Probability and Statistics

Radjesvarane ALEXANDRE

October 17, 2016

Chapter 1

Probability Spaces

Chapter 2

Random Variables

Chapter 3

2d Random Vectors

3.1 Generalities and dimension 2

Definition 3.1.1

Let Ω be a probability space. An n dimensional (real) random vector is a map X from Ω valued in $\mathbb{R}^n$, with each component being a rrv X_i.

Exemple: Let $\mathcal{E}$ be the experiment where one observes the length of a program submitted for running, together with its running time. Then elements of space Ω are of the form $\omega = (n, t)$ where n is the number of lines in the program and t the running time in seconds. Let $X = (X_1, X_2)$, with $X_1(\omega) = n$ and $X_2(\omega) = t$. This is a 2d random vector. Here X_1 is a drv and X_2 is a crv.

We say that a random vector is continuous if all its components are so, and discrete if this is so for tis components.

To simplify, we only consider in general the case of 2d random vectors.
Thus here $n = 2$.
Moreover, we shall only deal mostly with continuous rv.
Let us first introduce the definition

Definition 3.1.2

The joint distribution function (or more simply the joint distribution) $F_{X,Y}(x, y)$ of two arbitrary random variables X and Y is defined by

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

Properties

1.

$$F(-\infty, y) = 0,\ F(x, -\infty) = 0,\ F(+\infty, +\infty) = 1$$

2.

$$P(x_1 < X \leq x_2, Y \leq y) = F(x_2, y) - F(x_1, y)$$

et

$$P(X \leq x, y_1 < Y \leq y_2) = F(x, y_2) - F(x, y_1)$$

3.

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

Definition 3.1.3

If (X, Y) is a random vector, we say that X and Y are independent random variables if

$$P(X \in A, Y \in B) = P(X \in A)P(Y \in B)$$

for all events A and B.

On can show that

Proposition 3.1.1

If X and Y are two independent rv, then $g(X)$ and $h(Y)$ are also independent, for any continuous functions g and h.

et

Proposition 3.1.2

X and Y are independent iff

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

3.2 The case of a couple of continuous rv

In this section, (X,Y) is a couple of continuous rv.

Definition 3.2.1

If Z=(X,Y) is a continuous random vector, we say that $f_{X,Y}$ is a joint probability density function of the couple (X,Y) if, for any event A, we have

$$P(A) = P(Z \in A) = \int\int_A f_{X,Y}(x, y)dxdy$$

Note that

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

In particular $f_{X,Y}$ is positive. Moreover, its integral is 1, which means that it is a (physical) density.

In particular,

$$F_{X,Y}(x,y) = \int_{-\infty}^{x}\int_{-\infty}^{y} f_{X,Y}(u,v)dudv$$

Note that we get

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

From $F_{X,Y}$, we recover

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

and similarly

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

These are the marginal distribution functions.

Note also that

$$f_X(x) = \partial_x F(x,+\infty) \text{ et } f_Y(y) = \partial_Y F(+\infty, y)$$

Marginal statistics

Finally, note that

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

and similarly for $f_Y(y)$. These are the marginal probability density functions.

One can show that

Proposition 3.2.1

X and Y are independent iff

$$f_{X,Y}(x,y) = f_X(x)f_Y(y) \text{ if (X,Y) a continuous random vector}$$

Exemple : (Buffon needle) A thin needle of length $2a$ is thrown at random on an horizontal plan, recovered with parallel lines (to the y axis) with inter-distance of $2b$, with $b > a$. One can show that the probability that the needle touchs one of these lines is $2a/\pi b$.

In terms of rv, let us introduce the rv X, distance from the center of the needle to the closest line, and the rv Θ given by the angle between the needle and the direction orthogonal to the lines (that is in the direction of the x axis). We can assume that the rv X and Θ are independent, that X is uniform over $(0,b)$ and that Θ is uniform over $(0,\pi/2)$. We deduce that

$$f(x,\theta) = f_X(x)f_\Theta(\theta) = \frac{1}{b}\frac{2}{\pi}, 0 \leq x \leq b, 0 \leq \theta \leq \pi/2$$

and 0 elsewhere. Thus the probability that the point (X,Θ) will be in a region $D \subset R = [0,b] \times [0,\pi/2]$ will be the area of D multiplied by $2/\pi b$.

Here, the needle will intersect one of the lines if $X < a\cos\Theta$. Thus

$$p = P(X < a\cos\Theta) = \frac{2}{\pi b}\int_0^{\pi/2} a\cos\theta d\theta = \frac{2a}{\pi b}$$

One can use this result to find experimentally the number π using the frequency interpretation of p. If the needle is thrown n times, and if it intersects one of the lines n_i times, then

$$p = \frac{2a}{\pi b} \simeq \frac{n_i}{n} \text{ et donc } \pi \simeq \frac{2an}{bn_i}$$

Of course, experimentally means using numerical simulation.
$\odot$

In fact we have also

Proposition 3.2.2

Let (X,Y) be a continuous random vector. Assume that $X \times Y(\Omega)$ is a rectangle $]a,b[\times]c,d[$. Then X and Y are independent iff we can write

$$f_{X,Y}(x,y) = g(x)h(y)$$

with $g(x) > 0$ for $x \in]a,b[$ and $h(y) > 0$ for $y \in]c,d[$.

Exemple : Circular symmetry We say that the joint density of two rv X and Y is radial if it depends only on the distance, that is

$$f(x,y) = g(r) \text{ with } r = \sqrt{x^2+y^2}$$

Let us show that if the rv X and Y are circular symmetric and independent, then they are normal laws with null mean and equal variances.
Indeed, the independance implis that

$$g(\sqrt{x^2+y^2}) = f_X(x)f_Y(y)$$

Deriving this relation wrt x, and dividing the result by $xg(r) = xf_X(x)f_Y(y)$, we get

$$\frac{1}{r}\frac{g'(r)}{g(r)} = \alpha = \text{ constant}$$

thus we deduce that $g(r) = Ae^{\alpha r^2/2}$, and going back to f, we can show that X and Y are normal laws with null mean and variance $\sigma^2 = -1/\alpha$. $\odot$

3.3 Bi-normal law: introduction

Definition 3.3.1

We say that two rv are jointly normal or bi-normal if their joint density is given by

$$f(x,y) = Aexp\{-\frac{1}{2(1-r^2)}[\frac{(x-\mu_1)}{\sigma_1^2} - 2r\frac{(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2} + \frac{(y-\mu_2)^2}{\sigma_2^2}]\}$$

where the constant A is given

$$A = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-r^2}} \text{ avec } |r| < 1$$

and the other parameters are given.

We denote $(X, Y) \sim N(\mu_1, \mu_2; \sigma_1^2, \sigma_2^2; r)$.

Note that the quadratic form in the exponential function is a negative quadratic form, because $|r| < 1$.

One can show that μ_1 et μ_2 are the means of X and Y resp, while σ_1^2 and σ_2^2 are their resp. variances. The interpretation of the parameter r will be given later on (this is in fact the correlation coefficient).

All these results amount to show that the marginal densities of X and Y are given by

$$f_X(x) = \frac{1}{\sigma_1\sqrt{2\pi}} e^{-(x-\mu_1)^2/2\sigma_1^2} \text{ et } f_Y(y) = \frac{1}{\sigma_2\sqrt{2\pi}} e^{-(y-\mu_2)^2/2\sigma_2^2}$$

This fact comes by using the definition and integrating wrt x or y, and by writing that the term in brackets can be written as

$$[\ ] = [\frac{x-\mu_1}{\sigma_1} + \frac{y-\mu_2}{\sigma_2}]^2 + (1-r^2)\frac{(y-\mu_2)^2}{\sigma_2^2}$$

Remarque : We have seen that, if X and Y are bi-normal, then their marginal laws X and Y are (separately) normal. The reciproque is not true.

To see this,we shall construc two rv X_1 and Y_1 which are marginally normal but not bi-normal.

We start from two rv X and Y bi-normal, with a joint density $f(x, y)$ given by the previous definition.

Consider the set D made by four symmetric small disks located in each of the $1/4$ part of the plane. For small enough ε, introduce a new function $f_1(x, y)$ being $f_1(x, y) = f(x, y) \pm \varepsilon$ in D and $f(x, y)$ outside. Note that f_1 is a density by construction. So it can be associated with two new rv X_1 and Y_1.

Note that X_1 and Y_1 are not bi-normal, as their joint density f_1 cannot be written as a negative exponential. On the other hand, X_1 and Y_1 are marginally normal. Indeed, it is enough to recall that we just need to integrate wrt one of the variables x or y. ■

3.4 The case of a discrete couple of rv

Let $Z = (X, Y)$ be a 2d random vector, so that que $Z(\Omega)$ is at most countable. Then we can write

$$Z(\Omega)X \times Y(\Omega) = \{(x_j, y_k), ...\}$$

The notion of distribution function is defined similarly and in factLa notion de fonction de répartition se définit de la même facon, et plus précisement,

Proposition 3.4.1

If (X, Y) is a discrete random vector, then

$$F_{X,Y}(x, y) = \sum_{x_j \leq x, y_k \leq y} p_{X,Y}(x_j, y_k)$$

Definition 3.4.1

If (X,Y) is a discrete random vector, the joint probability mass function of this vector is defined by

$$p_{X,Y}(x_j, y_k) = P(X = x_j, Y = y_k)$$

These are positive numbers. Summation over the two indices is equal to 1.
If A is an event wrt $Z(\Omega)$, that $A \subset Z(\Omega)$, then

$$P(A) = \sum_{(x_j,y_k)\in A} p_{X,Y}(x_j, y_k)$$

Using the total probability rule, we get

$$p_X(x_j) = \sum_k P(X = x_j, Y = y_k) = \sum_k p_{X,Y}(x_j, y_k)$$

et de même

$$p_Y(y_k) = \sum_j P(X = x_j, Y = y_k) = \sum_j p_{X,Y}(x_j, y_k)$$

Functions p_X and p_y are called the fmarginal probability pass functions; they are obtained by summing iver one of the two indices.

Exemple: A box contains 6 transistors, among them 1 from brand A and 1 from brand B. Two transistors are taken at random and with reset. Let X, resp. Y, the number of transistors with brand A, resp, brand B among the two taken out. Set Z= (X,Y). Then the joint probability mass function $p_{X,Y}$ is given by the following table:

$y \mid x$	0	1	2
0	16/36	8/36	1/36
1	8/36	2/36	0
2	1/36	0	0

Note that the rv X and Y both follow a binomial law $B(n = 2, p = 1/6)$. Computing, we find for example that $P(X + Y \geq 1) = 5/9$.

⊙

Proposition 3.4.2

(X,Y) are two independent rv iff

$$p_{X,Y}(x_j, y_k) = p_X(x_j)p_Y(y_k)$$

3.5 Moments of a couple of rv

Let X and Y be two rv, let a function $g(x, y) : \mathbb{R}^2 \to \mathbb{R}$. Introduce the rv

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

Then the expectation of Z is given by

$$E(z) = \int_{-\infty}^{+\infty} z f_Z(z) dz$$

which seems to require the computation of f_Z in terms of the joint density $f_{X,Y}$. In fact, it is often useless because

Theorem 3.5.1

Let (X, Y) be a (continuous) randome vector, and $Z = g(X, Y)$. The expectation of Z is then given by

$$E(Z) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) f_{X,Y}(x, y) dxdy$$

This is coherent is the sense that if g depends only on x, then

$$E(g(X)) = \int \int g(x) f_{X,Y}(x, y) dxdy = \int g(x) f_X(x) dx$$

and we recover a known formula.

Remarque : If (X, Y) is discrete, then

$$E(Z) = \sum_{j=1}^{\infty} \sum_{k=1}^{+\infty} g(x_j, y_k) p_{X,Y}(x_j, y_k)$$

■

Remarque : We recover that the expectation is linear:

$$E(X + Y) = E(X) + E(Y)$$

and more generally that

$$E(\sum_{k=1}^{n} a_k g_k(X, Y)) = \sum_{k=1}^{n} a_k E(g_k(X, Y))$$

Note that in general

$$E(XY) \neq E(X)E(Y)$$

However, if X and Y are independent rv, then for all functions g_1 and g_2, we have

$$E(g_1(X) g_2(Y)) = E(g_1(X)) E(g_2(Y))$$

■

Definition 3.5.1

The correlation of X and Y is defined to be the number $E(XY)$. If this correlation is zero, we say *that the rv X and Y are orthogonal, denoted by $E \perp T$.*

The covariance of X and Y is defined by

$$Cov(X,Y) = \sigma_{X,Y} = E((X - E(X))(Y - E(Y))) = Cov(X,Y) = E(XY) - E(X)E(Y)$$

The correlation coefficient is defined by

$$Corr(X,Y) = \varrho_{X,Y} = \frac{cov(X,Y)}{\sigma_X \sigma_Y}$$

Remarque : i) Note that $Cov(X,X) = var(X)$. Thus the covariance generalizes the variance, but could change signe.

ii) If X and Y are independent, then $Cov(X,Y) = 0$.

ii') Note that the rv X et Y on one hand, and $X - \mu_X$ and $Y - \mu_Y$ on the other hand, have the same covariances and correlation coefficients.

iii) The correlation coefficient is a measure without units, measuring the linear link between X and Y. Ona can show that $|\varrho_{X,Y}| \leq 1$. Moreover if $Y = aX + b$, then $|\varrho_{X,Y}| = 1$, that is $|Cov(X,Y)| \leq \sigma_X \sigma_Y$.

iv) If X and Y are independent, then $\varrho_{X,Y} = 0$. The converse is not always true. If $\varrho_{X,Y} = 0$ but X and Y are not independent, we say that they are simply non correlated, that is

$$Cov(X,Y) = 0 \text{ that is } \varrho = 0 \text{ that is } E(XY) = E(X)E(Y)$$

For e.g. for $X \sim U(-1,1)$ and $Y \equiv X^2$, we find that they are not correlated. But thet of course not independent.

iv) If X and Y are not correlated, then $X - \mu_X$ et $Y - \mu_Y$ are orthogonal. If X and Y are not correlated, and if μ_X or μ_Y is zero, then X and Y are orthogonal. ■

Exemple : Let

$$f_{X,Y}(x,y) = 2 \text{ if } -y < x < y, 0 < y < 1$$

and 0 elsewhere. We find

$$f_X(x) = 1 - |x| \text{ if } -1 < x < 1 \text{ and } f_Y(y) = 2y \text{ if } 0 < y < 1$$

We find also that $E(X) = 0$ and that the two rv are not correlated. On the other hand, one can show that they are not independent, as

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

⊙

Exemple : Let us show that the correlation coefficient of a bi-normal couple (X,Y) is the parameter r which appears in the density. This explains why from now on, we shall denote it by ϱ, and no more by r.

With the previous remarks, we may assume that $\mu_X = \mu_Y = 0$. In this case as $Cov(X,Y) = E(XY)$, it is enough to show that $E(XY) = r\sigma_1\sigma_2$, to get the result on ϱ. As

$$\frac{x^2}{\sigma_1^2} - 2r\frac{xy}{\sigma_1\sigma_2} + \frac{y^2}{\sigma_2^2} = \left(\frac{x}{\sigma_1} - r\frac{y}{\sigma_2}\right)^2 + (1 - r^2)\frac{y^2}{\sigma_2^2}$$

we get

$$E(XY) = \frac{1}{\sigma_2\sqrt{2\pi}} \int y e^{-y^2/2\sigma_2^2} \int \frac{x}{\sigma_1\sqrt{2\pi(1-r^2)}} exp[-\frac{(x - ry\sigma_1/\sigma_2)^2}{2\sigma_1^2(1-r^2)}]dxdy$$

The inner integral is a normal density with mean $ry\sigma_1/\sigma_2$ multiplied by x. Thus it is equal to $ry\sigma_1/\sigma_2$. Thus

$$E(XY) = \frac{r\sigma_1/\sigma_2}{\sigma_2\sqrt{2\pi}} \int y^2 e^{-y^2/2\sigma_2^2} dy = r\sigma_1\sigma_2$$

⊙

Proposition 3.5.1

If X and Y are independent, that is if

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

then they are not correlated.

This follows from the fact that in that case $E(XY) = E(X)E(Y)$. More generally, if X and Y are independent, then we have also

$$E(g(X)h(Y)) = E(g(X))E(h(Y))$$

which is not true if we assume only that X and Y are not correlated.

Remarque : If two rv are not correlated, they are not necessarily independent.

However, for a bi-normal couple of rv, non correlation is equivalent to independence. Indeed, if X and Y are two bi-normal rv, with $\varrho = r = 0$, then $f(x,y) = f_X(x)f_Y(y)$. ■

To find joint statistics of X and Y, we need a priori to know their joint density. In practise, we know only their joint mean and variances, that is we only the five parameters

$$\mu_X, \mu_Y, \sigma_X, \sigma_Y \text{ et } \varrho_{X,Y}$$

If X and Y are bi-normal, then these five parameters suffice to determine uniquely $f(x,y)$.

Exemple : Assume that the va X and Y are bi-normal, with

$$\mu_X = 10, \mu_Y = 0, \sigma_X = 2, \sigma_Y = 1 \text{ et } \varrho_{X,Y} = 0,5$$

Let us look to the joint density of

$$Z = X + Y \text{ and } W = X * Y$$

We find

$$\mu_X = \mu_X + \mu_Y = 10,\ \mu_W = \mu_X - \mu_Y = 10$$

$$\sigma_Z^2 = \sigma_X^2 + \sigma_Y^2 + 2\varrho_{X,Y}\sigma_X\sigma_Y = 7,\ \sigma_W^2 = \sigma_X^2 + \sigma_Y^2 - 2\varrho_{X,Y}\sigma_X\sigma_Y = 3$$

$$E(ZW) = E(X^2 - Y^2) = 100 + 4 - 1 = 103$$

$$\varrho_{Z,W} = \frac{E(ZW) - E(Z)E(W)}{\sigma_Z\sigma_W} = \frac{3}{\sqrt{7 \times 3}}$$

Moreover, we know that Z and W are bi-normal, as they are linearly depending on X and Y. Thus their joint density is

$$N(10,10;7,3,\sqrt{3/7})$$

⊙

Variance

Let

$$Z = a_0 + a_1 X_1 + ... + a_n X_n$$

where the a_i are given constants, and the X_i rv.

One can show that

Proposition 3.5.2

We have

$$Var(Z) = \sum_{k=1}^{n} a_k^2 Var(X_k) + 2 \sum_{i=1}^{n} \sum_{k=1, i<k}^{n} a_i a_k Cov(X_i, X_k)$$

Note that the constant a_0 does not playany role in the variance of Z. We may also write the above formula as

$$Var(Z) = \sum_{i=1}^{n} \sum_{k=1}^{n} a_i a_k Cov(X_i, X_k)$$

Particular cases;

i) If the rv X_k are independent (or even only non correlated), then

$$Var(Z) = \sum_{k=1}^{n} a_k^2 Var(X_k)$$

ii) Assume that the rv X_k sont i.i.d (standing for independent and dientically distributed, that is with same distribution function. Then, if

$$S_n \equiv X_1 + ... + X_n$$

we have

$$E(S_n) = nE(X_1) \text{ and } Var(S_n) = nVar(X_1)$$

Remarque : On the whole, we have

$$E(X + Y) = E(X) + E(Y)$$

If moreover the rv are independent, then

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

and

$$Var(X + Y) =_{ind} Var(X) + Var(Y)$$

Note that in general $std(X + Y) \neq std(X) + std(Y)$ even if X and Y are independent. ■

Remark 3.5.1

We have seen that if X and Y were bi-normal, then the sum $aX + bY$ is also normal. We may show also the following special case: if X and Y are independent and normal, then their sum $X + Y$ is also normal. In fact, we have the more difficult result (Cramer): if the rv X and Y are independent,

if their sum is normal, then they are also normal.
⊙

One can also show that: if we know that the sum $aX + bY$ is normal for all a and b, then the rv X and Y are bi-normal. This is not true if we only admit a finite number of values for a and b.