre de « piles » obtenus lors des 8 premiers tirages d'un jeu de pile ou face est aussi une variable aléatoire discrète.*

• Loi de Dirac en $a \in \mathbb{R}$. *On fixe un nombre réel a. La loi de Dirac en a, généralement notée δ_a, est la loi de la variable aléatoire suivante :* $X(\Omega) = \{a\}$ *et* $\mathbf{P}(X = a) = 1$. *On dit que X vaut « presque-sûrement » a.*

Exercice 4 *Montrer que, si X suit la loi de Dirac en a, pour tout $A \subset \mathbb{R}$,* $\mathbf{P}(X \in A) = \mathbf{1}_A(a)$.

• Loi de Bernoulli. *La loi de Bernoulli $\mathcal{B}(p)$ de paramètre $p \in [0,1]$ est donnée par* $X(\Omega) = \{0,1\}$ *et* $\mathbf{P}(X = 1) = p = 1 - \mathbf{P}(X = 0)$. *Lors d'un tirage de pile ou face d'une pièce équilibrée, si on note $X = 1$ si la pièce tombe sur pile et 0 sinon, on obtient une variable aléatoire de loi de Bernoulli $\mathcal{B}(\frac{1}{2})$. Plus généralement, pour un événement A quelconque, la variable aléatoire* $\mathbf{1}_A$ *suit une loi de Bernoulli de paramètre* $\mathbf{P}(A)$.

• Loi binomiale. *La loi binomiale* $\mathrm{Bin}(n,p)$, *pour $n \in \mathbb{N}^*$ et $p \in [0,1]$ est donnée par* $X(\Omega) = \{0, \dots, n\}$ *et, pour tout $k \in \{0, \dots, n\}$,* $\mathbf{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$. *On retrouve ici la probabilité d'obtenir k fois exactement au cours de n tentatives (indépendantes) la réalisation d'un événement dont la probabilité est p. Par exemple, la probabilité de tirer exactement k 6 lors des n premiers lancers d'un dé est* $\binom{n}{k} 5^{n-k} 6^{-n}$.

• Loi uniforme sur $\{1, \dots, n\}$. *On a ici $X(\Omega) = \{1, \dots, n\}$ et cette loi affecte le même poids à chacun des éléments. On a donc* $\mathbf{P}(X = k) = 1/n$, *pour tout $k \in \{1, \dots, n\}$.*

• Loi géométrique $\mathcal{G}(p)$, $p \in]0,1[$ *: Cette loi est donnée par $X(\Omega) = \mathbb{N}^*$ et* $\mathbf{P}(X = k) = p(1-p)^{k-1}$ *pour tout $k \in \mathbb{N}^*$. On a vu plus haut que c'est la loi du numéro du tirage où la réussite survient pour la première fois (toujours dans le cadre d'une répétition indépendante des expériences de Bernoulli).*

• Loi de Poisson $\mathcal{P}(\lambda)$, $\lambda > 0$. *C'est la loi de la variable aléatoire X vérifiant $X(\Omega) = \mathbb{N}$ et* $\mathbf{P}(X = k) = e^{-\lambda} \lambda^k / k!$. *Elle est généralement utilisée pour modéliser le nombre d'appels reçus par un serveur au cours d'un laps de temps donné.*

• Loi hypergéométrique *: Soit r, b et n trois entiers naturel non nuls. La loi hypergéométrique $(b+r, r, n)$ est la loi du nombre de boules rouges que l'on obtient lorsque l'on tire simultanément n boules dans une urne contenant r boules rouges et b boules blanches. On a :*

$$\mathbf{P}(X = k) = \frac{\binom{r}{k}\binom{b}{n-k}}{\binom{r+b}{n}}$$

3.3 Exemples de variables aléatoires à densité

• Loi uniforme sur $[a,b]$ *: c'est la loi de la variable aléatoire X de densité* $\mathbf{1}_{[a,b]}/(b-a)$. *La probabilité qu'une variable aléatoire de loi uniforme sur $[a,b]$ appartienne à un sous-intervalle de $[a,b]$ est proportionnelle à la longueur de ce sous-intervalle. On a en particulier* $\mathbf{P}(X \in [a,b]) = 1$.

• Loi exponentielle de paramètre λ. *Il s'agit de la loi de densité* $f_\lambda(x) = \lambda \exp(-\lambda x)\mathbf{1}_{x>0}$. *Si* X *suit cette loi, on a* $\mathbf{P}(X \geq 0) = 1$. *La loi exponentielle est dite* **sans mémoire** *au sens où pour tous réels positifs* s *et* t, *on a* $\mathbf{P}(X > t+s|X > s) = \mathbf{P}(X > t)$. *C'est pour cette raison qu'elle est utilisée généralement pour modéliser des temps d'attente entre deux événements : par exemple entre deux pannes successives d'une machine, ou entre deux requêtes reçues par un serveur informatique.*

• Loi normale, ou loi de Gauss centrée réduite. *Il s'agit de la loi de la variable aléatoire* X *de densité* $f(x) = e^{-x^2/2}/\sqrt{2\pi}$. *C'est une loi très utilisée en statistique et en modélisation. Nous allons commencer par vérifier que c'est bien la densité d'une probabilité :* f *est une fonction continue positive, il reste à voir que* $I = \int_{\mathbb{R}} f(t)\ dt = 1$. *On ne connaît pas de primitive explicite de la fonction* f, *mais nous allons calculer* I^2. *On a*

$$\begin{aligned} I^2 &= \left(\int_{-\infty}^{+\infty} f(t)\ dt\right)^2 \\ &= \left(\int_{-\infty}^{+\infty} f(t)\ dt\right) \times \left(\int_{-\infty}^{+\infty} f(s)\ ds\right) \\ &= \int_{\mathbb{R}^2} e^{-(s^2+t^2)/2} \frac{ds\,dt}{2\pi}. \end{aligned}$$

Procédons à un changement de variables en coordonnées polaires en posant $s = r\cos\theta$ *et* $t = r\sin\theta$. *Il vient*

$$\begin{aligned} I^2 &= \int_0^\infty \int_{-\pi}^{\pi} e^{-r^2/2}\ \frac{r\,dr\,d\theta}{2\pi} \\ &= \int_0^\infty re^{-r^2/2} dr \\ &= 1. \end{aligned}$$

La loi normale $\mathcal{N}(m, \sigma^2)$ *est de densité*

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

3.4 Espérance

Donnons tout d'abord la définition générale de l'espérance, avant de l'appliquer aux variables aléatoires discrètes ou à densité.

Soit $X : \Omega \to \mathbb{R}$ *une variable aléatoire de loi* μ.

Définition 3.5 *Une variable aléatoire* X *est dite* **intégrable** *si la quantité*

$$\int_\Omega |X|\ d\mathbf{P} = \int_{\mathbb{R}} |x|\ d\mu$$

est finie. On définit alors son **espérance** *par*

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

Plus généralement, pour toute fonction continue par morceaux $h : \mathbb{R} \to \mathbb{R}$, on a

$$\mathbf{E}(h(X)) = \int_\Omega h(X)\ d\mathbf{P} = \int_\mathbb{R} h(x)\ d\mu,$$

lorsque la quantité

$$\int_\Omega |h(X)|\ d\mathbf{P} = \int_\mathbb{R} |h(x)|\ d\mu < \infty.$$

Dans le langage courant (et aussi probabiliste), l'espérance est appelée **moyenne**. *C'est un paramètre de position, qui indique autour de quelle valeur la variable aléatoire est répartie.*

Insistons dès maintenant sur le fait qu'une variable aléatoire X bornée par une constante M (c'est-à-dire que $\mathbf{P}(|X| \leq M) = 1$) est toujours intégrable, que l'on a dans ce cas $\mathbf{E}(|X|) \leq M$ et $\mathbf{E}(X) \in [-M, M]$.

Cette définition générale induit deux écritures différentes suivant que la variable aléatoire X est discrète ou à densité :

- *Si X est une variable aléatoire* **discrète**, *l'intégrabilité se traduit par*

$$\sum_{x \in X(\Omega)} |x| \mathbf{P}(X = x) < \infty$$

  et on a alors :

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

  Plus généralement, pour toute fonction $h : X(\Omega) \to \mathbb{R}$, on a

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

  si $h(X)$ est intégrable, c'est-à-dire, si

$$\mathbf{E}|h(X)| = \sum_{x \in X(\Omega)} |h(x)| \mathbf{P}(X = x) < \infty.$$

- *Si X est une variable aléatoire à valeurs réelle et de* **densité** *$f : \mathbb{R} \to \mathbb{R}^+$, elle est intégrable si*

$$\int_\mathbb{R} |x| f(x)\ dx < \infty$$

  et on a dans ce cas

$$\mathbf{E}(X) = \int_\mathbb{R} x f(x)\ dx.$$

  Plus généralement, pour toute fonction continue par morceaux $h : \mathbb{R} \to \mathbb{R}$, la variable aléatoire $h(X)$ est intégrable, si

$$\mathbf{E}|h(X)| = \int_\mathbb{R} |h(x)| f(x)\ dx < \infty$$

  et on a alors

$$\mathbf{E}(h(X)) = \int_\mathbb{R} h(x) f(x)\ dx.$$

Remarque importante *pour le cas discret : Supposons que l'espace* Ω *soit* **fini** *:* $\Omega = \{\omega_1, \ldots, \omega_k\}$. *Dans ce cas, toute variable aléatoire* $X : \Omega \to \mathbb{R}$ *est discrète. Notons* $n = card(X(\Omega))$ *et* $X(\Omega) = \{x_1, \ldots, x_n\}$. *On a*

$$\begin{aligned}
\mathbf{E}(X) &= \sum_{i=1}^{n} x_i \mathbf{P}(X = x_i) \\
&= \sum_{i=1}^{n} \left(x_i \sum_{\omega \in \Omega; X(\omega) = x_i} \mathbf{P}(\{\omega\}) \right) \\
&= \sum_{i=1}^{n} \left(\sum_{\omega \in \Omega; X(\omega) = x_i} x_i \mathbf{P}(\{\omega\}) \right) \\
&= \sum_{i=1}^{n} \left(\sum_{\omega \in \Omega; X(\omega) = x_i} X(\omega) \mathbf{P}(\{\omega\}) \right) \\
&= \sum_{\omega \in \Omega} X(\omega) \mathbf{P}(\{\omega\})
\end{aligned}$$

Cette expression permet de justifier très simplement que, pour toute variable aléatoire $X : \Omega \to \mathbb{R}$, *partant d'un espace fini* Ω, *et pour toute fonction* $h : \mathbb{R} \to \mathbb{R}$, *on a*

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

Cette expression de l'espérance de $h(X)$ *est appelée « thèorème du transfert ». Sa preuve peut donc se faire simplement dans le cas discret (avec* Ω *fini ou dénombrable) ; on peut la faire également pour certaines fonctions simples dans le cas des variables aléatoires à densité, mais le cas général nécessite des connaissances en théorie de la mesure.*

Exemple : *Calcul de l'espérance de la loi géométrique de paramètre* p. *On se donne une variable aléatoire* X *de loi* $\mathcal{G}(p)$. *On a vu que* $X(\Omega) = \mathbb{N}^*$ *: il est inutile de vérifier l'intégrabilité de* X *puisque* $X(\Omega) \subset \mathbb{R}^+$.

Puisque X *est une variable aléatoire discrète, on a*

$$\begin{aligned}
\mathbf{E}(X) &= \sum_{n \geq 1} n \mathbf{P}(X = n) \\
&= \sum_{n \geq 1} n p (1-p)^{n-1} \\
&= p \sum_{n \geq 0} n (1-p)^{n-1}
\end{aligned}$$

Or $n(1-p)^{n-1}$ *est la dérivée de* $-(1-p)^n$ *et* $\sum_{n \geq 0} (1-p)^n = 1/p$. *On obtient (en inversant une dérivation et une série, ce qui est licite car il s'agit d'une série entière) :*

$$\sum_{n \geq 0} n(1-p)^{n-1} = -\frac{d}{dp}\frac{1}{p} = \frac{1}{p^2}.$$

Finalement, $\mathbf{E}(X) = 1/p$.

Exemple : *Calcul de l'espérance d'une variable aléatoire* Y *de loi exponentielle, c'est-à-dire de densité* $f(y) = \lambda \exp(-\lambda y)\mathbf{1}_{\mathbb{R}^+}(y)$ *où* $\lambda > 0$ *est une constante. On constate que* Y *est une variable aléatoire positive (l'intégrale de sa densité sur* $\mathbb{R}^-$ *est nulle). On peut donc s'abstenir de vérifier l'intégrabilité, et passer directement au calcul de l'espérance.*

On a par définition

$$\mathbf{E}(Y) = \int_{\mathbb{R}} y f(y)\, dy = \int_{\mathbb{R}^+} \lambda y e^{-\lambda y}\, dy.$$

Cette intégrale s'intègre par parties :

$$\mathbf{E}(Y) = \Big[y \times (-\exp(-\lambda y)) \Big]_0^{+\infty} - \int_0^{+\infty} (-\exp(-\lambda y))\, dy = \frac{1}{\lambda}.$$

Proposition 3.6 • *Soient* a *une constante et* X *une variable aléatoire intégrable. Les variables aléatoires* $X + a$ *et* aX *sont intégrables et on a* $\mathbf{E}(X + a) = a + \mathbf{E}(X)$ *et* $\mathbf{E}(aX) = a\mathbf{E}(X)$.

• *Si* X *est une variable aléatoire positive et intégrable, alors* $\mathbf{E}(X) \geq 0$.

• *Si* X *et* Y *sont deux variables aléatoires intégrables vérifiant* $X \leq Y$ *alors* $\mathbf{E}(X) \leq \mathbf{E}(Y)$. *On déduit de cette propriété que toute variable aléatoire bornée par une constante (ou plus généralement par une variable aléatoire intégrable) est intégrable.*

• *L'espérance est une opération linéaire : si* X *et* Y *sont deux variables aléatoires intégrables et* a *et* b *deux nombres réels,* $\mathbf{E}(aX + bY) = a\mathbf{E}(X) + b\mathbf{E}(Y)$

Exercice 5 • *On se donne un événement* A. *Montrer que* $\mathbf{E}(\mathbf{1}_A) = \mathbf{P}(A)$.

• *On considère une variable aléatoire* X *et un réel* x. *Calculer* $\mathbf{E}(\mathbf{1}_{X \leq x})$.

3.5 Variance

Définition 3.7 *La* **variance** *d'une variable aléatoire de carré intégrable* X *est égale à*

$$\mathbf{var}\, X = \mathbf{E}\left[(X - \mathbf{E}(X))^2 \right].$$

On appelle **écart-type** *la quantité* $\sigma = \sqrt{\mathbf{var}\,(X)}$. *La variance et l'écart-type sont des paramètres de dispersion : plus ils sont grands, plus la variable aléatoire est dispersée au tour de sa moyenne (c'est-à-dir : prend des valeurs éloignées de la moyenne).*

Proposition 3.8

- **Formule de Koenig**. *Pour toute variable aléatoire de carré intégrable* X, *on a* $\mathbf{var}\, X = \mathbf{E}(X^2) - (\mathbf{E}(X))^2$.
- *Si* X *est une variable aléatoire de carré intégrable et si* a *est une constante, on a* $\mathbf{var}\,(X + a) = \mathbf{var}\,(X)$ *et* $\mathbf{var}\,(aX) = a^2\mathbf{var}\,(X)$.
- *La variance d'une variable aléatoire de carré intégrable est toujours une quantité positive. Elle n'est nulle que si la variable aléatoire suit une loi de Dirac.*

Preuve :

1) Notons $m = \mathbf{E}(X)$. *On a* $(X - m)^2 = X^2 - 2mX + m^2$. *Donc*

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

2) On pose $Y = X + a$. On a $\mathbf{E}(Y) = a + \mathbf{E}(X)$ et $\mathbf{var}\,(Y) = \mathbf{E}((Y - \mathbf{E}(Y))^2)$.
D'où $\mathbf{var}\,(Y) = \mathbf{E}((X - \mathbf{E}(X))^2) = \mathbf{var}\,(X)$.
Posons aussi $Z = aX$. On a $\mathbf{E}(Z) = a\mathbf{E}(X)$ et $\mathbf{var}\,(Z) = \mathbf{E}((aX - \mathbf{E}(Z))^2)$.
D'où $\mathbf{var}\,(Y) = a^2\mathbf{E}((X - \mathbf{E}(X))^2) = a^2\mathbf{var}\,(X)$.

3) La variance est l'espérance d'une variable aléatoire positive : elle est donc positive. Elle ne peut être nulle que si $X = \mathbf{E}(X)$ p.s. □

Remarque : *Rappelons les expressions usuelles de $\mathbf{E}(X^2)$:*

– *Si X est une variable aléatoire discrète,*

$$\mathbf{E}(X^2) = \sum_{x \in X(\Omega)} x^2 \mathbf{P}(X = x)$$

– *Si X est une variable aléatoire de densité f,*

$$\mathbf{E}(X^2) = \int_{-\infty}^{+\infty} x^2 f(x)\, dx.$$

Exemple : Calcul de la variance d'une variable aléatoire X de loi $\mathcal{G}(p)$. On a déjà vu que $\mathbf{E}(X) = 1/p$. Calculons maintenant $\mathbf{E}(X^2)$. On a :

$$\begin{aligned}\mathbf{E}(X^2) &= \sum_{n \geq 1} n^2 \mathbf{P}(X = n) \\ &= \sum_{n \geq 1} n^2 p(1-p)^{n-1}\end{aligned}$$

On va essayer de faire apparaître une dérivée seconde en écrivant $n^2 = n(n-1) + n$:

$$\begin{aligned}\mathbf{E}(X^2) &= p(1-p)\sum_{n \geq 1} n(n-1)(1-p)^{n-2} + p\sum_{n \geq 1} n(1-p)^{n-1} \\ &= p(1-p)\frac{d^2}{dp^2}\big(\sum_{n \geq 0}(1-p)^n\big) + \frac{1}{p} \\ &= p(1-p)\frac{d^2}{dp^2}\frac{1}{p} + \frac{1}{p} \\ &= \frac{2p(1-p)}{p^3} + \frac{1}{p} \\ &= \frac{2-p}{p^2}\end{aligned}$$

On en déduit maintenant la variance de X : $\mathbf{var}\,(X) = \mathbf{E}(X^2) - (\mathbf{E}(X))^2 = (1-p)/p^2$.

Exemple : Calcul de la variance d'une variable aléatoire Y de loi exponentielle de paramètre λ. On a vu (cf p.12) que l'espérance de Y est égale à $1/\lambda$. Calculons maintenant $\mathbf{E}(Y^2)$:

$$\begin{aligned}\mathbf{E}(Y^2) &= \int_{\mathbb{R}} x^2 \lambda e^{-\lambda x} \mathbf{1}_{x \geq 0}\, dx \\ &= \int_0^{+\infty} x^2 \lambda e^{-\lambda x}\, dx \\ &= \left[-x^2 e^{-\lambda x}\right]_0^{+\infty} + 2\int_0^{+\infty} x e^{-\lambda x}\, dx \\ &= 2/\lambda^2\end{aligned}$$

On a donc $\mathbf{var}\,(Y) = 1/\lambda^2$.

Exercice 6 *Calculer les variances des lois des paragraphes 3 et 3.3.*

3.6 Fonction de répartition d'une variable aléatoire

Définition 3.9 *On considère une variable aléatoire* $X : \Omega \to \mathbb{R}$. *La fonction de répartition de la loi de* X *est la fonction* $F_X : \mathbb{R} \to [0,1]$ *définie pour tout* $x \in \mathbb{R}$ *par*

$$F_X(x) = \mathbf{P}(X \leq x).$$

La fonction de répartition caractérise la loi d'une variable aléatoire.

Si X *suit une loi discrète, on a*

$$F_X(x) = \sum_{t \leq x} \mathbf{P}(X = t),$$

où la somme est prise sur tous les $t \in X(\Omega)$ *inférieurs ou égaux à* x. *On obtient une fonction constante par morceaux : ce n'est pas très maniable. La fonction de répartition est peu utilisée dans ce contexte.*

Si X *suit la loi de densité* f,

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

C'est alors une fonction continue, et en tout point où elle est dérivable, sa dérivée est égale à la densité de la loi de X *: c'est par conséquent un outil utilisé pour déterminer la loi de variables aléatoires que l'on pense à densité.*

Proposition 3.10 *La fonction de répartition d'une variable aléatoire* X *est toujours croissante et continue à droite. On a* $\lim_{x\to-\infty} F_X(x) = 0$ *et* $\lim_{x\to+\infty} F_X(x) = 1$.

Exemple : *Fonction de répartition de la loi uniforme sur l'intervalle* $[0,1]$. *Si* X *est de loi uniforme sur [0,1],* X *admet pour densité la fonction* $\mathbf{1}_{[0,1]}$. *Notons* F *sa fonction de répartition. Pour tout* $x \in \mathbb{R}$, *on a donc*

$$F(x) = \int_{-\infty}^{x} \mathbf{1}_{[0,1]}(t)\ dt.$$

- *Si* $x < 0$ *: on a* $F(x) = 0$, *car* $\mathbf{1}_{[0,1]}(t) = 0$ *pour tout* $t \in]-\infty, x]$.
- *Si* $x \in [0,1]$, *on a*

$$\begin{aligned} F(x) &= \int_{-\infty}^{0} \mathbf{1}_{[0,1]}(t)\ dt + \int_{0}^{x} \mathbf{1}_{[0,1]}(t)\ dt \\ &= 0 + \int_{0}^{x} 1 \times\ dt \\ &= x \end{aligned}$$

– *Si $x > 1$,*

$$\begin{aligned} F(x) &= \int_{-\infty}^{0} \mathbf{1}_{[0,1]}(t)\ dt + \int_{0}^{1} \mathbf{1}_{[0,1]}(t)\ dt + \int_{1}^{x} \mathbf{1}_{[0,1]}(t)\ dt \\ &= 0 + \int_{0}^{1} 1 \times\ dt + 0 \\ &= 1 \end{aligned}$$

La fonction de répartition F de la loi uniforme sur $[0,1]$ est donc nulle sur $\mathbb{R}^-$, égale à la fonction identité sur $[0,1]$ et constante égale à 1 sur $[1,+\infty[$. On peut remarquer que cette fonction F est continue, dérivable sur $\mathbb{R}\backslash\{0,1\}$, et que sa dérivée coïncide avec la densité de X là où elle existe.

3.7 Fonctions génératrice

La fonction génératrice est utilisée pour les variables aléatoires positives, et même surtout pour les variables aléatoires à valeurs dans $\mathbb{N}$.

Définition 3.11 *La* **fonction génératrice** *de la loi de la variable aléatoire $X : \Omega \to \mathbb{R}^+$ est la fonction G_X définie pour tout $s \in [0,1]$ par*

$$G_X(s) = \mathbf{E}(s^X) = \sum_{x \in X(\Omega)} s^x \mathbf{P}(X = x).$$

Cette fonction est définie (et finie) pour $s \in [0,1]$ puisque pour tout $s \in [0,1]$, la variable aléatoire s^X est positive et majorée par 1 : c'est donc une variable aléatoire intégrable.

Dans le cas où X est à valeurs dans $\mathbb{N}$, la fonction génératrice est une série entière, de rayon de convergence supérieur ou égal à 1. On peut alors également la définir pour $s \in [-1,0]$ et on a

$$G_X(s) = \mathbf{E}(s^X) = \sum_{n \in \mathbb{N}} s^n \mathbf{P}(X = n).$$

Remarque : *La fonction G_X est polynômiale de degré au plus n si et seulement si $X(\Omega) \subset \{0,\ldots,n\}$.*

Proposition 3.12
- *Soit X une variable aléatoire discrète positive et intégrable. On a $\mathbf{E}(X) = G'_X(1^-)$.*
- *Plus généralement, si $X : \Omega \to \mathbb{R}^+$ est une variable aléatoire vérifiant $\mathbf{E}(X^n) < \infty$, la dérivée n^ede G_X en 1^- est égale à $\mathbf{E}(X(X-1)\ldots(X-n+1))$.*
- *Si deux variables aléatoires positives ont la même fonction génératrice, alors elles suivent la même loi. On dit que la fonction génératrice caractérise la loi des variables aléatoires positives.*

Aide-mémoire : La dérivée de $s \to s^X$ est égale à Xs^{X-1}. En prenant l'espérance pour $s = 1$, on obtient donc $G'_X(1) = \mathbf{E}(X)$.

De même, la dérivée seconde de $s \to s^X$ est égale à $X(X-1)s^{X-1}$, et après la même opération que précédemment, $G''_X(1) = \mathbf{E}(X(X-1))$.

Exemple : Calcul de l'espérance et de la variance d'une loi de Poisson à partir de la fonction génératrice. Soit X une variable aléatoire de loi de Poisson de paramètre $m > 0$. Calculons sa fonction génératrice G_X.

Comme X est à valeurs dans $\mathbb{N}$, on a

$$\begin{aligned} G_X(s) &= \mathbf{E}(s^X) \\ &= \sum_{n\geq 0} s^n \mathbf{P}(X = n) \\ &= \sum_{n\geq 0} s^n e^{-m} \frac{m^n}{n!} \\ &= e^{-m} e^{ms} \end{aligned}$$

En dérivant, on obtient, $\mathbf{E}(X) = G'_X(1) = e^{-m} m e^{m\times 1} = m$.

En dérivant à nouveau : $\mathbf{E}(X(X-1)) = G''_X(1) = e^{-m} m^2 e^{m\times 1} = m^2$.

On en déduit alors la variance de X : par définition, $\mathbf{var}\, X = \mathbf{E}(X^2) - (\mathbf{E}(X))^2$ et en écrivant $\mathbf{E}(X^2) = \mathbf{E}(X(X-1)) + \mathbf{E}(X)$, on obtient $\mathbf{var}\,(X) = m$.

3.8 Comment calculer la loi

Le problème se pose fréquemment de calculer la loi d'une variable aléatoire Y définie par exemple comme fonction d'une autre variable aléatoire X : $Y = g(X)$, la fonction g étant continue par morceaux.

Si Y est une variable discrète, il suffit de déterminer l'ensemble des valeurs prises par Y, puis pour tout $y \in Y(\Omega)$, on aura $\mathbf{P}(Y = y) = \mathbf{P}(X \in g^{-1}(\{y\}))$, ce qui se calcule donc à l'aide de la loi de X.

Si on pense que Y va avoir une densité, on peut imaginer calculer sa fonction de répartition :

$$\mathbf{P}(Y \leq y) = \mathbf{P}(g(X) \leq y).$$

Si la fonction g est monotone, on imagine facilement la suite... mais sinon ? La méthode habituellement utilisée consiste à utiliser une fonction test $h : \mathbb{R} \to \mathbb{R}$ continue et bornée. Si on réussit alors à écrire $\mathbf{E}(h(Y))$ sous la forme

$$\mathbf{E}(h(Y)) = \int_{\mathbb{R}} h(y)\ d\mu(y),$$

on aura gagné : la mesure de probabilité μ obtenue sera la mesure image de Y (de la forme $d\mu = f_Y(y)\ dy$ si Y est à densité). En effet, en fixant $x \in \mathbb{R}$ et en prenant h de la forme $h(y) = \mathbf{1}_{y\leq x}$, on retrouve ainsi la fonction de répartition de la loi de Y. L'avantage est que le changement de variable auquel il faut procéder apparaît clairement.

Exemple : *Soit X une variable aléatoire de densité f_X Déterminons les lois de $Y = aX + b$ et de $Z = (1 + X)/(1 - X)$. Prenons donc une fonction test h continue par morceaux et bornée. On a*

$$\mathbf{E}(h(Y)) = \mathbf{E}(h(aX + b)) = \int_{\mathbb{R}} h(ax + b) f_X(x)\ dx.$$

Petite remarque préalable : la suite de la démonstration présentée ici est basée sur des notions de calcul intégral (programme L3) : ne vous avisez pas à utiliser de telles justifications devant *une classe de terminale ! Il est tout à fait possible de traiter séparément les cas $a > 0$ et $a < 0$ pour rentrer dans le cadre des programmes de lycée.*

On effectue alors le changement de variable $y = ax + b$. Ce changement de variable est bien un difféomorphisme de $\mathbb{R}$ et son jacobien vaut $dx = dy/|a|$. Il vient :

$$\mathbf{E}(h(Y)) = \int_{\mathbb{R}} h(y) f_X\left(\frac{y-b}{a}\right)\frac{dy}{|a|}.$$

On en déduit donc que Y admet pour densité la fonction f_Y définie sur $\mathbb{R}$ par

$$f_Y(Y) = f_X\left(\frac{y-b}{a}\right)\frac{1}{|a|}.$$

Procédons de même pour Z :

$$\begin{aligned}\mathbf{E}(h(Z)) &= \int_{\mathbb{R}} h\left(\frac{1+x}{1-x}\right) f_X(x)\ dx \\ &= \int_{\mathbb{R}} h(z) f_X\left(\frac{z-1}{z+1}\right)\frac{2}{(z+1)^2}\ dz\end{aligned}$$

On a en effet posé $z = (1+x)/(1-x)$, soit $x = (z-1)/(z+1) = 1 - 2/(z+1)$. Ce changement de variable est un difféomorphisme de $\mathbb{R}\backslash\{-1\}$ vers $\mathbb{R}\backslash\{1\}$. Remarquons qu'il est souvent plus pratique de raisonner en terme de jacobien plutôt qu'en terme de dérivée : il n'y a pas à s'occuper du sens des bornes des intégrales, mais il suffit de vérifier qu'un domaine s'envoie bien sur un autre domaine de $\mathbb{R}$ (et de ne pas oublier de mettre la valeur absolue de la dérivée et non la dérivée elle-même). On traite alors ce changement de variable comme un changement de variable dans une intégrale multiple.