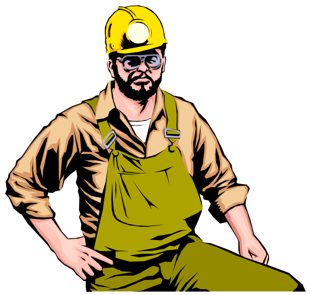
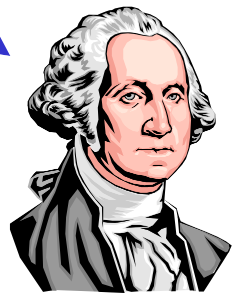
Random vector

POPULATION

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

Persons chosen at random

The sequence of random variables

$$X_1 = Height$$

$$X_2 = Weight$$

$$X_3 = Number\ of\ children$$

$$X_4 = Employed$$

is an example of a random vector.

Generally, a random vector $\boldsymbol{X} = \left(X_1, X_2, \ldots, X_n\right)$ assigns a vector $\left(x_1, x_2, \ldots, x_n\right)$ of real numbers to each outcome $\omega \in \Omega$

$$\Omega \xrightarrow{\boldsymbol{X}} \underbrace{R \times R \times \cdots \times R}_{n \text{ times}}$$

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

$$\boldsymbol{X} = (X_1, X_2, \ldots, X_n) : \Omega \to R^n$$

is called a random vector if

$$\{X_1 < a\} \cap \{X_2 < a\} \cap \ldots \cap \{X_n < a\} \in \mathcal{S}$$

for every $a \in R$

Probability distribution of a random vector

Let us consider a probabilistic space $\mathsf{P} = (\Omega, \mathsf{S}, P)$

for a random vector $\boldsymbol{X} = (X_1, X_2, \dots, X_n)$ we define its

probability distribution $F(x_1, x_2, \dots, x_n)$ as follows

$$F(x_1, x_2, \dots, x_n) = P(X_1 < x_1 \wedge X_1 < x_1 \wedge \dots \wedge X_n < x_n)$$

Properties of the distribution of a random vector

The probability distribution $F(x_1, x_2, \ldots, x_n)$ of a random vector $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ has the following properties

- $F(x_1, x_2, \ldots, x_n)$ is increasing and continuous on the left in each of its independent variables

- $\lim\limits_{x_i \to -\infty} F(x_1, x_2, \ldots, x_n) = 0,\ i = 1,2,\ldots,n$

- $\lim\limits_{\substack{x_1 \to \infty \\ x_2 \to \infty \\ \vdots \\ x_n \to \infty}} F(x_1, x_2, \ldots, x_n) = 1$

Discrete random vectors

A random vector $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ is called discrete if its

range is a finite or a countable set of real vectors

$$\left\{ \left(x_1^1, x_2^1, \ldots x_n^1\right), \left(x_1^2, x_2^2, \ldots x_n^2\right), \ldots \right\}$$

For a discrete random vector, we can define the probability function $p(x_1, x_2, \ldots, x_n)$

$$p(x_1, x_2, \ldots, x_n) = P(X_1 = x_1 \wedge X_2 = x_2 \wedge \ldots \wedge X_n = x_n)$$

The relationship between a probability distribution and probability function

$$F(x_1, x_2, \ldots x_n) = \sum_{t_1 < x_1 \wedge \ldots \wedge t_n < x_n} p(t_1, t_2, \ldots t_n)$$

Continuous random vectors

A random vector $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ is called continuous if

its range includes a Cartesian product of n intervals

$$[a_1, b_1] \times [a_2, b_2] \times \ldots \times [a_n, b_n] \subseteq R^n$$

If a function $f(x_1, x_2, \ldots, x_n)$ exists such that

$$F(x_1, x_2, \ldots, x_n) = \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_n} f(x_1, x_2, \ldots, x_n)\, dx_1 \ldots dx_n$$

we say that $f(x_1, x_2, \ldots, x_n)$ is the probability density of

the random vector $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$

Marginal distributions

For a random vector $\boldsymbol{X} = \left(X_1, X_2, \ldots, X_n\right)$ with a distribution

$F\left(x_1, x_2, \ldots, x_n\right)$ we define marginal distributions

$F_1\left(x_1\right), F_2\left(x_2\right), \ldots, F_n\left(x_n\right)$

$$F_i\left(x_i\right) = \lim_{\substack{x_1 \to \infty \\ \vdots \\ x_{i-1} \to \infty \\ x_{i+1} \to \infty \\ \vdots \\ x_n \to \infty}} F\left(x_1, x_2, \ldots, x_n\right)$$

$F_i\left(x_i\right)$ is a limit of $F\left(x_1, x_2, \ldots, x_n\right)$ with all variables except x_i tending to infinity

If a random vector $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ is discrete, we define

its marginal probability functions $p_1(x_1), p_2(x_2), \ldots, p_n(x_n)$

$$p_i(x_i) = \sum_{x_1,\ldots,x_{i-1},x_{i+1},\ldots,x_n} p(x_1, x_2, \ldots, x_n)$$

where the summation is done over all the values of the variables

$x_1, \ldots, x_{i-1}, x_{i+1}, \ldots, x_n$, If a random vector $\boldsymbol{X} = \left(X_1, X_2, \ldots, X_n\right)$ is continuous, we define

its marginal probability densities $f_1\left(x_1\right), f_2\left(x_2\right), \ldots, f_n\left(x_n\right)$

$$f_i\left(x_i\right) = \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f\left(x_1, x_2, \ldots, x_n\right) dx_1 \ldots dx_{i-1} dx_{i+1} \ldots dx_n$$

By considering a marginal distribution $F_i\left(x_i\right)$ of a random vector $\boldsymbol{X} = \left(X_1, X_2, \ldots, X_n\right)$ we actually define a single random variable X_i by "neglecting" all other variables.

Example for $n = 2$

Let (X,Y) be a discrete random variable with X taking on values from the set {1,2,3,4}, Y from the set {-1,1,3,5,7} and the probability function given by the below table. Calculate the marginal probability functions of the random variables X and Y.

Y \ X	**1**	**2**	**3**	**4**
-1	0,008	0,02	0,24	0,13
1	0,02	0,001	0,021	0,114
3	0,11	0,022	0,014	0,002
5	0,013	0,115	0,003	0,01
7	0,003	0,01	0,05	0,094

X / Y	**1**	**2**	**3**	**4**	**$p_2(y)$**
-1	0,008	0,02	0,24	0,13	**0,398**
1	0,02	0,001	0,021	0,114	**0,156**
3	0,11	0,022	0,014	0,002	**0,148**
5	0,013	0,115	0,003	0,01	**0,141**
7	0,003	0,01	0,05	0,094	**0,157**
$p_1(x)$	**0,154**	**0,168**	**0,328**	**0,35**	**1**

Let $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ be a random vector with a distribution

$F(x_1, x_2, \ldots, x_n)$ and marginal distributions

$$F_1(x_1), F_2(x_2), \ldots, F_n(x_n)$$

If $F(x_1, x_2, \ldots, x_n) = F_1(x_1) F_2(x_2) \ldots F_n(x_n)$ we say that

$X_1, X_2, \ldots, X_n$ are <u>independent random variables</u>.

If $X_1, X_2, \ldots, X_n$ are independent and have a probability

function p or density f, it can be proved that also

$p(x_1, x_2, \ldots, x_n) = p_1(x_1) p_2(x_2) \ldots p_n(x_n)$ or

$f(x_1, x_2, \ldots, x_n) = f_1(x_1) f_2(x_2) \ldots f_n(x_n)$

Correlation coefficient of two random variables

Let us consider a random vector (X,Y) with a distribution $F(x,y)$. Using the marginal distributions $F_x(x,y)$ and $F_y(x,y)$ we can define the expectancies $E(X)$ and $E(Y)$ and variances $D(X)$ and $D(Y)$. We define the covariance of the random vector (X,Y)

$$\boxed{\operatorname{cov}(X,Y) = E(XY) - E(X)E(Y)}$$

with

$$E(XY) = \sum_{x_i, x_j} x_i y_j \, p(x_i, y_j) \quad \text{or} \quad E(XY) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} xy \, f(x,y)\,dx\,dy$$

depending on whether (X,Y) is discrete or continuous. Then $p(x,y)$ or $f(x,y)$ is the probability function or density.

The correlation coefficient is then defined as

$$\rho(X,Y)=\frac{\operatorname{cov}(X,Y)}{\sqrt{D(X)D(Y)}}$$

$\rho(X,Y)$ has the following properties

- $-1\le \rho(X,Y)\le 1$
- X and Y are independent $\Rightarrow \rho(X,Y)=0$
- $Y=aX+b, a>0 \Rightarrow \rho(X,Y)=1$
- $Y=aX+b, a<0 \Rightarrow \rho(X,Y)=-1$

Example

Calculate the correlation coefficient of the discrete random vector (X,Y) with a probability function given by the below table

Y \ X	**1**	**2**	**3**	**4**
-1	0,008	0,02	0,24	0,13
1	0,02	0,001	0,021	0,114
3	0,11	0,022	0,014	0,002
5	0,013	0,115	0,003	0,01
7	0,003	0,01	0,05	0,094

X / Y	**1**	**2**	**3**	**4**	$p_2(y)$
-1	0,008	0,02	0,24	0,13	**0,398**
1	0,02	0,001	0,021	0,114	**0,156**
3	0,11	0,022	0,014	0,002	**0,148**
5	0,013	0,115	0,003	0,01	**0,141**
7	0,003	0,01	0,05	0,094	**0,157**
$p_1(x)$	**0,154**	**0,168**	**0,328**	**0,35**	**1**

E(X) =	2,874	**E(X^2) =**	9,378	**D(X) =**	1,118124
E(Y)=	2,006	**E(Y^2)=**	13,104	**D(Y)=**	9,079964
E(XY) =	5,168	**cov(X,Y) =**	-0,59724	**r =**	-0,18744

Note that, generally, it is not true that $\rho(X,Y)$ implies that X and Y are independent as proved by the following example

X \ Y	**1**	**2**	**3**	***p*** **$_2$(y)**
-1	0,1	0,15	0,05	**0,3**
0	0,2	0,1	0,1	**0,4**
1	0,15	0,05	0,1	**0,3**
p **$_1$(x)**	**0,45**	**0,3**	**0,25**	**1**

The correlation coefficient is zero as can be easily calculated, but we have, for example, $p_1(1)p_2(-1)=(0.45)(0.3)=0.135\neq 0.1$

so that X and Y cannot be independent.