Chapitre 2

Variables Aléatoires

Après avoir réalisé une expérience, on ne s'intéresse bien souvent à une certaine fonction du résultat et non au résultat en lui-même. Lorsqu'on regarde une portion d'ADN, au lieu de vouloir connaitre tout la suite de nucléotides, on peut vouloir juste connaitre le nombre d'apparition d'un "”mot“". Ces grandeurs (ou fonctions) auxquelles on s'intéresse sont en fait des fonctions réelles définies sur l'ensemble fondamental et sont appelées **variables aléatoires**.

On considère un ensemble Ω muni d'une probabilité $\mathbb{P}$.

Définition 0.1 Une **variable aléatoire** X est une fonction de l'ensemble fondamental Ω à valeurs dans $\mathbf{R}$, $X : \Omega \to \mathbf{R}$.
Lorsque la variable X ne prend que des valeurs discrètes, on parle de **variable aléatoire discrète**.
Un **vecteur aléatoire** $X : \Omega \to \mathbf{R}^d$ est une fonction $X = (X_1, \ldots, X_d)$ à valeurs dans $\mathbf{R}^d$ telle que les coordonnées X_i soient des variables aléatoires.

Pour tout intervalle $[a, b] \subset \mathbf{R}$, l'ensemble $\{X \in [a, b]\} = \{\omega \in \Omega : X(\omega) \in [a, b]\}$ est un événement.

Exemple 0.2

1. On jette deux dés distincts et on s'intéresse à la somme des points. On note X cette variable aléatoire, elle est définie par

$$\begin{array}{rccl} X : & \Omega & \to & \mathbf{R} \qquad \text{avec } \Omega = \{(1,1), (1,2), \ldots, (6,5), (6,6)\} \\ & (\omega_1, \omega_2) & \mapsto & \omega_1 + \omega_2. \end{array}$$

   L'ensemble des valeurs possibles de X est $\{2, 3, \ldots, 12\}$.

2. On lance toujours deux dés, mais cette fois on s'intéresse au plus grand chiffre Y obtenu. On a alors

$$\begin{array}{rccl} Y : & \Omega & \to & \mathbf{R} \qquad \text{avec } \Omega = \{(1,1), (1,2), \ldots, (6,5), (6,6)\} \\ & (\omega_1, \omega_2) & \mapsto & \max(\omega_1, \omega_2). \end{array}$$

   La variable Y est à valeurs dans $\{1, 2, \ldots, 6\}$.

3. On observe deux bactéries et on s'intéresse à la durée de vie T de la bactérie qui disparaîtra la première. L'ensemble fondamental est $\Omega = [0, +\infty[\times[0, +\infty[$. La variable T s'écrit alors

$$\begin{array}{rcl} T: \quad \Omega & \longrightarrow & \mathbf{R} \\ (\omega_1, \omega_2) & \mapsto & \inf\{\omega_1, \omega_2\}. \end{array}$$

Seul le dernier exemple n'est pas une variable discrète.

1 Loi de probabilité, Fonction de répartition

La loi de probabilité d'une variable aléatoire permet de connaitre les chances d'apparition des différentes valeurs de cette variable.
On se place sur l'espace de probabilité $(\Omega, \mathbb{P})$.

Définition 1.1 Soit X une variable aléatoire. La **loi de probabilité** de X est définie par la fonction F_X, appelée **fonction de répartition** de la variable X, définie par

$$\begin{array}{rcl} F_X: \quad \mathbf{R} & \to & [0,1] \\ x & \mapsto & \mathbb{P}(X \leq x). \end{array}$$

On dit que deux variables aléatoires X et Y ont la **même loi** si elles ont la même fonction de répartition $F_X = F_Y$.

Remarque 1.2 Soit I un intervalle de $\mathbf{R}$. L'événement $\{X \leq x\}$ représente l'ensemble des valeurs $\omega \in \Omega$ telles que $X(\omega)$ soit inférieur à x, i.e.$\{X \leq x\} = \{\omega \in \Omega : X(\omega) \leq x\}$.

La loi de X est en générale notée $\mathcal{L}(X)$ ou $\mathcal{L}oi(X)$.

Remarque 1.3 On a $\mathbb{P}(X \in \mathbf{R}) = 1$, car $\mathbb{P}(X \in \mathbf{R}) = \mathbb{P}(\{\omega \in \Omega : X(\omega) \in \mathbf{R}\}) = \mathbb{P}(\Omega) = 1$.

Propriétés 1.4 La fonction de répartition est une fonction croissante à valeur dans $[0, 1]$ telle que $\lim_{x \to -\infty} F_X(x) = 0$ et $\lim_{x \to +\infty} F_X(x) = 1$, mais elle n'est pas forcément continue.

Remarque 1.5 Soit $a \leq b$, on a $\mathbb{P}(X \in [a, b]) = \mathbb{P}(X \leq b) - \mathbb{P}(X < a)$.

1.1 Loi d'une variable discrète

La fonction de répartition d'une variable discrète est constante par morceaux. Si X est une variable discrète à valeurs dans $\{x_1, \ldots, x_n\}$ avec $x_1 < \ldots < x_n$ alors pour $x \in \mathbf{R}$

$$F_X(x) = \sum_{i=1}^{k} \mathbb{P}(X = x_i) \quad \text{avec } k \text{ tel que } x_k \leq x < x_{k+1}.$$

De même, si X prend une infinité de valeurs $\{x_1, \ldots, x_n, \ldots\}$ avec $x_1 < \ldots < x_n \ldots$, on a pour $x \in \mathbf{R}$

$$F_X(x) = \sum_{i=1}^{k} \mathbb{P}(X = x_i) \quad \text{avec } k \text{ tel que } x_k \leq x < x_{k+1}.$$

Les sauts de la fonction de répartition F_X ont lieu en les points x_i et la hauteur du saut au point x_i est égale à $\mathbb{P}(X = x_i)$. Il suffit donc de calculer la fonction de répartition aux points x_i.

Proposition 1.6 Si X est à valeurs discrètes dans $\{x_1, \ldots, x_n\}$ (ou $\{x_1, \ldots, x_n, \ldots\}$), la loi de X est entièrement caractérisée par $\{\mathbb{P}(X = x_i) : i \geq 1\}$. (En effet, voir p. 8)

On remarque que

1. pour tout $i \geq 1$, $\mathbb{P}(X = x_i) \in [0, 1]$,
2. $\displaystyle\sum_{i \geq 1} \mathbb{P}(X = x_i) = 1$. (En effet, $1 = \mathbb{P}(X \in \mathbf{R}) = \sum_{i \geq 1} \mathbb{P}(X = x_i)$.)

Exemple 1.7 Reprennons les deux premiers exemples précédents, qui décrivent effectivement des variables discrètes. Pour trouver la loi de ces variables on utilise la proposition 1.6.

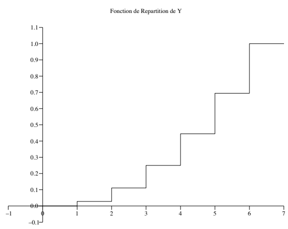
1. X est à valeurs dans $\{2, 3, \ldots, 12\}$, donc $\mathbb{P}(X = k) = 0$ pour $k \notin \{2, 3, \ldots, 12\}$.

k	2	3	4	5	6	7	8	9	10	11	12
$\mathbb{P}(X = k)$	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36

2. Y est à valeurs dans $\{1, 2, \ldots, 6\}$, donc $\mathbb{P}(Y = k) = 0$ pour $k \notin \{1, 2, \ldots, 6\}$.

k	1	2	3	4	5	6
$\mathbb{P}(Y = k)$	1/36	3/36	5/36	7/36	9/36	11/36
$F_Y(k)$	1/36	4/36	9/36	16/36	25/36	1

La fonction de répartition de Y est



1.2 Lois discrètes usuelles

Loi de Bernoulli, $\mathcal{B}(p)$, avec $p \in]0, 1[$.

Une variable aléatoire X de Bernoulli est une variable qui ne prend que deux valeurs : l'échec (au quel on associe la valeur 0) et le succès (auquel on associe la valeur 1) d'une expérience. Cette expérience est appelée **épreuve de Bernoulli**. Par exemple, on souhaite savoir si une cellule est atteinte par un virus. On associe la valeur 1 si elle est atteinte (succès) et la valeur 0 si elle est saine (échec).
La loi est donnée par : $P(X = 1) = p \quad P(X = 0) = 1 - p$.

Loi Binomiale, $\mathcal{B}(n, p)$, avec $p \in]0, 1[$, $n \in \mathbb{N}$, $n \geq 1$.

Exemple 1.8 On étudie les mésanges bleues parmi la population de mésange en Ille et Vilaine. La mésange n'est pas un oiseau migrateur et vit en groupes formés de plusieurs espèces de mésanges. On poste alors des observateurs à différents endroits dans le département. On note p la proportion de mésanges bleues parmi la population de mésange. Quelle est la probabilité que le nombre de mésanges bleues sur 1000 mésanges étudiées soit égal à k ?

La loi Binomiale est utilisée pour modéliser un “sondage avec remise“. C’est la loi du nombre de succès lorsqu’on renouvelle n fois de manière indépendante une épreuve de Bernoulli de paramètre p. On note X le nombre de succès obtenus à l’issue des n épreuves. Sa loi s’appelle loi Binomiale de paramètres n et p. On a

$$P(X=k) = \begin{pmatrix} n \\ k \end{pmatrix} p^k (1-p)^{n-k} \text{ avec } k \in \{0,1,..,n\}.$$

On peut écrire le nombre de succès X à l’aide des résultats de chaque épreuve de Bernoulli. On note X_i le résultat de la $i^{\text{ème}}$ expérience :

$$X_i = \begin{cases} 1 \text{ si la } i^{\text{ème}} \text{ expérience est réussie} \\ 0 \text{ si la } i^{\text{ème}} \text{ expérience est un échec} \end{cases}$$

On a alors $X = X_1 + \ldots + X_n$. Toute variable de Bernoulli peut s’écrire de cette façon.

Loi Hypergéométrique $H(N,m,n)$, avec $N \geq 1$, $(m,n) \in \{1,..,N\}^2$.

Exemple 1.9 Un lac contient N poissons. On en pêche m qu’on marque et qu’on remet à l’eau. On pêche à nouveau n poissons. Quelle est la probabilité de pêcher k poissons marqués ?

La loi hypergéométrique est utilisée pour modéliser un ”sondage sans remise“. C’est le cas de pratiquement tous les sondages (notamment lorsqu’on veut étudier la conformité d’un lot de médicaments, étudier le nombre de cellules atteintes par un virus, etc...).

La loi est donnée par : $P(X=k) = \dfrac{\begin{pmatrix} m \\ k \end{pmatrix}\begin{pmatrix} N-m \\ n-k \end{pmatrix}}{\begin{pmatrix} N \\ n \end{pmatrix}}$ si $k \in \{0,..,\min(m,n)\}$.

Remarque 1.10 Reprenons, l’exemple des poissons. Lorsque la taille N de la population de poissons dans le lac est très grande, le fait d’enlever 1 poisson ne changera quasiment pas la proportion $p = m/N$ de poissons marqués. De même, si on en enlève un nombre fini n. Par conséquent, chaque poisson pêché a à peu près la même probabilité p d’être marqué. On peut donc approcher la loi Hypergéométrique par la loi binomiale $\mathcal{B}(n,p)$ où p est la proportion de ”poissons marqués“.

De manière générale, lorsque la taille de la population est grande, on utilise souvent la binomiale même pour un ”sondage sans remise”.

Loi Géométrique, $\mathcal{G}(p)$, avec $p \in]0,1[$.
La loi géométrique est la loi du premier succès.

Exemple 1.11 On lance une pièce truquée jusqu’à ce qu’on obtienne une fois "Pile". On note p la probabilité de tomber sur “Pile”. On veut connaitre la probabilité d’avoir “Pile” au premier lancer, au deuxième, ..., au $k^{\text{ième}}$ lancer, ... On note X le nombre de lancers nécessaires pour avoir un succès.

La loi est donnée par : $P(X=k) = p(1-p)^{k-1}$ avec $k \in \mathbb{N}$, $k \geq 1$.
Une formule utile quand on veut faire des calculs avec la loi géométrique :

$$1 + x^2 + \ldots + x^n = \frac{1 - x^{n+1}}{1-x}.$$

On peut calculer explicitement la fonction de répartition : $F(i) = 1 - (1-p)^i$ avec $i \geq 1$.

Loi de Poisson, $\mathcal{P}(\lambda)$, avec $\lambda > 0$ un réel.
La loi de Poisson est utilisée pour modéliser le comptage d'événements rares, c'est à dire des événements ayant une faible probabilité de réalisation : maladies rares, accidents mortels rares, le titrage d'une solution virale, mutations ou recombinaisons dans une séquence génétique, pannes, radioactivité...

Exemple 1.12 Une plaque d'Agar est une boîte de Pétri stérile qui contient un milieu de culture utilisée pour la culture des micro-organismes. Elles sont traitées avec un antibiotique. Les bactéries qui sont réparties sur les plaques ne peuvent se multiplier sauf les rares résistantes aux antibiotiques. Elles forment des colonies. Le décompte des colonies suit la loi de Poisson.

La loi est donnée par $P(X=k) = \frac{\lambda^k}{k!} e^{-\lambda}$ avec $k \in \mathbb{N}$.
Une formule utile :

$$e^x = 1 + x + \frac{x^2}{2} + \frac{x^3}{3!} + \ldots = \sum_{k=0}^{\infty} \frac{x^k}{k!}$$

1.3 Loi d'une variable aléatoire à densité (ou continues)

Considérons la durée de vie d'une bactérie. On conçoit facilement que la probabilité que cette durée de vie vaille exactement une certaine valeur est nulle. Par exemple, il est quasiment impossible qu'une bactérie vive exactement 1 an 0 mois, 0 heure, 0 minute, ... La fonction de répartition d'une telle variable est par conséquent continue. On peut par contre s'intéresser à la probabilité que la bactérie vive moins d'un an.
On ne verra dans ce cours que des variables qui sont soit discrètes soit continues même s'il existe des variables plus complexes.

Définition 1.13 Une variable aléatoire X est **à densité**, ou **continue**, s'il existe une fonction f définie sur $\mathbf{R}$ telle que la fonction de répartition de X s'écrit

$$\forall x \in \mathbf{R} \quad F_X(x) = \int_{-\infty}^{x} f(t)dt$$

où f est une fonction intégrable sur $\mathbf{R}$ satisfaisant les conditions suivantes :

1. $f(t) \geq 0$ pour tout $t \in \mathbf{R}$,
2. $\displaystyle\int_{-\infty}^{+\infty} f(t)dt = 1$.

Une fonction qui vérifie les conditions 1. et 2. est appelée **densité de probabilité**.

Propriétés 1.14 Soit X une variable aléatoire à densité.
Alors pour tout $x \in \mathbf{R}$, $\mathbb{P}(X = x) = 0$.

Preuve. Soit $x \in \mathbf{R}$. On considère l'intervalle réduit à un point $I = \{x\}$. On a

$$\mathbb{P}(X = x) = \mathbb{P}(X \in I) = \int_x^x f(t)dt = 0.$$

△

Remarque 1.15
1. La probabilité $\mathbb{P}(X \in [a, b]) = \int_a^b f(t)dt$ correspond à l'aire de la surface comprise entre la courbe de f et l'axe des abcisses sur l'intervalle $[a, b]$.
2. La fonction de répartition d'une variable à densité est continue.

Proposition 1.16 Soit X une variable aléatoire de fonction de répartition F_X. Si F_X est continue sur $\mathbf{R}$ et dérivable sur $\mathbf{R}$ (sauf peut-être en un nombre fini de points), alors X est une variable à densité f donnée par $f(x) = F'_X(x)$.

1.4 Lois à densité usuelles

Loi Uniforme, $\mathcal{U}([a, b])$, avec $a, b \in \mathbb{R}$, $a < b$.
Cette loi est l'analogue continu de l'équiprobabilité dans le cas discret. Elle permet de modéliser le tirage d'un nombre aléatoire dans l'intervalle $[a, b]$.

Exemple 1.17 Lors d'une étude du comportement animal, on a relaché des oiseaux dont l'orientation a été rendue très difficile. On s'attend alors à ce que les oiseaux choisissent au hasard leur direction. On peut modéliser la direction prise par un oiseau de la façon suivante. On considère X l'angle entre le nord et la direction prise par l'oiseau (selon le sens des aiguilles d'une montre). La variable X suit une loi uniforme entre 0 et 360 degrés.

Densité :

$$\begin{aligned} f(x) &= \frac{1}{b-a} \quad \text{si } x \in [a, b] \\ &= 0 \quad \text{sinon} \end{aligned}$$

Fonction de répartition :

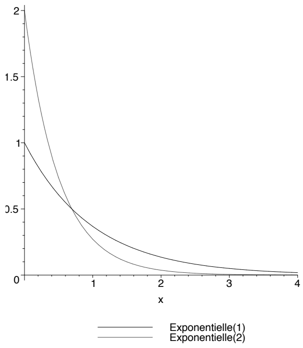
$$\begin{aligned} F(x) &= 0 \quad \text{si } x < a \\ &= \frac{x-a}{b-a} \quad \text{si } x \in [a, b] \\ &= 1 \quad \text{si } x > b \end{aligned}$$

Loi Exponentielle, $\mathcal{E}(\lambda)$, avec $\lambda > 0$ un réel.
Cette densité de probabilité permet en général de modéliser des durées de vie d'êtres non soumis au vieillissement (par exemple, la durée de vie d'une bactérie) ou des temps d'attente (par exemple, le temps d'attente entre deux signaux synaptiques).

Exemple 1.18 Dans une substance radioactive, la désintegration des noyaux se fait de façon spontanée. Le nombre de désintegration sur un intervalle de temps fixé suit une loi de Poisson. Par contre le temps d'attente entre deux désintégrations est modélisé par une loi exponentielle.

Densité :

$$\begin{aligned} f(x) &= \lambda e^{-\lambda x} && \text{si } x \geq 0 \\ &= 0 && \text{si } x < 0 \end{aligned}$$



Densité de lois exponentielles

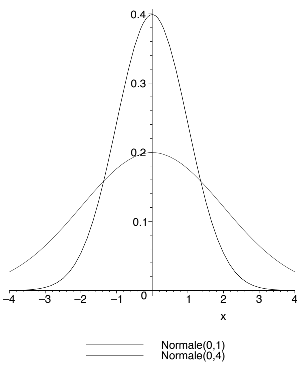
Fonction de répartition :

$$\begin{aligned} F(x) &= 0 && \text{si } x < 0 \\ &= 1 - e^{-\lambda x} && \text{si } x \geq 0 \end{aligned}$$

La loi exponentielle est la seule loi continue qui vérifie la propriété d'absence de mémoire : Si $X \sim \mathcal{E}(\lambda)$, alors pour tout $s, t > 0$ $\mathbb{P}(X > t + s | X > t) = \mathbb{P}(X > s)$.
La loi exponentielle est celle de la mortalité des êtres qui ne seraient pas soumis au veillissement : à chaque moment ils ont la même probabilité de mourrir dans l'unité de temps qu'il leur reste, quelque soit leur âge.

Loi Normale (ou loi Gaussienne), $\mathcal{N}(m, \sigma^2)$, avec $m \in \mathbb{R}$, $\sigma > 0$ réel.
La loi Normale est une loi centrale dans la théorie des probabilités. Elle est notamment très utilisée en statistique. Une grandeur influencée par un grand nombre de paramètres indépendants est souvent modélisée par une loi normale (par exemple, les erreurs de mesures lors d'une expérience).
Densité :

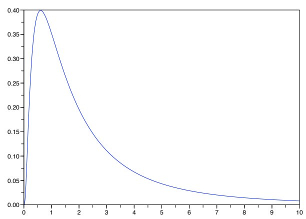
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}} \quad \text{avec } x \in \mathbb{R}.$$



Densité de lois normales

Loi Log-Normale, $LN\left(m, \sigma^2\right)$, avec $m \in \mathbb{R}$, $\sigma > 0$ réel.
Une variable X suit une loi Log-Normale $LN\left(m, \sigma^2\right)$ si $\ln(X)$ suit une loi normale $\mathcal{N}\left(m, \sigma^2\right)$.
Cette loi permet de modéliser par exemple le temps de survie des bactéries en présence de désinfectant, le dosage de certains médicaments ...
Densité :

$$f(x) = \frac{1}{\sigma x \sqrt{2\pi}} e^{-\dfrac{(\ln(x) - m)^2}{2\sigma^2}} \quad \text{avec } x \in > 0.$$



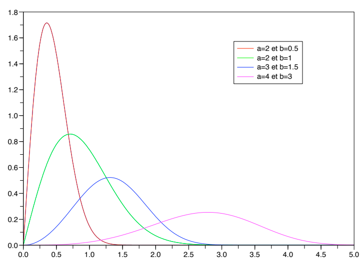
Densité d'une loi Log-Normale

Loi de Weibull, $W(a, b)$, avec $a > 0$, $b > 0$ réels.
La loi de Weibull est utilisée en démographie pour modéliser le veillissement et en épidémiologie pour modéliser la distribution de probabilité de la durée d'incubation d'une maladie infectieuse.
Densité :

$$f(x) = \frac{a}{b} \left(\frac{x}{b}\right)^{a-1} \exp\left(-\left(\frac{x}{b}\right)^a\right) \quad \text{avec } x \in > 0.$$

Fonction de répartition :

$$F(x) = 1 - \exp\left(-\left(\frac{x}{b}\right)^a\right) \quad \text{avec } x \in > 0.$$



Densité de lois de Weibull

On constate que pour $a = 1$, on retrouve la loi exponentielle. La loi exponentielle est celle de la mortalité des être vivants qui ne seraient pas soumis au veillissement : à chaque moment ils ont la même probabilité de mourrir dans l'unité de temps qu'il leur reste, quelque soit leur âge. Plus a est grand plus le veillissement se fait pesant (la mortalité augmente avec l'âge). Le cas $a < 1$ correspond à un monde dans lequel plus on veillirait, moins forte serait la probabilité de mourir dans l'unité de temps qui vient.

On va maintenant définir les paramètres de position et de dispersion d'une loi.

2 Espérance et variance d'une variable aléatoire

2.1 Valeur moyenne d'une variable : l'espérance

L'idée intuitive de l'espérance puise son origine dans les jeux de hasard. Considérons le jeu suivant : on lance un dé plusieurs fois de suite. Supposons que pour une mise de 1 euro, on gagne 1 euro si le résultat obtenu est pair, 2 euros si le résultat est 1 ou 3, et on perd 3 euros si le résultat est 5. Est-il intéressant de jouer à ce jeu ? Quel peut-être le gain moyen ? Soit X la variable aléatoire correspondant au nombre d'euros gagnés ou perdus. La loi de X est

k	-3	1	2
$\mathbb{P}(X=k)$	$1/6$	$1/2$	$1/3$

L'espérance de gain, noté $\mathbb{E}[X]$, est alors $\mathbb{E}[X] = -3\cdot 1/6 + 1\cdot 1/2 + 2\cdot 1/3 = 2/3$. Le joueur gagne donc en moyenne $2/3$ d'euros pour une mise de 1 euro...

Définition 2.1 L'**espérance** d'une variable aléatoire X est notée $\mathbb{E}[X]$. Elle représente la valeur moyenne prise par la variable X.

1. Si X est une variable discrète à valeurs dans $\mathbb{D} = \{x_1, \ldots, x_n\}$, son espérance est
$$\mathbb{E}[X] = x_1\mathbb{P}(X = x_1) + \ldots + x_n\mathbb{P}(X = x_n) = \sum_{i=1}^{n} x_i\mathbb{P}(X = x_i).$$
2. Si X est une variable discrète à valeurs dans l'ensemble infini $\mathbb{D} = \{x_i : i \geq 1\}$, lorsque la somme est bien définie, son espérance est
$$\mathbb{E}[X] = \sum_{i=1}^{+\infty} x_i\mathbb{P}(X = x_i).$$
3. Si X est une variable à densité f, lorsque l'intégrale est bien définie, son espérance est
$$\mathbb{E}[X] = \int_{-\infty}^{+\infty} xf(x)dx.$$

Lorsqu'une variable X vérifie $\mathbb{E}[X] = 0$, on dit que la variable est **centrée**.

Propriétés 2.2
1. L'espérance est linéaire :
soient a et $b \in \mathbf{R}$, deux variables aléatoires X et Y d'espérance finie alors
$$\mathbb{E}[aX + bY] = a\mathbb{E}[X] + b\mathbb{E}[Y].$$
2. Si $X \geq 0$, alors $\mathbb{E}[X] \geq 0$.
3. Si $X \leq Y$, alors $\mathbb{E}[X] \leq \mathbb{E}[Y]$.

Exemple 2.3 Supposons que la durée de vie T d'une bactérie est modélisée par la loi exponentielle de densité $f(t) = \lambda \exp(-\lambda t)$ pour $t \geq 0$ pour une certaine valeur de λ. Alors sa durée de vie moyenne est $\mathbb{E}(T) = 1/\lambda$.

2.2 Espérance d'une fonction

On considère une variable aléatoire X et une fonction h. On aimerait connaitre l'espérance de la nouvelle variable $Y = h(X)$.

Proposition 2.4 Soit X une variable aléatoire, h une fonction définie sur $\mathbf{R}$. Alors :

1. si X est une variable discrète à valeurs dans $\mathbb{D} = \{x_1, \ldots, x_n\}$, on a

$$\mathbb{E}\left[h(X)\right] = h(x_1)\mathbb{P}(X = x_1) + \ldots + h(x_n)\mathbb{P}(X = x_n) = \sum_{i=1}^{n} h(x_i)\mathbb{P}(X = x_i).$$

2. si X est une variable discrète à valeurs dans l'ensemble infini $\mathbb{D} = \{x_i : i \geq 1\}$, lorsque la somme est bien définie, on a

$$\mathbb{E}\left[h(X)\right] = \sum_{i=1}^{+\infty} h(x_i)\mathbb{P}(X = x_i).$$

3. si X est une variable à densité f, lorsque l'intégrale est bien définie, on a

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

2.3 Mesurer la diffusion d'une variable : variance et écart type

On a vu que l'espérance correspondait à la valeur moyenne d'une variable aléatoire. L'écart type représente l'écart moyen (la distance moyenne) entre la variable et sa moyenne. Elle mesure la dispersion d'une variable, plus l'écart-type est grand plus la variable prend des valeurs qui peuvent être éloignées les unes des autres, plus l'écart-type est petit plus la variable prend des valeurs proches de sa moyenne.

Définition 2.5 La **variance** d'une variable aléatoire X, notée $Var(X)$, est définie par

$$Var(X) = \mathbb{E}\left[(X - E(X))^2\right]$$

L'**écart type** est la racine carrée de la variance :

$$\sigma(X) = \sqrt{Var(X)}.$$

Lorsqu'une variable X vérifie $Var(X) = 1$, on dit que la variable est **réduite**.

Remarque 2.6 La variance s'écrit aussi $Var(X) = \mathbb{E}\left[X^2\right] - \mathbb{E}[X]^2$.

Preuve. Il suffit de développer le carré et d'utiliser la linéarité de l'espérance. △

Propriétés 2.7
1. $Var(X) = 0$ ssi X est constante.
2. Soient a et $b \in \mathbf{R}$, alors $Var(aX + b) = a^2 Var(X)$.

Preuve. Comme $\mathbb{E}[aX+b] = a\mathbb{E}[X]+b$, on a par définition $Var(aX+b) = \mathbb{E}[(aX - a\mathbb{E}[X])^2] = a^2 Var(X)$. △

Exemple 2.8 Supposons que la durée de vie T d'une bactérie est modélisée par la loi exponentielle de densité $f(t) = \lambda \exp(-\lambda t)$ pour $t \geq 0$ pour une certaine valeur de λ. La variance de la durée de vie de la bactérie étudiée est $Var(T) = 1/\lambda^2$.

3 Une fonction remarquable : la transformée de Laplace

Cette fonction est surtout utile en évolution des populations. En effet, si on considère une population dont chaque génération évolue de façon indépendante de la précédente mais selon la même loi, pour connaître s'il y a un risque d'extinction de la population, on utilise la transformée de Laplace.

Définition 3.1 Soit X une variable aléatoire à valeurs dans $\mathbf{R}$. Pour $t \in \mathbf{R}$, sa **transformée de Laplace** est définie par

$$L_X(t) = \mathbb{E}[e^{tX}]$$

pour $t \in \mathbf{R}$.

Proposition 3.2 La transformée de Laplace L_X d'une variable X vérifie les propriétés suivantes

1. L_X est à valeurs dans $[0, +\infty]$, $L_X(0) = 1$,
2. L_X caractérise la loi de X (i.e., si deux variables ont la même transformée de Laplace, alors elles ont la même loi),
3. si $\mathbb{E}[|X|^k] < \infty$ alors L_X est dérivable k fois en 0 et la dérivée $k^{\text{ième}}$ en zéro vaut $L_X^{(k)}(0) = \mathbb{E}[X^k]$.

4 Récapitulatif des lois usuelles

4.1 Lois de probabilité discrètes

Loi de Bernoulli, $\mathcal{B}(p)$, avec $p \in]0,1[$:
$P(X=1)=p \quad P(X=0)=1-p$.

Loi Binomiale, $\mathcal{B}(n,p)$, avec $p \in]0,1[$, $n \in \mathbb{N}^*$:
$P(X=k)=\begin{pmatrix} n \\ k \end{pmatrix} p^k (1-p)^{n-k}$ avec $k \in \{0,1,..,n\}$.

Loi Géométrique, $\mathcal{G}(p)$, avec $p \in]0,1[$:
$P(X=k)=p(1-p)^{k-1}$ avec $k \in \mathbb{N}^*$.

Loi de Poisson, $\mathcal{P}(\lambda)$, avec $\lambda > 0$:
$P(X=k)=\frac{\lambda^k}{k!}e^{-\lambda}$ avec $k \in \mathbb{N}$.

Loi	**Notation**	**Espérance**	**Variance**	**Transformée de Laplace**
Bernoulli	$\mathcal{B}(p)$	p	$p(1-p)$	$1-p+pe^t$
Binomiale	$\mathcal{B}(n,p)$	np	$np(1-p)$	$(1-p+pe^t)^n$
Géométrique	$\mathcal{G}(p)$	$\frac{1}{p}$	$\frac{1-p}{p^2}$	$\dfrac{pe^t}{1-(1-p)e^t}$
Poisson	$\mathcal{P}(\lambda)$	λ	λ	$e^{\lambda(e^t-1)}$

4.2 Lois de probabilité à densité

Loi	**Notation**	**Densité**	**Espérance**	**Variance**	**Transformée de Laplace**
Uniforme	$\mathcal{U}([a,b])$	$\dfrac{1}{b-a}$ pour $x \in [a,b]$	$\dfrac{a+b}{2}$	$\dfrac{(b-a)^2}{12}$	$\dfrac{e^{tb}-e^{ta}}{b-a}$
Exponentielle	$\mathcal{E}(\lambda)$, $\lambda > 0$	$\lambda e^{-\lambda x}$ pour $x \geq 0$	$\dfrac{1}{\lambda}$	$\dfrac{1}{\lambda^2}$	$\dfrac{\lambda}{\lambda-t}$ si $t < \lambda$
Normale	$\mathcal{N}(m,\sigma^2)$, $\sigma^2 > 0$	$\dfrac{1}{\sigma\sqrt{2\pi}} e^{-\dfrac{(x-m)^2}{2\sigma^2}}$	m	σ^2	$e^{tm+\frac{\sigma^2 t^2}{2}}$

Remarque 4.1 Dans *Méthodes statistiques* de P. Tassi (Ed. Economica), on retrouve les principales lois de probabilité ainsi que leurs caractéristiques (espérance, variance, etc...).

5 Exercices sur le chapitre 2

Exercice 2.7.
Dans les expériences suivantes, définir la variable aléatoire considérée et chercher la loi (ou les lois) la mieux adaptée au problème. Essayez d'expliquer d'où vient l'aléa.

1. Sur une plaque de verre quadrillée, on verse de façon homogène un liquide contenant des cellules simples. On fait attention à ce que le liquide ait une épaisseur constante sur la plaque. Sous le microscope nous comptons le nombre de carrés sans cellules, avec une cellule, avec deux cellules, etc. . .
2. Une plaque d'Agar est une boîte de Pétri stérile qui contient un milieu de culture utilisée pour la culture des micro-organismes. Elles sont traitées avec un antibiotique. Les bactéries qui sont réparties sur les plaques ne peuvent se multiplier sauf les rares résistantes aux antibiotiques. Elles forment des colonies. On décompte le nombre de colonies.
3. Dans une substance radioactive, la désintegration des noyaux se fait de façon spontanée. On note λ la constante radioactive du noyau, i.e. la probabilité de désintégration par unité de temps : entre les instants $t+h$ et t la probabilité de désintégration du noyau est λh. On s'intéresse au nombre de désintegration sur un intervalle de temps fixé et au temps d'attente entre deux désintégrations.

Exercice 2.8.
Une urne contient 3 sortes de boules de poids différents : 7 boules de poids 1kg, 5 boules de poids 3kg et 3 boules de poids 5kg.
On tire au hasard une boule de l'urne et on note X son poids.

1. Déterminer la loi de la variable X.
2. Calculer l'espérance et la variance de X.

Exercice 2.9.
Considérons deux parents hétérozygotes de génétopye Aa tels que leur enfants peuvent avoir les génotypes AA, Aa ou aa avec probabilité

$$\mathbb{P}(AA) = 1/4 \quad \mathbb{P}(Aa) = 1/2 \quad \mathbb{P}(aa) = 1/4.$$

Supposons qu'ils aient 4 enfants.

1. Quelle est la probabilité qu'exactement l'un deux aient le génotype aa ?
2. Quelle est la probabilité qu'au moins l'un deux ait le génotype aa ?

Exercice 2.10.
La proportion des groupes sanguins en France est environ :

$$A = 44\% \quad B = 13\% \quad AB = 3\% \quad O = 40\%$$

On considère la répartition de ces différents groupes sur 50 étudiants.

1. Donner la loi de la variable X égale au nombre d'étudiants de groupe O.
2. Donner la loi de la variable Y égale au nombre d'étudiants de groupe AB.
3. Calculer $\mathbb{P}(Y \leq 5)$, l'espérance et la variance de Y.

Exercice 2.11.
Considérons les enfants de parents hétérozygotes de génétopye Aa. La distribution des enfants est

$$\mathbb{P}(AA) = 1/4 \quad \mathbb{P}(Aa) = 1/2 \quad \mathbb{P}(aa) = 1/4.$$

On choisit de façon aléatoire 240 de ces enfants. On définit N_1, N_2, N_3 le nombre d'enfants de génotype AA,Aa et aa respectivement.

1. Quelle est la loi de N_1 ? N_2 ? N_3 ? Calculer l'espérance et la variance pour chaque génotype.
2. Quel est le lien entre ces différentes variables ?
3. Vérifier que pour $k_1, k_2, k_3 \in \mathbb{N}$,

$$\mathbb{P}(N_1 = k_1, N_2 = k_2, N_3 = k_3) \neq \mathbb{P}(N_1 = k_1)\mathbb{P}(N_2 = k_2)\mathbb{P}(N_3 = k_3).$$

Exercice 2.12. *Loi de Hardy-Weinberg*
En 1908, un mathématicien anglais G.H. Hardy et un médecin allemand W. Weinberg ont formulé une loi, connue sous le nom de loi de Hardy-Weinberg, qui concerne les fréquence alléliques pour un gène pouvant s'exprimer sous la forme de deux allèles A et a dans une population diploïde idéale. Une population diploïde est dite idéale si : la population est de grande taille (idéalement de taille infinie), panmictique (pas de choix du conjoint en fonction de son génotype), sans mutation, migration, ni sélection et dont les générations ne se chevauchent pas.
La loi de Hardy-Weinberg s'énonce de la façon suivante : dans une population idéale, les fréquences alléliques d'un gène s'exprimant sous la forme de deux allèles A et a sont constantes au cours du temps.
La distribution génotypique à la génération 0 est

$$P(AA) = r \quad P(Aa) = s \quad P(aa) = t.$$

1. Donner un exemple de population idéale (ou presque).
2. Montrer que les fréquences des génotypes AA, Aa et aa à la génération 1 sont $R = (r + s/2)^2$, $S = 2(r + s/2)(t + s/2)$ et $T = (t + s/2)^2$.
3. On pose $\theta = r + s/2$. Réexprimer les fréquences des génotypes AA, Aa et aa à la génération 1 en fonction de θ. En déduire que $S^2 = 4TR$.
4. Quelles sont les fréquences des génotypes AA, Aa et aa à la génération 2 ? à la génération n, pour $n \geq 2$?

Exercice 2.13.
Soit X une v.a. de densité f définie par $f(x) = 0$ si $x < 0$ et $f(x) = xe^{-x^2/2}$ sinon.

A-1 Vérifier que f est bien une densité de probabilité.

A-2 Chercher la loi et l'espérance de $Y = X^2$.

On s'intéresse maintenant aux précipitations à Brest. On suppose que la durée (en heure) T entre deux précipitations vérifie la relation $T = \lambda Y$, où λ est une constante. En consultant les relevés météo, on se rend compte qu'il s'écoule au moins 24h consécutives sans pluie avec une probabilité 0.09.

B-1 Déterminer λ.

B-2 Combien de temps s'écoule-t-il en moyenne entre deux averses ?

B-3 Il est midi. La pluie vient de s'arrêter. Quelle est la probabilite qu'il pleuve avant 14h ?

Exercice 2.14.
Soit X une variable aléatoire à valeurs dans $\mathbb{N}$ de loi de Poisson $\mathcal{P}(\lambda)$:

$$P\left(X = k\right) = \frac{\lambda^k}{k!} e^{-\lambda} \text{ avec } k \in \mathbb{N}$$

Donner l'expression de la transformée de Laplace de X.