6. Probability distributions

6.1. Random Variables

Example. Consider tossing four coins. The possible outcomes are then

$$S = \{\text{HHHH}, \text{HHHT}, \ldots, \text{THHH}, \\ \text{HHTT}, \text{HTHT}, \ldots, \text{TTHH}, \\ \text{HTTT}, \text{THTT}, \ldots, \text{TTTH}, \\ \text{TTTT}\}$$

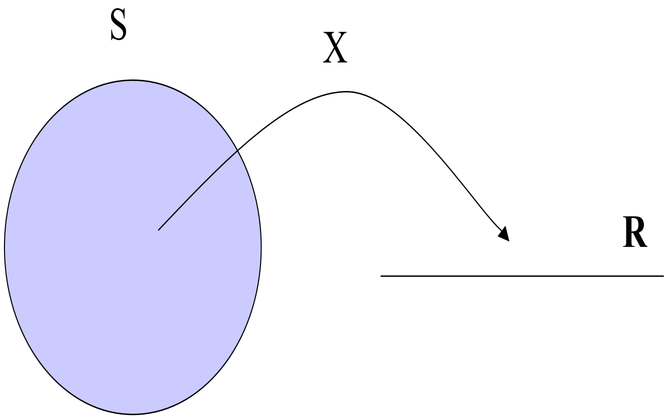
Suppose we are interested in the number X of heads up after 4 coin tosses, such that:

$X(\text{HHHH}) = 4$,
$X(\text{HHHT}) = \ldots = X(\text{THHH}) = 3$,
... and so on.

The value of X is determined by the random coin tossing. A variable whose value is determined by a random experiment is called a random variable (satunnaismuuttuja). It will become our mathematical tool for modelling statistical variables.

Mathematically a random variable is defined as a function from the sample space S to the line of real numbers $\mathbb{R}$, i.e.,

$$X : S \to \mathbb{R}$$



It is usual practice to distinguish the random variable from its value. Capital letters (like X above) for the random variables and lower case letters (like x) for the values are usual. Notations like $\tilde{X}$, $\tilde{Y}, \ldots$ or $\underline{x}$, $\underline{y}, \ldots$ are also used for random variables.

Random variables can often obtain values only in a subset of $\mathbb{R}$. The set of values that a random variable may possibly attain is called range space (arvojoukko) and often denoted by $S_X = X(S)$ or Ω_X. For example, $S_X = \{0, 1, 2, 3, 4\}$ in the coin tossing example above.

A discrete (diskreeti) random variable can assume only a countable number of values.

Example. Typical examples of discrete random variables are the number of children in a family, the result of tossing a die, the number of heads in the previous coin-toss example,etc. Also a random variable with range $\mathbf{Z} = \{\ldots -2, -1, 0, 1, 2, \ldots\}$ (the whole numbers) is discrete, and the values a discrete random variable can attain need not necessarily be equidistant.

A continuous (jatkuva) random variable can take any value in an interval of values, such as any interval on the real line $\mathbb{R}$ with finite length (the value set is uncountable).

The probability distribution (todennäköisyys-jakauma) of a discrete random variable consist of its attainable values and the corresponding probabilities (pistetodennäköisyys).

Example. In the previous example of 4 coin tosses, there were $2^4 = 16$ possible outcomes. We obtain the probability distribution of X by counting the number of outcomes with $X = 0, 1, 2, 3, 4$ heads and dividing by the possible number of outcomes:

x_i	0	1	2	3	4
$P(X = x_i)$	$\frac{1}{16}$	$\frac{1}{4}$	$\frac{3}{8}$	$\frac{1}{4}$	$\frac{1}{16}$

Note.
$\sum_{x_i \in S_X} P(X = x_i) = 1$, where $S_X = X(S)$, the set of possible values of the random variable (arvojoukko).

Mathematically, the probability distribution of a discrete random variable X is defined as a function $P : \Omega \to I\!R$ satisfying:

$$(1) \quad P(x_i) \geq 0, \text{ for all } x_i,$$

$$(2) \quad \sum_{x_i \in S_X} P(X = x_i) =: \sum_{i=1}^{k} p_i = 1.$$

The cumulative distribution function (kertymäfunktio) F of a discrete random variable is:

$$F(x) := P(X \leq x) = \sum_{x_i \leq x} P(X = x_i).$$

It has the important properties:

1. $F(x) \leq F(y)$ if $x < y$ with $F(-\infty) = 0$ and $F(\infty) = 1$.
2. $P(a < X \leq b) = P(X \leq b) - P(X \leq a) = F(b) - F(a)$ for $a \leq b$.
3. $P(X > x) = 1 - P(X \leq x) = 1 - F(x)$.

Property 2 implies that for discrete random variables:

$$P(X = x_i) = P(x_{i-1} < X \leq x_i) = F(x_i) - F(x_{i-1}).$$

Example: (Azcel, Example 3.2) Let

x:	0	1	2	3	4	5
$P(X = x)$:	0.1	0.2	0.3	0.2	0.1	0.1
Then $F(x)$:	0.1	0.3	0.6	0.8	0.9	1

$P(X \leq 3) = F(3) = P(0) + P(1) + P(2) + P(3) = 0.8$

$P(X \geq 2) = P(X > 1) = 1 - F(1) = 1 - 0.3 = 0.7$

$P(1 \leq X \leq 3) = P(0 < X \leq 3) = F(3) - F(0) = 0.7.$

Recall that a continuous random variable has uncountably many attainable values. Therefore we cannot define the probability distribution of a continuous random variable as a list of its attainable values and their associated probabilities, as we did for discrete random variables. Instead we define a so called probability density function (tiheysfunktio) f as a scaled histogram of infinitely many observations of the random variable in the limit of infinitesimal narrow class intervals. The histogram is scaled in such a way that the frequency density on the vertical axis is divided by the total number of observations. This changes the area under the histogram but not its shape. The probability density of a continuous random variable has the following properties:

(1) $f(x) \geq 0$ for all x

(2) $\int_{-\infty}^{\infty} f(x)\, dx = 1$,
(the total area under f is 1)

(3) $P(a < X \leq b) = \int_a^b f(x)\, dx$,
(the area under f between a and b).

The cumulative distribution function (kertymäfunktio) of a continuous random variable is then defined in anology to the discrete case:

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

This implies that for continuous random variables the probability $P(X \leq x)$ may be found from the area below the density function f (and above the x-axis) between $-\infty$ and x. The cumulative distribution function of a continuous random variable has the same properties (1–3) as that of a discrete random variable. Additionally for $a \leq b$:

$$\begin{aligned} & P(a < X < b) = P(a < X \leq b) \\ =& P(a \leq X < b) = P(a \leq X \leq b) \\ =& F(b) - F(a). \end{aligned}$$

This is because for continuous random variables

$$P(X = x) = \int_{x}^{x} f(x)\,dx = 0.$$

Note: The definition of the cumulative distribution function implies by the fundamental theorem of calculus that for continuous random variables: $F'(x) = f(x)$.

Example.

Let

$$f(x) = \begin{cases} \frac{1}{5} & \text{for } x \in [0,5], \\ 0 & \text{otherwise.} \end{cases}$$

This is an example of the (continuous) uniform distribution to be discussed soon.

Now, for $0 \le x \le 5$:

$$F(x) = \overbrace{\int_{-\infty}^{x} f(t)\,dt}^{P(X \le x)} = \overbrace{\int_{-\infty}^{0} f(t)\,dt}^{P(X \le 0)} + \overbrace{\int_{0}^{x} f(t)\,dt}^{P(0 < X \le x)}$$

$$= 0 + \int_0^x \frac{1}{5}\,dt = \left.\frac{t}{5}\right|_0^x = \frac{x}{5} - \frac{0}{5} = \frac{x}{5}.$$

Therefore:

$$F(x) = \begin{cases} 0, & x < 0 \\ \frac{x}{5}, & 0 \le x \le 5 \\ 1 & x > 5, \end{cases}$$

and, for example:

$$P(1 < X \le 3) = F(3) - F(1) = \frac{3}{5} - \frac{1}{5} = \frac{2}{5}.$$

Statistics of Random Variables

The <u>expected value</u> (odotusarvo) of a discrete random variable is defined as

$$\mu = \mathsf{E}\,[X] = \sum_{x \in S_X} x P(X = x).$$

If X is continuous then

$$\mu = \mathsf{E}\,[X] = \int_{-\infty}^{\infty} x f(x)\, dx.$$

Generally, if h is a function, and $Y = h(X)$,

$$\mathsf{E}\,[Y] = \mathsf{E}\,[h(X)] = \begin{cases} \sum_x h(x) P(X = x), & X \text{ discrete} \\ \int_{-\infty}^{\infty} h(x) f(x)\, dx, & X \text{ continuous} \end{cases}$$

The <u>median</u> (mediaani) is the smallest number M satisfying: $F(M) \geq 1/2$.

Draw independent observations at random from a population with finite mean μ. Then the <u>law of large numbers</u> (suurten lukujen laki) asserts that as the sample size increases, the mean of the sample $\bar{x}$ gets eventually closer and closer to the population mean μ.

The <u>variance</u> (varianssi) of a random variable is defined as

$$\mathsf{Var}\,[X] = \mathsf{E}\,(X-\mu)^2.$$

That is

$$\sigma^2 = \mathsf{Var}\,[X] = \begin{cases} \sum_x (x-\mu)^2 P(X=x), & X \text{ discrete,} \\ \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\,dx, & X \text{ continuous.} \end{cases}$$

The <u>standard deviation</u> (keskihajonta) is the positive square root of the variance

$$\sigma = \sqrt{\sigma^2}.$$

Example.

The probability distribution of casting a dice is:

x_i	1	2	3	4	5	6
$P(X = x_i)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

$$
\begin{aligned}
E(X) =& 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + 3 \cdot \frac{1}{6} + 4 \cdot \frac{1}{6} \\
& + 5 \cdot \frac{1}{6} + 6 \cdot \frac{1}{6} = \frac{21}{6} = 3.5. \\
V(X) =& (1 - 3.5)^2 \frac{1}{6} + (2 - 3.5)^2 \frac{1}{6} \\
& + (3 - 3.5)^2 \frac{1}{6} + (4 - 3.5)^2 \frac{1}{6} \\
& + (5 - 3.5)^2 \frac{1}{6} + (6 - 3.5)^2 \frac{1}{6} \approx 2.9167. \\
\sigma =& \sqrt{V(X)} \approx 1.7. \\
M =& 3, \text{ because } F(3) = 1/2.
\end{aligned}
$$

Note. The expected value $E(X)$ need not necessarily be a value which the random variable X can actually assume, but the median is always the smallest attainable value of X satisfying $F(X) \geq 1/2$.

Example.

Consider the uniform distribution:

$$f(x) = \begin{cases} \frac{1}{4} & \text{for } x \in [0,4], \\ 0 & \text{otherwise.} \end{cases}$$

$$\begin{aligned} E(X) &= \int_{-\infty}^{\infty} x f(x)\, dx = \int_0^4 x \cdot \frac{1}{4}\, dx = \frac{1}{4}\left[\frac{1}{2}x^2\right]_0^4 \\ &= \frac{1}{4}\left(\frac{1}{2}\cdot 4^2 - \frac{1}{2}\cdot 0^2\right) = 2. \end{aligned}$$

$$\begin{aligned} V(X) &= \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\, dx = \int_0^4 (x-2)^2 \cdot \frac{1}{4}\, dx \\ &= \frac{1}{4}\int_0^4 \left(x^2 - 4x + 4\right)\, dx = \frac{1}{4}\left[\frac{1}{3}x^3 - \frac{4}{2}x^2 + 4x\right]_0^4 \\ &= \frac{1}{4}\left(\frac{1}{3}4^3 - \frac{4}{2}4^2 + 4\cdot 4 - 0\right) = \frac{4}{3}. \end{aligned}$$

The distributions median is $M = 2$, because

$$F(2) = \int_{-\infty}^{2} f(t)\, dt = \int_0^2 \frac{1}{4}\, dt = \left[\frac{t}{4}\right]_0^2 = \frac{1}{2}.$$

Properties of Expected Value and Variance

1. $E(c) = c$ for c constant,

2. $E(cX) = cE(X)$ for c constant,

3. $E(X + Y) = E(X) + E(Y)$,

4. $E(X \cdot Y) = E(X) \cdot E(Y)$
   for X and Y independent,

5. $V(c) = 0$ for c constant,

6. $V(X) = E(X^2) - E(X)^2$,

7. $V(aX + b) = a^2V(X)$ for a, b constants,

8. $V(X + Y) = V(X) + V(Y)$
   for X and Y independent.