# Data Science Fundamentals

## Part I: Probability theory

ISEP 2nd year
2023-2024

Based on the course given by Nathalie Colin & Jean-Claude Guillerot

# Probability theory

- **5th session (October 27th 2023):**

- **Chapter 7: EXPECTED VALUE, CHARACTERISTIC FUNCTION AND MOMENTS FOR TWO-DIMENSIONAL RANDOM VARIABLES**

- **Chapter 6 (continuation): GAUSSIAN RANDOM VARIABLES**

# Probability theory

- **5th session (October 27th 2023):**
- **Chapter 7: EXPECTED VALUE, CHARACTERISTIC FUNCTION AND MOMENTS FOR TWO-DIMENSIONAL RANDOM VARIABLES**
- **Chapter 6 (continuation): GAUSSIAN RANDOM VARIABLES**

# Expected value of a function (1/3)

**Reminder : Expected value of a function of a random variable**

$$Y = g(X)$$

$$E(Y) = E(g(X)) = \int g(x) f_X(x) dx$$

**Definition : Expected value of a function of 2 variables**

$$Z = g(X, Y)$$

**Continuous case** $$\boxed{E(Z) = E(g(X, Y)) = \iint g(x, y) f_{X,Y}(x, y) dxdy}$$

**Discrete case** $$\boxed{E(Z) = E(g(X, Y)) = \sum_{i,j} g(x_i, y_j) P_{i,j}}$$

**$P_{i,j}$ : two-dimensional distribution**

## Expected value of a function (2/3)

**Remark :** linearity of the expected value

$$g(x,y)=\sum_k \lambda_k g_k(x,y)$$

$$\boxed{E\left(\sum_k \lambda_k g_k(x,y)\right)=\sum_k \lambda_k E\left(g_k(x,y)\right)}$$

*Expected value of a sum = sum of expected values*

**Example:** $$E(Z)=E(aX+bY)=aE(X)+bE(Y)$$

## Two-dimension conditional expected value :

So far, we have seen : $$f_Y\left(y \middle| X=x\right)=\frac{f_{X,Y}(x,y)}{f_X(x)}$$

So: $$E(Y|X=x)=\int y f_Y(y|X=x)dy = \frac{\int y f_{x,y}(x,y)dy}{f_X(x)}$$

## Expected value of a function (3/3)

### Two-dimension conditional expected value (continuation) :

**In general :**

$$E\big(g(X,Y)\big|X=x\big)=\int g(x,y)f_Y(y|X=x)dy = \frac{\int g(x,y)f_{x,y}(x,y)dy}{f_X(x)}$$

**Similarly:**

$$E\big(g(X,Y)\big|Y=y\big)=\int g(x,y)f_X(x|Y=y)dx = \frac{\int g(x,y)f_{x,y}(x,y)dx}{f_Y(y)}$$

# Characteristic function (1/2)

**Definition :**

$$\boxed{\varphi_{X,Y}(t_1,t_2) = E\left(e^{jt_1x+jt_2y}\right)}$$

$$\boxed{\varphi_{X,Y}(t_1,t_2) = \iint e^{jt_1x+jt_2y} f_{X,Y}(x,y)\,dx\,dy}$$

**with $t_1$ and $t_2$ : deterministic real variables**

$$\varphi_X(t_1) = \varphi_{X,Y}(t_1,0) \qquad \varphi_Y(t_2) = \varphi_{X,Y}(0,t_2) \qquad \varphi_{X,Y}(0,0) = 1$$

**Theorem: Inversion formulae :** $\boxed{f_{X,Y} \Leftrightarrow \varphi_{X,Y}}$

$$f_{X,Y}(x,y) = \frac{1}{(2\pi)^2}\iint e^{-(jt_1x+jt_2y)}\varphi_{X,Y}(t_1,t_2)\,dt_1\,dt_2$$

## Characteristic function (2/2)

## Moment-generating function:

$$\left.\frac{\partial^{k+l}\varphi_{X,Y}\left(t_1,t_2\right)}{\partial t_1^k \partial t_2^l}\right|_{t_1=t_2=0} = j^{k+l}E\left(X^k Y^l\right)$$

**The proof is similar to that of one-dimensional case: Taylor series of the characteristic function (around the point (0,0)).**

$$\varphi_{X,Y}(t_1,t_2) = 1 + jE(X)t_1 + jE(Y)t_2 - \frac{1}{2}E(X^2)t_1^2 - \frac{1}{2}E(Y^2)t_2^2 - E(XY)t_1t_2 + \ldots.$$

# Moments for a pair of random variables (1/5)

## Definition

**For 1 random variable, moment of order $n$ :**

$$m_n = E(X^n) = \int x^n f_X(x)dx$$

**For 2 random variables, moment of order $n$ such that $n = l + k$ :**

$$\boxed{m_{l,k} = E(X^l Y^k) = \iint x^l y^k f_{X,Y}(x,y)dxdy}$$

**For 2 discrete random variables :**

$$\boxed{m_{l,k} = E(X^l Y^k) = \sum_i \sum_j x_i^k y_j^l P_{i,j}}$$

**Remarks :**

- There are two 1st order moments: E (X) and E (Y)
- There are three 2nd order moments: E ($X^2$), E ($Y^2$), E (XY)
- In general, there are $(n + 1)$ moments of $n^{th}$ order

## Moments for a pair of random variables (2/5)

### Definition (continuation) :

**Central moments with $m_{10}$ = E(X) and $m_{01}$ = E(Y) :**

$$\mu_{l,k} = E((X - m_{1,0})^l (Y - m_{0,1})^k)$$

$$\mu_{l,k} == \iint (x - m_{1,0})^l (y - m_{0,1})^k f_{X,Y}(x, y)dxdy$$

$$\mu_{0,0} = 1$$

$$\mu_{1,0} = \mu_{0,1} = 0$$

$\mu_{2,0} = \sigma_X^2 \quad \mu_{0,2} = \sigma_Y^2$ **and** $\mu_{1,1} = \sigma_{X,Y}$ **the covariance of** $(X, Y)$

- **The covariance is also denoted** $cov\,(X, Y)$

# Moments for a pair of random variables (3/5)

## Covariance :

$$\mu_{1,1} = E\left(\left(X - m_{1,0}\right)\left(Y - m_{0,1}\right)\right) = \sigma_{X,Y}$$

$$\boxed{\mu_{1,1} = \sigma_{X,Y} = E(XY) - E(X)E(Y) = m_{1,1} - m_{1,0}m_{0,1}}$$

with $m_{10} = E(X)$ and $m_{01} = E(Y)$

Remark : $\sigma_{X,Y}$ it can be positive or negative.

**Schwartz inequality** : relation between the 2nd order moments

$$\boxed{E^2(XY) \leq E(X^2)E(Y^2)} \qquad m_{1,1}^2 \leq m_{2,0}m_{0,2}$$

For zero-mean random variables: $\mu_{1,1}^2 \leq \mu_{2,0}\mu_{0,2}$

In general: $\boxed{\sigma_{XY}^2 \leq \sigma_X^2 \sigma_Y^2}$

# Moments for a pair of random variables (4/5)

## Correlation coefficient :

$$\boxed{r = \frac{E\big((X - m_{1,0})(Y - m_{0,1})\big)}{\sqrt{E\big((X - m_{1,0})^2\big)E\big((X - m_{0,1})^2\big)}} = \frac{\mu_{1,1}}{\mu_{2,0}\mu_{0,2}} = \frac{\sigma_{X,Y}}{\sigma_X \sigma_Y}}$$

## Properties :

- The correlation coefficient measures the linear dependence between X and Y
- $\boxed{-1 \leq r \leq 1}$ (using the Cauchy-Schwartz inequality)
- Special values :
  - r = $\pm$ 1 : there is a perfect linear dependency between X and Y : Y=$a$X+ $b$
  - r = 0 : X and Y are non-correlated $\boxed{E(XY) = E(X)E(Y)}$

# Moments for 2 random variables (5/5)

## Independance and correlation :

**Independance :** $$f_{XY}(x,y) = f_X(x) f_Y(y)$$

$$\boxed{\begin{aligned} m_{l,k} &= E(X^l Y^k) = E(X^l)E(Y^k) \\ m_{l,k} &= m_{l,0} m_{0,k} \quad \forall\, l \text{ and } k \end{aligned}}$$

**For l = k = 1 :** $m_{1,1} = m_{1,0} m_{0,1}$ **=> r = 0 so X and Y decorrelated**

**Independance => (implies) decorrelation**

**Decorrelation ~~=>~~ Independance *unless (X, Y) is gaussian***

**Orthogonality : E(XY) = 0**

## Application: linear regression (1/2)

Objective: Given a pair of random variables $(X, Y)$, given $X$ fixed, we want to approach $Y$ by a linear function of $X$, denoted $g(X)$:

$$g(X) = aX + b$$

such that the mean square error between $Y$ and $g(X)$ is minimal.

We denote $Z = Y - (aX + b)$ and we look for $a$ and $b$ such that

$E(Z^2) = E((Y\text{-}g(X))^2\} = E((Y - (aX + b))^2)$ (*mean square error)*

is minimal.

By applying some properties of the expected value we get:

$$E(Z^2) = E((Y - aX)^2) + b^2 - 2bE(Y - aX)$$

# Application: linear regression (2/2)

Let us denote $\varepsilon = E(Z^2) = E(Y-g(X))^2$ the mean square error

***We calculate the derivatives to get the values of $a$ and $b$ that minimize $\varepsilon$***

$$\mathrm{E}(Z^2) = \varepsilon = \mathrm{E}(Y^2) + a^2\mathrm{E}(X^2) + b^2 - 2aE(XY) - 2bE(Y) + 2abE(X)$$

The derivatives : $\dfrac{\partial \varepsilon}{\partial a} = a\mathrm{E}(X^2) - E(XY) + bE(X) = 0$

$$\frac{\partial \varepsilon}{\partial b} = b - E(Y) + aE(X) = 0$$

which gives:

$$a = \frac{E(XY) - E(X)E(Y)}{E(X^2) - E^2(X)} = \frac{\sigma_{XY}}{\sigma_X^2} \qquad b = E(Y) - aE(X) \quad et\ r = \frac{\sigma_{XY}}{\sigma_X \sigma_Y}$$

$$a = \frac{r\sigma_Y}{\sigma_X} \qquad b = E(Y) - \frac{r\sigma_Y}{\sigma_X}E(X) \rightarrow \ \varepsilon\ min = \sigma_Y^2(1-r^2) \rightarrow \quad case\ r^2 = 1$$

**Regression line**

$$y = \frac{r\sigma_Y}{\sigma_X}(\,x - E(X)\,) + E(Y)$$

Likewise : if $Y=aX+b$ we find without difficulty that, $r=+-1$.
Conclusion, a necessary and sufficient condition for 2 random variables to be linearly related, is that $|r|=1$.

**What happens if the correlation coefficient is zero?**

*If $r^2$ = 0 the variables are uncorrelated, so*

$E(XY)=E(X)E(Y)$

- Application : calculation of the variance of a sum when the variables are decorrelated.

$$\boxed{Var(aX+bY)=a^2Var(X)+b^2Var(Y)=a^2\sigma_X^2+b^2\sigma_Y^2}$$

Remark : the correlation is a measure of the linear dependence between two variables.

The correlation between two variables may be weak even if the variables are very closely related. This means that, if a relation exists, it is not linear, for instance, it can be polynomial, exponential, etc.

## In summary

| Conditions | Then, |
| --- | --- |
| If | The variables X and Y are : |
| $f_{XY}(x,y) = f_X(x)\, f_Y(y)$ | Independent and, then, non-correlated. |
| $E(XY) = E(X)E(Y)$ | non-correlated |
| $E(XY) = 0$ | orthogonal |
| $r = \pm 1$ | Linearly dependent |

## Formula for the variance

- $Cov(X,Y) = E(XY) - E(X)E(Y)$
- $Var(aX + bY) = a^2Var(X) + b^2Var(Y) + 2abCov(X,Y)$
- $Cov(X,X) = Var(X)$
- $Cov(aX + bY, cU + dV) = ac.Cov(X,U) + ad.Cov(X,V) + bc.Cov(Y,U) + bd.Cov(Y,V)$
- $Cov(aX + bY, cX + dY) = ac.Var(X) + (ad + bc)Cov(X,Y) + bd.Var(Y)$

**Covariance matrix of a random vector or 2d-variable (X,Y)**

$$V(X,Y) = \begin{pmatrix} Var(X) & Cov(X,Y) \\ Cov(X,Y) & Var(Y) \end{pmatrix}$$

# Probability theory

- **5th session (October 27th 2023):**
- **Chapter 7: EXPECTED VALUE, CHARACTERISTIC FUNCTION AND MOMENTS FOR TWO-DIMENSIONAL RANDOM VARIABLES**
- **Chapter 6 (continuation): GAUSSIAN RANDOM VARIABLES**

## GAUSSIAN RANDOM VARIABLES

### Definition :

The pair of random variables or 2d-vector (X, Y) is Gaussian if its density is of the form:

$$f_{X,Y}(x,y)=e^{-\varphi(x,y)}$$

with

$$\varphi(x,y)=ax^2+bxy+cy^2+dx+ey+k$$

Remark : it depends on only 5 parameters

*k* is fixed by :

$$\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f_{X,Y}(x,y)\,dx\,dy=1$$

Usual presentation :

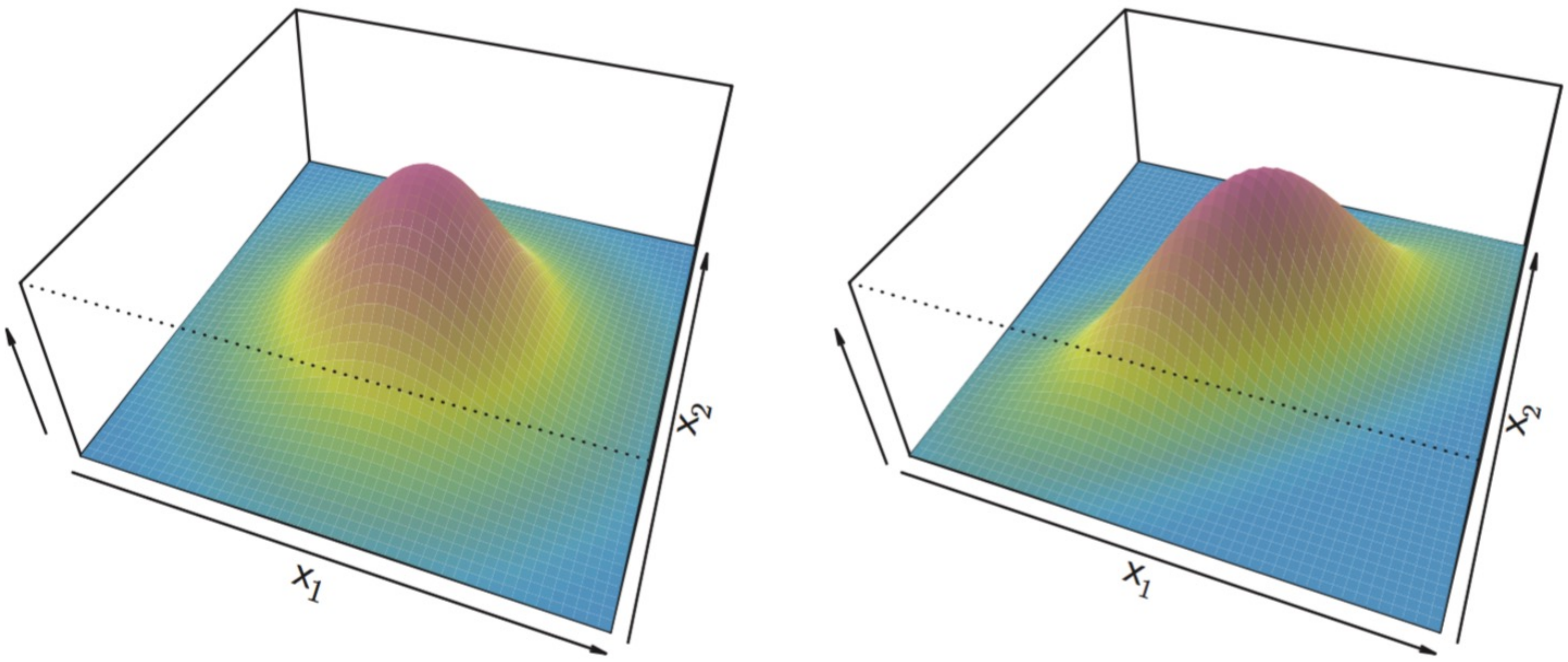
$$f_{X,Y}(x,y)=\frac{1}{2\pi\,\sigma_x\sigma_y\sqrt{1-r^2}}e^{-\frac{1}{2\left(1-r^2\right)}\varphi(x,y)}$$

The 5 parameters are

$m_x$, $m_y$, r, $\sigma_x$ and $\sigma_y$

$$\varphi(x,y)=\frac{(x-m_{x1})^2}{\sigma_x^2}-\frac{2r(x-m_{x1})(y-m_{y1})}{\sigma_x\sigma_y}+\frac{(y-m_{y1})^2}{\sigma_y^2}$$

## GAUSSIAN RANDOM VARIABLES




Left: Equal variance and zero correlation, Right: different variances and existing correlation.

For a *p-dimensional* vector, the density is: $f(x) = \frac{1}{(2\pi)^{\frac{p}{2}}|\boldsymbol{V}|^{\frac{1}{2}}} e^{-\frac{1}{2}(x-\mu)^T\boldsymbol{V}^{-1}(x-\mu)}$

Where $\mu \in \mathbb{R}^p$ is the vector of expected values (mean vector) and ***V*** is the covariance matrix.

# Marginal densities

Marginal for Y **:**

$$f_Y(y) = \int_{-\infty}^{+\infty} f_{X,Y}(x,y)dx$$

$$\boxed{f_Y(y) = \frac{1}{\sigma_Y\sqrt{2\pi}} e^{-\frac{(y-m_Y^2)^2}{2\sigma_Y^2}}}$$

Similarly for X :

$$\boxed{f_X(x) = \frac{1}{\sigma_X\sqrt{2\pi}} e^{-\frac{(x-m_X^2)^2}{2\sigma_X^2}}}$$

$m_x$ , $\sigma_x$ : expected value and standard deviation of X

$m_y$ , $\sigma_y$ : expected value and standard deviation of Y

**Two dimensional gaussian density ⇒ Gaussian marginal densities**

**the opposite is not necessarily true**

r is the ***"coefficient of correlation"*** between X and Y,

## <u>Independance</u> :

**X and Y gaussian each and independent from each other.** $\longleftrightarrow$ $f_{XY}(x,y) = f_X(x) f_Y(y)$

**(X,Y) Gaussian, r=0**

**Important remark: the non-correlation between 2 random variables (i.e. r = 0 ) implies the independance <u>only</u> in the case of a gaussian vector.**

# Summary table about Gaussian vector properties

In general

- X and Y independent $\Rightarrow$ cov(X,Y)=0
- cov(X,Y)=0 $\nRightarrow$ X and Y independent

Definition: (X,Y) is a Gaussian vector $\Leftrightarrow$ $\forall \boldsymbol{\alpha}, \boldsymbol{\beta} \in \mathbb{R}$, $\boldsymbol{\alpha}X+\boldsymbol{\beta}Y$ is a Gaussian random variable.

- If (X,Y) is a Gaussian vector and X and Y are independent $\Leftrightarrow$ cov(X,Y)=0.
- (X,Y) Gaussian vector $\Rightarrow$ X Gaussian and Y Gaussian (particular case if $\boldsymbol{\beta}$=0 and $\boldsymbol{\alpha}$=0 respectively).
- (X Gaussian and Y Gaussian) $\nRightarrow$ (X,Y) Gaussian vector.
- (X Gaussian, Y Gaussian and (X,Y) independent) $\Rightarrow$ (X,Y) Gaussian vector. (the opposite is true iff cov(X,Y)=0)
- (X Gaussian, Y Gaussian and (X,Y) independent) $\Rightarrow$ $\forall \boldsymbol{\alpha}, \boldsymbol{\beta} \in \mathbb{R}$, $\boldsymbol{\alpha}X+\boldsymbol{\beta}Y$ is a Gaussian random variable.

# Theorem concerning Gaussian vectors

**Theorem:** Let X be a Gaussian random vector in $\mathbb{R}^p$ that follows the p-dimensional Gaussian distribution with parameters $\mu$ (vector of means) and **V** (covariance matrix). If **A** is a deterministic *d*x*p*-matrix and a vector U $\in \mathbb{R}^d$, then:

$$E(\mathbf{A}X + U) = \mathbf{A}\, E(X) + U$$
$$\mathrm{Var}(\mathbf{A}X+U)=\mathbf{A}\mathbf{V}\mathbf{A}^T$$