# POSITION ET DISPERSION D'UNE VARIABLE ALÉATOIRE RÉELLE

Dans tout ce chapitre, $(\Omega, P)$ est un espace probabilisé **FINI**.

## 1 ESPÉRANCE D'UNE VARIABLE ALÉATOIRE RÉELLE

**Définition (Espérance d'une variable aléatoire réelle)** Soit $X$ une variable aléatoire réelle sur $\Omega$.

- On appelle *espérance de* $X$ le réel : $\mathrm{E}(X) = \displaystyle\sum_{x \in X(\Omega)} P(X=x)\, x$.
- On dit que $X$ est *centrée* si : $\mathrm{E}(X) = 0$.

L'espérance de $X$ est une moyenne des valeurs de $X$, donc un *indicateur de position*. Précisément, chaque valeur $x$, $x$ décrivant $X(\Omega)$, s'y trouve comptabilisée en proportion de sa probabilité d'occurrence $P(X=x)$. Ainsi, plus $P(X=x)$ est proche de 1, plus $x$ a d'importance dans le calcul.

**Exemple** Un dé à 6 faces numérotées de 1 à 6 a été truqué de telle sorte que les faces 1, 2 et 3 tombent avec probabilité $\dfrac{1}{6}$, les faces 4 et 5 avec probabilité $\dfrac{1}{12}$ et la face 6 avec probabilité $\dfrac{1}{3}$. Quel numéro obtient-on en moyenne ?

**Démonstration** Cet énoncé a du sens car : $3 \times \dfrac{1}{6} + 2 \times \dfrac{1}{12} + \dfrac{1}{3} = 1$. Notons $X$ le numéro obtenu avec un tel dé après un lancer. Par hypothèse : $P(X=1) = P(X=2) = P(X=3) = \dfrac{1}{6}$, $P(X=4) = P(X=5) = \dfrac{1}{12}$ et $P(X=6) = \dfrac{1}{3}$. Ainsi : $\mathrm{E}(X) = \displaystyle\sum_{k=1}^{6} P(X=k)\, k = \frac{1+2+3}{6} + \frac{4+5}{12} + \frac{6}{3} = 1 + \frac{3}{4} + 2 = 3,75$.

**Définition-théorème (Espérance des lois usuelles)** Soit $X$ une variable aléatoire réelle sur $\Omega$.

(i) **Variables aléatoires constantes :** Si $X$ est constante de valeur $m$ : $E(X) = m$.

(ii) **Loi uniforme :** Soit $E = \{x_1, \ldots, x_n\}$ une partie de $\mathbb{R}$. Si : $X \hookrightarrow \mathscr{U}(E)$, alors $\mathrm{E}(X)$ est la moyenne « naturelle » des valeurs $x_1, \ldots, x_n$ de $X$ : $\mathrm{E}(X) = \dfrac{1}{n}\displaystyle\sum_{k=1}^{n} x_k$.

(iii) **Loi de Bernoulli :** Soit $p \in [0,1]$. Si : $X \hookrightarrow \mathscr{B}(p)$, alors : $\mathrm{E}(X) = p$.

**Exemple fondamental :** Pour tout événement $A \in \mathscr{P}(\Omega)$ : $\mathrm{E}(\mathbb{1}_A) = P(A)$.

(iv) **Loi binomiale :** Soient $n \in \mathbb{N}^*$ et $p \in [0,1]$. Si : $X \hookrightarrow \mathscr{B}(n,p)$, alors : $\mathrm{E}(X) = np$.

**Démonstration**

(i) Par hypothèse : $X(\Omega) = \{m\}$ et $P(X=m) = P(\Omega) = 1$, donc : $\mathrm{E}(X) = m \times 1 = m$.

(ii) Pour tout $k \in [\![1,n]\!]$ : $P(X = x_k) = \dfrac{1}{n}$, donc : $\mathrm{E}(X) = \displaystyle\sum_{k=1}^{n} \frac{x_k}{n} = \frac{1}{n}\sum_{k=1}^{n} x_k$.

(iii) Par hypothèse : $X(\Omega) = \{0,1\}$, donc : $\mathrm{E}(X) = P(X=1) \times 1 + P(X=0) \times 0 = p$. Dans le cas où : $X = \mathbb{1}_A$ pour un certain $A \in \mathscr{P}(\Omega)$, nous savons déjà que : $\mathbb{1}_A \hookrightarrow \mathscr{B}\big(P(A)\big)$, donc : $\mathrm{E}(\mathbb{1}_A) = P(A)$.

(iv) Dans l'anneau $\mathbb{R}[T]$ : $(T+1-p)^n = \displaystyle\sum_{k=0}^{n} \binom{n}{k} T^k (1-p)^{n-k}$. Dérivons cette identité, puis multiplions par $T$ : $nT(T+1-p)^{n-1} = \displaystyle\sum_{k=0}^{n} \binom{n}{k} k T^k (1-p)^{n-k}$. Évaluons en $p$ : $np = \displaystyle\sum_{k=0}^{n} \binom{n}{k} k p^k (1-p)^{n-k}$.

Finalement : $\quad \mathrm{E}(X)=\sum_{k=0}^{n} P(X=k)\, k=\sum_{k=0}^{n}\binom{n}{k} k p^k(1-p)^{n-k}=np.$ ■

**Théorème (Propriétés de l'espérance)** Soient $X$ et $Y$ deux variables aléatoires réelles sur $\Omega$.

(i) **Une autre expression :** $\quad \mathrm{E}(X)=\sum_{\omega\in\Omega} P\big(\{\omega\}\big)\, X(\omega).$

(ii) **Linéarité :** Pour tous $\lambda,\mu\in\mathbb{R}$ : $\quad \mathrm{E}(\lambda X+\mu Y)=\lambda\mathrm{E}(X)+\mu\mathrm{E}(Y).$

(iii) **Positivité :** Si : $\quad X\geqslant 0,\quad$ alors : $\quad \mathrm{E}(X)\geqslant 0.\qquad$ **Croissance :** Si : $\quad X\leqslant Y,\quad$ alors : $\quad \mathrm{E}(X)\leqslant \mathrm{E}(Y).$

(iv) **Inégalité triangulaire :** $\quad \big|\mathrm{E}(X)\big|\leqslant \mathrm{E}\big(|X|\big).$

**Démonstration**

(i) Les événements $\{X=x\}$, $x$ décrivant $X(\Omega)$, forment un système complet d'événements, donc :

$$\sum_{\omega\in\Omega} P\big(\{\omega\}\big)X(\omega)=\sum_{x\in X(\Omega)}\sum_{\omega\in\{X=x\}} P\big(\{\omega\}\big)X(\omega)=\sum_{x\in X(\Omega)}\left(\sum_{\omega\in\{X=x\}} P\big(\{\omega\}\big)\right)x=\sum_{x\in X(\Omega)} P(X=x)\,x=\mathrm{E}(X).$$

(ii) $$\mathrm{E}(\lambda X+\mu Y)=\sum_{\omega\in\Omega} P\big(\{\omega\}\big)\,(\lambda X+\mu Y)(\omega)=\lambda\sum_{\omega\in\Omega} P\big(\{\omega\}\big)\,X(\omega)+\mu\sum_{\omega\in\Omega} P\big(\{\omega\}\big)\,Y(\omega)=\lambda\mathrm{E}(X)+\mu\mathrm{E}(Y).$$

(iii) Positivité évidente par définition de l'espérance. Pour la croissance, si : $\quad X\leqslant Y,\quad$ alors : $\quad Y-X\geqslant 0$, donc par linéarité : $\quad \mathrm{E}(Y)-\mathrm{E}(X)=\mathrm{E}(Y-X)\geqslant 0.$

(iv) Conséquence de la croissance, sachant que : $\quad -|X|\leqslant X\leqslant |X|.$ ■

**Exemple** Qu'obtient-on en moyenne quand on lance 2 fois un dé à 6 faces et qu'on additionne les résultats obtenus ?

**Démonstration** Notons $X_1$ (resp. $X_2$) le résultat du premier (resp. deuxième) lancer et $S=X_1+X_2$ leur somme. On cherche à calculer $\mathrm{E}(S)$. Bien sûr : $\quad X_1\hookrightarrow\mathscr{U}\big([\![1,6]\!]\big)\quad$ et $\quad X_2\hookrightarrow\mathscr{U}\big([\![1,6]\!]\big)$.

Il ne faut surtout pas ici calculer la loi de $S$ pour en déduire son espérance. Ce serait se donner beaucoup de peine pour rien. Plus simplement : $\quad \mathrm{E}(S)=\mathrm{E}(X_1)+\mathrm{E}(X_2)=\frac{1}{6}\sum_{k=1}^{6}k+\frac{1}{6}\sum_{k=1}^{6}k=2\times\frac{1}{6}\times\frac{6\times 7}{2}=7.$

**Exemple** Soient $p\in[0,1]$ et $X_1,\ldots,X_n$ des variables aléatoires indépendantes de même loi de Bernoulli $\mathscr{B}(p)$ définies sur un même espace probabilisé fini. Nous savons déjà que : $\quad \sum_{i=1}^{n} X_i\hookrightarrow\mathscr{B}(n,p).\quad$ Un nouveau calcul de l'espérance de la loi binomiale en découle : $\quad \mathrm{E}\left(\sum_{i=1}^{n}X_i\right)=\sum_{i=1}^{n}\mathrm{E}(X_i)=\sum_{i=1}^{n}p=np.$

**Exemple (Formule du crible)** Pour tous $A_1,\ldots,A_n\in\mathscr{P}(\Omega)$ : $\quad P\left(\bigcup_{i=1}^{n}A_i\right)=\sum_{k=1}^{n}(-1)^{k+1}\sum_{1\leqslant i_1<\ldots<i_k\leqslant n}P\big(A_{i_1}\cap\ldots\cap A_{i_k}\big).$

**Démonstration** Cette formule hors programme a été énoncée sans preuve au chapitre « Modélisation probabiliste des phénomènes finis ». On commence par un calcul d'indicatrices :

$$\mathbb{1}_{A_1\cup\ldots\cup A_n}=1-\mathbb{1}_{\overline{A_1\cup\ldots\cup A_n}}=1-\mathbb{1}_{\overline{A_1}\cap\ldots\cap\overline{A_n}}=1-\prod_{i=1}^{n}\mathbb{1}_{\overline{A_i}}=1-\prod_{i=1}^{n}\big(1-\mathbb{1}_{A_i}\big)=\sum_{k=1}^{n}(-1)^{k+1}\sum_{1\leqslant i_1<\ldots<i_k\leqslant n}\mathbb{1}_{A_{i_1}\cap\ldots\cap A_{i_k}}$$

et on conclut par linéarité de l'espérance : 
$$P\left(\bigcup_{i=1}^{n}A_i\right)=\mathrm{E}\big(\mathbb{1}_{A_1\cup\ldots\cup A_n}\big)=\sum_{k=1}^{n}(-1)^{k+1}\sum_{1\leqslant i_1<\ldots<i_k\leqslant n}\mathrm{E}\big(\mathbb{1}_{A_{i_1}\cap\ldots\cap A_{i_k}}\big)$$
$$=\sum_{k=1}^{n}(-1)^{k+1}\sum_{1\leqslant i_1<\ldots<i_k\leqslant n}P\big(A_{i_1}\cap\ldots\cap A_{i_k}\big).$$

**Théorème (Formule de transfert)** Soient $X$ une variable aléatoire sur $\Omega$ et $f:X(\Omega)\longrightarrow\mathbb{R}$ une fonction. L'espérance de $f(X)$ est entièrement déterminée par $f$ et la loi de $X$ : $\quad \mathrm{E}\big(f(X)\big)=\sum_{x\in X(\Omega)}P(X=x)\,f(x).$

**Démonstration** Les événements $\{X=x\}$, $x$ décrivant $X(\Omega)$, forment un système complet d'événements, donc :

$$\mathrm{E}\big(f(X)\big)=\sum_{\omega\in\Omega}P\big(\{\omega\}\big)\,f\big(X(\omega)\big)=\sum_{x\in X(\Omega)}\sum_{\omega\in\{X=x\}}P\big(\{\omega\}\big)\,f\big(X(\omega)\big)=\sum_{x\in X(\Omega)}\sum_{\omega\in\{X=x\}}P\big(\{\omega\}\big)\,f(x)=\sum_{x\in X(\Omega)}P(X=x)\,f(x).\quad\blacksquare$$

On dispose bien sûr d'une version pour les couples de la formule de transfert. Si $X$ et $Y$ sont deux variables aléatoires sur $\Omega$ et si $f:(X,Y)(\Omega)\longrightarrow F$ est une fonction, alors : $\mathrm{E}\big(f(X,Y)\big)=\sum_{(x,y)\in(X,Y)(\Omega)}P(X=x\text{ et }Y=y)\,f(x,y)$.

Par exemple : $\mathrm{E}(XY)=\sum_{(x,y)\in(X,Y)(\Omega)}P(X=x\text{ et }Y=y)\,xy$ dans le cas de la fonction $(x,y)\longmapsto xy$.

**Exemple** Pour toute variable aléatoire $X$ de loi uniforme sur $[\![1,n]\!]$ : $\mathrm{E}(X)=\dfrac{1}{n}\sum_{k=1}^{n}k=\dfrac{n+1}{2}\underset{n\to+\infty}{\sim}\dfrac{n}{2}$, et d'après la formule de transfert : $\mathrm{E}\big(X^2\big)=\sum_{k=1}^{n}P(X=k)\,k^2=\dfrac{1}{n}\sum_{k=1}^{n}k^2=\dfrac{(n+1)(2n+1)}{6}\underset{n\to+\infty}{\sim}\dfrac{n^2}{3}$.

**Théorème (Espérance du produit de deux variables aléatoires réelles indépendantes)** Soient $X$ et $Y$ deux variables aléatoires réelles sur $\Omega$. Si $X$ et $Y$ sont **INDÉPENDANTES**, alors : $\mathrm{E}(XY)=\mathrm{E}(X)\,\mathrm{E}(Y)$.

Ce résultat s'étend naturellement à un nombre fini quelconque de variables aléatoires indépendantes.

**Démonstration**

$$\mathrm{E}(X)\,\mathrm{E}(Y)=\left(\sum_{x\in X(\Omega)}P(X=x)\,x\right)\left(\sum_{y\in Y(\Omega)}P(Y=y)\,y\right)=\sum_{(x,y)\in X(\Omega)\times Y(\Omega)}P(X=x)\,P(Y=y)\,xy$$
$$\overset{\text{Indép.}}{=}\sum_{(x,y)\in(X,Y)(\Omega)}P(X=x\text{ et }Y=y)\,xy\overset{\text{Transfert}}{=}\mathrm{E}(XY).\quad\blacksquare$$

✖ **ATTENTION !** ✖ L'identité : $\mathrm{E}(XY)=\mathrm{E}(X)\,\mathrm{E}(Y)$ est fausse en général. Par exemple, pour toute variable aléatoire $X$ de loi uniforme sur $\{-1,1\}$ : $\mathrm{E}(X)=\dfrac{1}{2}\times1+\dfrac{1}{2}\times(-1)=0$ donc : $\mathrm{E}(X)^2=0$, mais : $\mathrm{E}\big(X^2\big)=\mathrm{E}(1)=1$.

## 2 VARIANCE, ÉCART-TYPE, COVARIANCE

**Définition (Moments, variance et écart-type d'une variable aléatoire réelle)** Soit $X$ une variable aléatoire réelle.

- Pour tout $k\in\mathbb{N}$, le réel $\mathrm{E}\big(X^k\big)$ est appelé le *moment d'ordre $k$ de $X$* et $\mathrm{E}\Big(\big(X-\mathrm{E}(X)\big)^k\Big)$ son *moment centré d'ordre $k$*.
- On appelle *variance de $X$* le réel positif : $\mathrm{V}(X)=\mathrm{E}\Big(\big(X-\mathrm{E}(X)\big)^2\Big)$ et *écart-type de $X$* le réel : $\sigma(X)=\sqrt{\mathrm{V}(X)}$.
- On dit que $X$ est *réduite* si : $\mathrm{V}(X)=1$.

L'espérance de $X$ est un indicateur de position, mais les valeurs de $X$ sont-elles plutôt proches de cette valeur moyenne ou plutôt éloignées ? Toute mesure de cette proximité à la moyenne est appelée un *indicateur de dispersion*. L'écart le plus naturel entre $X$ et son espérance est $\big|X-\mathrm{E}(X)\big|$, donc l'écart moyen de $X$ à sa moyenne est $\mathrm{E}\Big(\big|X-\mathrm{E}(X)\big|\Big)$, mais ce n'est pas là l'indicateur de dispersion que les mathématiciens ont choisi de mettre en avant. Ils lui ont préféré la variance, c'est-à-dire l'écart **QUADRATIQUE** moyen à la moyenne — « quadratique » parce qu'on passe au carré. Pourquoi ce choix moins naturel au premier abord ? Tout simplement parce que, comme on va le voir, la variance est plus facile à manipuler d'un point de vue calculatoire.

Et l'écart-type, quelle différence avec la variance ? Si par exemple $X$ représente une longueur, $\mathrm{V}(X)$ représente une longueur **AU CARRÉ**, donc il n'est pas possible de comparer directement la position moyenne $\mathrm{E}(X)$ et sa dispersion moyenne $\mathrm{V}(X)$ — un physicien dirait que l'espérance et la variance ne sont pas « homogènes ». L'écart-type, au contraire, est homogène à une longueur, donc comparable à l'espérance — d'où son intérêt.

**Théorème** **(Propriétés de la variance)** Soit $X$ une variable aléatoire réelle.

(i) **Expression développée :** $\mathrm{V}(X)=\mathrm{E}\big(X^2\big)-\mathrm{E}(X)^2$.

(ii) **Nullité :** $\mathrm{V}(X)=0$ si et seulement si : $P\big(X=\mathrm{E}(X)\big)=1$. On dit que $X$ est *presque sûrement constante*.

(iii) **Effet d'une transformation affine :** Pour tous $a,b\in\mathbb{R}$ : $\mathrm{V}(aX+b)=a^2\mathrm{V}(X)$.

En particulier, si : $\sigma(X)>0$, la variable aléatoire $\dfrac{X-\mathrm{E}(X)}{\sigma(X)}$ est centrée réduite.

**Démonstration**

(i) $\mathrm{V}(X)=\mathrm{E}\Big(X^2-2\overbrace{\mathrm{E}(X)}^{\text{constante}}X+\overbrace{\mathrm{E}(X)^2}^{\text{constante}}\Big)=\mathrm{E}\big(X^2\big)-2\mathrm{E}(X)\,\mathrm{E}(X)+\mathrm{E}(X)^2=\mathrm{E}\big(X^2\big)-\mathrm{E}(X)^2$.

(ii) D'après la formule de transfert : $\mathrm{V}(X)=\displaystyle\sum_{x\in X(\Omega)}\underbrace{P(X=x)\,\big(x-\mathrm{E}(X)\big)^2}_{\geqslant 0}$, donc :

$$\mathrm{V}(X)=0 \iff \forall x\in X(\Omega),\quad P(X=x)=0 \quad\text{ou}\quad x=\mathrm{E}(X) \iff \forall x\in X(\Omega)\setminus\big\{\mathrm{E}(X)\big\},\quad P(X=x)=0$$
$$\iff P\big(X=\mathrm{E}(X)\big)=1 \qquad\text{car :}\quad \sum_{x\in X(\Omega)}P(X=x)=1.$$

(iii) Comme $\mathrm{E}(aX+b)=a\mathrm{E}(X)+b$ : $\mathrm{V}(aX+b)=\mathrm{E}\Big(\big(aX+b-a\mathrm{E}(X)-b\big)^2\Big)=a^2\mathrm{E}\Big(\big(X-\mathrm{E}(X)\big)^2\Big)$.

Enfin, si : $\sigma(X)>0$, alors : $\mathrm{E}\left(\dfrac{X-\mathrm{E}(X)}{\sigma(X)}\right)=\dfrac{\mathrm{E}(X)-\mathrm{E}(X)}{\sigma(X)}=0$ et $\mathrm{V}\left(\dfrac{X-\mathrm{E}(X)}{\sigma(X)}\right)=\dfrac{\mathrm{V}(X)}{\sigma(X)^2}=1$. ■

**Définition-théorème** **(Covariance de deux variables aléatoires réelles)** Soient $X$ et $Y$ deux variables aléatoires réelles sur $\Omega$. On appelle *covariance de $X$ et $Y$* le réel : $\mathrm{cov}(X,Y)=\mathrm{E}\Big(\big(X-\mathrm{E}(X)\big)\big(Y-\mathrm{E}(Y)\big)\Big)$.

Clairement : $\mathrm{V}(X)=\mathrm{cov}(X,X)$ et $\mathrm{cov}(Y,X)=\mathrm{cov}(X,Y)$.

(i) **Expression développée :** $\mathrm{cov}(X,Y)=\mathrm{E}(XY)-\mathrm{E}(X)\,\mathrm{E}(Y)$.

(ii) **Variance d'une somme :** $\mathrm{V}(X+Y)=\mathrm{V}(X)+2\,\mathrm{cov}(X,Y)+\mathrm{V}(Y)$.

(iii) **Lien avec l'indépendance :** Si $X$ et $Y$ sont indépendantes : $\mathrm{cov}(X,Y)=0$ et $\mathrm{V}(X+Y)=\mathrm{V}(X)+\mathrm{V}(Y)$.

Les assertions (ii) et (iii) se généralisent au cas de variables aléatoires réelles $X_1,\ldots,X_n$ sur $\Omega$. Dans ce cas :

$$\mathrm{V}\left(\sum_{i=1}^n X_i\right)=\sum_{i=1}^n\mathrm{V}(X_i)+2\sum_{1\leqslant i<j\leqslant n}\mathrm{cov}(X_i,X_j),$$

et si $X_1,\ldots,X_n$ sont (seulement) DEUX À DEUX indépendantes : $\mathrm{V}\left(\displaystyle\sum_{i=1}^n X_i\right)=\displaystyle\sum_{i=1}^n\mathrm{V}(X_i)$.

✖ **ATTENTION !** ✖ La réciproque de l'assertion (iii) est fausse, la covariance de $X$ et $Y$ peut être nulle sans que $X$ et $Y$ soient indépendantes. C'est ce que montre le contre-exemple qui suit la preuve de la relation : $\mathrm{E}(XY)=\mathrm{E}(X)\,\mathrm{E}(Y)$.

**Démonstration**

(i) $\mathrm{cov}(X,Y)=\mathrm{E}\Big(XY-X\overbrace{\mathrm{E}(Y)}^{\text{constante}}-\overbrace{\mathrm{E}(X)}^{\text{constante}}Y+\overbrace{\mathrm{E}(X)\,\mathrm{E}(Y)}^{\text{constante}}\Big)=\mathrm{E}(XY)-\mathrm{E}(X)\,\mathrm{E}(Y)-\mathrm{E}(X)\,\mathrm{E}(Y)+\mathrm{E}(X)\,\mathrm{E}(Y)$
$=\mathrm{E}(XY)-\mathrm{E}(X)\,\mathrm{E}(Y)$.

(ii) $\mathrm{V}(X+Y)=\mathrm{E}\Big(\big(X-\mathrm{E}(X)+Y-\mathrm{E}(Y)\big)^2\Big)$
$=\mathrm{E}\Big(\big(X-\mathrm{E}(X)\big)^2+2\big(X-\mathrm{E}(X)\big)\big(Y-\mathrm{E}(Y)\big)+\big(Y-\mathrm{E}(Y)\big)^2\Big)=\mathrm{V}(X)+2\,\mathrm{cov}(X,Y)+\mathrm{V}(Y)$.

(iii) Si $X$ et $Y$ sont indépendantes : $\mathrm{cov}(X,Y)=\mathrm{E}(XY)-\mathrm{E}(X)\,\mathrm{E}(Y)=0$. ■

**Théorème** **(Variance des lois usuelles)** Soient $X$ une variable aléatoire réelle et $p\in[0,1]$.

(i) Si $X\hookrightarrow\mathscr{B}(p)$ : $\mathrm{V}(X)=p(1-p)$. (ii) Si $X\hookrightarrow\mathscr{B}(n,p)$ : $\mathrm{V}(X)=np(1-p)$.

**Démonstration**

(i) $\mathrm{V}(X)=\mathrm{E}\big(X^2\big)-\mathrm{E}(X)^2=\big(P(X=1)\times 1^2+P(X=0)\times 0^2\big)-p^2=p-p^2=p(1-p)$.

(ii) Soient $X_1, \ldots, X_n$ des variables aléatoires indépendantes sur $\Omega$ de même loi de Bernoulli $\mathscr{B}(p)$. Comme : $\sum_{i=1}^n X_i \hookrightarrow \mathscr{B}(n,p)$, alors par indépendance : $\mathrm{V}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \mathrm{V}(X_i) = \sum_{i=1}^n p(1-p) = np(1-p)$. Ce calcul ne concerne a priori pas la variable aléatoire $X$ de l'énoncé, mais l'espérance d'une variable aléatoire ne dépend que de sa loi, donc le calcul que nous venons de faire dans le cas particulier de la somme $\sum_{i=1}^n X_i$ est en fait emblématique. ■

## 3 Un premier pas vers les grands nombres

**Théorème (Inégalité de Markov)** Soit $X$ une variable aléatoire réelle. Pour tout $a > 0$ : $P\big(|X| \geqslant a\big) \leqslant \dfrac{\mathrm{E}\big(|X|\big)}{a}$.

Comme : $\big\{|X| > a\big\} \subset \big\{|X| \geqslant a\big\}$, on peut rendre l'inégalité stricte : $P\big(|X| > a\big) \leqslant \dfrac{\mathrm{E}\big(|X|\big)}{a}$.

**Démonstration** Tout repose sur les inégalités suivantes : $\mathbb{1}_{\{|X|\geqslant a\}} a \leqslant \mathbb{1}_{\{|X|\geqslant a\}} |X| \leqslant |X|$. Aussitôt, par croissance de l'espérance : $aP\big(|X| \geqslant a\big) = \mathrm{E}\big(\mathbb{1}_{\{|X|\geqslant a\}} a\big) \leqslant \mathrm{E}\big(|X|\big)$. ■

**Théorème (Inégalité de Bienaymé-Tchebychev)** Soit $X$ une variable aléatoire réelle. Pour tout $a > 0$ :

$$P\Big(\big|X - \mathrm{E}(X)\big| \geqslant a\Big) \leqslant \frac{\mathrm{V}(X)}{a^2}.$$

**Démonstration** D'après l'inégalité de Markov : $P\Big(\big(X - \mathrm{E}(X)\big)^2 \geqslant a^2\Big) \leqslant \dfrac{\mathrm{E}\Big(\big(X - \mathrm{E}(X)\big)^2\Big)}{a^2} = \dfrac{\mathrm{V}(X)}{a^2}$. Il reste à remarquer que les événements $\Big\{\big(X - \mathrm{E}(X)\big)^2 \geqslant a^2\Big\}$ et $\Big\{\big|X - \mathrm{E}(X)\big| \geqslant a\Big\}$ coïncident. ■

**Exemple** On dispose d'une pièce éventuellement truquée dont la probabilité d'obtention de pile est notée $p$. Pour connaître $p$, on lance cette pièce $n$ fois et on note $F$ la fréquence d'apparition de pile ainsi obtenue. À partir de quelle valeur de $n$ la probabilité pour que $F$ soit une approximation de $p$ à $10^{-2}$ près est-elle supérieure à 0,9 ?

**Démonstration** Nous noterons $N$ le nombre de piles obtenus — de sorte que : $F = \dfrac{N}{n}$. Par indépendance des lancers : $N \hookrightarrow \mathscr{B}(n,p)$, donc : $\mathrm{E}(N) = np$, puis : $\mathrm{E}(F) = p$. Ce résultat justifie à lui seul qu'on choisisse $F$ pour estimer $p$. D'après l'inégalité de Bienaymé-Tchebychev :

$$P\Big(|F-p| \geqslant 10^{-2}\Big) = P\Big(\big|N - \mathrm{E}(N)\big| \geqslant 10^{-2}n\Big) \leqslant \frac{\mathrm{V}(N)}{\big(10^{-2}n\big)^2} = \frac{np(1-p)}{10^{-4}n^2} = \frac{10^4}{n} \times p(1-p) \leqslant \frac{10^4}{n} \times \frac{1}{4} = \frac{2500}{n}.$$

Or : $P\Big(|F-p| < 10^{-2}\Big) \geqslant 0,9 \iff P\Big(|F-p| \geqslant 10^{-2}\Big) \leqslant 0,1$, donc il nous suffit de choisir $n$ avec la condition : $\dfrac{2500}{n} \leqslant 0,1$ pour garantir que $F$ soit une approximation de $p$ à $10^{-2}$ près. Conclusion : il faut quand même lancer 25 000 fois la pièce ! — ou trouver une majoration meilleure de $P\Big(|F-p| \geqslant 10^{-2}\Big)$.

Le paragraphe qui suit est l'aboutissement ultime de nos aventures probabilistes en MPSI. On s'intéresse à une grandeur d'espérance $m$ et d'écart-type $\sigma$. On la mesure dans un premier temps à l'occasion d'une unique expérience de résultat noté $X_1$. D'après l'inégalité de Bienaymé-Tchebychev, pour tout $\varepsilon > 0$ : $P\big(|X_1 - m| \geqslant \varepsilon\big) \leqslant \dfrac{\sigma^2}{\varepsilon^2}$. La mesure $X_1$ est donc à distance supérieure ou égale à $\varepsilon$ de $m$ avec probabilité au plus $\dfrac{\sigma^2}{\varepsilon^2}$.

Si l'on souhaite estimer $m$, l'expérience nous enseigne qu'une seule mesure a peu de chances de nous donner un résultat satisfaisant, nous obtiendrons une meilleure estimation de $m$ en augmentant le nombre de mesures. La théorie des probabilités confirme-t-elle cette intuition ? Quel résultat quantitatif pouvons-nous en tirer ? Nous effectuons à présent $n$ mesures $X_1,\ldots,X_n$ de la grandeur étudiée. Formellement, $X_1,\ldots,X_n$ sont des variables aléatoires réelles **INDÉPENDANTES** et **DE MÊME LOI**, d'espérance $m$ et d'écart-type $\sigma$. Nous noterons $\overline{X}_n$ la moyenne de ces $n$ mesures : $\overline{X}_n = \frac{1}{n}\sum_{k=1}^n X_k$. Par linéarité de l'espérance : $\mathrm{E}\big(\overline{X}_n\big) = \frac{1}{n}\sum_{k=1}^n \mathrm{E}(X_k) = \frac{1}{n}\sum_{k=1}^n m = m$ et par indépendance : $\mathrm{V}\big(\overline{X}_n\big) = \frac{1}{n^2}\sum_{k=1}^n \mathrm{V}(X_k) = \frac{1}{n^2}\sum_{k=1}^n \sigma^2 = \frac{\sigma^2}{n}$, donc : $\sigma\big(\overline{X}_n\big) = \frac{\sigma}{\sqrt{n}}$. En résumé :

> La moyenne de $n$ variables aléatoires **INDÉPENDANTES** et **DE MÊME LOI**, d'espérance $m$ et d'écart-type $\sigma$, admet toujours $m$ pour espérance, **MAIS** $\frac{\sigma}{\sqrt{n}}$ pour écart-type.

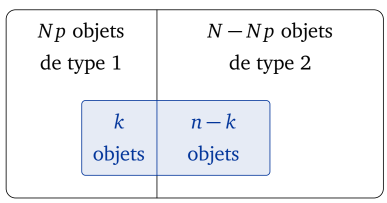
Avec $n$ mesures, nous avons ainsi gagné un facteur $\frac{1}{\sqrt{n}}$ en écart-type — et $\frac{1}{n}$ en variance. Intuitivement, cela veut dire que chaque nouvelle mesure nous rapproche de la certitude que la moyenne empirique $\overline{X}_n$ est proche de l'espérance $m$. Formellement, l'inégalité de Bienaymé-Tchebychev énonce cette fois que pour tout $\varepsilon > 0$ : $\boxed{P\Big(\big|\overline{X}_n - m\big| \geqslant \varepsilon\Big) \leqslant \frac{\sigma^2}{n\varepsilon^2}.}$

En particulier : $\lim\limits_{n\to+\infty} P\Big(\big|\overline{X}_n - m\big| \geqslant \varepsilon\Big) = 0$, ce que conforte notre intuition. Toute limite de ce genre est appelée une *loi des grands nombres*. Il en existe de nombreuses, dont l'inégalité ci-dessus — la plus faible de toutes, mais c'est déjà ça ! Les lois des grands nombres sont fondamentales car elles justifient l'interprétation *fréquentiste* que nous avons des probabilités. Qu'avons-nous en tête quand nous disons que la probabilité d'apparition de chaque face d'un dé équilibré à 6 faces vaut $\frac{1}{6}$ ? Nous signifions par là une vérité sur le monde qui n'a rien à voir avec le fait que de tels dés existent et qu'on peut les lancer un nombre indéfini de fois. Cette probabilité vaudrait $\frac{1}{6}$ même si l'humanité s'interdisait rigoureusement tout lancer de dé. Cela dit, le fait est que nous avons une autre notion de probabilité en tête simultanément, nous avons aussi l'habitude de constater que lorsqu'on lance un dé de nombreuses fois, la fréquence d'apparition de chaque face tend vers $\frac{1}{6}$ à mesure que ce nombre augmente. Les lois des grands nombres réconcilient les deux points de vue, la fréquence d'un événement lorsqu'on répète indéfiniment une expérience aléatoire tend vers la probabilité a priori de cet événement.

# 4 Un complément sur la loi hypergéométrique

La *loi hypergéométrique* est une loi usuelle hors programme, mais d'utilisation courante donc à étudier sérieusement.

Avec la loi hypergéométrique, on s'intéresse au tirage simultané de $n$ objets dans un ensemble de $N$ objets qui sont de deux types, disons de type 1 en proportion $p$ et de type 2 en proportion $1-p$. Parmi $N$ objets, $Np$ sont donc de type 1 et les $N-Np$ restants sont de type 2. Quelle est la loi du nombre $X$ des objets de type 1 parmi les $n$ tirés ?




On choisit pour univers $\Omega$ de cette expérience aléatoire l'ensemble $\mathscr{P}_n\big([\![1,N]\!]\big)$ des $n$-combinaisons de $[\![1,N]\!]$ — les objets de type 1 sont numérotés de 1 à $Np$ et les objets de type 2 de $Np+1$ à $N$ — et pour probabilité $P$ sur $\Omega$ la probabilité uniforme.

À présent, quelle est donc l'image de $X$ ? Lorsqu'on tire $k$ objets de type 1 et $n-k$ de type 2, deux inégalités sont nécessairement vraies : $0 \leqslant k \leqslant Np$ et $0 \leqslant n-k \leqslant N-Np$, équivalentes à :

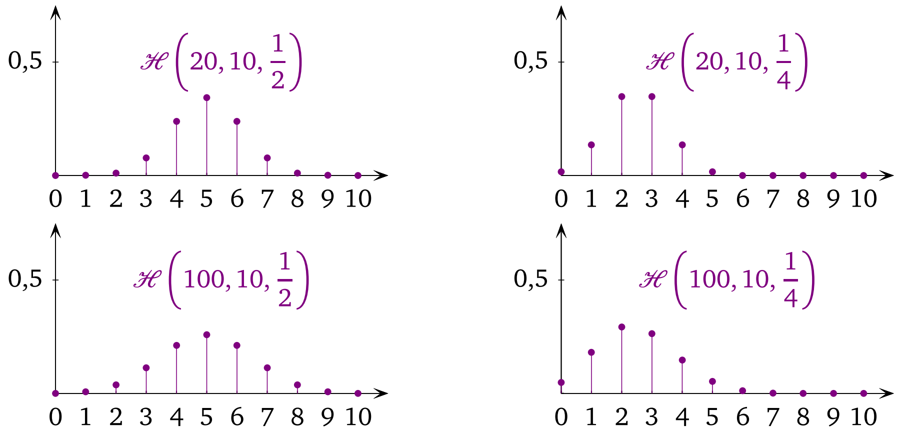
$$0 \leqslant k \leqslant Np \quad \text{et} \quad n-N+Np \leqslant k \leqslant n, \qquad \text{ou encore :} \quad k \in \Big[\!\Big[\max\big\{0, n-N+Np\big\}, \min\big\{Np, n\big\}\Big]\!\Big].$$

Pour un tel $k$, l'événement $\{X=k\}$ est réalisé quand on tire $k$ objets de type 1 parmi les $Np$ possibles ($\binom{Np}{k}$ possibilités) et $n-k$ de type 2 parmi les $N-Np$ restantes ($\binom{N-Np}{n-k}$ possibilités), donc : $P(X=k) = \frac{\big|\{X=k\}\big|}{|\Omega|} = \frac{\binom{Np}{k}\binom{N-Np}{n-k}}{\binom{N}{n}}$.

**Définition (Loi hypergéométrique ou loi des tirages SANS remise)** Soient $X$ une variable aléatoire et $N, n \in \mathbb{N}^*$ et $p \in [0,1]$ pour lesquels : $Np \in \mathbb{N}$.
On dit que $X$ *suit la loi hypergéométrique de paramètre* $(N,n,p)$, ce qu'on note : $X \hookrightarrow \mathscr{H}(N,n,p)$, si pour tout $k \in \mathbb{N}$ :

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

En réalité : $P(X=k) \neq 0$ si et seulement si $k$ appartient à $\left[\!\left[ \max\{0, n-N+Np\}, \min\{Np, n\} \right]\!\right]$.




Lorsqu'on tire $n$ objets simultanément dans un ensemble de $N$ objets dont $Np$ sont de type 1 et $N-Np$ de type 2, le **NOMBRE D'OBJETS DE TYPE 1 TIRÉS** suit la loi hypergéométrique $\mathscr{H}(N,n,p)$.

En pratique, la distinction de deux types distincts dans un même ensemble peut recouvrir des oppositions très variées telles que sain/malade, défectueux/non défectueux, fille/garçon, blanc/rouge, personnes ayant voté « oui »/personnes ayant voté « non », etc.

En outre, tout ceci définit bien une loi car : $\displaystyle\sum_{k=0}^{+\infty} \binom{Np}{k}\binom{N-Np}{n-k} = \binom{N}{n}$ — la somme est faussement infinie. En effet, dans l'anneau de polynômes $\mathbb{R}[T]$ : $(T+1)^{Np}(T+1)^{N-Np} = (T+1)^N$, ensuite il suffit ensuite d'identifier les coefficients de degré $n$ dans cette identité. Encore une généralisation de la formule de Vandermonde !

Enfin, l'expression « loi des tirages SANS remise » par laquelle on décrit parfois la loi hypergéométrique se comprend facilement, le tirage de $n$ objets simultanément est parfaitement équivalent au tirage de ces $n$ objets sans remise.

**Théorème (Espérance et variance de la loi hypergéométrique)** Soient $N, n \in \mathbb{N}^*$ et $p \in [0,1]$ pour lesquels : $Np \in \mathbb{N}$ et $X$ une variable aléatoire réelle de loi $\mathscr{H}(N,n,p)$. Alors : $\mathrm{E}(X) = np$ et $\mathrm{V}(X) = np(1-p)\dfrac{N-n}{N-1}$.

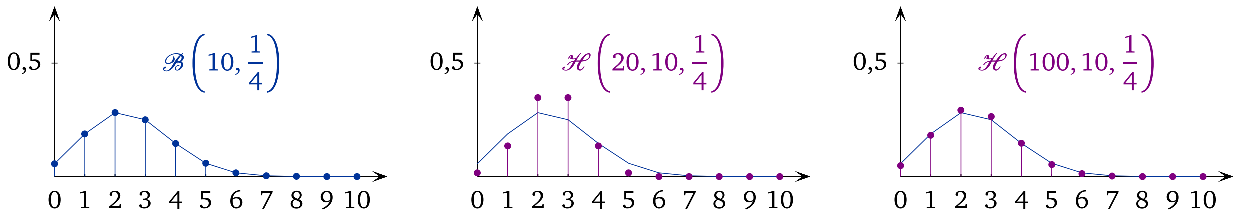
**Démonstration** Dans le calcul qui suit, toutes les sommes sont presque nulles.

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

$$\overset{i=k-1}{=} \frac{Np}{\binom{N}{n}} \sum_{i=0}^{+\infty} \binom{Np-1}{i}\binom{N-Np}{n-1-i} = \frac{Np}{\binom{N}{n}}\binom{N-1}{n-1} = Np \times \frac{n}{N} = np.$$

Nous admettrons le résultat pour la variance. ■

**Exemple (Approximation de la loi hypergéométrique par la loi binomiale)** Soit $X$ une variable aléatoire de loi hypergéométrique $\mathscr{H}(N,n,p)$ avec $N, n \in \mathbb{N}^*$ et $p \in ]0,1[$ et : $Np \in \mathbb{N}^*$. Plus $N$ est grand, plus la loi de $X$ est « proche » — en un sens que je laisse volontairement flou — de la loi binomiale $\mathscr{B}(n,p)$. Les figures ci-dessous sont à ce titre assez convaincantes.




**Démonstration** Tout d'abord, pour tout $k \in [\![0, n]\!]$ : $\displaystyle\binom{Np}{k} = \frac{(Np)(Np-1)\ldots(Np-k+1)}{k!} \underset{N\to+\infty}{\sim} \frac{(Np)^k}{k!}$,

et de même : $\displaystyle\binom{N-Np}{n-k} \underset{N\to+\infty}{\sim} \frac{(N-Np)^{n-k}}{(n-k)!} \underset{N\to+\infty}{\sim} \frac{\big(N(1-p)\big)^{n-k}}{(n-k)!}$ et $\displaystyle\binom{N}{n} \underset{N\to+\infty}{\sim} \frac{N^n}{n!}$.

En retour, par quotient : $\displaystyle\frac{\binom{Np}{k}\binom{N-Np}{n-k}}{\binom{N}{n}} \underset{N\to+\infty}{\sim} \binom{n}{k} p^k(1-p)^{n-k} \underset{N\to+\infty}{\longrightarrow} \binom{n}{k} p^k(1-p)^{n-k}$.

Conclusion : $P(X=k)$ tend donc vers $\displaystyle\binom{n}{k} p^k(1-p)^{n-k}$ pour tout $k \in [\![0, n]\!]$ lorsque $N$ tend vers $+\infty$, ce qui signifie que la loi hypergéométrique $\mathscr{H}(N, n, p)$ « tend » vers la loi binomiale $\mathscr{B}(n, p)$. On parle ici d'une *approximation de la loi hypergéométrique par la loi binomiale* et non le contraire, car la loi hypergéométrique est plus « compliquée » que la loi binomiale.

Quelle interprétation ? La loi hypergéométrique $\mathscr{H}(N, n, p)$ est la loi du nombre d'éléments de type 1 qu'on obtient en choisissant successivement **SANS** remise $n$ élements dans un ensemble de cardinal $N$ dont la proportion des éléments de type 1 est $p$. Or quand $N$ est grand devant $n$, le fait que les tirages soient effectués **SANS** remise compte assez peu car on a alors peu de chances de tirer plusieurs fois le même élément, c'est presque comme si les tirages étaient effectués **AVEC** remise selon la loi binomiale $\mathscr{B}(n, p)$.

**Exemple** Un institut de sondage cherche à anticiper le résultat d'un référendum en interrogeant un échantillon de $n$ personnes. Au moment du sondage, la proportion des « oui » dans la population complète est $p$ — valeur inconnue que l'institut essaie justement d'estimer. Si on note $\overline{p}$ la proportion des « oui » dans l'échantillon des $n$ personnes sondées, comment faut-il choisir $n$ pour que la probabilité de l'événement $\big\{|\overline{p}-p| \geqslant 0,05\big\}$ soit inférieure ou égale à 0,1 ?

**Démonstration** Notons $N$ le nombre total d'individus dans la population étudiée. À proprement parler, le nombre $n\overline{p}$ des « oui » collectés au cours du sondage suit la loi hypergéométrique $\mathscr{H}(N, n, p)$ ou loi des tirages **SANS** remise, mais quand on fait un sondage, l'échantillon $n$ est très petit devant la population totale $N$, si petit qu'en réalité la distinction **AVEC** ou **SANS** remise perd tout intérêt. Profitons donc de l'approximation au forceps de l'exemple précédent et faisons comme si la loi du nombre $n\overline{p}$ des « oui » était binomiale $\mathscr{B}(n, p)$.

L'inégalité de Bienaymé-Tchebychev donne aussitôt ceci :

$$P\big(|\overline{p}-p| \geqslant 0,05\big) = P\Big(\big|n\overline{p} - \mathrm{E}(n\overline{p})\big| \geqslant 0,05n\Big) \leqslant \frac{\mathrm{V}(n\overline{p})}{(0,05n)^2} = \frac{400p(1-p)}{n} \leqslant \frac{100}{n},$$

car pour tout $x \in \mathbb{R}$ : $x(1-x) \leqslant \dfrac{1}{4}$. À quelle condition a-t-on donc : $P\big(|\overline{p}-p| \geqslant 0,05\big) \leqslant 0,1$ ? Condition suffisante : $\dfrac{100}{n} \leqslant 0,1$, i.e. : $n \geqslant 1000$.