Lycée Henri IV
MPSI

Cours
Durée : 5 h 00

VARIABLES ALÉATOIRES FINIES

Les énoncés et parties suivis du symbole [✠] ne seront pas traités en cours.

MPSI – Lycée Henri IV – Sébastien Kerner

Prérequis

Revoir les chapitre sur :

- les sommes ;
- le dénombrement ;
- les probabilités finies.

Dans tout ce chapitre, $(\Omega, \mathscr{T}, P)$ désigne un espace probabilisé fini, où $\mathscr{T}$ désigne le clan complet $\mathscr{P}(\Omega)$.

Les trois lettres AQT signifient « Âne Qui Trotte » et sont utilisées pour désigner une démonstration facile laissée au lecteur.

a. Variables aléatoires finies

À l'issue d'une expérience aléatoire, on associe souvent une valeur numérique au résultat. Par exemple, si un joueur lance trois dés, on peut compter le nombre de 6 obtenus ou encore le nombre de valeurs paires sorties.

Mathématiquement, cette association entre un événement et une grandeur variable régie par le hasard fournit une « fonction aléatoire », que l'on désigne traditionnellement sous le nom de variable aléatoire.

Tout ce chapitre est consacré à l'étude de ces objets ; à leurs opérations relatives et aux différentes lois de probabilités qui régissent le comportement aléatoire de leurs résultats.

Définition 1

Soient $(\Omega, \mathscr{T})$ un espace probabilisable fini (avec $\mathscr{T} = \mathscr{P}(\Omega)$) et E un ensemble quelconque.

On appelle variable aléatoire sur $(\Omega, \mathscr{T})$ à valeurs dans E toute application X de Ω vers E. Pour toute partie A de E, on note $(X \in A)$ l'événement $X^{-1}(A) = \{\omega \in \Omega : X(\omega) \in A\}$. En particulier, on note $(X = x)$ l'événement $\{\omega \in \Omega : X(\omega) = x\}$.

Dans le cas où $E = \mathbb{R}$, on parle de variable aléatoire réelle, ou v.a.r. en abrégé. Pour tous $a, b \in \mathbb{R}$, on note alors $(X \leqslant b)$ l'événement $\{\omega \in \Omega : X(\omega) \leqslant b\}$, $(X \geqslant a)$ l'événement $\{\omega \in \Omega : X(\omega) \geqslant a\}$, $(a \leqslant X \leqslant b)$ l'événement $\{\omega \in \Omega : a \leqslant X(\omega) \leqslant b\}$, etc.

Les valeurs prises par X ont en fait beaucoup plus d'importance que les antécédents de ces valeurs. Ainsi, on s'intéressera davantage à l'univers image $X(\Omega)$ qu'à l'univers Ω lui-même. Dans certaines expériences, on verra qu'il est très simple de décrire explicitement $X(\Omega)$ alors que l'on ne saurait en faire de même pour Ω.

C'est dans cet esprit que l'on utilise la notation $(X = x)$ à la place de $\{\omega \in \Omega : X(\omega) = x\}$: cela permet d'insister sur le fait que X prend la valeur x, sans se soucier des issues ω pour lesquelles X prend cette valeur.

En première année, l'univers Ω étant toujours fini, on est certain que l'univers image $X(\Omega)$ l'est également. On parle alors de variable aléatoire finie. L'an prochain, vous rencontrerez des univers infinis sur lesquels existent des variables dont l'univers image est parfois fini (on parlera encore de variable aléatoire finie) et parfois infini.

Plus généralement, on peut définir des variables aléatoires réelles sur un espace probabilisable $(\Omega, \mathscr{T})$ où $\mathscr{T}$ n'est pas le clan complet. Mais (tout a un prix !), la définition se complique : une variable aléatoire réelle sur $(\Omega, \mathscr{T})$ est une application X de Ω dans $\mathbb{R}$ telle que, pour tout $x \in \mathbb{R}$, la partie $(X \leqslant x)$ est un événement de Ω (c'est-à-dire un élément de $\mathscr{T}$). Tant que l'on suppose que Ω est un univers fini, cette complication est inutile puisqu'on peut travailler avec le clan complet $\mathscr{P}(\Omega)$. Lorsque Ω est infini (vous envisagerez ce cas l'an prochain), il n'est pas toujours possible de choisir $\mathscr{P}(\Omega)$ pour ensemble des événements et l'on doit définir les variables aléatoires avec la condition qui assure que $(X \leqslant x)$ est bien un événement pour tout $x \in \mathbb{R}$.

Exemples :

- Une variable aléatoire constante est dite certaine.
- Exemple fil rouge : Un joueur lance deux fois de suite une pièce de monnaie. À chaque lancer, s'il obtient pile, il gagne 1 € et s'il obtient face, il perd 2 €. On note $\Omega = \{\text{Pile,Face}\}^2$ l'univers des possibles et X la variable aléatoire égale au gain algébrique du joueur. Alors $X(\Omega) = \{-4; -1; 2\}$ et l'on a (avec des notations transparentes)
$$(X = -4) = \{(\mathrm{F}, \mathrm{F})\}, \qquad (X = -1) = \{(\mathrm{F}, \mathrm{P}), (\mathrm{P}, \mathrm{F})\} \qquad \text{et} \qquad (X = 2) = \{(\mathrm{P}, \mathrm{P})\}.$$
- Si S désigne un événement, la fonction $\mathbb{1}_S$, appelée indicatrice de l'événement S, qui vaut 1 lorsque l'événement S se produit et 0 sinon, est une variable aléatoire finie.

À toute variable aléatoire est naturellement associé le système complet d'événements qui prend en compte, une par une, les valeurs prises par la variable.

Proposition 1

Soit X une variable aléatoire finie. Si $X(\Omega) = \{x_1, x_2, \ldots, x_n\}$ désigne l'ensemble des valeurs possibles de X, alors la famille d'événements $(X = x_1)$, $(X = x_2)$, $\ldots$, $(X = x_n)$ est un système complet d'événements, appelé système complet d'événements associé à X.

■ Les événements $(X = x_1), (X = x_2), \ldots, (X = x_n)$ sont clairement incompatibles deux à deux et l'on a $(X = x_1) \sqcup (X = x_2) \sqcup \cdots \sqcup (X = x_n) = \big(X \in X(\Omega)\big) = \Omega$, ce qui établit le résultat. ■

B. Loi d'une variable aléatoire

B.1. Loi et densité

Proposition 2

Soit X une variable aléatoire finie. On appelle loi de probabilité de X l'application P_X définie par

$$P_X \left\{ \begin{array}{ccc} \mathscr{P}(X(\Omega)) & \longrightarrow & [0;1] \\ A & \longmapsto & P(X \in A) \end{array} \right.$$

C'est une mesure de probabilité sur $X(\Omega)$ muni du clan complet $\mathscr{P}\big(X(\Omega)\big)$.

■ On a $P_X\big(X(\Omega)\big) = P\big(X \in X(\Omega)\big) = P(\Omega) = 1$. D'autre part, pour $A, B \in \mathscr{P}\big(X(\Omega)\big)$ incompatibles, on a $P_X(A \sqcup B) = P(X \in A \sqcup B) = P\big((X \in A) \sqcup (X \in B)\big) = P(X \in A) + P(X \in B) = P_X(A) + P_X(B)$. Ainsi, P_X est bien une mesure de probabilité sur $\big(X(\Omega), \mathscr{P}(X(\Omega))\big)$. ■

Nous verrons qu'une loi de probabilité ne détermine pas une variable aléatoire : il existe des variables associées à des phénomènes distincts qui ont la même loi de probabilité.

En pratique, il n'est pas simple de donner la loi de probabilité P_X. On préfère en général donner la densité de X.

Proposition 3

Soit X une variable aléatoire finie. On appelle densité (singulière) de probabilité de X l'application f_X définie par

$$f_X \left\{ \begin{array}{ccc} X(\Omega) & \longrightarrow & [0;1] \\ x & \longmapsto & P(X = x) \end{array} \right.$$

La donnée de f_X permet de reconstituer la loi P_X.

■ Pour $A \in \mathscr{P}\big(X(\Omega)\big)$, la formule $P_X(A) = \sum_{a \in A} f_X(a)$ permet de reconstituer P_X à partir de f_X. ■

On fait souvent la confusion entre f_X et P_X. Par conséquent, quand on donne f_X, on dit que l'on se donne la loi de probabilité de X.

Lorsque X n'est pas une variable finie, la définition de la densité de X est plus compliquée.

Présentation d'une densité

Pour définir la loi de probabilité d'une variable aléatoire finie X, on peut utiliser deux types de représentations de la densité de X :

- ▶ ou bien le tableau donnant les probabilités des valeurs prises par X :

$x \in X(\Omega)$	x_1	x_2	$\cdots$	x_n
$P(X = x)$	p_1	p_2	$\cdots$	p_n

- ▶ ou bien le graphe en bâtonnet de f_X : le bâtonnet d'abscisse x_k est de hauteur p_k.

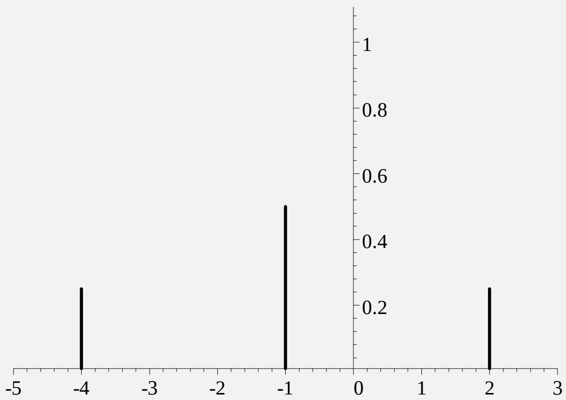
Comme $\big((X = x)\big)_{x \in X(\Omega)}$ est un système complet d'événements, la somme des nombres de la seconde ligne du tableau ou la somme des hauteurs des bâtonnets vaut 1. Autrement dit, on a

$$\sum_{x \in X(\Omega)} P(X = x) = 1.$$

Exemples :

- Exemple fil rouge : On reprend l'exemple du joueur qui lance deux fois une pièce de monnaie (de manière indépendante). À chaque fois qu'il obtient pile, il gagne 1 € et à chaque fois qu'il obtient face, il perd 2 €. On suppose en outre que la pièce est équilibrée. On note $\Omega = \{\text{Pile,Face}\}^2$ l'univers des possibles que l'on munit de la probabilité uniforme. Soit X la variable aléatoire égale au gain algébrique du joueur. Alors $X(\Omega) = \{-4; -1; 2\}$ et

x	-4	-1	2
$P(X = x)$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$



La proposition suivante donne une condition nécessaire et suffisante pour qu'une fonction soit une densité.

Proposition 4

Soient A un ensemble fini et $f : A \longrightarrow \mathbb{R}$ une application. Si f est à valeurs positives ou nulles et si $\sum_{x\in A} f(x) = 1$, alors f est une loi de probabilité, c'est-à-dire qu'il existe un espace probabilisé $(\Omega, \mathscr{T}, P)$ et une variable aléatoire X sur $(\Omega, \mathscr{T})$ telle que $X(\Omega) = A$ et $\forall x \in A,\ P(X = x) = f(x)$.

■ Posons $A = \{x_1, \ldots, x_n\}$. Comme $f(x_1) + \cdots + f(x_n) = 1$ par hypothèse, on sait, d'après le théorème de construction des probabilités finies, qu'il existe une probabilité P_f sur $\big(A; \mathscr{P}(A)\big)$ telle que $\forall k \in [\![1; n]\!],\ P_f(x_k) = f(x_k)$. Il suffit alors de prendre $\Omega = A$, $\mathscr{T} = \mathscr{P}(A)$, $P = P_f$ et $X = \mathrm{Id}_A$ pour avoir gain de cause. ■

Cette proposition permet de fabriquer autant de loi de probabilités qu'on le souhaite. Elle nous servira dans la section D. pour introduire les lois usuelles (loi uniforme, loi de Bernoulli, loi binomiale et loi hypergéométrique).

Exemples :

- Soient $n \geqslant 2$, $\alpha \in \mathbb{R}$ et $f : [\![0; n-1]\!] \longrightarrow \mathbb{R}$ telle que

$$\forall k \in [\![0; n-1]\!], \qquad f(k) = \alpha \sin \frac{k\pi}{n} \cdot$$

La fonction f est une loi de probabilité si, et seulement si, elle est à valeurs positives et $f(0) + f(1) + \cdots + f(n-1) = 1$. Or

$$\begin{aligned} \sum_{k=0}^{n-1} f(k) &= \alpha \Im\mathrm{m}\Big(\sum_{k=0}^{n-1} \mathrm{e}^{ik\pi/n}\Big) = \alpha \Im\mathrm{m}\Big(\frac{\mathrm{e}^{in\pi/n} - \mathrm{e}^{i0\pi/n}}{\mathrm{e}^{i\pi/n} - 1}\Big) = \alpha \Im\mathrm{m}\Big(\frac{-2\,\mathrm{e}^{-i\pi/(2n)}}{\mathrm{e}^{i\pi/(2n)} - \mathrm{e}^{-i\pi/(2n)}}\Big) \\ &= \alpha \Im\mathrm{m}\Big(\frac{-2\,\mathrm{e}^{-i\pi/(2n)}}{2i \sin\big(\pi/(2n)\big)}\Big) = \frac{\alpha}{\sin\big(\pi/(2n)\big)} \Im\mathrm{m}(i\,\mathrm{e}^{-i\pi/(2n)}) = \frac{\alpha \cos\big(\pi/(2n)\big)}{\sin\big(\pi/(2n)\big)}, \end{aligned}$$

donc f est une loi de probabilité si, et seulement si, $\alpha = \tan(\pi/(2n))$.

B.2. Fonction de répartition

Définition 2

Soit X une variable aléatoire réelle finie. On appelle fonction de répartition de X l'application F_X définie par

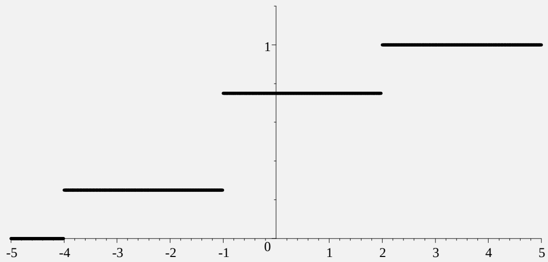
$$F_X \left\{ \begin{array}{ccc} \mathbb{R} & \longrightarrow & [0;1] \\ x & \longmapsto & P(X \leqslant x). \end{array} \right.$$

La fonction de répartition est particulièrement adaptée pour étudier des maximums (ou des minimums) de variables aléatoires réelles. En effet, l'événement $\big(\max(X,Y) = z\big)$ n'est pas simple à décrire (car $(X = z)$ et $(Y = z)$ ne sont pas incompatibles, y réfléchir ...). Par contre, l'événement $\big(\max(X,Y) \leqslant z\big)$ se réécrit tout simplement sous la forme $(X \leqslant z) \cap (Y \leqslant z)$.

Exemples :

- Exemple fil rouge : Dans l'exemple précédent du joueur effectuant deux lancers d'une pièce équilibrée avec un gain d'un euro pour chaque pile et une perte de deux euro pour chaque face, la fonction de répartition de la variable de gain X est

$$F_X \left\{ \begin{array}{ccc} \mathbb{R} & \longrightarrow & \mathbb{R} \\ x & \longmapsto & \left\{ \begin{array}{ll} 0 & \text{si } x < -4 \\ 1/4 & \text{si } -4 \leqslant x < -1 \\ 3/4 & \text{si } -1 \leqslant x < 2 \\ 1 & \text{si } 2 \leqslant x \end{array} \right. \end{array} \right.$$



Cet exemple illustre le fait que F_X est une fonction croissante qui, partant de 0, gloutonne une par une les valeurs de la densité, pour arriver jusqu'à la valeur 1.

La proposition suivante fait le lien entre la densité et la fonction de répartition d'une variable aléatoire. On y constate qu'il revient au même de se donner f_X ou F_X, autrement dit que la fonction de répartition caractérise la loi de probabilité comme le faisait la densité.

Proposition 5

Soit X une variable aléatoire réelle finie de densité f_X et de fonction de répartition F_X. Alors

$$\forall x \in \mathbb{R}, \quad F_X(x) = \sum_{\substack{y \in X(\Omega) \\ y \leqslant x}} f_X(y) \qquad \text{et} \qquad \forall x \in X(\Omega), \quad f_X(x) = F_X(x) - \lim_{t \to x^-} F_X(t).$$

Autrement dit,

- ▶ la fonction de répartition en x est égale à la somme des valeurs de la densité avant x ;
- ▶ la densité en x est égale à la hauteur de la « marche » de la fonction de répartition en x.

■ AQT ■

Exemples :

- Si X est à valeurs dans $\mathbb{Z}$, la relation $\forall k \in X(\Omega),\ f_X(k) = F_X(k) - F_X(k-1)$ donne

$$\forall k \in X(\Omega), \qquad P(X = k) = P(X \leqslant k) - P(X \leqslant k-1).$$

B.3. Image d'une variable aléatoire par une fonction

Définition 3

Soient X est une variable aléatoire finie sur $(\Omega, \mathscr{T})$ et g une fonction dont l'ensemble de définition contient $X(\Omega)$. La variable aléatoire $g \circ X$, obtenue en composant les applications X et g, est notée $g(X)$.

L'encadré suivant explique comment trouver la loi de $g(X)$.

Loi d'une composée

Soit X une variable aléatoire finie dont on connait l'univers image $X(\Omega) = \{x_1, \ldots, x_n\}$ et la loi de probabilité. Pour déterminer la loi de probabilité d'une variable composée $Y = g(X)$, on effectue les étapes suivantes :

- on détermine l'univers image de $Y = g(X)$ en calculant $y_1 = g(x_1)$, $\ldots$, $y_n = g(x_n)$ (certaines valeurs peuvent être égales) ;
- pour chaque $y \in Y(\Omega)$,
  - on note $x_{i_1}, \ldots, x_{i_k}$ tous les éléments de $X(\Omega)$ telles que $g(x_{i_1}) = \cdots = g(x_{i_k}) = y$ de sorte que $(Y = y) = (X = x_{i_1}) \sqcup \cdots \sqcup (X = x_{i_k})$;
  - on calcule
$$P(Y = y) = P(X = x_{i_1}) + \cdots + P(X = x_{i_k}).$$

On a donc

$$f_Y \left\{ \begin{array}{ccc} Y(\Omega) & \longrightarrow & [0;1] \\ y & \longmapsto & \displaystyle\sum_{\substack{x \in X(\Omega) \\ g(x) = y}} P(X = x) \end{array} \right.$$

Exemples :

- Exemple fil rouge : Reprenons l'exemple du joueur effectuant deux lancers d'une pièce équilibrée avec un gain d'un euro pour chaque pile et une perte de deux euros pour chaque face et cherchons la loi de la variable aléatoire $Y = X^2 + 2X$ où X désigne toujours le gain algébrique. Comme $X(\Omega) = \{-4; -1; 2\}$, on a $Y(\Omega) = \{-1; 8\}$ et
$$P(Y = -1) = P(X^2 + 2X = -1) = P(X = -1) = \frac{1}{2}$$
et
$$P(Y = 8) = P(X^2 + 2X = 8) = P(X = -4) + P(X = 2) = \frac{1}{4} + \frac{1}{4} = \frac{1}{2},$$
donc

i	-1	8
$P(Y = i)$	$\frac{1}{2}$	$\frac{1}{2}$

1 h 30

C. Paramétres de position et de dispersion

C.1. Espérance

Définition 4

Soit X une variable aléatoire réelle finie telle que $X(\Omega) = \{x_1, x_2, \ldots, x_n\}$. On appelle espérance de X le nombre réel $E(X)$ défini par

$$E(X) = \sum_{k=1}^{n} x_k P(X = x_k) = \sum_{x \in X(\Omega)} x P(X = x).$$

Une variable aléatoire d'espérance nulle est dite centrée.

L'espérance est la moyenne de chacune des valeurs prises par la variable aléatoire X pondérée par la probabilité que X prenne cette valeur. C'est un indicateur de position.

L'espérance ne dépend pas de la variable aléatoire mais seulement de sa loi. Ainsi, deux variables qui ont même loi ont même espérance.

Exemples :

- Une variable aléatoire certaine égale à m admet m comme espérance.
- Exemple fil rouge : Calculons les espérances du gain X et de la composée $Y = X^2 + 2X$, dont nous connaissons les lois de probabilité. On a

$$E(X) = -4 \times \frac{1}{4} + (-1) \times \frac{1}{2} + 2 \times \frac{1}{4} = -1 \qquad \text{et} \qquad E(Y) = -1 \times \frac{1}{2} + 8 \times \frac{1}{2} = \frac{7}{2}.$$

- Si $\mathbb{1}_S$ désigne l'indicatrice de l'événement S (c'est-à-dire la variable qui vaut 1 lorsque l'événement S se produit et 0 sinon), alors $E(\mathbb{1}_S) = P(S)$.
  En effet, on a $E(\mathbb{1}_S) = 0.P(\mathbb{1}_S = 0) + 1.P(\mathbb{1}_S = 1) = 0 + 1.P(S) = P(S)$.

Avant d'énoncer les propriétés de l'espérance, nous donnons une nouvelle expression d'icelle.

Lemme 1

Soit X une variable aléatoire réelle finie. L'espérance de X est donnée par la formule d'atomisation :

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

■ On a

$$\begin{aligned} E(X) &= \sum_{x \in X(\Omega)} x P(X = x) \\ &= \sum_{x \in X(\Omega)} x \sum_{\substack{\omega \in \Omega \\ X(\omega) = x}} P(\omega) \\ &= \sum_{x \in X(\Omega)} \sum_{\substack{\omega \in \Omega \\ X(\omega) = x}} X(\omega) P(\omega) \\ &= \sum_{\omega \in \Omega} X(\omega) P(\omega), \end{aligned}$$

car $\bigcup_{x \in X(\Omega)} (X = x) = \Omega$ puisque $\big((X = x)\big)_{x \in X(\Omega)}$ est un système complet d'événements. ■

La proposition suivante dresse la liste des propriétés de l'espérance. On notera l'analogie avec les propriétés de l'intégrale.

Proposition 6

Soient X, Y deux variables aléatoires réelles finies et $\lambda, \mu \in \mathbb{R}$. On a

(i) $E(\lambda X + \mu Y) = \lambda E(X) + \mu E(Y)$ (linéarité) ;

(ii) si $a \leqslant X \leqslant b$, alors $a \leqslant E(X) \leqslant b$,
en particulier, si X est positive, alors $E(X) \geqslant 0$ (positivité) ;

(iii) si $X \geqslant Y$, alors $E(X) \geqslant E(Y)$ (croissance) ;

(iv) $|E(X)| \leqslant E(|X|)$ (inégalité triangulaire).

■ (i) On a

$$E(\lambda X + \mu Y) = \sum_{\omega \in \Omega} (\lambda X + \mu Y)(\omega) P(\omega) = \lambda \sum_{\omega \in \Omega} X(\omega) P(\omega) + \mu \sum_{\omega \in \Omega} Y(\omega) P(\omega) = \lambda E(X) + \mu E(Y)$$

où la première égalité est la formule d'atomisation appliquée à $\lambda X + \mu Y$ et la dernière égalité découle d'une double utilisation de la formule d'atomisation pour X et Y.

(ii) Si $X(\Omega) \subset [a; b]$, on a $\forall x \in X(\Omega),\ aP(X = x) \leqslant xP(X = x) \leqslant bP(X = x)$. En sommant, on obtient

$$\sum_{x \in X(\Omega)} aP(X = x) \leqslant \sum_{x \in X(\Omega)} xP(X = x) \leqslant \sum_{x \in X(\Omega)} bP(X = x),$$

c'est-à-dire $a \leqslant E(X) \leqslant b$.

(iii) Supposons que $X \geqslant Y$ de sorte que $X - Y \geqslant 0$. La positivité nous dit alors que $E(X - Y) \geqslant 0$. La linéarité permet d'en déduire que $E(X) - E(Y) \geqslant 0$, ce qui donne le résultat.

(iv) En appliquant (iii) à l'encadrement $-|X| \leqslant X \leqslant |X|$, on obtient $E(-|X|) \leqslant E(X) \leqslant E(|X|)$, c'est-à-dire $-E(|X|) \leqslant E(X) \leqslant E(|X|)$ par linéarité. Cela signifie bien que $|E(X)| \leqslant E(|X|)$. ■

Exemples :

- Si $m = E(X)$, alors $X - m$ est centrée puisque $E(X - m) = E(X) - m = 0$.
- Si X et $-X$ ont même loi, alors X est centrée. En effet, puisqu'elles ont même loi, X et $-X$ ont même espérance : $E(X) = E(-X)$, d'où $E(X) = -E(X)$ et donc $E(X) = 0$.

On termine ce paragraphe avec un résultat naturel concernant les variables positives centrées. Là aussi, on peut noter l'analogie avec le résultat sur les intégrales qui dit qu'une fonction continue, positive et d'intégrale nulle est identiquement nulle.

Proposition 7

Soit X une variable aléatoire réelle finie. Si X est positive et centrée, alors X est presque sûrement nulle, c'est-à-dire que $P(X = 0) = 1$.

■ On a $0 = E(X) = \sum_{x \in X(\Omega)} xP(X = x)$. Comme tous les termes sont positifs ou nuls, ils sont tous nuls, c'est-à-dire $\forall x \in X(\Omega),\ xP(X = x) = 0$ ou encore $\forall x \in X(\Omega) \setminus \{0\},\ P(X = x) = 0$. Il s'ensuit que toute la masse de la loi de X est portée en 0, c'est-à-dire $P(X = 0) = 1$. ■

Exemples :

- La proposition 7 nous dit qu'il y a égalité dans l'inégalité triangulaire si, et seulement si, X est positive ou X est négative.

C.2. Espérance d'une composée - Moments

Pour déterminer l'espérance d'une variable aléatoire composée $g(X)$, on peut rechercher la loi de $g(X)$ et en déduire classiquement l'espérance. On peut aussi utiliser directement le théorème de transfert énoncé ci-dessous, ce qui ne nécessite pas la recherche préalable de la loi de $g(X)$.

Proposition 8

Soit X une variable aléatoire finie telle que $X(\Omega) = \{x_1, x_2, \ldots, x_n\}$. Pour toute fonction $g : X(\Omega) \longrightarrow \mathbb{R}$, l'espérance de la variable aléatoire composée $g(X)$ est donnée par la formule de transfert :

$$E\big(g(X)\big) = \sum_{k=1}^{n} g(x_k)P(X = x_k) = \sum_{x \in X(\Omega)} g(x)P(X = x).$$

■ On a

$$\begin{aligned}
E\big(g(X)\big) &= \sum_{\omega\in\Omega} g(X)(\omega)P(\omega) && \text{par la formule d'atomisation} \\
&= \sum_{x\in X(\Omega)} \sum_{\substack{\omega\in\Omega \\ X(\omega)=x}} g\big(X(\omega)\big)P(\omega) \\
&= \sum_{x\in X(\Omega)} g(x) \sum_{\substack{\omega\in\Omega \\ X(\omega)=x}} P(\omega) \\
&= \sum_{x\in X(\Omega)} g(x)P(X = x)
\end{aligned}$$

où la seconde égalité découle du fait que Ω est l'union des événements $\big((X = x)\big)_{x\in X(\Omega)}$. ■

On retiendra que pour calculer l'espérance de $g(X)$ on additionne les valeurs $g(x)$ prises par cette variable en pondérant chacune de ces valeurs par la probabilité que X prenne la valeur x.

Exemples :

- Exemple fil rouge : Reprenons l'exemple du joueur qui joue deux fois à pile ou face et retrouvons l'espérance de $Y = X^2 + 2X$ sans passer par la détermination de la loi de Y. On a

$$E(Y) = \sum_{x\in\{-4,-1,2\}} (x^2 + 2x)P(X = x) = 8 \times \frac{1}{4} + (-1) \times \frac{1}{2} + 8 \times \frac{1}{4} = \frac{7}{2},$$

  ce qui correspond bien au résultat que nous avions déjà trouvé.

La transformation la plus simple que l'on puisse faire subir à une variable aléatoire est l'élévation à une puissance. Cela conduit à la notion de moment d'une variable aléatoire.

Définition 5

Soient X une variable aléatoire réelle finie et $k \in \mathbb{N}$. Le réel $E(X^k)$ est appelé le moment d'ordre k de X. Le réel $E\Big[\big(X - E(X)\big)^k\Big]$ est appelé le moment centré d'ordre k de X.

Exemples :

- Le moment d'ordre 1 est donc l'espérance.
- Le paragraphe suivant est consacré au moment centré d'ordre 2, que l'on appelle variance.
- Exemple fil rouge : Pour notre lanceur de pièce, on a

$$E(X^2) = (-4)^2 \times \frac{1}{4} + (-1)^2 \times \frac{1}{2} + 2^2 \times \frac{1}{4} = \frac{11}{2}.$$

C.3. Variance et écart-type

Définition 6

Soit X une variable aléatoire réelle finie. On appelle variance de X le nombre réel $V(X)$ défini comme le moment centré d'ordre 2 de X, c'est-à-dire

$$V(X) = E\big[\big(X - E(X)\big)^2\big].$$

Elle est également définie comme la différence du moment d'ordre 2 et du carré du moment d'ordre 1, c'est-à-dire

$$V(X) = E(X^2) - E(X)^2 \qquad \text{(Formule de Kœnig–Huygens)}.$$

■ Si $m = E(X)$, on a $V(X) = E\big((X-m)^2\big) = E(X^2-2mX+m^2) = E(X^2)-2mE(X)+m^2 = E(X^2)-m^2$, ce qui justifie la formule de Kœnig–Huygens. ■

La variance est la moyenne quadratique des valeurs de la variable centrée $X - E(X)$ pondérées par leur probabilité. La variance mesure donc la dispersion quadratique de X par rapport à $E(X)$.

Mais pourquoi utilise-t-on la dispersion *quadratique* plutôt que la moyenne de la distance à l'espérance $E[|X - E(X)|]$? Tout simplement, parce que des complications techniques (liées aux valeurs absolues) rendent la manipulation de $E[|X-E(X)|]$ plus compliquée que celle de la variance.

On notera que la variance, tout comme l'espérance, ne dépend pas directement de la variable aléatoire mais seulement de sa loi. Ainsi, deux variables qui ont même loi ont même variance.

Dans la pratique, on utilise la formule de Kœnig–Huygens pour calculer la variance.

Exemples :

- Une variable certaine est de variance nulle.
- Exemple fil rouge : Pour notre joueur effectuant deux lancers d'une pièce, on a

$$V(X) = E(X^2) - E(X)^2 = \frac{11}{2} - (-1)^2 = \frac{9}{2}\cdot$$

La proposition suivante dresse la liste des propriétés de la variance.

Proposition 9

Soient X une variable aléatoire réelle finie et $\lambda, \mu \in \mathbb{R}$. Alors

(i) $V(X) \geqslant 0$;

(ii) $V(\lambda X + \mu) = \lambda^2 V(X)$ (la variance est dite quadratique).

■ (i) Comme $\big(X - E(X)\big)^2 \geqslant 0$, on a $V(X) = E\big[\big(X - E(X)\big)^2\big] \geqslant 0$ par la positivité de l'espérance.

(ii) $V(\lambda X+\mu) = E\big[\big(\lambda X+\mu-E(\lambda X+\mu)\big)^2\big] = E\big[\big(\lambda X+\mu-\lambda E(X)-\mu\big)^2\big] = E\big[\lambda^2\big(X-E(X)\big)^2\big] = \lambda^2 V(X)$. ■

La variance mesurant l'écart quadratique moyen à l'espérance, il semble naturel que la nullité de la variance traduise la (quasi-)constance de la variable.

Proposition 10

Si X est une variable aléatoire réelle finie d'espérance m et de variance nulle, alors X est presque sûrement égale à son espérance, c'est-à-dire que $P(X = m) = 1$.

■ Si $V(X) = 0$, alors $E[(X - m)^2] = 0$, ce qui signifie que la variable aléatoire positive $(X - m)^2$ est d'espérance nulle. En utilisant la proposition 7, on en déduit que $(X - m)^2$ et par suite $X - m$ sont presque surement nulles. Donc X est quasi-certainement égale à m. ■

À partir de la variance, on peut définir l'écart-type.

Définition 7

Soit X une variable aléatoire réelle finie. On appelle écart-type de X le nombre réel $\sigma(X)$ défini par

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

Une variable aléatoire d'écart-type égal à 1 (ou de variance égale à 1) est dite réduite.

L'écart-type, tout comme la variance, est une mesure de la dispersion des valeurs prises par une variable aléatoire par rapport à sa moyenne. Il possède en plus l'avantage de s'exprimer dans la même unité que la variable aléatoire. Ainsi, si X mesure des distances exprimées en mètres, la variance s'exprime en mètres carrés alors que l'écart-type est comparable aux valeurs de X, puisqu'il est homogène à des mètres.

Exemples :

- Une variable certaine est d'écart-type nul.
- Nicolas et Romain ont eu deux notes de maths ce trimestre. Leurs notes sont très différentes : Nicolas a eu deux fois 10 alors que Romain a eu 0 et 20. Leur moyenne est donc la même (c'est-à-dire 10) mais l'écart-type des notes de Nicolas est nul (ce qui exprime sa régularité) alors que l'écart-type de Romain est de 10 (ce qui exprime la forte variabilité de ses résultats).
- Exemple fil rouge : Pour notre joueur effectuant deux lancers d'une pièce, on a

$$\sigma(X) = \sqrt{V(X)} = \frac{3}{\sqrt{2}}.$$

Les propriétés de la variance implique celles de l'écart-type.

Proposition 11

Soient X une variable aléatoire réelle finie et $\lambda, \mu \in \mathbb{R}$. Alors

(i) $\sigma(X) \geqslant 0$;

(ii) $\sigma(\lambda X + \mu) = |\lambda|\sigma(X)$.

■ AQT ■

Exemples :

- Si X est d'espérance m et d'écart-type σ non nul, alors $(X - m)/\sigma$ est centrée et réduite (on dit centrée-réduite, histoire d'économiser deux lettres...) car

$$E\Big(\frac{X-m}{\sigma}\Big) = \frac{E(X)-m}{\sigma} = 0 \qquad \text{et} \qquad V\Big(\frac{X-m}{\sigma}\Big) = \frac{V(X)}{\sigma^2} = 1.$$

C.4. Inégalités de dispersion

En lemme apéritif, voici l'inégalité de dispersion de Markov.

Lemme 2

Soit X une variable aléatoire finie positive. Pour tout $a > 0$, on a l'inégalité de Markov :

$$P(X \geqslant a) \leqslant \frac{E(X)}{a}.$$

■ On a

$$\frac{E(X)}{a} = \sum_{x \in X(\Omega)} \frac{x}{a} P(X = x) \geqslant \sum_{\substack{x \in X(\Omega) \\ x \geqslant a}} \frac{x}{a} P(X = x) \geqslant \sum_{\substack{x \in X(\Omega) \\ x \geqslant a}} P(X = x) = P(X \geqslant a),$$

où, pour la première minoration, on a volontairement « oublié » des termes positifs ou nuls. ■

Exemples :

- Le nombre de jours de pluie au mois d'avril est une variable aléatoire finie d'espérance 6. Alors, d'après l'inégalité de Markov, $P(X \geqslant 10) \leqslant 6/10 = 3/5$, ce qui signifie que la probabilité qu'il y ait plus de 10 jours de pluie en avril est inférieur à $3/5$.

On peut maintenant donner l'inégalité de dispersion de Bienaymé–Tchebychev.

Proposition 12

Soit X une variable aléatoire finie d'espérance m et d'écart-type σ. Alors, pour tout $d > 0$, on a l'inégalité de Bienaymé–Tchebychev :

$$P(|X - m| \geqslant d) \leqslant \frac{\sigma^2}{d^2}.$$

■ En appliquant Markov à la v.a. positive $(X - m)^2$ avec $a = d^2$, on obtient immédiatement le résultat puisque les événements $\big((X - m)^2 \geqslant d^2\big)$ et $(|X - m| \geqslant d)$ sont égaux et $E\big((X - m)^2\big) = V(X) = \sigma^2$. ■

Bienaymé–Tchebychev : une inégalité de dispersion

L'inégalité de Bienaymé–Tchebychev permet de majorer la probabilité de l'événement de dispersion $(|X - m| \geqslant d)$. Elle permet ainsi d'estimer comment une variable aléatoire s'écarte de sa moyenne.

La qualité de cette inégalité est d'autant meilleure que le quotient σ/d est petit (c'est-à-dire lorsque la dispersion d est grande devant l'écart-type σ).

Si l'inégalité de Bienaymé–Tchebychev a l'avantage d'être universelle (elle s'applique à toute variable, indépendamment de sa loi), elle a l'inconvénient de fournir des majorations médiocres. Nous verrons cependant qu'elle est utile pour l'étude des phénomènes de convergence en loi.

Exemples :

- 400 élèves de MPSI résolvent un exercice. Chacun a 80 % de chance de réussir. On note X le nombre de succès. Nous verrons que, dans ces conditions, l'on a $E(X) = 320$ et $\sigma(X) = 8$. On veut minorer $P(300 < X < 340)$.
  L'inégalité de Bienaymé–Thebychev dit que $P(|X - 320| \geqslant 20) \leqslant 8^2/20^2$, ce qui donne, par passage à l'événement contraire, $P(300 < X < 340) \geqslant 1 - 8^2/20^2 = 0,84$. 3 h 30

D. Lois usuelles

D.1. Loi certaine

Définition 8

On dit qu'une variable aléatoire X est certaine, égale à m, lorsqu'elle est constante égale à m, autrement dit lorsqu'il existe $m \in \mathbb{R}$ telle que $X(\Omega) = \{m\}$. Dans ce cas, on a nécessairement $P(X = m) = 1$ et l'on dit que X suit la loi certaine égale à m.

Par extension, si X désigne une variable aléatoire finie telle que $X(\Omega) = \{x_1, \ldots, x_n\}$ avec $P(X = x_1) = 1$, on dit que X est quasi-certaine égale à x_1 ou encore qu'elle suit la loi quasi-certaine égale à x_1.

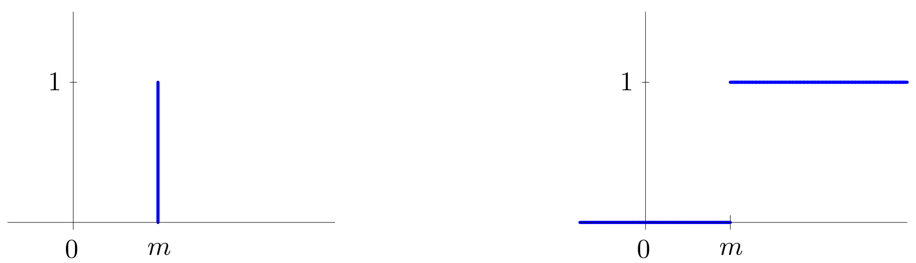
Schéma certain

La loi certaine régit les expériences déterministes, c'est-à-dire celles dont l'issue ne dépend pas du hasard. Elle intervient rarement en tant que telle. On l'utilise par contre souvent comme composante de lois plus complexes.

Exemples :

- Vous participez à un jeu dont la mise d'entrée est de m euro. Votre gain, au cours du jeu, et de G euro. Pour obtenir votre « vrai » gain, vous devez soustraire à G la variable certaine égale à m.

Densité et fonction de répartition



densité d'une loi certaine

fonction de répartition d'une loi certaine

L'espérance et la variance d'une loi certaine sont intuitives.

Proposition 13

Si X est une variable certaine égale à m, on a

$$E(X) = m \qquad \text{et} \qquad V(X) = 0.$$

■ On a

$$E(X) = mP(X = m) = m \times 1 = m$$

et

$$V(X) = E(X^2) - E(X)^2 = E(m^2) - E(m)^2 = m^2 - m^2 = 0.$$ ■

D.2. Loi uniforme

C'est la plus égalitaire de toutes les lois : celle de la pièce non truquée ; du dé non pipé ; de la roulette (russe ou de casino) et plus généralement de toutes les expériences qui négligent la nature de l'objet choisi. Avec elle, pas de racisme, la probabilité d'un événement est donnée par « le nombre de cas favorables sur le nombre de cas possibles ».

Définition 9

Soient $n \in \mathbb{N}^*$ et X une variable aléatoire finie telle que $X(\Omega) = \{x_1, \dots, x_n\}$. On dit que X suit la loi uniforme sur $X(\Omega)$ lorsque les événements du système complet $(X = x_1)$, $(X = x_2)$, $\dots$, $(X = x_n)$ sont équiprobables, c'est-à-dire

$$\forall k \in [\![1; n]\!], \qquad P(X = x_k) = 1/n.$$

En abrégé, on note $X \hookrightarrow \mathscr{U}\big(X(\Omega)\big)$ pour dire que X suit la loi uniforme sur $X(\Omega)$.

La loi uniforme sur un singleton (c'est-à-dire le cas $n = 1$) est une loi certaine.

Considérons une urne opaque contenant n boules indiscernables au toucher. On effectue un unique tirage d'une boule dans l'urne. On note X le numéro de la boule tirée. On a alors $X(\Omega) = [\![1; n]\!]$ et $\forall k \in [\![1; n]\!]$, $P(X = k) = 1/n$ puisque toutes les boules ont la même chance d'être tirée. Autrement dit, X suit la loi uniforme sur $[\![1; n]\!]$.

Cet exemple est représentatif du schéma uniforme décrit ci-dessous.

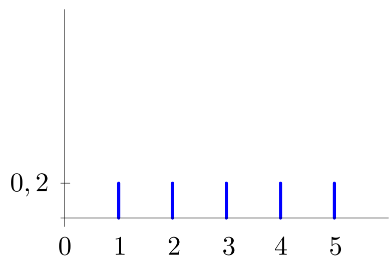
Schéma uniforme

La loi uniforme régit les expériences aléatoires dont toutes les issues sont équiprobables. Elle ne prend pas en compte la nature des résultats de la variable mais seulement le nombre de ces résultats.

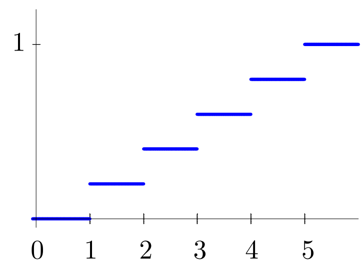
Exemples :

- On joue à pile ou face avec une pièce équilibrée. On pose $X = 1$ si pile sort et $X = 0$ si face sort. Alors $P(X = 0) = P(X = 1) = 1/2$, ce qui signifie que X suit la loi $\mathscr{U}(\{0; 1\})$.
- On lance un dé non pipé et on note X le nombre de points obtenus. Alors $X \hookrightarrow \mathscr{U}([\![1; 6]\!])$.
- On joue à la roulette de casino (dont les numéros vont de 0 à 36) et on note X le numéro qui sort. Alors $X \hookrightarrow \mathscr{U}([\![0; 36]\!])$.

Densité et fonction de répartition



densité de la loi uniforme $\mathscr{U}([\![1; 5]\!])$



fonction de répartition de la loi uniforme $\mathscr{U}([\![1; 5]\!])$

L'espérance et la variance de la loi uniforme sont données par la proposition suivante.

Proposition 14

Si X est une variable aléatoire réelle de loi uniforme sur $X(\Omega) = \{x_1, \dots, x_n\}$, alors l'espérance de X est donnée par la moyenne arithmétique usuelle :

$$E(X) = \frac{x_1 + \cdots + x_n}{n}\cdot$$

et sa variance vaut

$$V(X) = \frac{{x_1}^2 + \cdots + {x_n}^2}{n} - \left(\frac{x_1 + \cdots + x_n}{n}\right)^2\cdot$$

■ La formule de l'espérance est évidente. Celle de la variance découle du fait que X^2 est une variable aléatoire uniforme sur $\{{x_1}^2, \dots, {x_n}^2\}$ et de l'application de la formule de Kœnig–Huygens. ■

La formule de la variance ne doit pas être apprise par cœur. Il suffit de retenir que « la variance est la moyenne des carrés moins le carré de la moyenne ».

Exemples :

- Si X suit la loi uniforme sur $\{0; 2; 4; 10\}$, on a

  $$E(X) = 0 \times \frac{1}{4} + 2 \times \frac{1}{4} + 4 \times \frac{1}{4} + 10 \times \frac{1}{4} = 4$$

  et

  $$E(X^2) = 0^2 \times \frac{1}{4} + 2^2 \times \frac{1}{4} + 4^2 \times \frac{1}{4} + 10^2 \times \frac{1}{4} = 30,$$

  d'où

  $$V(X) = E(X^2) - E(X)^2 = 30 - 4^2 = 14.$$

- On rencontre souvent le cas où X suit la loi uniforme sur $[\![1; n]\!]$. Dans ce cas, on a

  $$E(X) = \frac{n+1}{2} \qquad \text{et} \qquad V(X) = \frac{n^2-1}{12}\cdot$$

  En effet, on a

  $$E(X) = \frac{1}{n}\sum_{k=1}^{n} k = \frac{1}{n}\frac{n(n+1)}{2} = \frac{n+1}{2},$$

  et

  $$E(X^2) = \frac{1}{n}\sum_{k=1}^{n} k^2 = \frac{1}{n}\frac{n(n+1)(2n+1)}{6} = \frac{(n+1)(2n+1)}{6}$$

  donc

  $$V(X) = E(X^2) - E(X)^2 = \frac{(n+1)(2n+1)}{6} - \frac{(n+1)^2}{4} = \text{calculs} = \frac{n^2-1}{12}\cdot$$

- Dans le cas de la roulette de casino, la variable X, égale au numéro sorti, suit la loi uniforme sur $[\![0; 36]\!]$. Par conséquent, la variable $X + 1$ suit la loi uniforme sur $[\![1; 37]\!]$, ce qui donne

  $$E(X+1) = \frac{37+1}{2} = 19 \qquad \text{et} \qquad V(X+1) = \frac{37^2-1}{12} = 114.$$

  Comme $E(X+1) = E(X) + 1$ et $V(X+1) = V(X)$, il s'ensuit que

  $$E(X) = 18 \qquad \text{et} \qquad V(X) = 114.$$

D.3. Loi de Bernoulli

C'est la plus simple de toutes les lois de probabilité (après la loi certaine), souvent matérialisée par le jet d'une pièce de monnaie. Avec elle, on gagne ou on perd, un point c'est tout, avec un seul paramètre : la probabilité du succès.

Définition 10

Soit $p \in [0;1]$. On dit que X suit une loi de Bernoulli de paramètre p et l'on note $X \hookrightarrow \mathscr{B}(p)$ lorsque

$$\begin{cases} X(\Omega) = \{0;1\} \\ P(X=1) = p \\ P(X=0) = q \end{cases} \qquad \text{où} \qquad q = 1-p.$$

Les lois $\mathscr{B}(0)$ et $\mathscr{B}(1)$ sont des lois certaines. La loi $\mathscr{B}(1/2)$ est aussi la loi $\mathscr{U}(\{0;1\})$.

Considérons une urne opaque contenant une proportion p de boules octarines (la couleur de la magie) et une proportion $q = 1-p$ de boules blanches. On effectue un unique tirage d'une boule dans l'urne. On note $X = 1$ si la boule est octarine et $X = 0$ sinon. On a alors

$$X(\Omega) = \{0;1\}, \qquad P(X=1) = p \qquad \text{et} \qquad P(X=0) = q,$$

ce qui signifie que X suit la loi $\mathscr{B}(p)$.

Cet exemple est représentatif du schéma de Bernoulli décrit ci-dessous.

Schéma de Bernoulli

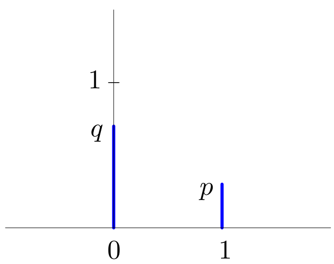
La loi $\mathscr{B}(p)$ régit les expériences aléatoires qui n'ont que deux issues possibles, dite « succès » et « échec » ; le succès se produisant avec probabilité p.

En pratique, il n'est pas toujours facile de connaître la probabilité p de succès d'une expérience donnée (peut-on être sûr qu'une pièce est parfaitement équilibrée ?). Ainsi, pour estimer p, on réalise un grand nombre de fois l'expérience en se disant que la proportion de succès devrait être voisine de p. Nous sommes alors aux portes de la loi des grands nombres.

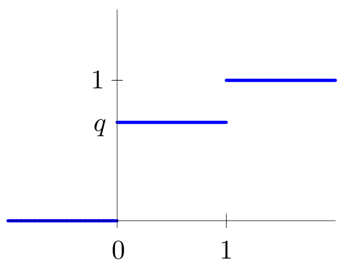
Exemples :

- On lance une pièce qui donne pile avec probabilité p et face avec probabilité $q = 1-p$. On pose $X = 1$ si le résultat est pile et $X = 0$ si c'est face. Alors $X \hookrightarrow \mathscr{B}(p)$.
- La variable indicatrice $\mathbb{1}_S$ d'un événement S (celle qui vaut 1 si S se produit et 0 sinon) suit une loi de Bernoulli dont le paramètre est la probabilité de succès $P(S)$.

Densité et fonction de répartition



densité d'une loi de Bernoulli



fonction de répartition d'une loi de Bernoulli

L'espérance et la variance d'une loi de Bernoulli, données ci-dessous, sont à connaître par ♡.

Proposition 15

Si X est une variable aléatoire de loi de Bernoulli de paramètre p, on a

$$E(X) = p \qquad \text{et} \qquad V(X) = pq.$$

■ On a

$$E(X) = 0 \times P(X = 0) + 1 \times P(X = 1) = 0 + 1 \times p = p.$$

Comme $X^2 = X$, on a

$$V(X) = E(X^2) - E(X)^2 = E(X) - E(X)^2 = p - p^2 = pq.$$ ■

Exemples :

- Si $\mathbb{1}_S$ est l'indicatrice de l'événement S, on a

$$E(\mathbb{1}_S) = P(S) \qquad \text{et} \qquad V(\mathbb{1}_S) = P(S)P(\overline{S}).$$

D.4. Loi binomiale

Le jeu du pile ou face est impitoyable : le couperet tombe d'un coup. Si l'on s'autorise plusieurs lancers, on gagne en indulgence, mais tout à un prix : la description probabiliste se complique.

Définition 11

Soient $n \in \mathbb{N}^*$ et $p \in [0;1]$. On dit que X suit une loi binomiale de paramètres n et p et l'on note $X \hookrightarrow \mathscr{B}(n;p)$ lorsque

$$\begin{cases} X(\Omega) = [\![0;n]\!] \\ \forall k \in [\![0;n]\!], \quad P(X=k) = \binom{n}{k} p^k q^{n-k} \end{cases} \qquad \text{où} \qquad q = 1-p.$$

■ On a
$$\sum_{k=0}^{n} P(X=k) = \sum_{k=0}^{n} \binom{n}{k} p^k q^{n-k} = (p+q)^n = 1^n = 1,$$
d'après la formule du binôme. Cela démontre que la loi binomiale est bien une loi. ■

Une variable aléatoire qui suit une loi binomiale de paramètres 1 et p est une loi de Bernoulli de paramètre p, autrement dit $\mathscr{B}(1;p) = \mathscr{B}(p)$.

Considérons à nouveau notre urne opaque contenant une proportion p de boules octarines et une proportion q de boules blanches. On effectue n tirages avec remise dans cette urne et l'on note X le nombre de boules octarines tirées.

On a clairement $X(\Omega) = [\![0;n]\!]$.

Pour tout $k \in [\![0;n]\!]$, l'événement $(X=k)$ est la réunion incompatible des événements $S_{i_1,i_2,\ldots,i_k}$: « on a tiré une boule octarine aux tirages i_1, i_2, ..., i_k et des boules blanches aux $n-k$ autres tirages » (où $1 \leqslant i_1 < i_2 < \cdots < i_k \leqslant n$).

Comme les résultats sont indépendants, on a $P(S_{i_1,i_2,\ldots,i_k}) = p^k q^{n-k}$.

Par ailleurs, pour dénombrer les événements de type $S_{i_1,i_2,\ldots,i_k}$, on choisit k positions parmi n pour les tirages de boules octarines, ce qui laisse $\binom{n}{k}$ choix (les boules blanches se plaçant sur les $n-k$ positions restantes). Il y a donc $\binom{n}{k}$ événements de type $S_{i_1,i_2,\ldots,i_k}$.

On en déduit que
$$P(X=k) = \binom{n}{k} p^k q^{n-k}.$$

On contate ainsi que X suit la loi $\mathscr{B}(n;p)$.

Cet exemple est représentatif du schéma binomial décrit ci-dessous.

Schéma binomial

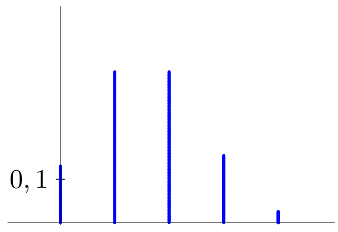
La loi $\mathscr{B}(n;p)$ régit le nombre de succès dans une suite de n expériences de Bernoulli identiques et indépendantes, chaque succès se produisant avec la probabilité p.

Dans le langage des boules et des urnes, la loi binomiale correspond à une suite de tirages avec remise.

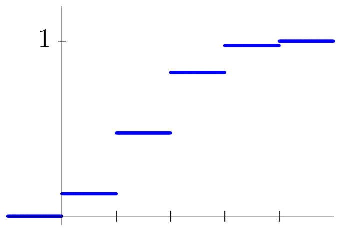
Exemples :

- On joue n fois à pile ou face avec une pièce donnant pile avec probabilité p. On note X le nombre de piles apparus. Alors X compte le nombre de succès (obtenir pile) dans une suite de n lancers indépendants avec, à chaque fois, une probabilité p de succès. Donc $X \hookrightarrow \mathscr{B}(n;p)$.
- Si $X \hookrightarrow \mathscr{B}(n;p)$, alors $n - X \hookrightarrow \mathscr{B}(n;q)$, puisque $n-X$ représente le nombre d'échecs, là où X représentait le nombre de succès.

Densité et fonction de répartition



densité de $\mathscr{B}(4; 0, 4)$



fonction de répartition de $\mathscr{B}(4; 0, 4)$

La densité de la loi binomiale présente une forme caractéristique en cloche. Plus p est proche de 0, plus le sommet de la cloche est décalé vers la gauche (les succès sont peu nombreux car difficiles à obtenir). Au contraire, plus p est proche de 1, plus le sommet de la cloche se décale vers la droite (les succès sont plus nombreux car plus faciles à obtenir).

L'espérance et la variance d'une loi binomiale, données ci-dessous, sont à connaître par ♡.

Proposition 16

Si X est une variable aléatoire de loi binomiale de paramètres n et p, on a

$$E(X) = np \qquad \text{et} \qquad V(X) = npq.$$

■ Pour tout $\ell \in [\![1; n+1]\!]$, on a

$$\begin{aligned}
E\big(X(X-1)\cdots(X-\ell+1)\big) &= \sum_{k=\cancel{0}\,\ell}^{n} k(k-1)\cdots(k-\ell+1)\binom{n}{k}p^k q^{n-k} && \text{formule de transfert}\\
&= \sum_{k=\ell}^{n} n(n-1)\cdots(n-\ell+1)\binom{n-\ell}{k-\ell}p^k q^{n-k} && \text{formule du pion (ℓ fois)}\\
&= n(n-1)\cdots(n-\ell+1)p^\ell \sum_{j=0}^{n-\ell}\binom{n-\ell}{j}p^j q^{n-\ell-j} && \text{en posant } j = k-\ell\\
&= n(n-1)\cdots(n-\ell+1)p^\ell (p+q)^{n-\ell} && \text{formule du binôme}\\
&= n(n-1)\cdots(n-\ell+1)p^\ell.
\end{aligned}$$

Pour $\ell = 1$, on a immédiatement

$$E(X) = np.$$

Pour $\ell = 2$, on a

$$E\big(X(X-1)\big) = n(n-1)p^2,$$

donc, comme

$$V(X) = E(X^2) - E(X)^2 = E\big(X(X-1)\big) + E(X) - E(X)^2,$$

cela donne

$$V(X) = n(n-1)p^2 + np - n^2p^2 = npq.$$

■

Nous verrons plus tard que l'on peut simplifier cette preuve en utilisant le fait qu'une variable de loi $\mathscr{B}(n; p)$ a la même loi que la somme de n variables indépendantes de même loi $\mathscr{B}(p)$.

Exemples :

- Reprenons l'exemple de ces 400 élèves qui résolvaient un exercice, avec 80 % de chance de succès pour chaque élève. Si X désigne le nombre de succès (réussir l'exercice) dans cette série de 400 résolutions indépendantes avec, à chaque fois, une probabilité de succès égale à $0, 8$, alors $X \hookrightarrow \mathscr{B}(400; 0, 8)$. Cela fournit les résultats que nous avions admis :

$$E(X) = 400 \times 0,8 = 320 \qquad \text{et} \qquad \sigma(X) = \sqrt{V(X)} = \sqrt{400 \times 0,8 \times 0,2} = 8.$$

D.5. Loi hypergéométrique [✠]

a) Relation de Vandermonde

Lemme 3

Pour tous $a, b, n \in \mathbb{N}$, on a la formule de Vandermonde :

$$\sum_{k=0}^{n} \binom{a}{k} \binom{b}{n-k} = \binom{a+b}{n}.$$

■ Soient $a, b, n \in \mathbb{N}$. On considère un paquet de bonbons consitué de a arlequins et de b caramels. On prélève dans ce paquet une poignée constituée de n bonbons. Pour dénombrer le nombre de façons différentes d'effectuer ce prélèvement, on peut procéder de deux manières :

- on peut choisir directement n bonbons parmi les $a+b$ du paquet : $\binom{a+b}{n}$ possibilités ;
- on peut choisir :
  - ▷ ou bien 0 arlequin parmi a et n caramels parmi b : $\binom{a}{0}\binom{b}{n}$ possibilités ;
  - ▷ ou bien 1 arlequin parmi a et $n-1$ caramels parmi b : $\binom{a}{1}\binom{b}{n-1}$ possibilités ;
  - ▷ ou bien 2 arlequins parmi a et $n-2$ caramels parmi b : $\binom{a}{2}\binom{b}{n-2}$ possibilités ;
  - ▷ etc
  - ▷ ou bien $n-1$ arlequins parmi a et 1 caramel parmi b : $\binom{a}{n-1}\binom{b}{1}$ possibilités ;
  - ▷ ou bien n arlequins parmi a et 0 caramel parmi b : $\binom{a}{n}\binom{b}{0}$ possibilités ;

  ce qui donne $\sum_{k=0}^{n} \binom{a}{k}\binom{b}{n-k}$ possibilités de choisir n bonbons parmi les $a+b$ du paquet.

En rapprochant ces deux méthodes, on obtient la formule de Vandermonde. ■

Lorsque $a \leqslant n$ ou $b \leqslant n$, la formule reste valide mais certains coefficients binomiaux (et donc aussi certains termes de la somme) sont nuls.

Exemples :

- Pour $a = b = n$, on a
$$\sum_{k=0}^{n} \binom{n}{k}\binom{n}{n-k} = \binom{2n}{n},$$
c'est-à-dire, en vertu de la formule de symétrie,
$$\sum_{k=0}^{n} \binom{n}{k}^2 = \binom{2n}{n}.$$

b) Description de la loi hypergéométrique

Vous voulez connaître la probabilité de gagner au loto ? Alors, la loi hypergéométrique est faite pour vous. À défaut de gagner, vous saurez au moins pourquoi vous avez perdu !

Définition 12

Soient $n, N \in \mathbb{N}^*$ et $p \in [0;1]$ tels que $n \leqslant N$ et $Np \in \mathbb{N}$. On dit que X suit une loi hypergéométrique de paramètres N, n et p et on note $X \hookrightarrow \mathscr{H}(N, n; p)$ lorsque

$$\begin{cases} X(\Omega) = [\![0; n]\!] \\ \forall k \in [\![0; n]\!], \quad P(X = k) = \dfrac{\dbinom{Np}{k}\dbinom{Nq}{n-k}}{\dbinom{N}{n}} \qquad \text{où} \qquad q = 1 - p. \end{cases}$$

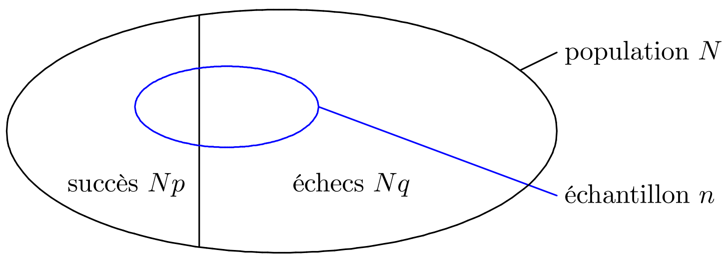
■ On a

$$\sum_{k=0}^{n} P(X = k) = \sum_{k=0}^{n} \frac{\binom{Np}{k}\binom{Nq}{n-k}}{\binom{N}{n}} = \frac{1}{\binom{N}{n}} \sum_{k=0}^{n} \binom{Np}{k}\binom{Nq}{n-k} = \frac{1}{\binom{N}{n}}\binom{N}{n} = 1,$$

où l'avant dernière égalité découle de la relation de Vandermonde. ■

Certaines des probabilités $P(X = k)$, où $k \in [\![0; n]\!]$, peuvent être nulles (il suffit que $Np < n$ et $k > Np$ ou que $Nq > n$ et $n - k > Nq$).

Considérons encore une fois une urne opaque contenant une population totale de N boules, dont une proportion p de boules octarines et une proportion q de boules blanches. On a donc Np boules octarines (les succès) et Nq boules blanches (les échecs). Dans cette population, on prélève un échantillon de n individus, c'est-à-dire qu'on effectue une suite de n tirages <u>sans remise</u>. On note X le nombre de boules octarines tirées.



On a $X(\Omega) = [\![0; n]\!]$ (certaines valeurs de $[\![0; n]\!]$ ne sont éventuellement pas atteintes par X).

Prenons $k \in [\![0; n]\!]$ et déterminons la probabilité de l'événement $(X = k)$.

L'univers Ω associé à cette expérience est l'ensemble des combinaisons de n boules prises parmi N. On a donc $\operatorname{card} \Omega = \binom{N}{n}$. On le munit de la loi uniforme puisque chaque combinaison est équiprobable.

Un tirage pour lequel $X = k$ est constitué de k boules octarines choisies parmi Np, ce qui laisse $\binom{Np}{k}$ possibilités, et de $n - k$ boules blanches parmi Nq, ce qui donne $\binom{Nq}{n-k}$ possibilités. Le nombre de tirages vérifiant $X = k$ vaut donc $\binom{Np}{k}\binom{Nq}{n-k}$.

On en déduit que

$$P(X = k) = \frac{\binom{Np}{k}\binom{Nq}{n-k}}{\binom{N}{n}}.$$

On contate ainsi que X suit la loi $\mathscr{H}(N, n; p)$.

Cet exemple est représentatif du schéma hypergéométrique décrit ci-dessous.

Schéma hypergéométrique

Le cadre de la loi $\mathscr{H}(N, n; p)$ est celui d'un ensemble de N objets séparés en deux catégories (Np objets du premier type et Nq du second type) dans lequel on extrait un échantillon de n objets. La loi $\mathscr{H}(N, n; p)$ permet alors de connaître la probabilité du nombre de représentants de l'une (ou l'autre) des deux catégories.

Dans le langage des boules et des urnes, la loi hypergéométrique correspond à une suite de tirages sans remise.

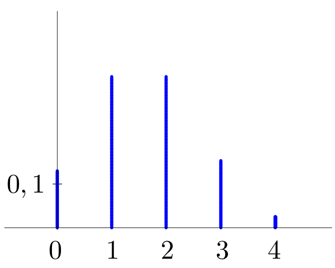
La loi hypergéométrique n'est pas au programme de MPSI. Cependant, on la rencontre souvent en exercice. Il faut donc savoir reconnaître le schéma hypergéométrique mais aussi être capable d'expliquer la formule donnant les valeurs de la densité.

Les tirages étant sans remise, la probabilité de succès au deuxième tirage, sachant ce qui a été tiré au premier tirage, n'est plus celle du premier tirage. Autrement dit, si le premier tirage donne un succès (respectivement un échec), la probabilité de succès au second tirage est plus petite (respectivement plus grande) qu'au premier tirage. Il peut donc paraître curieux de considérer, dans la loi hypergéométrique, un paramètre de succès p constant. En fait, nous avons vu (dans l'exercice sur les urnes de Pólya) que, si l'on calcule la probabilité de succès au second tirage, inconditionnellement à ce qui s'est passé au premier tirage, on constate que celle-ci est égale à la probabilité de succès du premier tirage.

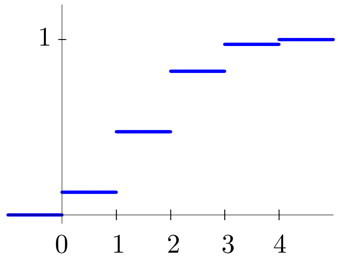
Exemples :

- Dans l'ancien Loto, on cochait 6 numéros parmi les numéros de 1 à 49. Au tirage, 6 numéros gagnants étaient désignés. En fonction du nombre X de numéros gagnants que l'on avait cochés, on gagnait un peu (ou beaucoup) de pognon. Ainsi, X comptait le nombre de succès (avoir coché un numéro gagnant) dans une suite de 6 cochages, parmi 49, sans remise (il est bien sûr impossible de cocher deux fois la même case, au risque de faire un trou dans le papier !), avec une probabilité de succès égale à $6/49$. Donc X suivait la loi $\mathscr{H}(49, 6; 6/49)$.
- Dans un paquet de bonbons contenant 10 arlequins et 33 berlingots, on pioche une poignée de 6 bonbons. On note X le nombre d'arlequins dans la poignée. Alors X compte le nombre de succès dans une suite de 6 choix de bonbons parmi 43, sans remise, avec une probabilité de succès de $10/43$. Donc $X \hookrightarrow \mathscr{H}(43, 6; 10/43)$.

Densité et fonction de répartition



densité de $\mathscr{H}(80, 4; 0, 4)$



fonction de répartition de $\mathscr{H}(80, 4; 0, 4)$

Tout comme celle de la loi binomiale, la densité de la loi hypergéométrique présente une forme caractéristique en cloche. Là encore, plus p est proche de 0, plus le sommet de la cloche est décalé vers la gauche (les succès sont peu nombreux car difficiles à obtenir) et plus p est proche de 1, plus le sommet de la cloche se décale vers la droite (les succès sont plus nombreux car plus faciles à obtenir).

L'énoncé suivant donne l'espérance et la variance d'une loi hypergéométrique. L'espérance est à savoir par ♡. La variance, par contre, est hors-programme (et n'est donnée qu'à titre culturel).

Proposition 17

Si X est une variable aléatoire hypergéométrique de paramètres N, n et p, on a

$$E(X) = np \qquad \text{et} \qquad V(X) = npq\frac{N-n}{N-1}.$$

■ Pour tout $\ell \in [\![1; n+1]\!]$, on a

$$\begin{aligned}
& E\big(X(X-1)\cdots(X-\ell+1)\big) \\
= {} & \sum_{k=\cancel{0}\,\ell}^{n} k(k-1)\cdots(k-\ell+1)\frac{\binom{Np}{k}\binom{Nq}{n-k}}{\binom{N}{n}} && \text{formule de transfert} \\
= {} & \frac{1}{\binom{N}{n}}\sum_{k=\ell}^{n} Np(Np-1)\cdots(Np-\ell+1)\binom{Np-\ell}{k-\ell}\binom{Nq}{n-k} && \text{formule du pion (ℓ fois)} \\
= {} & \frac{Np(Np-1)\cdots(Np-\ell+1)}{\binom{N}{n}}\sum_{j=0}^{n-\ell}\binom{Np-\ell}{j}\binom{Nq}{n-\ell-j} && \text{en posant } j = k-\ell \\
= {} & \frac{Np(Np-1)\cdots(Np-\ell+1)}{\binom{N}{n}}\binom{N-\ell}{n-\ell} && \text{formule du Vandermonde} \\
= {} & \frac{Np(Np-1)\cdots(Np-\ell+1)}{\binom{N}{n}}\,\frac{(n-\ell+1)(n-\ell+2)\cdots n}{(N-\ell+1)(N-\ell+2)\cdots N}\binom{N}{n} && \text{formule du pion (ℓ fois)} \\
= {} & \ell!\binom{n}{\ell}\frac{Np(Np-1)\cdots(Np-\ell+1)}{N(N-1)\cdots(N-\ell+1)}.
\end{aligned}$$

Pour $\ell = 1$, on a immédiatement

$$E(X) = 1 \times n \times \frac{Np}{N} = np.$$

Pour $\ell = 2$, on a

$$E\big(X(X-1)\big) = 2 \times \frac{n(n-1)}{2} \times \frac{Np(Np-1)}{N(N-1)} = n(n-1)\frac{p(Np-1)}{N-1},$$

donc, comme

$$V(X) = E(X^2) - E(X)^2 = E\big(X(X-1)\big) + E(X) - E(X)^2,$$

cela donne

$$V(X) = n(n-1)\frac{p(Np-1)}{N-1} + np - n^2p^2 = npq\frac{N-n}{N-1}.$$ ■

On constate que l'espérance de la loi $\mathscr{H}(N, n; p)$ est la même que celle de la loi $\mathscr{B}(n; p)$. Nous expliquerons pourquoi dans le chapitre sur les vecteurs aléatoires.

On remarque par ailleurs que la variance de la loi $\mathscr{H}(N, n; p)$ est obtenue en multipliant celle de la loi $\mathscr{B}(n; p)$ par le terme « correctif » $(N-n)/(N-1)$. Comme celui-ci est inférieur à 1, on constate que la variance de $\mathscr{H}(N, n; p)$ est inférieure à celle de $\mathscr{B}(n; p)$.

c) *Approximation d'une loi hypergéométrique par une loi binomiale*

Proposition 18

Soient $n \in \mathbb{N}$ et $p \in [0;1]$. On a

$$\frac{\binom{Np}{k}\binom{Nq}{n-k}}{\binom{N}{n}} \xrightarrow[N\to+\infty]{} \binom{n}{k}p^k q^{n-k},$$

où N tend vers $+\infty$ tout en respectant la condition $Np \in \mathbb{N}$. Autrement dit, à n et p fixés, on a

$$\mathscr{H}(N,n;p) \xrightarrow[N\to+\infty]{} \mathscr{B}(n;p).$$

■ On a

$$\begin{aligned}
\frac{\binom{Np}{k}\binom{Nq}{n-k}}{\binom{N}{n}} &= \frac{\dfrac{(Np)(Np-1)\cdots(Np-k+1)}{k!} \times \dfrac{(Nq)(Nq-1)\cdots(Nq-n+k+1)}{(n-k)!}}{\dfrac{(N)(N-1)\cdots(N-n+1)}{n!}} \\
&= \frac{n!}{k!(n-k)!}\,\frac{(Np)(Np-1)\cdots(Np-k+1)\times(Nq)(Nq-1)\cdots(Nq-n+k+1)}{(N)(N-1)\cdots(N-n+1)} \\
&\underset{N\to\infty}{\sim} \binom{n}{k}\frac{(Np)^k(Nq)^{n-k}}{N^n} \\
&\underset{N\to\infty}{\sim} \binom{n}{k}p^k q^{n-k}
\end{aligned}$$

ce qui donne le résultat. ■

La convergence de la loi $\mathscr{H}(N,n;p)$ vers la loi $\mathscr{B}(n;p)$ quand N tend vers $+\infty$ (avec n et p constants) est parfaitement cohérente avec les modélisations des lois binomiale et hypergéométrique. En effet, si la population totale est grande devant la taille de l'échantillon extrait, les proportions des différents types d'objets ne varient pas, que l'on procède à des tirages avec ou sans remise.

On constate par ailleurs que l'espérance et la variance d'une loi hypergéométrique tendent vers l'espérance et la variance d'une loi binomiale lorsque N tend vers $+\infty$.

En pratique, dès que $N \geqslant 10n$, nous pourrons approcher la loi hypergéométrique de paramètre N, n et p par la loi binomiale de paramètres n et p.

Exemples :

- Pour $N = 50$, $n = 4$ et $p = 1/10$, si $B \hookrightarrow \mathscr{B}(4;1/10)$ et $H \hookrightarrow \mathscr{H}(50;4;1/10)$, on a

k	0	1	2	3	4
$P(B=k)$	$0,6561$	$0,2916$	$0,0486$	$0,0036$	$0,0001$
$P(H=k)$	$0,6470$	$0,3081$	$0,0430$	$0,0020$	$0,0002$

- Lorsqu'un prédateur prélève, dans la nature, des proies pour les manger, on considère en général que les prélèvements sont suffisamment faibles pour que la population de proies reste stable. Dans ce cas, on considère que les prélèvements sont indépendants, bien qu'ils soient sans remise. Cela revient à remplacer la loi hypergéométrique par la loi binomiale. Si les proies viennent à manquer (en général parce que le prédateur est un bipède du type sapiens sapiens qui ne sait pas exploiter raisonnablement son milieu) ou si les prédateurs pullulent (en général parce que le même bipède a zigouillé tous les prédateurs du prédateur), l'approximation de la loi hypergéométrique par la loi binomiale n'est plus valable. 5 h 00