# Probabilité: CM6 - Variance et covariance

Toutes les variables aléatoires considérées sont de carré intégrable:$\mathbb{E}[X^2] < \infty$

> **Définition**:
> On note variance de la variable aléatoire $X$ :
>
> $$Var(X) = \mathbb{E}[(X - \mathbb{E}[X])^2]$$
>
> $$Var(X) = \mathbb{E}[X^2] - \mathbb{E}[X]^2$$
>
> et covariance des variables $X$ et $Y$:
>
> $$cov(X,Y) = \mathbb{E}[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])]$$
>
> $$cov(X,Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$$

> **Propriétés**:
> 1. $Var(X+Y) = Var(Y) + Var(Y) + 2cov(X,Y)$
> 2. Si $X$ et $Y$ sont indépendantes, $cov(X,Y) = 0$
> 3. $cov(X,Y)$ est symétrique et bilinéaire

> **Inégalité de Tchebyshev**:
>
> $$\forall t > 0, P(|X - \mathbb{E}[X]| > t) \leq \frac{Var(X)}{t^2}$$

> **Loi (faible) des grands nombres**:
> Soient $X_1, X_2, \ldots$ une suite de variable aléatoire de même loi, indépendantes et de carré intégrable ( $\mathbb{E}[X_i^2] < \infty$ ),
>
> $$\overline{X_n} = \frac{1}{n}\sum_{i=1}^{n} X_i$$
>
> $$Var(\overline{X_n}) = \frac{Var(X_1)}{n} \to 0$$
>
> ($\overline{X_n}$ converge vers $\mathbb{E}[X_1]$)

## Exercice 1

Soient $X, Y$ deux variables aléatoires telles que

$$Y = a + bX + \epsilon$$

1. $\mathbb{E}[\epsilon] = 0,\ \epsilon$ et $X$ sont indépendants
2. On suppose de plus que $Var(X) = 1$ (on divise $X$ par $\sqrt{Var(X)}$ de sorte que $X$ soit sans unité)

Calculer la covariance entre $X$ et $Y$:

**Solution**

$cov(X,Y) = cov(X, a + bX + \epsilon)$
$cov(X,Y) = cov(X,a) + b.cov(X,X) + cov(X,\epsilon)$
$cov(X,Y) = b.cov(X,X)$
$cov(X,Y) = b$

**Interprétation:**

La covariance est (la taille de) l'effet de X sur Y

## Exercice 2

Soit $Z$ une variable aléatoire de loi $\mathcal{N}(0,1)$

$$\forall z \in \mathbb{R}, f_Z(z) = \frac{1}{\sqrt{2\pi}} e^{\frac{-z^2}{2}}$$

On pose $X = m + \sigma Z$:

1. Espérance et variance de $Z$
2. Espérance et variance de $X$
3. Densité de la loi de $X$

**Solution**

1. $\mathbb{E}[Z] = 0$ et

$$Var(Z) = \mathbb{E}[Z^2] - \mathbb{E}[Z]^2 = 1$$

2. $\mathbb{E}[X] = \mathbb{E}[m + \sigma Z] = m + \sigma\mathbb{E}[Z] = m$ et

$$Var(X) = Var(m + \sigma Z) = Var(\sigma Z) = \sigma^2 Var(Z) = \sigma^2$$

3. $Z = \frac{X-m}{\sigma} = \frac{X - \mathbb{E}[X]}{\sqrt{Var(X)}}$ *(normalisation)*

$$P(X \leq t) = \int_{-\infty}^{t} f_X(x)dx$$

$$=P(Z < \quad \frac{t-m}{\sigma}) = \int_{-\infty}^{t} f_Z(\frac{x-m}{\sigma})\frac{dx}{\sigma}$$

La densité est

$$\forall x \in \mathbb{R}, f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-m)^2}{2\sigma^2}}$$

## Exercice 3

Soit $p$ une fonction de densité sur $\mathbb{R}$.
On appelle entropie de $p$ la fonction suivante:

$$H(p) = -\mathbb{E}[\ln(p(X))]$$

où $X$ suit la loi de densité p

1. Calculer $h(m, \sigma^2) = H(\mathcal{N}(m, \sigma^2))$
2. On admet l'inégalité de Gibbs,
   Soit $p, q$ deux densités sur $\mathbb{R}$, et X une va de loi de densité q

$$d_{KL}(q||p) = \mathbb{E}_q[\ln(\frac{q(X)}{p(X)})] \geq 0$$

Soit $q(x)$ une densité définie sur $\mathbb{R}$, de moyenne m et de variance $\sigma^2$, alors:

$$H(q) \leq h(m, \sigma^2)$$

Parmi toutes les lois sur $\mathbb{R}$ qui ont une variance $\sigma^2$, l'incertitude est maximisée pour la loi $\mathcal{N}(m, \sigma^2)$

1. Soit $p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-m)^2}{2\sigma^2}}$ alors,
   $H(p) = -\int \ln(p(x))p(x)dx$
   $H(p) = -\mathbb{E}_p[\ln(p(X)]$
   $H(p) = -\mathbb{E}_p[\ln(\frac{1}{\sqrt{2\pi\sigma^2}}) - \frac{(X-m)^2}{2\sigma^2}]$
   $H(p) = \frac{1}{2}[\ln(2\pi\sigma^2) + \mathbb{E}[Z^2]]$
   $H(p) = \frac{1}{2}[\ln(2\pi\sigma^2) + 1]$
   $\mathbb{E}_q[\ln(\frac{q(X)}{p(X)}] = -H(q) - \mathbb{E}_q[\ln(p(X))]$
   $\mathbb{E}_q[\ln(\frac{q(X)}{p(X)}] = -H(q) + h(m, \sigma^2)$

D'après l'inégalité de Gibbs

$$H(q) \leq h(m, \sigma^2)$$