3. RANDOM VARIABLES

Study time: 80 minutes

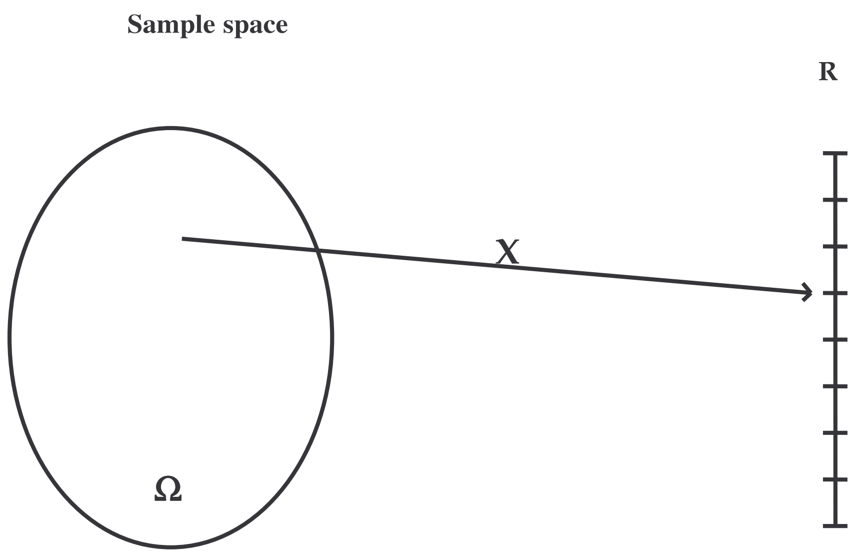
Aim - you will be able to

- describe the random variable by the distribution function
- characterize a discrete and a continuous random variable
- understand the hazard rate function
- determine the numerical characteristics of the random variable
- transform the random variable

Explication

3.1. Definition of a random variable

Let us consider a probability space (Ω, S, P). A **random variable X** (RV) on a sample space Ω is such real function X(ω) that for each real x$\in$ R is the set $\{\omega \in \Omega \mid X(\omega < x)\} \in S$, i.e. it is a random event. Therefore, the random variable is such function X: $\Omega \rightarrow$ R that for each x$\in$ R holds: $X^{-1}((-\infty, x)) = \{\omega \in \Omega \mid X(\omega < x)\} \in S$. From definition implies that we can determine a probability of $X(\omega) < x$ for any x$\in$ R.



A group of all values $\{x = X(\omega),\ \omega \in \Omega\}$ is called **sample space**.

3.2. Distribution function

Definition: The distribution function of a random variable X is written $F(t)$ and, for each $t \in R$ has the value:

$$F(t) = P\{X \in (-\infty, t)\} = P(X \le t).$$

Properties of the probability distribution function:

1. $0 \le F(x) \le 1$ for $-\infty < x < +\infty$

2. the distribution function is a monotonic increasing function of x, i.e. $\forall x_1, x_2 \in R: x_1 < x_2 \Rightarrow F(x_1) \le F(x_2)$

3. the distribution function F(x) is left-continuous

4. $\lim_{x \to +\infty} F(x) = 1;\ \lim_{x \to -\infty} F(x) = 0$
5. $\forall$ a, b $\in$ R; a < b : $P(a \le X < b) = F(b) - F(a)$

6. $P(x = x_0) = \lim_{x \to x_0+} F(x) - F(x_0)$

If the range of the random variable function is discrete, then the random variable is called a discrete random variable. Otherwise, if the range includes a complete interval on the real line, the random variable is continuous.

3.3. Discrete random variable

We speak about discrete random variable if a random variable is from some finite and enumerable set. The most often it is an integer random variable e.g. a number of student that entered the main building of VŠB TUO before midday (0,1,2,...), a number of house members (1,2,3,...), a number of car accidents during one day on a Prague - Brno highway (0,1,2,...), etc..

Definition
We say that a random **variable X has a discrete probability distribution** when:

1. $\exists$ finite or enumerable set of real numbers M={ $x_1, \dots, x_{n}, \dots$} that
   $P(X = x_i) > 0$ i = 1, 2, ...
2. $\sum_i P(X = x_i) = 1$

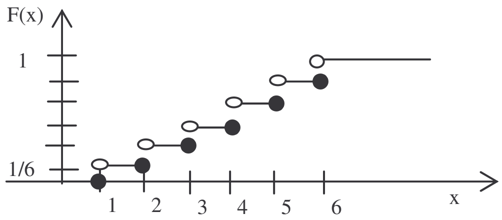
Function $P(X = x_i) \Leftrightarrow P(x_i)$ is called **probability function of random variable** X.
A distribution function of such distribution is a step function with steps in $x_1, \dots, x_{n}, \dots$
For a distribution function of discrete random variable it holds:

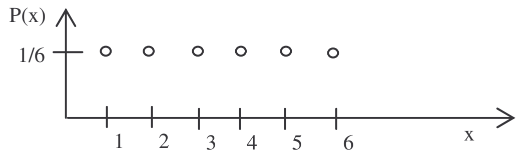
$$F(x) = \sum_{x_i < x} P(X = x_i)$$

Example

A throwing dice, X … a number of obtained dots

x_i	$P(X = x_i)$	$F(x_i)$
1	1/6	0
2	1/6	1/6
3	1/6	2/6
4	1/6	3/6
5	1/6	4/6
6	1/6	5/6





3.4. Continuous random variable

If a random variable may have any value from a certain interval we speak about a random variable with continuous distribution. As an example we can name a service life of a product (0, ∞), the length of specific object etc. In such case, we can use a density function as well as distribution function to describe a distribution of random variable.

Definition
Random variable has a **continuous probability distribution** when a function f(x) exists that

$$F(x) = \int_{-\infty}^{x} f(t)dt \qquad \text{pro } -\infty < x < \infty$$

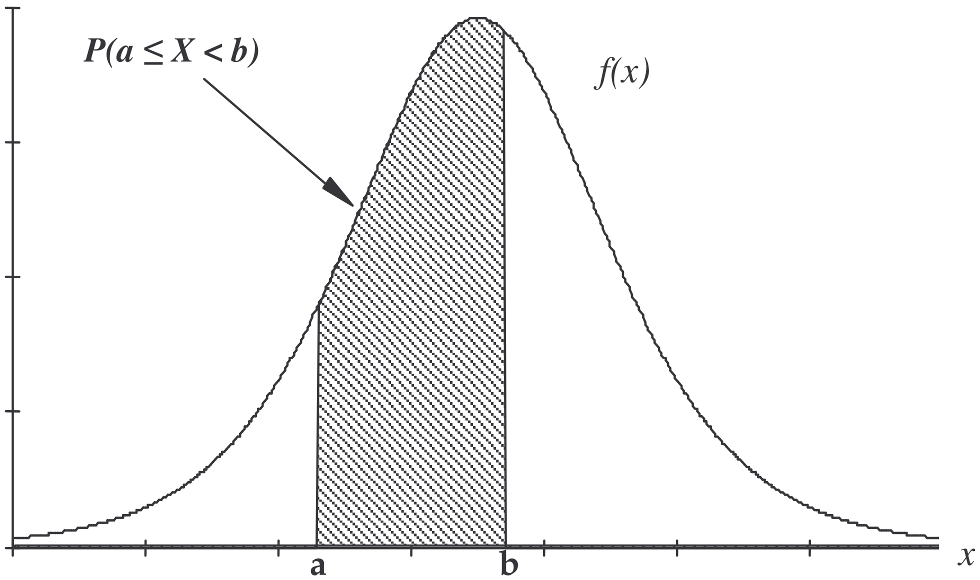
Function *f(x)* is called a **probability density function** of continuous random variable X. It is non-negative real function.
We can show that in all points where a derivation of distribution function exists it holds:

$$f(x) = \frac{dF(x)}{dx}$$

If we know a distribution function we can easily determine a probability density function vice versa.

The area below *f(x)* spline for $x \in \langle a; b)$; $(a, b \in R)$ in any interval is the probability that X will gain the value of this interval. It also fully corresponds with our density definition.

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



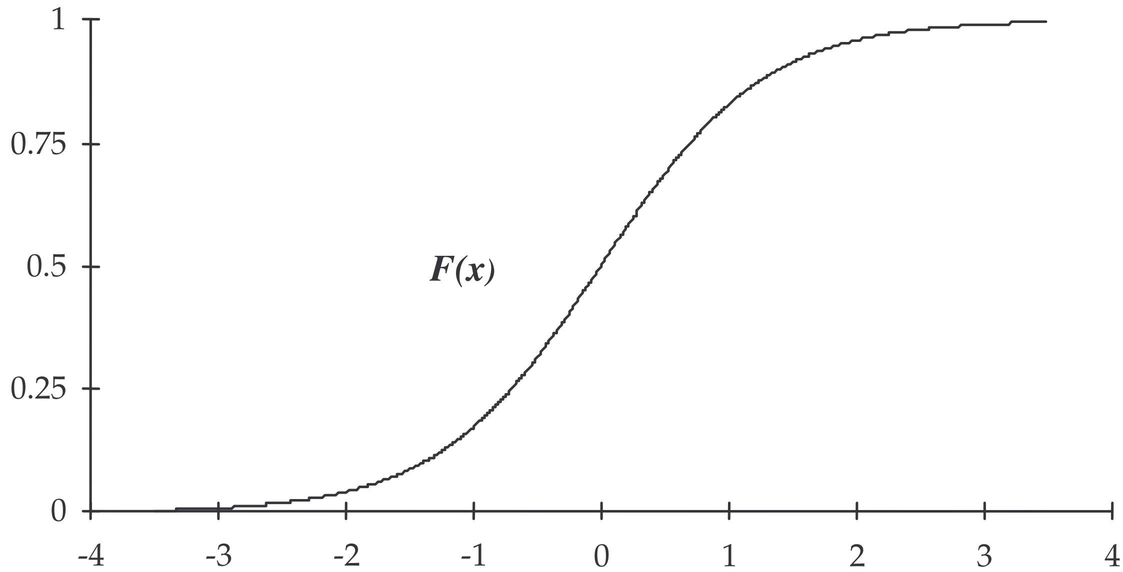
One of attributes for each density probability is the fact that the whole area under curve is equal to one. It is analogical to a discrete random variable where the sum of probabilities for all possible results is also equal to one. We can describe this attribute by the following equation:

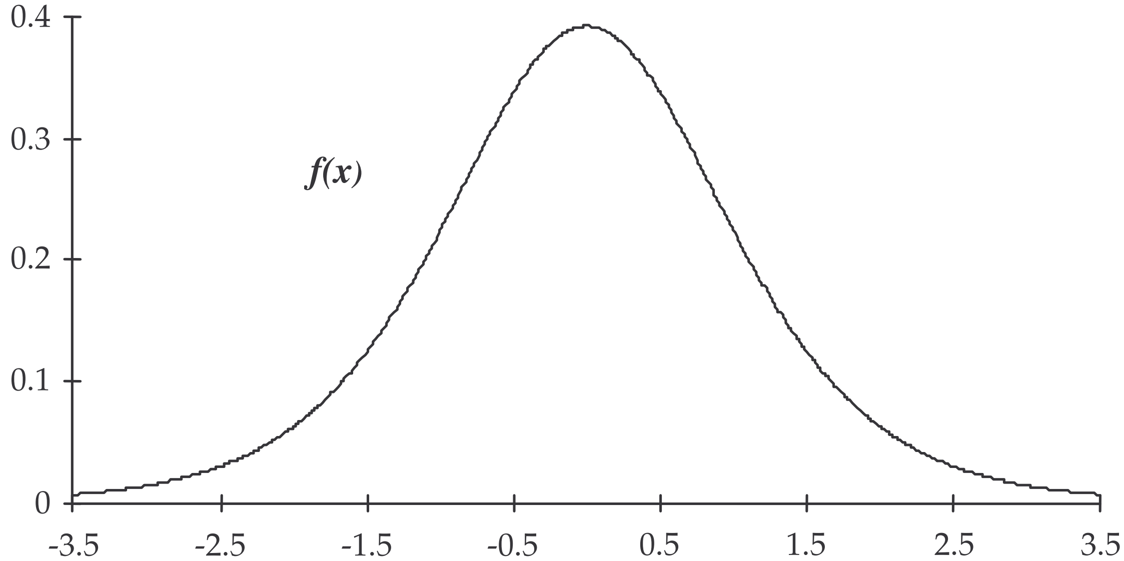
$$\int_{-\infty}^{\infty} f(x)\,dx = 1$$

Example

Logistic probability distribution has a following distribution function *F(x)* and a probability density *f(x):*

$$F(x) = \frac{1}{1+e^{-(\beta_0+\beta_1 x)}} \qquad f(x) = \frac{\beta_1\, e^{-(\beta_0+\beta_1 x)}}{(1 + e^{-(\beta_0+\beta_1 x)})^2}$$





3.5. Failure rate

Let X be a non-negative random variable with continuous distribution. Then, we define a **failure rate** for $F(t) < 1$

$$\lambda(t) = \frac{f(t)}{1 - F(t)}.$$

We can easily derive the following formula:

$$\lambda(t) = \lim_{\Delta t \to 0+} \frac{P(t < X \le t + \Delta t | X > t)}{\Delta t} = \frac{f(t)}{1 - F(t)}$$

Let X be a **mean time to failure** of any system. Then, the failure rate expresses that if in the t-time there was no failure the probability of failure in a small following time Δt is approximately $\lambda(t).\Delta t$:

$$P(t < X \le t + \Delta t | X > t) \approx \frac{f(t)}{1 - F(t)} \Delta t \; = \; \lambda(t).\Delta t$$

The failure rate characterizes the probability distribution of non-negative random variable.

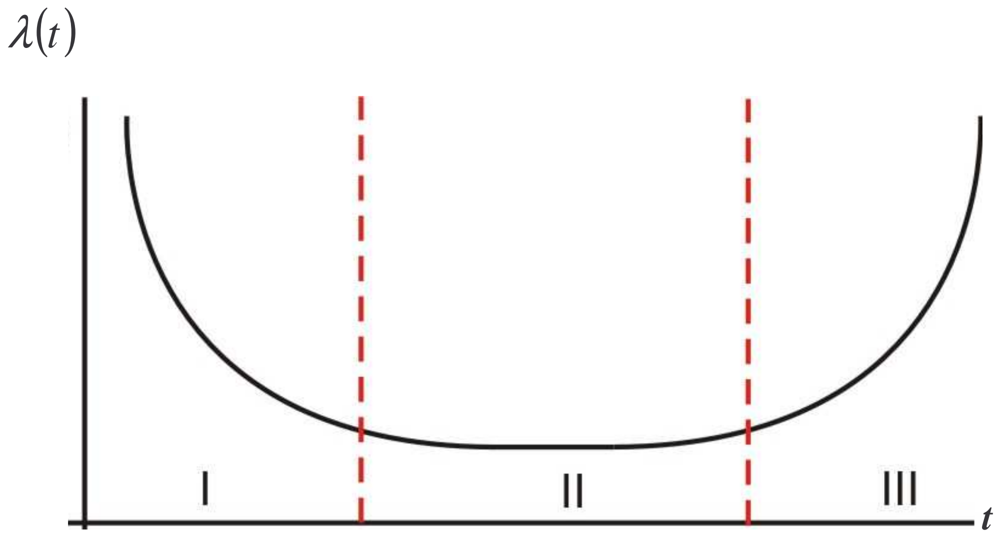
Table 1 shows the mutual conversions between $f(t), F(t), \lambda(t)$:

	$F(t)$	$f(t)$	$\lambda(t)$
$F(t)$	$F(t)$	$\int_0^t f(x)dx$	$1 - \exp\left[-\int_0^t \lambda(x)dx\right]$
$f(t)$	$\frac{dF(t)}{dt}$	$f(t)$	$\lambda(t) \cdot \exp\left[-\int_0^t \lambda(x)dx\right]$
$\lambda(t)$	$\frac{\frac{dF(t)}{dt}}{1 - F(t)}$	$\frac{f(t)}{1 - \int_0^t f(x)dx}$	$\lambda(t)$

Table 1

- **The most often graphical interpretation of failure rate**

Let a random variable X be a **mean time to failure** of any system. Then, a typical form of failure rate is shown in the following figure. The curve in this figure is called the **bathtub curve.**



I … The first part is a decreasing failure rate, known as early failures or **infant mortality.**

II … The second part is a constant failure rate, known as random failures.

III … The third part is an increasing failure rate, known as wear-out failures.

3.6. Numerical characteristics of random variable

The probability distribution of each random variable X is fully described by its distribution function F(x). In many cases we can summarize the total information to several numbers. These numbers are called the **numerical characteristics of the random variable X.**

1. **Moments**

r-th general moment is denoted $\mu_r' = EX^r$ r = 0,1,2, …

discrete RV: $\mu_r' = \sum_i x_i^r \cdot P(x_i)$

continuous RV: $\mu_r' = \int_{-\infty}^{\infty} x^r \cdot f(x)\, dx$ r = 0,1,2, …

if stated progression or integral tend absolutely.

r-th central moment is denoted $\mu_r = E[X - EX]^r$ r = 0,1,2, …

discrete RV: $\mu_r = \sum_i [x_i - EX]^r \cdot P(x_i)$

continuous RV: $\mu_r = \int_{-\infty}^{\infty} (x - EX)^r \cdot f(x)\, dx$

if stated progression or integral tend absolutely.

2. Expected value (mean) $EX = \mu_1'$

discrete RV: $EX = \sum_i x_i \, . \, P(X = x_i)$

continuous RV: $EX = \int_{-\infty}^{\infty} x . f(x)dx$

Properties:

1. $E(aX + b) = a. EX + b \quad a,b \in R$
2. $E(X_1 + X_2) = EX_1 + EX_2$
3. X_1, X_2 ... independent RV $\Rightarrow E(X_1 . X_2) = EX_1 \, . \, EX_2$
4. $Y = g(X);$ g(X) is a continuous function: $EY = E(g(X))$

   Y is a continuous RV: $EY = \int_{-\infty}^{\infty} g(x). f(x) \, dx$

   Y is a discrete RV: $EY = \sum_i g(x_i) \, . \, P(X = x_i)$

3. Variance $DX = \mu_2 = E(X - EX)^2 = EX^2 - (EX)^2$

discrete RV: $DX = \sum_i x_i^2 . P(x_i) - (\sum_i x_i . P(x_i))^2$

continuous RV: $DX = \int_{-\infty}^{\infty} x^2 . f(x)dx - (\int_{-\infty}^{\infty} x . f(x)dx)^2$

Properties::

1. $D(aX + b) = a^2 . DX$
2. X_1, X_2 ... independent $\Rightarrow D(X_1 + X_2) = DX_1 + DX_2$

4. **Standard deviation** $\sigma_x = \sqrt{DX}$

5. **Skewness** $a_3 = \mu_3 / \sigma_x^3$

Is a level of symmetry for the given probability distribution and it is hold:

$a_3 = 0$ … symmetrical distribution
$a_3 < 0$ .... negative skewed set
$a_3 > 0$ .... positive skewed set

6. **Kurtosis** $a_4 = \mu_4 / \sigma_x^4$

Is a level of kurtosis (flatness):

$a_4 = 3$ .... normal kurtosis (i.e. kurtosis of a normal distribution)
$a_4 < 3$ .... lower kurtosis than normal distribution one (flatter)

$a_4 > 3$ .... greater kurtosis than normal distribution one (sharper)

7. Quantiles

$p \in (0,1)$

x_p ... 100p% quantile $\quad x_p = \sup\{x \mid F(x) \le p\}$

continuous RV: $\quad F(x_p) = p$

Special types of the quantiles:

$x_{0,5}$… 50% quantile is called a median
$x_{0,25}$ and $x_{0,75}$… 25% quantile is called a lower quartile and 75% quantile is called an upper quartile
$x_{k/10}$ … k =1,2,… 9 the k-th decile
$x_{k/100}$ … k =1,2,… 99 the k-th percentile

8. Mode

The mode $\hat{x}$ of a discrete RV X is such value that holds:

$$P(X = \hat{x}) \geq P(X = x_i) \qquad i = 1, 2, \ldots$$

It means that the mode is a value in which the discrete RV comes with the biggest probability.

The mode $\hat{x}$ of a continuous RV X is such value that holds:

$$f(\hat{x}) \geq f(x) \qquad \text{pro } -\infty < x < \infty$$

It is a value where the probability density has a maximum value.

4. RANDOM VECTOR

Study time: 60 minutes

Aim - you will be able to

- describe a random vector and its joint distribution
- explain the notions of marginal and conditional probability distribution
- explain a stochastic independence of random variables

Explication

4.1. Random vector

For continuous random variables, the joint distribution can be represented either in the form of a distribution function or of a probability density function.

$$F(x_1, x_2, \ldots, x_n) = P(X_1 < x_1, X_2 < x_2, \ldots, X_n < x_n), \quad F\colon R^n \to R$$

$$f(x_1, x_2, \ldots, x_n) = \frac{\partial^n F(x_1, x_2, \ldots, x_n)}{\partial x_1\, \partial x_2 \ldots \partial x_n}$$

The two forms are again equivalent. In terms of the joint probability density function, the joint distribution function of $X_1, \ldots, X_n$ is

$$F(x_1, x_2, \ldots, x_n) = \int_{-\infty}^{x_n} \ldots \int_{-\infty}^{x_2} \int_{-\infty}^{x_1} f(t_1, t_2, \ldots, t_n)\, dt_1\, dt_2 \ldots dt_n.$$

Although in theory the joint distribution of a discrete variable with a continuous variable does exist, there is no practical algebraic formulation of such a distribution. Such distributions are only represented in conditional form.

4.2. Marginal distribution

Definition

Let $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ be a random vector. The random vector $\boldsymbol{Y} = (X_{i_1}, X_{i_2}, \ldots, X_{i_k})$, where $k < n,\ i_j \in \{1, 2, \ldots, n\},\ i_u \neq i_v\ pro\ u \neq v,$ we called the **marginal random vector.** Especially,

X_i is the **marginal random variable** for every $i=1,2,\dots,n$. The probability distribution of $\boldsymbol{Y}$ we called the **marginal probability distribution**.

Let $\boldsymbol{X} = (X_1, X_2)$ be a bivariate random variable with given distribution function $F(x_1, x_2)$.

$F_1(x_1) = \lim\limits_{x_2 \to +\infty} F(x_1, x_2) = F(x_1, +\infty)$ …marginal distribution function of random variable X_1

$F_2(x_2) = \lim\limits_{x_1 \to +\infty} F(x_1, x_2) = F(+\infty, x_2)$ …marginal distribution function of random variable X_2

For continuous random variables, the marginal probability density of one jointly distributed variable is found by integrating the joint density function with respect the other variable.

$$f_1(x_1) = \int_{x_2} f(x_1, x_2)\,dx_2 \quad \text{for } X_1$$
$$f_2(x_2) = \int_{x_1} f(x_1, x_2)\,dx_1 \quad \text{for } X_2 .$$

For discrete random variables, the marginal distributions are given by:

$P_1(x_1) = \sum\limits_{x_2} P(X_1 = x_1, X_2 = x_2)$ … marginal distribution function of X_1

$P_2(x_2) = \sum\limits_{x_1} P(X_1 = x_1, X_2 = x_2)$ … marginal distribution function of X_2

4.3. Conditional distribution

The conditional distribution is the distribution of one variable at a fixed value of the other jointly distributed random variable. For two discrete variables, the conditional distribution is given by the ratio of the joint probabilities to the corresponding marginal probability.

$$p(x_1 \mid x_2) = P(X_1 = x_1 \mid X_2 = x_2) = \frac{P(X_1 = x_1, X_2 = x_2)}{P_2(x_2)} = \frac{p(x_1, x_2)}{P_2(x_2)} .$$

For continuous random variables, the conditional densities are given analogously by the ratio of the joint density to the corresponding marginal density.

$$f(x_1 \mid x_2) = \frac{f(x_1, x_2)}{f_2(x_2)} .$$

$f_2(x_2) = \int_{-\infty}^{+\infty} f(x_1, x_2)\,dx_1$ is the corresponding marginal density of X_2.

4.4. Independence of Random Variables

Definition

$X_1 \dots X_n$ are **mutually independent** $\Leftrightarrow$ the random events $\{X_i < x_i\}$, ($i=1,2,\dots,n$, where $x_i \in R$) are mutually independent.

Therefore, $X_1 \dots X_n$ are mutually independent $\Leftrightarrow$ $F(x_1,\dots,x_n) = F_1(x_1) \dots F_n(x_n)$.

It is true, because
$F(x_1,\ldots,x_n) = P(X_1<x_1, \ldots , X_n<x_n) = P(X_1< x_1) \,.\, P(X_2<x_2) \ldots\ldots P(X_n<x_n) = F_1(x_1) \,.\, F_2(x_2) \ldots\ldots F_n(x_n)$.

This implies the following rule::
$X_1 \ldots X_n$ are mutually independent $\Leftrightarrow$ $f(x_1,\ldots,x_n) = f_1(x_1) \,. \ldots . f_n(x_n)$.

Example:
X_1, X_2 are mutually independent. Determine the variance $X_1 + X_2$.

In general, if X_1 and X_2 are not independent, the variance of their sum is given by

$$D(X_1 + X_2) = D(X_1) + D(X_2) + 2\,Cov(X_1, X_2)$$

where the covariance of X_1 and X_2 is defined by

$$Cov(X_1, X_2) = E[(X_1 - E(X_1))(X_2 - E(X_2))]$$

When X_1 and X_2 are independent, the covariance is zero.
An alternate expression for the covariance similar to that for the variance and simpler for computation is

$$Cov(X_1, X_2) = E(X_1 X_2) - E(X_1)E(X_2).$$

Correlation coefficient

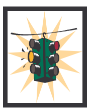
The correlation coefficient measures the strength of the relation between two random variables, X_1 and X_2. The correlation coefficient is defined by

$$\rho_{X_1X_2} = \frac{Cov(X_1, X_2)}{\sigma_{X_1}\sigma_{X_2}}.$$

The correlation coefficient properties are:

1. $-1 \le \rho \le 1$
2. $\rho(X,Y) = \rho(Y,X)$

The correlation assumes values between -1 and 1. A value close to 1 implies a strong positive relationship, a value close to -1 implies a strong negative relationship, and a value close to zero implies little or no relationship.



Solved example

Imagine that we will repeat the trial for three times (we known the success probability, e.g. coin throws).

Lets write all possible combinations: (S - success, F - false):

{ FFF; SFS; SSF; FSS; FSF; FFS; SFF; SSS }

Now we specify the following random variables:

Y ... a number of attempts to the first success
Z ... a number of the following successes

a) determine the probability function P(Y), P(Z)
b) set the joint probability function Y, Z
c) determine the marginal distribution function and P(Y | Z), P(Z | Y)

Solution:
ada) Y and Z are the discrete RV and that is why Y and Z can gain the values: 0, 1, 2, 3
Let's name all element events of the sample space:

A1 ... FFF $P(A1) = (1 - p)^3$
A2 ... SFS $P(A2) = p^2.(1 - p)$
A3 ... SSF $P(A3) = p^2.(1 - p)$
A4 ... FSS $P(A4) = p^2.(1 - p)$
A5 ... FSF $P(A5) = p.(1 - p)^2$
A6 ... FFS $P(A6) = p.(1 - p)^2$
A7 ... SFF $P(A7) = p.(1 - p)^2$
A8 ... SSS $P(A8) = p^3$

For our calculation we use the fact that the F and S variables are independent.

Y ... a number of attempts to first success			
0	1	2	3
SFS, SSF, SFF, SSS	FSS, FSF	FFS	FFF

Z ... a number of following successes			
0	1	2	3
FFF	SFS, FSF, FFS, SFF	SSF,FSS	SSS

Since A1, ... , A8 events are disjoint we can simply determine the probablity function (p=0.5).

Y ... a number of attempts to first success			
P(Y= 0)	P(Y= 1)	P(Y= 2)	P(Y= 3)
0.5	0.25	0.125	0.125

Z ... a number of following successes			
P(Z = 0)	P(Z = 1)	P(Z = 2)	P(Z = 3)
0.125	0.5	0.25	0.125

adb) we will proceed in the same way like we did in probability function finding
Z

Y	0	1	2	3
0	-	SFS, SFF	SSF	SSS
1	-	FSF	FSS	-
2	-	FFS	-	-
3	FFF	-	-	-

	Z			
Y	0	1	2	3
0	0	0.25	0.125	0.125
1	0	0.125	0.125	0
2	0	0.125	0	0
3	0.125	0	0	0

adc) marginal probability functions - P(Y), P(Z)

	Z				
Y	0	1	2	3	P(Y)
0	0	0.25	0.125	0.125	0.5
1	0	0.125	0.125	0	0.25
2	0	0.125	0	0	0.125
3	0.125	0	0	0	0.125
P(Z)	0.125	0.5	0.25	0.125	1

$P(Y \mid Z) = P(Y \wedge Z) / P(Z)$

	Z			
Y	0	1	2	3
0	0	0.5	0.5	1
1	0	0.25	0.5	0
2	0	0.25	0	0
3	1	0	0	0

$P(Z \mid Y) = P(Y \wedge Z) / P(Y)$

	Z			
Y	0	1	2	3
0	0	0.5	0.25	0.25
1	0	0.5	0.5	0
2	0	1	0	0
3	1	0	0	0

Summary of notions

Random variable X is a real function which can be characterized by a **distribution function** *F(t)*.

Distribution function is a function that assigns to each rea number a probability that the random variable will be less then this real number. Distribution function has some general properties like $\forall$ a, b $\in$ R; a < b platí $P(a \leq X < b) = F(b) - F(a)$.

According values the random variable may become we distinguish **continuous** and **discrete variable.**

The discrete random variable is also characterized by a **probability function**, the continuous one by a **density function**.

In many cases, it is useful to cover the whole information about random variable into several numbers that characterize some properties of random variable while allowing the comparison of different random variables. These numbers are called the **numerical characteristics** of random variable.

A **random vector** is a vector consisted of random variables $\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ that is characterized by **joint distribution function.**

From joint distribution function of random vector we can easily determine a **marginal probability distribution** of particular random variables the vector is composed of.

Questions

1. What is a mutual relationship between a distribution function and probability function of discrete random variable?
2. What is a mutual relationship between a distribution function and probability density function of continuous random variable?
3. What is a median and a mode?
4. Explain term of conditional probability distribution.
5. Explain term of stochastic independence of random variables.
6. What does a value of correlation coefficient tell us?

Problems

Example 1: Let Y be a continuous variable defined by a probability density function:

$$f(y) = \begin{cases} c.(1+y).(1-y); & -1 < y < 1 \\ 0 & \text{elsewhere} \end{cases}$$

Find a constant c, a distribution function, an expected value and a variance of this variable.

{*Answer:* c=0.75; F(y) = 0.25 $(3y - y^3+2)$; EY = 0; DY = 0.2}

Example 2: Let random variable W is defined as a linear transformation of random variable Y, defined in previous example.

$$W = 5Y + 6$$

Find a probability density function, a distribution function, an expected value and a variance of random variable W.

{*Answer:* $f(w) = -\frac{3}{500}(w^2-12w+11)$; $F(w) = 0.25\,[3(\frac{w-6}{5}) - (\frac{w-6}{5})^3 + 2]$; EW = 6; DW = 5}

Example 3: Let random variable Z be defined as:

$$f(z) = 1 / [(1+e^z).(1+e^{-z})]; \qquad -\infty < z < \infty$$

Find a distribution function of random variable Z.

{*Answer:* $F(z) = \frac{e^z}{1+e^z}$ }