
Course:	Mathematical Statistics
Term:	Fall 2017
Instructor:	Gordan Žitković

Lecture 5

Probability review - joint distributions

5.1 Random vectors

So far we talked about the distribution of a single random variable Y . In the discrete case we used the notion of a pmf (or a probability table) and in the continuous case the notion of the pdf, to describe that distribution and to compute various related quantities (probabilities, expectations, variances, moments).

Now we turn to distributions of several random variables put together. Just like several real numbers in an order make a vector, so n random variables, defined in the same setting (same probability space), make a **random vector**. We typically denote the components of random vectors by subscripts, so that (Y_1, Y_2, Y_3) make a typical 3-**dimensional** random vector. We can also think of a random vector as a random point in an n -dimensional space. This way, a random pair (Y_1, Y_2) can be thought of as a random point in the plane, with Y_1 and Y_2 interpreted as its x - and y -coordinates.

There is a significant (and somewhat unexpected) difference between the distribution of a random vector and the pair of distributions of its components, taken separately. This is not the case with non-random quantities. A point in the plane is uniquely determined by its (two) coordinates, but the distribution of a random point in the plane is not determined by the distributions of its projections onto the coordinate axes. The situation can be illustrated by the following example:

Example 5.1.1.

1. Let us toss two unbiased coins, and let us call the outcomes Y_1 and Y_2 . Assuming that the tosses are unrelated, the probabilities of the following four outcomes

$$\{Y_1 = H, Y_2 = H\}, \{Y_1 = H, Y_2 = T\}, \\ \{Y_1 = T, Y_2 = H\}, \{Y_1 = T, Y_2 = T\}$$

are the same, namely $1/4$. In particular, the probabilities that the first coin lands on H or T are the same, namely $1/2$. The distribution

tables for both Y_1 and Y_2 are the same and look like this

	H	T
	$1/2$	$1/2$

Let us now repeat the same experiment, but with two coins attached to each other (say, welded together) as in the picture:



Figure 1. Two quarters welded together, so that when one falls on heads the other must fall on tails, and vice versa.

We can still toss them and call the outcome of the first one Y_1 and the outcome of the second one Y_2 . Since they are welded, it can never happen that $Y_1 = H$ and $Y_2 = H$ at the same time, or that $Y_1 = T$ and $Y_2 = T$ at the same time, either. Therefore of the four outcomes above only two “survive”

$$\{Y_1 = H, Y_2 = H\}, \{Y_1 = H, Y_2 = T\}, \\ \{Y_1 = T, Y_2 = H\}, \{Y_1 = T, Y_2 = T\}$$

and each happens with the probability $\frac{1}{2}$. The distribution of Y_1 considered separately from Y_2 is the same as in the non-welded case, namely

	H	T
	$1/2$	$1/2$

and the same goes for Y_2 . This is one of the simplest examples, but it already strikes the heart of the matter: randomness in one part of the system may *depend* on the randomness in the other part.

- Here is an artistic (geometric) view of an analogous phenomenon. The projections of a 3D object on two orthogonal planes do not

determine the object entirely. Sculptor Markus Raetz used that fact to create the sculpture entitled “Yes/No”:

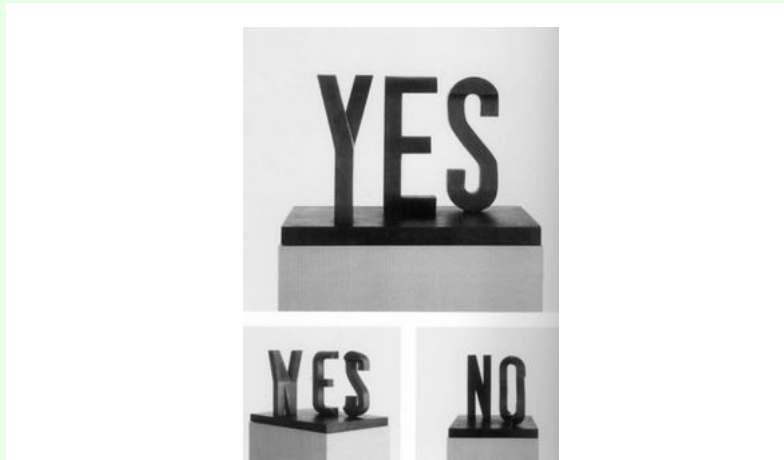


Figure 2. Yes/No - A “typographical” sculpture by Markus Raetz

Not to be outdone, I decided to create a different typographical sculpture with the same projections (I could not find the exact same font Markus is using so you need to pretend that my projections match his completely). It is not hard to see that my sculpture differs significantly from Marcus’s, but they both have (almost) the same projections, namely the words “Yes” and “No”.

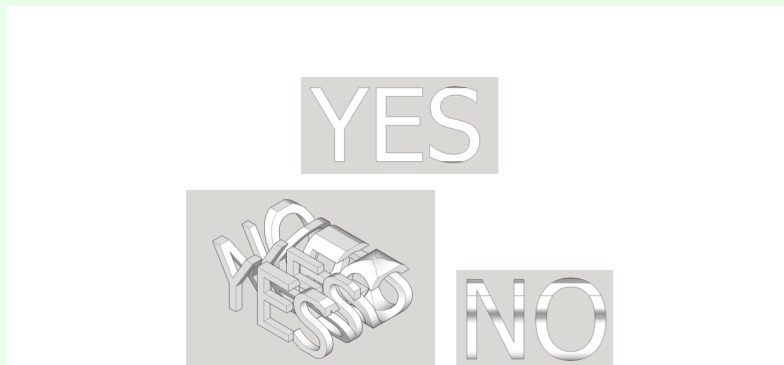


Figure 3. My own attempt at a “typographical” sculpture, using *SketchUp*. You should pretend that Markus’s and mine fonts are the same.

5.2 Joint distributions - the discrete case

So, in order to describe the distribution of the random vector (Y_1, \dots, Y_n) , we need more than just individual distributions of its components $Y_1 \dots Y_n$. In the discrete case, the events whose probabilities finally made their way into the distribution table were of the form $\{Y = i\}$, for all i in the support \mathcal{S}_Y of Y . For several random variables, we need to know their **joint distribution**, i.e., the probabilities of all combinations

$$\{Y_1 = i_1, Y_2 = i_2, \dots, Y_n = i_n\},$$

over the set of all combinations (i_1, i_2, \dots, i_n) of possible values our random variables can take. These numbers cannot comfortably fit into a table, except in the case $n = 2$, where we talk about the **joint distribution table** which looks like this

	j_1	j_2	\dots
i_1	$\mathbb{P}[Y_1 = i_1, Y_2 = j_1]$	$\mathbb{P}[Y_1 = i_1, Y_2 = j_2]$	\dots
i_2	$\mathbb{P}[Y_1 = i_2, Y_2 = j_1]$	$\mathbb{P}[Y_1 = i_2, Y_2 = j_2]$	\dots
\vdots	\vdots	\vdots	\ddots

Example 5.2.1. Two dice are thrown (independently of each other) and their outcomes are denoted by Y_1 and Y_2 . Since $\mathbb{P}[Y_1 = i, Y_2 = j] = 1/36$ for any $i, j \in \{1, 2, \dots, 6\}$, the joint distribution table of (Y_1, Y_2) looks like this

	1	2	3	4	5	6
1	1/36	1/36	1/36	1/36	1/36	1/36
2	1/36	1/36	1/36	1/36	1/36	1/36
3	1/36	1/36	1/36	1/36	1/36	1/36
4	1/36	1/36	1/36	1/36	1/36	1/36
5	1/36	1/36	1/36	1/36	1/36	1/36
6	1/36	1/36	1/36	1/36	1/36	1/36

The situation is more interesting if Y_1 still denotes the outcome of the first die, but Z now stands for the *sum* of the numbers on two dies. It is not hard to see that the joint distribution table of (Y_1, Z) now looks like this:

	2	3	4	5	6	7	8	9	10	11	12
1	1/36	1/36	1/36	1/36	1/36	1/36	0	0	0	0	0
2	0	1/36	1/36	1/36	1/36	1/36	1/36	0	0	0	0
3	0	0	1/36	1/36	1/36	1/36	1/36	1/36	0	0	0
4	0	0	0	1/36	1/36	1/36	1/36	1/36	1/36	0	0
5	0	0	0	0	1/36	1/36	1/36	1/36	1/36	1/36	0
6	0	0	0	0	0	1/36	1/36	1/36	1/36	1/36	1/36

Going from the joint distribution of the random vector (Y_1, Y_2, \dots, Y_n) to individual (called **marginal**) distributions of Y_1, Y_2, \dots, Y_n is easy. To compute $\mathbb{P}[Y_1 = i]$ we need to sum $\mathbb{P}[Y_1 = i, Y_2 = i_2, \dots, Y_n = i_n]$, over all combinations (i_2, \dots, i_n) where i_2, \dots, i_n range through all possible values Y_2, \dots, Y_n can take.

Example 5.2.2. Continuing the previous example, let us compute the marginal distribution of Y_1, Y_2 and Z . For Y_1 we sum the probabilities in each row in the joint distribution table of (Y_1, Y_2) to obtain

	1	2	3	4	5	6
	1/6	1/6	1/6	1/6	1/6	1/6

The same table is obtained for the marginal distribution of Y_2 (even though we sum over columns this time). For the marginal distribution of Z , we use the joint distribution table for (Y_1, Z) and sum over columns:

	2	3	4	5	6	7	8	9	10	11	12
	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36

Once the distribution table of a random vector (Y_1, \dots, Y_n) is given, we can compute (in theory) the probability of any event concerning the random variables Y_1, \dots, Y_n , but simply summing over the set of appropriate entries in the joint distribution.

Example 5.2.3. We continue with random variables Y_1, Y_2 and Z defined above and ask the following question: what is the probability that two dice have the same outcome? In other words, we are interested in $\mathbb{P}[Y_1 = Y_2]$? The entries in the table corresponding to this event are boxed:

	1	2	3	4	5	6
1	1/36	1/36	1/36	1/36	1/36	1/36
2	1/36	1/36	1/36	1/36	1/36	1/36
3	1/36	1/36	1/36	1/36	1/36	1/36
4	1/36	1/36	1/36	1/36	1/36	1/36
5	1/36	1/36	1/36	1/36	1/36	1/36
6	1/36	1/36	1/36	1/36	1/36	1/36

so that

$$\mathbb{P}[Y_1 = Y_2] = 1/36 + 1/36 + 1/36 + 1/36 + 1/36 + 1/36 = 1/6.$$

5.3 Joint distributions - the continuous case

Just as in the univariate case (the case of a single random variable), the continuous analogue of the distribution table (or the pmf) is the pdf. Recall that pdf $f_Y(y)$ of a single random variable Y is the function with the property that

$$\mathbb{P}[Y \in [a, b]] = \int_a^b f_Y(y) dy.$$

In the multivariate case (the case of a random vector, i.e., several random variables), the pdf of the random vector (Y_1, \dots, Y_n) becomes a function of several variables $f_{Y_1, \dots, Y_n}(y_1, \dots, y_n)$ and it is characterized by the property that

$$\begin{aligned} \mathbb{P}[Y_1 \in [a_1, b_1], Y_2 \in [a_2, b_2], \dots, Y_n \in [a_n, b_n]] &= \\ &= \int_{a_1}^{b_1} \int_{a_2}^{b_2} \dots \int_{a_{n-1}}^{b_{n-1}} \int_{a_n}^{b_n} f_{Y_1, \dots, Y_n}(y_1, \dots, y_n) dy_n dy_{n-1} \dots dy_2 dy_1. \end{aligned}$$

This formula is better understood if interpreted geometrically. The left-hand side is the probability that the random vector (Y_1, \dots, Y_n) (think of it as a random point in \mathbb{R}^n) lies in the region $[a_1, b_1] \times \dots \times [a_n, b_n]$, while the right-hand side is the integral of f_{Y_1, \dots, Y_n} over the same region.

Example 5.3.1. A point is randomly and uniformly chosen inside a square with side 1. That means that any two regions of equal area inside the square have the same probability of containing the point. We denote the two coordinates of this point by Y_1 and Y_2 (even though X and Y would be more natural), and their joint pdf by f_{Y_1, Y_2} . Since the probabilities are computed by integrating f_{Y_1, Y_2} over various regions in the square, there is no reason for f to take different values on different points inside the square; this makes $f_{Y_1, Y_2}(y_1, y_2) = c$ for some constant $c > 0$, for all points (y_1, y_2) in the square. Our random point never falls outside the square, so the value of f outside the square should be 0. Pdfs (either in one or in several dimensions) integrate to 1, so we conclude that f should be given by

$$f_{Y_1, Y_2}(y_1, y_2) = \begin{cases} 1, & (y_1, y_2) \in [0, 1]^2 \\ 0, & \text{otherwise.} \end{cases}$$

Once a pdf of a random vector (Y_1, \dots, Y_n) is given, we can compute all kinds of probabilities with it. For any region $A \subset \mathbb{R}^n$ (not only for rectangles of the form $[a_1, b_1] \times \dots \times [a_n, b_n]$), we have

$$\mathbb{P}[(Y_1, \dots, Y_n) \in A] = \int \dots \int_A f_{Y_1, \dots, Y_n}(y_1, \dots, y_n) dy_n \dots dy_1.$$

As it is almost always the case in the multivariate setting, this is much better understood through an example:

Example 5.3.2. Let (Y_1, Y_2) be the random uniform point in the square $[0, 1]^2$ from the previous example. To compute the probability that the distance from (Y_1, Y_2) to the origin $(0, 0)$ is at most 1, we define

$$A = \left\{ (y_1, y_2) \in [0, 1]^2 : \sqrt{y_1^2 + y_2^2} \leq 1 \right\},$$

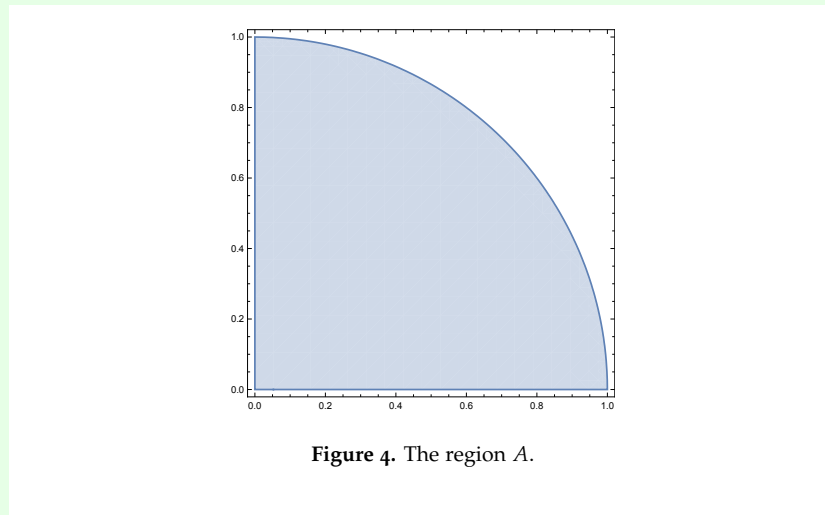


Figure 4. The region A .

Therefore, since $f_{Y_1, Y_2}(y_1, y_2) = 1$, for all $(y_1, y_2) \in A$, we have

$$\begin{aligned} \mathbb{P}[(Y_1, Y_2) \text{ is at most 1 unit away from } (0,0)] &= \mathbb{P}[(Y_1, Y_2) \in A] \\ &= \iint_A f_{Y_1, Y_2}(y_1, y_2) dy_2 dy_1 = \iint_A 1 dy_1 dy_2 = \text{area}(A) = \frac{\pi}{4}. \end{aligned}$$

The calculations in the previous example sometimes fall under the heading of *geometric probability* because the probability $\pi/4$ we obtained is simply the ratio of the area of A and the area of $[0, 1]^2$ (just like one computes a uniform probability in a finite setting by dividing the number of “favorable” cases by the total number). This works only if the underlying pdf is uniform. In practice, pdfs are rarely uniform.

Example 5.3.3. Let (Y_1, Y_2) be a random vector with the pdf

$$f_{Y_1, Y_2}(y_1, y_2) = \begin{cases} 6y_1, & 0 \leq y_1 \leq y_2 \leq 1 \\ 0, & \text{otherwise,} \end{cases}$$

or, in the indicator notation,

$$f_{Y_1, Y_2}(y_1, y_2) = 6y_1 \mathbf{1}_{\{0 \leq y_1 \leq y_2 \leq 1\}}.$$

Here is a sketch of what f looks like

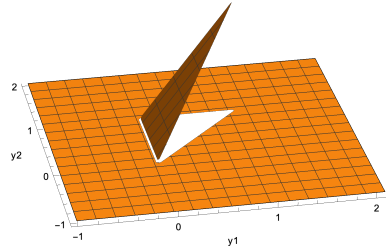


Figure 5. The pdf of (Y_1, Y_2)

This still corresponds to a distribution of a random point in the unit square, but this distribution is no longer uniform - the point can only appear in the upper left triangle, and the larger the value of y_1 the more likely the point. To compute, e.g., the probability $\mathbb{P}[Y_1 \geq 1/2, Y_2 \geq 1/2]$ we integrate f_{Y_1, Y_2} over the region $[1/2, 1] \times [1/2, 1]$:

$$\mathbb{P}[Y_1 \geq 1/2, Y_2 \geq 1/2] = \int_{1/2}^1 \int_{1/2}^1 6y_1 \mathbf{1}_{\{0 \leq y_1 \leq y_2 \leq 1\}} dy_2 dy_1.$$

We would like to compute this double integral as an iterated integral, i.e., by evaluation the “inner” y_2 -integral first, and then integrate the result with respect to y_1 . The indicator notation helps us with the bounds. Indeed, when doing the inner integral, we can think of y_1 as a constant and interpret the indicator function as a function of y_1 only; it will tell us to integrate from y_1 to 1. Since the outer integral is over the region in which $y_1 \geq 1/2$, the inner integral is given by

$$\int_{1/2}^1 6y_1 \mathbf{1}_{\{0 \leq y_1 \leq y_2 \leq 1\}} dy_2 = 6y_1 \int_{y_1}^1 dy_2 = 6y_1(1 - y_1).$$

It remains to integrate the obtain result from 1/2 to 1:

$$\mathbb{P}[Y_1 \geq 1/2, Y_2 \geq 1/2] = \int_{1/2}^1 6y_1(1 - y_1) dy_1 = \frac{1}{2}.$$

the integrals become more complicated, as the following example shows. It also sheds some light on the usefulness of indicators.

Example 5.3.4. Suppose that the pair (Y_1, Y_2) has a pdf as above given by

$$f_{Y_1, Y_2}(y_1, y_2) = 6y_1 \mathbf{1}_{\{0 \leq y_1 \leq y_2 \leq 1\}}.$$

Let us compute now the same probability as in Example (5.3.1), namely that the point (Y_1, Y_2) is at most 1 unit away from $(0,0)$. It is still the case that

$$\begin{aligned} \mathbb{P}[(Y_1, Y_2) \text{ is at most 1 unit away from } (0,0)] &= \mathbb{P}[(Y_1, Y_2) \in A] = \\ &= \iint_A f_{Y_1, Y_2}(y_1, y_2) dy_2 dy_1, \end{aligned}$$

but, since f is no longer uniform, this integral is no longer just the area of A . We use the indicator notation to write

$$f_{Y_1, Y_2}(y_1, y_2) = 6y_1 \mathbf{1}_{\{0 \leq y_1 \leq y_2 \leq 1\}},$$

and replace the integration over A by multiplication by the indicator of A , i.e., $\mathbf{1}_{\{y_1^2 + y_2^2 \leq 1\}}$ to obtain

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} 6y_1 \mathbf{1}_{\{0 \leq y_1 \leq y_2 \leq 1\}} \mathbf{1}_{\{y_1^2 + y_2^2 \leq 1\}} dy_2 dy_1.$$

This can be rewritten as

$$\int_0^1 \int_0^1 6y_1 \mathbf{1}_{\{y_2 \geq y_1, y_2^2 \leq 1 - y_1^2\}} dy_2 dy_1.$$

We do the inner integral, first, and interpret the indicator as if y_1 were a constant:

$$\int_0^1 6y_1 \mathbf{1}_{\{y_2 \geq y_1, y_2 \leq \sqrt{1 - y_1^2}\}} dy_2.$$

The value of this integral is simply $6y_1$ multiplied by the length of the interval $[y_1, \sqrt{1 - y_1^2}]$, when it is nonempty. Graphically, the situation looks like this:

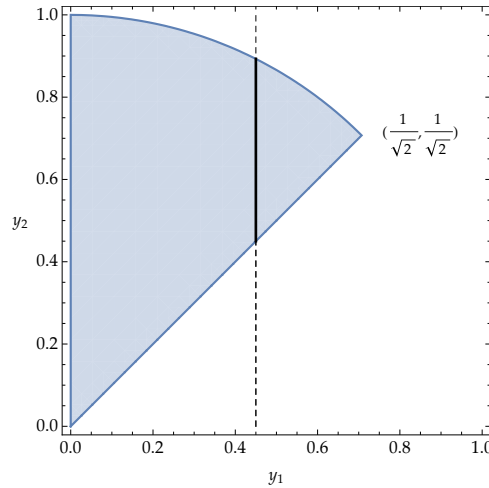


Figure 6. The region where $0 \leq y_1 \leq y_2 \leq 1$ and $y_1^2 + y_2^2 \leq 1$.

The curves $y_2 = y_1$ and $y_2 = \sqrt{1 - y_1^2}$ intersect to the right of 0 at the point $(1/\sqrt{2}, 1/\sqrt{2})$, and, so, the inner integral becomes $6y_1(\sqrt{1 - y_1^2} - y_1)$ for $y_1 \in [0, 1/\sqrt{2}]$ and 0 otherwise, i.e.,

$$\begin{aligned} \int_0^1 6y_1 \mathbf{1}_{\{y_2 \geq y_1, y_2 \leq \sqrt{1-y_1^2}\}} dy_2 &= 6y_1(\sqrt{1-y_1^2} - y_1) \mathbf{1}_{\{0 \leq y_1 \leq 1/\sqrt{2}\}} \\ &= \begin{cases} 6y_1(\sqrt{1-y_1^2} - y_1), & y_1 \in [0, 1/\sqrt{2}] \\ 0, & \text{otherwise.} \end{cases} \end{aligned}$$

We continue with the outer integral

$$\begin{aligned} \int_0^1 6y_1(\sqrt{1-y_1^2} - y_1) \mathbf{1}_{\{0 \leq y_1 \leq 1/\sqrt{2}\}} dy_1 &= \int_0^{1/\sqrt{2}} 6y_1(\sqrt{1-y_1^2} - y_1) dy_1 \\ &= (-2(1-y_1^2)^{3/2} - 2y_1^3) \Big|_0^{1/\sqrt{2}} \\ &= 2 - \sqrt{2}. \end{aligned}$$

Therefore $\mathbb{P}[Y_1^2 + Y_2^2 \leq 1] = 2 - \sqrt{2}$, when (Y_1, Y_2) have the joint distribution $f_{Y_1, Y_2}(y_1, y_2) = 6y_1 \mathbf{1}_{\{0 \leq y_1 \leq y_2 \leq 1\}}$.

So far we focused on computing probabilities involving two random variables. One can compute expected values of such quantities, as well, using a

familiar (fundamental) formula:

Theorem 5.3.5. Let (Y_1, \dots, Y_n) be a continuous random vector, let g be a function of n variables, and let $W = g(Y_1, \dots, Y_n)$. Then

$$\mathbb{E}[W] = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} g(y_1, \dots, y_n) f_{Y_1, \dots, Y_n}(y_1, \dots, y_n) dy_n \dots dy_1,$$

provided the multiple integral on the right is well-defined.

Example 5.3.6. Continuing from Example 5.3.4, the pair (Y_1, Y_2) has a pdf given by

$$f_{Y_1, Y_2}(y_1, y_2) = 6y_1 \mathbf{1}_{\{0 \leq y_1 \leq y_2 \leq 1\}}.$$

Let us compute the expected square of the distance of the point (Y_1, Y_2) to the origin, i.e., $\mathbb{E}[g(Y_1, Y_2)]$, where $g(y_1, y_2) = y_1^2 + y_2^2$. According to Theorem 5.3.5 above, we have

$$\begin{aligned} \mathbb{E}[g(Y_1, Y_2)] &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (y_1^2 + y_2^2) f_{Y_1, Y_2}(y_1, y_2) dy_2 dy_1 \\ &= \int_0^1 \int_0^1 (y_1^2 + y_2^2) 6y_1 \mathbf{1}_{\{y_1 \leq y_2\}} dy_2 dy_1 \\ &= \int_0^1 \int_{y_1}^1 (y_1^2 + y_2^2) 6y_1 dy_2 dy_1 \\ &= \int_0^1 (2y_1 + 6y_1^3 - 8y_1^4) dy_1 = \frac{9}{10}. \end{aligned}$$

5.4 Marginal distributions and independence

While the distributions of the components Y_1 and Y_2 , considered separately, do not tell the whole story about the distribution of the random vector (Y_1, Y_2) , going the other way is quite easy.

Proposition 5.4.1. If the random vector (Y_1, \dots, Y_n) has the pdf $f_{Y_1, \dots, Y_n}(y_1, \dots, y_n)$, then each Y_i is a continuous random variable with the pdf f_{Y_i} given by

$$f_{Y_i}(y) = \int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f_{Y_1, \dots, Y_n}(y_1, \dots, y_{i-1}, y, y_{i+1}, \dots, y_n) dy_n \dots dy_{i+1} dy_{i-1} \dots dy_1.$$

In words, the pdf of the i -th component is obtained by integrating the multivariate pdf f_{Y_1, \dots, Y_n} over $(-\infty, \infty)$ in all variables except for y_i . This is sometimes referred to as **integrating out** the other variables.

Example 5.4.2. Let (Y_1, Y_2) have the pdf

$$f_{Y_1, Y_2}(y_1, y_2) = 6y_1 \mathbf{1}_{\{0 \leq y_1 \leq y_2 \leq 1\}},$$

from example 5.3.4. To obtain the (marginal) pdfs of Y_1 and Y_2 , we follow Proposition 5.4.1 above:

$$\begin{aligned} f_{Y_1}(y) &= \int_{-\infty}^{\infty} f_{Y_1, Y_2}(y, y_2) dy_2 = \int_{-\infty}^{\infty} 6y_1 \mathbf{1}_{\{0 \leq y \leq y_2 \leq 1\}} dy_2 \\ &= 6y(1 - y) \mathbf{1}_{\{0 \leq y \leq 1\}}. \end{aligned}$$

To compute the marginal pdf of Y_2 , we proceed in a similar way

$$\begin{aligned} f_{Y_2}(y) &= \int_{-\infty}^{\infty} f_{Y_1, Y_2}(y_1, y) dy_1 = \int_{-\infty}^{\infty} 6y_1 \mathbf{1}_{\{0 \leq y_1 \leq y \leq 1\}} dy_1 \\ &= \mathbf{1}_{\{0 \leq y \leq 1\}} \int_0^y 6y_1 dy_1 = 3y^2 \mathbf{1}_{\{0 \leq y \leq 1\}}. \end{aligned}$$

The simplest way one can supply the information that can be used to construct the joint pdf from the marginals is to require that the components be *independent*.

Definition 5.4.3. Two random variables Y_1 and Y_2 are said to be **independent** if

$$\mathbb{P}[Y_1 \in [a_1, b_1], Y_2 \in [a_2, b_2]] = \mathbb{P}[Y_1 \in [a_1, b_1]] \times \mathbb{P}[Y_2 \in [a_2, b_2]],$$

for all $a_1 < b_1$ and all $a_2 < b_2$.

One defines the notion of independence for n random variables Y_1, \dots, Y_n by an analogous condition:

$$\begin{aligned} \mathbb{P}[Y_1 \in [a_1, b_1], Y_2 \in [a_2, b_2], \dots, Y_n \in [a_n, b_n]] &= \\ &= \mathbb{P}[Y_1 \in [a_1, b_1]] \times \mathbb{P}[Y_2 \in [a_2, b_2]] \times \dots \times \mathbb{P}[Y_n \in [a_n, b_n]], \end{aligned}$$

for all $a_1 < b_1, a_2 < b_2, \dots$, and all $a_n < b_n$.

Broadly speaking, independence comes into play in two ways:

1. It is a *modeling choice*. That means that the situation modeled makes it plausible that the random variables have nothing to do with each other, and that the outcome of one of them does not affect the outcome of the

other. The basic example here is the case of two different coins tossed separately.

2. It is a *mathematical consequence*. Sometimes, random variables that are defined in complicated ways from other random variables happen to be independent, even though it is intuitively far from obvious. An example here is the fact (which we will talk about later in more detail) that the sample mean and the sample standard deviation in a random sample from a normal distribution are independent, even though both of them are functions of the same set random variables (the sample).

There is an easy way to check whether continuous random variables are independent:

Theorem 5.4.4 (The factorization criterion). *Continuous random variables Y_1, Y_2, \dots, Y_n are independent if and only if*

$$f_{Y_1, \dots, Y_n}(y_1, y_2, \dots, y_n) = f_{Y_1}(y_1)f_{Y_2}(y_2) \dots f_{Y_n}(y_n),$$

for all y_1, \dots, y_n , where f_{Y_1}, \dots, f_{Y_n} are the marginal pdfs of Y_1, \dots, Y_n .

It gets even better. One does not need to compute the marginals to apply the theorem above:

Theorem 5.4.5 (The factorization criterion 2.0). *Continuous random variables Y_1, Y_2, \dots, Y_n are independent if and only if there exist nonnegative functions f_1, f_2, \dots, f_n (which are not necessarily the marginals) such that*

$$f_{Y_1, \dots, Y_n}(y_1, y_2, \dots, y_n) = f_1(y_1)f_2(y_2) \dots f_n(y_n),$$

for all y_1, \dots, y_n .

A similar factorization criterion holds for discrete random variables, as well. One simply needs to replace pdfs by pmfs.

Example 5.4.6.

1. In the example of a random point chosen uniformly over the square $[0, 1]^2$, the pdf of the two coordinates Y_1, Y_2 was given by the expression

$$f_{Y_1, Y_2}(y_1, y_2) = \mathbf{1}_{\{y_1, y_2 \in [0, 1]\}} = \mathbf{1}_{\{0 \leq y_1 \leq 1\}} \mathbf{1}_{\{0 \leq y_2 \leq 1\}}.$$

the functions $f_1(y_1) = \mathbf{1}_{\{0 \leq y_1 \leq 1\}}$ and $f_2(y_2) = \mathbf{1}_{\{0 \leq y_2 \leq 1\}}$ have the property that $f_{Y_1, Y_2}(y_1, y_2) = f_1(y_1)f_2(y_2)$. Therefore, the factorization criterion 2.0 can be used to conclude that Y_1 and Y_2 are independent. This makes intuitive sense. If we are told the x -coordinate

of this point, we are still just as ignorant about its y coordinate as before.

2. Consider now the case where the distribution is no longer uniform, but comes with the pdf

$$f_{Y_1, Y_2}(y_1, y_2) = 6y_1 \mathbf{1}_{\{0 \leq y_1 \leq y_2 \leq 1\}}.$$

If we forgot the indicator, the remaining part, namely $6y_1$, can be easily factorized. Indeed, $6y_1 = f_1(y_1)f_2(y_2)$, where $f_1(y_1) = 6y_1$ and $f_2(y_1) = 1$. The presence of the indicator, however, prevents us from doing the same for f_{Y_1, Y_2} . We have already computed the marginals f_{Y_1} and f_{Y_2} in Example 5.4.2; if we multiply them together, we obtain

$$\begin{aligned} f_{Y_1}(y_1)f_{Y_2}(y_2) &= 6y_1(1 - y_1)\mathbf{1}_{\{0 \leq y_1 \leq 1\}}3y_2^2\mathbf{1}_{\{0 \leq y_2 \leq 1\}} \\ &= 18y_1y_2^2\mathbf{1}_{\{0 \leq y_1, y_2 \leq 1\}}, \end{aligned}$$

which is clearly not equal to f_{Y_1, Y_2} . Using the original factorization criterion, we may conclude that Y_1 and Y_2 are not independent. Like above, this makes perfect intuitive sense. The indicator $\mathbf{1}_{\{0 \leq y_1 \leq y_2 \leq 1\}}$ forces the value of Y_1 to be below that of Y_2 . Therefore, the information that, e.g., $Y_2 = 0.1$ would change our belief about Y_1 a great deal. We would know with certainty that $Y_1 \in [0, 0.1]$ - a conclusion we would not have been able to reach without the information that $Y_2 = 0.1$.

The factorization theorem(s), i.e., Theorems 5.4.4 and 5.4.5 have another, extremely useful consequence:

Proposition 5.4.7. *Let Y_1, \dots, Y_n be independent random variables, and let g_1, \dots, g_n be functions. Then*

$$\mathbb{E}[g_1(Y_1) \dots g_n(Y_n)] = \mathbb{E}[g_1(Y_1)] \dots \mathbb{E}[g_n(Y_n)],$$

provided all expectations are well defined.

We do not give proofs in these notes, but it is not hard to derive Proposition 5.4.7 in the case of continuous random variables by combining the factorization criterion (Theorem 5.4.4) with the fundamental formula (Theorem 5.3.5).

Example 5.4.8. In the context of the random, uniformly distributed point (Y_1, Y_2) in the unit square, let us compute $\mathbb{E}[\exp(Y_1 + Y_2)]$. One

approach would be to multiply the function $g(y_1, y_2) = e^{y_1+y_2}$ by the pdf f_{Y_1, Y_2} of (Y_1, Y_2) and integrate over the unit square. The other is to realize that Y_1 and Y_2 are independent and use Proposition 5.4.7 with $g_1(y_1) = e^{y_1}$ and $g_2(y_2) = e^{y_2}$

$$\mathbb{E}[e^{Y_1+Y_2}] = \mathbb{E}[e^{Y_1}] \mathbb{E}[e^{Y_2}] = \int_0^1 e^y dy \times \int_0^1 e^y dy = (e-1)^2.$$

5.5 Functions of two random variables

When we talked about a function of a single random variable and considered the transformation $W = g(Y)$, we listed several methods, like the cdf-method and the h -method. These methods can be extended further to the case of several random variables, but we only deal briefly with the case of two random variables in these notes. The special case of the sum of several independent random variable will be dealt with later.

The main approach remains the cdf-method. To find the pdf f_W of the function $W = g(Y_1, Y_2)$ of a pair of random variables with the joint pdf f_{Y_1, Y_2} we write down the expression for the cdf F_W :

$$F_W(w) = \mathbb{P}[W \leq w] = \mathbb{P}[g(Y_1, Y_2) \leq w] = \mathbb{P}[(Y_1, Y_2) \in A],$$

where A is the set of all pairs $(y_1, y_2) \in \mathbb{R}^2$ such that $g(y_1, y_2) \leq w$. Unfortunately, no nice function from \mathbb{R}^2 to \mathbb{R} admits an inverse, so we have to “solve” the inequality $g(y_1, y_2) \leq w$ on a case by case basis. Supposing that we can do that, it remains to remember that

$$\mathbb{P}[(Y_1, Y_2) \in A] = \iint_A f_{Y_1, Y_2}(y_1, y_2) dy_1 dy_2.$$

Here are some examples:

Example 5.5.1. Suppose that Y_1 and Y_2 are both uniformly distributed on $(0, 1)$ and independent. Their joint pdf is then given by

$$f_{Y_1, Y_2}(y_1, y_2) = \mathbf{1}_{\{0 \leq y_1 \leq 1, 0 \leq y_2 \leq 1\}}.$$

If we are interested in the distribution of $W = Y_1 + Y_2$, we need to describe the set $A = \{(y_1, y_2) \in \mathbb{R}^2 : y_1 + y_2 \leq w\}$. These are simply the regions below the 45-degree line $y_2 \leq w - y_1$, passing through the point $(0, w)$ on the y_2 -axis. Since the density f_{Y_1, Y_2} is positive only in the unit square $[0, 1] \times [0, 1]$, we are only interested in the intersection of A with it. For $w > 1$, this intersection is the entire $[0, 1] \times [0, 1]$. For $w < 0$, this intersection is empty. The typical cases with $w \in (0, 1)$ and $w \in (1, 2)$ are given in the two pictures below:

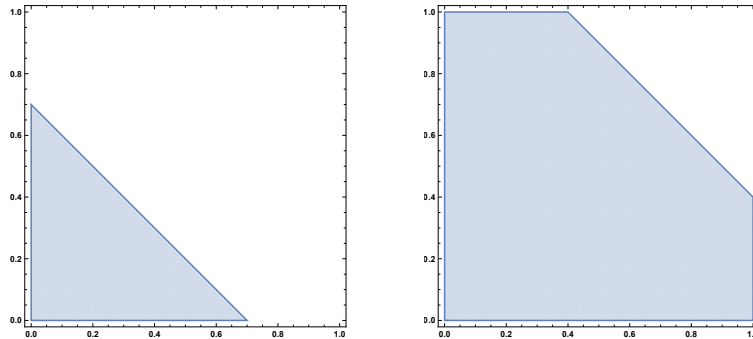


Figure 7. Typical regions of integration for $w \in (0, 1]$ on the left and $w \in (1, 2)$ on the right.

Once we have the region corresponding to $\mathbb{P}[Y_1 + Y_2 \leq x]$, the integral of the pdf over it is not hard to evaluate - indeed, the pdf f_{Y_1, Y_2} is constant over it, with value 1, so all we need to do is calculate its area. In the case $w \in (0, 1]$ we have the area of a right equilateral triangle with sides w and w , making its area $\frac{1}{2}w^2$. In the case $w > 1$, the area under consideration is the area of the square $[0, 1] \times [0, 1]$ minus the area of the (white) triangle in the top right corner, i.e., $1 - (2 - w)^2/2 = -1 + 2w - w^2/2$. Putting all together, we get

$$F_W(w) = \mathbb{P}[W \leq w] = \begin{cases} 0, & w < 0 \\ \frac{1}{2}w^2, & w \in [0, 1] \\ -1 + 2w - \frac{1}{2}w^2, & w \in [1, 2] \\ 1, & w \geq 2 \end{cases}$$

It remains to differentiate $F_W(w)$ to obtain:

$$f_W(w) = \begin{cases} 0, & w < 0 \\ w, & w \in [0, 1] \\ 2 - w, & w \in [1, 2] \\ 0, & w \geq 2 \end{cases}$$

The obtained distribution is sometimes called the **triangle distribution** and its pdf is depicted below:

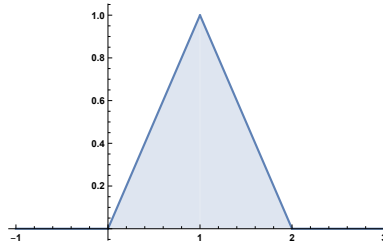


Figure 8. The pdf $f_W(w)$ of the sum $W = Y_1 + Y_2$ of two independent $U(0, 1)$ random variables.

Here is another example which yields a special case of a distribution very important in hypothesis testing in statistics:

Example 5.5.2. Let Y_1 and Y_2 be two independent χ^2 random variables, and let $W = Y_2/Y_1$ be their quotient, i.e., $W = g(y_1, y_2)$ where $g(y_1, y_2) = y_2/y_1$. Remembering that χ^2 takes only positive values and that its pdf is $\frac{1}{\sqrt{2\pi y}}e^{-\frac{1}{2}y}\mathbf{1}_{\{y>0\}}$, we easily obtain that

$$f_{Y_1, Y_2}(y_1, y_2) = \frac{1}{2\pi\sqrt{y_1 y_2}}e^{-\frac{1}{2}(y_1 + y_2)}\mathbf{1}_{\{y_1 > 0, y_2 > 0\}}.$$

To get a handle on the region where $g(y_1, y_2) \leq w$, we first note that it is enough to consider $w > 0$, as W takes only positive values. For such a w , we have

$$\begin{aligned} A &= \{(y_1, y_2) \in (0, \infty) \times (0, \infty) : y_2/y_1 \leq w\} \\ &= \{(y_1, y_2) \in (0, \infty) \times (0, \infty) : y_2 \leq w y_1\}, \end{aligned}$$

so that A is simply the (infinite) region bounded by the y_1 -axis from below and the line $y_2 = w y_1$ from above. We integrate the joint pdf over that region:

$$\begin{aligned} F_W(w) &= \iint_A f_{Y_1, Y_2}(y_1, y_2) dy_1 dy_2 \\ &= \int_0^\infty \frac{1}{\sqrt{2\pi y_1}} e^{-\frac{1}{2}y_1} \int_0^{w y_1} \frac{1}{\sqrt{2\pi y_2}} e^{-\frac{1}{2}y_2} dy_2 dy_1 \\ &= \int_0^\infty \frac{1}{\sqrt{2\pi y_2}} e^{-\frac{1}{2}y_2} F_{Y_1}(w y_2) dy_2 \end{aligned}$$

At this point we are stuck and we do not have a closed form expression for the cdf of the χ^2 -distribution. On the other hand, what we are after

is the pdf and not the cdf of W , so there may be some hope, after all. If we differentiate both sides with respect to w (and accept without proof that we can differentiate inside the integral on the right-hand side), we get, for $w > 0$

$$\begin{aligned} f_W(w) &= \int_0^\infty \frac{1}{\sqrt{2\pi y_2}} e^{-\frac{1}{2}y_2} f_{Y_1}(wy_2) y_2 dy_2 \\ &= \int_0^\infty \frac{1}{2\pi y_2 \sqrt{w}} e^{-\frac{1}{2}(y_2 + wy_2)} f_{Y_1}(wy_2) y_2 dy_2 \\ &= \frac{1}{2\pi\sqrt{w}} \int_0^\infty e^{-\frac{1}{2}(1+w)y_2} dy_2 = \frac{1}{\pi\sqrt{w(1+w)}} \end{aligned}$$

The distribution of the random variable with the pdf $\frac{1}{\pi(1+w)\sqrt{w}} \mathbf{1}_{\{w>0\}}$ is called the **F-distribution** (or, more precisely, the $F(1,1)$ -distribution) and the graph of its pdf is given below:

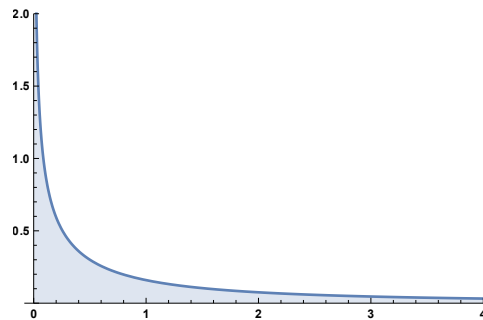


Figure 9. The pdf of the $F(1,1)$ -distribution.

5.6 Problems

Problem 5.6.1. Three (fair and independent) coins are thrown; let Y_1 , Y_2 and Y_3 be the outcomes (encoded as H or T). Player 1 gets \$1 if H shows on coin 1 ($Y_1 = H$) and/or \$2 if H shows on coin 2 ($Y_2 = H$). Player 2, on the other hand, gets \$1 when $Y_2 = H$ and/or \$2 when $Y_3 = H$. With W_1 and W_2 denoting the total amount of money given to Player 1 and Player 2, respectively,

1. Write down the marginal distributions (pmfs) of W_1 and W_2 ,
2. Write down the joint distribution table of (W_1, W_2) .
3. Are W_1 and W_2 independent?

Problem 5.6.2. Let (Y_1, Y_2) be a random vector with the following distribution table

	-1	1
1	$\frac{1}{6}$	$\frac{1}{2}$
2	*	o

If it is known that Y_1 and Y_2 are independent, the values * and o in the second row are

- (a) $* = 1/6, o = 3/4$
- (b) $* = 1/12, o = 1/4$
- (c) $* = 1/6, o = 1/6$
- (d) $* = 1/24, o = 7/24$
- (e) none of the above

Problem 5.6.3. Let $Z_1 \sim N(1, 1)$, $Z_2 \sim N(2, 2)$ and $Z_3 \sim N(3, 3)$ be independent random variables. The distribution of the random variable $W = Z_1 + \frac{1}{2}Z_2 + \frac{1}{3}Z_3$ is

- (a) $N(5/3, 7/6)$
- (b) $N(3, 3)$
- (c) $N(3, \sqrt{3})$
- (d) $N(3, \sqrt{5/3})$
- (e) none of the above

(Note: In our notation $N(\mu, \sigma)$ means normal with mean μ and standard deviation σ .)

Problem 5.6.4. A point is chosen uniformly over a 1-yard wooden stick, and a mark is made. The procedure is repeated, independently, and another mark is made. The stick is then sawn at the two marks, yielding three shorter sticks. What is the probability that at least one of those sticks is at least 1/2 yard long?

Problem 5.6.5. The random vector (Y_1, Y_2) has the pdf

$$f_{Y_1, Y_2}(y_1, y_2) = 6y_1 \mathbf{1}_{\{0 \leq y_1 \leq y_2 \leq 1\}}.$$

Then,

- (a) The pdf of Y_1 is $2y\mathbf{1}_{\{0 \leq y \leq 1\}}$.
- (b) The pdf of Y_2 is $3y^2\mathbf{1}_{\{0 \leq y \leq 1\}}$.
- (c) Y_1 and Y_2 are independent.
- (d) $\mathbb{P}[Y_1 = 1/12, Y_2 = 1/6] = 1/2$
- (e) none of the above

Problem 5.6.6. Let Y_1 and Y_2 be independent exponential random variables with parameters τ_1 and τ_2 .

1. What is the joint density of (Y_1, Y_2) ?
2. Compute $\mathbb{P}[Y_1 \geq Y_2]$.

Problem 5.6.7. A dart player throws a dart at a dartboard - the board itself is always hit, but any region of the board is as likely to be hit as any other of the same area. We model the board as the unit disc $\{y_1^2 + y_2^2 \leq 1\}$, and the point where the board is hit by a pair of random variables (Y_1, Y_2) . This means that (Y_1, Y_2) is uniformly distributed on the unit disc, i.e., the joint pdf is given by

$$f_{Y_1, Y_2}(y_1, y_2) = \frac{1}{\pi} \mathbf{1}_{\{y_1^2 + y_2^2 \leq 1\}} = \begin{cases} \frac{1}{\pi}, & y_1^2 + y_2^2 \leq 1 \\ 0, & \text{otherwise.} \end{cases}$$

1. Do you expect the random variables Y_1 and Y_2 to be independent? Explain why or why not (do not do any calculations).
2. Find the marginal pdfs of Y_1 and Y_2 . Are Y_1 and Y_2 independent?
3. Compute $\mathbb{P}[Y_1 \geq Y_2]$.
4. In a simplified game of darts, the score S associated with the dart falling at the point (Y_1, Y_2) is $S = 1 - (Y_1^2 + Y_2^2)$, i.e., one minus the square of the distance to the origin (bull's eye). Compute the expected score of our player. (Note: In order not to make this a problem on integration, you can use the fact that $\int_{-1}^1 (1 - y_1^2)^{3/2} dy_1 = \frac{3\pi}{8}$.)

Problem 5.6.8. Two random numbers, Y_1 and Y_2 are chosen independently of each other, according to the uniform distribution $U(-1, 2)$ on $[-1, 2]$. The probability that their product is positive is

- (a) 1/3 (b) 2/3 (c) 1/9 (d) 5/9 (e) none of the above

Problem 5.6.9. Let (Y_1, Y_2) have the joint pdf given by

$$f_{Y_1, Y_2}(y_1, y_2) = cy_1y_2\mathbf{1}_{\{0 \leq y_1 \leq 1, y_1 \leq y_2 \leq 2y_1\}}$$

1. What is the value of c ?

2. What are the expected values of Y_1 and Y_2 ?
3. What is the expected value of the product $Y_1 Y_2$?
4. What is the covariance between Y_1 and Y_2 ? Are they independent?

Problem 5.6.10. Let Y_1 and Y_2 be two independent exponential distributions with parameter $\tau = 1$. Find the pdfs of the following random variables:

1. $Y_1 + Y_2$.
2. Y_2/Y_1