Statistics I: Chapter 4: Multivariate Random Variables

Carlos Oliveira

E-mail: carlosoliveira@iseg.ulisboa.pt

ISEG - Lisbon School of Economics and Management

Multivariate random variable: A random variable $k-$ dimensional is a function with domain $\mathbf{S}$ and codomain $\mathbb{R}^k$:

$$(X_1, \ldots, X_k) : s \in \mathbf{S} \to (X_1(s), \ldots, X_k(s)) \in \mathbb{R}^k \, .$$

The function $(X_1(s), \ldots, X_k(s))$ is usually written for simplicity as $(X_1, \ldots, X_k)$.

Remark: If $k = 2$ we have the bivariate random variable or two dimensional random variable

$$(X, Y) : s \in \mathbf{S} \to (X(s), Y(s)) \in \mathbb{R}^2 \, .$$

Joint cumulative distribution function: Let (X, Y) be a bivariate random variable. The real function of two real variables with domain $\mathbb{R}^2$ and defined by

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

is the joint cumulative distribution function of the two dimensional random variable (X, Y) .

Example (Dice casting)

Random Experiment: Roll two different dice (one red and one green) and write down the number of dots on the upper face of each die.

Random vector: (X_{red}, X_{green}), where X_i is the of dots the i die, with $i =$ green or red.

Some probabilities:

$$P(X_{red} = 2, X_{green} = 4) = \frac{1}{36}$$

$$P(X_{red} + X_{green} > 10) = \frac{1}{12}$$

$$P\left(\frac{X_{red}}{X_{green}} \leq 2\right) = P\left(\frac{X_{red}}{X_{green}} = 1\right) + P\left(\frac{X_{red}}{X_{green}} = 2\right)$$

$$= \frac{6}{36} + \frac{3}{36} = \frac{1}{4}$$

Example (Coin Tossing)

Random experiment: Two different and fair coins are tossed once.

Random vector: (X_1, X_2), where X_i represents the number of heads obtained with coin i, with $i = 1, 2$.

Some probabilities:

$$P(X_1 = 0, X_2 = 0) = P(X_1 = 0, X_2 = 1) = P(X_1 = 1, X_2 = 0) = P(X_1 = 1, X_2 = 1) = \frac{1}{4}$$

$$P(X_1 + X_2 \geq 1) = \frac{3}{4}$$

Properties of the joint cumulative distribution function:

- $0 \leq F_{X,Y}(x,y) \leq 1$
- $F_{X,Y}(x,y)$ is non decreasing with respect to x and y :
  - $\Delta_x > 0 \Rightarrow F_{X,Y}(x+\Delta x, y) \geq F_{X,Y}(x,y)$
  - $\Delta_y > 0 \Rightarrow F_{X,Y}(x, y+\Delta y) \geq F_{X,Y}(x,y)$
- $\lim_{x\to-\infty} F_{X,Y}(x,y) = 0$, $\lim_{y\to-\infty} F_{X,Y}(x,y) = 0$ and $\lim_{x\to+\infty, y\to+\infty} F_{X,Y}(x,y) = 1$
- $P(x_1 < X \leq x_2, y_1 < Y \leq y_2) = F_{X,Y}(x_2,y_2) - F_{X,Y}(x_1,y_2) - F_{X,Y}(x_2,y_1) + F_{X,Y}(x_1,y_1)$.
- $F_{X,Y}(x,y)$ is right continuous with respect to x and y: $\lim_{x\to a^+} F_{X,Y}(x,y) = F_{X,Y}(a,y)$ and $\lim_{y\to b^+} F_{X,Y}(x,y) = F_{X,Y}(x,b)$.

Example (Coin Tossing)

Random experiment: Two different and fair coins are tossed once.

Random vector: (X_1, X_2), where X_i represents the number of heads obtained with coin i, with $i = 1, 2$.

Joint cumulative distribution function:

$$F_{X_1,X_2}(x_1,x_2) = P(X_1 \leq x_1, X_2 \leq x_2) = \begin{cases} 0, & x_1 < 0 \\ 0, & x_2 < 0 \\ \frac{1}{4}, & 0 \leq x_1 < 1, 0 \leq x_2 < 1 \\ \frac{1}{2}, & 0 \leq x_1 < 1, x_2 \geq 1 \\ \frac{1}{2}, & 0 \leq x_2 < 1, 0 \leq x_1 < 1 \\ 1, & x_1 \geq 1, x_2 \geq 1 \end{cases}$$

The (marginal) cumulative distribution functions of X and Y can be obtained form the Joint cumulative distribution functions of (X, Y) :

- The Marginal cumulative distribution function of X :
  $F_X(x) = P(X \leq x) = P(X \leq x, Y \leq +\infty) = \lim_{y \to +\infty} F_{X,Y}(x, y)$.
- The Marginal cumulative distribution function of Y :
  $F_Y(y) = (Y \leq y) = P(X \leq +\infty, Y \leq y) = \lim_{x \to +\infty} F_{X,Y}(x, y)$.

Remark: The joint distribution uniquely determines the marginal distributions, but the inverse is not true.

Example (Coin Tossing)

Random experiment: Two different and fair coins are tossed once.
Random vector: (X_1, X_2), where X_i represents the number of heads obtained with coin i, with $i = 1, 2$.
Joint distribution function:

$$F_{X_1,X_2}(x_1,x_2) = \begin{cases} 0, & x_1 < 0 \text{ or } x_2 < 0 \\ \frac{1}{4}, & 0 \leq x_1 < 1, 0 \leq x_2 < 1 \\ \frac{1}{2}, & 0 \leq x_1 < 1, x_2 \geq 1 \\ \frac{1}{2}, & 0 \leq x_2 < 1, 0 \leq x_1 < 1 \\ 1, & x_1 \geq 1, x_2 \geq 1 \end{cases}$$

Marginal distribution function for X_1:

$$F_{X_1}(x_1) = \lim_{x_2 \to +\infty} F_{X_1,X_2}(x_1,x_2) = \begin{cases} 0, & x_1 < 0 \\ \frac{1}{2}, & 0 \leq x_1 < 1 \\ 1, & x_1 \geq 1 \end{cases}$$

Example

Let (X, Y) be a jointly distributed random variable with CDF:

$$F_{X,Y}(x,y) = \begin{cases} 1 - e^{-x} - e^{-y} + e^{-x-y}, & x \geq 0, y \geq 0 \\ 0, & x < 0, y < 0 \end{cases}.$$

Marginal cumulative distribution function of the random variable X is:

$$F_X(x) = \begin{cases} 1 - e^{-x}, & x \geq 0 \\ 0, & x < 0 \end{cases}.$$

Marginal cumulative distribution function of the random variable Y is:

$$F_Y(y) = \begin{cases} 1 - e^{-y}, & y \geq 0 \\ 0, & y < 0 \end{cases}.$$

Independence of jointly distributed random variables

Definition: The jointly distributed random variables X and Y are said to be independent if and only if for any two sets $B_1 \in \mathbb{R}$, $B_2 \in \mathbb{R}$ we have

$$P(X \in B_1, Y \in B_2) = P(X \in B_1)P(Y \in B_2)$$

Remark: Independence implies that $F_{X,Y}(x, y) = F_X(x)F_Y(y)$, for any $(x, y) \in \mathbb{R}^2$.

Theorem: If X and Y are independent random variables and if $h(X)$ and $g(Y)$ are two functions of X and Y respectively, then the random variables $U = h(X)$ and $V = g(Y)$ are also independent random variables.

Example (Coin Tossing)

Random experiment: Two different and fair coins are tossed once.

Random vector: (X_1, X_2), where X_i represents the number of heads obtained with coin i, with $i = 1, 2$.

Are these random variables independent?

$$F_{X_1}(x_1) = \begin{cases} 0, & x_1 < 0 \\ \frac{1}{2}, & 0 \leq x_1 < 1 \\ 1, & x_1 \geq 1 \end{cases}, \quad F_{X_2}(x_2) = \begin{cases} 0, & x_2 < 0 \\ \frac{1}{2}, & 0 \leq x_2 < 1 \\ 1, & x_2 \geq 1 \end{cases}$$

One can easily verify that $F_{X_1,X_2}(x_1, x_2) = F_{X_1}(x_1) \times F_{X_1}(x_1)$.

$$F_{X_1,X_2}(x_1, x_2) = \begin{cases} 0, & x_1 < 0 \text{ or } x_2 < 0 \\ \frac{1}{4}, & 0 \leq x_1 < 1, 0 \leq x_2 < 1 \\ \frac{1}{2}, & 0 \leq x_1 < 1, x_2 \geq 1 \\ \frac{1}{2}, & 0 \leq x_2 < 1, 0 \leq x_1 < 1 \\ 1, & x_1 \geq 1, x_2 \geq 1 \end{cases}$$

Example

Let (X, Y) be a jointly distributed random variable with CDF:

$$F_{X,Y}(x,y) = \begin{cases} 1 - e^{-x} - e^{-y} + e^{-x-y}, & x \geq 0, y \geq 0 \\ 0, & x < 0, y < 0 \end{cases}.$$

Marginal cumulative distribution function of the random variable X and Y are:

$$F_X(x) = \lim_{y \to +\infty} F_{X,Y}(x,y) = \begin{cases} 1 - e^{-x}, & x \geq 0 \\ 0, & x < 0 \end{cases}$$

$$F_Y(y) = \lim_{x \to +\infty} F_{X,Y}(x,y) = \begin{cases} 1 - e^{-y}, & y \geq 0 \\ 0, & y < 0 \end{cases}.$$

X and Y are independent random variables because:

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

since

$$1 - e^{-x} - e^{-y} + e^{-x-y} = (1 - e^{-y})(1 - e^{-x}).$$

Let $D_{(X,Y)}$ be the set of of discontinuities of the joint cumulative distribution function $F_{(X,Y)}(x,y)$, that is

$$D_{(X,Y)} = \left\{(x,y) \in \mathbb{R}^2 : P(X = x, Y = y) > 0\right\}$$

Definition: (X, Y) is a two dimensional discrete random variable if and only if

$$\sum_{(x,y)\in D_{(X,Y)}} P\left(X = x, Y = y\right) = 1.$$

Remark: As in the univariate case, a multivariate discrete random variable can take a finite number of possible values (x_i, y_j), where $i = 1, 2, ..., k_1$ and $j = 1, 2, ..., k_2$, where k_1 and k_2 are finite integers, or a countably infinite (x_i, y_j), where $i = 1, 2, ...$ and $j = 1, 2, ...$ For the sake of generality we consider the latter case. That is $D_{(X,Y)} = \{(x_i, y_j), i = 1, 2, ..., j = 1, 2, ...\}$

Joint probability distribution/ function: If X and Y are discrete random variables, then the function given by

$$f_{X,Y}(x,y) = P\left(X = x, Y = y\right)$$

for $(x,y) \in D_{(X,Y)}$ is called the joint probability function of (X,Y) or joint probability distribution of the random variables X and Y.

Theorem: A bivariate function $f_{X,Y}(x,y)$ can serve as joint probability distribution of the pair of discrete random variables X and Y if and only if its values satisfy the conditions:

- $f_{X,Y}(x,y) \geq 0$ for any $(x,y) \in \mathbb{R}^2$
- $\sum_{(x,y)\in D_{(X,Y)}} f_{X,Y}(x,y) = \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} f_{X,Y}(x_i,y_j) = 1$

Remark: We can calculate any probability using this function. For instance $P\left((x,y) \in B\right) = \sum_{(x,y)\in B} f_{X,Y}(x,y)$

Example

Let X and Y be the random variables representing the population of monthly wages of husbands and wives in a particular community. Say, there are only three possible monthly wages in euros: 0, 1000, 2000. The joint probability distribution is

Y \ X	0	1000	2000
0	0.05	0.15	0.10
1000	0.10	0.10	0.30
2000	0.05	0.05	0.10

The probability that a husband earns 2000 euros and the wife earns 1000 euros is given by

$$\begin{aligned} f_{X,Y}(2000, 1000) &= P(X = 2000, Y = 1000) \\ &= 0.30 \end{aligned}$$

Joint cumulative distribution function: If X and Y are discrete random variables, the function given by

$$F_{X,Y}(x\ y) = \sum_{s \leq x} \sum_{t \leq y} f_{X,Y}(s,t)$$

for $(x,y) \in \mathbb{R}^2$ is called the joint distribution function or joint cumulative distribution of X and Y.

Marginal probability distribution/function: If Y and X are discrete random variables and $f_{X,Y}$ is the value of their joint probability distribution at (x,y) the function given by

$$\begin{aligned} P\left(X=x\right) &= \begin{cases} \sum_{y \in D_y} f(x,y) = \sum_{y \in D_Y} f_{X,Y}(x,y), & \text{for } x \in D_x \\ 0, & \text{for } x \notin D_x \end{cases} \\ P\left(Y=y\right) &= \begin{cases} \sum_{x \in D_x} f(x,y) = \sum_{x \in D_X} f_{X,Y}(x,y), & \text{for } y \in D_y \\ 0, & \text{for } y \notin D_y \end{cases} \end{aligned} .$$

are respectively is the Marginal probability distribution of the r.v. X and Y, where D_x and D_y are the range of X and Y respectively.

Example

$$P(X=x) = P(X=x, Y=0) + P(X=x, Y=1000) + P(X=x, Y=2000)$$

$$P(Y=y) = P(X=0, Y=y) + P(X=1000, Y=y) + P(X=2000, Y=y)$$

Applying these formulas we have:

Y \ X	0	1000	2000	$P(Y=y)$
0	0.05	0.15	0.10	0.30
1000	0.10	0.10	0.30	0.50
2000	0.05	0.05	0.10	0.20
$P(X=x)$	0.20	0.30	0.50	1

$$F_{X,Y}(1000,1000) = P(X=0, Y=0) + P(X=0, Y=1000) + P(X=1000, Y=0) + P(X=1000, Y=1000)$$

$$F_{X,Y}(0,1000) = P(X=0, Y=0) + P(X=0, Y=1000)$$

Independence of random variables: Two discrete random variables X and Y are independent if and only if, for all $(x, y) \in D_{X,Y}$,

$$P(X = x, Y = y) = P(X = x) P(Y = y).$$

Example

	X	0	1000	2000	$P(Y = y)$
Y					
0		0.05	0.15	0.10	0.30
1000		0.10	0.10	0.30	0.50
2000		0.05	0.05	0.10	0.20
$P(X = x)$		0.20	0.30	0.50	1

Are these two random variables independent?

$P(X = 2000, Y = 2000) = P(X = 2000) \times P(Y = 2000) = 0.1$

Is this sufficient to say that X and Y are independent? **NO!**

$P(X = 0, Y = 0) = 0.05$ but $P(X = 0)P(Y = 0) = 0.06$

thus X and Y are not independent.

Conditional probability function of Y given X: A *conditional probability function* of a discrete random variable Y given another discrete variable X taking a specific value is defined as

$$f_{Y|X=x}(y) = P(Y=y|X=x) = \frac{P(Y=y, X=x)}{P(X=x)}$$
$$= \frac{f_{X,Y}(x,y)}{f_X(x)}, \quad f_X(x) > 0, \text{ for a fixed } x.$$

The *conditional probability* function of X given Y is defined by

$$f_{X|Y=y}(x) = \frac{f_{X,Y}(x,y)}{f_y(y)}, \quad f_Y(y) > 0.$$

Remarks:

- The conditional probability functions satisfy all the properties of probability functions, and therefore $\sum_{i=1}^{\infty} f_{Y|X}(y_i) = 1$.
- If X and Y are independent $f_{Y|X=x}(y) = f_y(y)$ and $f_{X|Y=y}(x) = f_X(x)$

Example

Consider the joint probability function

	X	0	1000	2000	$P(Y=y)$
Y					
0		0.05	0.15	0.10	0.30
1000		0.10	0.10	0.30	0.50
2000		0.05	0.05	0.10	0.20
$P(X=x)$		0.20	0.30	0.50	1

Compute $P(Y=y|X=0)$.

$$P(Y=0|X=0) = \frac{P(Y=0,X=0)}{P(X=0)} = \frac{0.05}{0.2} = 0.25$$

$$P(Y=1000|X=0) = \frac{P(Y=1000,X=0)}{P(X=0)} = \frac{0.1}{0.2} = 0.5.$$

$$P(Y=2000|X=0) = \frac{P(Y=2000,X=0)}{P(X=0)} = \frac{0.05}{0.2} = 0.25.$$

Definition: The conditional CDF of Y given X by defined by

$$F_{Y|X=x}(y) = P\left(Y \leq y | X = x\right) = \sum_{y' \in D_Y \wedge y' \leq y} \frac{P\left(Y = y, X = x\right)}{P\left(X = x\right)}$$

for a fixed x, with $P(X = x) > 0$.

Remark: It can be checked that $F_{Y|X=x}$ is indeed a CDF.

Exercice: Verify that $F_{Y|X=x}$ is non-decreasing and and $\lim_{y \to +\infty} F_{Y|X=x}(y) = 1$.

Definition: The conditional CDF of Y given X by defined by

$$F_{Y|X=x}(y) = P\left(Y \leq y|X=x\right) = \sum_{y' \in D_Y \wedge y' \leq y} \frac{P\left(Y=y, X=x\right)}{P\left(X=x\right)}$$

for a fixed x, with $P(X=x) > 0$.

Remark: It can be checked that $F_{Y|X=x}$ is indeed a CDF.

Exercice: Verify that $F_{Y|X=x}$ is non-decreasing and and $\lim_{y \to +\infty} F_{Y|X=x}(y) = 1$.

1) $F_{Y|X=x}(y+\delta) - F_{Y|X=x}(y)$

$$= P\left(Y \leq y+\delta, X=x\right) - P\left(Y \leq y, X=x\right) \geq 0.$$

2) $\lim_{y \to +\infty} F_{Y|X=x}(y) = \lim_{y \to +\infty} P\left(Y \leq y|X=x\right)$

$$= \lim_{y \to +\infty} \frac{P\left(Y \leq y, X=x\right)}{P\left(X=x\right)} = \frac{P\left(Y \leq \infty, X=x\right)}{P\left(X=x\right)} = \frac{P\left(X=x\right)}{P\left(X=x\right)} = 1.$$

Example

Consider the conditional probability of Y given that $X = 0$ previously deduced:

$$\begin{aligned} P(Y = 0|X = 0) &= \frac{P(Y = 0, X = 0)}{P(X = 0)} = \frac{0.05}{0.2} = 0.25 \\ P(Y = 1000|X = 0) &= \frac{P(Y = 1000, X = 0)}{P(X = 0)} = \frac{0.1}{0.2} = 0.5. \\ P(Y = 2000|X = 0) &= \frac{P(Y = 2000, X = 0)}{P(X = 0)} = \frac{0.05}{0.2} = 0.25. \end{aligned}$$

Then the conditional CDF of Y given that $X = 0$ is

$$F_{Y|X=0}(y) = \begin{cases} 0, & y < 0 \\ 0.25, & 0 \le y < 1000 \\ 0.75, & 1000 \le y < 2000 \\ 1, & y \ge 0 \end{cases}.$$

Definition: (X, Y) is a two-dimensional continuous random variable with a joint cumulative distribution function $F_{X,Y}(x, y)$, if and only if X and Y are continuous random variables and there is a non-negative real function $f_{X,Y}(x, y)$, such that

$$F_{X,Y}(x, y) = P(X \leq x, Y \leq y) = \int_{-\infty}^{y} \int_{-\infty}^{x} f_{X,Y}(t, s) dtds.$$

The function $f_{X,Y}(x, y)$ is the joint (probability) density of X and Y.

Remark: Let A be a set in the $\mathbb{R}^2$. Then,

$$P((X, Y) \in A) = \int \int_{A} f_{X,Y}(t, s) dtds.$$

Example

Joint probability density function of the two dimensional random variable (P_1, S) where P_1 represents the price and S the total sales (in 10000 units).

Joint density function:

$$f_{P_1,S}(p,s) = \begin{cases} 5pe^{-ps}, & 0.2 < p < 0.4, \quad s > 0 \\ 0, & \text{otherwise} \end{cases}$$

Joint cumulative distribution function:

$$\begin{aligned} F_{P_1,S}(p,s) &= P(P_1 \leq p, S \leq s) \\ &= \begin{cases} 0, & p < 0.2 \text{ or } s < 0 \\ -1 + 5p - 5\frac{e^{-0.2s} - e^{-ps}}{s}, & 0.2 < p < 0.4, s \geq 0 \\ 1 - 5\frac{e^{-0.2s} - e^{-0.4s}}{s}, & p \geq 0.4, s \geq 0 \end{cases} \end{aligned}$$

To get the CDF we need to make the following computations:

Example

- If $p < 0.2$ or $s < 0$, then $f_{P_1,S}(p,s) = 0$ and

$$P(P_1 \leq p, S \leq s) = \int_{-\infty}^{p} \int_{-\infty}^{s} f_{P_1,S}(t,u) dtdu = 0$$

- If $0.2 < p < 0.4$ and $s \geq 0$, then

$$P(P_1 \leq p, S \leq s) = \int_{-\infty}^{p} \int_{-\infty}^{s} f_{P_1,S}(t,u) dtdu$$

$$= \int_{0.2}^{p} \int_{0}^{s} f_{P_1,S}(t,u) dtdu = -1 + 5p - 5\frac{e^{-0.2s} - e^{-ps}}{s}$$

- If $p \geq 0.4$ and $s \geq 0$, then

$$P(P_1 \leq p, S \leq s) = \int_{-\infty}^{p} \int_{-\infty}^{s} f_{P_1,S}(t,u) dtdu$$

$$= \int_{0.2}^{0.4} \int_{0}^{s} f_{P_1,S}(t,u) dtdu = 1 - 5\frac{e^{-0.2s} - e^{-0.4s}}{s}$$

Theorem: A bivariate function can serve as a joint probability density function of a pair of continuous random variables X and Y if its values, $f_{X,Y}(x,y)$, satisfy the conditions:

- $f_{X,Y}(x,y) \geq 0$, for all $(x,y) \in \mathbb{R}^2$
- $\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f_{X,Y}(x,y)dxdy = 1$.

Property: Let (X,Y) be a bivariate random variable and $B \in \mathbb{R}^2$, then

$$P((X,Y) \in B) = \int\int_B f_{X,Y}(x,y)dxdy.$$

Example

Let (X, Y) be a continuous bi-dimensional random variable with density function $f_{X,Y}$ given by

$$f_{X,Y}(x,y) = \begin{cases} kx + y, & 0 < x < 1, 0 < y < 1 \\ 0, & \text{otherwise} \end{cases}.$$

Find k.

Solution: From the first condition, we know that $f_{X,Y}(x,y) \geq 0$. Therefore $k \geq 0$. Additionally,

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

This is equivalent to

$$\int_0^1 \int_0^1 kx + ydxdy = 1 \Leftrightarrow \frac{1+k}{2} = 1 \Leftrightarrow k = 1.$$

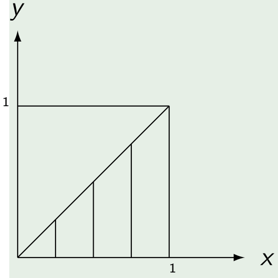
Example

Let (X, Y) be a continuous bi-dimensional random variable with density function $f_{X,Y}$ given by

$$f_{X,Y}(x,y) = \begin{cases} x+y, & 0<x<1, 0<y<1 \\ 0, & \text{otherwise} \end{cases}.$$

Compute $P(X > Y)$.

Solution: Firstly, we notice that



$$P(X > Y) = \int_0^1 \int_0^x (x+y)dydx$$
$$= \int_0^1 \frac{3}{2}x^2 dx = \frac{1}{2}$$

Properties: Let (X, Y) be a continuous bivariate random variable. If $f_{X,Y}$ represents the density function of (X, Y) and $F_{X,Y}$ represents respectively joint CDF of (X, Y). Then,

- $f_{X,Y}(x, y) = \frac{\partial^2 F_{X,Y}(x,y)}{\partial x \partial y} = \frac{\partial^2 F_{X,Y}(x,y)}{\partial y \partial x}$, almost everywhere.
- Marginal density functions of the random variable X

$$f_X(x) = \int_{-\infty}^{+\infty} f_{X,Y}(x, v)dv,$$

- Marginal density functions of the random variable Y

$$f_Y(y) = \int_{-\infty}^{+\infty} f_{X,Y}(u, y)du.$$

- Marginal CDF of the random variable X

$$F_X(x) = \lim_{y \to +\infty} F_{X,Y}(x,y) = \int_{-\infty}^{x} \underbrace{\int_{-\infty}^{+\infty} f_{X,Y}(u,y)dy}_{=f_X(u)}\, du,$$

- Marginal CDF of the random variable Y

$$F_Y(y) = \lim_{x \to +\infty} F_{X,Y}(x,y) = \int_{-\infty}^{y} \underbrace{\int_{-\infty}^{+\infty} f_{X,Y}(x,v)dv}_{=f_Y(v)}\, dx.$$

Example

Joint density function:

$$f_{P,S}(p,s) = \begin{cases} 5pe^{-ps}, & 0.2 < p < 0.4, \quad s > 0 \\ 0, & \text{otherwise} \end{cases}$$

Marginal density function of P:

$$f_P(p) = \int_{-\infty}^{+\infty} f_{P,S}(p,s)ds = \begin{cases} 5\underbrace{\int_0^{+\infty} pe^{-ps} ds}_{=1}, & 0.2 < p < 0.4 \\ 0, & \text{otherwise} \end{cases}$$

$$= \begin{cases} 5, & 0.2 < p < 0.4 \\ 0, & \text{otherwise} \end{cases}$$

Marginal cumulative distribution function:

$$F_S(s) = \lim_{p \to +\infty} F_{P,S}(p,s) = \begin{cases} 0, & s < 0 \\ 1 - 5\frac{e^{-0.2s} - e^{-0.4s}}{s^2}, & s \geq 0 \end{cases}$$

Definition: If $f_{X,Y}(x,y)$ is the joint probability density function of the continuous random variables X and Y and $f_Y(y)$ is the marginal density function of Y, the function given by

$$f_{X|Y=y}(x) = \frac{f_{X,Y}(x,y)}{f_Y(y)}, x \in \mathbb{R} \text{ (for fixed } y\text{)}, f_Y(y) > 0$$

is the **conditional probability function of X given $\{\mathbf{Y} = \mathbf{y}\}$.** Similarly if $f_X(x)$ is the marginal density function of X

$$f_{Y|X=x}(y) = \frac{f_{X,Y}(x,y)}{f_X(x)}, y \in \mathbb{R} \text{ (for fixed } x\text{)}, f_X(x) > 0$$

is the conditional probability function of Y given $\{X = x\}$.

Remark: Note that

$$P(X \in B|Y = y) = \int_B f_{X|Y=y}(x)\, dx$$

for any $B \subset \mathbb{R}$.

Example

(X, Y) is a random vector with the following joint density function:

$$f_{X,Y}(x,y) = \begin{cases} (y+x) & \text{for } (x,y) \in (0,1)\times(0,1) \\ 0 & \text{otherwise} \end{cases} .$$

Conditional density function of Y given that $X = x$ (with $x \in (0,1)$):

$$f_X(x) = \int_0^1 (y+x)dy = x + \frac{1}{2}$$

$$f_{Y|X=x}(y) = \frac{f_{X,Y}(x,y)}{f_X(x)} = \frac{x+y}{x+\frac{1}{2}},\ y \in (0,1)$$

Probability of $Y \geq 0.7 | X = 0.5$

$$P(Y \geq 0.7 | X = 0.5) = \int_{0.7}^1 f_{Y|X=0.5}(y)\,dy = \int_{0.7}^1 (y+0.5)\,dy = 0.405.$$

Remark:

- The conditional density functions of X and Y verify all the properties of a density function of a univariate random variable.
- Note that we can always decompose a joint density function in the following way

$$f_{X,Y}(x,y) = f_X(x) f_{Y|X=x}(y) = f_Y(y) f_{X|Y=y}(x).$$

- If X and Y are independent $f_{Y|X=x}(y) = f_Y(y)$ and $f_{X|Y=y}(x) = f_X(x)$.

Example

Consider the conditional density function of Y given that $X = x$ (with $x \in (0,1)$):

$$f_{Y|X=x}(y) = \frac{f_{X,Y}(x,y)}{f_X(x)} = \frac{x+y}{x+\frac{1}{2}}, \; y \in (0,1).$$

$f_{Y|X=x}$ is indeed a density function:

$$f_{Y|X=x}(y) \geq 0 \quad \text{and} \quad \int_0^1 \frac{x+y}{x+\frac{1}{2}} dy = 1.$$

Example

Consider the conditional density function of Y given that $X = x$ (with $x \in (0,1)$) and the marginal density function of Y.

$$f_{Y|X=x}(y) = \frac{f_{X,Y}(x,y)}{f_X(x)} = \frac{x+y}{x+\frac{1}{2}},\, y \in (0,1)$$

$$f_Y(y) = y + \frac{1}{2},\, y \in (0,1).$$

The random variables are not independent because

$$f_{Y|X=x}(y) \neq f_Y(y), \quad \text{for some } y \in (0,1).$$

Definition: The conditional CDF of Y given X by defined by

$$F_{Y|X=x}(y) = \int_{-\infty}^{y} f_{Y|X=x}(s)ds = \int_{-\infty}^{y} \frac{f_{Y,X}(s,x)}{f_X(x)}ds$$

for a fixed x, with $f_X(x) > 0$.

Remark: It can be checked that $F_{Y|X=x}$ is indeed a CDF.

Example

Consider the conditional density function of Y given that $X = x$ (with $x \in (0,1)$):

$$f_{Y|X=x}(y) = \frac{f_{X,Y}(x,y)}{f_X(x)} = \frac{x+y}{x+\frac{1}{2}},\ y \in (0,1).$$

For $x \in (0,1)$, the conditional cumulative density function is given by:

$$F_{Y|X=x}(y) = \begin{cases} 0, & y < 0 \\ \frac{y(2x+y)}{1+2x}, & 0 \leq y < 1 \\ 1, & y \geq 1 \end{cases}$$

where, $\frac{y(2x+y)}{1+2x} = \int_0^y \frac{x+s}{x+\frac{1}{2}}ds$.