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

➢ **4th session (October 20th 2023):**

Chapter 5 (bis): TRANSFORMATION OF A REAL-VALUED RANDOM VARIABLE

Chapter 6 : TWO-DIMENSIONAL RANDOM VARIABLES

Chapter 5 (bis): TRANSFORMATION OF A REAL RANDOM VARIABLE

- Deterministic function of a random variable
  - Transformation of the density $f_X(x)$
  - Examples
  - Particular cases

Description of the problem

Let X be a random variable of density $f_X(x)$ and g a deterministic real function of the variable X.

$$X \longrightarrow \boxed{\text{Transformation } g} \longrightarrow Y$$

Random variable of probability density $f_X(x)$: known

Random variable of probability density unknown

Transformation : deterministic function, Y=g(X)

Examples of transformation : $Y = X^2$; $Y = aX + B$

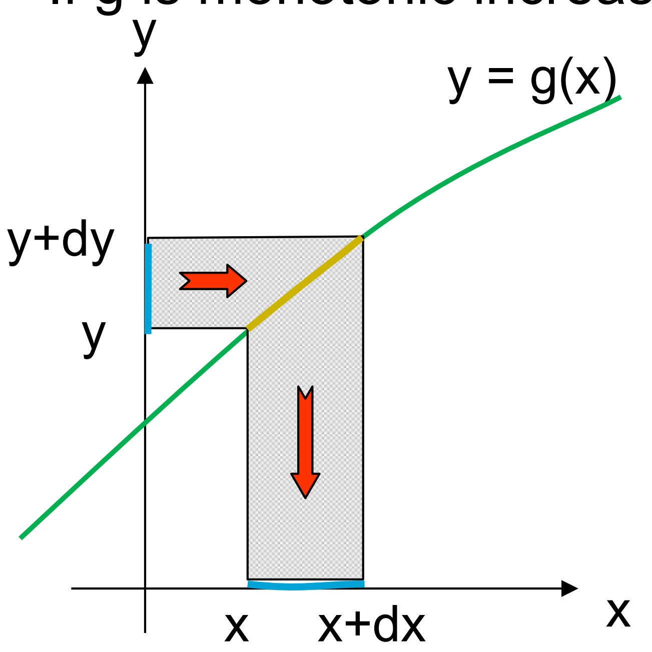
Three situations can take place :

- One-to-one correspondance between X and Y $\quad y \longleftrightarrow x$
- Non- One-to-one correspondance $\quad y \longleftrightarrow (x_i)$ for i = 1, 2, ,.,, n
- Particular cases

Case 1: One-to-one correspondance between X and Y

Subcase 1.1:

If g is monotonic increasing



the event: *is created by:*

$$\{y \le Y \le y+dy\} \Leftrightarrow \{x \le X \le x+dx\}$$

Given the fact that:

y fixed has only one antecedent x

dy>0 and dx>0

- $P\{y \le Y \le y+dy\} = P\{x \le X \le x+dx\}$
- $f_Y(y)\,dy = f_X(x)\,dx$

Then, the density of Y is:

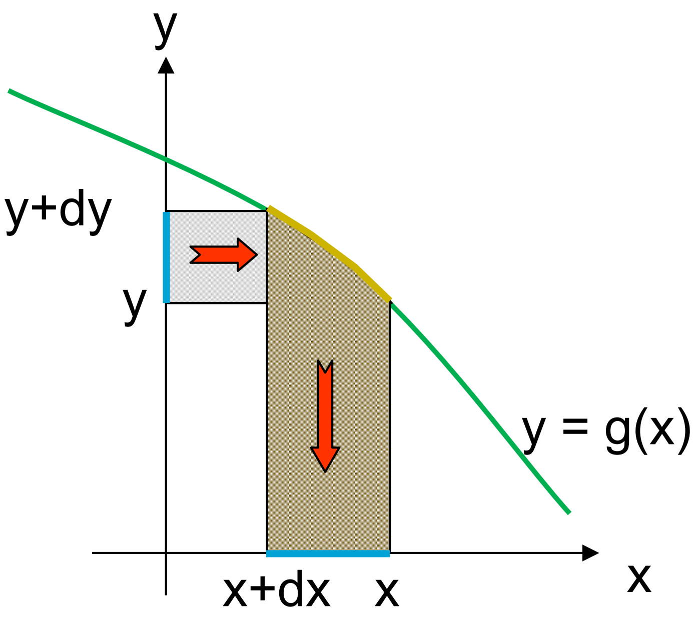
$$\boxed{f_Y(y) = f_X(x)\frac{dx}{dy} = f_X(g^{-1}(y))\frac{dx}{dy}}$$

Where g^{-1} is the inverse function of g

Case 1: One-to-one correspondance between X and Y

Subcase 1.2:

If *g* is monotonic decreasing



the event: $\{y \le Y \le y+dy\}$ $\Leftrightarrow$ *is created by:* $\{x+dx \le X \le x\}$

Given the fact that:

y fixed has only one antecedent x

dy > 0 but dx < 0

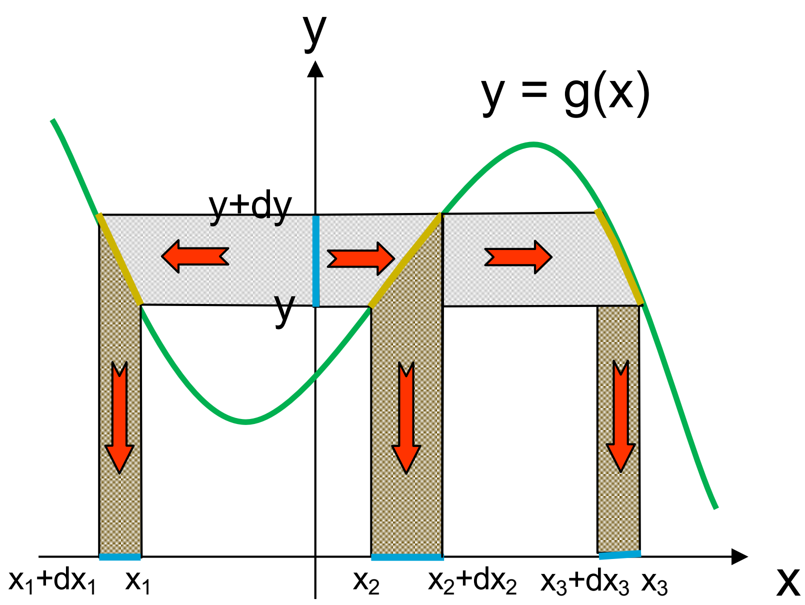
- $P\{y \le Y \le y+dy\} = P\{x+dx \le X \le x\}$
- $f_Y(y)\,dy = -f_X(x)\,dx$

Then, the density of Y is:

$$\boxed{f_Y(y) = f_X(x)\left|\frac{dx}{dy}\right| = f_X(g^{-1}(y))\left|\frac{dx}{dy}\right|}$$

Case 2: the correspondance between X and Y is not one-to-one (1/2)

There might be more than one value x_i of X corresponding to the same value y



y can have more than one antecedent

$$y = g(x_i) \qquad i = 1,2,....$$

(for example (x_1, x_2, x_3))

dy>0 but $dx_1<0$, $dx_2>0$, $dx_3<0$

the event: *is created by:*

$$\{y \le Y \le y + dy\} \Leftrightarrow \{x_1 + dx_1 \le X \le x_1\} or \{x_2 \le X \le x_2 + dx_2\} or \{x_3 + dx_3 \le X \le x_3\}$$

Case 2: the correspondance between X and Y is not one-to-one (2/2)

$$\{y \le Y \le y + dy\} \Leftrightarrow \{x_1 + dx_1 \le X \le x_1\} or \{x_2 \le X \le x_2 + dx_2\} or \{x_3 + dx_3 \le X \le x_3\}$$

Since the events

$$\{x_1 + dx_1 \le X \le x_1\},\ \{x_2 \le X \le x_2 + dx_2\},\ \{x_3 + dx_3 \le X \le x_3\}$$

are disjoint $\quad f_Y(y)|dy| = \sum_i f_X(x_i)\,|dx_i|$

According to the axiom 3:

$$\boxed{f_Y(y) = \sum_i f_X(x_i) \left|\frac{dx}{dy}\right|_{x = x_i}}$$

where x_i are the solutions of the equation $x = g^{-1}(y)$ when y is fixed.

Example (1/2) :

Let X a random variable with density $f_X(x)$ and the transformation $Y = X^2$. The aim is to calculate $f_Y(y)$?

The inverse function is: $x = \pm\sqrt{y}$ with y>0

Calculation of the derivative : $dy = 2 \cdot x \cdot dx$

$$\left[g_i^{-1}(y)\right]' = \frac{dx}{dy} = \pm\frac{1}{2\sqrt{y}}$$

The density $f_Y(y)$ will be :

$$f_Y(y) = \frac{1}{2\sqrt{y}} \cdot \left[f_X\left(\sqrt{y}\right) + f_X\left(-\sqrt{y}\right)\right]$$

Example (2/2):

Application : $f_X(x)$ is a density of Rayleigh :

$$f_X(x) = \frac{x}{\sigma^2} \cdot \exp\left(-\frac{x^2}{2\sigma^2}\right) \qquad x \geq 0$$

$$f_X(x) = 0 \quad \textit{elsewhere}$$

Warning : x ≥ 0 only the positive root belongs to the domain of definition

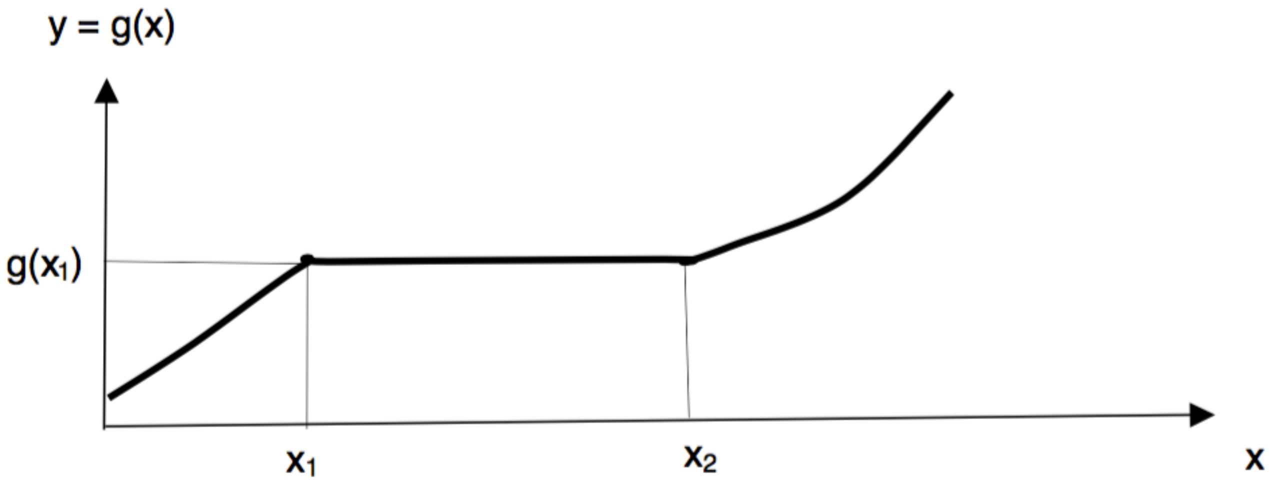
$$f_Y(y) = \frac{1}{2\sqrt{y}} \cdot \frac{\sqrt{y}}{\sigma^2} \cdot \exp\left(-\frac{\left(\sqrt{y}\right)^2}{2\sigma^2}\right) = \frac{1}{2\sigma^2} \cdot \exp\left(-\frac{y}{2\sigma^2}\right)$$

with y >=0

It is an exponential distribution with parameter $\frac{1}{2\sigma^2}$

Particulier case 1 : g(x) is constant in an interval

Let us suppose that *g* is continuous, non-decreasing for all x except in the interval $[x_1, x_2]$



$$g(x_1) = g(x_2)$$

$$P\{Y = g(x_1)\} = P\{x_1 \le X \le x_2\} = \int_{x_1}^{x_2} f_X(x) \cdot dx$$

All the values of x in the interval $[x_1,x_2]$ are transformed in a single value of $Y=g(x_1)$. And $P(Y=g(x_1)) \ge 0$ (probability that Y belongs to an interval of length 0). Therefore:

- The density, $f_Y(y)$ is the Dirac distribution at point $y = g(x_1)$
- Discontinuity in the CDF, $F_Y(y)$

$$F_Y\{y = g(x_1^+)\} = F_Y\{y = g(x_1^-)\} + P\{Y = g(x_1)\}$$

Example of the particular case (1/2)

The random variable X follows the uniform distribution in the interval $[-1,1]$.

Consider the transformation $g(x) = \begin{cases} x \text{ if } x \geq 0 \\ 0 \text{ if } x < 0 \end{cases}$

For $x < 0$: $P(Y = 0) = P(X < 0) = \frac{1}{2}\int_{-1}^{0} dx = \frac{1}{2}$

For $x \geq 0$ $f_Y(y) = f_X(x) = \frac{1}{2}$

Therefore, the probability density function of Y is:

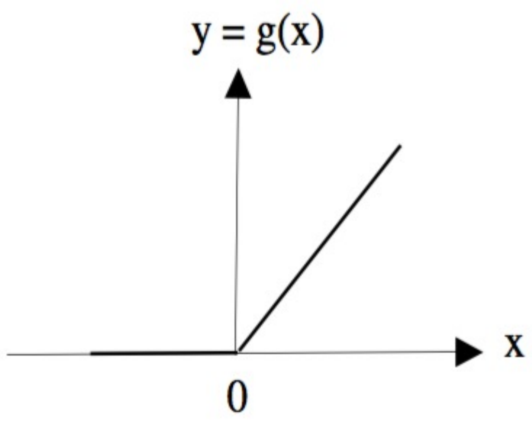
$$f_Y(y) = \frac{1}{2}\delta(y) \Leftrightarrow P(Y = y) = \frac{1}{2}, \qquad y=0$$

$$f_Y(y) = \frac{1}{2}, \quad 0 < y \leq 1$$

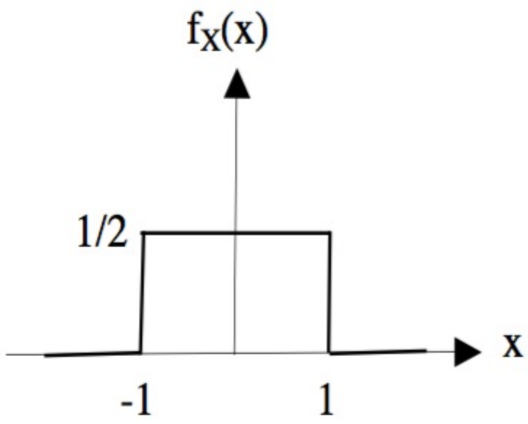
And the CDF of Y is:

$F_Y(y) = \frac{1}{2} + \frac{y}{2}$, $0 \leq y \leq 1$

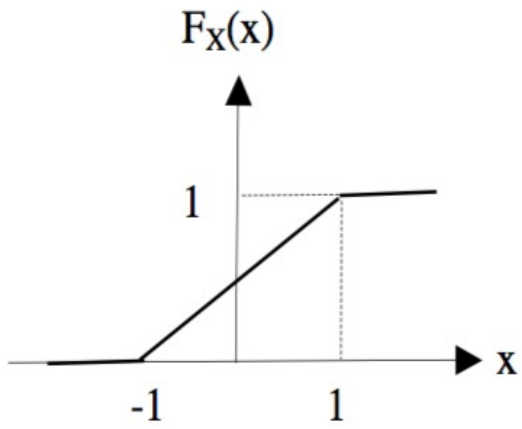
Example of the particular case (2/2)



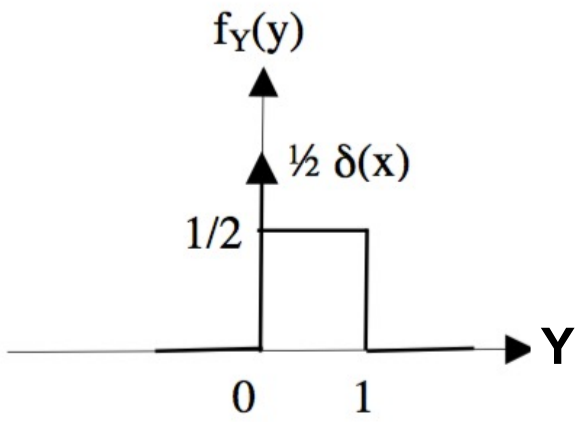
Transformation



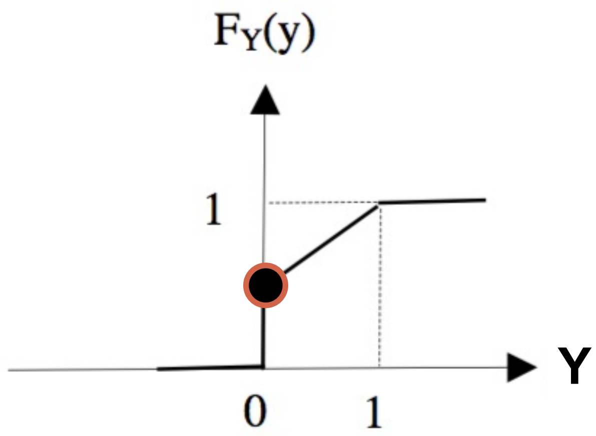
Density of X



CDF of X



Density of Y



CDF of Y

Chapter 6 (beggining): TWO-DIMENSIONAL RANDOM VARIABLES

- **Two-dimensional discrete random variables**
- **Two-dimensional continuous random variables**

Chapter 6 (beggining): TWO-DIMENSIONAL RANDOM VARIABLES

- **Two-dimensional discrete random variables**
- **Two-dimensional continuous random variables**

Reminder for a one-dimensional random variable

A one-dimensional random variable X is completely defined by one of the following data:

- » Its Cumulative distribution function (CDF) $F_x(x)$
- » Its probability density funcion $f_x(x)$
- » Its characteristic function $\varphi_x(x)$

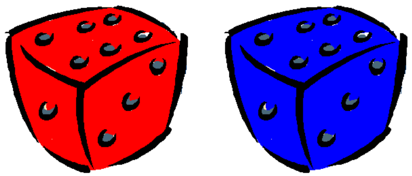
Described by its moments m_n, μ_n

Possibility of creation of new random variables via a transformation function.

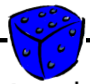
All these notions will be extended for a couple of two random variables, (X, Y)

Example: experiment

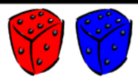
Rolling two dice, a red one and a blue one



- 36 possible outcomes : (numbers appearing on the upper face of each die)
- The dice are symmetric => equiprobability (all the outcomes have the same probability)



	1	2	3	4	5	6
1	(1,1)	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)
2	(2,1)	(2,2)	(2,3)	(2,4)	(2,5)	(2,6)
3	(3,1)	(3,2)	(3,3)	(3,4)	(3,5)	(3,6)
4	(4,1)	(4,2)	(4,3)	(4,4)	(4,5)	(4,6)
5	(5,1)	(5,2)	(5,3)	(5,4)	(5,5)	(5,6)
6	(6,1)	(6,2)	(6,3)	(6,4)	(6,5)	(6,6)



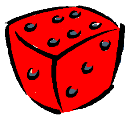
Red die *(l=1,2,3,4,5,6)*, blue die *(k= 1,2,3,4,5,6)*

$$P\{\omega_n\} = \frac{1}{36}, \quad n = 1 \text{ à } 36$$

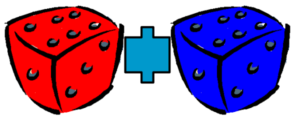
Definition of two-dimensional random variables (1/2)

To define the pair of random variables (X,Y), we define two mappings from Ω to the set of real numbers:

X (red die)

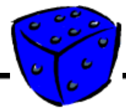


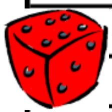
Y (sum for the 2 dice)



$$
\begin{aligned}
X(\omega_n) &= X(l,k) &&= l &&= x_i \\
Y(\omega_n) &= Y(l,k) &&= l + k &&= y_j
\end{aligned}
$$

Possible outcomes of the pair (X,Y):





	1	2	3	4	5	6
1	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)	(1,7)
2	(2,3)	(2,4)	(2,5)	(2,6)	(2,7)	(2,8)
3	(3,4)	(3,5)	(3,6)	(3,7)	(3,8)	(3,9)
4	(4,5)	(4,6)	(4,7)	(4,8)	(4,9)	(4,10)
5	(5,6)	(5,7)	(5,8)	(5,9)	(5,10)	(5,11)
6	(6,7)	(6,8)	(6,9)	(6,10)	(6,11)	(6,12)

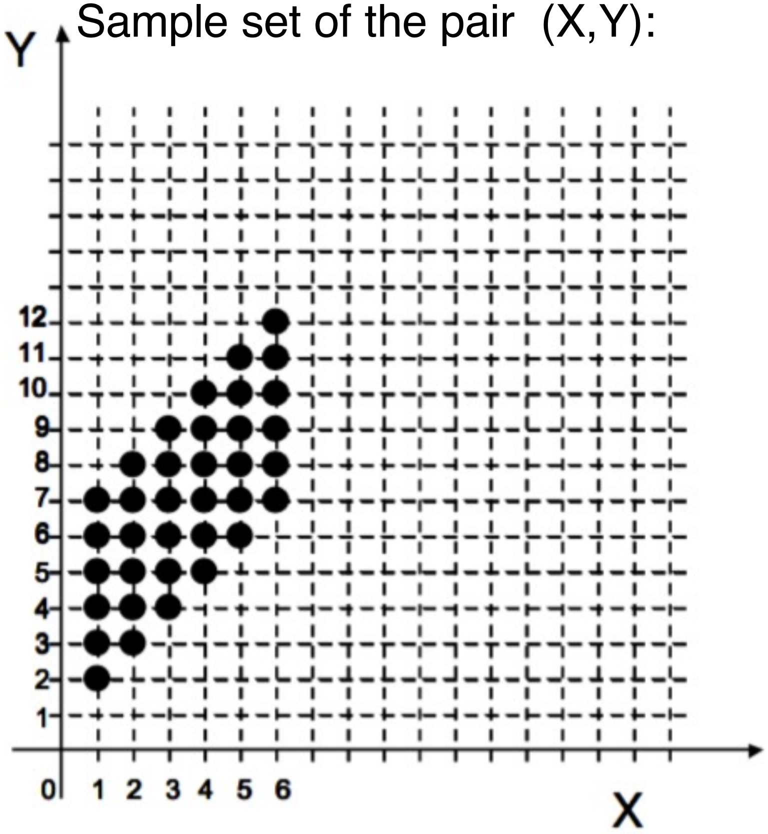
Definition of two-dimensional random variables (2/2)

Possible outcomes of X : $x_i = 1,2,...6$ for i=1 to 6

Possible outcomes of Y : $y_j = 2,3,...12$ for j=1 to 11

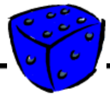
The couple (X, Y) cannot take the values of any combination (x_i, y_j). For example the values (4,3) or (2,10) will never occur!

Sample set of the pair (X,Y):



Probability density of the pair (X,Y): $P(X=x_i \text{ and } Y=y_j)=P_{ij}$

The two mappings stablish a one-to-one relation (bijection) between the outcomes (ω_n) and the ordered pairs (x_i, y_j).



	1	2	3	4	5	6
1	(1,2)	(1,3)	(1,4)	(1,5)	(1,6)	(1,7)
2	(2,3)	(2,4)	(2,5)	(2,6)	(2,7)	(2,8)
3	(3,4)	(3,5)	(3,6)	(3,7)	(3,8)	(3,9)
4	(4,5)	(4,6)	(4,7)	(4,8)	(4,9)	(4,10)
5	(5,6)	(5,7)	(5,8)	(5,9)	(5,10)	(5,11)
6	(6,7)	(6,8)	(6,9)	(6,10)	(6,11)	(6,12)

36 outcomes for each value of the couple: $(\boldsymbol{x_i}, \boldsymbol{y_j})$

$$\boldsymbol{p_{i,j}} = \boldsymbol{P}\{\omega_n\} = \frac{1}{36}, \quad \text{for } n=1 \text{ to } 36$$

Contingency table

A table that defines the probability distribution of the pair (X, Y)

Y	x1=1	x2=2	x3=3	x4=4	x5=5	x6=6	P Yj
y1=2	1/36	0	0	0	0	0	1/36
y2=3	1/36	1/36	0	0	0	0	2/36
y3=4	1/36	1/36	1/36	0	0	0	3/36
y4=5	1/36	1/36	1/36	1/36	0	0	4/36
y5=6	1/36	1/36	1/36	1/36	1/36	0	5/36
y6=7	1/36	1/36	1/36	1/36	1/36	1/36	6/36
y7=8	0	1/36	1/36	1/36	1/36	1/36	5/36
y8=9	0	0	1/36	1/36	1/36	1/36	4/36
y9=10	0	0	0	1/36	1/36	1/36	3/36
y10=11	0	0	0	0	1/36	1/36	2/36
y11=12	0	0	0	0	0	1/36	1/36
P Xi	6/36	6/36	6/36	6/36	6/36	6/36	1

<u>joint law (X,Y)</u>

$$P_{ij} \geq 0$$

$$\sum_i \sum_j P_{ij} = 1$$

Marginal law of Y

$$P(Y = y_j) = \sum_i P_{ij}$$

The sum of all the entries of the table must be <u>always</u> 1!

Marginal law of X $P(X = x_i) = \sum_j P_{ij}$

Conditional distribution

Given X equal to a fixed value, for example $X = x_3 = 3$, we can calculate the conditional probability of Y if X takes the value of 3.

Y	x1=1	x2=2	x3=3	x4=4	x5=5	x6=6	P Yj
y1=2	1/36	0	0	0	0	0	1/36
y2=3	1/36	1/36	0	0	0	0	2/36
y3=4	1/36	1/36	1/36	0	0	0	3/36
y4=5	1/36	1/36	1/36	1/36	0	0	4/36
y5=6	1/36	1/36	1/36	1/36	1/36	0	5/36
y6=7	1/36	1/36	1/36	1/36	1/36	1/36	6/36
y7=8	0	1/36	1/36	1/36	1/36	1/36	5/36
y8=9	0	0	1/36	1/36	1/36	1/36	4/36
y9=10	0	0	0	1/36	1/36	1/36	3/36
y10=11	0	0	0	0	1/36	1/36	2/36
y11=12	0	0	0	0	0	1/36	1/36
P Xi	6/36	6/36	6/36	6/36	6/36	6/36	1

Conditional probability of X given Y fixed

$$P(X = x_i | Y = y_j) = \frac{P_{ij}}{P_{Y_j}} = \ldots \frac{1}{3}$$

Conditional probability of Y given X

$$P(Y = y_j | X = x_i) = \frac{P_{ij}}{P_{X_i}} = \ldots \frac{1}{6}$$

In summary for a couple of discrete random variables (X,Y)

The pair (X,Y) having the following outcomes :

$$X = x_i, \qquad i = 1,2,\dots$$
$$Y = y_j, \qquad j = 1,2,\dots$$

Joint probability function $\quad P\left(X = x_i \text{ and } Y = y_j\right) = p_{i,j}$

Marginal distribution of *X* $\quad P(X = x_i) = \sum_j p_{i,j} = PX_i$

Marginal distribution of *Y* $\quad P\left(Y = y_j\right) = \sum_i p_{i,j} = PY_j$

Conditional distribution of X given Y $\quad P\left(X = x_i \mid Y = y_j\right) = \dfrac{p_{i,j}}{\sum_i p_{i,j}} = \dfrac{p_{i,j}}{PY_j}$

Conditional distribution of Y given X $\quad P\left(Y = y_j \mid X = x_i\right) = \dfrac{p_{i,j}}{\sum_j p_{i,j}} = \dfrac{p_{i,j}}{PX_i}$

Independance

The 2 variables X and Y are independent if for any i, j :

$$P\left(X = x_i \cap Y = y_j\right) = P\left(X = x_i\right)P\left(Y = y_j\right) \text{ or } p_{ij} = PX_i PY_j \ \forall i, j$$

If they are independent the conditional distributions are identical to the marginal distributions.

$$P\left(X = x_i \mid Y = y_j\right) = PX_i\,, \qquad \forall\, y_j$$

$$P\left(Y = y_j \mid X = x_i\right) = PY_j\,, \qquad \forall\, x_i$$

Chapter 6 (beggining): TWO-DIMENSIONAL RANDOM VARIABLES

- **Two-dimensional discrete random variables**
- **Two-dimensional continuous random variables**

Cumulative distribution function (CDF)

Definition :

Reminder for one-dimensional variable: $F_X(x) = P(X \leq x\ )$

For two-dimensional random variables (X,Y), the CDF is the probability of the intersection of the events : $\{X \leq x\ \}$ and $\{Y \leq y\}$

$$\boxed{F_{X,Y}(x, y) = P(\{X \leq x\} \cap \{Y \leq y\})}$$

Notation :

F : Joint cumulative distribution function

X, Y : random variables

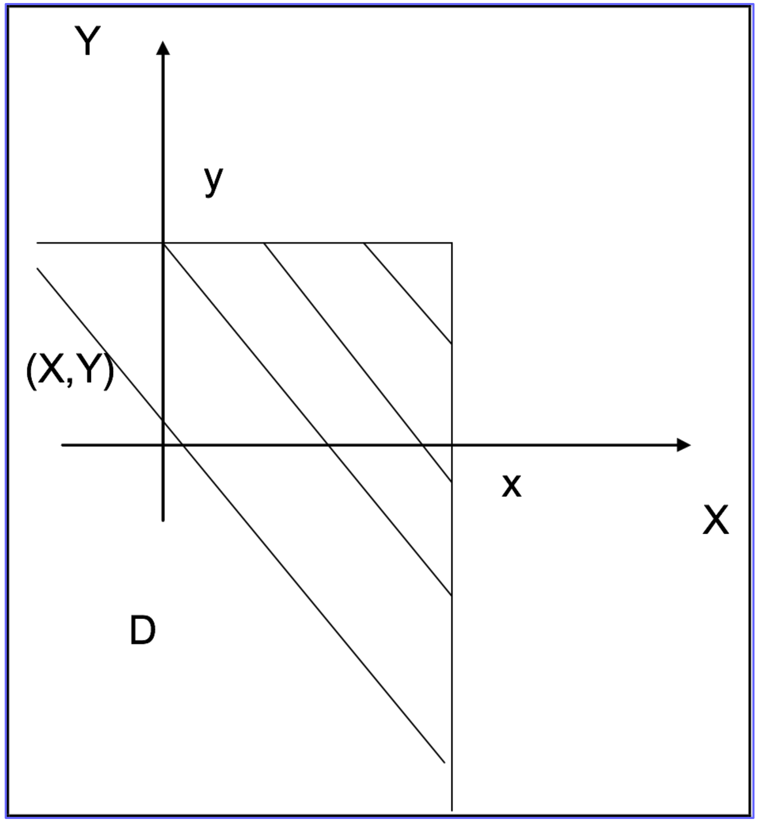
x, y : real thresholds

P : Probability

Graphical represention (very useful!) :

Representation in the cartesian coordinate system

$F_{XY}(x, y)$: Probability that the pair $(X,Y) \in D$



Marginal cumulative distribution functions

Marginal cumulative distribution function of X

It is the cumulative distribution function of X only

$$F_X(x) = P(\{X \leq x\}) = P(\{X \leq x\}, \forall Y)$$

Just consider the two-dimensional cumulative distribution function and set:

$y = +\infty$ (or $y = y_{max}$ in the domain $D_{x,y}$)

$$\boxed{F_X(x) = P(\{X \leq x, \forall\, Y < +\infty\}) = F_{X,Y}(x, +\infty)}$$

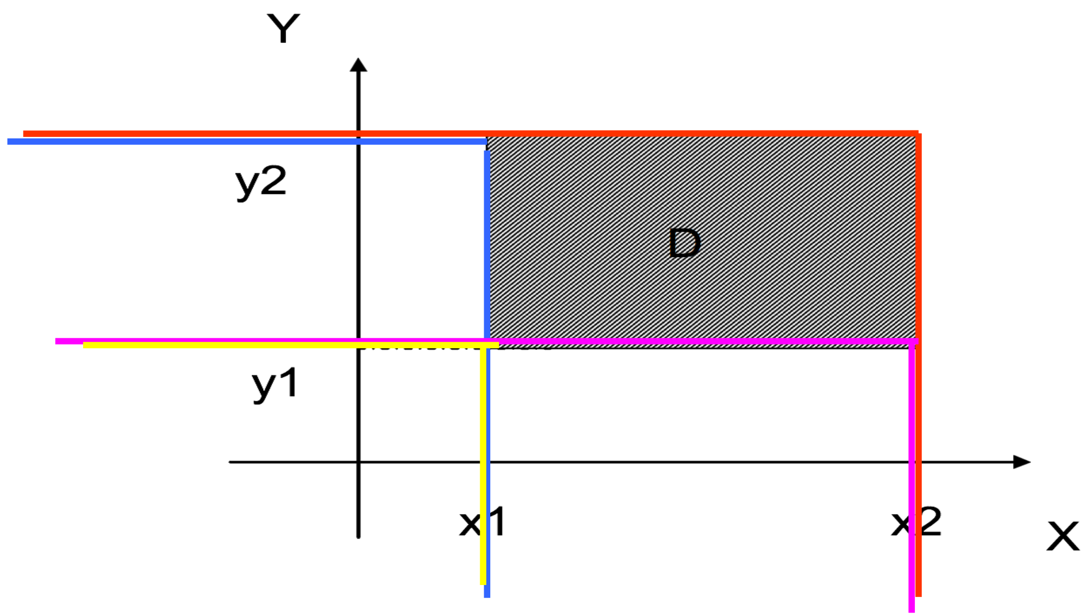
Marginal cumulative distribution function of Y

$$\boxed{F_Y(y) = F_{X,Y}(+\infty, y)}$$

Calculation of the probability of a rectangle (two-dimension interval)

We aim to calculate in terms of $F_{X,Y}(x,y)$ the probability of the pair (X,Y) to belong to the rectangle D:

$$\boldsymbol{D}: \{\boldsymbol{X} \in [\boldsymbol{x_1}, \boldsymbol{x_2}] \text{ and } \boldsymbol{Y} \in [\boldsymbol{y_1}, \boldsymbol{y_2}]\}$$



$$P(\{x_1 < X \leq x_2, y_1 < Y \leq y_2\})$$
$$= F_{X,Y}(x_2,y_2) - F_{X,Y}(x_1,y_2) - F_{X,Y}(x_2,y_1) + F_{X,Y}(x_1,y_1)$$

Joint probability density function

The one-dimensional case: $f_X(x) = \dfrac{dF_X(x)}{dx}$

The two-dimensional case :

$$f_{X,Y}(x,y) = \frac{\partial}{\partial y}\left(\frac{\partial F_{X,Y}(x,y)}{\partial x}\right) = \frac{\partial^2 F_{X,Y}(x,y)}{\partial y \partial x}$$

Notation :

f : probability density function

X, Y : random Variables

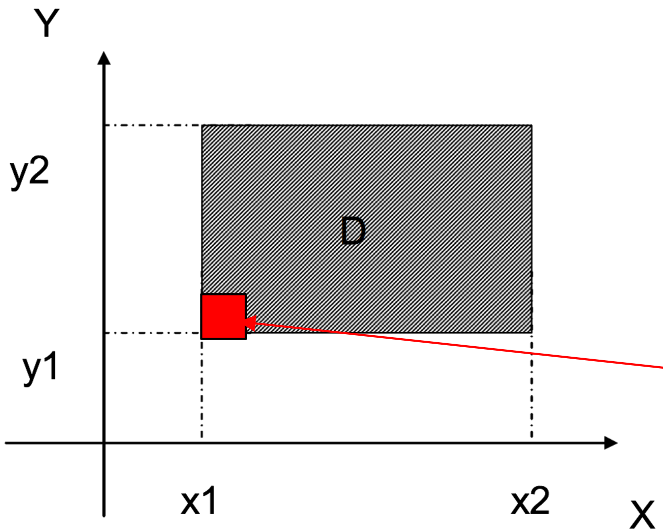
x, y : real thresholds

$\dfrac{\partial^2}{\partial y \partial x}$: derivatives with respect to x (threshold), then y (threshold)

Infinitesimal interpretation of the probability density

Let us calculate the integral over the domain D (rectangle) :

$$\int_{x1}^{x2}\int_{y1}^{y2} f_{X,Y}(x,y)dxdy = \underbrace{F_{X,Y}(x_2,y_2) - F_{X,Y}(x_1,y_2) - F_{X,Y}(x_2,y_1) + F_{X,Y}(x_1,y_1)}_{P(\{x_1 < X \leq x_2, y_1 < Y \leq y_2\})}$$



In the infinitesimal domain *dS=dxdy* the density can be considered constant

$$x_1 = x,\ x_2 = x + dx$$

$$y_1 = y,\ y_2 = y + dy$$

$$\boxed{P(\{x < X \leq x + dx, y < Y \leq y + dy\}) = f_{X,Y}(x,y)dxdy}$$

Interpretation: the probability of the pair (X, Y) of belonging to dS at the neighborhood of the point (x, y) is proportional to the value of the joint density function at that point.

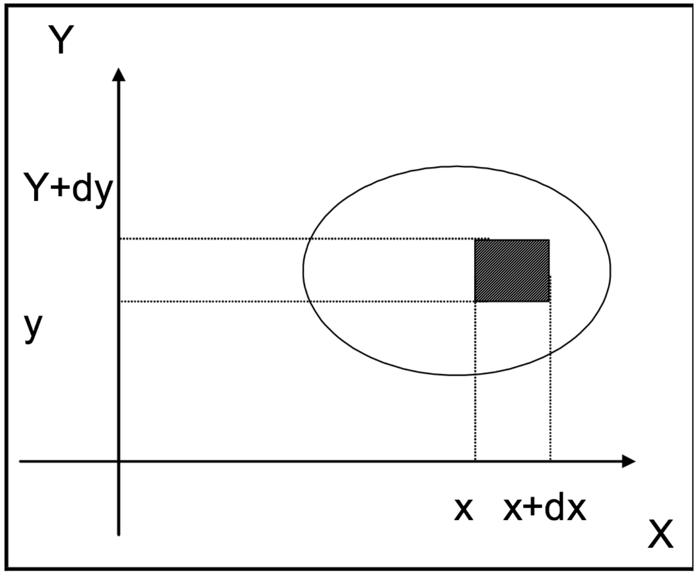
Conclusion: for a continuous random variable, the probability of having {X = x, Y = y} is zero because dS is considered zero.

Use of the density :

The density allows to calculate the probability for a couple of random variables (X, Y) to belong to any domain D

$$\boxed{P(\{(X,Y)\in D\})=\int_D f_{X,Y}(x,y)dxdy}$$

Just decompose D into disjoint elements and apply the axiom 3 of probabilities.



The integral represents the volume between the surface defined by the joint probability density function and the domain D.

Relation with the cumulative distribution function:

$$\boxed{\int_{-\infty}^{x_2}\int_{-\infty}^{y_2} f_{X,Y}(x,y)\,dx\,dy = F_{X,Y}(x_2,y_2)}$$

By taking $x_2 = +\infty,\ y_2 = +\infty$, we obtain :

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

Properties:

- the volume under the surface is 1.

- The density is non-negative.

$$\boxed{f_{X,Y}(x,y) \geq 0}$$

Marginal densities of X and Y :

Marginal density of X

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

Marginal density of Y

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

Hint: calculate the integral over the definition domain of the other variable to make it disappear

Conditional distributions

Conditional distribution of Y given X

$$f_Y(y|X=x)=\frac{f_{X,Y}(x,y)}{\int_{-\infty}^{+\infty} f_{X,Y}(x,y)dy}=\frac{f_{X,Y}(x,y)}{f_X(x)}$$

Conditional distribution of X given Y

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

What happens if X and Y are independent?

Reminder : If the events A and B are independent :

$$P(A \cap B) = P(A)P(B)$$

If A = {X ≤ x} and B = {Y≤ y} then $P(X \leq x,\ Y \leq y) = P(X \leq x)P(Y \leq y)$:

$$\boxed{F_{X,Y}(x,y) = F_X(x)F_Y(y)}$$

The joint cumulative distribution function of the couple is equal to the product of the marginal cumulative distribution functions.

What about the density?

$$\boxed{f_{X,Y}(x,y) = f_X(x)f_Y(y)}$$

The joint density is also the product of the marginal densities.