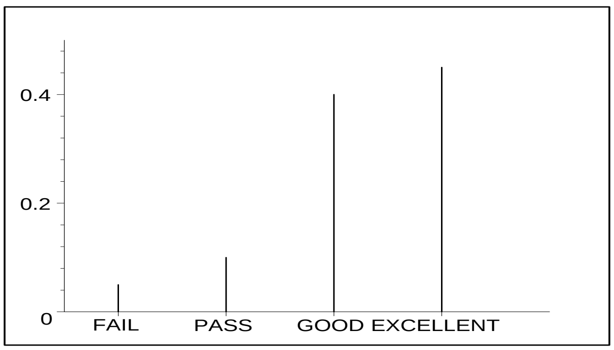
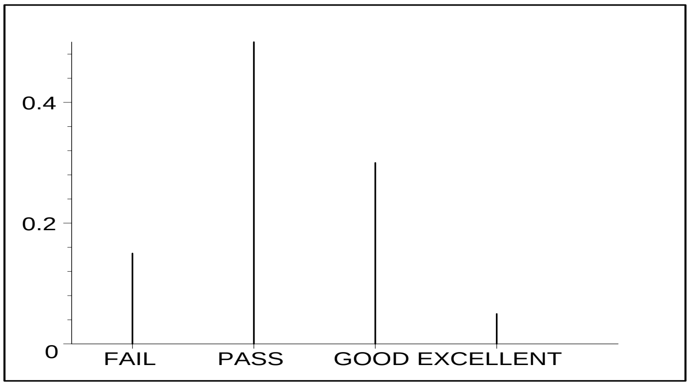
3 Random Variables and Distributions

Most quantities that we deal with every day are more or less of a random nature. The height of the person we first meet on leaving home in the morning, the grade we will earn in the next exam, the cost of a bottle of wine we will buy in the nearest shopping centre, and many more, serve as examples of so called random variables. Any such variable has its own specifications or characteristics. For example, the height of an adult male may take all values from 150 cm to 230 cm or even beyond this interval. Besides, the interval $(175, 180)$ is more likely than other intervals of the same length, say $(152, 157)$ or $(216, 221)$. On the other hand, the grade we will earn during the coming exam will take finite many values: excellent, good, pass, fail, and for the reader some of these values are more likely than others.

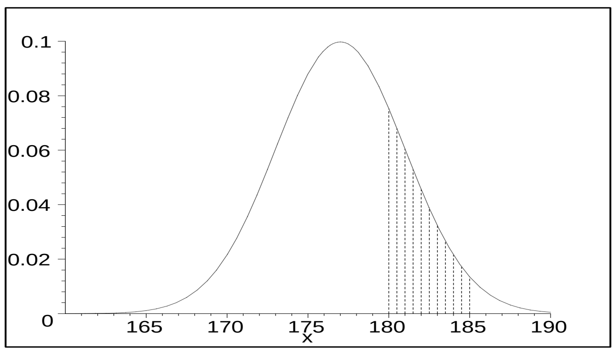


Graphically, the distribution of the grades of the reader may be described by the above figure. The heights of the bars are equal to the corresponding probabilities and their sum equals 1.

For an average student the distribution of his grades might have a different shape.



A hypothetical distribution of the height of an adult male may appear as the following picture. The probability here is represented by a corresponding area under the graph. For example, the shaded area in the figure is just the probability that the height of an adult male is greater then 180 cm and less than 185 cm. The whole area under the graph equals 1.



Distributions of random variables can often be characterised by parameters. This makes it possible to compare two or more distributions with each other. Also, the problem of determining and measuring the dependence or correlation of random variables with each other is of great importance. In fact, we believe that the height and weight of a student is more strongly correlated than her height and examination mark.

We will discuss the problems mentioned above in the following sections. First, we will consider probability distributions, then random variables and vectors. In the next chapter basic characteristics of probability distributions and the problem of correlation will be considered.

3.1 Probability Distributions

The real line, real plane, three dimensional space and, generally, the space $\mathbb{R}^n$ are often considered as sets Ω of elementary events. In most cases one considers σ-algebra of all Borel sets $\mathcal{B}(\mathbb{R}^n)$, i.e. the smallest σ-algebra containing all of the open sets. On the other hand, the probability measures P defined on $\mathcal{B}(\mathbb{R}^n)$ can be chosen in a variety of ways.

Definition 3.1.1 *A probability distribution (n-dimensional), briefly a distribution, is a measure P such that the triple*

$$(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n), P)$$

is a probability space.

Two distinct classes of probability distributions are usually of interest: discrete distributions and continuous distributions. There are also other less typical distributions beyond those contained in these two classes.

3.1.1 Discrete Distributions

Definition 3.1.2 *A distribution $P : \mathcal{B}(\mathbb{R}^n) \longrightarrow \mathbb{R}$ is discrete if there exist sequences $\{x_i\} \subset \mathbb{R}^n$ and $\{p_i\} \subset \mathbb{R}$, $i = 1, 2, \ldots, N$ with $N \leq \infty$ such that:*

$$p_i > 0, \qquad \sum_{i=1}^{N} p_i = 1, \qquad \textit{and} \qquad P(\{x_i\}) = p_i \qquad \textit{for all } i = 1, 2, \ldots, N.$$

This definition fully determines the distribution. Namely, for any Borel set A we have:

$$P(A) = P(A \cap \bigcup_{i=1}^{N} \{x_i\}) = \sum_{i=1}^{N} P(A \cap \{x_i\}) = \sum_{i: x_i \in A} P(\{x_i\}).$$

The first equality follows from the fact that $P(\bigcup_{i=1}^{N} \{x_i\}) = 1$ and from Problem 1.6. Thus we have:

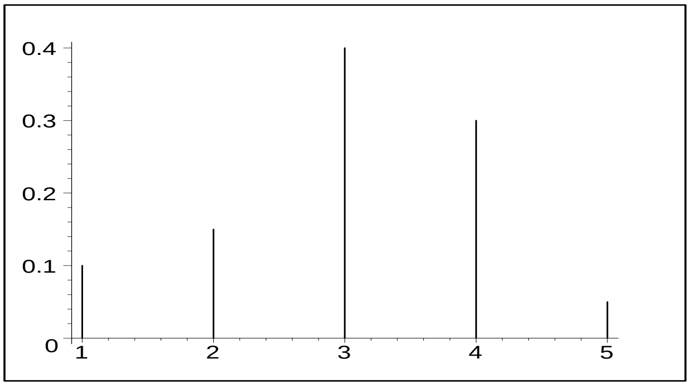
$$P(A) = \sum_{i: x_i \in A} p_i. \tag{3.1}$$

The grades of a student just discussed had discrete distributions.

MAPLE ASSISTANCE 19

We graph a discrete one dimensional distribution determined by the sequences: $x_1 = 1$, $x_2 = 2$, $x_3 = 3$, $x_4 = 4$, $x_5 = 5$ and $p_1 = 0.1$, $p_2 = 0.15$, $p_3 = 0.4$, $p_4 = 0.3$, $p_5 = 0.05$.

```
> with(plottools):
> PLOT( CURVES([[1,0],[1,0.1]],[[2,0],[2,0.15]],
  [[3,0],[3,0.4]],[[4,0],[4,0.3] ],[[5,0],[5,0.05]],
  THICKNESS(3)));
```



Note carefully that capital letters have been used in the MAPLE command.

MAPLE ASSISTANCE 20

We graph a discrete two dimensional distribution concentrated in points $x_1 = (1,2)$, $x_2 = (1,-1)$, $x_3 = (1,1)$, $x_4 = (0,0)$, $x_5 = (2,0)$, $x_6 = (2,1)$ with corresponding probabilities $p_1 = 0.1$, $p_2 = 0.2$, $p_3 = 0.15$, $p_4 = 0.2$, $p_5 = 0.3$, $p_6 = 0.05$. This time we apply a slightly different approach to that in the previous example.

```
> x := [1,2],[1,-1],[1,1],[0,0],[2,0],[2,1];
```

$$x := [1,\ 2],\ [1,\ -1],\ [1,\ 1],\ [0,\ 0],\ [2,\ 0],\ [2,\ 1]$$

```
> p := 0.1,0.2,0.15,0.2,0.3,0.05;
```

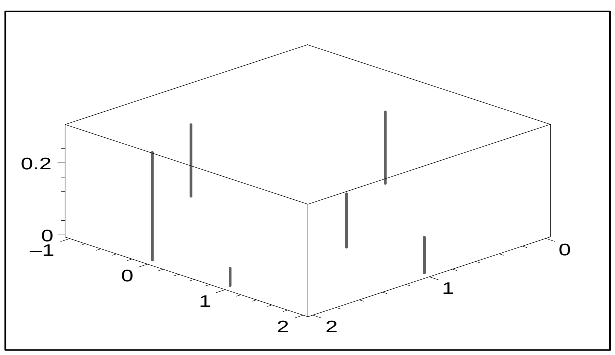
$$p := .1,\ .2,\ .15,\ .2,\ .3,\ .05$$

We make the list of probability bars:

```
> zip((x,p)->[[op(x),0],[op(x),p]], [x],[p]);
```

$$[[[1,\ 2,\ 0],\ [1,\ 2,\ .1]],\ [[1,\ -1,\ 0],\ [1,\ -1,\ .2]],\ [[1,\ 1,\ 0],\ [1,\ 1,\ .15]],$$
$$[[0,\ 0,\ 0],\ [0,\ 0,\ .2]],\ [[2,\ 0,\ 0],\ [2,\ 0,\ .3]],\ [[2,\ 1,\ 0],\ [2,\ 1,\ .05]]]$$

```
> PLOT3D( CURVES(op(%),THICKNESS(5),COLOR(HUE,0.99)),
  AXESSTYLE(BOX),AXESTICKS(3,3,2));
```



Let us check that it is actually a probability distribution.

```
>  add(p[i],i=1..nops([p]));
```

$$1.00$$

. . .

A rather trivial discrete distribution is a one-point distribution, i.e. a distribution P concentrated at a single point, say in point $c \in \mathbb{R}^n$. Then:

$$P(\{c\}) = 1.$$

Despite its simplicity an important discrete distribution is the two-point distribution, i.e. a distribution P concentrated at two points. Usually one considers a one dimensional two-point distribution concentrated in points 0 and 1. Then we have:

$$P(\{0\}) = q, \qquad P(\{1\}) = p,$$

where $p, q > 0$ and $p + q = 1$.

In future, for convenience, we will just write $P(a)$ instead of $P(\{a\})$ for singleton events. Later in this book we will discuss more interesting discrete distributions.

3.1.2 Continuous Distributions

Definition 3.1.3 *A distribution $P : \mathcal{B}(\mathbb{R}^n) \longrightarrow \mathbb{R}$ is called continuous if there exists a function $f : \mathbb{R}^n \longrightarrow \mathbb{R}$, called its density function, such that for any Borel set $A \subset \mathbb{R}^n$ we have:*

$$P(A) = \int_A f(x)\,dx = \int_A f\,dx, \tag{3.2}$$

where the integral is taken with respect to the Lebesgue measure $\mu = \mu_n$ restricted to $\mathcal{B}(\mathbb{R}^n)$ (see Section 2.2).

An example of a continuous distribution is height of a man shown in the figure on page 62. The probability of any set A equals the area under the graph over A.

Note that a density function f has to be integrable and satisfy the two following conditions:

1. $f \geq 0$ a.e.
2. $\displaystyle\int_{\mathbb{R}^n} f(x)\,dx = 1$

One can show that an integrable function f satisfying these two conditions is the density function of some continuous distribution (see E 2.7.

Later in the book we will examine some basic continuous distributions. For now we consider the simplest such distribution.

Example 3.1.1 Let G be a Borel set of positive and finite Lebesgue measure. Define a function

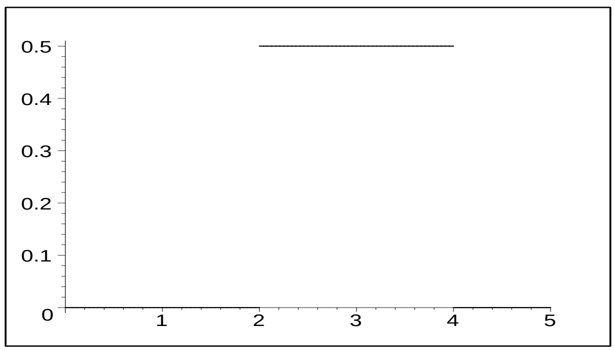
$$f(x) = \begin{cases} 0, & \text{if } x \notin G \\ \dfrac{1}{\mu(G)} & \text{if } x \in G \end{cases}$$

This f clearly satisfies the conditions of a density function. The resulting distribution is called the *uniform distribution* on the set G, and sometimes is denoted by $U(G)$. Compare this example with the definition of the geometric probability (see Definition 2.1.3).

Let us graph the density of the uniform distribution on the interval $(2,4)$, $U(2,4)$.

Maple Assistance 21

```
> f := x->piecewise(x>2 and x<4,1/2):
> plot(f,0..5,discont=true,thickness=3,color=BLACK);
```



As we have already mentioned in the case of a continuous one dimensional distribution, it is easy to "see" the probability of a given set. For an n-dimensional distribution with density f it is the Lebesgue measure of the set:

$$\{(x,y) \in \mathbb{R}^n \times \mathbb{R} : x \in A, 0 \le y \le f(x)\}.$$

This indicates that the probability of all singletons and thus that of any finite or countable set is zero. In fact, formula (3.2) applied to a singleton set A immediately implies this assertion.

3.1.3 Distribution Function

One dimensional probability distributions play an exceptional role amongst probability distributions. Here the use of distribution functions greatly simplifies the examination of distributions.

Definition 3.1.4 *A function $F : \mathbb{R} \longrightarrow \mathbb{R}$ is called a distribution function or a cumulative distribution function if the following conditions are satisfied:*

1. *F is increasing: $x < y \Rightarrow F(x) \leq F(y)$;*
2. *F is continuous from the right: $\lim_{x \to a^+} F(x) = F(a)$, for all $a \in \mathbb{R}$;*
3. *$\lim_{x \to \infty} F(x) = 1$;*
4. *$\lim_{x \to -\infty} F(x) = 0$.*

There is a close correspondence between probability distributions and distribution functions.

Theorem 3.1.1 *If P is a one dimensional probability distribution, then the function $F = F_P$ defined by*

$$F(x) = P((-\infty, x]) = P(-\infty, x] \tag{3.3}$$

is a distribution function. ($P((-\infty, x])$ will usually be written $P(-\infty, x]$).

Conversely: for any distribution function F there exists a unique probability distribution P such that formula (3.3) is satisfied.

Proof. Let P be a probability distribution.

1. For $x < y$ we have: $(-\infty, x] \subset (-\infty, y]$, so $F(x) = P(-\infty, x] \leq P(-\infty, y] = \text{F(y)}$.

2. Let $x_n \to a^+$, as $n \to \infty$, and assume it is a decreasing sequence (i.e. $x_n > x_{n+1} > a$). To prove Point 2 we will show that $F(x_n) \to F(a)$. We have $F(x_n) = P(-\infty, x_n]$, and the sets $(-\infty, x_n]$ form a decreasing sequence with the intersection $= (-\infty, a]$. Property 7 in Proposition 1.1.1 completes the proof.

3. The arguments are quite similar to those above. We just have to note that for an increasing sequence with x_n going to infinity, the sets $(-\infty, x_n]$ form an increasing sequence with their union equal to $\mathbb{R}$.

4. As in Point 3, but we take a decreasing sequence x_n tending to $-\infty$.

The proof of the second statement is more advanced and therefore will be omitted. Readers interested in the proof are referred to Halmos [13]. ⊡

Incidentally, some authors define a distribution function with condition 2 replaced by the condition that F is continuous from the left. In such a case formula (3.3) becomes $F(x) = P(-\infty, x)$. Certainly, both approaches are equally valid.

. . .

An important question concerns the nature of a discontinuity of a distribution function at a given point. The following proposition brings the answer.

Proposition 3.1.1 *Let P be a probability distribution, $F = F_P$ be a corresponding distribution function, and $a \in \mathbb{R}$. Then*

$$P(a) = F(a) - F(a^-), \tag{3.4}$$

where $F(a^-)$ denotes the left hand side limit of F at the point a. (This limit exists since F is increasing).

Hence F is continuous at a if and only if $P(a) = 0$.

Proof. Let x_n be an increasing sequence. Then $(-\infty, a) = \bigcup_n (-\infty, x_n]$ and hence $F(a^-) = \lim_{n\to\infty} F(x_n) = \lim_{n\to\infty} P(-\infty, x_n] = P(-\infty, a)$. Hence $P(a) = P((-\infty, a] \setminus (-\infty, a)) = P(-\infty, a] - P(-\infty, a) = F(a) - F(a^-)$. □

In other words, a continuity point of a distribution function has zero probability, and a discontinuity point has probability equal to the "jump" of the distribution function.

. . .

The distribution function is easy to find for both discrete and continuous probability distribution. We have:

1. In the discrete case:
$$F_P(x) = \sum_{i: x_i \le x} p_i. \tag{3.5}$$

2. In the continuous case:
$$F_P(x) = \int_{-\infty}^{x} f(s)\, ds. \tag{3.6}$$

The distribution function of a continuous probability distribution is thus continuous at every point. The converse is not true, however. Moreover, (3.6) implies that

$$\frac{d}{dx} F_P(x) = f(x) \tag{3.7}$$

at any continuity point x of the density f.

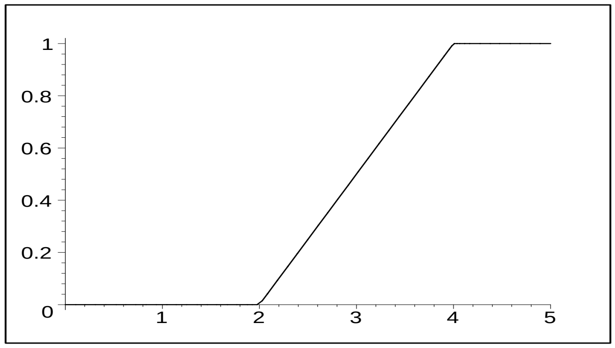
Example 3.1.2 The distribution function of the uniform distribution $U(a, b)$ on the interval (a, b) is

$$F(x) = \begin{cases} 0, & x < a \\ \dfrac{x-a}{b-a}, & a \le x < b \\ 1, & b \le x \end{cases}$$

We graph the distribution function of the uniform distribution on the interval $(2, 4)$.

MAPLE ASSISTANCE 22

```
> F := x->piecewise(x>2 and x<4,(x-2)/2,x>4,1):
> plot(F,0..5,thickness=3,color=BLACK);
```



We check the validity of formula (3.7) in the above case.

MAPLE ASSISTANCE 23

```
> diff(F(x),x);
```

$$\begin{cases} 0 & x < 2 \\ \mathit{undefined} & x = 2 \\ \frac{1}{2} & x < 4 \\ \mathit{undefined} & x = 4 \\ 0 & 4 < x \end{cases}$$

. . .

The definition and some properties of a distribution function can also be established in n-dimensional space with a similar correspondence between distribution functions and probability distributions to that in the one dimensional case. The only two complications are with an appropriate formulation of properties that in one dimension are expressed by means of the ordering "$\leq$" of the real line and by means of ∞ and $-\infty$.

3.2 Random Variables and Random Vectors

We will start with the definition of a random variable and then generalise it to the definition of a random vector. Let (Ω, Σ, P) be a probability space.

Definition 3.2.1 *A function $\xi : \Omega \longrightarrow \mathbb{R}$ is called a random variable if it is measurable, i.e.:*

$$\xi^{-1}(B) = \{\omega \in \Omega : \xi(\omega) \in B\} \in \Sigma,$$

for any set $B \in \mathcal{B}(\mathbb{R})$.

Sets of the form $\xi^{-1}(B)$, where $B \in \mathcal{B}(\mathbb{R})$, are said to be specified by ξ. We will also use the shorter notation $\{\xi \in B\}$ for $\xi^{-1}(B)$. For example: $P(\xi < \varepsilon)$ means $P(\{\omega \in \Omega : \xi(\omega) < \varepsilon\})$.

3.2.1 A Problem of Measurability

We would like to explain the idea of the above definition, especially since many applied statistics texts omit or postpone the assumption on measurability and call any function $\xi : \Omega \longrightarrow \mathbb{R}$ a random variable. This may not cause difficulties in some simple cases, but may lead to serious mistakes in more advanced cases. We can think of measurability as being consistency with the information in the probability space under consideration.

Let us note first that in the case $\Sigma = \mathcal{P}(\Omega)$, as in the classical scheme $\xi^{-1}(B) \in \Sigma$ for any B, any function $\xi : \Omega \longrightarrow \mathbb{R}$ is a random variable. On the other hand, if Σ is smaller than $\mathcal{P}(\Omega)$, then one can find functions $\xi : \Omega \longrightarrow \mathbb{R}$ that are not random variables. In the extreme case, where Σ is the σ-algebra containing only the empty set $\emptyset$ and the whole space Ω, only the constant functions $\Omega \longrightarrow \mathbb{R}$ are random variables. The following example illustrates what is and what is not a random variable.

Example 3.2.1 Consider an experiment of tossing two dice where we are interested in the maximum of pips. The set of elementary events is then $\Omega = \{(i,j) : i,j = 1,\ldots,6\}$, the σ-algebra $\Sigma = \sigma(A_1, \ldots A_6)$ is the σ-algebra of all unions of sets $A_k = \{(i,j) : \max(i,j) = k\}$. (The measure P here may be defined quite arbitrarily.) Define three functions $\Omega \longrightarrow \mathbb{R}$:

$$\xi_1(i,j) = \max(i,j),$$

$$\xi_2(i,j) = \begin{cases} 1, & \text{if } i > 3 \text{ or } j > 3, \\ 0, & \text{otherwise,} \end{cases}$$

$$\xi_3(i,j) = i + j.$$

It is easy to see that the first and the second function are random variables on (Ω, Σ, P), while the third function is not. In fact, the set $\xi_3^{-1}(-\infty, 3] = \{(1,1),(1,2),(2,1)\}$ is not a union of sets A_k, so does not belong to Σ.

The idea here is that the sets specified by the first two functions provide some information about the experiment. For example $\xi_2 = 0$ means the maximum ≤ 3, and therefore they are events. On the other hand, sets specified by ξ_3 may not be events and we are not able to say here whether or not $\xi_3^{-1}(-\infty, 3] = \{(1,1),(1,2),(2,1)\}$ occurs.

3.2.2 Distribution of a Random Variable

A random variable induces a probability distribution on the measurable space $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$, which is often more convenient to work with.

Definition 3.2.2 *Let (Ω, Σ, P) be a probability space and let $\xi : \Omega \longrightarrow \mathbb{R}$ be a random variable. The probability distribution P_ξ defined by*

$$P_\xi(B) = P(\xi^{-1}(B)), \text{ for } B \in \mathcal{B}(\mathbb{R}) \tag{3.8}$$

is called the distribution induced by the random variable ξ

We note that the measurability of ξ guarantees the validity of the expression $P(\xi^{-1}(B))$. Also, one easily checks that P_ξ is in fact the probability measure. Thus $(\mathbb{R}, \mathcal{B}(\mathbb{R}), P_\xi)$ is a probability space.

Example 3.2.2 Find the distribution of the random variable ξ_2 from Example 3.2.1. Assume that the measure P is defined according to the classical scheme. We have: $P_\xi(x) = P(\xi^{-1}(x)) = 0$ for $x \neq 0$ and $x \neq 1$, and $P_\xi(0) = P(A_1 \cup A_2 \cup A_3) = \frac{9}{36}$, $P_\xi(1) = P(A_4 \cup A_5 \cup A_6) = \frac{27}{36}$. So P_ξ is a two-point distribution.

Some examples of random variables and their distributions have already been mentioned on page 61. However, we said nothing about the probability space the random variables were defined on. Actually, it is a typical situation. Most often we deal with a random quantity and what we really see is a probability distribution characterising this quantity. Then we believe that the quantity is a random variable ξ on some probability space (Ω, Σ, P) and the probability distribution we see is just P_ξ. This, perhaps rather naive, reasoning is fortunately justified by the following:

Theorem 3.2.1 *Let $Q : \mathcal{B}(\mathbb{R}) \longrightarrow \mathbb{R}$ be a probability distribution. Then there is a probability space (Ω, Σ, P) and a random variable $\xi : \Omega \longrightarrow \mathbb{R}$ such that*

$$Q = P_\xi.$$

Proof. Simply take $\Omega = \mathbb{R}$, $\Sigma = \mathcal{B}(\mathbb{R})$, $P = Q$ and $\xi(\omega) = \omega$ for $\omega \in \mathbb{R}$. □

• • •

Random variables share a lot of useful algebraic properties with general measurable functions (see Theorem 2.2.2 on page 43). We just recall that:

1. The indicator I_A of a set $A \in \Sigma$ is a random variable.
2. The sum, difference, product, quotient (if defined) of random variables is a random variable.
3. The minimum and maximum of two random variables is a random variable.

4. The supremum and infimum of a sequence of random variables is a random variable.
5. The limit of a sequence of random variables is a random variable.
6. The composition $g \circ \xi$, of a random variable ξ with a continuous function g is a random variable.

3.2.3 Random Vectors

A random vector is a straightforward generalisation of a random variable.

Definition 3.2.3 *A function $\xi : \Omega \longrightarrow \mathbb{R}^n$ is called a random vector if it is measurable, i.e.*

$$\xi^{-1}(B) = \{\omega \in \Omega : \xi(\omega) \in B\} \in \Sigma, \quad \textit{for any set } B \in \mathcal{B}(\mathbb{R}^n).$$

A random variable is a one dimensional random vector, so there were no formal reasons to distinguish the random variables case. We did so, however, due to tradition and because of the special role random variables play among random vectors.

As with random variables we can define the distribution of a random vector $X : \Omega \longrightarrow \mathbb{R}^n$ by:

$$P_X(B) = P(X^{-1}(B)), \text{ for } B \in \mathcal{B}(\mathbb{R}^n). \tag{3.9}$$

3.3 Independence

A pair (X, Y) of random vectors $X,\ Y$ defined on the same probability space, in particular a pair of random variables, is a random vector. What is the relationship between the distributions P_X, P_Y and $P_{(X,Y)}$? In general, there is only a partial relationship, namely, the definition of the distribution of the random vector implies:

$$P_X(A_1) = P_{(X,Y)}(A_1 \times \mathbb{R}^n), \qquad P_Y(A_2) = P_{(X,Y)}(\mathbb{R}^m \times A_2). \tag{3.10}$$

Hence, the knowledge of the distribution of the pair of random vectors implies the knowledge of the distributions of its "coordinates". Note that knowing the probability space that the random vectors are defined on is not essential here. On the other hand, the probability space is important for questions like the converse problem: determine the distribution $P_{(X,Y)}$ from the distributions P_X and P_Y. The problem is illustrated with the following:

Example 3.3.1 Consider two different situations.

Case 1. Let (Ω, Σ, P) be a classical scheme with $\Omega = \{1, 2, 3, 4, 5, 6\}$, $\xi, \eta : \Omega \longrightarrow \mathbb{R}$ defined as: $\xi(\omega) = \eta(\omega) = \omega$. Here the measures $P_\xi = P_\eta$ are concentrated at the points $1, 2, 3, 4, 5, 6$ with equal probabilities $\frac{1}{6}$, while the measure $P_{(\xi,\eta)}$ is concentrated at the points $x_i = (i,i)$, $i = 1,\dots,6$ with $P_{(\xi,\eta)}(x_i) = \frac{1}{6}$.

Case 2. Let (Ω, Σ, P) be the classical scheme with $\Omega = \{1,2,3,4,5,6\}^2$, $\xi, \eta : \Omega \longrightarrow \mathbb{R}$ defined as: $\xi(\omega_1,\omega_2) = \omega_1$, $\eta(\omega_1,\omega_2) = \omega_2$. As above $P_\xi = P_\eta$ are concentrated at the points $1,2,3,4,5,6$ with equal probabilities $\frac{1}{6}$, while the measure $P_{(\xi,\eta)}$ is concentrated at the points $x_{ij} = (i,j)$, $i,j = 1,\dots,6$ with $P_{(\xi,\eta)}(x_{ij}) = \frac{1}{36}$.

In both cases the random variables ξ, η had the same distributions, but the distributions of the pair (ξ, η) were different.

Note that in Case 2 we have: $P_{(\xi,\eta)}(i,j) = P_\xi(i)P_\eta(j)$ for all i,j, which means that the distribution $P_{(\xi,\eta)}$ is the (Cartesian) product of P_ξ and P_η. The fundamental reason is that the random variables in Case 2 are independent. What independence of the random variables means will be formally defined below, though we already believe that the toss of a die is independent of the toss of another die, so ξ, η "should be" independent.

Definition 3.3.1 *Let (Ω, Σ, P) be a probability space and let $X_1,\dots,X_k$ be random vectors defined on this space. We say that $X_1,\dots,X_k$ are independent if for every choice of Borel sets $B_1,\dots,B_k$ contained in the appropriate spaces we have:*

$$P(X_1 \in B_1,\dots,X_k \in B_k) = P(X_1 \in B_1)\cdot\ldots\cdot P(X_k \in B_k). \tag{3.11}$$

In other words, random vectors are independent if the events specified by these vectors are independent.

We say that the random vectors $X_1, X_2, X_3, \dots$ are independent, if for any k, the random vectors $X_1,\dots,X_k$ are independent.

In the last example it is obvious that we have a pair of dependent random variables in Case 1 and a pair of independent random variables in Case 2. This example suggests the following:

Theorem 3.3.1 *Let X, Y be random vectors defined on the same probability space (Ω, Σ, P). Then*

$$X,\ Y \text{ are independent} \iff P_{(X,Y)} = P_X \times P_Y.$$

Moreover, we then have:

1. *If random vectors X and Y have discrete distributions and*

$$P(X = x_i) = p_i, \qquad P(Y = y_j) = q_j,$$

then (X,Y) has a discrete distribution with

$$P(X = x_i, Y = y_j) = p_i \cdot q_j \tag{3.12}$$

2. *If random vectors X and Y have continuous distributions with densities f and g, then (X,Y) has the continuous distribution with density*

$$h(x,y) = f(x)\cdot g(y) \tag{3.13}$$

The above formulas are very useful for determining distributions of functions of random variables and vectors and will be used later in the book.

3.4 Functions of Random Variables and Vectors

The following situation is often encountered. Given a random vector X with known distribution P_X and given a function φ, find the distribution $P_{\varphi(X)}$ of the random vector $\varphi(X) = \varphi \circ X$. For example, we may want to find the distribution of ξ^2 given the distribution of a random variable ξ, or we may want find the distribution of $\max(\xi, \eta)$ given the distributions of ξ and η.

The theoretical solution to this problem is remarkably easy, but its practical meaning is fairly limited. Nevertheless, we present it first and will then consider a more practical approach.

Let (Ω, Σ, P) be a probability space, $X : \Omega : \longrightarrow \mathbb{R}^n$ a random vector and $\varphi : \mathbb{R}^n : \longrightarrow \mathbb{R}^k$ a function. We are interested in the composition

$$\varphi \circ X : \Omega \longrightarrow \mathbb{R}^k \quad \text{defined by} \quad (\varphi \circ X)(\omega) = \varphi(X(\omega)).$$

For a broad class of functions φ one can prove that this composition is a random vector, for example, if φ is a Borel function, i.e. measurable with respect to the space $(\mathbb{R}^n, \mathcal{B}(\mathbb{R}^n))$. For historical reasons we often write $\varphi(X)$ instead of $\varphi \circ X$. Its distribution is then:

$$P_{\varphi(X)}(C) = P((\varphi \circ X)^{-1}(C)) = P(X^{-1}(\varphi^{-1}(C))) = P_X(\varphi^{-1}(C))$$

for any Borel subset C of the space $\mathbb{R}^k$. As we have already mentioned this nice and general formula is not very convenient in practice because one usually considers distribution functions or densities rather than probability distributions.

In the following we offer a few examples illustrating common methods of finding distributions of functions of random vectors.

Example 3.4.1 Let ξ be a random variable having discrete distribution concentrated at the points -1, 0, 1, 2, 3 with equal probability $\frac{1}{5}$. The distribution of the random variable ξ^2 is found as follows: ξ^2 takes values 0, 1, 4, 9 with probabilities, respectively, $\frac{1}{5}, \frac{2}{5}, \frac{1}{5}, \frac{1}{5}$, since $(\pm 1)^2 = 1$.

Example 3.4.2 Let ξ be a random variable, $F = F_\xi$ its distribution function and let $a, b \in \mathbb{R}$ be fixed numbers with $a \neq 0$. We seek the distribution function of the random variable $\eta = a\xi + b$.

For $a > 0$ we have

$$F_\eta(x) = P(\eta \leq x) = P(a\xi + b \leq x) = P\left(\xi \leq \frac{x-b}{a}\right) = F_\xi\left(\frac{x-b}{a}\right).$$

Similarly, for $a < 0$ we have

$$F_\eta(x) = P(\eta \le x) = P(a\xi + b \le x) = P\left(\xi \ge \frac{x-b}{a}\right)$$

$$= 1 - P\left(\xi < \frac{x-b}{a}\right) = 1 - F_\xi\left(\left(\frac{x-b}{a}\right)^-\right).$$

Assume additionally that ξ has a density f, where f is a continuous function. This is not necessary as the next example shows, but simplifies arguments here. Then (3.7) means that F_ξ is differentiable, so is F_η from the above formula too. By (3.7) η has a continuous distribution with density:

$$g(x) = \frac{1}{|a|} f\left(\frac{x-b}{a}\right). \tag{3.14}$$

Example 3.4.3 We generalise Example 3.4.2 to some extent. Let X be a random vector having an n-dimensional continuous probability distribution with density $f : \mathbb{R}^n \longrightarrow \mathbb{R}$. Assume $\varphi : \mathbb{R}^n \longrightarrow \mathbb{R}^n$ is a diffeomorphism, i.e. φ is invertible with both φ and φ^{-1} having continuous first derivatives. Then, the random vector $\varphi(X)$ also has continuous distribution with density g given by:

$$g(x) = \left|Jac_x \varphi^{-1}\right| \, f(\varphi^{-1}(x)),$$

where the Jacobian $Jac_x \varphi = \det \left[\frac{\partial \varphi_i}{\partial x_j}\right]$ is nonzero everywhere.

In fact, by Theorem 2.3.3, for any Borel set A we have

$$P(\varphi(X) \in A) = P(X \in \varphi^{-1}(A))$$

$$= \int_{\varphi^{-1}(A)} f(y)\, dy = \int_A f(\varphi^{-1}(x))\, |Jac_x \varphi^{-1}|\, dx.$$

Example 3.4.4 Let ξ be a random variable having uniform distribution on the interval $[-1, 1]$, i.e. its density is $f(x) = \frac{1}{2} I_{[-1,1]}$, where $I_{[-1,1]}$ is the indicator function of the interval $[-1, 1]$. We determine the distribution of the random variable $\frac{1}{\xi}$ by computing its distribution function G.

We have to compute $G(a)$ for any $a \in \mathbb{R}$. Let $a < 0$. Then

$$G(a) = P\left(\frac{1}{\xi} \le a\right) = P\left(\frac{1}{\xi} \le a, \xi < 0\right) = P\left(\frac{1}{a} \le \xi < 0\right)$$

$$= \int_{\frac{1}{a}}^{0} f(x)\, dx = \begin{cases} -\dfrac{1}{2a}, & \text{for } a \le -1 \\ \dfrac{1}{2}, & \text{for } -1 < a \le 0. \end{cases}$$

The cases $a > 0$ and $a = 0$ are even simpler and hence are left to the reader. We then have

$$G(a) = \begin{cases} -\dfrac{1}{2a}, & \text{for } a \leq -1 \\ \dfrac{1}{2}, & \text{for } -1 < a \leq 1 \\ 1 - \dfrac{1}{2a}, & \text{for } 1 < a. \end{cases}$$

We see that the distribution function G is differentiable at any point except at -1 and 1. Formula (3.7) can be used to determine the density:

$$g(a) = G'(a) = \begin{cases} \dfrac{1}{2a^2}, & \text{for } a < -1 \\ 0, & \text{for } -1 < a < 1 \\ \dfrac{1}{2a^2}, & \text{for } 1 < a. \end{cases}$$

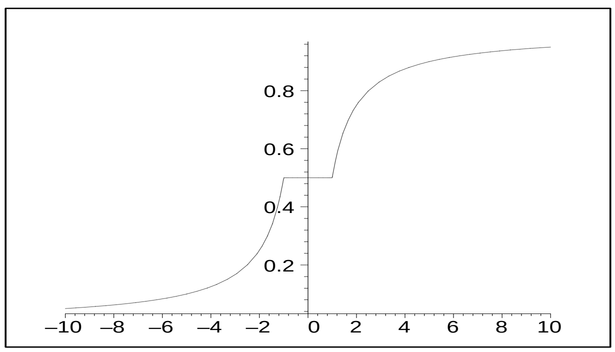
We sketch the graph of the distribution function G and the density g:

MAPLE ASSISTANCE 24

```
> G := a->piecewise(a<=-1,-1/(2*a),
  -1<a and a <=1,1/2,
  1<a,1-1/(2*a));
```

$$G := a \to \text{piecewise}(a \leq -1,\ -\frac{1}{2}\frac{1}{a},\ -1 < a \textbf{ and } a \leq 1,\ \frac{1}{2},\ 1 < a,\ 1 - \frac{1}{2}\frac{1}{a})$$

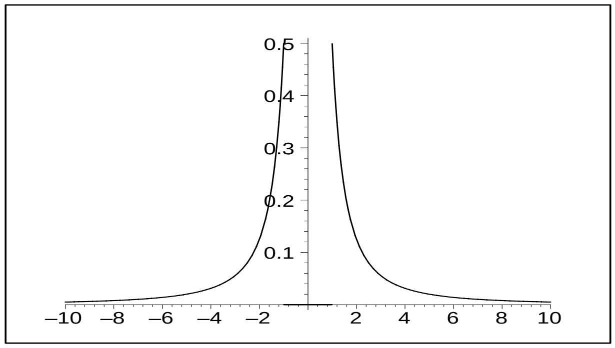
```
> plot(G);
```




```
> g := D(G);
```

$$g := a \to \text{piecewise}(a \leq -1,\ \frac{1}{2}\frac{1}{a^2},\ -1 < a \textbf{ and } a \leq 1,\ 0,\ 1 < a,\ \frac{1}{2}\frac{1}{a^2})$$

```
> plot(g,discont=true,thickness=3,color=BLACK);
```



3.4.1 Distributions of a Sum

Example 3.4.5 We will discuss the distribution of the sum of a random variable for the continuous case only since the discrete one is simpler and is thus left to the reader (see Problems 3.10 and 3.11).

Assume that a two dimensional random vector (ξ, η) has continuous distribution with density f. First we find the distribution function of the sum $F = F_{\xi+\eta}$. Fix $a \in \mathbb{R}$. Then, $F(a) = P(\xi + \eta \leq a) = P((\xi, \eta) \in A)$, where $A = \{(x, y) : x + y \leq a\}$. Hence,

$$F(a) = \int_A f(x, y)\, d(x, y) = \int_{-\infty}^{a} \left(\int_{-\infty}^{\infty} f(t, s - t)\, dt \right) ds.$$

We have used the substitution $t = x$, $s = x + y$. Differentiating with respect to a and setting $x = a$ we obtain the formula for the density of the sum:

$$f_{\xi+\eta}(x) = \int_{-\infty}^{\infty} f(t, x - t)\, dt.$$

Assume now that the random variables ξ and η are independent. Then, by formula (3.13), the last equality simplifies to:

$$f_{\xi+\eta}(x) = \int_{-\infty}^{\infty} f_\xi(t)\, f_\eta(x - t)\, dt. \tag{3.15}$$

. . .

To complete this section we state a useful theorem.

Theorem 3.4.1 *Let* $X : \Omega \longrightarrow \mathbb{R}^n$ *and* $Y : \Omega \longrightarrow \mathbb{R}^m$ *be independent random vectors and let* $g : \mathbb{R}^n \longrightarrow \mathbb{R}^k$, $h : \mathbb{R}^m \longrightarrow \mathbb{R}^l$ *be Borel functions, i.e. measurable with respect to appropriate* σ*-algebras of Borel sets. Then the random vectors* $g(X)$ *and* $h(Y)$ *are independent.*

3.5 MAPLE Session

M 3.1 *Write* MAPLE *commands for the figure on page 62 of the distribution of height.*

```
> sigma := 4: m := 177:
> w := 1/(sqrt(2*Pi)*sigma);
```

$$w := \frac{1}{8}\,\frac{\sqrt{2}}{\sqrt{\pi}}$$

```
> f := x -> w*exp((-((x-m)/sigma)^2/2));
```

$$f := x \to w\,e^{(-1/2\,\frac{(x-m)^2}{\sigma^2})}$$

```
> with(plottools):
> an := plot(f(x),x = 160..190):
  bn := seq(curve([[i/2,0],[i/2,f(i/2)]],linestyle = 2),
  i = 360..370):
> plots[display]({an,bn},axes=FRAMED,
  view=[160..190,0.. 0.1]);
```

M 3.2 *Verify that the function f used in the previous problem really is a density function.*

```
> int(f,-infinity..infinity);
```

$$1$$

M 3.3 *Write* MAPLE *commands for the figure on page 61.*

```
> PLOT(CURVES([[2,0],[2,0.05]],[[3,0],[3,0.1]],
  [[4,0],[4,0.4]],[[5,0],[5,0.45]],THICKNESS(3)),
  AXESTICKS([2=`NOPASS`,3=`PASS`,4=`GOOD`,5=`EXCELLENT`],2),
  VIEW(1.5..6,0..0.5));
```

M 3.4 *Write a procedure that determines the distribution function for any finite sequences* $\{x_i\}$ *and* $\{p_i\}$ *defining a discrete distribution.*

```
> FF := proc(x::list,p::list,xi)
  f[0] := 0:
  for i from 1 to nops(x) do
  f[i] := f[i-1] + p[i]
  od:
  if xi < x[1] then 0
  elif xi >= x[nops(x)] then 1
  else
  for i from 1 to nops(x)-1 do
  if x[i] <= xi and xi < x[i+1] then RETURN(f[i]) fi
  od
  fi
  end:

Warning, `f` is implicitly declared local
```

```
Warning, 'i' is implicitly declared local
```

Specify some sequences:

```
> x := [1,3,4,6]: p := [0.2,0.3,0.4,0.1]:
```

Define the distribution:

```
> F := xi->FF(x,p,xi);
```

$$F := \xi \to \mathrm{FF}(x,\, p,\, \xi)$$

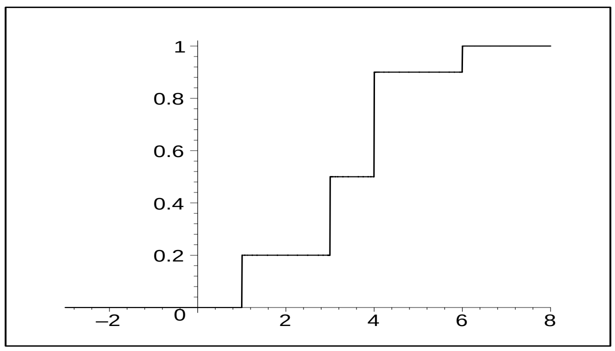
Does it work?

```
> F(3);
```

$$.5$$

Make the plot:

```
> plot(F,-3..8,thickness=3,color=BLACK);
```



The above method works under an assumption that the sequence x is given in the natural order. If not, then we have to order it first, remembering that the correspondence between the x_i's and p_i's is essential.

Fix some sequences:

```
> x := [1,4,3,5,2]: p := [0.05,0.3,0.4,0.1,0.05]:
```

Define the criterion function which is necessary to indicate the order of pairs by their first elements.

```
> crit := (w,z)->(w[1] <= z[1]);
```

$$crit := (w,\, z) \to w_1 \le z_1$$

Combine the lists x and p into a list of pairs:

```
> a := [seq([x[i],p[i]],i=1..nops(x))];
```

$$a := [[1,\, .05],\, [4,\, .3],\, [3,\, .4],\, [5,\, .1],\, [2,\, .05]]$$

Sort the list a by applying `crit`. Pay particular attention to the use of the composition operator @ and the command `is`.

```
> a := sort(a,is@crit);
```

$$a := [[1,\ .05],\ [2,\ .05],\ [3,\ .4],\ [4,\ .3],\ [5,\ .1]]$$

Return now to the lists x and p.

```
> x := [seq(a[i][1],i=1..nops(x))];
```

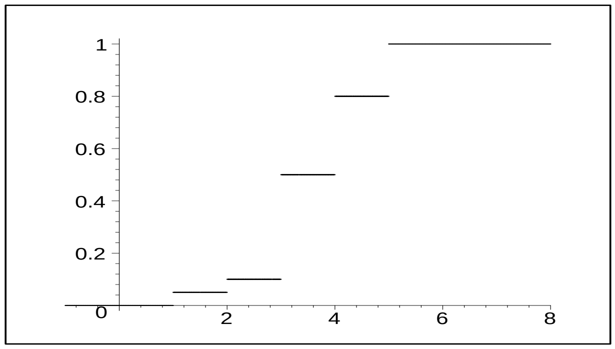
$$x := [1,\ 2,\ 3,\ 4,\ 5]$$

```
> p := [seq(a[i][2],i=1..nops(x))];
```

$$p := [.05,\ .05,\ .4,\ .3,\ .1]$$

Plot the distribution removing unwanted vertical segments.

```
> plot(F,-1..8,discont=true,thickness=3,color=BLACK);
```



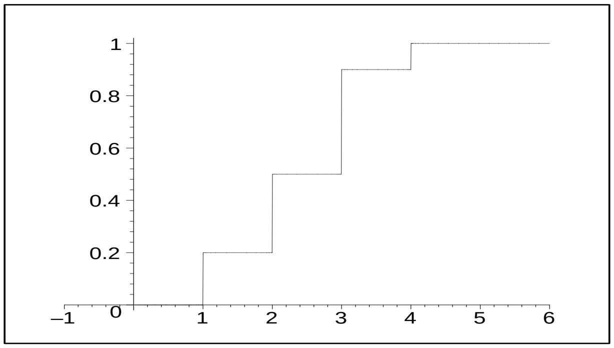
M 3.5 *Plot the distribution function of the random variable taking points* $1, 2, 3, 4$ *with probabilities* $0.2,\ 0.3,\ 0.4,\ 0.1$.

This time we will use the `stats package`.

```
> restart:
> with(stats):
> F := x ->
  statevalf[dcdf,empirical[0.2,0.3,0.4,0.1]](floor(x));
```

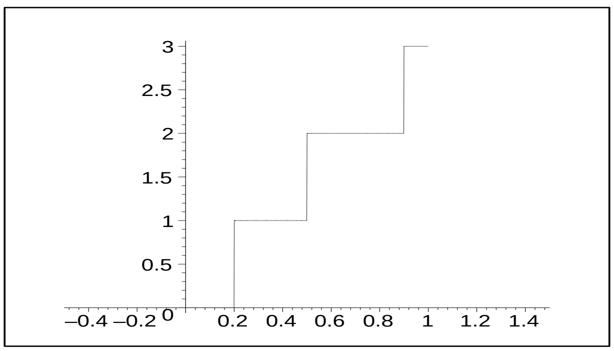
$$F := x \rightarrow \mathit{statevalf}_{\mathit{dcdf},\ \mathit{empirical}_{.2,\ .3,\ .4,\ .1}}(\mathrm{floor}(x))$$

```
> plot(F,-1..6,thickness=2);
```



M 3.6 *For any given $p \in [0,1]$ define the inverse distribution function by $invF(p) = \inf\{x \in \mathbb{R} : p \le F(x)\}$. Plot $invF(p)$ for a previous example.*

```
> invF := x->statevalf[idcdf,empirical[0.2,0.3,0.4,0.1]](x);
> plot(invF,-0.5..1.5,thickness=2);
```



M 3.7 *Find the densities of sums $S_n = \xi_1 + \ldots + \xi_n$ for $n = 2, \ldots, 7$, when $\xi_1, \ldots, \xi_n$ are independent random variables with the common density*

$$f(x) = \begin{cases} 0, & x < 0 \\ 3\mathrm{e}^{-3x}, & 0 \le x. \end{cases}$$

Define the convolution using formula (3.15). Note that for a function that is zero for $x < 0$ the limits of integration are in fact 0 and x in formula (3.15).

```
> conv := proc(f,g);
  x -> int(f(x-t)*g(t),t=0..x);
  end;
```

$$conv := \mathbf{proc}(f,\, g)\, x \to \mathrm{int}(f(x-t) \times g(t),\, t = 0..x)\, \mathbf{end}$$

Define the required density.

```
> f := x -> lambda*exp(-lambda*x);
```

$$f := x \to \lambda\, e^{(-\lambda\, x)}$$

Note that $S_{i+1} = S_i + \xi_{i+1}$ and use the result from Example 3.4.5.

```
> h1 := f:
> for i from 1 to 6 do
  conv(h.i,f)(x):
> h.(i+1) := unapply(%,x)
  od:
```

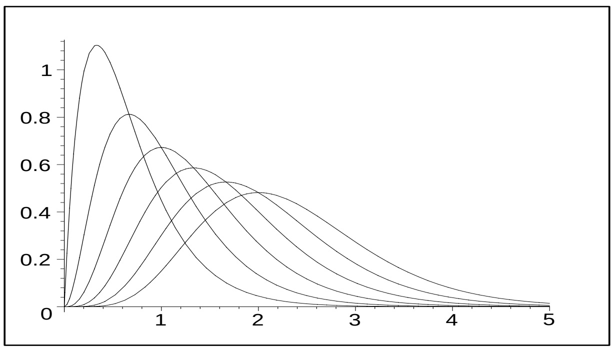
Put $\lambda = 3$. Check that $h7$ really is a density.

```
> lambda := 3:
> int(h.7,0..infinity);
```

$$1$$

Plot all the densities on the same axes:

```
> plot({h.($2..7)},0..5,color=BLACK);
```



3.6 Exercises

E 3.1 Find a constant C so that the function $P : \mathcal{B}(\mathbb{R}) \longrightarrow \mathbb{R}$ taking values $P(k) = \frac{C}{k(k+1)}$ for $k = 1, 2, 3, \ldots$, is a distribution.

E 3.2 Find a constant C so that the function

$$f(x) = \begin{cases} 0, & \text{for } x < 1 \\ \dfrac{C}{x^4}, & \text{for } x \geq 1 \end{cases}$$

is a density of a probability distribution P and determine $P(\{x \in \mathbb{R} : \frac{1}{x} < a\})$ for a given a.

E 3.3 Given a distribution P on $\mathbb{R}^2$ define the marginal distributions:

$$P_1(A) = P(A \times \mathbb{R}) \quad \text{and} \quad P_2(B) = P(\mathbb{R} \times B),$$

where A and B are Borel sets. Is it true that $P = P_1 \times P_2$?

E 3.4 Find $F(0)$, $F(0.3)$ and $F(3)$ for a distribution function F defined by $F = \frac{1}{3}F_1 + \frac{2}{3}F_2$, where F_1 is the distribution function of the discrete distribution with common value $\frac{1}{5}$ at points $0, 1, 2, 3, 4$, and F_2 is the distribution function of the uniform distribution on the interval $(0, 2)$. Sketch the graph F. Does F correspond to a continuous or discrete distribution?

Hint. Use the Law of Total Probability.

E 3.5 Let $\xi : \Omega \longrightarrow \mathbb{R}$ and assume that Ω is a union of pairwise disjoint sets $A_1, A_2, \ldots, A_k$. Let Σ be the σ-algebra consisting of all unions formed by these sets. Prove that ξ is Σ-measurable if and only if ξ is constant on each of the sets A_i, $i = 1, \ldots, k$.

E 3.6 Find $P(|\xi| > a)$, for $a = 0, \frac{1}{2}, 1, 2$, with the distribution F of a random variable ξ given by:

$$F(x) = \begin{cases} 0, & \text{for } x < -1 \\ \dfrac{1}{3}, & \text{for } -1 \le x < 0 \\ \dfrac{1}{3}(x+1), & \text{for } 0 \le x < 1 \\ 1, & \text{for } 1 \le x. \end{cases}$$

E 3.7 We choose an integer a randomly from the interval $[1, 10]$ according to the classical scheme and then we choose an integer ξ randomly from the interval $[1, a]$, again according to the classical scheme. Find the distribution of ξ. Determine two probability spaces on which the random variable ξ could be defined.

E 3.8 The independent random variables ξ and η are uniformly distributed on the interval $[0.1]$. Find the distributions of $\min(\xi, \eta)$ and $\max(\xi, \eta)$. Are these two variables independent?

E 3.9 Find the distribution of the sum of independent random variables that are uniformly distributed on the interval $[0, 1]$.

E 3.10 Let independent variables ξ and η have discrete distributions $P(\xi = k) = p_k$, $P(\eta = k) = q_k$, $k = 0, 1, 2, 3, \ldots$. Prove that the sum $\xi + \eta$ has the distribution

$$P(\xi + \eta = k) = \sum_{i=0}^{k} p_i q_{k-i}, \quad k = 0, 1, 2, 3, \ldots \tag{3.16}$$

E 3.11 Derive a formula like (3.16) assuming that both variables take all their values in a set $\{1, \ldots, N\}$, where N is a fixed integer.

E 3.12 Does the independence of ξ and η imply the independence of $\xi + \eta$ and $\xi - \eta$?

E 3.13 There are 5 white balls and 3 black balls in a box. Let T_w and T_b denote the number of drawings in which we first get, respectively, a white ball or a black ball. Find (a) The distributions of T_w and T_b. (b) The distribution of the random vector (T_w, T_b). Are T_w and T_b independent? Consider both cases: drawing without and with replacement.

E 3.14 Random variables ξ and η are independent and have common uniform distribution on the interval $(c - \frac{1}{2}, c + \frac{1}{2})$. Find the density of the difference $\xi - \eta$.

E 3.15 Given independent random variables ξ and η with common distribution Q find $P(\xi = \eta)$. Consider both cases, where Q is first discrete and then continuous.

E 3.16 Is it true that the random vector (ξ, η) has a continuous distribution when both ξ and η do?

E 3.17 Find the distribution of the sum of independent random variables ξ, η if ξ has the uniform distribution on the interval $[0, 1]$ and $P(\eta = 0) = P(\eta = 1) = \frac{1}{2}$.

http://www.springer.com/978-3-540-42666-0

From Elementary Probability to Stochastic Differential Equations with MAPLE®
Cyganowski, S.; Kloeden, P.; Ombach, J.
2002, XVI, 310 p. 19 illus., Softcover
ISBN: 978-3-540-42666-0