Covariance, corrélation.

Définition 1

On appelle covariance de deux variables aléatoires finies X et Y, le réel

$$\text{Cov}(X,Y) = \mathbb{E}\left((X - \mathbb{E}(x))(Y - \mathbb{E}(Y))\right).$$

Remarques.

1. La covariance caractérise les variations simultanées de deux variables aléatoires : elle sera positive lorsque les écarts entre les variables et leurs moyennes ont tendance à être de même signe, négative dans le cas contraire.
2. Si la covariance est nulle alors les variables aléatoires sont dites non corrélées.

Proposition 1☣

$\text{Cov}(X,Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$.

Démonstration

$$\begin{aligned}\text{Cov}(X,Y) &= \mathbb{E}\left[(X - \mathbb{E}(X))(Y - \mathbb{E}(Y))\right]\\ &= \mathbb{E}\left[XY - X\mathbb{E}(Y) - \mathbb{E}(X)Y + \mathbb{E}(X)\mathbb{E}(Y)\right]\\ &= \mathbb{E}(XY) - \mathbb{E}(Y)\mathbb{E}[X] - \mathbb{E}(X)\mathbb{E}[Y] - \mathbb{E}(X)\mathbb{E}(Y)\\ &= \mathbb{E}(XY) - \mathbb{E}(X) - \mathbb{E}(Y)\end{aligned}$$

■

Définition 2

Le coefficient de corrélation de deux variables aléatoires réelles finies est le nombre

$$\rho(X,Y) = \frac{\text{Cov}(X,Y)}{\sqrt{V(X)V(Y)}}.$$

Remarques.

1. $\rho(X,Y) \in [-1;1]$.

Proposition 2 - Invariance d'échelle.

$\rho(aX + b, cY + d) = \rho(X,Y)$.