Probabilités et variables aléatoires
Préparation à l'agrégation interne

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

1 Une petite liste de révisions

1.1 Convergence de suites et séries

Suite croissante majorée
Série à termes positifs
Suite convergente, série convergente, critère de Cauchy
Série absolument convergente
Série semi-convergente
Transformation d'Abel
Famille sommable
Série de fonctions : convergence uniforme, convergence normale ; séries entières.
Quelques sommes et séries à connaître : binôme de Newton, série(s) géométrique(s), série exponentielle

1.2 Convergence d'intégrales

Intégrale d'une fonction continue par morceaux sur un segment
Intégrale impropre : fonction continue par morceaux sur un intervalle I quelconque (ie sur tout segment inclus dans I).
Intégrale absolument convergente
Intégrale semi-convergente
Intégration terme à terme d'une série de fonctions.

1.3 Intégrales multiples

Intégrale sur un rectangle (Tonelli, Fubini) ; Chasles ; linéarité.
Intégrale d'une fonction continue sur un ensemble « géométriquement simple »
Changement de variables, notamment passage en coordonnées polaires

1.4 Intégrale d'une suite/famille de fonctions

Limite d'intégrales : convergences dominée et monotone
Continuité et dérivation sous l'intégrale

2 Tribu et probabilité

2.1 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 stable par intersection dénombrable.

Une méthode usuelle pour construire une tribu est de procéder par **engendrement** : on se donne une famille (A_i) de sous-ensemble de Ω et on considère la plus petite tribu les contenant. On peut remarquer que c'est l'intersection de toutes les tribus contenant les (A_i). Montrer que deux tribus sont égales se ramène ainsi souvent à démontrer que les ensembles qui engendrent l'une peuvent s'écrire comme réunions/intersections/complémentaires d'ensembles qui engendrent l'autre.

Exercice 1 1. *Expliciter la tribu engendrée par un événement A (non vide), puis par deux événements disjoints et non vides A et B.*

2. *Sur un ensemble Ω quelconque, quelle est la tribu engendrée par les singletons ?*

Le plus souvent, si Ω est dénombrable, la tribu utilisée sera $\mathcal{P}(\Omega)$. Sur $\mathbb{R}$, on travaille presque exclusivement avec la tribu borélienne : c'est la tribu engendrée par les ouverts (ou par les fermés, ou par les intervalles $]-\infty, a]$, a décrivant $\mathbb{R}$, ou par les intervalles $[a, b[$, a et b décrivant $\mathbb{R}$, avec $a \leq b$).

Exercice 2 *Montrer l'équivalence de ces définitions de la tribu borélienne.*

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 2.1 *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 (qui appartient donc à Σ), $\mathbf{P}(A) \geq 0$.*
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 2.2 *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, $\mathbf{P}(A \cup B) = \mathbf{P}(A) + \mathbf{P}(B) - \mathbf{P}(A \cap B)$.

Exercice 3 *Entrainez-vous en rédigeant les preuves de ces propriétés ! N'hésitez pas à faire des petits dessins pour alimenter votre réflexion.*

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 tribu Σ et d'une mesure de probabilité P.

N.B. Dans un problème concret, le choix de l'univers Ω, et de la mesure de probabilité dont on le munit, est une question parfois difficile ! Ne pas savoir expliciter l'espace Ω sur lequel on travaille est rarement un frein pour mener les calculs... On fera le plus souvent possible abstraction de toutes les questions liées à la tribu : si Ω est explicite, la tribu sera ~~le~~ plus souvent toujours soit l'ensemble des parties de Ω lorsque Ω est dénombrable, soit la tribu borélienne dans le cas où Ω est $\mathbb{R}$ ou $\mathbb{R}^n$.

2.2 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.

Les éléments ω de Ω tels que $\mathbf{P}(\{\omega\})$ est non nul sont appelés les atomes de la probabilité, et la valeur de $\mathbf{P}(\{\omega\})$ est le poids de ω.

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

Proposition 2.3 *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, qui doivent de plus appartenir à la tribu Σ.

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, qui constitue la tribu dont on munit Ω, 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** $\Omega = \{\omega_1, \ldots, \omega_n\}$, 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 équilibré à six faces de façon répétée jusqu'à obtenir un 6, et on note le numéro du lancer du premier 6. L'univers est constitué de l'ensemble de tous les résultats possibles (et c'est donc $\mathbb{N}^*$). On a évidemment $\mathbf{P}(\{1\}) = 1/6$.

  Pour que le premier 6 sorte au deuxième lancer, il faut avoir obtenu un nombre entre 1 et 5 au premier lancer, puis un 6 au deuxième lancer. Si on considère l'ensemble des résultats (équiprobables) de lancers successifs de deux dés, 5 couples de résultats conviennent, alors qu'il y a 36 résultats possibles, et qu'ils sont équiprobables. Il vient donc :

  $$\mathbf{P}(\{2\}) = \frac{5}{36}$$

  De même, pour tout $k \geq 2$, 5^{k-1} lancers de k dés produisent le premier 6 au k-ième lancer, pour 6^k résultats équiprobables de lancers successifs de k dés. On a donc :

  $$\mathbf{P}(\{k\}) = \frac{5^{k-1}}{6^k} = \left(\frac{5}{6}\right)^{k-1} \frac{1}{6}.$$

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.

★*Quel univers pour cette expérience ? Et quelle tribu, quelle probabilité ?* Ici, on a considéré comme univers l'ensemble des résultats possibles de l'expérience (numéro du lancer du premie "6"), mais on pourrait chercher à écrire un univers U permettant de décrire l'intégralité des résultats des tirages successifs, et le munir d'une tribu puis d'une probabilité $\widetilde{P}$.

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 $(x_i)_{i\leq n} \in \{1, \ldots, 6\}^n$ (on fixe les n premières composantes et on laisse les autres libres). Par ailleurs, on sait calculer la probabilité de tout événement ne faisant intervenir qu'un nombre fini de lancers, c'est-à-dire que l'on connaît la probabilité de chacun des événements qui engendrent la tribu : cela suffit pour caractériser complètement la probabilité sur U.

2.3 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 peut définir 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. Elle est définie sur la tribu borélienne de $\mathbb{R}$, notée $\mathcal{B}(\mathbb{R})$, qui est la tribu engendrée (par passage au complémentaire, réunions, intersections dénombrables) par les intervalles.

Des exemples

- La mesure uniforme sur l'intervalle $[a, b]$, où $a < b$: On définit une probabilité μ par

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

  La densité de μ est la fonction $\mathbf{1}_{[a,b]}/(b - a)$.

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

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

  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.

3 Probabilité conditionnelle, indépendance

Définition 3.1 *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*

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

On rencontre typiquement des probabilités conditionnelles dans les expériences aléatoires obtenues par des tirages successifs de boules dans des urnes, avec ou sans remise, mais il y a bien sûr d'autres cadres où elles apparaissent naturellement : en pratique dès que deux expériences ont lieu successivement ou simultanément. On résout ces questions le plus souvent en dressant un arbre de probabilité : les données figurant sur les arêtes de l'arbre sont des probabilités conditionnelles.

Exercice 4 *On procède à un lancer de dé à six faces équilibré. L'univers considéré est $\{1, \dots, 6\}$ muni de la probabilité uniforme. On note X le résultat du lancer et A l'événement « X est pair ». Déterminer $\mathbf{P}_A(\{\omega\})$ pour chacun des événements élémentaires, puis la probabilité que le résultat du lancer soit multiple de trois, sachant qu'il est pair. Posez-vous les mêmes questions avec un dé tétraédrique.*

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 incompatibles, $\mathbf{P}_B(A_1 \cup A_2) = \mathbf{P}_B(A_1) + \mathbf{P}_B(A_2)$, $\mathbf{P}_B(\Omega \backslash A) = 1 - \mathbf{P}_B(A)$...

On peut donc énoncer la proposition suivante :

Proposition 3.2 *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 $A \subset \Omega$, $\mathbf{P}_B(A) \geq 0$.*
- $\mathbf{P}_B(\Omega) = 1$
- *Si $(A_n)_n$ est une famille dénombrable d'événements deux à deux disjoints, $\mathbf{P}_B(\cup_n A_n) = \sum_n \mathbf{P}_B(A_n)$*

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 3.3 *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 dénombrable, 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 3.4 (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, tels que $\mathbf{P}(\bar{A}) > 0$. 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 $\mathbf{P}(A_i \cap B)$, d'où le résultat. □

Exercice 5 *On dispose d'un sac contenant dix jetons (3 noirs et 7 rouges).*

1. *On pioche simultanément trois jetons. On s'intéresse au nombre de jetons noirs piochés. Expliciter un univers permettant de modéliser cette situation, ainsi que la probabilité dont on le munit. Quelle est la probabilité d'obtenir 2 jetons noirs ? Quelle est la probabilité de tirer au moins un jeton noir sachant que l'on tire au moins un jeton rouge ?*

2. *On change légèrement l'expérience : on ne pioche plus simultanément les trois jetons, mais successivement sans remise. Expliciter un univers et une mesure de probabilité décrivant cette situation. Quelle est la probabilité d'obtenir un jeton noir sur chacun des deux premiers tirages ? Et d'obtenir exactement deux jetons noirs parmi les trois tirages ? Quelle est la probabilité de tirer un jeton rouge au 3ième tirage ? Quelle est la probabilité de tirer un jeton noir au premier tirage sachant que l'on tirera un jeton rouge au 3ième tirage ? Quelle est la probabilité de tirer au moins un jeton noir sachant que l'on tire au moins un jeton rouge ?*

Définition 3.5 *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 6

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 » peuvent généralement être supposés indépendants (sauf en cas de tricherie...)

- Lorsque l'on lance un dé à six faces, les événements « Le résultat est multiple de 3 » et « Le résultat est pair » sont indépendants, mais cela devient faux si on considère un dé tétraédrique, ou si le dé n'est pas équilibré.

- *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 $A = \{\text{tirer un bouton noir au premier tirage}\}$ et $B = \{\text{tirer un bouton jaune au deuxième tirage}\}$ sont-ils indépendants ?

- *Tirage sans remise.* On considère la même urne que ci-dessus, et on effectue deux tirages successifs et sans remise d'un jeton. Les événements « Obtenir un jeton jaune au 1er tirage »et « Obtenir un jeton jaune au 2ième tirage »sont de même probabilité mais ne sont pas 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 3.6 *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 :

• 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.

4 Schéma de Bernoulli

Un schéma de Bernoulli permet de modéliser la répétition (un nombre fini de fois) d'une expérience n'ayant que deux issues possibles (succès/échec), dans des conditions identiques et indépendantes. On notera (TOUJOURS) n le nombre de répétitions de l'expérience, et p la probabilité d'obtenir un succès lors de la première expérience. Par exemple, on lance cinq fois (successivement) un dé, le succès étant l'obtention d'un 6. On aura alors $p = 1/6$, et on pourra étudier la probabilité d'obtenir un 6 au 3e tirage, à celle d'obtenir quatre fois une autre face, puis finalement un 6, ou encore à la probabilité d'avoir obtenu exactement 3 fois un 6 lors de ces cinq répétitions.

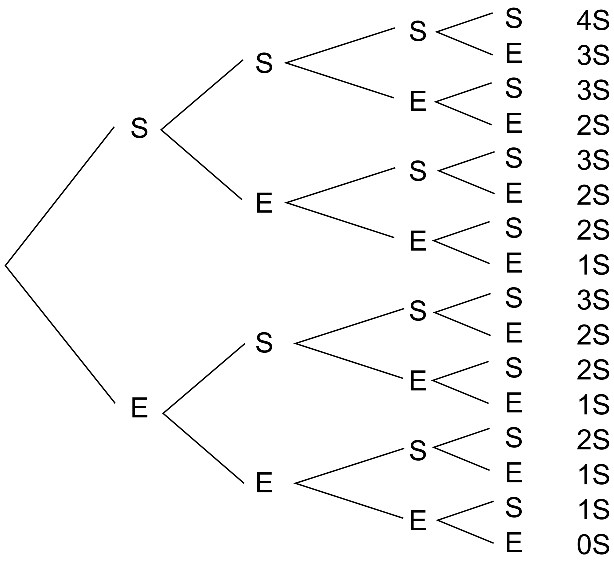
L'univers Ω le plus souvent choisi pour décrire les résultats de l'ensemble des n expériences sera $\Omega = \{S, E\}^n$ (ou $\{0, 1\}^n$), où $n \in \mathbb{N}^*$ désigne le nombre de répétitions de l'expérience. Il reste à décrire la probabilité dont on le munit, et pour cela, à déterminer la probabilité de chacun des événements élémentaires.

Un événement élémentaire est une suite de longueur n de succès et d'échec(s). Si p est la probabilité succès lors de la première expérience, la probabilité d'un événement élémentaire comportant exactement k succès (avec $0 \leq k \leq n$) est alors $p^k(1-p)^{n-k}$, peu importe l'ordre dans lequel interviennent les succès et les échecs.

Étudions maintenant la probabilité de l'événement

$$A_k = \{\text{il y a exactement } k \text{ succès parmi les } n \text{ tirages}\}.$$

Cet événement est constitué d'un certain nombre d'événements élémentaires qui ont tous pour probabilité $p^k(1-p)^{n-k}$. Reste à déterminer le nombre de tels événements élémentaires. Pour cela, on dresse un arbre faisant apparaître les résultats des tirages successifs, et on met en regard de chacun des chemins le nombre de succès obtenus.



Le nombre d'événements élémentaires de A_k correspond au nombre de chemins dans l'arbre comportant exactement k succès : il y en a $\binom{n}{k}$.

On peut alors conclure que $\mathbf{P}(A_k) = \binom{n}{k} p^k (1-p)^{n-k}$.

La loi binomiale de paramètre (n, p) est la loi de probabilité μ donnant le nombre de succès dans un schéma de Bernoulli de paramètre (n, p) : c'est une loi de probabilité sur $\{0, \dots, n\}$ et on a, pour tout $k \in \{0, \dots, n\}$,

$$\mu(\{k\}) = \binom{n}{k} p^k (1-p)^{n-k}.$$

Exercice 7 *Sur l'arbre représentant un schéma de Bernoulli, retrouver que* $\binom{n}{k} = \binom{n}{n-k}$*, ainsi que la formule du triangle de Pascal. Démontrer également la formule du binôme de Newton. Comment faire le lien entre le nombre de chemins comportant exactement k succès et le nombre de parties à k éléments dans un ensemble en contenant n ?*

5 Variables aléatoires réelles

5.1 La loi

5.1.1 Définition

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

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 $\mathbf{P}(I)$.

Pour tout intervalle I de $\mathbb{R}$, on note

$$\mu(I) = \mathbf{P}(\{X \in I\}) = \mathbf{P}(X^{-1}(I))$$

Énonçons la propriété fondamentale de μ :

Proposition 5.1 *La fonction μ ainsi définie induit une probabilité sur $\mathbb{R}$ (ou sur l'ensemble $X(\Omega)$).*

Définition 5.2 *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 $\mu(I) = \mathbf{P}(X^{-1}(I))$ 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 $X(\omega)$... mais, si on connaît la loi de X, on connaît l'ensemble $X(\Omega)$.

Définition 5.3

- *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 les résultats de 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.

5.2 Exemples de variables aléatoires discrètes

Définition 5.4 *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 8 *On se donne une variable aléatoire* $X : (\Omega, \Sigma) \to (\mathbb{N}, \mathcal{P}(\mathbb{N}))$. *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. On peut remarquer 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 $X(\Omega) = \{0, \ldots, n\}$ et, pour tout $k \in \{0, \ldots, n\}$, $\mathbf{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k}$. 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 $\{1, \ldots, n\}$. On a ici $X(\Omega) = \{1, \ldots, n\}$ et cette loi affecte le même poids à chacun des éléments. On a donc $\mathbf{P}(X = k) = 1/n$, pour tout $k \in \{1, \ldots, n\}$.

- *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 d'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, ou le nombre de véhicules passant à un carrefour pendant un quart d'heures, ou lr nombre de requêtes reçues par un serveur informatique pendant une seconde.

- *Loi hypergéométrique* : Soit r, b et n trois entiers naturel non nuls, avec $n \leq r + b$. 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, pour tout entier naturel k vérifiant $n - b \leq k \leq \min(r, n)$ et

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

5.3 Exemples de variables aléatoires à densité

- *Loi uniforme sur* $[a, b]$: c'est la loi d'une 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, ce qui n'empêche pas de calculer la valeur de cette intégrale. Il y a (au moins) trois méthodes classiques pour mener ce calcul : nous en présentons ici une basée sur une intégrale double. Une autre méthode est par exemple décrite dans la deuxième épreuve du CAPES externe 2014.

  Calculons donc I^2

$$
\begin{aligned}
I^2 &= \left(\int_{-\infty}^{+\infty} f(t)\ dt\right)^2 \\
&= \left(\int_{-\infty}^{+\infty} f(t)\ dt\right) \times \left(\int_{-\infty}^{+\infty} f(s)\ ds\right) \\
&= \int_{-\infty}^{+\infty} \left(\int_{-\infty}^{+\infty} f(s)\ ds\right) f(t)\ dt \qquad \text{[On a juste fait rentrer une constante dans une intégrale]} \\
&= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(s) f(t)\ ds\, dt \qquad \text{[Théorème de Tonelli]} \\
&= \int_{\mathbb{R}^2} e^{-(s^2+t^2)/2} \frac{ds\, dt}{2\pi}.
\end{aligned}
$$

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

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

On l'obtient par changement de variables à partir de la loi $\mathcal{N}(0,1)$: Y suit la loi $\mathcal{N}(m, \sigma^2)$ si et seulement si $(Y-m)/\sigma$ suit la loi $\mathcal{N}(0,1)$

Attention : la notation $\mathcal{N}(m, \sigma^2)$ n'est pas universelle ; c'est celle utilisée dans les programmes de terminale, mais dans la plupart des logiciels, on utilise la notation $\mathcal{N}(m, \sigma)$.

- Loi de Cauchy : c'est la loi de densité $x \mapsto 1/(\pi(1+x^2))$ sur $\mathbb{R}$.

6 Fonction de répartition d'une variable aléatoire

Définition 6.1 *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 : si deux variables aléatoires ont la même fonction de répartition, elles suivent la même loi.

Proposition 6.2 *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 $\lim_{x\to+\infty} F_X(x) = 1$.*

Proposition 6.3 • *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.

- *Si X suit la loi de densité f,*

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

La fonction de répartition de la loi de X est alors la primitive de sa densité, nulle en $-\infty$.

Remarque :

- Lorsque la variable aléatoire X est discrète, sa fonction de répartition est une fonction constante par morceaux : ce n'est pas très maniable. La fonction de répartition est peu utilisée dans ce contexte.
- Lorsque la variable aléatoire X admet une densité, sa fonction de répartition 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.

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.

7 Espérance d'une variable aléatoire discrète

7.1 Espérance de X

Soit $X : \Omega \to \mathbb{R}$ une variable aléatoire discrète.

Définition 7.1 *Une variable aléatoire X est dite* **intégrable** *si la quantité*

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

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

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

L'espérance est également appelée « moyenne » de la variable aléatoire X.

Proposition 7.2

1. *Si X est positive (c'est-à-dire si $\mathbf{P}(X \geq 0) = 1$) et intégrable, alors $\mathbf{E}(X) \geq 0$.*
2. *Si X est bornée par une constante M (c'est-à-dire si $\mathbf{P}(X \in [-M, M]) = 1$), alors X est intégrable et $\mathbf{E}(X) \in [-M, M]$.*
3. *Si X est constante, égale à a, alors $\mathbf{E}(X) = a$.*
4. *Si X et Y sont deux variables aléatoires intégrables, alors $X+Y$ est une variable aléatoire intégrable et $\mathbf{E}(X+Y) = \mathbf{E}(X) + \mathbf{E}(Y)$.*
5. *Si X et Y sont deux variables aléatoires intégrables vérifiant (en tout point de Ω) $X \leq Y$, alors $\mathbf{E}(X) \leq \mathbf{E}(Y)$.*

La preuve des deux premières propriétés est immédiate : on peut en effet supposer que $X(\Omega)$ est inclus dans $\mathbb{R}^+$ ou borné, ce qui induit ces deux résultats.

La preuve du troisième point est moins triviale et sera faite dans le paragraphe sur les couples aléatoires discrets. Celle du dernier point consiste à remarquer que la variable aléatoire $Y - X$ est positive et intégrable ; on utilise alors les points 1 et 3 pour conclure.

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

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

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 à l'intérieur de son disque de convergence) :

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

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

Exercice 9 *Soit X une variable aléatoire. On fixe un réel x. Quelle est la loi de $\mathbf{1}_{\{X\leq x\}}$? Quelle est son espérance ?*

7.2 Espérance d'une fonction de X

Soit X une variable aléatoire discrète et $h : X(\Omega) \to \mathbb{R}$ une fonction. Notons Y la variable aléatoire $Y = h(X)$. Y est une variable aléatoire discrète et, pour calculer son espérance et/ou vérifier son intégrabilité, on peut bien entendu commencer par déterminer sa loi, mais le théorème du transfert permet de se passer de cette étape parfois laborieuse :

Théorème 7.3 (Théorème du transfert) *Soit X une variable aléatoire et $h : X(\Omega) \to \mathbb{R}$ une fonction. Alors la variable aléatoire $h(X)$ est intégrable si et seulement si*

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

et on a dans ce cas

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

Preuve : Commençons par déterminer la loi de Y. Cette variable aléatoire est discrète, à valeurs dans $h(X(\Omega))$ et pour tout $y \in h(X(\Omega))$, on a

$$\mathbf{P}(Y = y) = \mathbf{P}(h(X) = y) = \mathbf{P}(X \in h^{-1}(\{y\}))$$

L'ensemble $h^{-1}(\{y\})$ est un sous-ensemble de $X(\Omega)$:

$$h^{-1}(\{y\}) = \left\{x \in h^{-1}(\{y\})\right\}$$

On a donc :

$$\mathbf{P}(Y = y) = \mathbf{P}\left(\bigcup_{x\in h^{-1}(\{y\})} (X^{-1}(\{x\}))\right)$$

Or les événements $((X^{-1}(\{x\}))_{x\in h^{-1}(\{y\})}$ sont deux à deux disjoints. On en déduit

$$\mathbf{P}(Y = y) = \sum_{x\in h^{-1}(\{y\})} \mathbf{P}(X^{-1}(\{x\})) = \sum_{x\in h^{-1}(\{y\})} \mathbf{P}(X = x)$$

Revenons à l'intégrabilité de Y : par définition, Y est intégrable si et seulement si la série

$$\sum_{y\in h(X(\Omega))} |y|\mathbf{P}(Y = y)$$

converge.

Or

$$\begin{aligned}\sum_{y\in h(X(\Omega))}|y|\mathbf{P}(Y=y) &= \sum_{y\in h(X(\Omega))}\left(|y|\sum_{x\in h^{-1}(\{y\})}\mathbf{P}(X=x)\right)\\ &= \sum_{y\in h(X(\Omega))}\left(\sum_{x\in h^{-1}(\{y\})}|y|\mathbf{P}(X=x)\right)\\ &= \sum_{y\in h(X(\Omega))}\left(\sum_{x\in h^{-1}(\{y\})}|h(x)|\mathbf{P}(X=x)\right)\end{aligned}$$

Penchons-nous sur les indices des deux séries : ce découpage en deux sommes successives correspond à une partition de $X(\Omega)$ suivant les valeurs prises par h. Puisque ces séries sont à termes positifs, on a donc

$$\sum_{y\in h(X(\Omega))}|y|\mathbf{P}(Y=y)=\sum_{x\in X(\Omega)}|h(x)|\mathbf{P}(X=x)$$

c'est-à-dire que $h(X)$ est intégrable si et seulement si la dernière série converge.

Les mêmes calculs sans valeurs absolues conduisent au calcul de l'espérance de $h(X)$. □

Le théorème du transfert permet de montrer très simplement les deux résultats suivants :

Proposition 7.4 *Soit X une variable aléatoire intégrable et a une constante. Alors les variables aléatoire $X+a$ et aX sont intégrables et on a $\mathbf{E}(X+a)=\mathbf{E}(X)+a$ et $\mathbf{E}(aX)=a\mathbf{E}(X)$.*

Notons également l'expression de l'espérance de X^2 lorsqu'elle existe :

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

La linéarité partielle énoncée ci-dessus se généralise la somme de deux variables aléatoires intégrables :

Proposition 7.5 *Si X et Y sont deux variables aléatoires intégrables (discrètes ou non), alors $X+Y$ est intégrable et on a $\mathbf{E}(X+Y)=\mathbf{E}(X)+\mathbf{E}(Y)$.*

7.3 Espérance et probabilité

Il est parfois compliqué de calculer ou majorer une probabilité, alors que le calcul de l'espérance peut s'avérer plus simple.

Proposition 7.6 *Inégalité de Markov. Soit X une variable aléatoire intégrable et λ un réel strictement positif. On a*

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

et plus généralement, pour tout $k\in\mathbb{R}^{+*}$ tel que $\mathbf{E}(|X|^k)<\infty$,

$$\mathbf{P}(|X|\geq\lambda)\leq\frac{\mathbf{E}(|X|^k)}{\lambda^k}$$

Preuve : Il suffit de remarquer que $|X|\geq\lambda\mathbf{1}_{|X|\geq\lambda}$ et d'intégrer pour obtenir le premier point. Le deuxième point s'en déduit par croissance de la fonction $x\mapsto x^p$ sur $\mathbb{R}^+$.

7.4 Variance et écart-type

Considérons une variable aléatoire X telle que $\mathbf{E}(X^2)$ soit fini ; cette variable aléatoire est dite de **carré intégrable**.

Remarque : Si X est de carré intégrable (c'est-à-dire $\mathbf{E}(X^2) < \infty$), alors X est intégrable : en effet, pour tout x réel, l'inégalité suivante est vérifiée :

$$|x| \leq x^2 + 1$$

(et on a même $|x| \leq (x^2+1)/2$).

On en déduit que $\mathbf{E}(|X|) \leq \mathbf{E}(X^2) + 1$, ce qui implique l'intégrabilité de X.

Utilisons la linéarité de l'espérance pour le calcul de l'espérance de $(X-\alpha)^2$, où α est une constante. On a

$$\mathbf{E}((X-\alpha)^2) = \mathbf{E}(X^2 - 2\alpha X + \alpha^2) = \mathbf{E}(X^2) - 2\alpha\mathbf{E}(X) + \alpha^2.$$

Cette expression est donc un polynôme positif et de degré 2 en α, qui atteint son minimum en $\mathbf{E}(X)$.

Définition 7.7 *La quantité* $\mathbf{E}((X-\mathbf{E}(X))^2)$ *est appelée la variance de* X*. On note* $\mathbf{var}\,(X) = \mathbf{E}((X-\mathbf{E}(X))^2)$*. L'écart-type est la racine de la variance.*

Remarque :

- L'espérance d'une variable aléatoire est un paramètre dit de position (si on translate toutes les valeurs de X d'une même quantité, on translate son espérance de la même quantité). L'écart-type et la variance sont des paramètres de dispersion : plus la variance est grande, plus les valeurs prises par X ont tendance à s'éloigner de la moyenne de X. Pour observer ce phénomène, calculer la variance de la variable aléatoire X_a, à valeurs dans $\{-a, a\}$ et prenant chacune de ces deux valeurs avec probabilité 1/2.
- La variance est toujours positive.

Proposition 7.8

- **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.*

Pour déterminer la variance d'une variable aléatoire discrète X, on sera fréquemment amené à calculer $\mathbf{E}(X^2)$; par le théorème du transfert, on a

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

Néanmoins, dans le cas des variables aléatoires à valeurs entières, il est souvent plus simple de commencer par calculer $\mathbf{E}(X(X-1))$:

$$\mathbf{E}(X(X-1)) = \sum_{x \in X(\Omega)} x(x-1)\mathbf{P}(X=x) \qquad \text{puis} \qquad \mathbf{E}(X^2) = \mathbf{E}(X(X-1)) + \mathbf{E}(X).$$

Exemple : Calcul de la variance d'une variable aléatoire de loi géométrique. On a ici :

$$\begin{aligned}
\mathbf{E}(X(X-1)) &= \sum_{n\geq 1} n(n-1)\mathbf{P}(X=n) \\
&= \sum_{n\geq 1} n(n-1)p(1-p)^{n-1} \\
&= p(1-p)\sum_{n\geq 0} \frac{d^2(1-p)^n}{dp^2} \\
&= p(1-p)\frac{d^2(1/p)}{dp^2} \qquad \text{[série entière]} \\
&= p(1-p)\frac{2}{p^3}
\end{aligned}$$

Puis

$$\mathbf{var}\,(X) = \mathbf{E}(X(X-1)) + \mathbf{E}(X) - \mathbf{E}(X^2) = \frac{2(1-p)}{p^2} + \frac{1}{p} - \frac{1}{p^2} = \frac{1-p}{p}.$$

□

Appliquons l'inégalité de Markov à la variable aléatoire $X - \mathbf{E}(X)$ et $k = 2$:

Proposition 7.9 (Inégalité de Bienaymé-Tchebychev) *Soit X une variable aléatoire de carré intégrable. Pour tout λ strictement positif, on a*

$$\mathbf{P}(|X - \mathbf{E}(X)| \geq \lambda) \leq \frac{\mathbf{var}\,(X)}{\lambda^2}$$

7.5 Inégalité de Cauchy-Schwarz

L'inégalité de Cauchy-Schwarz est une inégalité classique de produits scalaires : si u et v sont deux vecteurs, alors $\langle u, \rangle v \leq \|u\|\|v\|$. Cette inégalité permet notamment de définir le cosinus d'un angle. Elle s'étend aux variables aléatoires :

Théorème 7.10 *Soient X et Y deux variables aléatoires de carré intégrable. Alors la variable aléatoire XY est intégrable et on a*

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

Preuve : La première inégalité provient de la croissance de l'espérance : on a $-|XY| \leq XY \leq |XY|$ ce qui donne en intégrant : $-\mathbf{E}(|XY|) \leq \mathbf{E}(XY) \leq \mathbf{E}(|XY|)$ d'où le résultat.

Pour démontrer la deuxième inégalité, considérons le polynôme $R : \lambda \mapsto \mathbf{E}((|X| + \lambda|Y|)^2)$. En développant, il vient , pour tout λ,

$$R(\lambda) = \lambda^2\mathbf{E}(Y^2) + 2\lambda\mathbf{E}(|XY|) + \mathbf{E}(X^2)$$

Ce polynôme de degré 2 est de signe constant : son discriminant est par conséquent négatif ou nul

$$4\mathbf{E}(|XY|)^2 - 4\mathbf{E}(X^2)\mathbf{E}(Y^2) \leq 0$$

ce qui est le résultat souhaité. □

Pour varier les plaisirs, une autre preuve, suggérée par Denis Roussillat. On peut supposer que X et Y ne sont pas constamment nulles. On a alors $\mathbf{E}(X^2) > 0$ et $\mathbf{E}(Y^2) > 0$. Utiliser alors la positivité de

$$\mathbf{E}\left(\left(\frac{|X|}{\sqrt{\mathbf{E}(X^2)}} - \frac{|Y|}{\sqrt{\mathbf{E}(Y^2)}}\right)^2\right)$$

pour conclure. □

7.6 Fonction génératrice

Soit X une variable aléatoire à valeurs dans $\mathbb{N}$. Remarquons que pour tout $s \in [-1, 1]$, la variable aléatoire s^X est bornée par 1, donc intégrable, et considérons la fonction $s \mapsto \mathbf{E}(s^X)$, appelée **série génératrice de la loi de** X et notée G_X.

Proposition 7.11
- *Pour tout $s \in [-1, 1]$, on a*

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

- *Si X est à valeurs dans $\mathbb{N}$, alors $s \mapsto \mathbf{E}(s^X)$ est une série entière.*

Théorème 7.12 *Soit X une variable aléatoire à valeurs entières positives.*

1. *La série génératrice est une série entière de rayon de convergence supérieur ou égal à 1 et elle caractérise la loi de X au sens où deux variables aléatoires de même série génératrice sont de même loi.*

2. *G_X est dérivable en 1 si et seulement si X est intégrable et on a $G'_X(1) = \mathbf{E}(X)$. Plus généralement, G_X est k fois dérivable en 1 (où k est un entier non nul) si et seulement si X^k est intégrable et on a $G_X^{(k)}(1) = \mathbf{E}(X(X-1)\dots(X-k+1))$*

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

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

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 $\mathbf{var}\,(X) = m$.

8 Espérance d'une variable aléatoire à densité

8.1 Espérance

On se donne une variable aléatoire X de densité f.

Définition 8.1 *On dit que X est intégrable si l'intégrale $\int_{-\infty}^{+\infty} |x|f(x)\ dx$ est convergente et on définit dans ce cas l'espérance de X comme étant le réel $\mathbf{E}(X)$ suivant :*

$$\mathbf{E}(X) = \int_{-\infty}^{+\infty} xf(x)\ dx$$

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

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

Cette intégrale s'intègre par parties :

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

On retrouve dans le cadre des variables aléatoires à densité les mêmes propriétés que pour les variables aléatoires discrètes : la linéarité, la croissance, les inégalités de Markov, Bienaymé-Tchebychev et Cauchy-Schwarz...

On admet le théorème du transfert pour les variables aléatoires à densité :

Théorème 8.2 *Pour toute fonction h continue par morceaux sur $\mathbb{R}$ et telle que*

$$\int_{-\infty}^{+\infty} |h(t)|f(t)\ dt$$

converge, on a

$$\mathbf{E}(h(X)) = \int_{-\infty}^{+\infty} h(t)f(t)\ dt$$

Exercice 10 *On considère une variable aléatoire X et un réel x. Calculer $\mathbf{E}(\mathbf{1}_{X\leq x})$.*

8.2 Variance

Rappelons la définition de la variance : si X est une variable aléatoire de carré intégrable, $\mathbf{var}\,(X) = \mathbf{E}((X - \mathbf{E}(X))^2)$.

Que X soit discrète ou à densité, la variance est une quantité positive, et on a $\mathbf{var}\,(X) = \mathbf{E}(X^2) - (\mathbf{E}(X))^2$, ainsi que, pour toute constante a, $\mathbf{var}\,(X + a) = \mathbf{var}\,(X)$ et $\mathbf{var}\,(aX) = a^2\mathbf{var}\,(X)$.

Rappelons, à toutes fins utiles, le calcul de l'espérance de X^2, qui se déduit du théorème du transfert :

Proposition 8.3 *Si X est une variable aléatoire de densité f, alors X est de carré intégrable si et seulement si l'intégrale $\int_{\mathbb{R}} x^2 f(x)\ dx$ est convergente et on a alors*

$$\mathbf{E}(X^2) = \int_{\mathbb{R}} x^2 f(x)\ dx$$

Exemple : Calcul de la variance d'une variable aléatoire Y de loi exponentielle de paramètre λ. On a vu (cf p.22) 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 $\mathbf{var}\,(Y) = 1/\lambda^2$.

9 Comment calculer la loi d'une variable aléatoire

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. Les mêmes techniques s'appliquent lorsque la variable aléatoire Y est définie comme une fonction explicite d'un couple (ou d'un vecteur) aléatoire dont on connaît la loi.

9.1 Loi discrète

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. Peu importe ici que la loi de X soit discrète ou non.

9.2 Fonction de répartition

On a vu que la fonction de répartition n'est pas un outil très maniable pour les variables aléatoires discrète, mais si on pense que Y va avoir une densité, on peut imaginer calculer sa fonction de répartition :

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

Pour obtenir la densité, il restera à dériver cette fonction de répartition en tout point où cela est possible.

Si la fonction g est monotone, on imagine facilement comment évaluer $\mathbf{P}(g(X) \leq y)$... mais sinon ?

9.3 Une fonction test

La méthode habituellement utilisée, et qui a l'avantage de s'appliquer en toute dimension, 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, mais il est tout à fait possible de traiter séparément les cas $a > 0$ et $a < 0$.

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.

9.4 Laplace, Fourier, fonction génératrice

Nous avons déjà vu que, pour les variables aléatoires positives, la fonction génératrice $s \mapsto \mathbf{E}(s^X)$ caractérise la loi de X, et que cette transformation est utilisée essentiellement dans le cas des variables aléatoires discrètes à valeurs dans $\mathbb{N}$. Des transformations plus générales existent : la transformée de Laplace $\lambda \mapsto \mathbf{E}(e^{-\lambda X})$, définie sur $\mathbb{R}^+$, caractérise la loi des variables aléatoires positives (discrètes ou à densité) et la fonction caractéristique, ou transformée de Fourier, $\alpha \mapsto \mathbf{E}(e^{i\alpha X})$, définie sur $\mathbb{R}$, caractérise la loi de toute variable aléatoire, au sens où deux variables aléatoires positives de même transformée de Laplace sont de même loi, et deux variables aléatoires de même fonction caractéristique sont de même loi. Ceci est basé, au moins pour les variables aléatoires à densité, sur l'injectivité de la transformée de La place (ou de Fourier).

On peut vérifier que lorsque X est une variable aléatoire positive et λ est un réel positif, la variable aléatoire $e^{-\lambda X}$ est bien intégrable, et que c'est également le cas de $e^{i\alpha X}$ pour une variable aléatoire X quelconque. Ces transformations sont utiles par exemple pour déterminer la loi de la somme de deux variables aléatoires (de préférence indépendantes), mais aussi pour étudier des convergences de suites de variables aléatoires.

10 lois usuelles

10.1 Lois discrètes

Nom	Paramètres	Loi	Espérance	Variance	$\mathbf{E}s^X,\ s \in [-1,1]$
Bernoulli	$p \in [0,1]$	$\mathbf{P}(X=0) = 1-p,\ \mathbf{P}(X=1) = p$	p	$p(1-p)$	$1-p(1-s)$
Binomiale	$n \in \mathbb{N}^*, p \in [0,1]$	$\mathbf{P}(X=k) = \binom{n}{k} p^k (1-p)^{n-k}, k \in \{0,\dots,n\}$	np	$np(1-p)$	$(1-p(1-s))^n$
Géométrique	$p \in]0,1[$	$\mathbf{P}(X=k) = p(1-p)^{k-1},\ k \in \mathbb{N}^*$	$\dfrac{1}{p}$	$\dfrac{1-p}{p^2}$	$\dfrac{ps}{1-(1-p)s}$
Uniforme	$n \in \mathbb{N}^*$	$\mathbf{P}(X=k) = 1/n$ pour tout $k \in \{1,\dots,n\}$	$\dfrac{n+1}{2}$	$\dfrac{n^2-1}{12}$	$s\dfrac{1-s^n}{n(1-s)}$
Poisson	$\mu \in [0,+\infty[$	$\mathbf{P}(X=k) = e^{-\mu}\dfrac{\mu^k}{k!}$ pour tout $k \in \mathbb{N}$	μ	μ	$e^{-\mu(1-s)}$
Hypergéométrique	$n, r, b \in \mathbb{N}, n \leq r+b$ ou $n, N = r+b, p = r/(r+b)$	$\mathbf{P}(X=k) = \binom{b}{n-k} / \binom{r+b}{n}$	np	$np(1-p)\frac{N-n}{N-1}$	

Interprétation de ces lois

- **Loi de Bernoulli.** On joue à pile ou face et on note le résultat du tirage sous la forme : $X = 1$ si la pièce tombe sur « pile » et $X = 0$ si la pièce tombe sur « face ». La probabilité de tomber sur « pile » est alors $\mathbf{P}(X=1) = p$.

- **Loi binomiale.** On relance cette fois-ci n fois la même pièce et on comptabilise le nombre de fois où elle tombe sur « pile ». La loi de ce nombre de « piles » parmi les n premiers lancers est alors une loi binomiale de paramètres n et p : en effet, s'il y a eu exactement k « piles » au cours des n premiers lancers (d'où le p^k), il y a eu également $n-k$ « faces » (d'où le $(1-p)^{n-k}$), et il faut tenir compte de tous les ordres de tirages possibles, ce qui explique le $\binom{n}{k}$.

- **Loi géométrique.** Toujours au jeu de pile ou face, on étudie maintenant le numéro du premier lancer où un « pile » est obtenu : si le premier « pile » arrive au n^{e} lancer, c'est que la pièce est tombée sur « face » lors des $n-1$ premiers lancers, et sur « pile » au n^{e}. La probabilité de cet événement est donc $(1-p)^{n-1} \times p$.

- **Loi uniforme.** La loi uniforme sur un ensemble à n éléments intervient dès que ces éléments sont équiprobables, par exemple si l'on joue avec un dé non pipé, ou une pièce équilibrée.

- **Loi de Poisson.** Cette loi permet de modéliser le nombre de personnes se présentant à un guichet au cours d'une journée : si, en moyenne, μ personnes arrivent dans le service chaque jour, on utilisera la loi de Poisson de paramètre μ.

- **Loi hypergéométrique.** C'est la loi obtenue lorsque l'on étudie le nombre de boules rouges obtenues en piochant n boules dans une urne contenant r boules rouges et b boules blanches.

10.2 Lois à densité

nom	paramètres	densité	$\mathbf{E}X$	**var** X	$\mathbf{E}\exp -\lambda X$	$\mathbf{E}\exp i\alpha X$
Uniforme	$a < b$	$\mathbf{1}_{[a,b]}(x)/(b-a)$	$\frac{a+b}{2}$	$\frac{(b-a)^2}{12}$	$\frac{e^{-\lambda a}-e^{-\lambda b}}{\lambda(b-a)}$	$\frac{e^{i\alpha b}-e^{i\alpha a}}{i\alpha(b-a)}$
Exponentielle	μ	$\mu e^{-\mu x}\mathbf{1}_{\mathbb{R}^+}(x)$	$\frac{1}{\mu}$	$\frac{1}{\mu^2}$	$\frac{\mu}{\mu+\lambda}$ pour $\lambda > -\mu$	$\frac{\mu}{\mu-i\alpha}$
Normale	m, σ^2	$\frac{\exp -((x-m)^2/(2\sigma^2))}{\sqrt{2\pi}\ \sigma}$	m	σ^2	$e^{-\lambda m}e^{\lambda^2\sigma^2/2}$	$e^{i\alpha m}e^{-\alpha^2\sigma^2/2}$
Cauchy	μ	$\frac{\mu}{\pi(\mu^2+x^2)}$	non définie	non définie	$+\infty$	$e^{-\mu\|\alpha\|}$
Student	n	$\frac{\Gamma((n+1)/2)}{\sqrt{n\pi}\Gamma(n/2)(1+x^2/n)^{(n+1)/2}}$	0, si $n \geq 2$	$\frac{n}{n-2}$, si $n \geq 3$	$+\infty$	compliquée
Gamma	μ, n	$\frac{\mu^n x^{n-1}e^{-\mu x}}{\Gamma(n)}\mathbf{1}_{\mathbb{R}^+}(x)$	$\frac{n}{\mu}$	$\frac{n}{\mu^2}$	$\frac{\mu^n}{(\mu+\lambda)^n}$	$\frac{\mu^n}{(\mu-i\alpha)^n}$
Chi 2 (χ^2)	n	$\frac{x^{n/2-1}e^{-x/2}}{2^{n/2}\Gamma(n/2)}\mathbf{1}_{\mathbb{R}^+}(x)$	n	$2n$	$(1+2\lambda)^{-n/2}$	$(1-2i\alpha)^{-n/2}$

Rappel : $\Gamma(x) = \int_0^\infty t^{x-1}e^{-t}\,dt$, pour tout $x \in \mathbb{R}_+^*$ et $\Gamma(n) = (n-1)!$ pour tout $n \in \mathbb{N}^*$

Les paramètres : σ, μ sont des nombres réels positifs, a, b et m sont des nombres réels, n est un nombre entier strictement positif.

Remarques :

- La fonction caractéristique permet toujours de caractériser la loi d'une variable aléatoire, qu'elle soit discrète, à densité, ou d'une autre forme. La transformée de Laplace permet de le faire lorsque l'intégrale (ou la somme) est convergente au moins pour tout $\lambda \geq 0$: la loi de Cauchy et la loi de Student (pour tous les choix de paramètre) ont la même transformée de Laplace (égale à $+\infty$), mais sont bien des lois différentes pour $n > 1$.

- La somme de n vaiid exponentielles de paramètre λ suit une loi Gamma (n, λ).
- La somme des carrés de n vaiid normales $\mathcal{N}(0,1)$ suit une loi du χ^2_n.
- Le quotient de deux variables normales centrées indépendantes suit une loi de Cauchy.
- Les lois normales, de Student et du χ^2 interviennent en statistique. Leur fonction de répartition est donnée dans des tables.
- Une variable aléatoire de loi de Cauchy n'est pas intégrable, et sa transformée de Laplace n'existe pas : l'intégrale est divergente.
- La loi de Student de paramètre $n = 1$ est une loi de Cauchy.