# Cas particulier de la loi gaussienne

**G. Perrin**

guillaume.perrin@univ-eiffel.fr

Année 2022-2023






<

# Plan de la séance

# Partie 1 : introduction

- Lors des séances précédentes, nous avons vu qu'on pouvait caractériser complètement un vecteur aléatoire $\boldsymbol{X}$ à partir :
  - de ses CDFs marginales, $F_{X_i}$,
  - d'une fonction copule $C$ caractérisant les dépendances entre les différentes composantes de $\boldsymbol{X}$.
- Pour la modélisation des fonctions $F_{X_i}$ et $C$, on distingue :
  - les approches **paramétriques** : $F_{X_i}$ et $C$ sont ainsi supposées appartenir à des ensembles paramétrés de fonctions. Les caractériser revient à identifier au mieux les paramètres de ces lois à partir de l'information disponible.
  - les approches **non-paramétriques** : $F_{X_i}$ et $C$ sont directement construits à partir de l'information disponible sur $\boldsymbol{X}$ (théorie de l'information, construction à noyaux,...)

Aujourd'hui, on va s'attarder sur une représentation particulière bien connue : la distribution gaussienne !

# Partie 1 : introduction

## Gauss et la loi gaussienne

- L'origine de la loi gaussienne date de Bernoulli (1654-1705), en travaillant sur le pari basé sur un jeu de pile ou face, le menant à la loi des grands nombres.
- Laplace (1747-1829) poursuit ces travaux en travaillant sur l'estimation d'erreurs (en astronomie notamment).
- Gauss (1777-1855) étudie également cette loi dans le cadre d'estimation moindres carrés et de minimisation d'erreur. La loi gaussienne apparaît alors comme densité permettant de garantir que l'estimateur du maximum de vraisemblance à partir d'une série de mesures indépendantes est la moyenne empirique.
- C'est finalement Laplace qui publie en 1812 le théorème centrale limite.

# Partie 2 : présentation de la loi gaussienne

## Définition

Un vecteur aléatoire $\boldsymbol{X}$ est dit gaussien ssi sa PDF $f_{\boldsymbol{X}}$ s'écrit sous la forme :

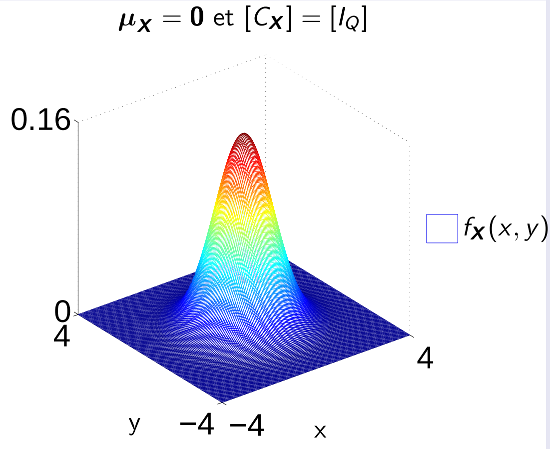
$$f_{\boldsymbol{X}}(\boldsymbol{x}) = \frac{1}{\sqrt{(2\pi)^d \det[C_{\boldsymbol{X}}]}} \exp\left\{ -\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu}_{\boldsymbol{X}})^T [C_{\boldsymbol{X}}]^{-1} (\boldsymbol{x} - \boldsymbol{\mu}_{\boldsymbol{X}}) \right\},$$

$$\mathbb{E}\left[\boldsymbol{X}\right] = \boldsymbol{\mu}_{\boldsymbol{X}}, \quad \mathbb{E}\left[(\boldsymbol{X} - \mathbb{E}\left[\boldsymbol{X}\right]) \otimes (\boldsymbol{X} - \mathbb{E}\left[\boldsymbol{X}\right])\right] = [C_{\boldsymbol{X}}].$$

- Supposer que $\boldsymbol{X}$ est gaussien revient à dire que sa distribution est complètement caractérisée par sa moyenne et sa matrice de covariance (que l'on supposera toujours inversible).
- Pour caractériser un vecteur gaussien, il suffit d'identifier sa moyenne et sa matrice de covariance.

# Partie 2 : présentation de la loi gaussienne

## Visualisation graphique 2D



# Partie 2 : présentation de la loi gaussienne

## Copule gaussien (rappel)

Si $\boldsymbol{X}$ est un vecteur aléatoire gaussien de moyenne $\boldsymbol{\mu}$ et de matrice de covariance $[R]$, alors son copule est défini par :

$$C(u_1, \ldots, u_d) = \Phi\left(\phi^{-1}(u_1), \ldots, \phi^{-1}(u_d)\right),$$

$$\phi(x) = \int_{-\infty}^{x_i} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{y^2}{2}\right) dy,$$

$$\Phi(\boldsymbol{x}) = \frac{1}{(2\pi)^{d/2}\sqrt{\det([R])}} \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_d} \exp\left(-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T [R]^{-1} (\boldsymbol{x}-\boldsymbol{\mu})\right).$$

- le copule gaussien est non-explicite.
- la loi gaussienne est un cas particulier de loi paramétrique où l'expression du copule est plus compliquée que la loi elle-même.

# Partie 2 : présentation de la loi gaussienne

## Définition

La **fonction caractéristique** de $\boldsymbol{X}$ est la fonction $\boldsymbol{t} \mapsto \Phi_{\boldsymbol{X}}(\boldsymbol{t})$ telle que :

$$\Phi_{\boldsymbol{X}}(\boldsymbol{t}) = \mathbb{E}\left[\exp\left(i\left\langle \boldsymbol{t}, \boldsymbol{X}\right\rangle\right)\right] = \int_{\mathbb{R}^d} \mathrm{e}^{i\langle \boldsymbol{t}, \boldsymbol{X}\rangle} P_{\boldsymbol{X}}(d\boldsymbol{x}).$$

Si $\boldsymbol{X}$ admet une densité $f_{\boldsymbol{X}}$, et si $\Phi_{\boldsymbol{X}}$ est intégrable sur $\mathbb{R}^d$, alors :

$$f_{\boldsymbol{X}}(\boldsymbol{x}) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \mathrm{e}^{-i\langle \boldsymbol{t}, \boldsymbol{X}\rangle} \Phi_{\boldsymbol{X}}(\boldsymbol{t}) d\boldsymbol{t}.$$

## Propriété

Si $\boldsymbol{X}$ est d'ordre $p$, alors pour tout $\boldsymbol{\alpha}$ tel que $|\boldsymbol{\alpha}| = \alpha_1 + \ldots + \alpha_d \leq p$,

$$\left\{ \frac{\partial^{\alpha_1}}{\partial t_1^{\alpha_1}} \times \cdots \times \frac{\partial^{\alpha_d}}{\partial t_d^{\alpha_d}} \Phi_{\boldsymbol{X}}(\boldsymbol{t}) \right\}_{\boldsymbol{t}=0} = m_{\boldsymbol{\alpha}} i^{|\boldsymbol{\alpha}|}.$$

# Partie 2 : présentation de la loi gaussienne

Exercice :

1. Montrer le résultat de la propriété précédente.
2. A partir de la propriété précédente, indiquer à quelle distribution est associée la fonction caractéristique $t \mapsto \exp(iat)$.
3. A partir de la définition précédente, indiquer à quelle distribution est associée la fonction caractéristique $t \mapsto \exp(iat + (ibt)^2/2)$.

# Partie 3 : propriétés remarquables

## Stabilité par transformation affine

Si $\boldsymbol{X}$ est un vecteur gaussien de dimension $d$, de moyenne $\boldsymbol{\mu_X}$ et de matrice de covariance $[C_{\boldsymbol{X}}]$, alors pour toute matrice $[M]$ de dimension $(n, d)$ et tout vecteur $\boldsymbol{z}$ de dimension $n$, alors $\boldsymbol{Y} = [M]\boldsymbol{X} + \boldsymbol{z}$ est également un vecteur gaussien, et on a :

$$\mathbb{E}[\boldsymbol{Y}] = [M]\boldsymbol{\mu_X} + \boldsymbol{z}, \quad [Cov(\boldsymbol{Y})] = [M][C_{\boldsymbol{X}}][M]^T.$$

Exercice : en remarquant l'invariance par rotation d'un vecteur gaussien centré de composantes indépendantes, expliquer comment générer des réalisations indépendantes dont la distribution est uniformément répartie sur une sphère de rayon $r$.

# Partie 3 : propriétés remarquables

Preuve.

**Lemme**

Si $X_1, \ldots, X_d$ sont $d$ variables aléatoires indépendantes de fonctions caractéristiques respectives $\Phi_i$, alors la fonction caractéristique $\Phi_Z$ de la variable aléatoire $Z = X_1 + \ldots + X_d$ est égale à :

$$\Phi_Z = \prod_{i=1}^{d} \Phi_i.$$

Exercice :

- Montrer le lemme précédent.
- Déduire du lemme précédent que la somme de variables gaussiennes indépendantes est une variable gaussienne.
- En déduire ensuite que la somme de variables gaussiennes quelconques reste une variable aléatoire gaussienne.
- Conclure sur la propriété précédente.

# Partie 3 : propriétés remarquables

## Génération de réalisations indépendantes

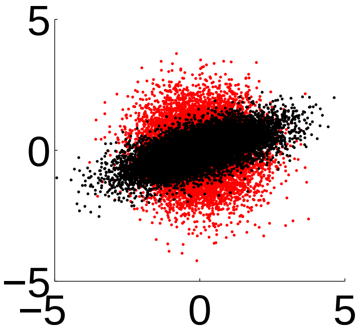
Si :

- $\boldsymbol{X}$ est un vecteur gaussien de moyenne $\boldsymbol{\mu_X}$ et de matrice de covariance $[C_{\boldsymbol{X}}]$,
- $\boldsymbol{\xi}$ est un vecteur gaussien centré, dont toutes les composantes sont de variance 1 et sont indépendantes,
- $[R]$ est une matrice symétrique telle que $[R][R] = [C_{\boldsymbol{X}}]$,

alors, comme conséquence de la propriété précédente, on peut montrer que $\boldsymbol{X}$ et $[R]\boldsymbol{\xi} + \boldsymbol{\mu_X}$ ont la même distribution.

$\Rightarrow$ c'est le moyen le plus utilisé pour générer des réalisations indépendantes d'un vecteur gaussien quelconque à partir d'un générateur de variable gaussienne centrée réduite.

# Partie 3 : propriétés remarquables




Exercice : sur cette figure ont été tracées $10^4$ réalisations indépendantes de deux vecteurs gaussiens centrés $\boldsymbol{X}$ bidimensionnels. L'un tel que $\mathbb{E}\left[X_1X_2\right] = 0$ et l'autre tel que $\mathbb{E}\left[X_1X_2\right] = 0.8$. Associer les couleurs aux vecteurs aléatoires correspondants.

# Partie 3 : propriétés remarquables

## Stabilité par conditionnement

Soient $\boldsymbol{X}$ et $\boldsymbol{Y}$ deux vecteurs gaussiens de tailles $n$ et $p$, de moyennes $\boldsymbol{\mu_X}$ et $\boldsymbol{\mu_Y}$, de covariances $[C_{\boldsymbol{X}}]$ et $[C_{\boldsymbol{Y}}]$ respectivement, et tels que
$\mathbb{E}\left[(\boldsymbol{X}-\boldsymbol{\mu_X})\otimes(\boldsymbol{Y}-\boldsymbol{\mu_Y})\right]=[C_{\boldsymbol{XY}}]$.
Alors la loi de $\boldsymbol{Y}$ **conditionnellement** à $\boldsymbol{X}$ est également gaussienne :

$$(\boldsymbol{Y}\mid\boldsymbol{X}=\boldsymbol{x})\ \sim\ \mathcal{N}(\boldsymbol{\mu}^{\mathrm{Cond}}(\boldsymbol{x}),[C^{\mathrm{Cond}}(\boldsymbol{x})]),$$

$$\begin{cases}\boldsymbol{\mu}^{\mathrm{Cond}}(\boldsymbol{x})=\mathbb{E}\left[\boldsymbol{Y}\mid\boldsymbol{X}=\boldsymbol{x}\right]=\boldsymbol{\mu_Y}+[C_{\boldsymbol{XY}}][C_{\boldsymbol{X}}]^{-1}\left(\boldsymbol{x}-\boldsymbol{\mu_X}\right),\\ [C^{\mathrm{Cond}}(\boldsymbol{x})]=[C_{\boldsymbol{Y}}]-[C_{\boldsymbol{XY}}][C_{\boldsymbol{X}}]^{-1}[C_{\boldsymbol{XY}}]^T.\end{cases}$$

Exercices :

1. Montrer la propriété précédente. On rappelle que :

$$\begin{bmatrix}A & B\\ C & D\end{bmatrix}^{-1}=\begin{bmatrix}A^{-1}+A^{-1}B(D-CA^{-1}B)^{-1}CA^{-1} & -A^{-1}B(D-CA^{-1}B)^{-1}\\ -(D-CA^{-1}B)^{-1}CA^{-1} & (D-CA^{-1}B)^{-1}\end{bmatrix}.$$

2. On pose $X\sim\mathcal{N}(0,1)$, $Y\sim\mathcal{N}(0,1)$, $\mathbb{E}\left[XY\right]=\rho\leq 1$. Quantifier l'influence de l'observation de $X=x$ sur la distribution de $Y$.

# Partie 3 : propriétés remarquables

Illustration du conditionnement gaussien : $X \sim \mathcal{N}(0,1)$, $Y \sim \mathcal{N}(0,1)$, $-1 \leq \mathbb{E}[XY] = \rho \leq 1 \Rightarrow (Y \mid X = x) \sim \mathcal{N}(\rho x, 1-\rho^2)$.

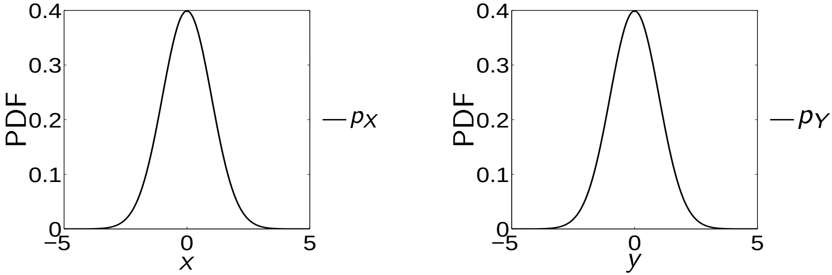



FIGURE: Distributions initiales

Le conditionnement modifie la moyenne et réduit la variance !

# Partie 3 : propriétés remarquables

Illustration du conditionnement gaussien : $X \sim \mathcal{N}(0,1)$, $Y \sim \mathcal{N}(0,1)$, $-1 \leq \mathbb{E}[XY] = \rho \leq 1 \Rightarrow (Y \mid X = x) \sim \mathcal{N}(\rho x, 1-\rho^2)$.

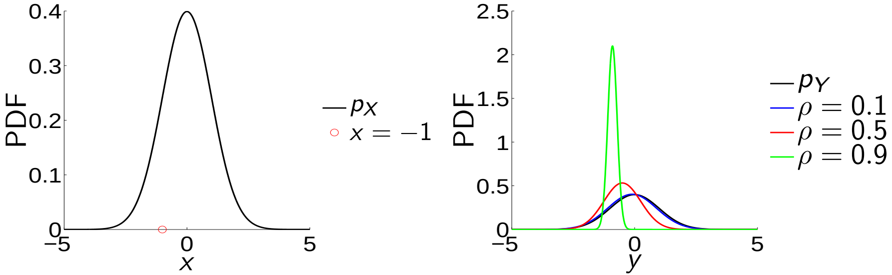



FIGURE: On conditionne $Y$ par le fait que $X = -1$

Le conditionnement modifie la moyenne et réduit la variance !

# Partie 3 : propriétés remarquables

Illustration du conditionnement gaussien : $X \sim \mathcal{N}(0,1)$, $Y \sim \mathcal{N}(0,1)$, $-1 \leq \mathbb{E}[XY] = \rho \leq 1 \Rightarrow (Y \mid X = x) \sim \mathcal{N}(\rho x, 1-\rho^2)$.

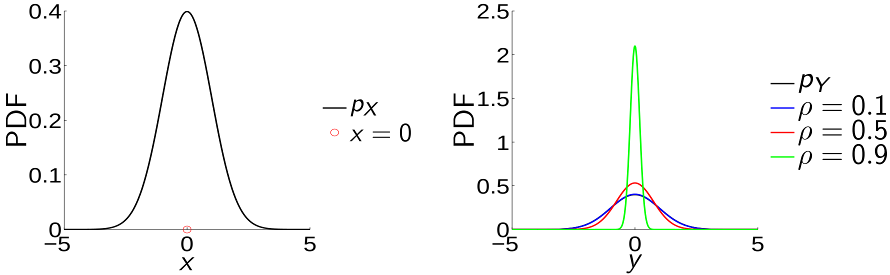



FIGURE: On conditionne $Y$ par le fait que $X = 0$

Le conditionnement modifie la moyenne et réduit la variance !

# Partie 3 : propriétés remarquables

Illustration du conditionnement gaussien : $X \sim \mathcal{N}(0,1)$, $Y \sim \mathcal{N}(0,1)$, $-1 \leq \mathbb{E}[XY] = \rho \leq 1 \Rightarrow (Y \mid X = x) \sim \mathcal{N}(\rho x, 1-\rho^2)$.

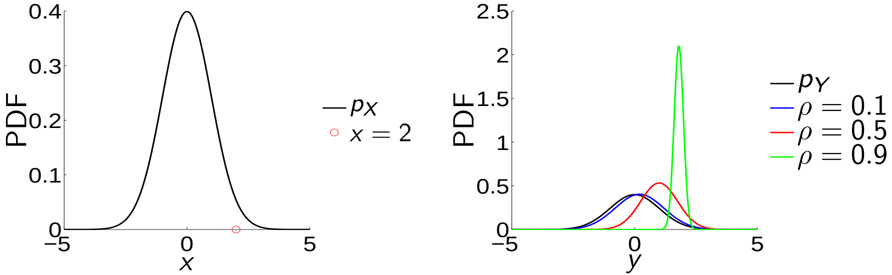



FIGURE: On conditionne $Y$ par le fait que $X = 2$

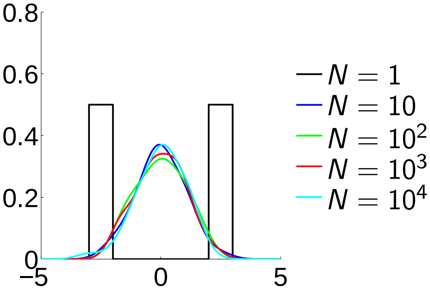
Le conditionnement modifie la moyenne et réduit la variance !

# Partie 3 : propriétés remarquables

## La loi gaussienne comme limite du théorème centrale limite

Si $X_1, \ldots, X_N$ sont $N$ v.a. indépendantes de même loi, et de moyenne $\mu$ et d'écart type $\sigma$, alors :

$$\frac{1}{N}\sum_{n=1}^{N} X_n \rightsquigarrow \mathcal{N}(\mu, \sigma^2/N).$$



# Partie 3 : propriétés remarquables

## Les polynômes de Hermite

Soit $t \mapsto H_n(t)$ les polynômes tels que :

$$\widetilde{H}_n(t) = (-1)^n \exp(x^2/2) \frac{d^n \exp(-x^2/2)}{dx^n}, \;\; H_n = \widetilde{H}_n / \left\| \widetilde{H}_n \right\|.$$

Les polynômes $\{H_n, \;\; n \geq 0\}$ sont appelés polynômes de Hermite.

## Propriété

Les polynômes de Hermite sont orthogonaux vis-à-vis de la mesure gaussienne :

$$\langle H_n(\xi), H_m(\xi) \rangle := \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} H_n(t) H_m(t) \exp\left(-\frac{t^2}{2}\right) dt = \delta_{n,m}.$$

Exercice : vérifier le caractère orthogonal des polynômes de Hermite pour n=0,1,2.

# Partie 3 : propriétés remarquables

## Base des variables aléatoires du second ordre

- Soit $X$ une v.a. quelconque d'ordre 2 ($\leftrightarrow \mathbb{E}[X^2] < +\infty$ ), et $\xi$ une v.a. gaussienne centrée réduite. Alors $X$ peut s'écrire sous la forme :

$$X = \sum_{n=1}^{+\infty} \langle X, H_n(\xi) \rangle \, H_n(\xi).$$

- Autrement dit, la famille $\{H_n, \ n \geq 1\}$ forme une base orthonormée des vecteurs aléatoires d'ordre 2.
- Ce résultat se généralise à l'ordre $d$ pour la dimension du vecteur aléatoire $\boldsymbol{X}$ considéré.

# Plan de la séance

# Partie 4 : ouvertures

## Vecteur gaussien et processus gaussien

- Un processus aléatoire est une fonction $\{X(t), \ t \in \Omega\}$, telle qu'en tout $t \in \Omega$, $X(t)$ est une variable aléatoire.
- Caractériser la distribution d'un processus aléatoire revient à identifier l'infinité des lois jointes de $(X(t_1), \dots, X(t_n))$, $n \geq 1$.
- La notion de vecteur gaussien se généralise à la notion de processus aléatoire gaussien en faisant tendre $n$ vers $+\infty$.

Exercice : généraliser pour les processus gaussiens les propriétés de paramétrage minimal, de stabilité par transformation affine et par conditionnement.

# Partie 4 : ouvertures

## La distribution gaussienne au coeur des différentes séances de cours

- estimation statistique par approche spectrale,
- propagation d'incertitudes pour l'estimation d'intervalles de confiance,
- métamodélisation pour quantifier la précision des prédictions,
- approches non-paramétriques comme distribution élémentaire des représentations à noyaux,
- calibration pour la modélisation des erreurs de mesure et de modèle,
- fiabilité pour l'approximation de probabilités de dépassement de seuils.

# Plan de la séance