MA400. September Introductory Course
(Financial Mathematics, Risk & Stochastics)
P2. Random Variables and Distribution Functions

Preliminary considerations

1. Consider the random choice of a person from among N people. Assuming that all people in the group are equally likely to be chosen,

$$\Omega = \{\omega_1, \ldots, \omega_N\}, \quad \mathcal{F} = \mathcal{P}(\Omega) \quad \text{and} \quad \mathbb{P}\big(\{\omega_i\}\big) = \frac{1}{N}, \quad \text{for } i = 1, \ldots, N,$$

where ω_i is the i-th representative of the group and $\mathcal{P}(\Omega)$ is the power set of Ω (i.e., the set of all subsets of Ω), provide an appropriate probability space.

There are many quantities that can be associated with this probability space. For example, each individual $\omega \in \Omega$ is associated with their height $X(\omega)$, their weight $Y(\omega)$, their gender $Z(\omega)$ or their cultural background $U(\omega)$. Each of these quantities is a *random variable*. The random variables X and Y take values in the set of positive real numbers, the random variable Z takes values in the set of all possible genders, say $\{\text{male}, \text{female}\}$, while the random variable U takes values in the set of all possible cultural backgrounds. Unlike U, the random variables X, Y and Z can be quantified in practice.

Since mathematical modelling involves mathematical objects, we concentrate our attention on random variables that take values in a space of mathematical objects such as, e.g., the real numbers $\mathbb{R}$, the Euclidean space $\mathbb{R}^n$, or the set $C([0, \infty))$ of all continuous real-valued functions on $[0, \infty)$. In the context of the example that we discuss here, we therefore focus on random variables such as X, Y, and Z with a definition such as

$$Z(\omega) = \begin{cases} \sqrt{2}, & \text{if } \omega \text{ is female}, \\ -\pi, & \text{if } \omega \text{ is male}. \end{cases}$$

After the random choice has been made, the value of every random variable is known. On the other hand, *before* the random choice happens, every random variable is a *function* on Ω with values in the appropriate space: each individual $\omega \in \Omega$ is associated with a height $X(\omega)$, a weight $Y(\omega)$ and a gender $Z(\omega)$. (Of course, this story involves some idealisation: if somebody is weighed before and after a big lunch, then the scale's readings are most likely to be different numbers.)

2. Generalising the example above, a random variable X defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ is a function mapping Ω into a set of (mathematical) objects S. Accordingly, each sample $\omega \in \Omega$ is associated with a unique (mathematical) object $X(\omega) \in S$.

3. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be the probability space of Paragraph 1 and let X be the random variable that associates a person's rounded at the first decimal point height in metres $X(\omega)$ with each person $\omega \in \Omega$. For instance, $X(\omega) = 1.7\text{m}$ if the actual height of person ω is in the range $[1.65\text{m}, 1.75\text{m})$. Accordingly, the random variable X takes values in the set $\{\ldots, 1.5, 1.6, 1.7, 1.8, \ldots\}$.

   On the measurable space $(S, \mathcal{S})$ given by

$$S = \{\ldots, 1.5, 1.6, 1.7, 1.8, \ldots\} \quad \text{and} \quad \mathcal{S} = \mathcal{P}(S),$$

   which is associated with the set S in which X takes values, we can define a function $\overline{\mathbb{P}} : \mathcal{S} \to [0, 1]$ by

$$\begin{aligned} \overline{\mathbb{P}}(\emptyset) &= 0, \\ \overline{\mathbb{P}}\big(\{x\}\big) &= \frac{\text{number of people } \omega \in \Omega \text{ with height in } [x - 0.5, x + 0.5)}{N} \\ &= \mathbb{P}\big(\{\omega \in \Omega \mid \ X(\omega) = x\}\big), \qquad \text{for } x \in S, \end{aligned}$$

   and

$$\overline{\mathbb{P}}(A) = \sum_{x_j \in A} \overline{\mathbb{P}}\big(\{x_j\}\big), \quad \text{for } A \subseteq S \text{ with } A \neq \emptyset.$$

   We can then verify that $\overline{\mathbb{P}}$ is a probability measure on $(S, \mathcal{S})$ (see Exercise 2.4). This probability measure is called the *distribution* of X, and, given any $A \in \mathcal{S}$, $\overline{\mathbb{P}}(A)$ is the answer to the question "what is the probability that the recorded value of X turns out to be in the range given by the set A? ".

4. Motivated by the discussion above, we consider a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ and a function $X : \Omega \to S$, where S is a set. If $\mathcal{S}$ is a σ-algebra on S such that

$$X^{-1}(C) \equiv \{\omega \in \Omega \mid \ X(\omega) \in C\} \in \mathcal{F} \quad \text{for all } C \in \mathcal{S}, \tag{1}$$

   then we can define a probability measure $\overline{\mathbb{P}}$ on the measurable space $(S, \mathcal{S})$ by

$$\overline{\mathbb{P}}(C) = \mathbb{P}\left(X^{-1}(C)\right) \equiv \mathbb{P}\left(\{\omega \in \Omega \mid \ X(\omega) \in C\}\right), \quad \text{for } C \in \mathcal{S}. \tag{2}$$

   This probability measure, which is often denoted by $\mathbb{P}X^{-1}$ is called the *distribution* of X, or the *image law* of $\mathbb{P}$ under X, or just the *law* of X.

5. *The definition of* $\overline{\mathbb{P}}$ *in (2) is well-posed because (1) is satisfied.*

   Since the distributions of random variables are of fundamental importance in probability theory, this observation suggests that the following is the appropriate mathematical definition of a random variable:

   Definition. A *random variable* X defined on a measurable space $(\Omega, \mathcal{F})$ and with values in a measurable space $(S, \mathcal{S})$ is any function $X : \Omega \to S$ such that

$$X^{-1}(A) = \big\{\omega \in \Omega \mid \ X(\omega) \in A\big\} \in \mathcal{F} \quad \text{for every set } A \in \mathcal{S}.$$

   In particular, a *real-valued random variable* X defined on a measurable space $(\Omega, \mathcal{F})$ is any function $X : \Omega \to \mathbb{R}$ such that

$$X^{-1}(A) = \big\{\omega \in \Omega \mid \ X(\omega) \in A\big\} \in \mathcal{F} \quad \text{for every set } A \in \mathcal{B}(\mathbb{R}),$$

   where $\mathcal{B}(\mathbb{R})$ is the Borel σ-algebra on $\mathbb{R}$.

6. We now consider the following issue: what information do we get by observing a given random variable? To fix ideas, let us consider the probability space of Example 1 above, and let us denote by X the random variable considered in Paragraph 3.

   What do we learn by observing X? In other words, what information do we get if a measurement is made and we are reported the value of X only? Assuming that the population size is large (e.g., $N \geq 100$), the observed value of X does not necessarily identify ω: for instance, there may be several people with height in the range $[1.65\text{m}, 1.75\text{m})$, and all of them are associated with the same reading 1.7 of X. However, the value of X can tell us whether or not the person involved in the measurement belongs to several classes, namely subsets of Ω. Indeed, as soon as we are reported the observed value of X, we can provide a clear yes or no as answer to each question of the form: does the person involved in the measurement belong to the set of people with an X-measurement in the set $A \subseteq \mathbb{R}$. For instance, if we are reported that the value of X is somewhere in the range $(1.367, 1.5] \cup [1.77,\ 1.823]$, then we are certain that person involved in the measurement belongs to the family of those people with height in the range $[1.35\text{m}, 1.55\text{m}) \cup [1.75\text{m}, 1.85)$. Therefore, we can see that learning the value of X will inform us on whether the person chosen for the measurement belongs or not to any of the groups of people in the collection

$$\left\{ \{\omega \in \Omega \mid \ X(\omega) \in A\} \,\middle|\, A \in \mathcal{B}(\mathbb{R}) \right\} \subseteq \mathcal{F},$$

   which is a σ-algebra on Ω.

7. Generalising the previous example, we can identify the information we obtain by the observation of an $(S, \mathcal{S})$-valued random variable X defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with the σ-algebra

$$\sigma(X) = \left\{ \{\omega \in \Omega \mid \ X(\omega) \in A\} \;\middle|\; A \in \mathcal{S} \right\} \subseteq \mathcal{F}. \tag{3}$$

In particular, we can identify the information we obtain by the observation of a real-valued random variable X defined on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with the σ-algebra

$$\sigma(X) = \left\{ \{\omega \in \Omega \mid \ X(\omega) \in A\} \;\middle|\; A \in \mathcal{B}(\mathbb{R}) \right\} \subseteq \mathcal{F}, \tag{4}$$

where $\mathcal{B}(\mathbb{R})$ is the Borel σ-algebra on $\mathbb{R}$.

8. *It is worth stressing that the information set $\sigma(X)$ is associated with the random variable X and* not *with its eventually observed value.* To appreciate this comment, let us consider the following very simple example.

   Suppose that $\Omega = \{1, 2, 3, 4, 5\}$ and $\mathcal{F} = \mathcal{P}(\Omega)$. Also, let $A = \{1, 2\}$ and let X be the random variable defined by

$$X(\omega) = \mathbf{1}_A(\omega) = \begin{cases} 1, & \text{if } \omega \in A, \\ 0, & \text{if } \omega \notin A, \end{cases} \quad \text{for } \omega \in \Omega.$$

   We can check that, in this case,

$$\sigma(X) = \{\Omega, \emptyset, A, A^c\} = \big\{\Omega, \emptyset, \{1, 2\}, \{3, 4, 5\}\big\}.$$

   Before observing the actual value of X, we have certainty that we will be able to say whether each event in this information set has occurred or not as soon as we observe X. Furthermore, there is *no* event outside this information set for which we can have such a certainty.

   Of course, *after* we observe X, we have more information. Indeed, if we have been reported that $X = 1$, then the information we possess is

$$\begin{aligned} \big\{&\Omega, \emptyset, \{1, 2\}, \{3\}, \{4\}, \{5\}, \{3, 4\}, \{3, 5\}, \{4, 5\}, \\ &\{3, 4, 5\}, \{1, 2, 3\}, \{1, 2, 4\}, \{1, 2, 5\}, \{1, 2, 3, 4\}, \{1, 2, 3, 5\}, \{1, 2, 4, 5\}\big\}, \end{aligned}$$

   while, if we have been reported that $X = 0$, then we know that each event in

$$\big\{\Omega, \emptyset, \{1\}, \{2\}, \{1, 2\}, \{3, 4, 5\}, \{1, 3, 4, 5\}, \{2, 3, 4, 5\}\big\}$$

   has occurred or not.

9. Building on the ideas discussed above, we can now ask the following question: what is the information set we get if we observe two random variables, say X and Y? It might be tempting to say that this could be identified with $\sigma(X) \cup \sigma(Y)$, where $\sigma(\cdot)$ is a σ-algebra defined as in (4). However, a closer scrutiny shows that this is not a good idea.

   To fix ideas, suppose that a fair coin that lands heads (H) or tails (T) is tossed three times. The choices

$$\Omega = \big\{\mathrm{HHH}, \mathrm{THH}, \mathrm{HTH}, \mathrm{HHT}, \mathrm{TTH}, \mathrm{THT}, \mathrm{HTT}, \mathrm{TTT}\big\} \quad \text{and} \quad \mathcal{F} = \mathcal{P}(\Omega)$$

   provide a measurable space that is appropriate for the modelling of this random situation. Now, let X be the outcome of the first toss, and let Y be the total number of heads observed in the three coin tosses: the random variable X takes values in the set $\{\mathrm{H}, \mathrm{T}\}$, while the random variable Y takes values in the set $\{0, 1, 2, 3\}$. We can check that

$$\sigma(X) = \big\{\Omega, \emptyset, \{\mathrm{HHH}, \mathrm{HTH}, \mathrm{HHT}, \mathrm{HTT}\}, \{\mathrm{THH}, \mathrm{TTH}, \mathrm{THT}, \mathrm{TTT}\}\big\}$$

   and

$$\begin{aligned}\sigma(Y) = \big\{&\Omega, \emptyset, \{\mathrm{TTT}\}, \{\mathrm{HTT}, \mathrm{THT}, \mathrm{TTH}\}, \{\mathrm{THH}, \mathrm{HTH}, \mathrm{HHT}\}, \{\mathrm{HHH}\}\\ &\{\mathrm{TTT}, \mathrm{HTT}, \mathrm{THT}, \mathrm{TTH}\}, \{\mathrm{TTT}, \mathrm{THH}, \mathrm{HTH}, \mathrm{HHT}\}, \{\mathrm{TTT}, \mathrm{HHH}\},\\ &\{\mathrm{HTT}, \mathrm{THT}, \mathrm{TTH}, \mathrm{THH}, \mathrm{HTH}, \mathrm{HHT}\}, \{\mathrm{HTT}, \mathrm{THT}, \mathrm{TTH}, \mathrm{HHH}\},\\ &\{\mathrm{THH}, \mathrm{HTH}, \mathrm{HHT}, \mathrm{HHH}\}, \{\mathrm{TTT}, \mathrm{HTT}, \mathrm{THT}, \mathrm{TTH}, \mathrm{THH}, \mathrm{HTH}, \mathrm{HHT}\},\\ &\{\mathrm{TTT}, \mathrm{THH}, \mathrm{HTH}, \mathrm{HHT}, \mathrm{HHH}\},\\ &\{\mathrm{HTT}, \mathrm{THT}, \mathrm{TTH}, \mathrm{THH}, \mathrm{HTH}, \mathrm{HHT}, \mathrm{HHH}\}\big\}.\end{aligned}$$

   In this context, the choice $\sigma(X) \cup \sigma(Y)$ for the information set generated by the observation of the random variables X and Y is not appropriate. Indeed, we can be sure that we will be able to say whether the event $\{\mathrm{HHT}, \mathrm{HTH}\}$ has occurred or not as soon as we observe X and Y:

$$\text{if } X = \mathrm{H} \text{ and } Y = 0 \text{ or } 1 \text{ or } 3, \text{ then } \{\mathrm{HHT}, \mathrm{HTH}\} \text{ has not occurred,}$$
$$\text{if } X = \mathrm{T} \text{ and } Y = 0 \text{ or } 1 \text{ or } 2 \text{ or } 3, \text{ then } \{\mathrm{HHT}, \mathrm{HTH}\} \text{ has not occurred}$$

   and

$$\text{if } X = \mathrm{H} \text{ and } Y = 2, \text{ then } \{\mathrm{HHT}, \mathrm{HTH}\} \text{ has occurred.}$$

   The event $\{\mathrm{HHT}, \mathrm{HTH}\}$ does not belong to $\sigma(X) \cup \sigma(Y)$. However, it does belong to the σ-algebra generated by $\sigma(X) \cup \sigma(Y)$, namely, $\sigma\big(\sigma(X) \cup \sigma(Y)\big)$.

10. In light of the discussion in the previous paragraph, we can identify the information we obtain by the observation of the random variables in a given family $(X_i,\ i \in I)$, where $I \neq \emptyset$ is an index set, with the σ-algebra

$$\sigma(X_i,\ i \in I) := \sigma\left(\bigcup_{i \in I} \sigma(X_i)\right).$$

Random variables

11. *Definition.* A *real-valued random variable* X defined on a measurable space $(\Omega, \mathcal{F})$ is a real-valued function on Ω (i.e., $X : \Omega \to \mathbb{R}$) such that

$$X^{-1}(A) = \big\{\omega \in \Omega \mid \ X(\omega) \in A\big\} \in \mathcal{F} \quad \text{for every set } A \in \mathcal{B}(\mathbb{R}), \tag{5}$$

where $\mathcal{B}(\mathbb{R})$ is the Borel σ-algebra on $\mathbb{R}$.

Note that we often write $\{X \in A\}$ to denote the event $X^{-1}(A)$.

12. *Definition.* Given a measurable space $(S, \mathcal{S})$, an $(S, \mathcal{S})$-*valued random variable* X defined on a measurable space $(\Omega, \mathcal{F})$ is a function mapping Ω into S (i.e., $X : \Omega \to S$) such that

$$X^{-1}(A) = \big\{\omega \in \Omega \mid \ X(\omega) \in A\big\} \in \mathcal{F} \quad \text{for every set } A \in \mathcal{S}. \tag{6}$$

13. *Definition.* The σ-algebra $\sigma(X)$ generated by a real-valued random variable X is the collection of all sets $X^{-1}(A) = \{\omega \in \Omega \mid \ X(\omega) \in A\}$, where $A \in \mathcal{B}(\mathbb{R})$, i.e.,

$$\sigma(X) = \big\{X^{-1}(A) \mid \ A \in \mathcal{B}(\mathbb{R})\big\}. \tag{7}$$

14. *Lemma.* The family of events $\sigma(X)$ defined by (7) is indeed a σ-algebra.

Proof. We use the fact that $\mathcal{B}(\mathbb{R})$ is a σ-algebra on $\mathbb{R}$ to check that $\sigma(X)$ satisfies the three properties that characterise a σ-algebra on Ω:

(i) $\Omega \in \sigma(X)$ because $\Omega = X^{-1}(\mathbb{R})$ and $\mathbb{R} \in \mathcal{B}(\mathbb{R})$.

(ii) Let any event $C \in \sigma(X)$. We need to show that $\Omega \setminus C \in \sigma(X)$.
To this end, observe that the definition (7) of $\sigma(X)$ implies that there exists $A \in \mathcal{B}(\mathbb{R})$ such that

$$C = X^{-1}(A) \equiv \{\omega \in \Omega \mid \ X(\omega) \in A\}.$$

Now, we calculate

$$\begin{aligned} \Omega \setminus C &= \Omega \setminus \{\omega \in \Omega \mid \ X(\omega) \in A\} \\ &= \{\omega \in \Omega \mid \ X(\omega) \notin A\} \\ &= \{\omega \in \Omega \mid \ X(\omega) \in \mathbb{R} \setminus A\} \\ &= X^{-1}(\mathbb{R} \setminus A) \ \in \ \sigma(X), \end{aligned}$$

because $\mathbb{R} \setminus A \in \mathcal{B}(\mathbb{R})$.

(iii) Consider any sequence of events $C_1, C_2, \ldots, C_n, \ldots \in \sigma(X)$. We need to prove that $\bigcup_{n=1}^{\infty} C_n \in \sigma(X)$.

Since $C_n \in \sigma(X)$ for all n, the definition (7) of $\sigma(X)$ implies that there exists a sequence of events $A_1, A_2, \ldots, A_n, \ldots \in \mathcal{B}(\mathbb{R})$ such that

$$C_n = X^{-1}(A_n) \equiv \{\omega \in \Omega \mid \ X(\omega) \in A_n\} \quad \text{for all } n = 1, 2, \ldots.$$

Now, we calculate

$$\begin{aligned}
\bigcup_{n=1}^{\infty} C_n &= \bigcup_{n=1}^{\infty} \{\omega \in \Omega \mid \ X(\omega) \in A_n\} \\
&= \left\{\omega \in \Omega \mid \ X(\omega) \in \bigcup_{n=1}^{\infty} A_n\right\} \\
&= X^{-1}\left(\bigcup_{n=1}^{\infty} A_n\right) \ \in \ \sigma(X),
\end{aligned}$$

because $\bigcup_{n=1}^{\infty} A_n \in \mathcal{B}(\mathbb{R})$.

15. *Definition*. The σ-algebra $\sigma(X)$ generated by an $(S, \mathcal{S})$-valued random variable X is the collection of all sets $X^{-1}(A) = \{\omega \in \Omega \mid \ X(\omega) \in A\}$, where $A \in \mathcal{S}$, i.e.,

$$\sigma(X) = \big\{X^{-1}(A) \mid \ A \in \mathcal{S}\big\}. \tag{8}$$

16. *Definition*. The σ-algebra generated by a collection of random variables $(X_i, \ i \in I)$, where $I \neq \emptyset$, is

$$\sigma\big(X_i, \ i \in I\big) = \sigma\big(\sigma(X_i), \ i \in I\big) \equiv \sigma\left(\bigcup_{i \in I} \sigma(X_i)\right).$$

17. *Definition*. Given a random variable X and a σ-algebra $\mathcal{H}$ on Ω, we say that X is *$\mathcal{H}$-measurable* if $\sigma(X) \subseteq \mathcal{H}$.

18. With the terminology introduced by this definition, note that:

    Given a random variable X, $\sigma(X)$ is the *smallest* σ-algebra with respect to which X is measurable.

    Given a family of random variables $(X_i, \ i \in I)$, $\sigma\big(X_i, \ i \in I\big)$ is the *smallest* σ-algebra with respect to which every X_i is measurable.

As we have discussed above, σ-algebras can be considered as models of information. Informally, this definition says that a random variable X is $\mathcal{H}$-measurable if the information provided by X is a subset of the information contained in $\mathcal{H}$.

19. *Lemma.* Consider two measurable spaces $(\Omega, \mathcal{F})$ and $(S, \mathcal{S})$, and a family of sets $\mathcal{C}$ such that $\sigma(\mathcal{C}) = \mathcal{S}$. If a function $X : \Omega \to S$ satisfies

$$X^{-1}(A) = \big\{\omega \in \Omega \mid \; X(\omega) \in A\big\} \in \mathcal{F} \quad \text{for all } A \in \mathcal{C}, \tag{9}$$

then X is an $(S, \mathcal{S})$-valued random variable.

[To appreciate the significance of this result, do compare (9) with (6)!]

Proof. We will prove that X is an $(S, \mathcal{S})$-valued random variable if we show that

$$X^{-1}(A) \in \mathcal{F} \quad \text{for all } A \in \mathcal{S},$$

or equivalently, if we show that

$$\Big\{A \in \mathcal{S} \mid \; X^{-1}(A) \in \mathcal{F}\Big\} = \mathcal{S}. \tag{10}$$

To this end, we define

$$\mathcal{H} = \Big\{A \in \mathcal{S} \mid \; X^{-1}(A) \in \mathcal{F}\Big\},$$

and we note that

$$\mathcal{C} \subseteq \mathcal{H} \subseteq \mathcal{S}, \tag{11}$$

where the first inclusion follows thanks to (9).

Furthermore, we note that $\mathcal{H}$ is a σ-algebra on S, because:

(i) $S \in \mathcal{H}$ because $X^{-1}(S) = \Omega \in \mathcal{F}$.

(ii) Given an event $A \in \mathcal{H}$,

$$X^{-1}(S \setminus A) = \Omega \setminus X^{-1}(A) \in \mathcal{F},$$

so, $S \setminus A \in \mathcal{H}$.

(iii) Given a sequence of events $A_1, A_2, \ldots, A_n, \ldots \in \mathcal{H}$,

$$X^{-1}\left(\bigcup_{n=1}^{\infty} A_n\right) = \bigcup_{n=1}^{\infty} X^{-1}(A_n) \in \mathcal{F},$$

so, $\bigcup_{n=1}^{\infty} A_n \in \mathcal{H}$.

Now, in view of the assumption that $\sigma(\mathcal{C}) = \mathcal{S}$, (11), and the fact that $\mathcal{H}$, $\mathcal{S}$ are σ-algebras on S, we can see that

$$\mathcal{S} = \sigma(\mathcal{C}) \subseteq \sigma(\mathcal{H}) = \mathcal{H} \subseteq \mathcal{S},$$

which proves that $\mathcal{H} = \mathcal{S}$, and establishes (10).

20. *Lemma.* Suppose that X and Y are real-valued random variables defined on a measurable space $(\Omega, \mathcal{F})$, and let λ be a real number. Then, $X+Y$, XY and λX are all real-valued random variables.

Proof. In view of Lemma 19 and the fact that the family of sets

$$\mathcal{C}_1 = \big\{(a, \infty) \mid \ a \in \mathbb{R}\big\}$$

generates the Borel σ-algebra, i.e., $\sigma(\mathcal{C}_1) = \mathcal{B}(\mathbb{R})$, we will prove that the sum $X+Y$ of two random variables X and Y is also a random variable if we show that

$$\{\omega \in \Omega \mid \ X(\omega) + Y(\omega) > a\} \in \mathcal{F} \quad \text{for all } a \in \mathbb{R}. \tag{12}$$

To this end, we note that, given any $\omega \in \Omega$ and any $a \in \mathbb{R}$, $X(\omega) > a - Y(\omega)$ if and only if we can find a rational number q such that $X(\omega) > q > a - Y(\omega)$. Therefore,

$$\begin{aligned}\{\omega \in \Omega \mid \ X(\omega) + Y(\omega) > a\} &= \bigcup_{q \in \mathbb{Q}} \{\omega \in \Omega \mid \ X(\omega) > q > a - Y(\omega)\} \\ &= \bigcup_{q \in \mathbb{Q}} \big(\{\omega \in \Omega \mid \ X(\omega) > q\} \cap \{\omega \in \Omega \mid \ Y(\omega) > a - q\}\big).\end{aligned}$$

However, the expression on the right hand side of this expression is a *countable* union of events in $\mathcal{F}$ (because X and Y are random variables), and (12) follows.

Now, we use Lemma 19 and the fact that the family of sets

$$\mathcal{C}_2 = \big\{(-\infty, a] \mid \ a \in \mathbb{R}\big\}$$

generates the Borel σ-algebra $\mathcal{B}(\mathbb{R})$ to show that, given a constant $\lambda \in \mathbb{R}$ and a random variable X, the function λX mapping Ω into $\mathbb{R}$ is a random variable by proving that

$$\{\omega \in \Omega \mid \ \lambda X(\omega) \leq a\} \in \mathcal{F} \quad \text{for all } a \in \mathbb{R}. \tag{13}$$

Indeed, given any $a \in \mathbb{R}$,

$$\{\omega \in \Omega \mid \ \lambda X(\omega) \leq a\} = \begin{cases} \{\omega \in \Omega \mid \ X(\omega) \leq a/\lambda\}, & \text{if } \lambda > 0, \\ \emptyset, & \text{if } \lambda = 0 \text{ and } a < 0, \\ \Omega, & \text{if } \lambda = 0 \text{ and } a \geq 0, \\ \{\omega \in \Omega \mid \ X(\omega) \geq a/\lambda\}, & \text{if } \lambda < 0. \end{cases}$$

All of the events on the right hand side of this expression belong to $\mathcal{F}$ (because X is a random variable), and (13) follows.

Similarly, if X is a random variable, then, given any $a \in \mathbb{R}$,

$$\big\{\omega \in \Omega \mid \ X^2(\omega) \leq a\big\} = \begin{cases} \emptyset, & \text{if } a < 0, \\ \{\omega \in \Omega \mid \ X(\omega) \in [-a, a]\}, & \text{if } a \geq 0. \end{cases}$$

Since either of the two events appearing on the right hand side of this expression belong to $\mathcal{F}$ (because X is a random variable), it follows that X^2 is a random variable.

Using what we have proved up to now, we can see that, given any random variables X and Y, the product XY is also a random variable because the identity

$$XY = \tfrac{1}{2}(X+Y)^2 - \tfrac{1}{2}X^2 - \tfrac{1}{2}Y^2$$

expresses XY as a sum of random variables.

21. *Lemma.* Suppose that $X_1, X_2, \dots, X_n, \dots$ is a sequence of real-valued random variables defined on a measurable space $(\Omega, \mathcal{F})$. The functions

$$\inf_{n\geq 1} X_n, \quad \sup_{n\geq 1} X_n, \quad \liminf_{n\to\infty} X_n \quad \text{and} \quad \limsup_{n\to\infty} X_n$$

mapping Ω into $[-\infty, \infty]$, defined by

$$\left(\inf_{n\geq 1} X_n\right)(\omega) = \inf_{n\geq 1} X_n(\omega), \quad \left(\sup_{n\geq 1} X_n\right)(\omega) = \sup_{n\geq 1} X_n(\omega),$$
$$\left(\liminf_{n\to\infty} X_n\right)(\omega) = \liminf_{n\to\infty} X_n(\omega) \quad \text{and} \quad \left(\limsup_{n\to\infty} X_n\right)(\omega) = \limsup_{n\to\infty} X_n(\omega),$$

respectively, are $\big([-\infty,\infty], \mathcal{B}([-\infty,\infty])\big)$-valued random variables, where $\mathcal{B}\big([-\infty,\infty]\big)$ is the Borel σ-algebra on $[-\infty,\infty]$, so that

$$\mathcal{B}\big([-\infty,\infty]\big) = \sigma\big(\{[-\infty, a] \mid a \in [-\infty,\infty]\}\big) \supseteq \mathcal{B}\big(\mathbb{R}\big). \tag{14}$$

Furthermore,

$$\big\{\omega \in \Omega \mid \lim_{n\to\infty} X_n(\omega) \text{ exists in } \mathbb{R}\big\} \in \mathcal{F}. \tag{15}$$

Proof. In view of Lemma 19 and (14), we can see that the inclusion

$$\begin{aligned}\left(\sup_{n\geq 1} X_n\right)^{-1}\big([-\infty, a]\big) &= \left\{\omega \in \Omega \mid \sup_{n\geq 1} X_n(\omega) \leq a\right\} \\ &= \bigcap_{n=1}^{\infty} \{\omega \in \Omega \mid X_n(\omega) \leq a\} \in \mathcal{F} \quad \text{for all } a \in [-\infty,\infty],\end{aligned}$$

implies that $\sup_{n\geq 1} X_n$ is an $\big([-\infty,\infty], \mathcal{B}([-\infty,\infty])\big)$-valued random variable.

Recalling that, if Z is a random variable, then $-Z$ is also a random variable (see Lemma 20), we can see that the result we have just proved and the identity $\inf_{n\geq 1} X_n = -\sup_{n\geq 1}(-X_n)$ imply that $\inf_{n\geq 1} X_n$ is an $\big([-\infty,\infty], \mathcal{B}([-\infty,\infty])\big)$-valued random variable.

Now, if we define

$$\underline{Z}_n = \inf_{k \geq n} X_k \quad \text{and} \quad \overline{Z}_n = \sup_{k \geq n} X_k, \quad \text{for } n \geq 1,$$

then $\underline{Z}_n$ and $\overline{Z}_n$ are $\big([-\infty, \infty], \mathcal{B}([-\infty, \infty])\big)$-valued random variables for all $n \geq 1$. It follows that

$$\liminf_{n\to\infty} X_n = \lim_{n\to\infty} \inf_{k\geq n} X_k = \sup_{n\geq 1} \inf_{k\geq n} X_n = \sup_{n\geq 1} \underline{Z}_n$$

and

$$\limsup_{n\to\infty} X_n = \lim_{n\to\infty} \sup_{k\geq n} X_k = \inf_{n\geq 1} \sup_{k\geq n} X_k = \inf_{n\geq 1} \overline{Z}_n$$

are $\big([-\infty, \infty], \mathcal{B}([-\infty, \infty])\big)$-valued random variables.

Finally, we note that (15) follows immediately from the identity

$$\begin{aligned}
&\big\{\omega \in \Omega \mid \lim_{n\to\infty} X_n(\omega) \text{ exists in } \mathbb{R}\big\} \\
&\quad = \big\{\omega \in \Omega \mid \limsup_{n\to\infty} X_n(\omega) < \infty\big\} \cap \big\{\omega \in \Omega \mid \liminf_{n\to\infty} X_n(\omega) > -\infty\big\} \\
&\qquad \cap \big\{\omega \in \Omega \mid \Big(\limsup_{n\to\infty} X_n - \liminf_{n\to\infty} X_n\Big)(\omega) = 0\big\}
\end{aligned}$$

and the fact that the events on the right-hand side of this expression belong to $\mathcal{F}$.

22. *Lemma.* Suppose that X and Y are real-valued random variables such that Y is $\sigma(X)$-measurable. Then there exists a Borel function $g : \mathbb{R} \to \mathbb{R}$ such that $Y = g(X)$.

    More generally, given n real-valued random variables $X_1, \ldots, X_n$, suppose that a random variable Y is $\sigma(X_1, \ldots, X_n)$-measurable. Then there exist a Borel function $g : \mathbb{R}^n \to \mathbb{R}$ such that $Y = g(X_1, \ldots, X_n)$.

Distributions

23. *Definition.* The distribution function F_X of a real-valued random variable X is defined by

$$F_X(a) = \mathbb{P}(X \le a) \equiv \mathbb{P}\big(X^{-1}\big((-\infty, a]\big)\big), \quad \text{for } a \in \mathbb{R}.$$

Provided there is no possibility of confusion, we often write $F(a)$ instead of $F_X(a)$.

24. *Lemma.* The following are simple properties of distribution functions:

(i) Every distribution function F is an increasing function.

Proof. Observing that, given any $a \le b$,

$$\{\omega \in \Omega \mid \ X(\omega) \le a\} \subseteq \{\omega \in \Omega \mid \ X(\omega) \le b\},$$

we can see that

$$\begin{aligned} F(a) &= \mathbb{P}\big(\{\omega \in \Omega \mid \ X(\omega) \le a\}\big) \\ &\le \mathbb{P}\big(\{\omega \in \Omega \mid \ X(\omega) \le b\}\big) \\ &= F(b). \end{aligned}$$

Here, we have used the monotonicity of a probability measure, i.e., that, given any $A, B \in \mathcal{F}$,

$$A \subseteq B \quad \Rightarrow \quad \mathbb{P}(A) \le \mathbb{P}(B).$$

(ii) Every distribution function F satisfies

$$\lim_{a \to -\infty} F(a) = 0 \quad \text{and} \quad \lim_{a \to \infty} F(a) = 1.$$

Proof. Since F is an increasing function, both limits exist. Therefore, we only have to show that

$$\lim_{n \to \infty} F(-n) = 0 \quad \text{and} \quad \lim_{n \to \infty} F(n) = 1.$$

To this end, we first consider the decreasing sequence of events $A_1 \supseteq A_2 \supseteq \cdots \supseteq A_n \supseteq \cdots$ defined by

$$A_n = \{\omega \in \Omega \mid \ X(\omega) \le -n\},$$

and we observe that $\bigcap_{n=1}^{\infty} A_n = \emptyset$. Using the "continuity" of a probability measure, we can calculate

$$\begin{aligned} \lim_{n \to \infty} F(-n) &= \lim_{n \to \infty} \mathbb{P}(A_n) \\ &= \mathbb{P}\left(\bigcap_{n=1}^{\infty} A_n\right) \\ &= \mathbb{P}(\emptyset) \\ &= 0. \end{aligned}$$

Next, we consider the increasing sequence of events $B_1 \subseteq B_2 \subseteq \cdots \subseteq B_n \subseteq \cdots$ defined by

$$B_n = \{\omega \in \Omega \mid \ X(\omega) \leq n\}.$$

and we observe that $\bigcup_{n=1}^{\infty} B_n = \Omega$. In view of the "continuity" of a probability measure, it follows that

$$\begin{aligned}
\lim_{n\to\infty} F(n) &= \lim_{n\to\infty} \mathbb{P}(B_n) \\
&= \mathbb{P}\left(\bigcup_{n=1}^{\infty} B_n\right) \\
&= \mathbb{P}(\Omega) \\
&= 1.
\end{aligned}$$

(*iii*) Every distribution function F is right-continuous.

Proof. Since F is increasing, both of the limits $\lim_{x\downarrow a} F(x)$ and $\lim_{x\uparrow a} F(x)$ exist at every point $a \in \mathbb{R}$. Therefore, to see that F is right-continuous we observe that, given any $a \in \mathbb{R}$,

$$\begin{aligned}
\lim_{n\to\infty} F\left(a + \frac{1}{n}\right) &= \lim_{n\to\infty} \mathbb{P}\left(\left\{\omega \in \Omega \mid \ X(\omega) \leq a + \frac{1}{n}\right\}\right) \\
&= \mathbb{P}\left(\bigcap_{n=1}^{\infty} \left\{\omega \in \Omega \mid \ X(\omega) \leq a + \frac{1}{n}\right\}\right) \\
&= \mathbb{P}\left(\{\omega \in \Omega \mid \ X(\omega) \leq a\}\right) \\
&= F(a).
\end{aligned}$$

25. *Example*. Suppose that we roll a fair dice once. Let X be the number we observe. The distribution of X is

$$F(a) = \begin{cases} 0, & \text{if } a < 1, \\ \frac{1}{6}, & \text{if } 1 \leq a < 2, \\ \frac{2}{6}, & \text{if } 2 \leq a < 3, \\ \frac{3}{6}, & \text{if } 3 \leq a < 4, \\ \frac{4}{6}, & \text{if } 4 \leq a < 5, \\ \frac{5}{6}, & \text{if } 5 \leq a < 6, \\ 1, & \text{if } 6 \leq a. \end{cases}$$

26. *Example*. The distribution function of a random variable X is given by

$$F(x) = \begin{cases} 0, & \text{if } -\infty < x < 0, \\ 1 - 0.5e^{-x}, & \text{if } 0 \leq x. \end{cases}$$

We can compute

$$\begin{aligned}\mathbb{P}(X=0) &= \mathbb{P}(0)-\mathbb{P}(0-)\\ &= 0.5,\end{aligned}$$

and

$$\begin{aligned}\mathbb{P}(1<X\le 2) &= \mathbb{P}(X\le 2)-\mathbb{P}(X\le 1)\\ &= F(2)-F(1)\\ &= 0.5\left(e^{-1}-e^{-2}\right).\end{aligned}$$

27. *Definition*. The joint distribution of n random variables $X_1,\ldots,X_n$ is defined to be

$$F_{X_1\ldots X_n}(a_1,\ldots,a_n)=\mathbb{P}(X_1\le a_1,\ldots,X_n\le a_n)=\mathbb{P}\left(\bigcap_{i=1}^n X_i^{-1}\big((-\infty,a_i]\big)\right).$$

Discrete random variables

28. *Definition.* A real-valued random variable X is *discrete* if it maps Ω into a countable subset of $\mathbb{R}$. The *probability mass function* of a discrete random variable X is the collection of all pairs (x_j, p_j) such that

$$p_j = \mathbb{P}\big(X = x_j\big) > 0. \tag{16}$$

29. In view of (16), the distribution function of a discrete random variable X is given by

$$F(a) = \sum_{j \text{ such that } x_j \le a} p_j.$$

Also,

$$p_j = F(x_j) - F(x_j-).$$

30. *Example.* Given an event $A \in \mathcal{F}$, the random variable

$$\mathbf{1}_A(\omega) = \begin{cases} 1, & \text{if } \omega \in A \quad (\text{“success”}), \\ 0, & \text{if } \omega \in A^c \quad (\text{“failure”}), \end{cases}$$

is called the *indicator* of A. We say that the random variable $\mathbf{1}_A$ is *Bernoulli*. The probability mass function of the random variable $\mathbf{1}_A$ is given by

$$p_1 = \mathbb{P}\big(\mathbf{1}_A = 1\big) \equiv \mathbb{P}(A) \quad \text{and} \quad p_0 = \mathbb{P}\big(\mathbf{1}_A = 0\big) \equiv \mathbb{P}(A^c).$$

31. *Example.* A discrete random variable X has the *binomial* distribution with parameters n, p if its probability mass function is characterised by

$$p_i \equiv \mathbb{P}\big(X = i\big) = \binom{n}{i} p^i (1-p)^{n-i}, \quad \text{for } i = 0, 1, \ldots, n,$$

where

$$\binom{n}{i} = \frac{n!}{i!(n-i)!}.$$

Here, n is a positive integer and $p \in (0, 1)$. We often write $X \sim B(n, p)$,

Suppose that a coin that lands heads with probability p is tossed n times. If we define the random variable X to be the total number of heads observed in the n tosses, then X has the binomial distribution. More generally, the total number of “successes” in a fixed number of independent trials has the binomial distribution.

32. *Example.* A random variable X has the *Poisson* distribution with parameter λ if its probability mass function is given by

$$p_i = \mathbb{P}\big(X = i\big) = e^{-\lambda} \frac{\lambda^i}{i!}, \quad i = 0, 1, \ldots.$$

Here, $\lambda > 0$. In such a case, we write $X \sim \mathbb{P}(\lambda)$.

Continuous random variables

33. *Definition.* A real-valued random variable X is *continuous* if there exists a function f, called the *probability density function* of X, such that

$$\mathbb{P}\big(X \in A\big) = \int_A f(x)\,dx \quad \text{for all } A \in \mathcal{B}(\mathbb{R}). \tag{17}$$

34. Since probabilities are positive, every probability density function f satisfies

$$f(a) \geq 0 \quad \text{for all } a \in \mathbb{R}.$$

Since $\mathbb{P}(\Omega) \equiv \mathbb{P}(X \in \mathbb{R}) = 1$, every probability density function f satisfies

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

Also, observe that (17) implies

$$\mathbb{P}(a \leq X \leq b) = \mathbb{P}(a < X \leq b) = \mathbb{P}(a \leq X < b) = \mathbb{P}(a < X < b) = \int_a^b f(x)dx.$$

35. *Example.* A random variable X has the *normal* distribution with mean m and variance σ^2 if its probability density function is given by

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

Here, $m \in \mathbb{R}$ and $\sigma > 0$. We often write $X \sim N(m, \sigma^2)$. Normal random variables are also called *Gaussian*.

The probability distribution function of a normal random variable satisfies

$$F(a) = \Phi\left(\frac{a-m}{\sigma}\right),$$

where Φ is the "standard normal distribution function" defined by

$$\Phi(a) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{a} e^{-\frac{x^2}{2}}\,dx.$$

To see this, observe first that

$$F(a) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{a} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right) dx.$$

If we make the change of variables $y = (x-m)/\sigma$, then

$$\begin{aligned} F(a) &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{a-m}{\sigma}} \exp\left(-\frac{y^2}{2}\right) dy. \\ &= \Phi\left(\frac{a-m}{\sigma}\right). \end{aligned} \tag{18}$$

36. *Example.* A random variable X has the *exponential* distribution with parameter μ if its probability density function is given by

$$f(a) = \begin{cases} \mu e^{-\mu a}, & \text{if } a \geq 0, \\ 0, & \text{if } a < 0. \end{cases}$$

Here, $\mu > 0$.

37. *Definition.* n real-valued random variables $X_1, \ldots, X_n$ are said to be *jointly continuous* if there exists a function $f : \mathbb{R}^n \to \mathbb{R}$, called the *joint probability density function* of $X_1, \ldots, X_n$, such that

$$\mathbb{P}\big(X_1 \in A_1, \ldots, X_n \in A_n\big) = \int_{A_1} \cdots \int_{A_n} f(x_1, \ldots, x_n)\, dx_1 \cdots dx_n$$

for all $A_1, \ldots, A_n \in \mathcal{B}(\mathbb{R})$.