Random variable

POPULATION

Persons chosen at random

Height:	115 cm
Weight:	17 kg
No of children:	0
Employed:	No

Height:	195 cm
Weight:	98 kg
No of children:	4
Employed:	Yes

Height:	170 cm
Weight:	80 kg
No of children:	2
Employed:	No

Such quantities as

Height (cm)

Weight (cm)

Number of children (non-negative integer)

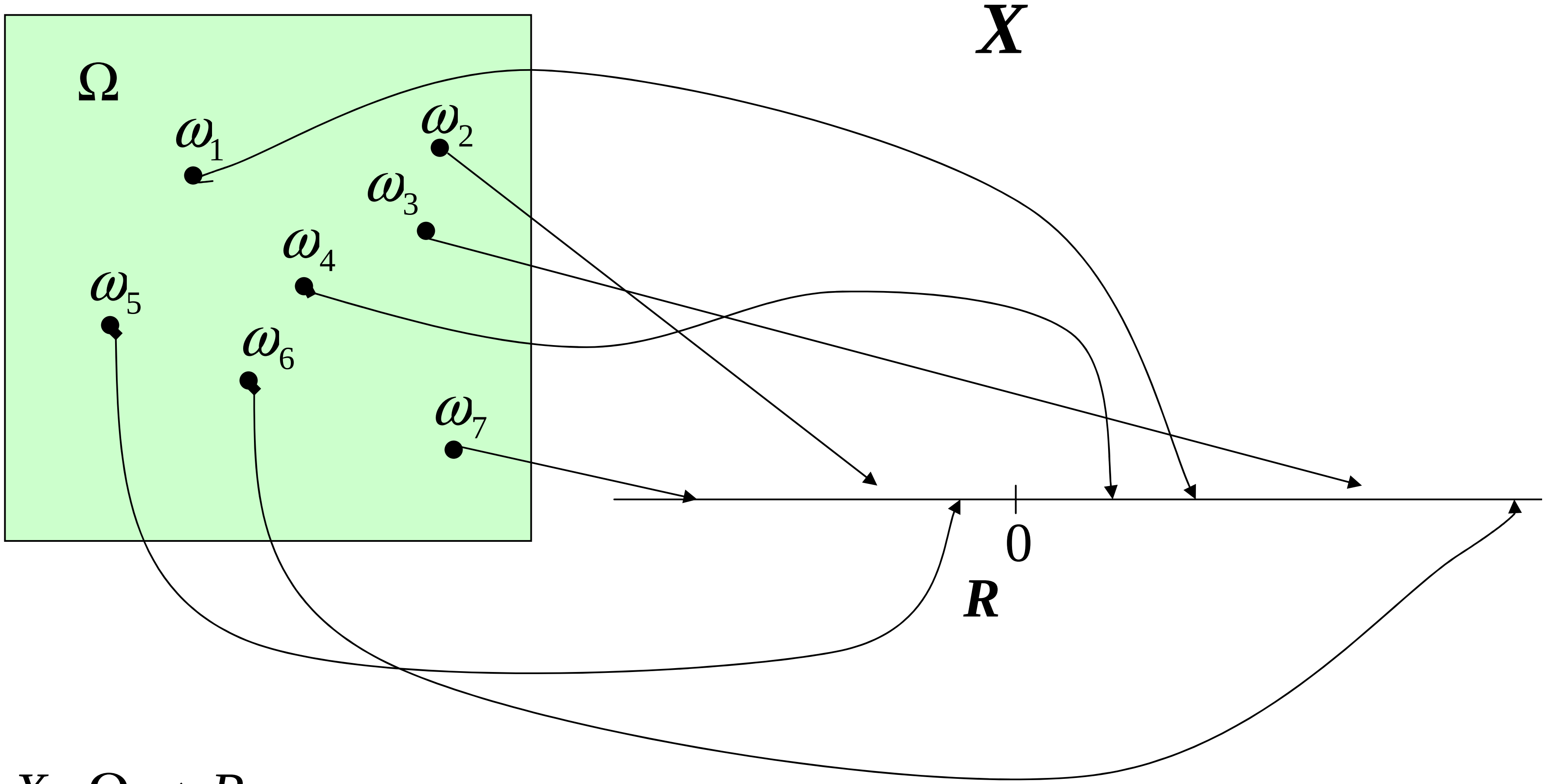
Employed (Yes/No)

are examples of random variables.

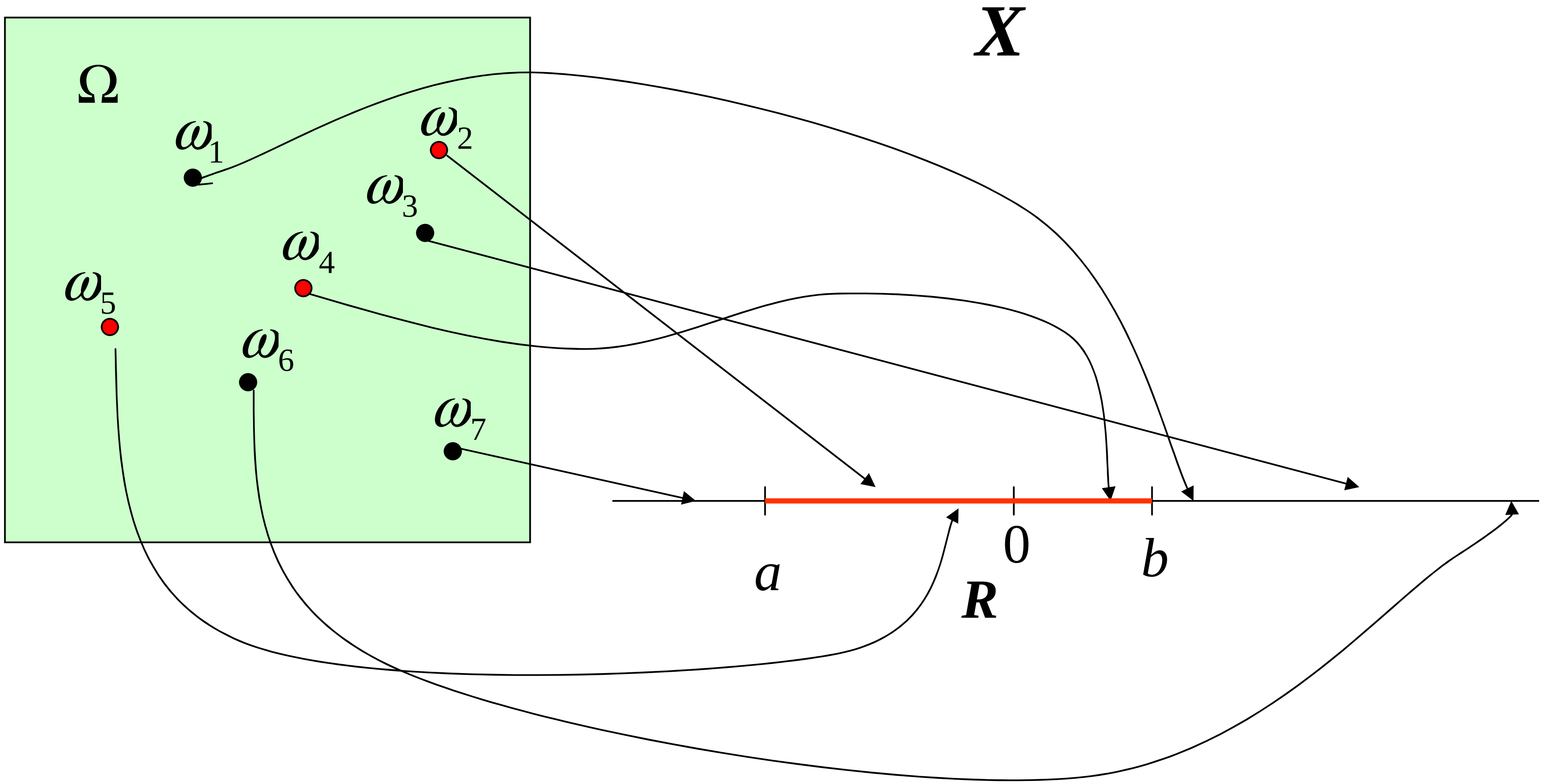
A random variable X assignes a real number to each outcome in the sample space

Thus each random variable X is a <u>mapping</u> not a number !!!.

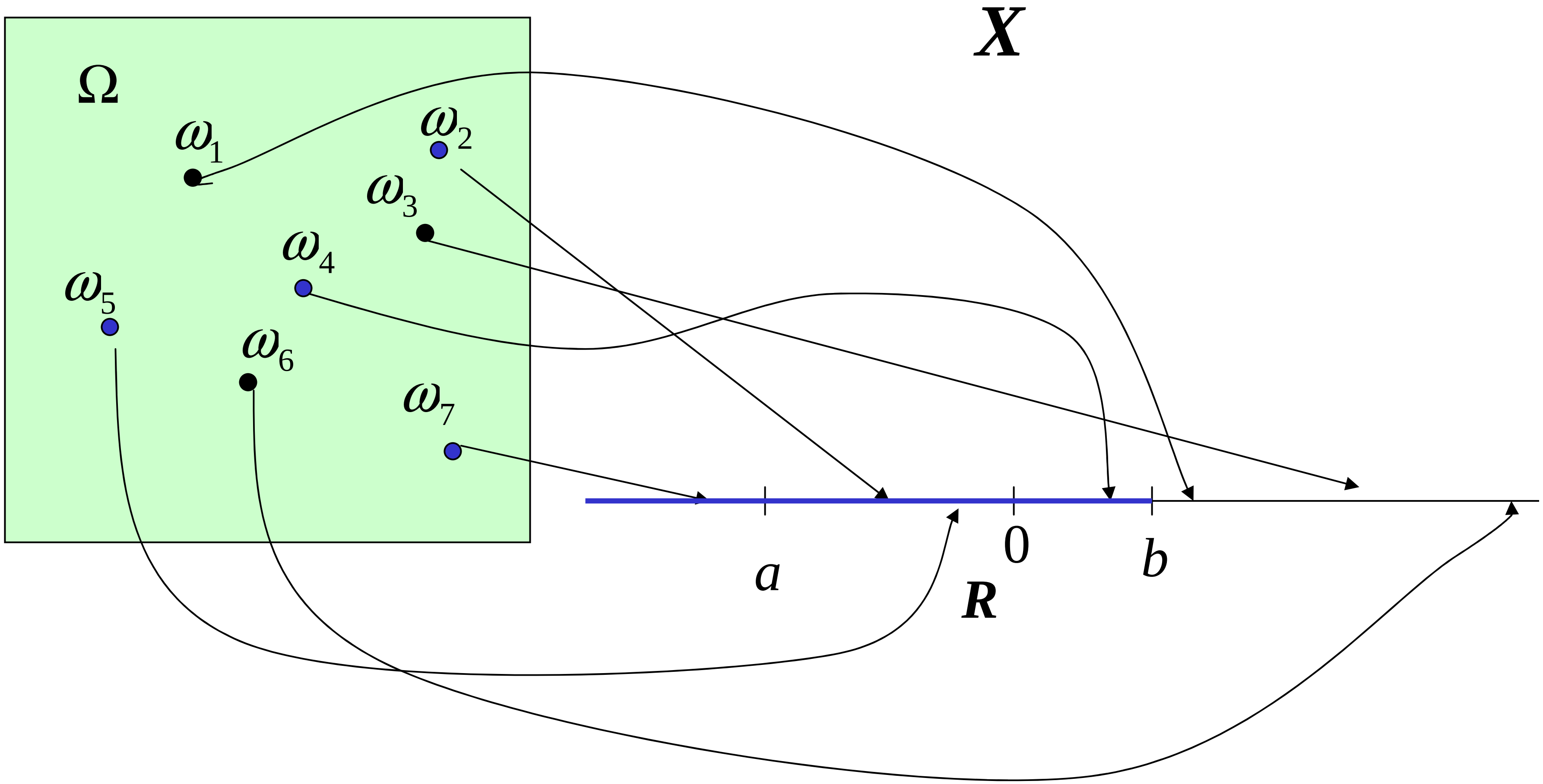
For a $\omega \in \Omega$, $X(\omega)$ is called an implementaion of the random variable X.



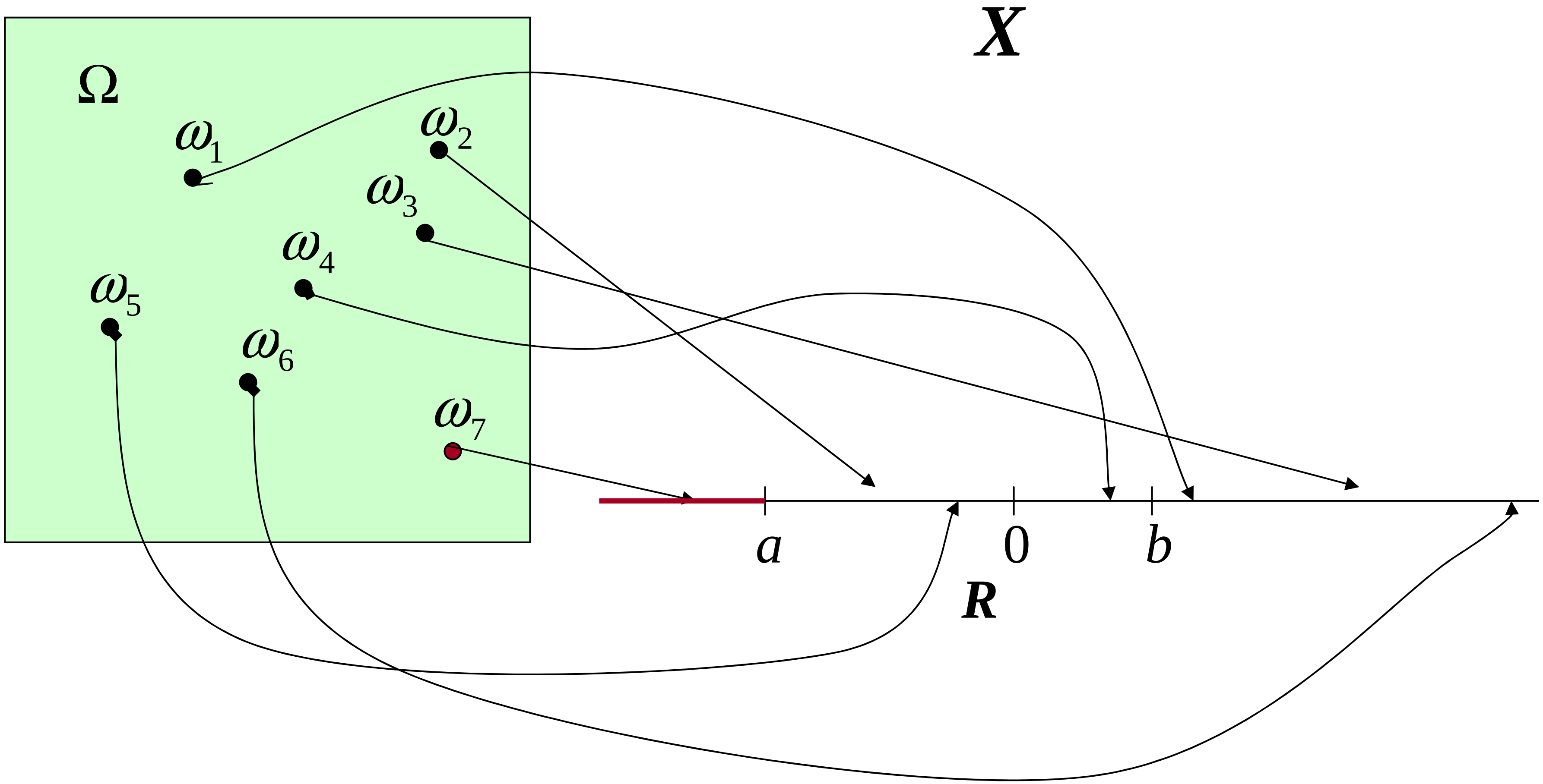
$$X : \Omega \rightarrow R$$



$$\{\ \omega_2, \omega_4, \omega_5\ \} = \{\ \omega \in \Omega \mid X(\omega) \in [a,b]\ \} = \{\ X \in [a,b]\ \}$$



$$\{\omega_2, \omega_4, \omega_5, \omega_7\} = \{\omega \in \Omega \mid X(\omega) < b\} = \{X < b\}$$



$$\{\,\omega_7\,\} = \{\,\omega \in \Omega \mid X(\omega) < a\,\} = \{\,X < a\,\}$$

$$\{X \in [a,b)\} = \{X < b\} - \{X < a\}$$

Random variables are very often used to define subsets of a sample space. We need these subsets to be events in a predefined σ-field of events so we we want a random variable to satisfy an additional condition:

Given a probabilistic space $\mathsf{P} = (\Omega, \mathsf{S}, P)$

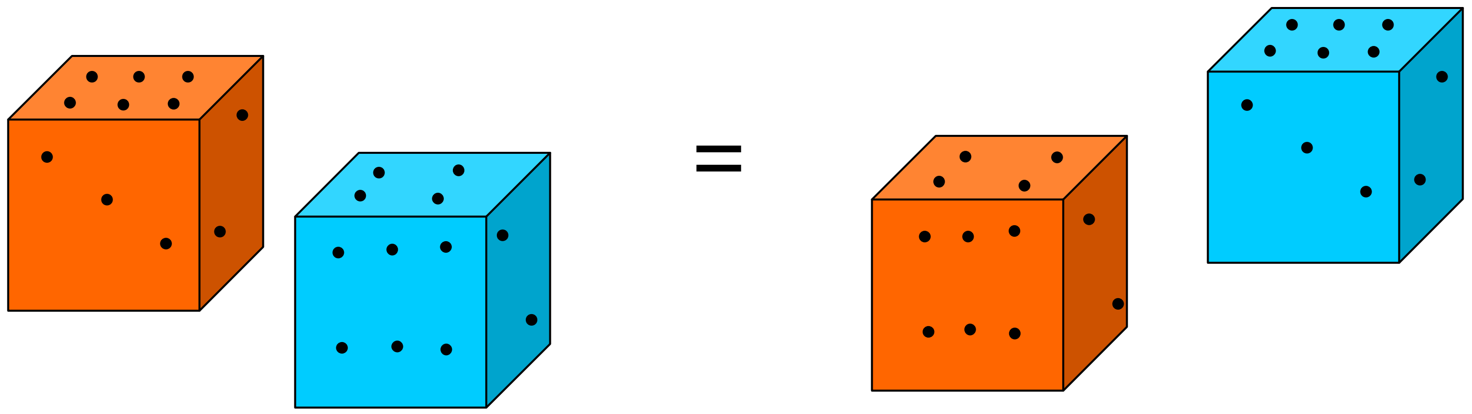
a mapping $X : \Omega \to R$ is called a random variable if

$\{X < a\} \in \mathsf{S}$ for every $a \in R$

Not every mapping $X : \Omega \rightarrow R$ is a random variable as illustrated by the following example

Experiment

We are rolling two dice, one red and one blue. In determining what is and what is not an event, we don't differentiate between the colurs, that is, if for example the outcome Red=6, Blue=4, is favourable to an event *E*, then so is the outcome Red=4, Blue=6. We take every subset satisfying this condition to be an event. It can be proved that, in such a way, we have de fined a σ-field S .



Let us define a mapping *X* assigning to each outcome a number in this way:

$$X(R=m, B=n) = 2m+n$$

In other words, the red die counts twice as much.

Now it is easy to see that

$$E = \{X < 17\} = \Omega - \{(\text{Red} = 6, \text{Blue} = 5), (\text{Red} = 6, \text{Blue} = 6)\}$$

However, X ({Red=5, Blue=6}) = 16 and so {Red=5, Blue=6}$\in E$, which means that the subset E, not being in S, is not an event and X is not a random variable.

Distribution function

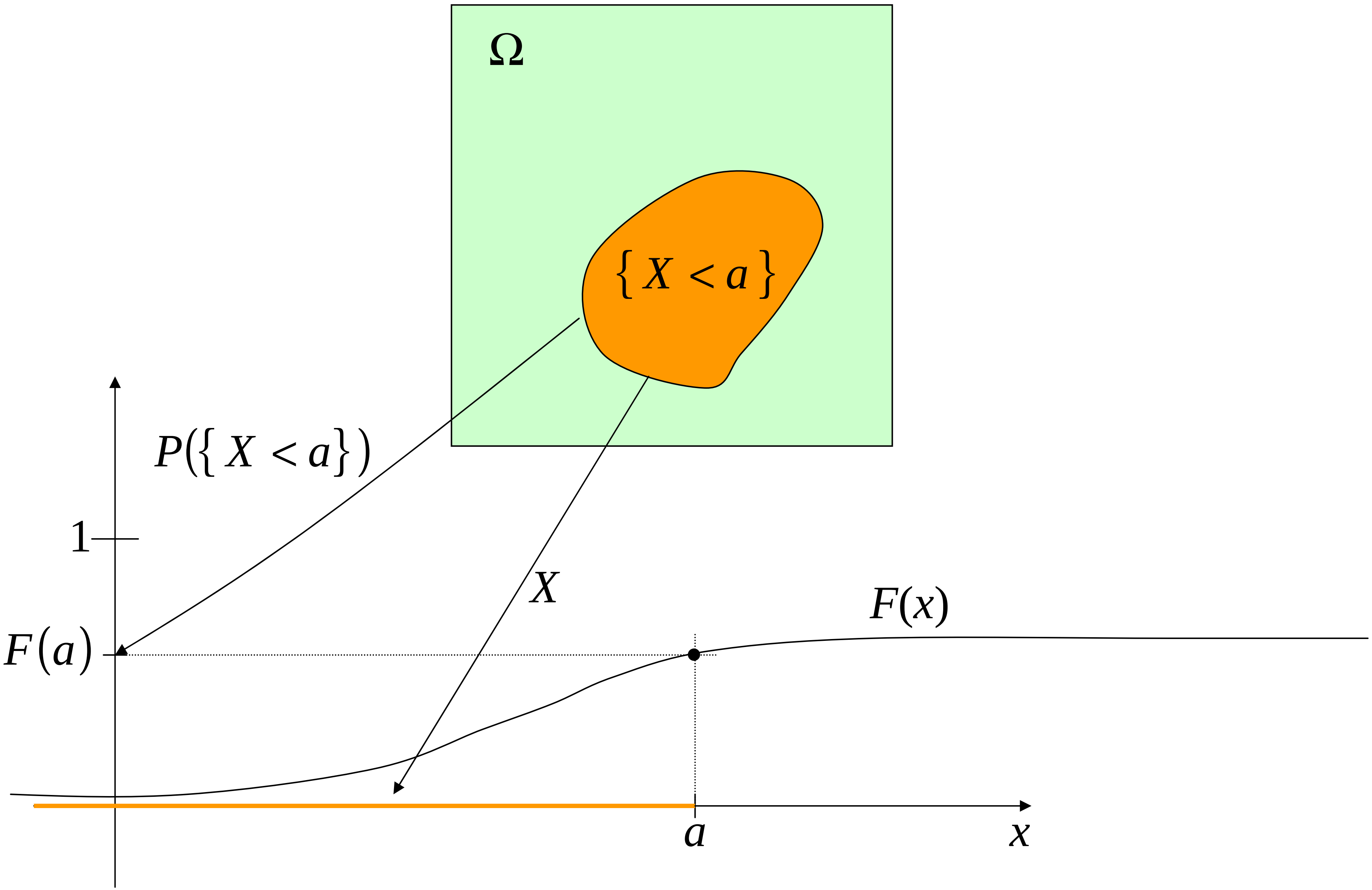
Let a probabilistic space $\mathsf{P} = (\Omega, \mathsf{S}, P)$ and a random variable X be given. Define a real function $F(x)$ of a real variable as follows:

$$F(x) = P(\{\omega \in \Omega \mid X(\omega) < x\}) = P(X < x)$$

$F(x)$ is called the distribution function of the random variable X. In other words, $F(x)$ is the probability of the outcome of an experiment to be in the interval $(-\infty, x)$.

Let $(\Omega, \mathcal{S}, P)$ be a probabilistic space, X a random variable and $F(x)$ its distribution function It follows from the definition of $F(x)$ and from the properties of P that we can write

$$P(\{\omega \in \Omega \mid a \leq X(\omega) < b\}) = P(a \leq X < b) = F(b) - F(a)$$



Ω
{ X < a }
P({ X < a})
1
X
F(x)
F(a)
a
x

Properties of a distribution function

Every distribution function $F(x)$ has the following properties:

1) $F : R \to [0,1]$

2) $F(x)$ is non-decreasing

3) $F(x)$ is continuous on the left

4) $\lim_{x \to -\infty} F(x) = 0$

5) $\lim_{x \to \infty} F(x) = 1$

The distribution function $F(x)$ of a random variable X is a very useful tool since it can be used to determine the probability of the outcome of an experiment falling into an arbitrary interval $[a,b)$, which in pratice is all we need to determine the probability of every event of practical importance.

In fact, in most cases the probability P connected with a random variable X is actually defined through its distribution function $F(x)$.

For practical purposes, we often define two special types of a random variable:

- discrete random variable

- continuous random variable

Discrete random variable

The range $R(X)$ of a discrete random variable X is a countable subset of discrete real numbers, that is, $R(X)$ does not contain any interval.

Discrete random variables

POPULATION

Persons chosen at random

~~Height: 115 cm~~

~~Weight: 17 kg~~

No of children: 0

Employed: No

~~Height: 195 cm~~

~~Weight: 98 kg~~

No of children: 4

Employed: Yes

~~Height: 170 cm~~

~~Weight: 80 kg~~

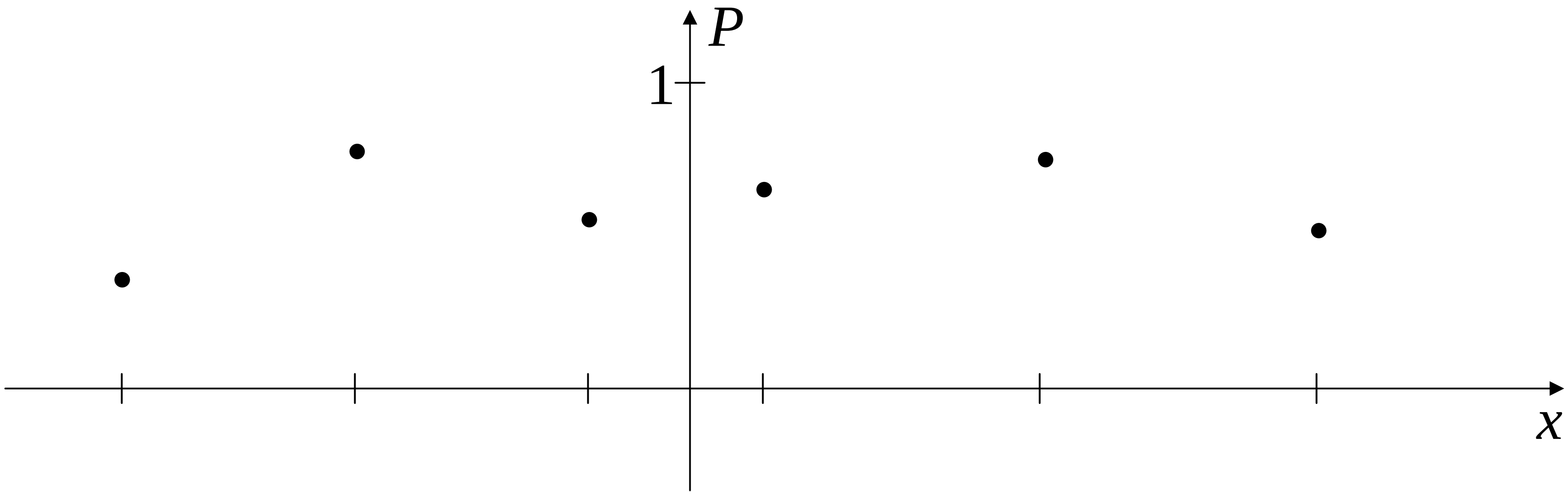
No of children: 2

Employed: No

For a discrete random variable X, we define the ***probability function*** $p(x)$. For every $a \in R$ we put

$$p(a) = P(\{\omega \in \Omega \mid X(\omega) = a\}) = P(X = a)$$

The probability function $p(x)$ of a random variable X is non-zero only at points belonging to $R(X)$.

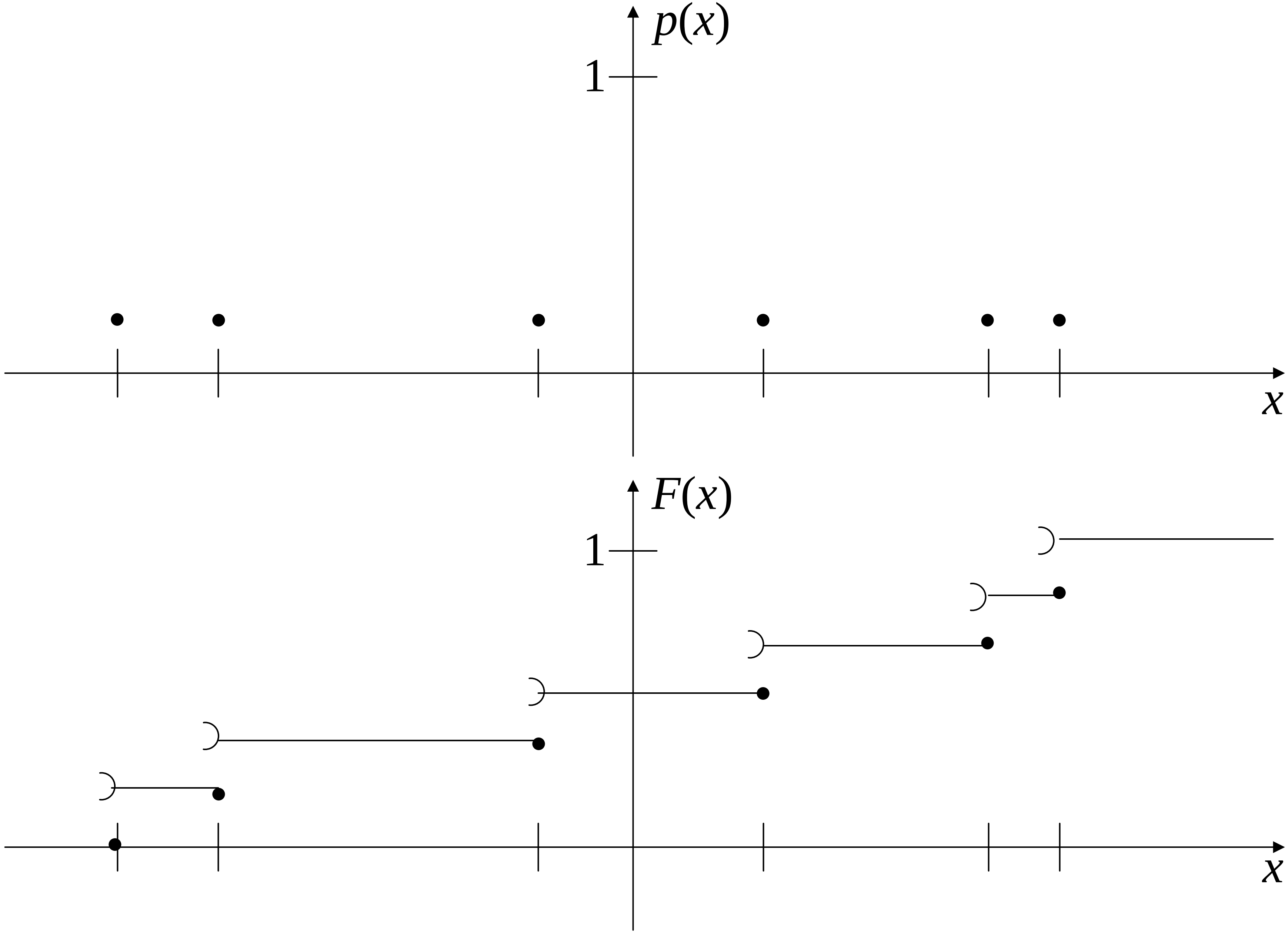


Let $(\Omega, \mathcal{S}, P)$ be a probabilistic space, X a discrete random variable, $p(x)$ its probability function and $F(x)$ its distribution function. Denoting by $R(X)$ the range of X, we can write for any $a \le b$:

$$\sum_{x \in R(X)} p(x) = 1$$

$$\sum_{x \in R(X) \wedge x < a} p(x) = F(a)$$

$$\sum_{x \in R(X) \wedge x \ge a \wedge x < b} p(x) = F(b) - F(a) = P(a \le X < b)$$



p(x)
1
x
F(x)
1
x

Continuous random variable

The range $R(X)$ of a continuous random variable X is an interval or the union of several intervals.

Discrete random variables

POPULATION

Persons chosen at random

Height: 115 cm

Weight: 17 kg

~~No of children: 0~~

~~Employed: No~~

Height: 195 cm

Weight: 98 kg

~~No of children: 4~~

~~Employed: Yes~~

Height: 170 cm

Weight: 80 kg

~~No of children: 2~~

~~Employed: No~~

Let X be a continuous random variable with $F(x)$ as a distribution function. If a function $f(x)$ exists such that

$$F(x) = \int_{-\infty}^{x} f(t)\,dt$$

we say that $f(x)$ is the density of random variable X.

Let $\left(\Omega, S, P\right)$ be a probabilistic space, X a continuous random variable, $f\,(x)$ its density function and $F\,(x)$ its distribution function. Then we can write for any $a \leq b$:

$$\int_{-\infty}^{\infty} f\left(x\right) dx = 1$$

$$\int_{a}^{b} f\left(x\right) dx = F\left(b\right) - F\left(a\right) = P\left(a \leq X < b\right)$$