Intégration et Probabilités

S. Nonnenmacher
USTC
28/04/2026

Lois classiques de probabilité.

a) Si E est de cardinal n $\left(E=\{1,\dots,n\}\right)$,

la **loi uniforme** sur E est la loi telle que

$$\forall x\in E,\quad \mathbb{P}(X=x)=\frac{1}{n}.$$

Tous les points de E ont la même probabilité.

b) Loi de Bernoulli sur $E=\{0,1\}$ ou $E=\{\text{pile},\text{face}\}$

1 ☺

paramètre de la loi : $p\in[0,1]$.

$\mathbb{P}(X=\text{pile})=\mathbb{P}(X=1)=p,\quad \mathbb{P}(X=0)=1-p.$

C'est la loi uniforme si $p=\frac{1}{2}$, sinon cette loi n'est pas uniforme.

c) Loi binomiale $\mathcal{B}(n,p)$, $n\in\mathbb{N}^*$, $p\in[0,1]$.

$E=\{0,1,\dots,n\}$

$$\mathbb{P}(X=k) = \binom{n}{k} p^k (1-p)^{n-k}, \qquad 0 \leq k \leq n$$

Cette loi binomiale est reliée à la loi de Bernoulli :

si on tire n fois une loi de Bernoulli avec paramètre p,

$\mathbb{P}(X=k) =$ probabilité d'obtenir k fois "pile".

et donc $(n-k)$ fois "face".

$\Omega = \{0,1\}^n$ l'espace probabilisé pour n tirage de la pièce.

Parmi les 2^n suites possibles, compter combien de suites contiennent exactement p "pile".

d) Loi géométrique de paramètre $p \in]0,1[$.

v.a. à valeur dans $\mathbb{N}$, de loi :

$$\mathbb{P}(X=k) = (1-p)\,p^k \quad , \quad \forall k \in \mathbb{N}.$$

Lien avec la loi de Bernoulli : si on tire une infinité de fois une pièce de Bernoulli-p, la loi géométrique décrit le nombre de "pile" avant le premier "face".

Même question que pour le tirage du dé : combien de tirages faut-il effectuer avant d'obtenir le premier "face" ?

"géométrique" car la suite $(C p^k)_{k \geq 0}$ est appelée une suite géométrique de paramètre p.

e) Loi de Poisson de paramètre $\lambda \in \mathbb{R}_+^*$.

v.a. dans $\mathbb{N}$, de probabilité

$$\mathbb{P}(X=k) = \frac{\lambda^k}{k!} e^{-\lambda} \quad , \; k \in \mathbb{N}.$$

Cette loi décrit les "évènements rares" pendant un grand nombre de tirages (ou pendant une longue période).

Lien avec la loi de Bernoulli : suite de v.a. X_n de lois binomiales $\mathcal{B}(n, \underbrace{p_n = \frac{\lambda}{n}}_{\text{évènement rare}})$.

évènement rare : obtenir "pile" a, à chaque fois, la probabilité $\frac{\lambda}{n}$.

$$\lim_{n\to\infty} \mathbb{P}(X_n = k) = \lim_{n\to\infty} \binom{n}{k} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k}$$

$$\underset{\text{Stirling}}{=} \frac{\lambda^k}{k!} e^{-\lambda}$$

$$\frac{n!}{(n-k)!\,k!} \sim \frac{\left(n/e\right)^n \sqrt{\cancel{2\pi} n}}{\left(\frac{n-k}{e}\right)^{n-k} \sqrt{\cancel{2\pi}(n-k)}} \; \frac{1}{k!}$$

$$\sim \frac{1}{e^k} \frac{n^{n-k}}{(n-k)^{n-k}} \frac{\lambda^k}{\cancel{n^k}} \frac{\left(1-\frac{\lambda}{n}\right)^n}{\left(1-\frac{\lambda}{n}\right)^k}$$

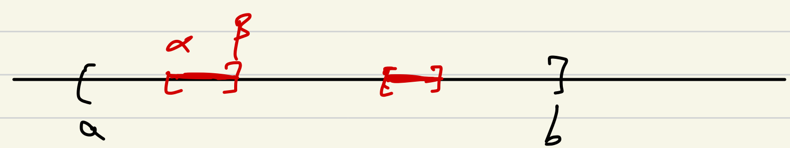
$$\sim \frac{1}{e^k} \frac{\left(1-\frac{k}{n}\right)^k}{\left(1-\frac{k}{n}\right)^n} \qquad \frac{\lambda^k}{k!} e^{-\lambda}$$

$$\sim \cancel{e^{-k}} \cdot \cancel{e^{k}}$$

Lois de variables continues (ex : sur $\mathbb{R}$)

a) Loi <u>uniforme sur un intervalle</u> borné $[a,b]$: loi de densité

$$p(x) = \frac{1}{b-a} \mathbb{1}_{[a,b]}(x) .$$



$$\mathbb{P}([\alpha,\beta]) = \frac{\beta-\alpha}{b-a}$$ ne dépend que de la longueur de $[\alpha,\beta]$, et non de ~~sa~~ position dans $[a,b]$.

b) Loi exponentielle de paramètre $\lambda \in \mathbb{R}_+^*$: loi de densité $p(x) = \lambda e^{-\lambda x}$ sur $\mathbb{R}_+$.

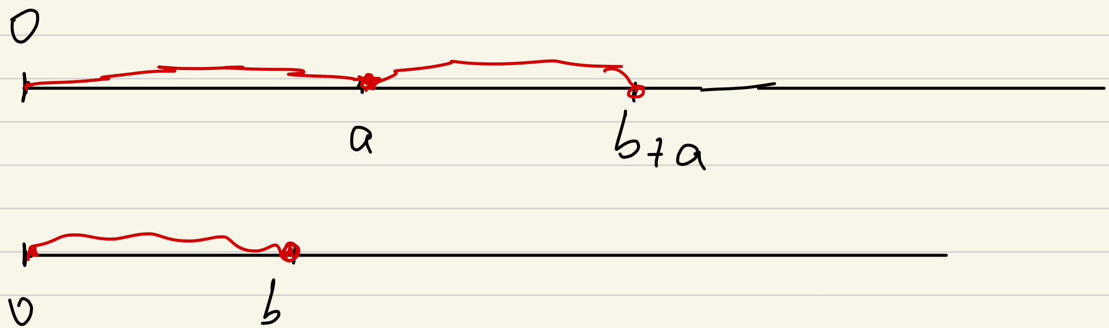
$$= \lambda e^{-\lambda x} \mathbb{1}_{\mathbb{R}_+}(x) \quad \text{sur } \mathbb{R}$$

Propriété intéressante de cette loi : $\forall a, b > 0$,

$$\mathbb{P}(X > a+b) = \mathbb{P}(X > a) . \mathbb{P}(X > b).$$

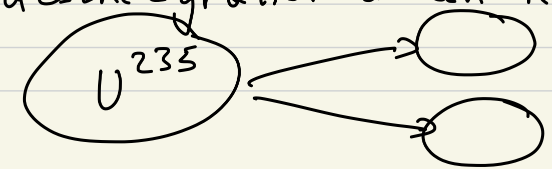
$$\int_{a+b}^{+\infty} p(x)\,dx = e^{-\lambda(a+b)} = e^{-\lambda a} . e^{-\lambda b}$$

Interprétation : pour que X dépasse $(a+b)$, il faut d'abord dépasser a, ce qui est fait avec proba $e^{-\lambda a}$, puis il faut encore franchir une distance b pour atteindre $a+b$. Cette 2^e étape, est indépendante de la 1ère, et elle a la même probabilité que pour aller de 0 à b.



Propriété de factorisation reflète l'<u>indépendance</u> des 2 parties du chemin. <u>Absence de mémoire</u> de la particule.

Exemple d'application en physique atomique : processus de désintégration d'un noyau atomique radioactif.



désintégration au bout d'1 temps aléatoire.

La loi de ce temps de désintégration est une loi exponentielle. Le paramètre λ dépend du type de noyau considéré. $\frac{1}{\lambda}$ = temps de demi-vie de l'atome.

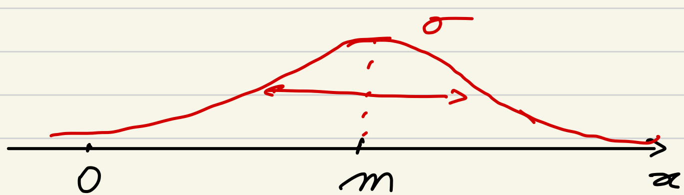
c) Loi gaussienne, ou loi normale $\mathcal{N}(m, \sigma^2)$.

. $m \in \mathbb{R}$ le centre de la gaussienne

. $\sigma \in \mathbb{R}_+$ l'écart-type de la gaussienne.

v.a. sur $\mathbb{R}$, de densité

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-m)^2}{2\sigma^2}}$$



$$m = \mathbb{E}[X] = C\int x\, e^{-\frac{(x-m)^2}{2\sigma^2}}\,dx = C\int (y+m)\, e^{-\frac{y^2}{2\sigma^2}}\,dy$$

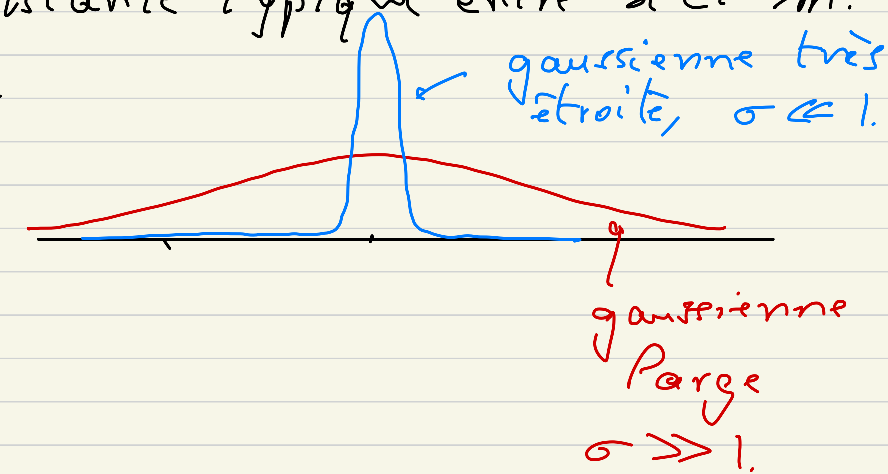
$y = x - m$

$$= \underbrace{C\int y\, e^{-y^2/2\sigma^2}\,dy}_{0} + m\,\underbrace{C\int e^{-\frac{y^2}{2\sigma^2}}\,dy}_{1}$$

$\sigma^2 = \mathbb{E}[(X-m)^2]$ la variance de la loi.

$\sigma = \sqrt{\mathbb{E}[(X-m)^2]}$ distance typique entre x et m.

$=$ Ecart-type.



Si X suit la loi $\mathcal{N}(m, \sigma^2)$, $(X-m)$ suit la <u>loi centrée</u> $\mathcal{N}(0, \sigma^2)$.

La v.a. $\dfrac{X-m}{\sigma}$ suit la loi $\mathcal{N}(0,1)$ = la loi normale <u>centrée réduite</u>

- Si X suit la loi $\mathcal{N}(m, \sigma^2)$, alors la v.a. $Y = \lambda X + \mu$ suit la v.a. normale $\mathcal{N}(\lambda m + \mu, \lambda^2 \sigma^2)$.
- La loi normale est très importante : le théorème central limite décrit la <u>somme</u> de N variables

aléatoires indépendantes et identiques.

- Fonction de répartition d'1 v.a. réelle X

$$F_X(t) = \mathbb{P}(X \leq t) = P_X(]-\infty, t]) , \quad \forall t \in \mathbb{R}.$$

$$F_X(t) \xrightarrow[t \to +\infty]{} 1 , \quad \text{cf. loi de probabilité.}$$

- F fonction $F \xrightarrow[t \to -\infty]{} 0$, $\uparrow$, continue à droite

$$F \xrightarrow[t \to +\infty]{} 1$$

$\longrightarrow$ correspond une unique loi de probab. sur $\mathbb{R}$.

$$\mathbb{P}(a \leq X \leq b) = F_X(b) - F_X(a-)$$

$$\mathbb{P}(a < X < b) = F_X(b_-) - F_X(a).$$

Points de discontinuité de F_X = atomes de P_X.

- Tribu engendrée par une v. a.

Ω, $\mathcal{A} = ?$ Pour choisir la tribu $\mathcal{A}$, on demandait que certaines fonctions $f : \Omega \to \mathbb{R}$ soient mesurables.

Si $X : \Omega \to (E, \mathcal{E})$, la tribu engendrée par X, notée $\sigma(X)$, est la plus petite tribu telle que X soit mesurable entre $(\Omega, \sigma(X))$ et $(E, \mathcal{E})$.

On trouve que

$$\sigma(X) = \underbrace{\{A = X^{-1}(B),\ B \in \mathcal{E}\}}_{\text{une tribu.}}.$$

- $(X_i)_{i \in I}$ famille de fonctions $\Omega \longrightarrow (E_i, \mathcal{E}_i)$, alors on peut définir la tribu engendrée par la famille de ces fonctions :

$$\mathcal{F} = \sigma\left((X_i)_{i \in I}\right) = \sigma\left(\{\underbrace{X_i^{-1}(B)}_{\subset \Omega} \; ; \; B \in \mathcal{E}_i \, , \; \forall i \in I\}\right)$$

La plus petite tribu telle que tous les X_i soient mesurables $\mathcal{F} \longrightarrow \mathcal{E}_i$.

- <u>Proposition</u> Soit X une une v.a. à valeur dans $(E, \mathcal{E})$, et soit Y une v.a. réelle. On a alors équivalence entre :

i) la v.a. Y est $\sigma(X)$-mesurable.

$\iff$ $\sigma(Y)$ est une sous-tribu de $\sigma(X)$.

ii) il existe une fonction $f : (E, \mathcal{E}) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ mesurable,

telle que $\boxed{Y = f(X).}$

$$\begin{tikzcd}
\Omega \arrow[r, "X"] \arrow[dr, "Y"'] & E \arrow[d, "f"] \\
& \mathbb{R}
\end{tikzcd}$$

On dit alors que Y est totalement dépendante de X.

"si je connais la valeur de X, alors la valeur de Y est fixée.

Preuve : ii) $\Rightarrow$ i) si $Y = f \circ X$, alors, $\forall A \in \mathcal{B}(\mathbb{R})$,

$$\sigma(Y) \ni Y^{-1}(A) = (f \circ X)^{-1}(A) = X^{-1}\big(\underbrace{f^{-1}(A)}_{\in \mathcal{E}}\big) \in \sigma(X).$$

i) $\Rightarrow$ ii) Supposons que Y est $\sigma(X)$-mesurable.

1^{er} cas : si Y est étagée : $Y = \sum_{i=1}^{n} \lambda_i \mathbb{1}_{A_i}$, avec

$A_i \in \sigma(X)$ disjoints 2 à 2.

$\Rightarrow \exists B_i \in \mathcal{E}$ tq $A_i = X^{-1}(B_i)$, B_i disjoints 2 à 2

$$\Rightarrow Y = \sum_{i=1}^{n} \lambda_i \mathbb{1}_{X^{-1}(B_i)} = \sum_{i=1}^{n} \lambda_i \mathbb{1}_{B_i} \circ X$$

$$= \underbrace{\left(\sum_{i=1}^{n} \lambda_i \mathbb{1}_{B_i}\right)}_{f \,:\, E_{\mathcal{E}} \to \mathbb{R}_{\mathcal{B}(\mathbb{R})} \text{ mesurable.}} \circ X$$

- Y une fonction quelconque, $\sigma(X)$-mesurable.

$Y = Y^+ - Y^-$,

$Y^+ = \lim_{n \to \infty} Y_n \longrightarrow$ fonctions étagées ≥ 0 $\sigma(X)$-mesurables

Chaque $Y_n = f_n \circ X$ pour une fonction étagée ≥ 0 f_n, mesurable $\mathcal{E} \to \mathcal{B}(\mathbb{R})$.

$\forall \omega \in \Omega$, $Y_n(\omega) \longrightarrow Y(\omega)$ simplement $\Longrightarrow$

$$f_n(\underbrace{X(\omega)}_{x \in \mathrm{Ran}(X) \subset E}) \longrightarrow Y(\omega).$$

$\Rightarrow \forall x \in \mathrm{Ran}(X)$, $f_n(x)$ a une limite.

$\Rightarrow$ définissons la fonction f sur E, tq.

est mesurable $\leftarrow$ $f(x) = \begin{cases} \lim_{n\to\infty} f_n(x) & \text{si la limite existe} \\ 0 & \text{sinon.} \end{cases}$

On sait que la limite existe si $x \in \mathrm{Ran}(X)$.

$\Rightarrow \forall \omega \in \Omega$, $f(X(\omega)) = \lim_{n\to\infty} f_n(X(\omega)) = Y(\omega)$.

$\Rightarrow Y = f \circ X$.

$\square$.

- TCM : $X_n \geq 0, \quad X_n \uparrow X \implies \mathbb{E}[X_n] \uparrow \mathbb{E}[X]$.

Fatou : $X_n \geq 0 \implies \mathbb{E}\left[\liminf_n X_n\right] \leq \liminf_n \mathbb{E}[X_n]$

TCD : $|X_n| \leq Z$ et $\mathbb{E}[Z] < 0, \quad X_n \longrightarrow X$ p.p.

$$\implies \mathbb{E}[X_n] \longrightarrow \mathbb{E}[X].$$

- Une v.a. réelle X est centrée si elle est intégrable ($\mathbb{E}[|X|] < \infty$) et $\mathbb{E}[X] = 0$.

- En théorie des probabilités, "presque partout" se dit "presque sûrement" (p.s.) = "avec probabilité 1".

$\leq$ de Hölder : $\mathbb{E}[|XY|] \leq \underbrace{\mathbb{E}[|X|^p]^{1/p}}_{\|X\|_{L^p}} \underbrace{\mathbb{E}[|Y|^q]^{1/q}}_{\|Y\|_{L^q}}$

q, p conjuguées.

- $\|X\|_r \leq \|X\|_p$ si $p \geq r$

- Cauchy-Schwarz : $\mathbb{E}[|XY|] \leq \mathbb{E}[X^2]^{1/2} \, \mathbb{E}[Y^2]^{1/2}$

$(Y=1) \qquad \Longrightarrow \mathbb{E}[|X|]^2 \leq \mathbb{E}[X^2]$.

- <u>Variance et matrice de covariance.</u>

<u>Déf :</u> Soit $X \in L^2(\Omega, \mathbb{P})$ une v.a. réelle. On définit sa variance par :

$$\mathrm{var}(X) = \mathbb{E}\Big[(\underbrace{X - \mathbb{E}[X]}_{\text{v.a. centrée}})^2\Big] < \infty$$

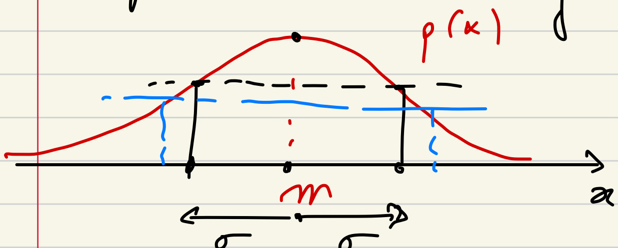
et son écart-type $\sigma_X = \sqrt{\mathrm{var}(X)}$.

largeur de la distribution de X

dispersion de X

Si $\mathrm{Var}(X) = 0 \Rightarrow X = E[X]$ p.s.

. Si X suit la loi normale $\mathcal{N}(m, \sigma^2)$, alors $\sigma_X = \sigma$ représente la "largeur à mi-hauteur" de la gaussienne :



$$|x-m| \le \underset{\sigma\sqrt{2\ln 2}}{\sigma} \iff p(x) \ge \underbrace{e^{-\frac{1}{2}}}_{\substack{1/2 \\ 0{,}606\ldots}} \cdot p_{max}$$

Proposition : $X \in L^2(\Omega, \mathbb{P})$. Alors

$$\mathrm{var}(X) = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 .$$

$\mathbb{E}[X^2]$: moment d'ordre 2 de la v.a. X.

$$\forall a \in \mathbb{R},\quad \mathbb{E}\left[(X-a)^2\right] = \mathrm{var}(X) + \underbrace{(\mathbb{E}[X]-a)^2}_{\geq 0}$$

$$\Rightarrow \quad \mathrm{var}(X) = \inf_{a\in\mathbb{R}} \mathbb{E}\left[(X-a)^2\right],$$ avec le minimum atteint pour $a = \mathbb{E}[X]$.

Preuve : développer les carrés

$$\mathbb{E}\left[(X-a)^2\right] = \mathbb{E}\left[X^2 - 2aX + a^2\right]$$

$$= \mathbb{E}[X^2] - 2a\,\mathbb{E}[X] + a^2\,\mathbb{E}[1]$$

$$= \underbrace{\mathbb{E}[X^2] - \mathbb{E}[X]^2}_{\mathrm{var}(X)} + (\mathbb{E}[X]-a)^2$$

- Inégalités :

  - Inégalité de Markov : X v.a. ≥ 0, $\forall a > 0$

$$\Rightarrow \mathbb{P}(X \geq a) \leq \frac{1}{a} \mathbb{E}[X].$$

  - Inégalité de Bienaymé-Chebychev :

    si $X \in L^2(\Omega, \mathbb{P})$, $a > 0$,

$$\mathbb{P}(|X - \mathbb{E}[X]| \geq a) \leq \frac{1}{a^2} \operatorname{var}(X).$$

pr : utiliser Markov pour la v.a. $Y = (X - \mathbb{E}[X])^2$

Ces $\leq$ permettent de borner les probabilités de grandes déviations en terme d'espérances.

- Covariance entre plusieurs v.a. réelles.

<u>Déf</u>: Soit $X, Y \in L^2(\Omega, \mathbb{P})$. La covariance de X et Y est définie par :

$$\mathrm{cov}(X,Y) = \mathbb{E}\Big[\underbrace{(X - \mathbb{E}[X])}_{\text{v.a. centrée}}(Y - \mathbb{E}[Y])\Big]$$

$$= \mathbb{E}\left[X(Y - \mathbb{E}[Y])\right]$$

$$= \mathbb{E}\left[(X - \mathbb{E}[X])\,Y\right]$$

$$= \mathbb{E}[XY] - \mathbb{E}[X]\,\mathbb{E}[Y]$$

Soit un vecteur aléatoire $X = (X_1, X_2, \dots X_d)$ v.a. à valeurs dans $\mathbb{R}^d$.

On suppose que $X_i \in L^2(\Omega, \mathbb{P})$, $\forall i = 1, \dots, d$.

On définit alors la matrice de covariance de X :

$$K_X = \left(\operatorname{cov}(X_i, X_j) \right)_{i,j=1,\dots,d}$$

Propriétés de $\operatorname{cov}(X,Y)$:

1) Si $X = Y$, $\operatorname{cov}(X,Y) = \operatorname{var}(X) \geq 0$.

Par contre, si $X \neq Y$, on peut avoir $\operatorname{cov}(X,Y) \leq 0$.

ex : $Y = -X \Rightarrow \operatorname{cov}(X,-X) = -\operatorname{var}(X) \leq 0$.

2) Cauchy-Schwarz : $|\operatorname{cov}(X,Y)| \leq \sqrt{\operatorname{var}(X)}\sqrt{\operatorname{var}(Y)}$

3) $(X,Y) \longmapsto \operatorname{cov}(X,Y)$ est une forme bilinéaire symétrique sur $L^2(\Omega)$.

4) Pour $X = (X_1, \dots, X_d)$, la matrice de covariance K_X est symétrique positive :

$$(K_X)_{ij} = (K_X)_{ji}$$

$\forall \lambda = (\lambda_1, \dots, \lambda_d) \in \mathbb{R}^d$,

$$\langle \lambda, K_X \lambda \rangle = \sum_{i,j=1}^{d} \lambda_i \lambda_j (K_X)_{ij} = \operatorname{var}\left(\sum_{i=1}^{d} \lambda_i X_i\right) \geq 0$$

Exercice : si A matrice $n \times d$, et $X' = AX$ une v.a. dans $\mathbb{R}^n$, alors la matrice de covariance

sur $\mathbb{R}^d$

$$\underbrace{K_{X'}}_{n \times n} = \underbrace{A}_{n \times d} \underbrace{K_X}_{d \times d} \underbrace{{}^tA}_{d \times n}$$

La covariance donne une mesure de la dépendance, ou la corrélation, entre les v.a. X et Y.

Si $Y \simeq X \Rightarrow \mathrm{cor}(X,Y) \geq 0$

$Y \approx -X \Rightarrow$ " " ≤ 0

Si Y et X sont indépendantes $\Rightarrow \mathrm{cor}(X,Y) = 0$.