Chapitre 4

Mesure de la variabilité

Le propre d'une v.a. c'est précisément de pouvoir prendre plus qu'une valeur, l'importance de ces valeurs étant contrôlée par la probabilité choisie. Nous considérons ici le cas des v.a. unidimensionnelles. Cette étude est également utile pour des v.a. à d dimensions, dans la mesure où elle s'applique à chacune des composantes. Toutefois, le cas multidimensionnel ouvre la question du comportement des diverses composantes l'une vis-à-vis de l'autre qui fera l'objet du chapitre suivant. Dans tout ce chapitre on supposera systématiquement que X *est de carré intégrable* (on note cela $X \in L^2(\Omega)$), c'est-à-dire que $\mathbb{E}(X^2) \leq \infty$, ce qui est toujours vrai lorsque X est une v.a. élémentaire.

4.1 Variance

Soit X une v.a. (unidimensionelle) et $\lambda \in \mathbb{R}$ un nombre. Considérons la quantité $\mathrm{Var}_\lambda(X)$ définie par

$$\mathrm{Var}_\lambda(X) := \mathbb{E}((X-\lambda)^2).$$

Il s'agit de l'espérance de la v.a. $(X-\lambda)^2$ qui est positive, donc d'espérance positive, et nous voyons que si $\mathrm{Var}_\lambda(X) = 0$ l'évènement $\{X \neq \lambda\}$ est de probabilité nulle. Voilà pourquoi on peut comprendre $\mathrm{Var}_\lambda(X)$ comme une mesure[1] de la variabilité de la v.a. X vis-à-vis de la valeur fixée λ. Une question naturelle maintenant est de déterminer (si possible) une valeur de λ vis-à-vis de laquelle la variablité de X est minimale. Le problème est simple ; en voici la solution :

Proposition 4.1 *Parmi toutes les valeurs de $\lambda \in \mathbb{R}$, celle pour pour laquelle $Var_\lambda(X)$ est la plus petite est $\lambda^* := \mathbb{E}(X)$.*

Preuve : Soit $\varphi(\lambda) := \mathrm{Var}_\lambda(X) = \mathbb{E}((X-\lambda)^2) = \lambda^2 - 2\lambda\mathbb{E}(X) + \mathbb{E}(X^2)$; le minimum λ^* de cette fonction polynôme de degré 2 de λ est la solution de $\varphi'(\lambda^*) = 0$. Or $\varphi'(\lambda) = 2(\lambda - \mathbb{E}(X))$, ce qui montre que $\lambda^* = \mathbb{E}(X)$ comme annoncé. □

Il est donc naturel de retenir $\mathrm{Var}_{\lambda_*}(X)$ comme mesure intrinsèque de la variabilité de la v.a. X ; c'est ce que l'on fait en posant la définition suivante :

Définition : On appelle *variance* de la v.a. unidimensionnelle X et on note $\mathrm{Var}(X)$ le nombre défini par

$$\boxed{\mathrm{Var}(X) := \mathbb{E}((X - \mathbb{E}(X))^2)}$$

On appelle é*cart-type* de la v.a. réelle X et on note $\sigma(X)$ le nombre défini par

$$\boxed{\sigma(X) := \sqrt{\mathrm{Var}(X)} = \sqrt{\mathbb{E}((X-\mathbb{E}(X))^2)}}$$

Proposition 4.2 *Pour tout $a \in \mathbb{R}$, on a* $\boxed{Var(aX) = a^2\, Var(X)}$ *et* $\boxed{\sigma(aX) = |a|\sigma(X)}$.

[1]Notons que si X et λ ont une "dimension" au sens de la Physique, c'est-à-dire représentent par exemple un nombre de mètres (m) ou d'euros (EUR), alors $\mathrm{Var}_\lambda(X)$ s'exprime en m^2 ou EUR^2, et il serait plus naturel de considérer la racine de $\mathrm{Var}_\lambda(X)$; les calculs en seraient en revanche compliqués et il est aisé de vérifier conceptuellement qu'ils aboutiraient aux mêmes conclusions.

Exercice : On suppose que $\sigma(X) = 0$; que peut-on dire de $\mathbb{P}(\{X \neq \mathbb{E}(X)\})$? (on supposera que la v.a. X est élémentaire)

Proposition 4.3 (formule de Huygens) $\boxed{Var(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2}$

Preuve : . Notons $x := \mathbb{E}(X)$; on a $\mathrm{Var}\,(X) := \mathbb{E}((X-x)^2) = \mathbb{E}(X^2-2xX+x^2) = \mathbb{E}(X^2)-2x\mathbb{E}(X)+x^2 = \mathbb{E}(X^2) - 2\mathbb{E}(X)^2 + (\mathbb{E}(X))^2 = \mathbb{E}(X^2) - (\mathbb{E}(X))^2$. □

Remarque : La formule de Huygens peut encore s'écrire

$$\mathbb{E}(X^2) = (\mathbb{E}(X))^2 + \mathrm{Var}\,(X). \tag{4.1}$$

Il s'agit en fait d'une formule de Pythagore : nous verrons au chapitre suivant en quoi la v.a. $\mathbb{E}(X)$ (en fait, non aléatoire) est toujours orthogonale à la v.a. $X - \mathbb{E}(X)$. Si l'on comprend que l'application $Y \mapsto \mathbb{E}(Y^2)$ comme le carré d'une "norme" $\|Y\|_{L^2}$, nous voyons que la formule de Huygens-Pythagore (4.1) n'est autre que $\|X\|_{L^2}^2 = \|\mathbb{E}(X)\|_{L^2}^2 + \|X - \mathbb{E}(X)\|_{L^2}^2$. En fait $Y \mapsto \sqrt{\mathbb{E}(X^2)}$ n'est "pas tout-à-fait" une norme, puisque $\|Y\|_{L^2} = 0$ n'entraine pas que $Y = 0$, mais seulement que $\mathbb{P}(\{Y \neq 0\}) = 0$, une petite subtilité à laquelle il faut s'habituer : on dit que $Y = 0$ *presque sûrement* (p.s.).

4.2 Variance de quelques lois

4.2.1 Loi de Bernoulli

Définition : On dit que la v.a. X suit une *loi de Bernoulli* et on note $X \rightsquigarrow \mathcal{B}(1,p)$ où $p \in]0,1[$ si et seulement si $X(\Omega) = \{0,1\}$ et $\mathbb{P}(\{X = 1\}) = p$.

Proposition 4.4 $\boxed{\textit{Si } X \rightsquigarrow \mathcal{B}(1,p), \textit{ alors } \mathbb{E}(X) = p \textit{ et } Var(X) = p(1-p).}$

4.2.2 Loi binomiale

La *loi binomiale* $\mathcal{B}(n,p)$ est la loi suivie par toute somme de n v.a. de Bernoulli $\mathcal{B}(1,p)$ indépendantes. La proposition suivante en précise quelques caractéristiques :

Proposition 4.5 *Soient $X_1, \ldots, X_n$ n v.a. de Bernoulli indépendantes, $X_i \in \mathcal{B}(1,p)$ pour tout $i = 1..n$. Soit $X = X_1 + \ldots + X_n \rightsquigarrow \mathcal{B}(n,p)$; alors,* $\boxed{\textit{pour tout } k = 0..n,\ \mathbb{P}(X = k) = C_n^k = \frac{n\,!}{k\,!(n-k)\,!}.}$

De plus $\boxed{\mathbb{E}(X) = np}$ *et* $\boxed{Var(X) = np(1-p)}$.

Preuve : Posons $\mathcal{C}_n^k := \{x = (x_1, \ldots, x_n) \in \{0,1\}^n \mid x_1 + \ldots + x_n = k\}$. On sait que $\mathrm{Card}\,(\mathcal{C}_n^k) = C_n^k = \frac{n\,!}{k\,!(n-k)\,!}$. En effet, on a $\mathcal{C}_n^k = \mathcal{Y}_{n-1}^k \,\dot\cup\, \mathcal{Z}_{n-1}^{k-1}$ avec $\mathcal{Y}_{n-1}^k = \{x = (y_1, \ldots, y_{n-1}, 0)$ avec $y = (y_1, \ldots, y_{n-1}) \in \mathcal{C}_{n-1}^k\}$ et $\mathcal{Z}_{n-1}^{k-1} = \{x = (z_1, \ldots, z_{n-1}, 1)$, avec $z = (z_1, \ldots, z_{n-1}) \in \mathcal{C}_{n-1}^{k-1}\}$. Donc $\mathrm{Card}\,(\mathcal{C}_n^k) = \mathrm{Card}\,(\mathcal{Y}_{n-1}^k) + \mathrm{Card}\,(\mathcal{Z}_{n-1}^{k-1})$. On en déduit facilement, par récurrence, que $\mathrm{Card}\,(\mathcal{C}_n^k) = C_n^k := \frac{n\,!}{k\,!(n-k)\,!}$. A présent

$$\begin{aligned}
\mathbb{P}(\{X = k\}) &= \mathbb{P}(\{X_1 = x_1, \ldots, X_n = x_n \text{ , avec } x = (x_1, \ldots, x_n) \in \mathcal{C}_n^k\}) \\
&= \mathbb{P}\left(\dot{\bigcup}_{x \in \mathcal{C}_n^k} \{X_1 = x_1, \ldots, X_n = x_n\}\right) = \sum_{x \in \mathcal{C}_n^k} \mathbb{P}(\{X_1 = x_1, \ldots, X_n = x_n\}) \\
&= \sum_{x \in \mathcal{C}_n^k} \mathbb{P}(\{X_1 = x_1\} \cap \ldots \cap \{X_n = x_n\}) \\
&= \sum_{x \in \mathcal{C}_n^k} \mathbb{P}(\{X_1 = x_1\}) \cdots \mathbb{P}(X_n = x_n\}) \text{ car les } X_i \text{ sont indépendantes} \\
&= \sum_{x \in \mathcal{C}_n^k} p^k (1-p)^{n-k} \text{ puisque } x_1 + \ldots + x_n = k \\
&= p^k (1-p)^{n-k} \sum_{x \in \mathcal{C}_n^k} 1 = p^k (1-p)^{n-k} \mathrm{Card}\,(\mathcal{C}_n^k) = C_n^k p^k (1-p)^{n-k}.
\end{aligned}$$

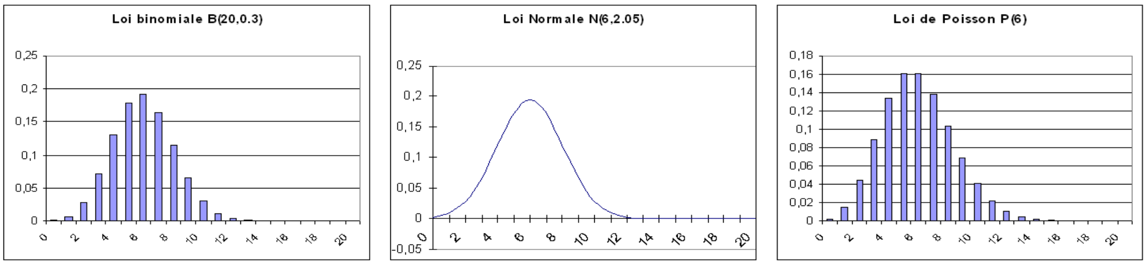

FIG. 4.1 – Histogramme et graphe de trois lois ayant même espérance 6 : la loi Binomiale $\mathcal{B}(20, 0.3)$, la loi normale $\mathcal{N}(6, \sqrt{4.2})$, et la loi de Poisson $\mathcal{P}(6)$. Les deux première ont la même variance 4.2 ; la valeur de la variance (6) de la loi de Poisson est imposée par le choix de son espérance. Nous verrons dans quelle mesure une loi binômiale $\mathcal{B}(n,p)$ peut être approchée, pour n grand, par une loi normale (théorème central-limit) de même espérance et même variance, ou par une loi de Poisson $\mathcal{P}(\lambda)$, pour $\lambda = \lim_n np_n$.

Finalement $\mathbb{E}(X_1+\ldots+X_n) = \mathbb{E}(X_1)+\ldots+\mathbb{E}(X_n) = np$; le fait que $\mathrm{Var}\,(X_1+\ldots+X_n) = \mathrm{Var}\,(X_1)+\ldots+\mathrm{Var}\,(X_n) = np(1-p)$ résulte du fait que les v.a. sont indépendantes et de la proposition 4.9 ci-dessous. □

4.2.3 Loi de Poisson

Définition : Soit $\lambda > 0$; on dit que la v.a. X suit une *loi de Poisson* de paramètre λ et on note $X \rightsquigarrow \mathcal{P}(\lambda)$ si et seulement si $X(\Omega) = \mathbb{N}$ et $\mathbb{P}(X = k) = c\frac{\lambda^k}{k\,!}$, avec $c = e^{-\lambda}$.

Proposition 4.6 $\boxed{\textit{Si } X \rightsquigarrow \mathcal{P}(\lambda), \textit{ alors } \mathbb{E}(X) = \lambda \textit{ et } Var(X) = \lambda.}$

Preuve : laissée en exercice. Indication : pour le calcul de $\mathrm{Var}\,(X)$ commencer par calculer $\mathbb{E}(X(X-1))$ puis en déduire $\mathbb{E}(X^2)$ et finalement $\mathrm{Var}\,(X)$ en appliquant la formule de Huygens. □

4.2.4 Loi Exponentielle

Définition : Soit $\lambda > 0$; on dit que la v.a. X suit une *loi exponentielle* de paramètre λ et on note $X \rightsquigarrow \mathcal{E}(\lambda)$ si et seulement si $X(\Omega) = \mathbb{R}^+$ et X admet pour densité la fonction $f_X(x) = ce^{-\lambda x}\mathbb{I}_{\mathbb{R}^+}(x)$, avec $c = \lambda$.

Proposition 4.7 $\boxed{\textit{Si } X \rightsquigarrow \mathcal{E}(\lambda), \textit{ alors } \mathbb{E}(X) = \frac{1}{\lambda} \textit{ et } Var(X) = \frac{1}{\lambda^2}.}$

Preuve : laissée en exercice. □

4.2.5 Loi normale

Définition : Soient $\mu \in \mathbb{R}$ et $\sigma > 0$; on dit que la v.a. X suit une *loi normale* d'espérance μ et écart-type σ et on note $X \rightsquigarrow \mathcal{N}(\mu, \sigma)$ si et seulement si $X(\Omega) = \mathbb{R}$ et X admet pour densité la fonction $f_X(x) = ce^{-\frac{(x-\mu)^2}{2\sigma^2}}$, avec $c = \frac{1}{\sigma\sqrt{2\pi}}$.

Proposition 4.8 $\boxed{\textit{Si } X \rightsquigarrow \mathcal{N}(\mu, \sigma), \textit{ alors } \mathbb{E}(X) = \mu \textit{ et } Var(X) = \sigma^2.}$

Preuve : laissée en exercice. Indication : considérer d'abord le cas de $Y = \frac{1}{\sigma}(X-\mu) \rightsquigarrow \mathcal{N}(0,1)$. □

Exercice : Que pouvez-vous dire de $\mathbb{P}(\{X = 21\})$ et $\mathbb{P}(\{X \leq 21\})$ pour X suivant chacune des trois lois de la figure 4.1.

4.3 Variance et indépendance

Proposition 4.9 *Soient $X_1, \ldots, X_n$ n v.a.. Alors*

$$\boxed{Var(X_1 + \ldots + X_n) = Var(X_1) + \ldots + Var(X_n) \quad \text{si les v.a. } X_1, \ldots, X_n \text{ sont indépendantes.}}$$

Preuve : Posons $Y_i := X_i - \mathbb{E}(X_i)$ de telle sorte que $\mathbb{E}(Y_i) = \mathbb{E}(X_i - \mathbb{E}(X_i)) = \mathbb{E}(X_i) - \mathbb{E}(\mathbb{E}(X_i)) = \mathbb{E}(X_i) - \mathbb{E}(X_i) = 0$. A présent, nous voyons que

$$\begin{aligned} \mathrm{Var}\,(X_1 + \ldots + X_n) &= \mathbb{E}((X_1 + \ldots + X_n - \mathbb{E}(X_1 + \ldots + X_n))^2) = \mathbb{E}((X_1 - \mathbb{E}(X_1) + \ldots + X_n - \mathbb{E}(X_n))^2) \\ &= \mathbb{E}((Y_1 + \ldots + Y_n)^2) = \mathbb{E}\left(\sum_{i,j=1}^{n} Y_i Y_j\right) = \sum_{i,j=1}^{n} \mathbb{E}(Y_i Y_j). \end{aligned}$$

C'est à présent que nous utilisons l'hypothèse que pour $i \neq j$ les v.a. Y_i et Y_j sont indépendantes (puisque fonction chacune d'une des v.a. indépendantes X_i et X_j) et donc $\mathbb{E}(Y_iY_j) = \mathbb{E}(Y_i)\mathbb{E}(Y_j) = 0{\cdot}0 = 0$, puisque $\mathbb{E}(Y_i) = 0 = \mathbb{E}(Y_j)$. Donc, dans la sommes $\sum_{i,j=1}^{n} \mathbb{E}(Y_iY_j)$, ne subsistent que les termes tels que $i = j$, et on obtient

$$\mathrm{Var}\,(X_1 + \ldots + X_n) = \sum_{i=1}^{n} \mathbb{E}(Y_iY_i) = \sum_{i=1}^{n} \mathbb{E}(Y_i^2) = \sum_{i=1}^{n} \mathbb{E}((X_i - \mathbb{E}(X_i))^2) = \sum_{i=1}^{n} \mathrm{Var}\,(X_i).$$

□

Remarque : l'hypothèse d'indépendance dans la proposition 4.9 est essentielle ; en effet, soit par exemple X une v.a. de variance non nulle ; posons $X_1 := +X$ et $X_2 := -X$. Alors

$$\begin{aligned} \mathrm{Var}\,(X_1 + X_2) &= \mathrm{Var}\,(X - X) = \mathrm{Var}\,(0) = 0 \\ &\neq 2\mathrm{Var}\,(X) = \mathrm{Var}\,(X) + \mathrm{Var}\,(X) = \mathrm{Var}\,(X) + \mathrm{Var}\,(-X) = \mathrm{Var}\,(X_1) + \mathrm{Var}\,(X_2). \end{aligned}$$

Le corollaire suivant est une version probabiliste de l'adage que veut qu'il vaille mieux mettre ses oeufs dans deux paniers :

Corollaire 4.10 *Soient X_1, X_2, …, X_n n v.a. indépendantes ; soit $M := \frac{1}{n}(X_1 + \ldots + X_n)$ leur moyenne. Alors $Var(M) = \frac{1}{\sqrt{n}} Var(X_1)$.*

Exercice : Montrer le corollaire. Donner un modèle probabiliste étayant l'adage ; **indication :** postuler que la seule manière de briser un oeuf est de faire tomber le panier qui le contient et dans ce cas tous les oeufs du panier sont brisés. Discuter ce postulat et l'adage en termes de variance.

Exercice : Soient X et Y deux v.a. de même loi. Lorsqu'on découvre les probabilités et la notion (essentielle) de loi d'une v.a. il vient immanquablement le moment où l'on se pose la question de savoir si pour deux v.a. X et Y avoir la même loi est synonyme d'être égales... Qu'en pensez-vous ? Montrer que si X et Y sont indépendantes, $\mathrm{Var}\,(X - Y) = \mathrm{Var}\,(X) + \mathrm{Var}\,(Y)$. Voyez-vous le lien avec l'interrogation évoquée juste avant ? Si l'on renonce à l'hypothèse d'indépendance, peut-on avoir $\mathrm{Var}\,(X - Y) > \mathrm{Var}\,(X) + \mathrm{Var}\,(Y)$? (N'hésitez pas à supposer que $\mathbb{E}(X) = 0 = \mathbb{E}(Y)$). Il serait naturel que vous vous demandiez alors si la variance $\mathrm{Var}\,(X - Y)$ peut être arbitrairement grande ; la réponse est non : au chapitre suivant, l'inégalité de Cauchy-Schwarz nous permettra de montrer que $\mathrm{Var}\,(X - Y) \leq (\sigma(X) + \sigma(Y))^2$.