

**Probability Calculus 2019/2020**  
**Lecture 4**

1. REAL-VALUED RANDOM VARIABLES

We already know how to describe the results of a random experiment in terms of a formal mathematical construction, i.e. the probability triple  $(\Omega, \mathcal{F}, \mathbb{P})$ . More often than not, however, we are not interested in the precise outcome of the experiment, but rather in a function of the result. For example, the stock market processes may be thought of in terms of random experiments (i.e. we assume that stocks rise and fall randomly, according to some rules). What an investor is interested in is, perhaps, not the outcome of the experiment as such, but rather the value of his portfolio – a function of the price movements, but weighted by the stocks owned. Formally, we will be interested in a function  $X$  defined over  $\Omega$ , which will transform elementary events into real numbers (as for now – one dimensional). If what is interesting to us are real values, then natural questions that may also be of interest are of the type: what is the probability, that the value of the portfolio is not more than \$1 mln? This means that we want “ $X$  does not exceed  $a$ ” to be events of interest, i.e.

$$X^{-1}((-\infty, a]) = \{\omega \in \Omega : X(\omega) \leq a\} \in \mathcal{F},$$

which leads us to the following definition:

**Definition 1.** A real-valued random variable is any function  $X : \Omega \rightarrow \mathbb{R}$ , such that for all  $a \in \mathbb{R}$  the set  $X^{-1}((-\infty, a])$  is an event, i.e.  $X^{-1}((-\infty, a]) \in \mathcal{F}$ .

If  $\Omega$  is countable (finite or infinite), and we have  $\mathcal{F} = 2^\Omega$ , then any function  $X : \Omega \rightarrow \mathbb{R}$  will be a random variable.

Let us now look at some very simple examples.

- (1) We toss a coin twice. Let  $X$  denote the number of heads. We have  $\Omega = \{(H, H), (H, T), (T, H), (T, T)\}$ , and the function  $X((H, H)) = 2$ ,  $X((H, T)) = X((T, H)) = 1$ ,  $X((T, T)) = 0$ .
- (2) We roll a die twice, and let  $X$  denote the sum of numbers obtained. We have  $\Omega = \{(x, y) : x, y \in \{1, 2, 3, 4, 5, 6\}\}$ ,  $X((x, y)) = x + y$ .
- (3) We randomly choose a point from the interval  $[0, 2]$ . Let  $X$  denote the distance from this point to the nearest integer. We have  $\Omega = [0, 2]$  and

$$X(\omega) = \begin{cases} \omega & \text{if } \omega \in [0, \frac{1}{2}], \\ |\omega - 1| & \text{if } \omega \in (\frac{1}{2}, \frac{3}{2}], \\ 2 - \omega & \text{if } \omega \in (\frac{3}{2}, 2]. \end{cases}$$

If we have more than one random variable defined over the same space  $\Omega$ , we may perform any sensible operation of our choice – such as addition, subtraction, multiplication, division (if not by zero) – and as a result, we will obtain random variables. Also, if we transform a random variable with a “decent” function  $f : \mathbb{R} \rightarrow \mathbb{R}$  (a function transforming Borel sets into Borel sets),  $f(X)$  will also be a random variable. That is, if  $X$  and  $Y$  are random variables, then  $Z_1 = X \cdot Y^2$ ,  $Z_2 = \sin(X) + e^Y$ ,  $Z_3 = \frac{\ln(X^4+1)}{Y^2+1}$  etc. are also random variables.

We have hinted above that we may be interested in asking questions about the probability that the value of  $X$  will be equal to or fall in a specific interval, etc. In order to be able to ask these questions, we have to make sure that we will be able to comfortably assign probability to the values of  $X$ . Since we have probability  $\mathbb{P}$  over  $(\Omega, \mathcal{F})$ , this assignment will be quite straightforward, as illustrated by the following examples:

- (1) We toss a symmetric coin twice, and let  $X$  denote the number of heads. What are the probabilities of the specific outcomes?

We will adopt a useful notation:  $\mathbb{P}(X \in A) = \mathbb{P}(X^{-1}(A))$ .

We have  $\mathbb{P}(X = 0) = \frac{1}{4}$ ,  $\mathbb{P}(X = 1) = \frac{1}{2}$  and  $\mathbb{P}(X = 2) = \frac{1}{4}$  – i.e. we get a probability distribution on  $\mathbb{R}$ . What would happen if we were interested in  $Y$ , the number of tails?

We would also have  $\mathbb{P}(Y = 0) = \frac{1}{4}$ ,  $\mathbb{P}(Y = 1) = \frac{1}{2}$  and  $\mathbb{P}(Y = 2) = \frac{1}{4}$ . We see that two different random variables –  $X$  and  $Y$  – have the same probability distribution.

- (2) From a circle of radius equal to 1 we randomly draw a point. Let  $X$  denote the distance of this point from the center of the circle. This random variable may assume only values from the range  $[0, 1]$ . For any  $a \in [0, 1]$ , we have  $\mathbb{P}(X \in [0, a]) = \frac{\pi a^2}{\pi} = a^2$ , which means that we can assign probability to intervals of the type  $[0, a]$ . This probability may be extended to probability over any events on the real line.

In the above examples, we have seen how to transform the probability over the initial sample space  $(\Omega, \mathcal{F})$  to a probability over  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$  with the use of the random variable  $X$ . Different  $\Omega$ s may lead to substantially different probability models on  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$  (such as discrete or continuous). For a given  $\Omega$ , we may have different probabilities over  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$  for various random variables  $X$  (or the same probabilities for different random variables...). This leads us to the definition of a probability distribution connected with the variable  $X$ :

**Definition 2.** *The probability distribution of a random variable  $X$  (real-valued) is the probability  $\mu_X$  on  $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ , such that  $\mu_X(A) = \mathbb{P}(X \in A)$ .*

We will often write  $X \sim \mu$  to denote the fact that the distribution of  $X$  is given by the function  $\mu$ . Note that – as we have already seen – different random variables may have the same probability distributions, so that from a distribution of a random variable one can not derive the definition of the random variable. This is the reason why in most cases we will “forget” about the way a random variable was defined (i.e. forget about  $\Omega$  etc.), and concentrate on the probability distribution of this random variable.

Let us now look at some common distribution types.

- (1) We roll a die. Let  $X$  denote the number obtained. In this case,  $\mu_X$  is a probability distribution *concentrated* on the set  $\{1, 2, 3, 4, 5, 6\}$ , i.e. for  $k = 1, 2, \dots, 6$  we have

$$\mu_X(\{k\}) = \frac{1}{6} > 0,$$

and for any other values of  $k$  this probability would be null. Note that when describing probability over countable sample spaces, it was sufficient to describe probability over simple events only, and the probability of all other events could be easily calculated on this basis. It is the same in the case of distributions concentrated on countable subsets of  $\mathbb{R}$ : the description of the probability distribution may be derived from a description of probabilities of the outcomes of non-negative probability. Therefore, the formula from above may be expanded to

$$\mu_X(A) = \frac{1}{6} \sum_{k=1}^6 1_A(k),$$

for all  $A \in \mathcal{B}(\mathbb{R})$ .

- (2) The above is an example of a discrete probability distribution: we will call a distribution **discrete**, if there exists a countable (at most) subset  $S$  of  $\mathbb{R}$  such that  $\mu(S) = 1$ . This distribution is unequivocally defined by the probabilities of the elements of  $S$ : for any  $A \in \mathcal{B}(\mathbb{R})$ , we have

$$\mu(A) = \sum_{s \in A} \mu(\{s\}).$$

This is why, when we deal with a discrete probability distribution, this distribution is described by means of a set of pairs of values and corresponding (non-negative) probabilities  $\{(s_i, p_i) : s_i \in S, p_i > 0, \sum p_i = 1\}$  (typically presented in a table).

- (3) **Binomial distribution** with parameters  $n$  and  $p$  (denoted  $\text{Bin}(n, p)$ ) – this is the distribution of the number of successes in a Bernoulli Scheme with  $n$  trials and a probability of success in a single trial equal to  $p$ . As we already know, the probabilities are:

$$\mu(\{k\}) = \mathbb{P}(X = k) = \binom{n}{k} p^k (1-p)^{n-k}, \quad k = 0, 1, 2, \dots, n.$$

- (4) **Geometric distribution** with parameter  $p \in (0, 1)$ , denoted  $\text{Geom}(p)$  – this is the distribution of the number of the trial in which a success appeared for the first time in a series of Bernoulli trials with probability of success in a single trial equal to  $p$ . This distribution is concentrated on the set  $\{1, 2, \dots\}$ , and we have

$$\mu_X(\{k\}) = \mathbb{P}(X = k) = (1 - p)^{k-1}p, \quad k = 1, 2, \dots$$

Sometimes the geometric distribution is defined slightly differently (shifted – the number of failures before the first success appeared). This latter distribution is concentrated on the set  $\{0, 1, 2, \dots\}$ , and may be written as

$$\mu_Y(\{k\}) = \mathbb{P}(Y = k) = (1 - p)^k p, \quad k = 0, 1, 2, \dots$$

(We have  $Y = X - 1$ .)

- (5) **Poisson distribution** with parameter  $\lambda > 0$ , denoted  $\text{Pois}(\lambda)$  – this is a distribution concentrated on the set  $\{0, 1, 2, \dots\}$ , such that

$$\mu(\{k\}) = \mathbb{P}(X = k) = \frac{\lambda^k}{k!} e^{-\lambda}.$$

As we already know, this is the limit distribution for binomial distributions

- (6) **Uniform distribution** over interval  $[a, b]$ , denoted  $\mathcal{U}(a, b)$  – this corresponds to randomly drawing a number from the interval  $[a, b]$ . From geometric models of probability, we know that for an interval  $[c, d] \subseteq [a, b]$  we have

$$\mu_X([c, d]) = \mathbb{P}(X \in [c, d]) = \frac{|[c, d]|}{|[a, b]|} = \frac{d - c}{b - a},$$

which may be extended to any set  $A \in \mathcal{B}(\mathbb{R})$ :

$$\mu_X(A) = \mu_X(A \cap [a, b]) = \mathbb{P}(X \in A \cap [a, b]) = \frac{|A \cap [a, b]|}{|[a, b]|} = \frac{|A \cap [a, b]|}{b - a}.$$

It will be useful to rewrite the probability density in what, as for now, seems a more complicated form:

$$\mu_X(A) = \int_{A \cap [a, b]} \frac{1}{b - a} dx.$$

This last example is different from the previous ones in that we did not describe the probability that  $X$  will assume a given value; rather, we talked about the probabilities of sets (just as in the geometric probability case). It is the most simple example of a random variable assuming uncountably many values (i.e. values from a given range). This example may be complicated slightly:

- (7) Assume a wild bee leaves its nest (located at point 0 on a real line) and flies left with probability  $\frac{1}{3}$ , and right with probability  $\frac{2}{3}$ . If it flies left, it comes across a stick lying on  $[-3, -1]$ , covered uniformly with honey. It then randomly chooses a point from this stick to sit upon (all places are “equally attractive” in terms of honey). If the wild bee flies right, it comes across a stick lying on  $[2, 3]$ , also uniformly covered with honey, and also chooses a location to sit upon randomly. Using the same mechanism as in the previous example, we may write that for any Borel subset  $A \subseteq [-3, -1]$ , we have

$$\mu_X(A) = \mathbb{P}(X \in A) = \int_A \frac{1}{3} \cdot \frac{1}{-1 - (-3)} dx,$$

and for any Borel subset  $A \subseteq [2, 3]$ , we have

$$\mu_X(A) = \mathbb{P}(X \in A) = \int_A \frac{2}{3} \cdot \frac{1}{3 - 2} dx.$$

In the general case, we may therefore write

$$\mu_X(A) = \int_A g(x) dx,$$

where

$$g(x) = \begin{cases} \frac{1}{6} & \text{if } x \in [-3, -1), \\ \frac{2}{3} & \text{if } x \in [2, 3], \\ 0 & \text{otherwise.} \end{cases}$$

In this latter case, it became slightly more evident why we may wish to use a notation with integrals (but still, due to the fact that drawing from ranges was uniform, these integrals were not necessary). There is no reason, however, why we should constrain ourselves only to cases where probability is distributed uniformly (the layer of honey has the same thickness everywhere); in this case, the integral notation, where we describe the probability connected with a given range(set) on the real line with a function which is not necessarily constant, becomes crucial.

**Definition 3.** A random variable  $X$  has a **continuous distribution**, if there exists a function  $g : \mathbb{R} \rightarrow \mathbb{R}_+$ , such that for any set  $A \in \mathcal{B}(\mathbb{R})$ ,  $\mu_X(A) = \mathbb{P}(X \in A) = \int_A g(x)dx$ .  $g$  is called the **probability density function of  $X$** .

A density function must be nonnegative (it would be acceptable to change some isolated values to negative – all we will do with a density function will be integrating, and the result of integration is not vulnerable to a change in the value of the integrated function in an isolated point). Another feature of the density function is that since the probability of the whole space – in this case  $\mathbb{R}$  – must be equal to 1, we must have that

$$\int_{-\infty}^{\infty} g(x)dx = 1.$$

A density function determines the distribution of a random variable unequivocally. Going back to examples:

(6) – cont. For a uniform distribution  $\mathcal{U}(a, b)$  we have

$$g(x) = \frac{1}{b-a} 1_{[a,b]}(x).$$

(8) An **Exponential distribution** with parameter  $\lambda > 0$ , denoted  $\text{Exp}(\lambda)$ . This is a distribution with density

$$g(x) = \lambda e^{-\lambda x} 1_{[0,\infty)}(x).$$

Note that we would get the same distribution if the density function was equal to

$$g(x) = \lambda e^{-\lambda x} 1_{(0,\infty)}(x).$$

(9) **Standard normal distribution**, denoted  $\mathcal{N}(0, 1)$ . This is a distribution with density

$$g(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}.$$

This distribution is the distribution with what is probably the widest-known density function: the bell-shape curve. This distribution (or rather the more general versions) is a very common distribution which often appears when we describe the weight, size, IQ etc. distributions in a given population. This is not a coincidence – the Central Limit Theorem, which we will talk about in due time, provides an answer as to why this distribution is so popular in nature.

(10) The more general version of the normal distribution, denoted by  $\mathcal{N}(a, \sigma^2)$ , is defined for any  $a \in \mathbb{R}$  (the parameter of location) and  $\sigma > 0$  (the parameter of scale) by a density function

$$g_{a,\sigma^2}(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-a)^2}{2\sigma^2}\right).$$

(11) Additionally, for  $a \in \mathbb{R}$  and  $\sigma = 0$  we define  $\mathcal{N}(a, 0)$  as a single-point distribution  $\delta_a$  (Dirac delta), such that

$$\delta_a(A) = 1_A(a) = \begin{cases} 1 & \text{if } a \in A \\ 0 & \text{otherwise} \end{cases}$$