Random Variable

Aim - you will be able to

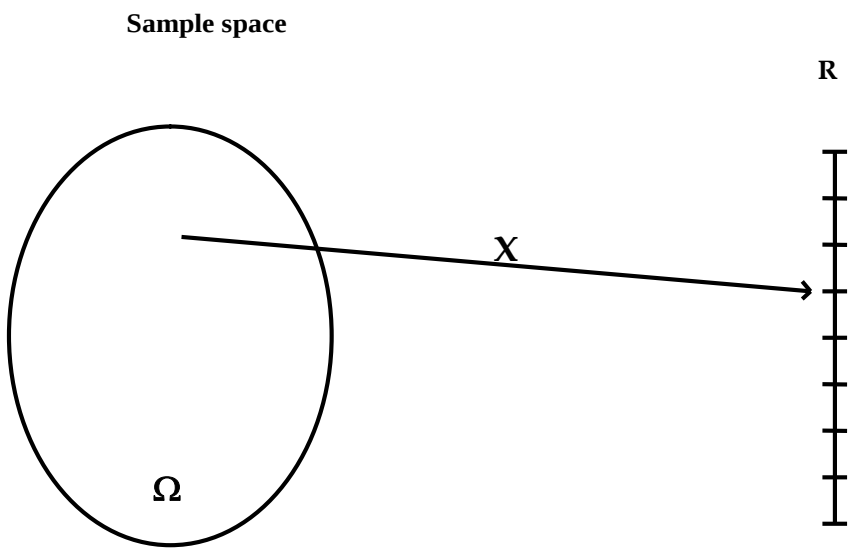
- describe the random variable by the distribution function
- characterize a discrete and a continuous random variable
- determine the numerical characteristics of the random variable
- understanding to mode and median

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*: Ω → R, that for each $x \in$ R we are able determine a probability:

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



A group of all values

$$\{x = X(\omega),\ \omega \in \Omega\}$$

is a range of values of the random variable

Distribution function

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

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

We call the *F(t)* as **distribution function** of *X*.

Distribution function is a function *X*: R → R

Which for each $t \in$ R assigns the probability:

$$P(X < t)$$

General properties of distribution function

1. $0 \le F(x) \le 1$

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. $\lim\limits_{x \to +\infty} F(x) = 1 \quad \lim\limits_{x \to -\infty} F(x) = 0$

5. $\forall$ a, b $\in$ R; a < b must be: $P(a \le X < b) = F(b) - F(a)$

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

Discrete random variable

Definition

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

$\exists$ finite or enumerable set of real numbers M={ $x_1, \dots, x_{n,\dots}$} that

$P(X = x_i) > 0$ i = 1, 2, ...

$$\sum_i \mathrm{P}(X = x_i) = 1$$

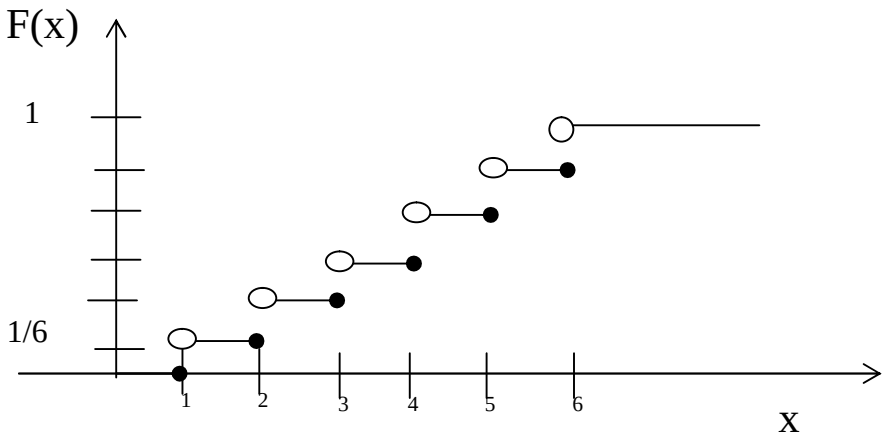
Function $\mathrm{P}(X = x_i) \Leftrightarrow \mathrm{P}(x_i)$ is called **probability function of random variable** *X*.

A distribution function of such distribution is a step function with steps in $x_1, \dots, x_{n,..}$

For a distribution function of discrete random variable it holds:

$$F(x) = P(X < x) =$$

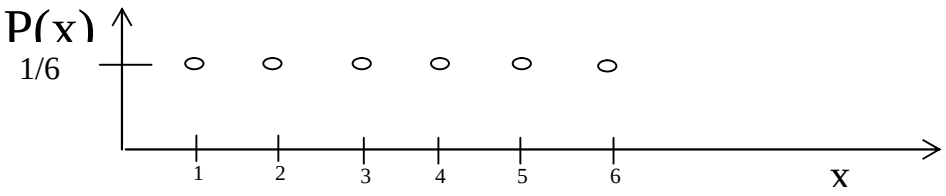
$$\sum_{x_i < x} P(X = x_i)$$



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



Continuous random variable

Definition

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

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

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:

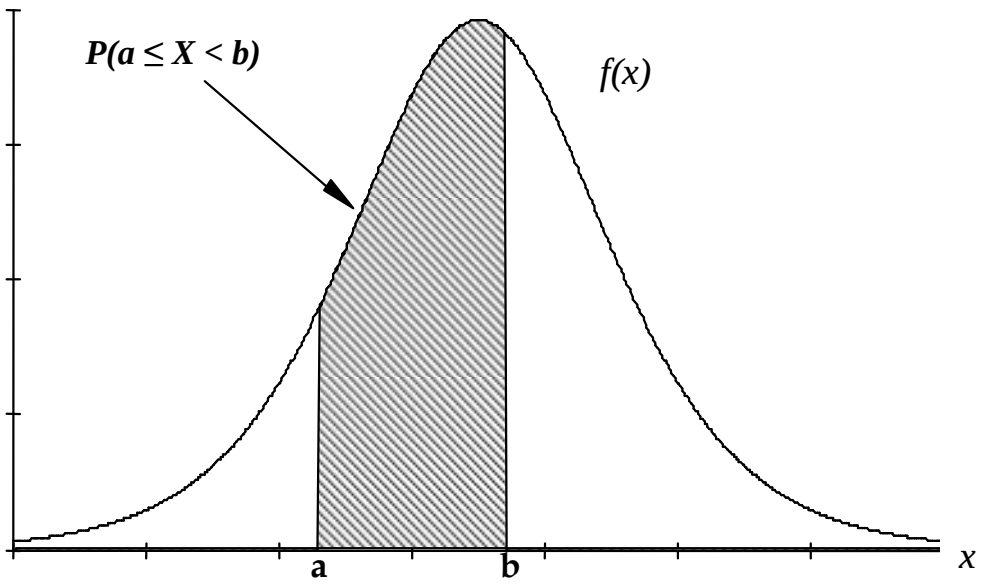
$$f(x) = \frac{dF(x)}{dx}$$

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

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

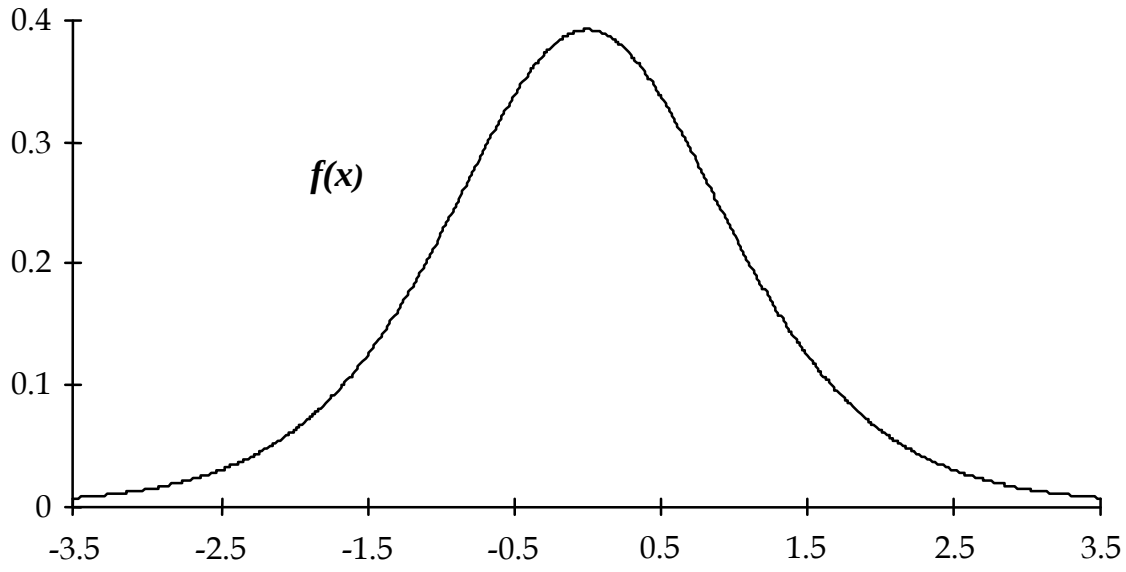
$$P(a \le X < b) = F(b) - F(a) =$$

$$\int_{-\infty}^{b} f(t)dt - \int_{-\infty}^{a} f(t)dt = \int_{a}^{b} f(t)dt$$



One of attributes for each density probability is the fact that the whole area under curve is equal to one.

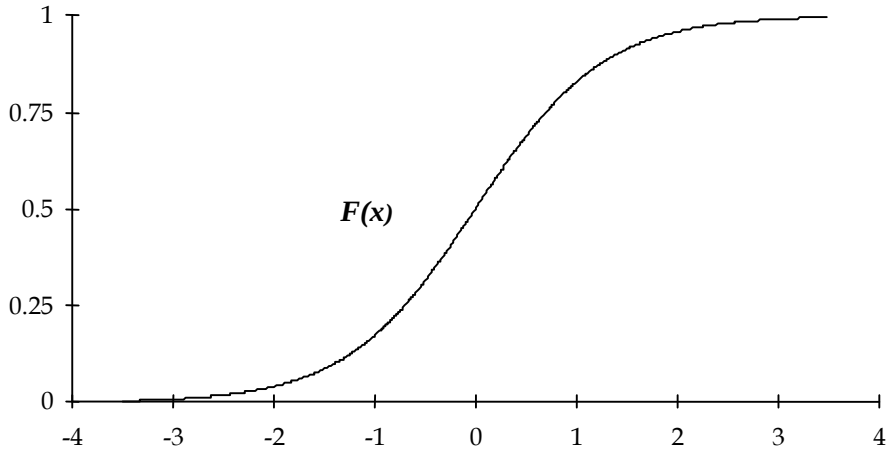
$$\int_{-\infty}^{\infty} f(x)dx = 1$$

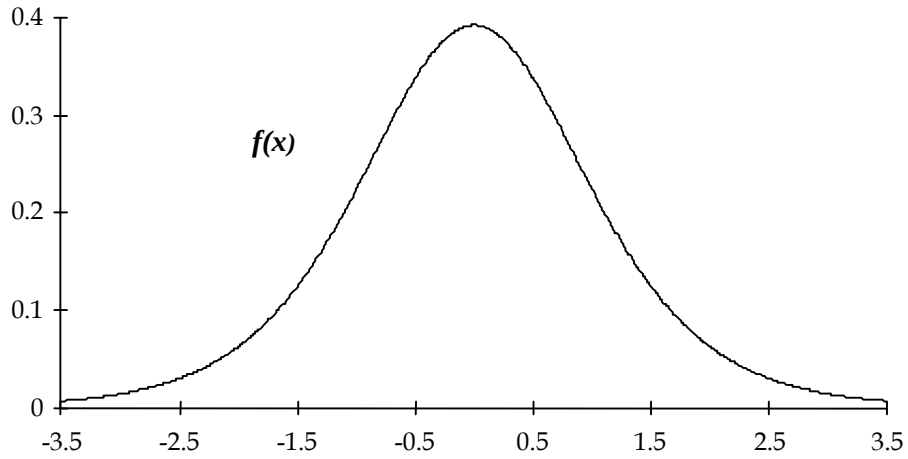


Example: Logistic probability distribution has a following distribution function *F(x)* and a probability density *f(x):*

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

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





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

Numerical characteristics of random variable

Expected value … EX

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

Y is a discrete RV: $$EY = \sum_{i} g(x_i) \,.\, P(X = x_i)$$

Moments

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

discrete RV:

$$\mu_r' = \sum_i x_i^r \,.\, P(x_i)$$

Continuous RV:

$$\mu_r' = \int_{-\infty}^{\infty} x^r f(x)\, dx \qquad r = 0,1,2, \ldots$$

r-th central moment is denoted μ_r

$$\mu_r = E[\,X - EX\,]^r \qquad r = 0,1,2,\ldots$$

Discrete RV: $\mu_r = \sum_{i} [\,x_i - EX\,]^r \,.\, P(\,x_i\,)$

Continuous RV: $\mu_r = \int_{-\infty}^{\infty} (\,x - EX\,)^r \,.\, f(\,x\,)\,dx \qquad r = 0,1,2,\ \ldots$

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 $\Rightarrow D(X_1 + X_2) = DX_1 + DX_2$

Standard deviation

$$\sigma_x = \sqrt{DX}$$

ILLUSTRATION OF SKEWNESS

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

The Gamma distribution, a theoretical distribution commonly used to study reliability of manufactured products, has **postive skew**. As a result, the mean is greater than the median and both are much greater than the mode.

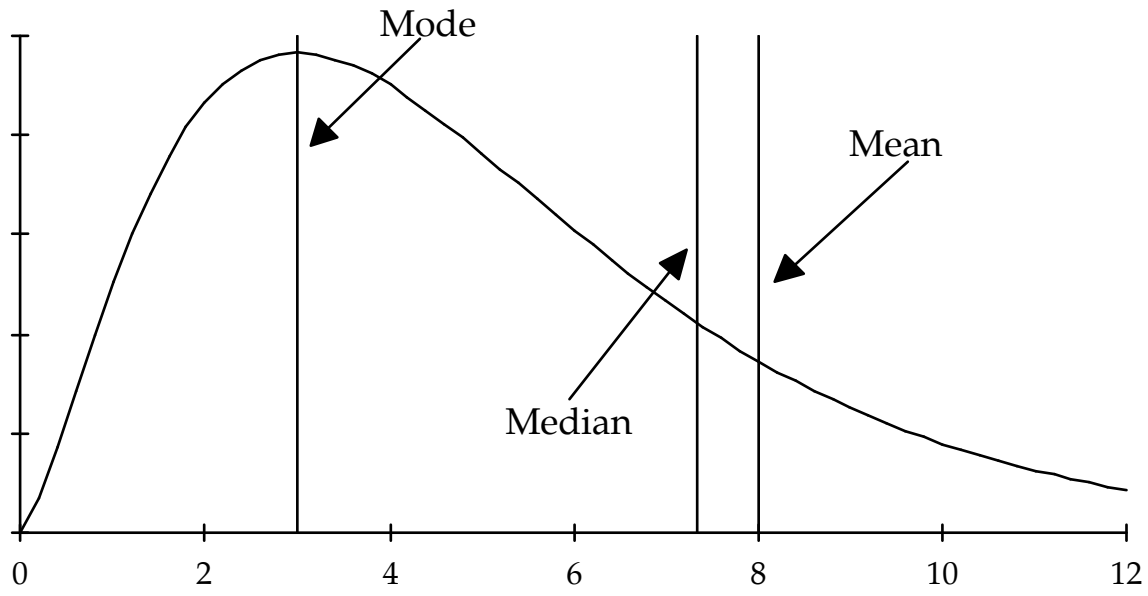


ILLUSTRATION OF KURTOSIS

Kurtosis $a_4 = \mu_4 / \sigma_x^4$

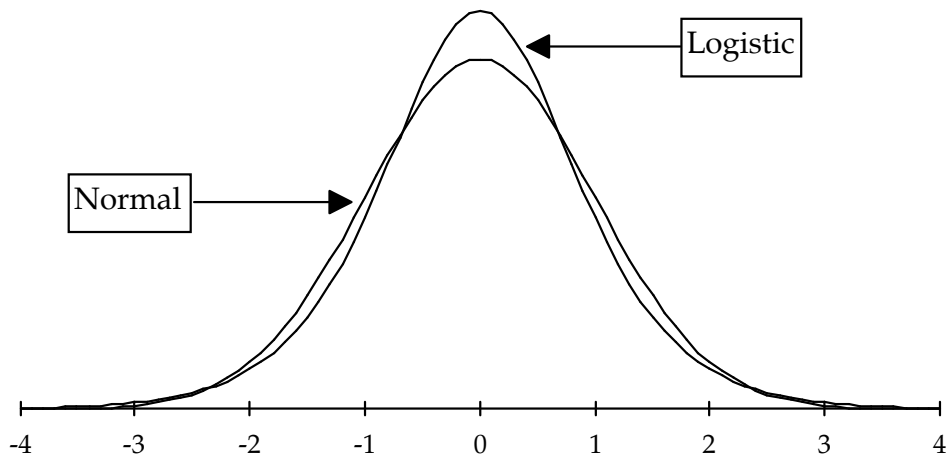
Is a level of kurtosis (flatness):

$a_4 = 3$ .... 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)

Two theoretical symmetric distributions with different kurtosis are the normal and the logistic. In the following illustration, both distributions have the same mean and standard deviation, but the logistic distribution has **greater kurtosis than the standard normal**.



Quantiles x_p

$$p \in (0,1)$$

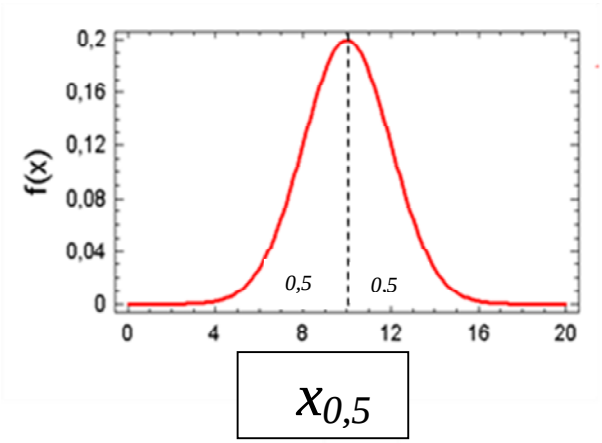
x_p ... 100p% quantile $\quad x_p = \sup\{x \mid F(x) \leq p\}$

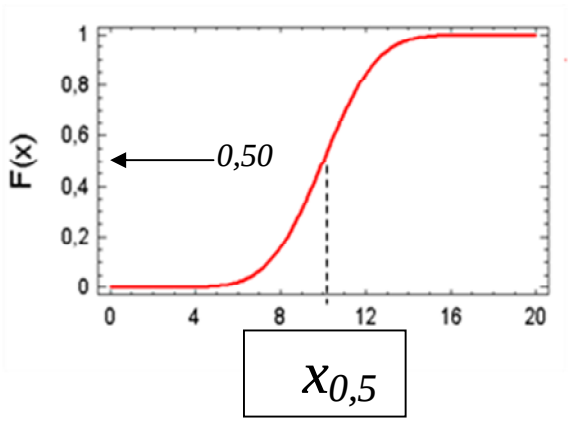
continuous RV: $\quad F(x_p) = p$

<u>Special types of the quantiles:</u>

$x_{0,5}$… 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**

Median





Median $x_{0,5}$**:** $F(x_{0,5}) = 0{,}5$

lower quartile $F(x_{0,25}) = 0{,}25$

upper quartile $F(x_{0,75}) = 0{,}75$

Mode

$\hat{x}$

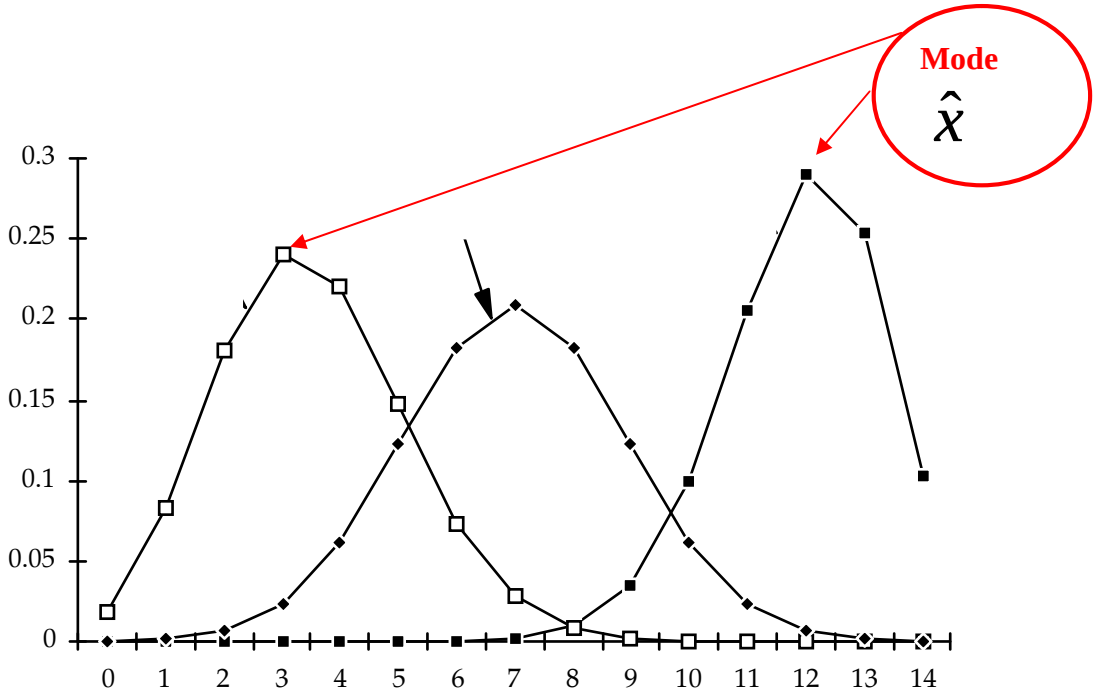
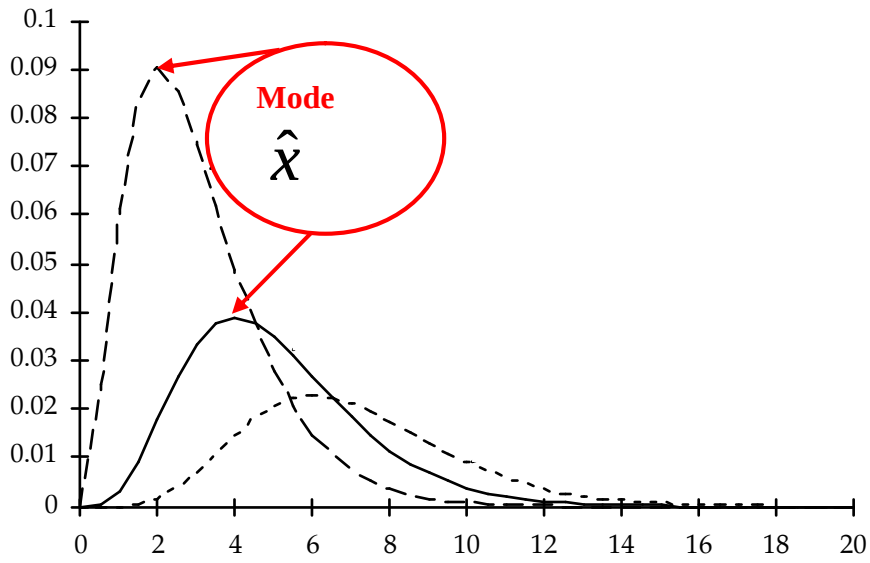
The mode $\hat{x}$ of a discrete RV X is such value that holds:

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

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

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

Mode of discrete variable



Mode of continuous variable