Random Data

In statistics there are 3 types of random variables:

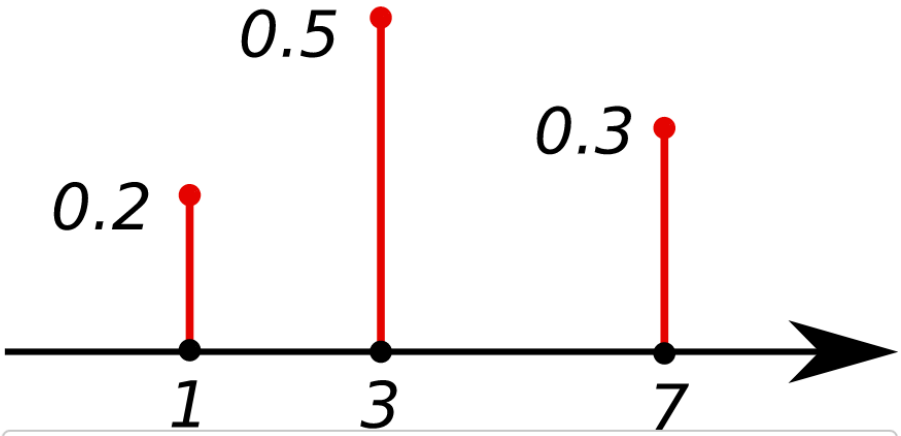
- **Discrete /Дискретни/**
- **Continuous /Непрекъснати/**
- **Mixtures /Смеси/**

In this course we are going to review only the first two.

Let's first explain some terms that we will need to use.

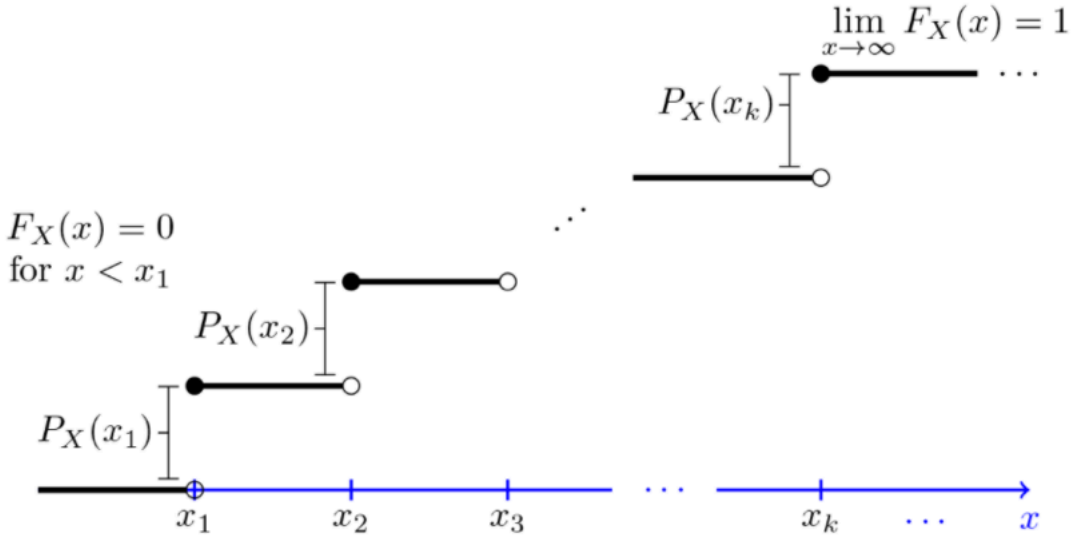
- For discrete random variable:
  - **Probability mass function (PMF) /Функция (Ред) на разпределение/** – is a function that gives the probability that a discrete random variable is exactly equal to some value. $\mathbb{P}_X(x_i) := \mathbb{P}(X = x_i)$

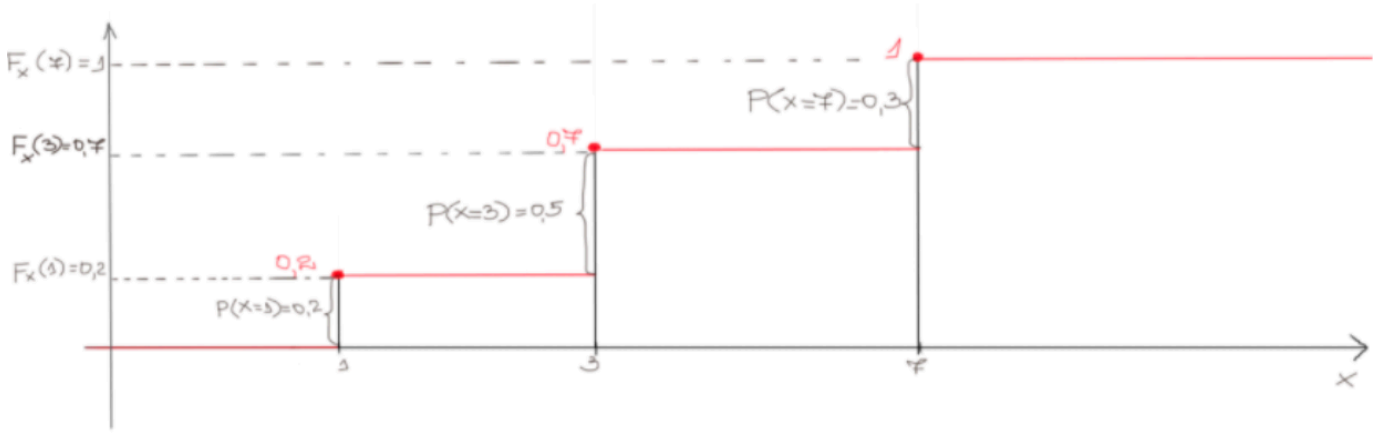
Example: $\mathbb{P}_X(x_i) = \mathbb{P}(X = x_i) = \begin{cases} 0.2, & x_i = 1 \\ 0.5, & x_i = 3 \\ 0.3, & x_i = 7 \\ 0, & \text{otherwise} \end{cases}$




```
* **Cumulative distribution function (CDF) /Кумул
ативна функция на разпределение/** - is the probabilit
y that a random variable $X$ will take a value le
ss than or equal to $x$.
```

$$F_X(x) := \mathbb{P}(X \leq x) = \sum_{i:x_i \leq x} \mathbb{P}(X = x_i)$$

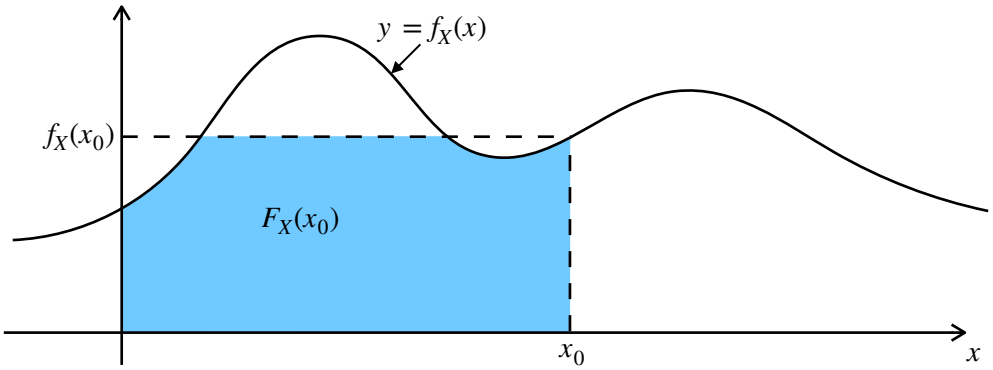




For absolutely continuous random variable:

- **Probability density function (PDF) /Плътност на разпределение/** – If the probability density function $f_X(x)$ exist, it is a non-negative function such that:

$$F_X(x_0) = \int_{-\infty}^{x_0} f_X(x)\mathrm{d}\,x \quad \text{and} \quad \int_{-\infty}^{\infty} f_X(x)\mathrm{d}\,x = 1.$$

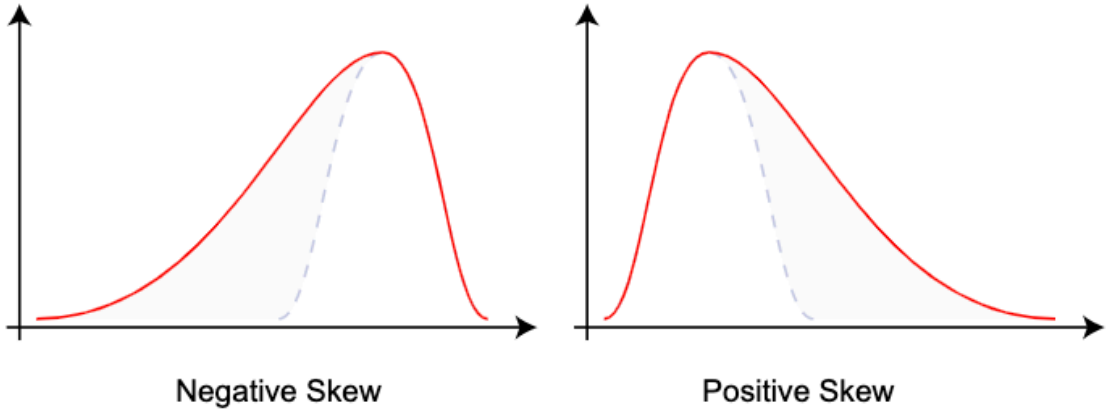


- **Cumulative distribution function (CDF) /Кумулативна функция на разпределение/** – is the probability that a random variable X will take a value less than or equal to x. $F_X(x) := \mathbb{P}(X \leq x)$

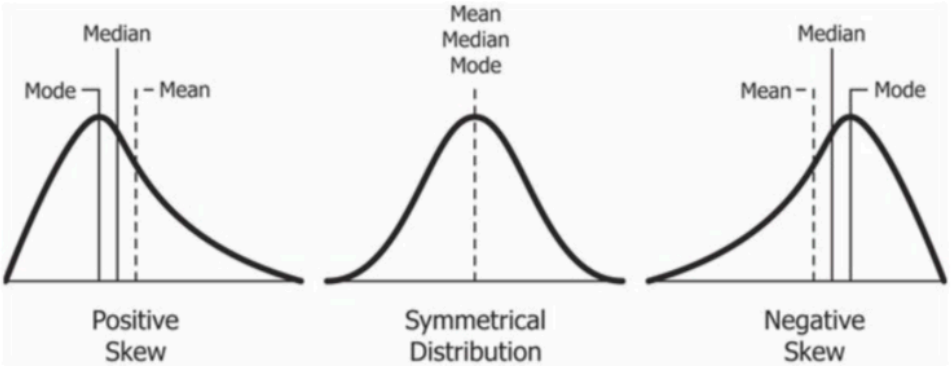
- **Moments** – quantitative measures related to the shape of the function
  - **First moment /Expectation, Mean/** $\mu = \mathbb{E}[X] = \int_{-\infty}^{+\infty} x \, \mathrm{d}\, F_X(x)$
  - **Second central moment /Variance/**

$$\begin{aligned}\sigma^2 = Var[X] = \mathbb{D}[X] = \mathbb{E}[(X - \mathbb{E}[X])^2] = \\ = \mathbb{E}[X^2 - 2X\mathbb{E}[X] + (\mathbb{E}[X])^2] = \\ = \mathbb{E}[X^2] - 2\mathbb{E}X\mathbb{E}[X] + (\mathbb{E}X)^2 = \\ = \mathbb{E}[X^2] - (\mathbb{E}[X])^2\end{aligned}$$

  - **Third standardized moment /Skewness/** $\mathbb{E}\left[\left(\frac{X-\mu}{\sigma}\right)^3\right] = \frac{\mathbb{E}(X-\mu)^3}{\sigma^3}$.

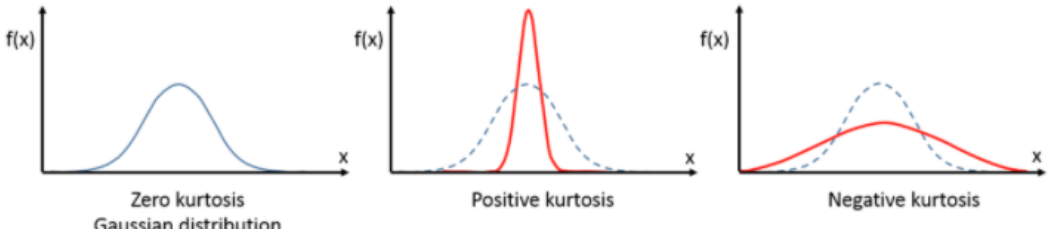


We also say that the data on the first graphic is left-skewed and the data on the second graphic is right-skewed.



- **Fourth standardized moment** $\mathbb{E}\left[\left(\frac{X-\mu}{\sigma}\right)^4\right] = \frac{\mathbb{E}(X-\mu)^4}{\sigma^4}$

Kurtosis = Fourth standardized moment – 3



They are also named: **Mesokurtic /Normal/, Leptokurtic, Platykurtic**

Discrete Distriutions

Discrete uniform distribution /Дискретно равномерно разпределение/

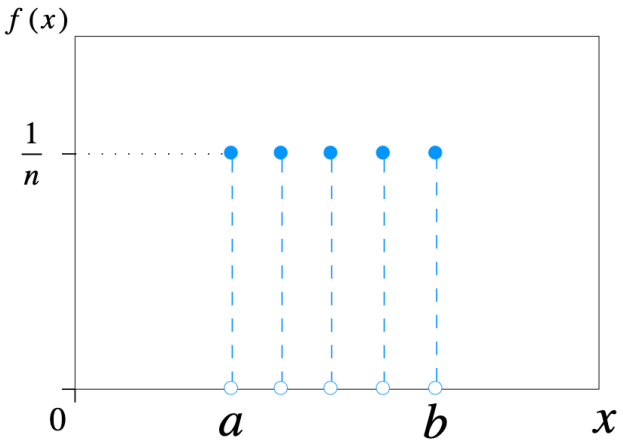
Discrete uniform distribution contains a known, finite number of outcomes equally likely to happen. In statistics we use the notation **DU** and **DUnif**.

$$X \in DU(a,b), X \in DUnif(a,b)$$

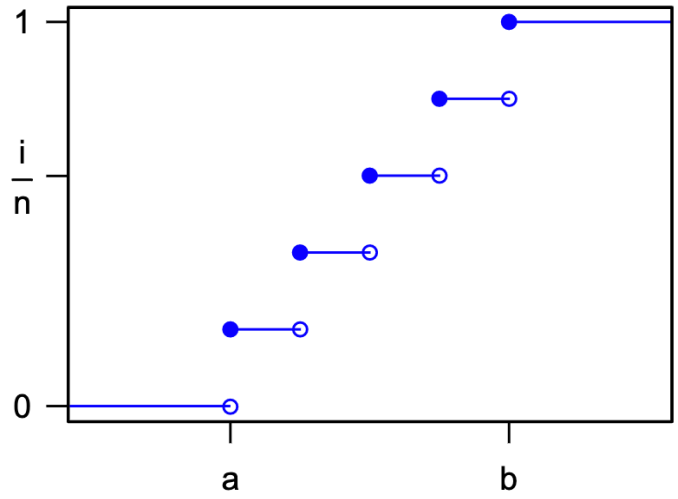
Where a is the min, b is the max possible value and $n = b - a + 1$

Probability mass function (PMF)

$$\mathbb{P}_X(x_i) = \mathbb{P}(X = x_i) = \begin{cases} \frac{1}{n}, & x_i = a+1, \ldots, b \\ 0, & \text{otherwise} \end{cases}$$



Cumulative distribution function (CDF)



Mean

$$\mu = \mathbb{E}[X] = \int_{-\infty}^{+\infty} x \, \mathrm{d}\, F_X(x) = \sum_{i=a}^{b} i\mathbb{P}(X = i) =$$

$$= \sum_{i=a}^{b} i\frac{1}{n} = \frac{1}{n}\sum_{i=a}^{b} i = \frac{1}{n}(a + \ldots + b) =$$

$$= \frac{[(a-1)+1] + \ldots + [(a-1)+(b-(a-a))]}{n} =$$

$$= \frac{n(a-1) + \sum_{i=1}^{b-a+1}}{n} = \frac{n(a-1) + \sum_{i=1}^{n} i}{n} =$$

$$= (a-1) + \frac{n(n+1)}{2n} = a - 1 + \frac{b-a+2}{2} =$$

$$= \frac{a+b}{2}$$

Variance

$$\sigma^2 = \frac{(b-a+1)^2 - 1}{12}$$

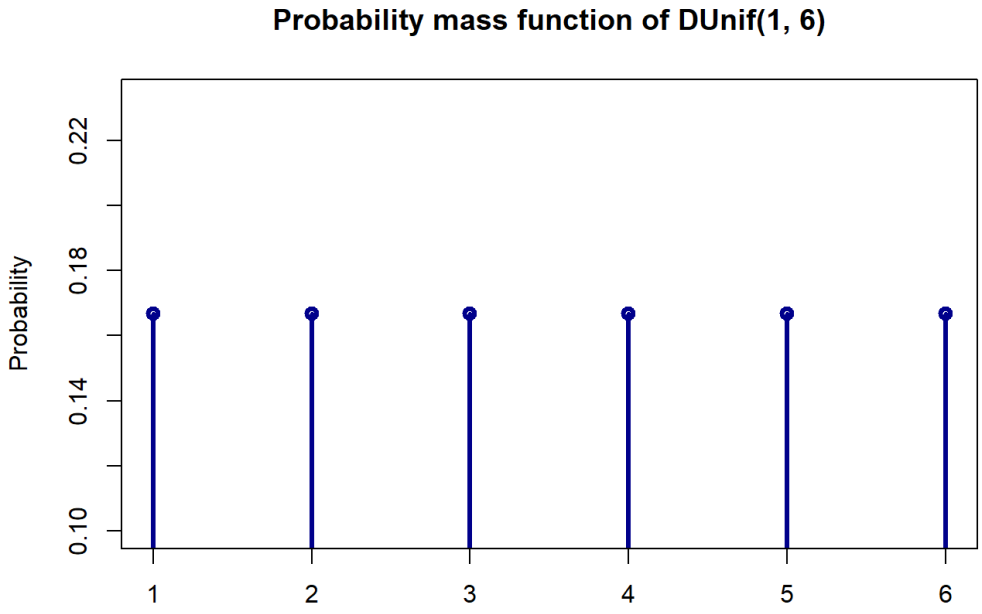
Example 1

As a first example let's consider rolling a regular dice. It will have discrete uniform distribution $DUnif(1,6)$.

Probability mass function (PMF)

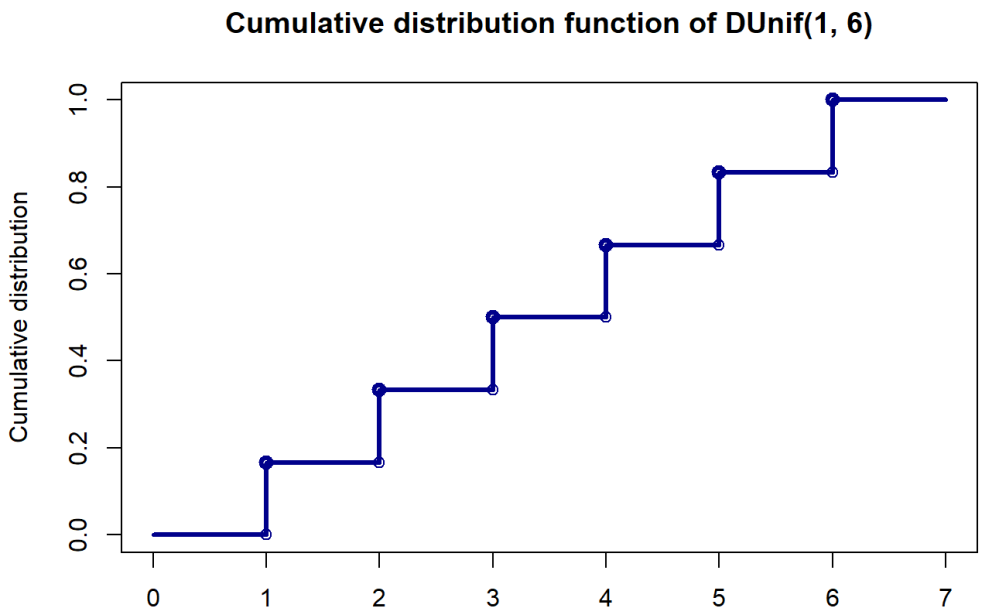
```
> min <- 1; max <- 6
> n <- max - min + 1
> 1/n
[1] 0.1666667
```

```
> plot(min:max, rep(1/n, n),
+      main = "Probability mass function of DUnif(1, 6)",
+      xlab = "", ylab = "Probability",
+      type = "h", lwd = 3, col = "darkblue")
> points(min:max, rep(1/n, n),
+        type = "p", lwd = 3, col = "darkblue")
```



Cumulative distribution function (CDF)

```
> 1/n
[1] 0.1666667
> plot((min-1):(max+1), c(0, cumsum(rep(1/n, n)), 1),
+      main = "Cumulative distribution function of DUnif(1, 6)",
+      xlab = "", ylab = "Cumulative distribution",
+      type = "s", lwd = 3, col = "darkblue")
> points(min:max, cumsum(rep(1/n, n)),
+        type = "p", lwd = 3, col = "darkblue")
> points(min:max, c(0, cumsum(rep(1/n, n-1))),
+        type = "p", lwd = 1, col = "darkblue")
```



We can simulate an outcome from a dice rolling using the sample function.

```
> sample(1:6, size = 1, replace = TRUE)
[1] 4
```

You can see that every time it generates a new number

```
> sample(1:6, size = 1, replace = TRUE)
[1] 3
> sample(1:6, size = 1, replace = TRUE)
[1] 3
> sample(1:6, size = 1, replace = TRUE)
[1] 4
```

But if we want for example to save our work and to continue later on or to send it to a friend or you need someone to check it it is useful always to generate the same random numbers. You can do this using the set.seed function.

```
> set.seed(10)
> sample(1:6, size = 1, replace = TRUE)
[1] 3
> set.seed(10)
> sample(1:6, size = 1, replace = TRUE)
[1] 3
> set.seed(10)
> sample(1:6, size = 1, replace = TRUE)
[1] 3
```

Let's simulate 10 outcomes from rolling a regular dice.

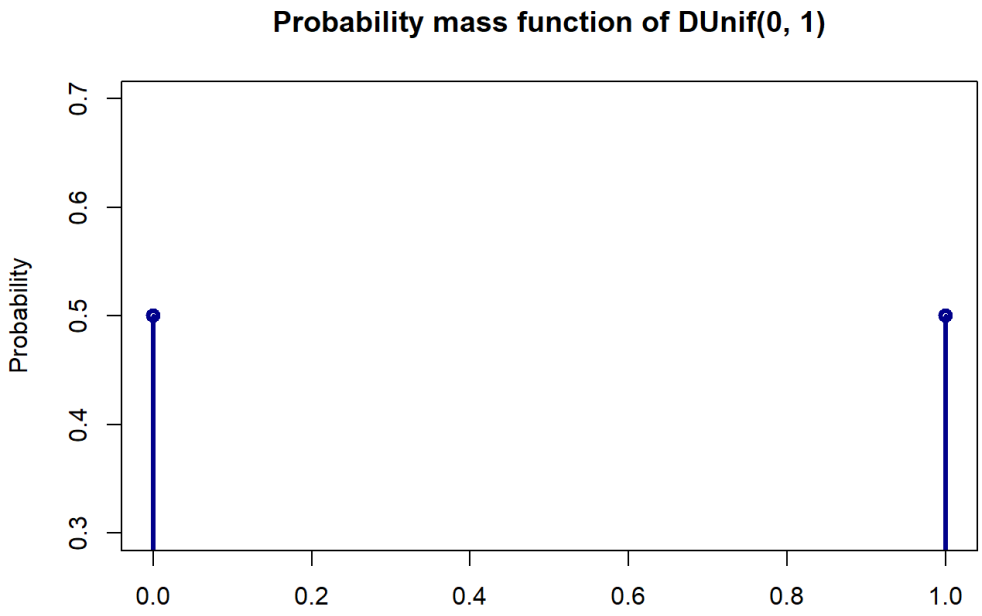
```
> sample(1:6, size = 10, replace = TRUE)
 [1] 1 2 4 6 3 2 2 2 5 6
```

Example 2

Another example can be a regular coin toss. It will have discrete uniform distribution $DUnif(0,1)$.

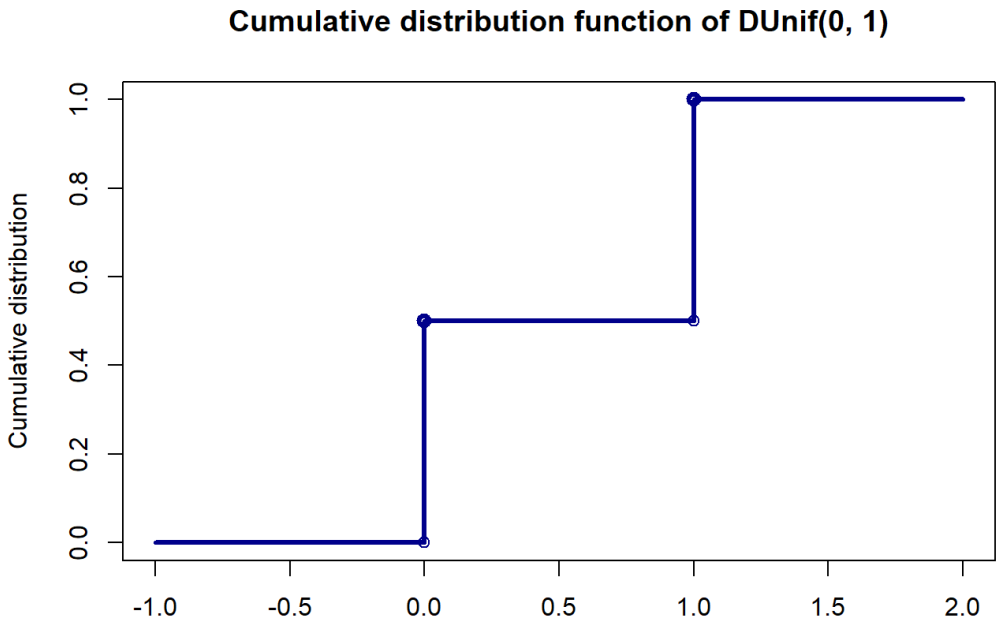
Probability mass function (PMF)

```
> min <- 0; max <- 1
> n <- max - min + 1
> plot(min:max, rep(1/n, n),
+      main = "Probability mass function of DUnif(0, 1)",
+      xlab = "", ylab = "Probability",
+      type = "h", lwd = 3, col = "darkblue")
> points(min:max, rep(1/n, n),
+        type = "p", lwd = 3, col = "darkblue")
```



Cumulative distribution function (CDF)

```
> plot((min-1):(max+1), c(0, cumsum(rep(1/n, n)), 1),
+      main = "Cumulative distribution function of DUnif(0, 1)",
+      xlab = "", ylab = "Cumulative distribution",
+      type = "s", lwd = 3, col = "darkblue")
> points(min:max, cumsum(rep(1/n, n)),
+        type = "p", lwd = 3, col = "darkblue")
> points(min:max, c(0, cumsum(rep(1/n, n-1))),
+        type = "p", lwd = 1, col = "darkblue")
```



We can simulate 10 outcomes from tossing a symmetric coin.

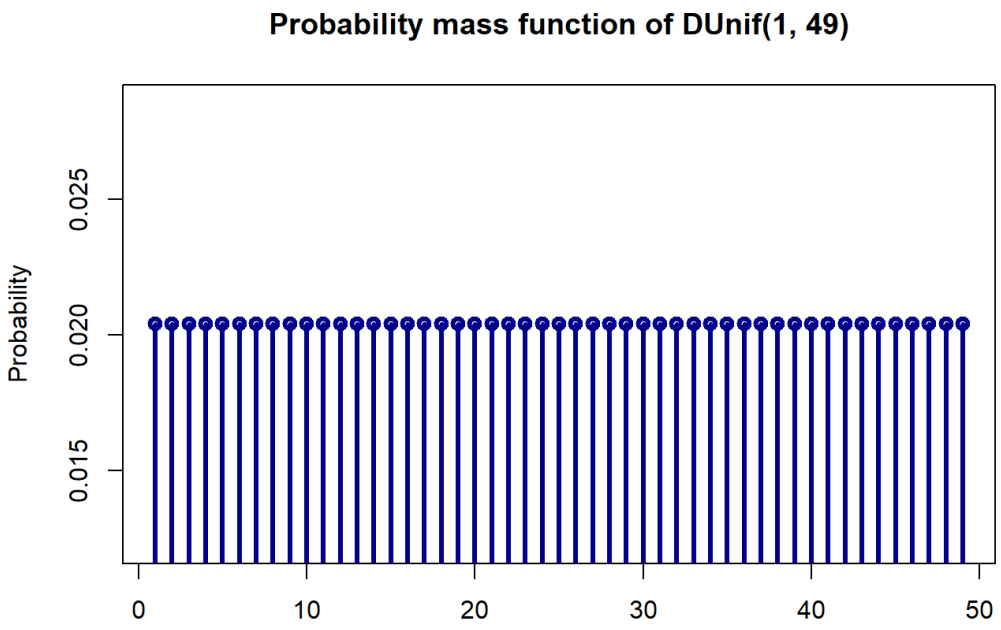
```
> sample(c("Head","Tail"), size = 10, replace = TRUE)
 [1] "Head" "Tail" "Head" "Head" "Tail" "Tail" "Head" "Head" "Head" "Head"
```

Example 3

Another example can be the drawing of a number from 6 of 49 lottery numbers. If we have equal probability to draw each number we will have discrete uniform distribution $DUnif(1,49)$.

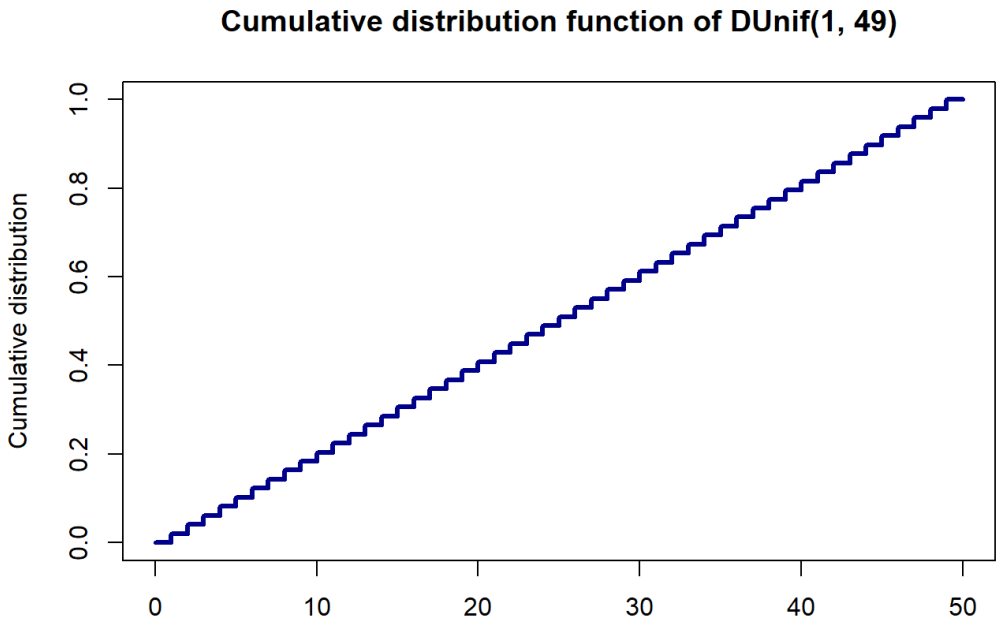
Probability mass function (PMF)

```
> min <- 1; max <- 49
> n <- max - min + 1
> plot(min:max, rep(1/n, n),
+      main = "Probability mass function of DUnif(1, 49)",
+      xlab = "", ylab = "Probability",
+      type = "h", lwd = 3, col = "darkblue")
> points(min:max, rep(1/n, n),
+        type = "p", lwd = 3, col = "darkblue")
```



Cumulative distribution function (CDF)

```
> plot((min-1):(max+1), c(0, cumsum(rep(1/n, n)), 1),
+      main = "Cumulative distribution function of DUnif(1, 49)",
+      xlab = "", ylab = "Cumulative distribution",
+      type = "s", lwd = 3, col = "darkblue")
```



We can simulate drawing a number of six 6 of 49 lottery numbers.

```
> sample(1:49, size = 1)
[1] 28
```

Example 4

Simulating drawing 1 card from a deck at random.

First let's construct our deck of cards.

```
> cards <- paste(rep(c("Ace",2:10,"Jack","Queen","King"), 4),
+                c("Heart","Diamond","Spade","Club"))
> cards
 [1] "Ace Heart"     "2 Diamond"     "3 Spade"       "4 Club"
 [5] "5 Heart"       "6 Diamond"     "7 Spade"       "8 Club"
 [9] "9 Heart"       "10 Diamond"    "Jack Spade"    "Queen Club"
[13] "King Heart"    "Ace Diamond"   "2 Spade"       "3 Club"
[17] "4 Heart"       "5 Diamond"     "6 Spade"       "7 Club"
[21] "8 Heart"       "9 Diamond"     "10 Spade"      "Jack Club"
[25] "Queen Heart"   "King Diamond"  "Ace Spade"     "2 Club"
[29] "3 Heart"       "4 Diamond"     "5 Spade"       "6 Club"
[33] "7 Heart"       "8 Diamond"     "9 Spade"       "10 Club"
[37] "Jack Heart"    "Queen Diamond" "King Spade"    "Ace Club"
[41] "2 Heart"       "3 Diamond"     "4 Spade"       "5 Club"
[45] "6 Heart"       "7 Diamond"     "8 Spade"       "9 Club"
[49] "10 Heart"      "Jack Diamond"  "Queen Spade"   "King Club"
```

Now we can draw 5 cards from it with replacement. These are 5 realizations of $DUnif(\{deck\})$.

```
> sample(x = cards, size = 5, replace = TRUE)
[1] "Jack Diamond" "6 Heart"      "King Diamond" "Jack Heart"   "7 Heart"
```

Example 5

Let's simulate rolling 2 dice 3 times independently.

Using the outer function we can take the product of two arrays. This is our **sample space**.

```
> dice <- as.vector(outer(1:6, 1:6, paste)); dice
 [1] "1 1" "2 1" "3 1" "4 1" "5 1" "6 1" "1 2" "2 2" "3 2" "4 2" "5 2" "6 2"
[13] "1 3" "2 3" "3 3" "4 3" "5 3" "6 3" "1 4" "2 4" "3 4" "4 4" "5 4" "6 4"
[25] "1 5" "2 5" "3 5" "4 5" "5 5" "6 5" "1 6" "2 6" "3 6" "4 6" "5 6" "6 6"
```

These are ordered couples.

Now we can take 3 elementary outcomes from our sample space with equal probability.

```
> sample(x = dice, size = 3, replace = TRUE)
[1] "1 2" "2 5" "6 5"
```

Bernoulli distribution /Бернулиево разпределение/

Bernoulli distribution, named after Swiss mathematician **Jacob Bernoulli** is the probability distribution of a random variable which takes the value 1 with probability p and the value 0 with probability $q = 1 - p$

$$X \in Bernoulli(p)$$

If A is an event with probability p then X is frequently called **indicator** of the event A and is denoted by I_A.

A special case of binomial distribution random variable $X \in Bi(1, p)$.

Probability mass function (PMF)

$$\mathbb{P}_X(x_i) = \mathbb{P}(X = x_i) = \begin{cases} p, & x_i = 1 \\ q, & x_i = 0 \\ 0, & \text{otherwise} \end{cases}$$

Mean

$$\mu = p$$

Variance

$$\sigma^2 = pq$$

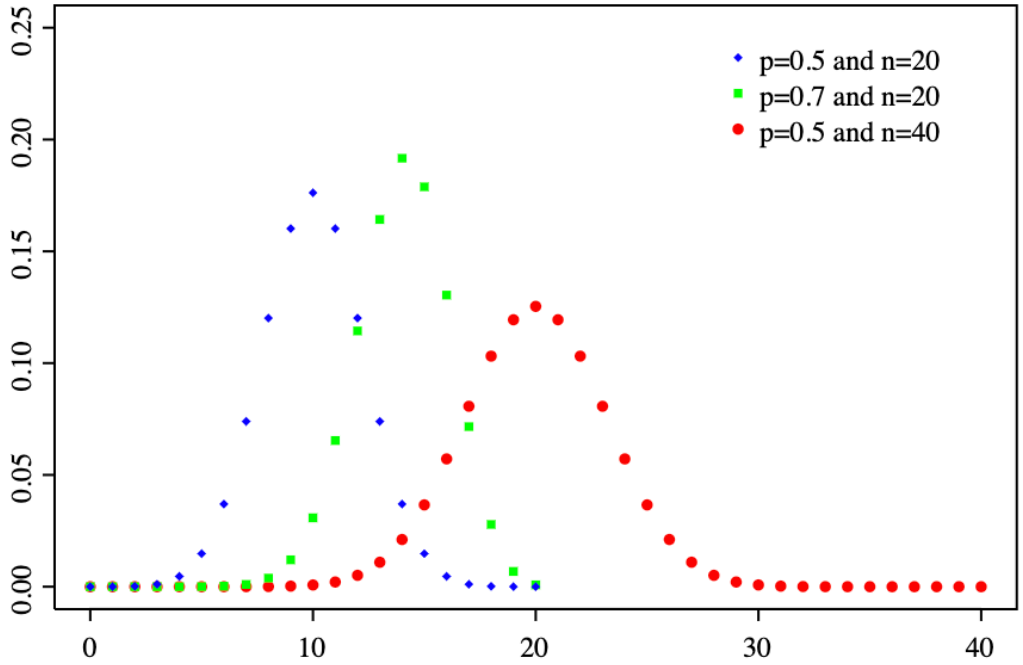
Binomial distribution /Биномно разпределение/

Binomial distribution with parameters n and p represents probability distribution of the number of successes in a sequence of n independent experiments each with probability for success p.

$$X \in Bi(n, p)$$

Probability mass function (PMF)

$$\mathbb{P}_X(k) = \mathbb{P}(X = k) = \binom{n}{k} p^k q^{n-k}, k = 0, 1, \ldots, n$$



X can be represented as sum of n independent indicators of events with probability

$$p = \mathbb{P}(A_i), i = 1, \ldots, n.$$
$$X = I_{A_1} + I_{A_2} + \ldots + I_{A_n}.$$

Mean

$$\mu = \mathbb{E}[X] = \mathbb{E}[I_{A_1} + I_{A_2} + \ldots I_{A_n}] =$$
$$= \mathbb{E}[I_{A_1}] + \mathbb{E}[I_{A_2}] + \ldots + \mathbb{E}[I_{A_n}] = p + \ldots + p = np$$

Variance

$$\sigma^2 = D[I_{A_1}] + D[I_{A_2}] + \ldots + D[I_{A_n}] = pq + \ldots + pq = npq$$

Using the rbinom function we can generate a sample of observations of a binomially distributed random variable. The first parameter of the function is how many numbers we want to generate and the next two parameters coincide with the parameters of the binomial distribution.

Let's generate a realization of $Bi(n, p)$ random variable. This is the same as making $n = 1$ experiment with probability for success $p = 0.5$ and counting the number of successes.

```
> n <- 1
> p <- 0.5
> rbinom(1, n, p)
[1] 0
```

Let's generate a realization of $Bi(10, 0.5)$ random variable. This is the same as making ten experiments with probability for success $p = 0.5$ and counting the number of successes.

```
> n <- 10
> p <- 0.5
> rbinom(1, n, p)
[1] 3
```

Let's generate 15 realizations of $Bi(10, 0.5)$ random variable. Each of them is the same as making ten experiments with probability for success $p = 0.5$ and counting the number of successes.

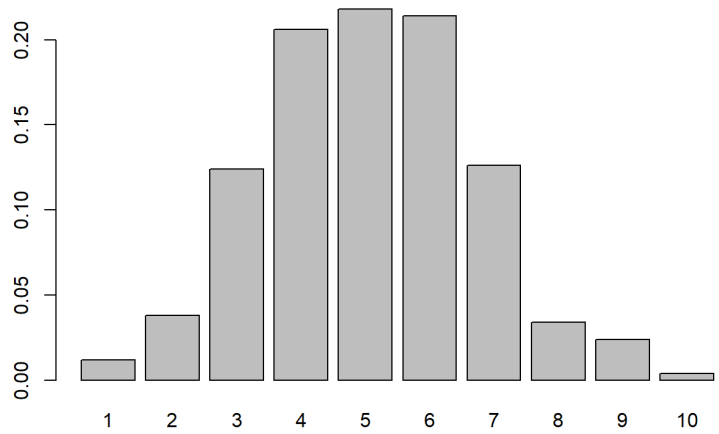
```
> n <- 10
> p <- 0.5
> rbinom(15, n, p)
 [1] 2 5 3 6 4 7 4 5 4 5 5 5 5 5 2
```

Let's generate 500 realizations of $Bi(10, 0.5)$ random variable.

```
> x <- rbinom(500, n, p)
```

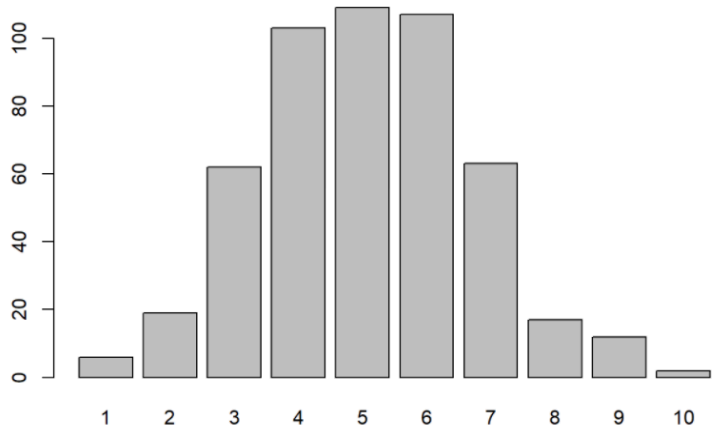
What is the range of the observations? Estimate the probability for occurrence of each one of them.

```
> table(x)
x
  1   2   3   4   5   6   7   8   9  10
  6  19  62 103 109 107  63  17  12   2

> prop.table(table(x))
x
    1     2     3     4     5     6     7     8     9    10
0.012 0.038 0.124 0.206 0.218 0.214 0.126 0.034 0.024 0.004

> barplot(table(x))
```



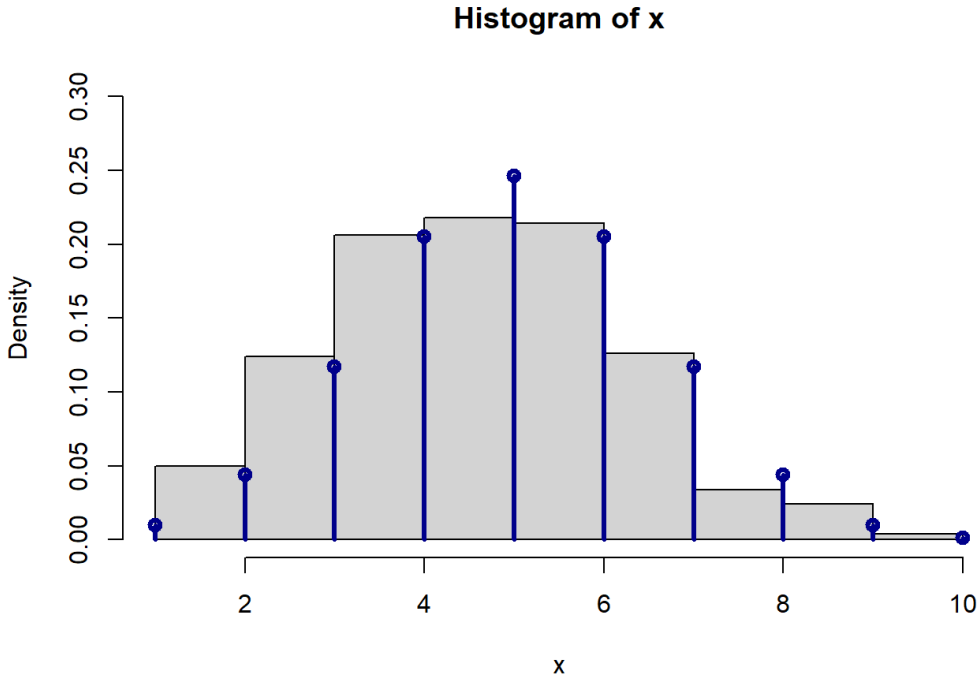
The corresponding empirical probability mass function (PMF) is

```
> barplot(prop.table(table(x)))
```

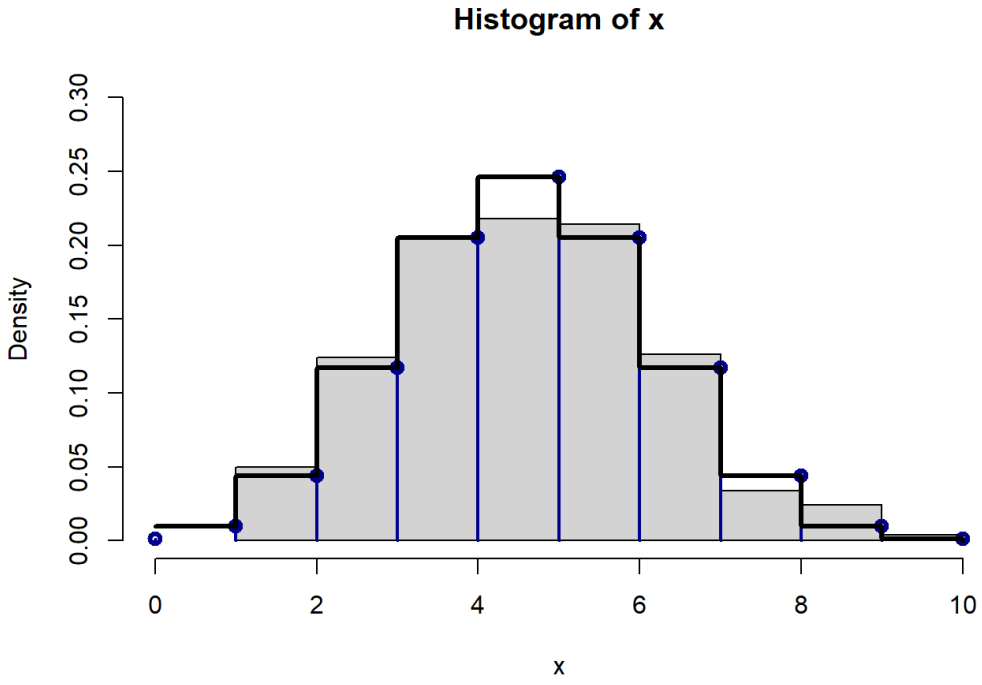


The corresponding **theoretical probability mass function (PMF)** of $Bi(10, 0.5)$ can be computed by using the dbinom function

```
> hist(x, probability = TRUE, ylim = c(0, 0.3))
> points(0:n, dbinom(0:n, n, p), type = "h", lwd = 3, col = "darkblue")
> points(0:n, dbinom(0:n, n, p), type = "p", lwd = 3, col = "darkblue")
```

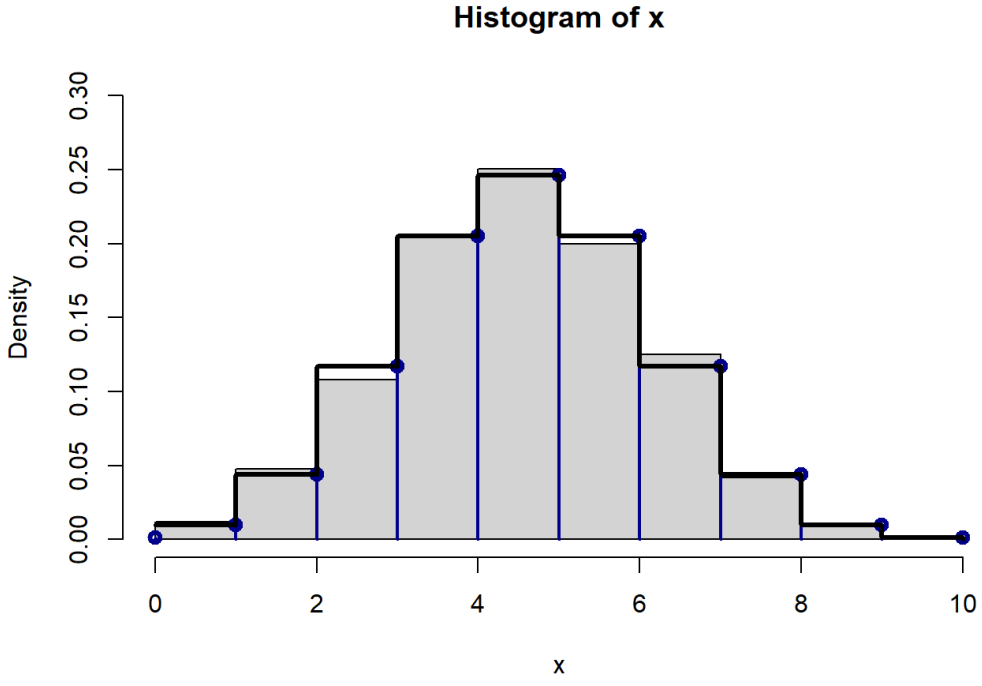



```
> hist(x, probability = TRUE, xlim = c(0, n), ylim = c(0, 0.3))
> points(0:n, dbinom(0:n, n, p), type = "h", lwd = 2, col = "darkblue")
> points(0:n, dbinom(0:n, n, p), type = "p", lwd = 3, col = "darkblue")
> points(0:n, dbinom(0:n + 1, n, p), type = "s", lwd = 3)
```



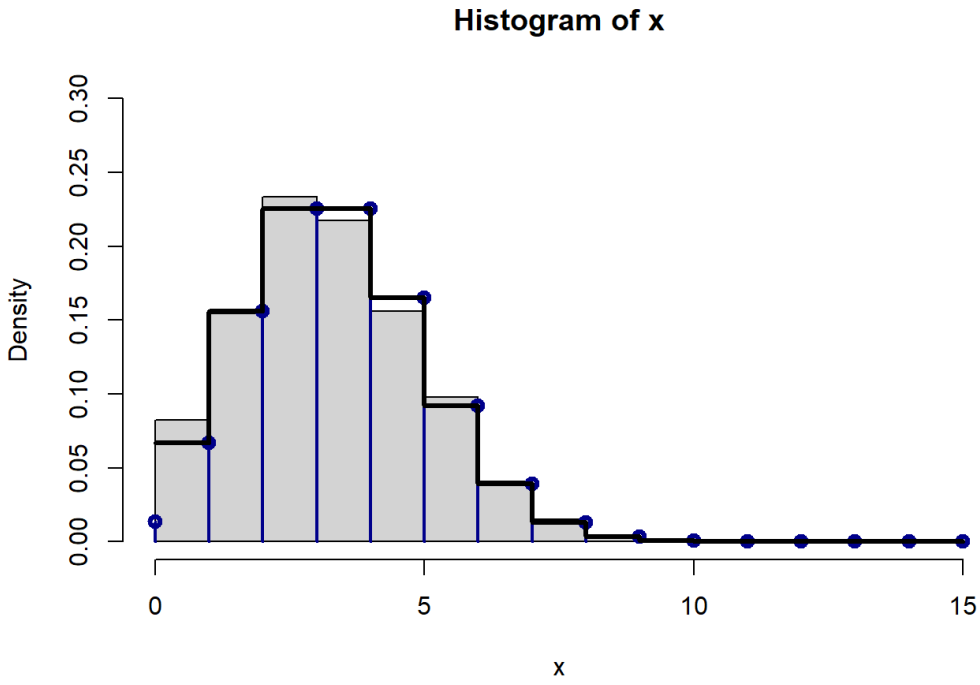
Now let's generate 5000 realizations of $Bi(10, 0.5)$ random variable.

```
> x <- rbinom(5000, n, p)
> hist(x, probability = TRUE, xlim = c(0, n), ylim = c(0, 0.3))
> points(0:n, dbinom(0:n, n, p), type = "h", lwd = 2, col = "darkblue")
> points(0:n, dbinom(0:n, n, p), type = "p", lwd = 3, col = "darkblue")
> points(0:n, dbinom(0:n + 1, n, p), type = "s", lwd = 3)
```



Let's generate 5000 numbers, that are the number of successes in 15 experiments each with probability for success 0.25.

```
> n <- 15; p <- 0.25
> x <- rbinom(5000, 15, 0.25)
> table(x)
x
   0    1    2    3    4    5    6    7    8    9   10   11
  71  340  771 1168 1086  781  489  203   74   13    3    1
> hist(x, probability = TRUE, xlim = c(0, n), ylim = c(0, 0.3))
> points(0:n, dbinom(0:n, n, p), type = "h", lwd = 2, col = "darkblue")
> points(0:n, dbinom(0:n, n, p), type = "p", lwd = 3, col = "darkblue")
> points(0:n, dbinom(0:n + 1, n, p), type = "s", lwd = 3)
```



Now let's see how many of the observations are less than or equal to 3.

```
> sum(x <= 3)
[1] 2350
```

What is the proportion of this observations in the sample?

```
> sum(x <= 3) / length(x)
[1] 0.47
```

Knowing that the random variable follows $X \in Bi(15, 0.25)$ distribution we can also see the theoretical probability to have an observation less than or equal to 3 using the pbinom function.

$$\mathbb{P}(X \leq 3)$$

```
> pbinom(q = 3, size = 15, prob = 0.25)
[1] 0.4612869
```

Is this number similar to the number observed above?

We can also see the theoretical probability to have an observation greater than 3.

$$\mathbb{P}(X > 3)$$

```
> pbinom(q = 3, size = 15, prob = 0.25, lower.tail = FALSE)
[1] 0.5387131
```

or also we can calculate it by using $1 - \mathbb{P}(X \leq 3)$

```
> 1 - pbinom(q = 3, size = 15, prob = 0.25)
[1] 0.5387131
```

The probability $\mathbb{P}(X \in (3,5]) = \mathbb{P}(X \leq 5) - \mathbb{P}(X \leq 3)$ can be computed by

```
> pbinom(q = 5, size = 15, prob = 0.25) - pbinom(q = 3, size = 15, prob = 0.25)
[1] 0.390345
```

The probability $\mathbb{P}(X \in [3,5]) = \mathbb{P}(X \leq 5) - \mathbb{P}(X \leq 2)$ can be computed by

```
> pbinom(q = 5, size = 15, prob = 0.25) - pbinom(q = 2, size = 15, prob = 0.25)
[1] 0.6155441
```

or by using $\mathbb{P}(X \in [3,5]) = \mathbb{P}(X = 3) + \mathbb{P}(X = 4) + \mathbb{P}(X = 5)$

```
> sum(dbinom(x = c(3,4,5), size = 15, prob = 0.25))
[1] 0.6155441
```

By using qbinom function we can compute 0.4 quantile of X. It is the smallest number when the CDF of X is bigger than or equal to 0.4.

$$X_{0.4} = \min\{x \in \mathbb{R} : \mathbb{P}(X \leq x \geq 0.4)$$

```
> qbinom(p = 0.4, size = 15, prob = 0.25)
[1] 3
```

and also the quantile from which we have 0.4 probability to be higher

```
> qbinom(p = 0.4, size = 15, prob = 0.25, lower.tail = FALSE)
[1] 4
```

Geometric distribution /Геометрично разпределение/

There are two definitions for the geometric distribution.

Let's say we are making independent repetitions of an experiment. The first definition of **geometric distribution** with parameter p represents probability distribution of the

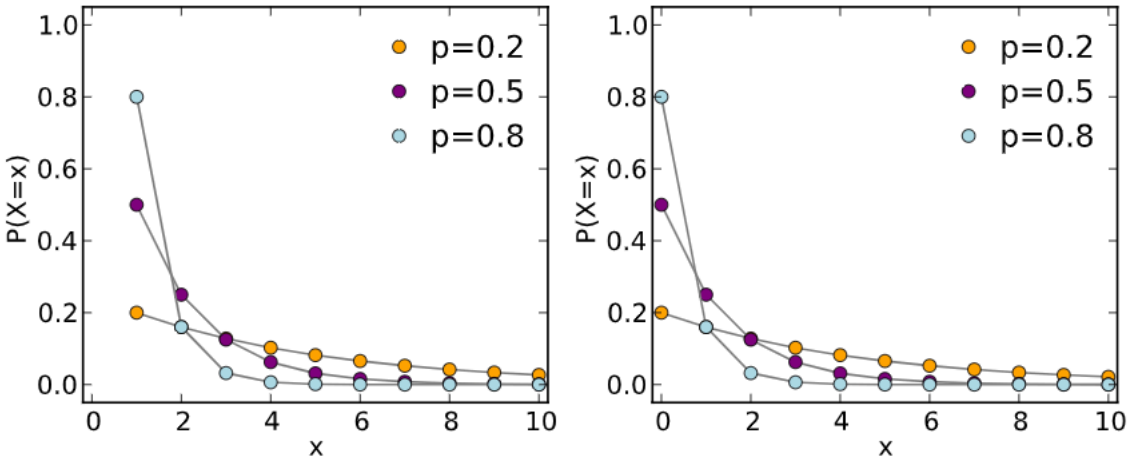
number of the first success. The second definition represents probability distribution of the number of failures before the first success. We are going to use the second one.

$$X \in Ge(p)$$

Geometric distribution is a special case of Negative binomial distribution $NegBi(1,p)$.

Probability mass function (PMF)

$$\mathbb{P}_X(k) = \mathbb{P}(X = k) = q^k p, k = 0, 1, \ldots$$



Mean

$$\mu = \mathbb{E}[X] = \int_{-\infty}^{+\infty} x \, \mathrm{d}\, F_X(x) = \sum_{i=0}^{\infty} i\mathbb{P}(X = i) =$$

$$= \sum_{i=0}^{\infty} iq^i p = pq \sum_{i=0}^{\infty} iq^{i-1} =$$

$$= pq \frac{\partial}{\partial q} \left(\sum_{i=0}^{\infty} q^i \right) = pq \frac{\partial}{\partial q} \left(\frac{1}{1-q} \right) =$$

$$= pq \frac{-1}{(1-q)^2}(-1) = \frac{pq}{p^2} = \frac{q}{p}$$

Variance

$$\sigma^2 = \frac{q}{p^2}$$

Similarly as above we can use the functions dgeom, rgeom, pgeom and qgeom to find the theoretical probability mass function, to generate random numbers, to find the cumulative distribution function and quantiles of geometrically distributed random variable.

Let's generate a realization of $Geom(0.2)$ random variable.

```
> rgeom(1, 0.2)
[1] 5
```

This number represents the number of failures before the first success.

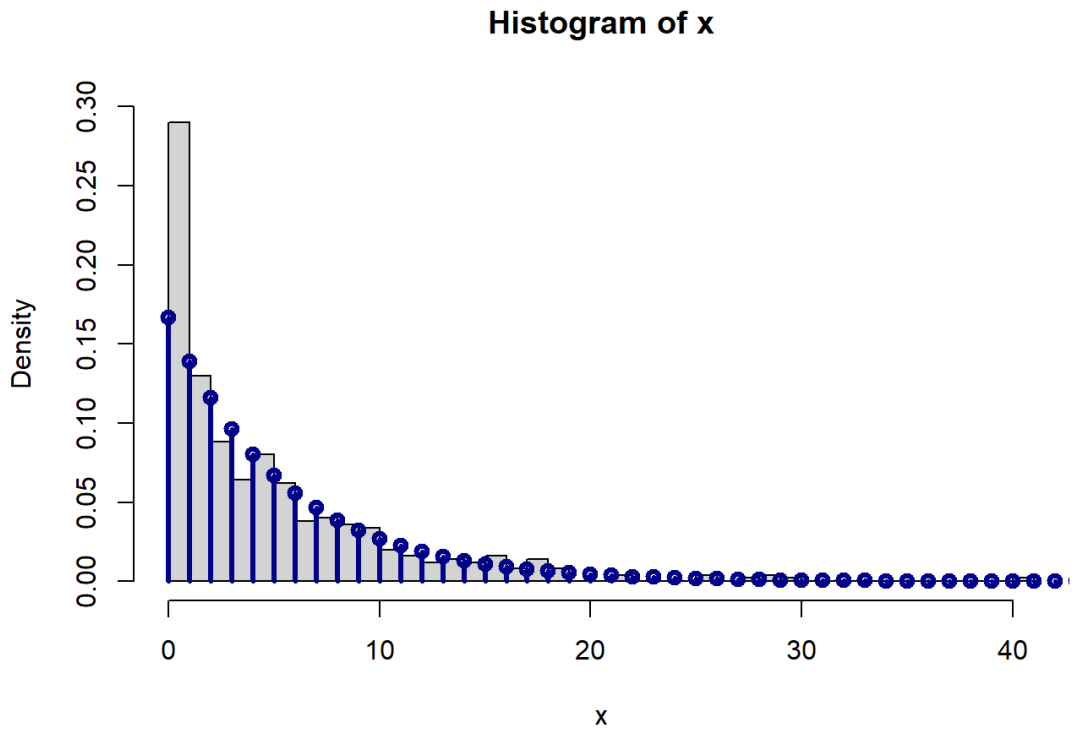
Let's generate 10 realizations of $Geom(0.2)$ random variable.

```
> rgeom(10, 0.2)
 [1]  2 19  2  0  1  2 15  1 10  0
```

Example

Let's generate 500 random numbers representing the number of failures before the first row of 6 on a regular die.

```
> p <- 1/6
> x <- rgeom(500, p)
> hist(x, breaks = 30, probability = TRUE, ylim = c(0, 0.3))
> points(0:100, dgeom(0:100,p), type = "h", lwd = 3, col = "darkblue")
> points(0:100, dgeom(0:100,p), type = "p", lwd = 3, col = "darkblue")
```



Now let's see how many time the 6 is observed before the 3rd row of the die.

```
> sum(x <= 2)
[1] 210
```

What is the proportion of this outcomes in the sample?

```
> sum(x <= 2) / length(x)
[1] 0.42
```

Knowing that the random variable follows $X \in Geom(0.2)$ distribution we can also see the theoretical probability for the first time to row 6 before the 3rd row of the die using the pgeom function. $\mathbb{P}(X \leq 2)$

```
> pgeom(q = 2, p)
[1] 0.4212963
```

We can also see the probability to row 6 for the first time between the 6th and 3rd row of the die.

```
> pgeom(q = 5, p) - pgeom(q = 2, p)
[1] 0.2438057
```

Or we can see the theoretical quantile from which we have 0.5 probability to have 6 for the first time before that row.

```
> qgeom(p = 0.5, p)
[1] 3
```

Negative binomial distribution /Отрицателно биномно разпределение/

Let's say we are making independent repetitions of an experiment, each with probability for success p. **Negative binomial distribution** with parameters n and p represents probability distribution of the number of failures before the n-th success occur.

$$X \in NegBi(n, p)$$

Probability mass function (PMF)

$$\mathbb{P}_X(k) = \mathbb{P}(X = k) = \binom{n + k - 1}{k} q^k p^n, \quad k = 0, 1, \ldots$$

X can be represented as sum of n independent geometrically distributed random variables $Y_i \in Ge(p), i = 1, \ldots, n$

$$X = Y_1 + Y_2 + \ldots + Y_n$$

Mean

$$\mu = \mathbb{E}[X] = \mathbb{E}[Y_1 + Y_2 + \ldots + Y_n] =$$
$$= \mathbb{E}[Y_1] + \mathbb{E}[Y_2] + \ldots + \mathbb{E}[Y_n] = \frac{q}{p} + \ldots + \frac{q}{p} = n\frac{q}{p}$$

Variance

$$\sigma^2 = \mathbb{D}[Y_1 + Y_2 + \ldots + Y_n] =$$

$$= \mathbb{D}[Y_1] + D[Y_2] + \ldots + D[Y_n] = \frac{q}{p^2} + \ldots + \frac{q}{p^2} = n\frac{q}{p^2}$$

As far as the geometric random variables are independent

Similarly as above we can use the functions dnbinom, rnbinom, pnbinom and qnbinom to find the theoretical probability mass function, to generate random numbers, to find the cumulative distribution function and quantiles of negative binomial distributed variable.

Let's generate a realization of $NegBi(5, 0.2)$ representing the number of failures before the 5th success, when the probability for success in every one of the experiments is 0.2.

```
> rnbinom(1, 5, 0.2)
[1] 21
```

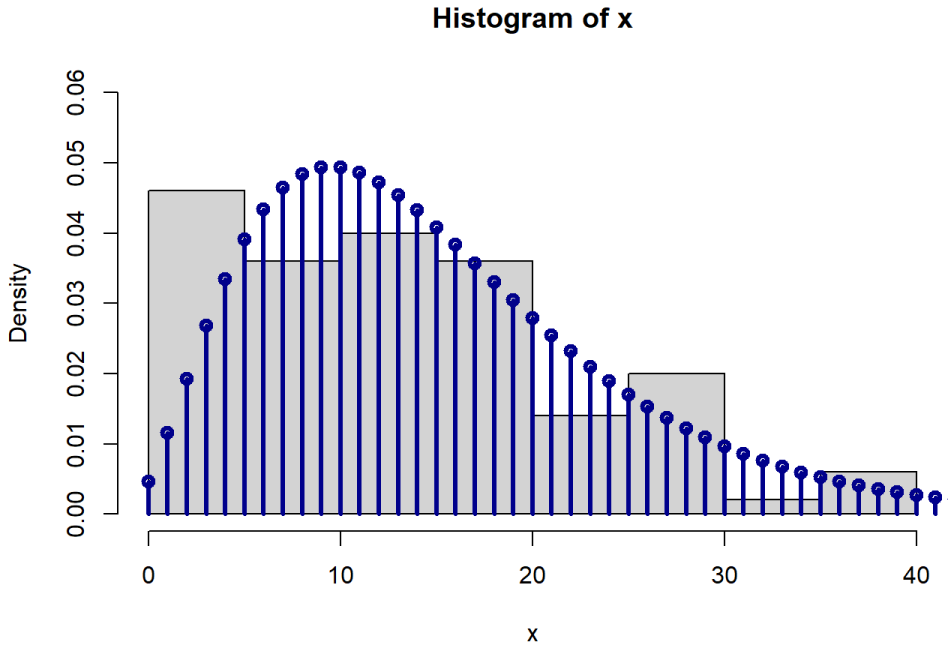
Let's generate 10 realizations of $NegBi(5, 0.2)$ random variable.

```
> rnbinom(10, 5, 0.2)
 [1] 29 13 19 11 28 26 13 18  6 28
```

Example

Let's have a regular die and say rolling of 6 is a success and event X models the number of failures before the 3rd row of 6.

```
> n <- 3; p <- 1/6
> x = rnbinom(100, n, p)
> hist(x, probability = TRUE, ylim = c(0, 0.06))
> points(0:100, dnbinom(0:100, n, p), type = "h", lwd = 3, col = "darkblue")
> points(0:100, dnbinom(0:100, n, p), type = "p", lwd = 3, col = "darkblue")
```

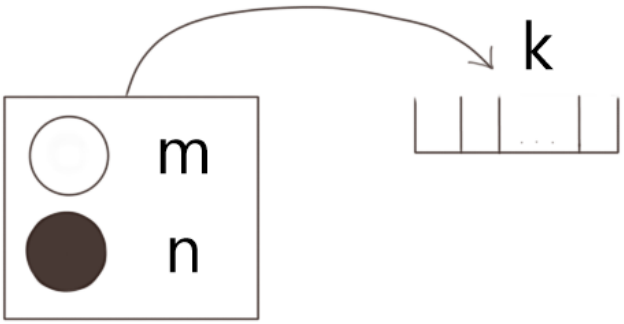


```
> pnbinom(5, n, p)
[1] 0.1348469
> pnbinom(10, n, p) - pnbinom(4, n, p)
[1] 0.2761473
> pnbinom(0.5, n, p)
[1] 0.00462963
```

Hypergeometric distribution /Хипергеометрично разпределение/

Hypergeometric distribution describes the number X of successes in k draws, **without replacement (dependent examples)**, from a finite population which contains exactly m objects with that feature and n objects without that feature, wherein each draw is either a success or a failure.

$$X \in HG(m, n, k)$$



Probability mass function (PMF)

$$\mathbb{P}_X(s) = \mathbb{P}(X = s) = \frac{\binom{m}{s}\binom{n}{k-s}}{\binom{m+n}{k}}, \qquad \max(0, k-n) \leq s \leq \min(k, m)$$

X can be represented as sum of k dependent indicators of events with probability $\mathbb{P}(A_i) = \frac{m}{m+n}, i = 1, \ldots, k.$

$$X = I_{A_1} + I_{A_2} + \ldots + I_{A_k}$$

Mean

$$\mu = \mathbb{E}[X] = \mathbb{E}[I_{A_1} + I_{A_2} + \ldots + I_{A_k}] =$$
$$= \mathbb{E}[I_{A_1}] + \mathbb{E}[I_{A_2}] + \ldots + \mathbb{E}[I_{A_k}] = \frac{m}{m+n} + \ldots + \frac{m}{m+n} = k\frac{m}{m+n}$$

Variance

$$\sigma^2 = \frac{kmn(m+n-k)}{(m+n)^2(m+n-1)}$$

Similarly as above we can use the functions dhyper, rhyper, phyper and qhyper to find the theoretical probability mass function, to generate random numbers, to find the cumulative distribution function and quantiles of hypergeometrically distributed variable.

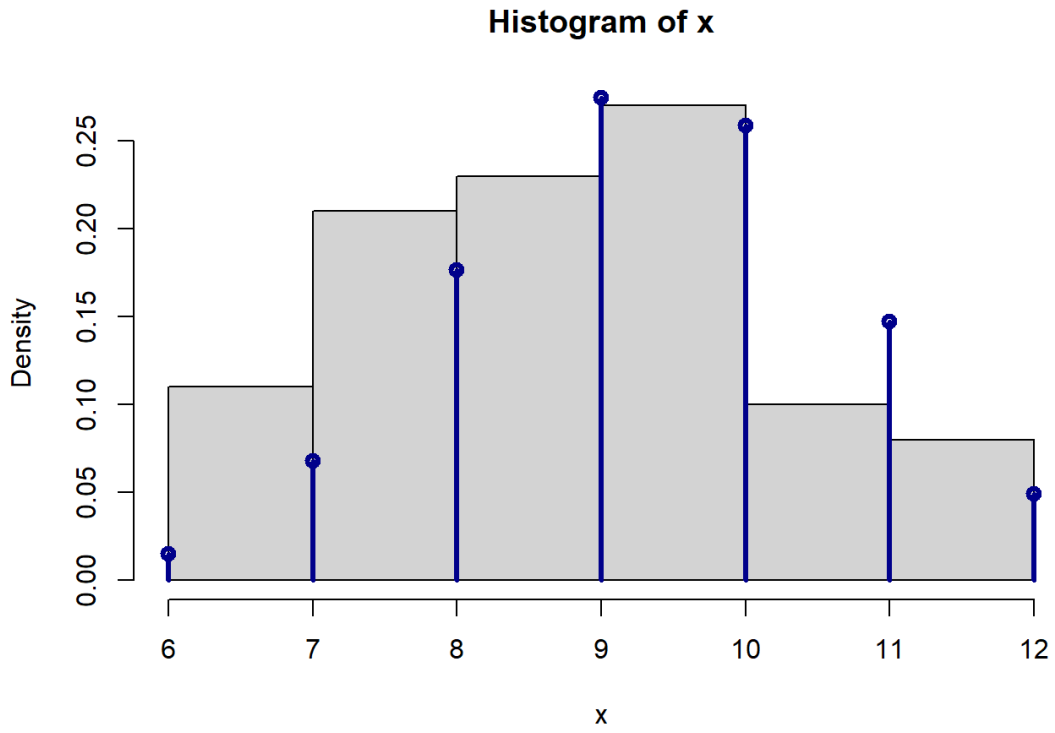
Example

Let's generate a realization of $HG(10,12,15)$ representing the number of white balls drawn from a box, if in the beginning of the experiment in the box there were 10 white and 12 black balls and we have drawn 15 balls without turning them back to the box. So, the probability to take white ball has changed after every drawn.

```
> rhyper(1, m = 10, n = 12, k = 15)
[1] 7
```

Let's generate 10 realizations of $HG(10,12,15)$ random variable.

```
> rhyper(10, m = 10, n = 12, k = 15)
 [1] 6 6 8 6 6 5 7 7 8 6
> m <- 20; n <- 12; k <- 15
> x <- rhyper(100, m, n, k)
> hist(x, probability = TRUE)
> points(0:100, dhyper(0:100, m, n, k), type = "h", lwd = 3, col = "darkblue")
> points(0:100, dhyper(0:100, m, n, k), type = "p", lwd = 3, col = "darkblue")
```



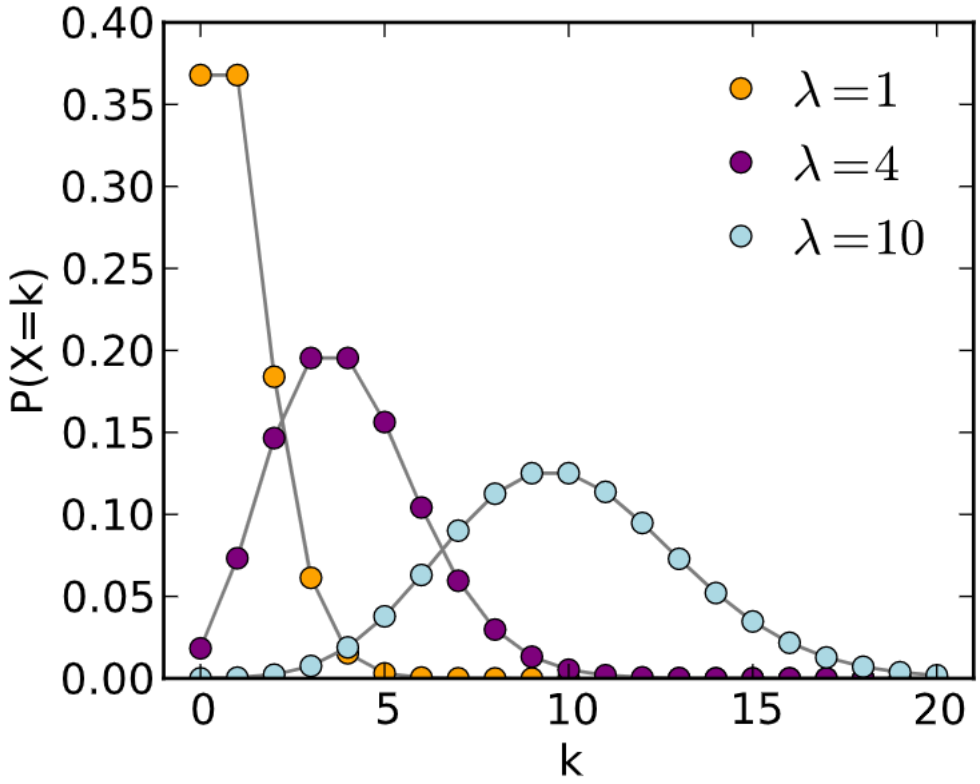
Poisson distribution /Поасоново разпределение/

Named after French mathematician **Siméon Denis Poisson**. **Poisson distribution** expresses the probability of a given number of events occurring in a fixed interval of time or space if these events occur with a known constant rate λ and independently of the time(space) since the last event.

$$X \in Po(\lambda), \text{ where } \lambda \in (0, +\infty)$$

Probability mass function (PMF)

$$\mathbb{P}_X(k) = \mathbb{P}(X = k) = \frac{\lambda^k e^{-\lambda}}{k!}, k = 0, 1, \ldots$$



Mean

$$\mu = \mathbb{E}[X] = \int_{-\infty}^{+\infty} x \, \mathrm{d}\, F_X(x) = \sum_{i=0}^{\infty} i\mathbb{P}(X = i) =$$

$$= \sum_{i=0}^{\infty} i\frac{\lambda^i}{i!}e^{-\lambda} = \sum_{i=1}^{\infty} \frac{\lambda^i}{(i-1)!}e^{-\lambda} =$$

$$= \lambda e^{-\lambda} \sum_{i=1}^{\infty} \frac{\lambda^{i-1}}{(i-1)!} = \lambda e^{-\lambda} \sum_{s=0}^{\infty} \frac{\lambda^s}{s!} = \text{(from Maclaurin series for } e^x) = \lambda e^{-\lambda} e^{\lambda} = \lambda$$

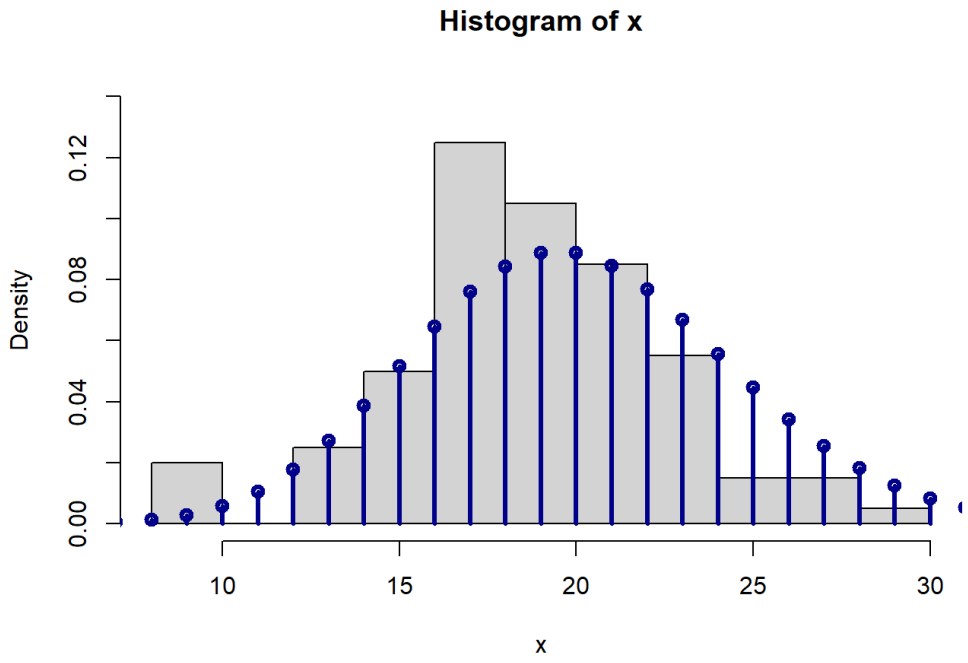
Variance

$$\sigma^2 = \lambda$$

Similarly as above we can use the functions dpois, rpois, ppois and qpois find the theoretical probability mass function, to generate random numbers, to find the cumulative distribution function and quantiles of Poisson distributed variable.

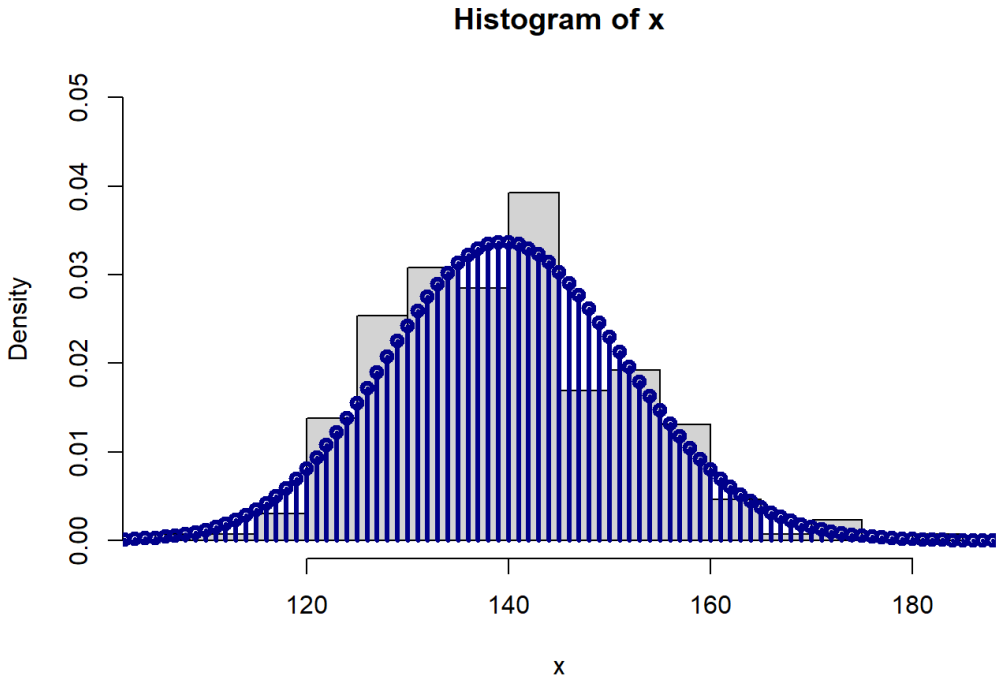
Example 1
The number of cars arriving at a car wash per day is Poisson distributed with mean 20. Generate the numbers of cars arriving at this car wash daily, for 100 days. $Po(\lambda = 20)$

```
> lambda <- 20
> x <- rpois(100, lambda)
> hist(x, probability = TRUE, ylim = c(0, 0.14))
> points(0:100, dpois(0:100, lambda), type = "h", lwd = 3, col = "darkblue")
> points(0:100, dpois(0:100, lambda), type = "p", lwd = 3, col = "darkblue")
```



Generate how many cars are going to arrive weekly for 260 weeks?

```
> x <- rpois(260, 7*lambda)
> hist(x, probability = TRUE, breaks = 15, ylim = c(0, 0.05))
> points(0:200, dpois(0:200, 7*lambda), type = "h", lwd = 3, col = "darkblue")
> points(0:200, dpois(0:200, 7*lambda), type = "p", lwd = 3, col = "darkblue")
```



Example 2
The average number of homes sold by the Home2U company is 20 homes per day. What is the probability that exactly 21 homes will be sold tomorrow?

Calculate it using probability mass function: $X \in Po(\lambda = 21)$

$$\mathbb{P}_X(21) = \mathbb{P}(X = 21) = \frac{20^{21}e^{-20}}{21!} \approx 0.0846$$

Calculate it by using R:

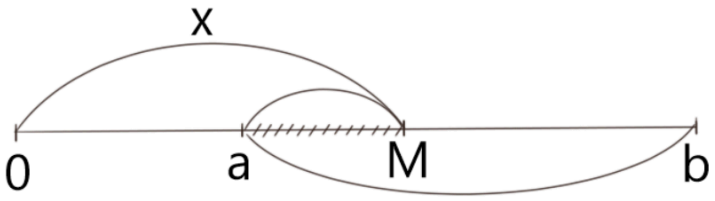
```
> lambda <- 20
> dpois(21, lambda)
[1] 0.08460506
```

Continuous Distributions

Continuous uniform distribution /Непрекъснато равномерно разпределение/

$X \in U(a, b)$ or similarly $X \in Unif(a, b)$

If we choose a point M at random in the interval between a and b and X is the length of the interval between M and a by geometric definition of probability for $x \in [a, b]$

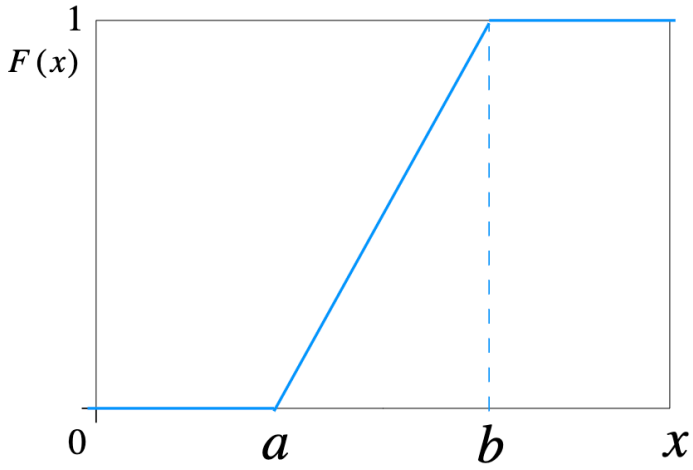


$$\mathbb{P}(X \leq x) = \frac{x-a}{b-a}$$

therefore

Cumulative distribution function (CDF) is

$$F_X(x) = \begin{cases} 0, & x < a \\ \frac{x-a}{b-a}, & x \in [a,b] \\ 1, & x > b \end{cases}$$

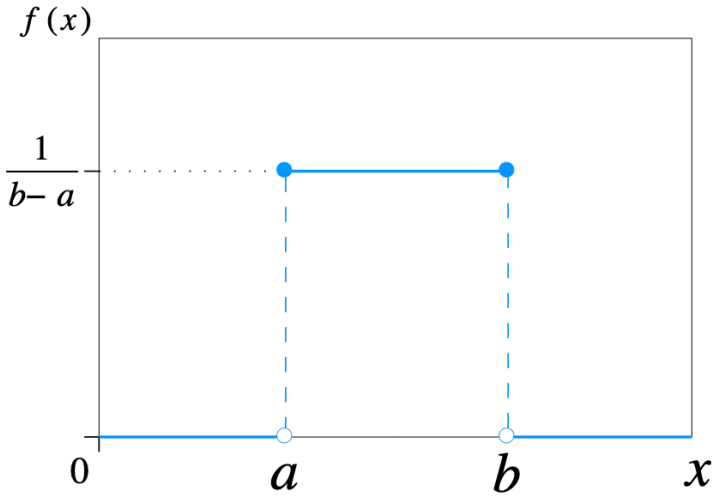


For $p \in (0,1)$ the corresponding

Quantile function is $X_p = F_{\overleftarrow{X}}(p) = \min\{x \in \mathbb{R} : \mathbb{P}(X \leq x) \geq p\} = p(b-a) + a$

As far as the cumulative distribution function (CDF) is differentiable in (a,b)
Probability density function (PDF) is

$$f_X(x) = \begin{cases} \frac{1}{b-a}, & x \in [a,b] \\ 0, & \text{otherwise} \end{cases}$$



Mean /First initial moment/

$$\mu = \mathbb{E}[X] = \int_{-\infty}^{+\infty} x \,\mathrm{d}\, F_X(x) = \int_{-\infty}^{+\infty} x f_X(x) \mathrm{d}\, x =$$

$$= \int_{-\infty}^{a} x f_X(x) \mathrm{d}\, x + \int_{a}^{b} x f_X(x) \mathrm{d}\, x + \int_{b}^{+\infty} x f_X(x) \mathrm{d}\, x =$$

$$= \int_{-\infty}^{a} x \times 0 \,\mathrm{d}\, x + \int_{a}^{b} x \frac{1}{b-a} \,\mathrm{d}\, x + \int_{b}^{+\infty} x \times 0 \,\mathrm{d}\, x =$$

$$= \frac{1}{b-a} \int_{a}^{b} x \,\mathrm{d}\, x = \frac{1}{b-a} \frac{x^2}{2} \Big|_a^b = \frac{1}{b-a} \left(\frac{b^2 - a^2}{2} \right) =$$

$$= \frac{a+b}{2}$$

Second initial moment

$$\mathbb{E}[X^2] = \int_{-\infty}^{+\infty} x^2 \,\mathrm{d}\, F_X(x) = \int_{-\infty}^{+\infty} x^2 f_X(x) \mathrm{d}\, x =$$

$$= \int_{-\infty}^{a} x^2 f_X(x) \mathrm{d}\, x + \int_{a}^{b} x^2 f_X(x) \mathrm{d}\, x + \int_{b}^{+\infty} x^2 f_X(x) \mathrm{d}\, x =$$

$$= \int_{-\infty}^{a} x^2 \times 0 \,\mathrm{d}\, x + \int_{a}^{b} x^2 \frac{1}{b-a} \,\mathrm{d}\, x + \int_{b}^{+\infty} x^2 \times 0 \,\mathrm{d}\, x =$$

$$= \frac{1}{b-a} \int_{a}^{b} x^2 \,\mathrm{d}\, x = \frac{1}{b-a} \frac{x^3}{3} \Big|_a^b = \frac{1}{b-a} \left(\frac{b^3 - a^3}{3} \right) =$$

$$= \frac{a^2 + ab + b^2}{3}$$

Variance

$$\sigma^2 = \mathbb{D}[X] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 =$$
$$= \frac{a^2 + ab + b^2}{3} - \frac{(a+b)^2}{4} =$$
$$= \frac{(b-a)^2}{12}$$

Similarly as above we can use the functions dunif, runif, punif and qunif to find the theoretical probability density function, to generate random numbers, to find the cumulative distribution function and quantiles of continuous uniform distributed variable.
Let's generate a realization of continuous $Unif(0,2)$ random variable.

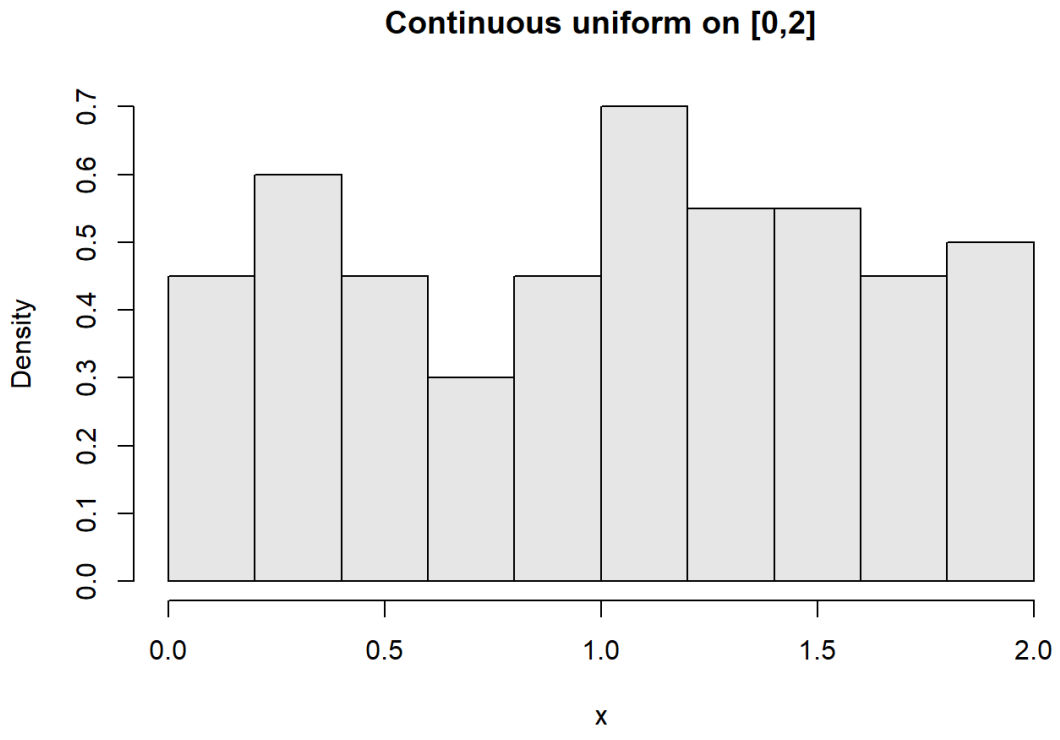
```
> runif(1, min = 0, max = 2)
[1] 1.234092
```

Let's generate 5 realizations of continuous $Unif(0,2)$ random variable.

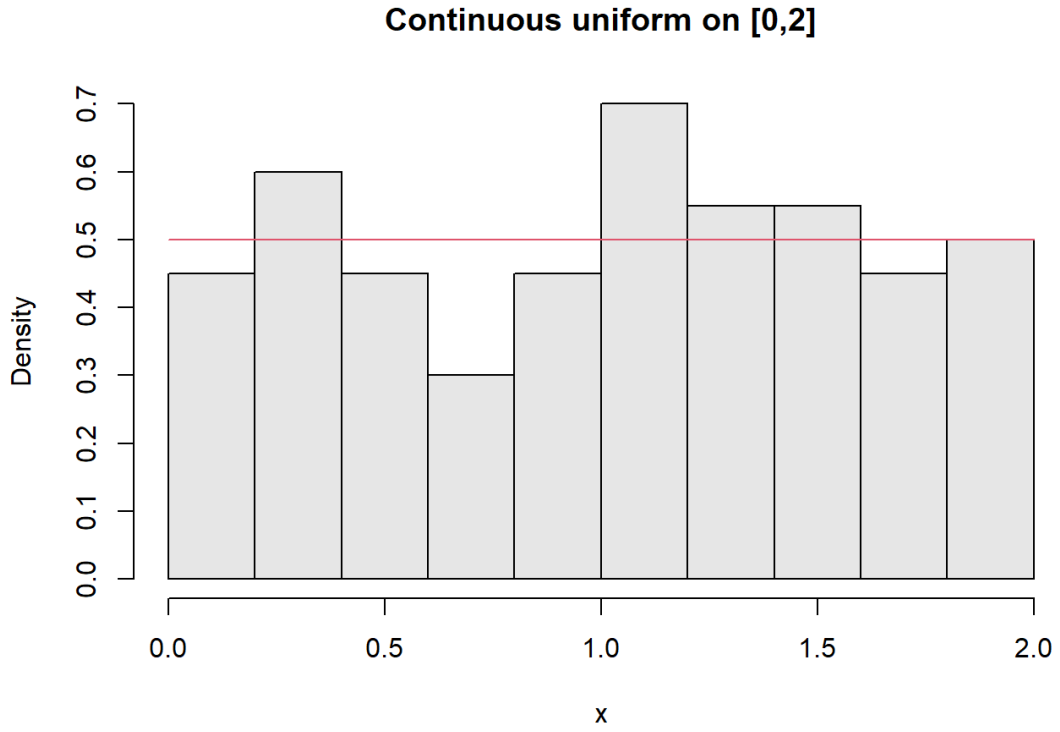
```
> runif(5, min = 0, max = 2)
[1] 1.3361018 0.6545223 0.5515193 0.8480260 0.6518157
```

Let's generate 100 realizations of continuous $Unif(0,2)$ random variable.

```
> x <- runif(100, min = 0, max = 2)
> hist(x, probability = TRUE,
+      main = "Continuous uniform on [0,2]",
+      col = gray(.9))
```



```
> hist(x, probability = TRUE,
+      main = "Continuous uniform on [0,2]",
+      col = gray(.9))
> curve(dunif(x, min = 0, max = 2),
+       add = TRUE,
+       col = 2)
```



We can calculate the skewness and kurtosis of the sample using the skewness and kurtosis from EnvStats package.

```
> library(EnvStats)
Warning: package 'EnvStats' was built under R version 4.0.3

Attaching package: 'EnvStats'
The following objects are masked from 'package:stats':

    predict, predict.lm
The following object is masked from 'package:base':

    print.default

> skewness(x)
[1] -0.1370806
> kurtosis(x)
[1] -1.19603
```

As we can see the skewness is close to 0, because the uniform distribution is symmetric. As far as the kurtosis is negative, the distribution is platykurtic.

Knowing that the random variable follows $X \in Unif(0,1)$ distribution we can also see the theoretical probability to have an observation less than or equal to 0.2 using the punif function.

$$\mathbb{P}(X \leq 0.2)$$

```
> punif(0.2, min = 0, max = 2)
[1] 0.1
```

$$\mathbb{P}(X > 0.2)$$

```
> punif(0.2, min = 0, max = 2, lower.tail = FALSE)
[1] 0.9
```

By using the "qunif" function, for the same random variable X, we can compute the theoretical 0.6 quantile. It is the smallest number where the cumulative distribution function (CDF) is bigger than or equal to 0.6.

$$X_{0.6} = \min\{x \in \mathbb{R} : \mathbb{P}(X \leq x) \geq 0.6\} =$$
$$= p(b-a) + a = 0.6(2-0) + 0 = 1.2$$

```
> qunif(0.6, min = 0, max = 2)
[1] 1.2
```

Exponential distribution /Експоненциално разпределение/

Exponential distribution is the probability distribution of the time between events in a Poisson point process.

$$X \in Exp(\lambda)$$

Cumulative distribution function (CDF)

$$F_X(x) = 1 - e^{-\lambda x}, x > 0$$

For $p \in (0,1)$

$$1 - e^{-\lambda x} = p$$
$$1 - p = e^{-\lambda x}$$
$$\ln(1-p) = -\lambda x$$
$$x = -\frac{\ln(1-p)}{\lambda}$$

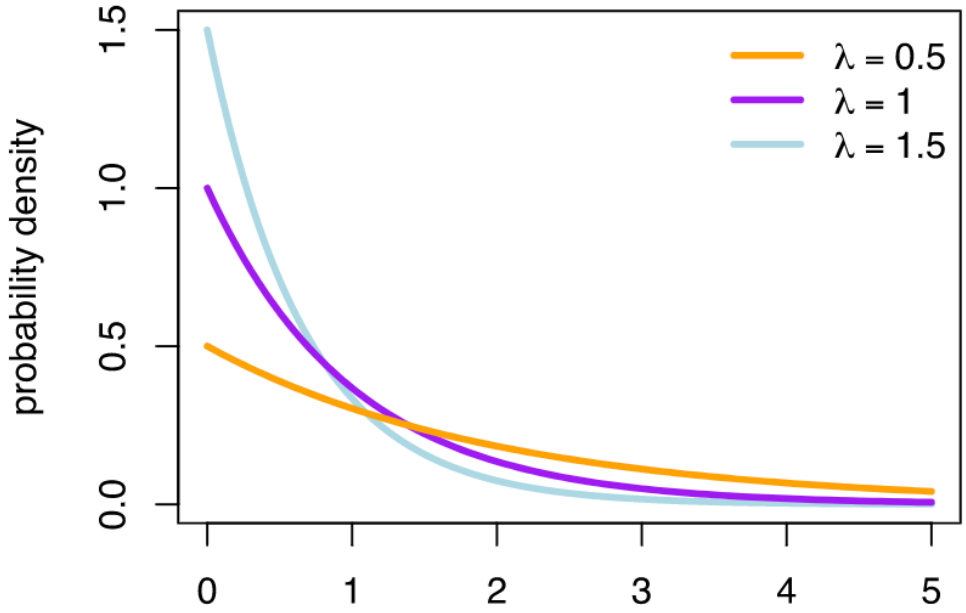
therefore

Quantile function

$$X_p = F_{\overleftarrow{X}}(p) = \min\{x \in \mathbb{R} : \mathbb{P}(X \leq x) \geq p\} = -\frac{\ln(1-p)}{\lambda}$$

Probability density function (PDF)

$$f_X(x) = \lambda e^{-\lambda x}, x > 0$$



Mean

$$\mu = \frac{1}{\lambda}$$

Variance

$$\sigma^2 = \frac{1}{\lambda^2}$$

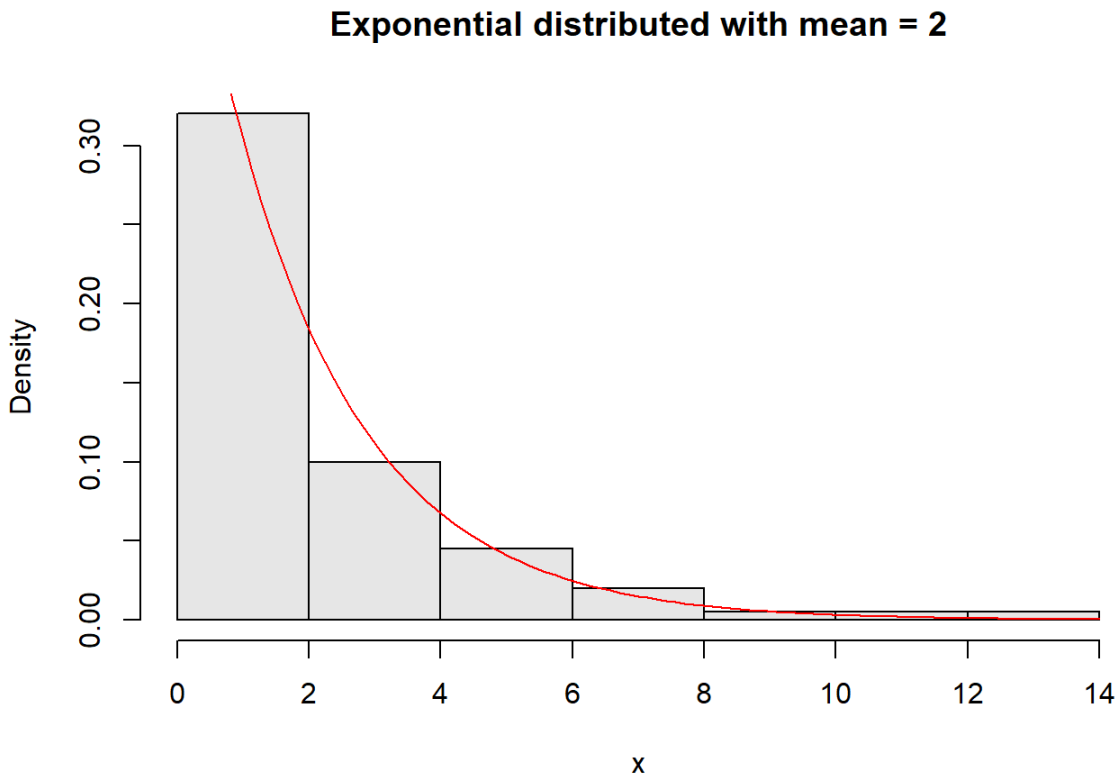
We can use the functions dexp, rexp, pexp and qexp to find the theoretical probability density function, to generate random numbers, to find the cumulative distribution function and quantiles of exponentially distributed random variable.

Example

The average life of a bulb is 2 years. Simulate the length of the life of 100 bulbs and summarize the results.

$$X \in Exp\left(\frac{1}{2}\right)$$

```
> x <- rexp(100, 1/2)
> summary(x)
    Min.  1st Qu.   Median     Mean  3rd Qu.     Max.
0.001296 0.535708 1.415715 2.083163 2.217798 12.183554
> hist(x, probability = TRUE, col = gray(.9), main = "Exponential distributed with mean = 2")
> curve(dexp(x, 1/2), add = TRUE, col = "red")
```



We can calculate the skewness and kurtosis of the sample.

```
> skewness(x)
[1] 2.276738
> kurtosis(x)
[1] 5.961678
```

As we can see the skewness is positive, because the exponential distribution is right-skewed. As far as the kurtosis is positive, the distribution is leptokurtic.
The theoretical probability a bulb to live less than or equal to 1.5 years could be found by using the pexp function.

$$\mathbb{P}(X \leq 1.5) = 1 - e^{-2\times 1.5} \approx 0.9502$$

```
> pexp(1.5, rate = 2)
[1] 0.9502129
```

A bulb to live more than 1.5 years

$$\mathbb{P}(X > 1.5)$$

```
> pexp(1.5, rate = 2, lower.tail = FALSE)
[1] 0.04978707
```

$$\mathbb{P}(X > 1.5) = 1 - \mathbb{P}(X \leq 1.5)$$

```
> 1 - pexp(1.5, rate = 2)
[1] 0.04978707
```

What is the 0.6 quantile of this distribution or with other words what is the length of the life that will be reached by 60% of the bulbs.

$$X_{0.6} = \min\{x \in \mathbb{R} : \mathbb{P}(X \leq x) \geq 0.6\} =$$
$$= -\frac{\ln(1-p)}{\lambda} = -\frac{\ln(1-0.6)}{2} \approx 0.4581$$

```
> qexp(0.6, rate = 2)
[1] 0.4581454
```

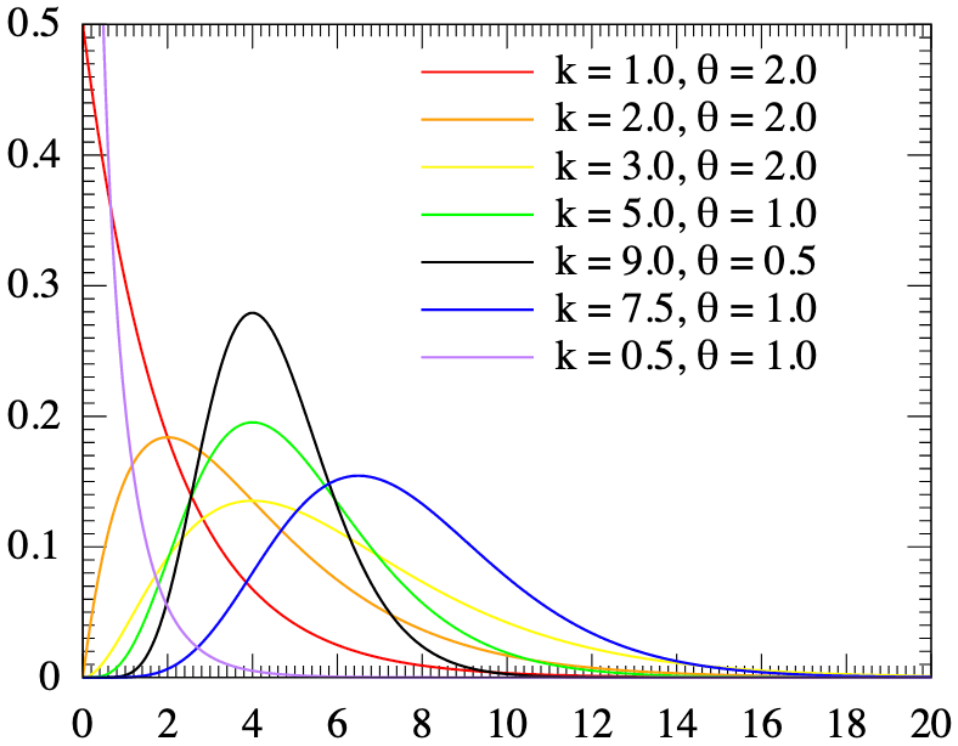
Gamma distribution /Гама разпределение/

Gamma distribution is a generalization of the exponential distribution.
$X \in \Gamma(\alpha, \beta)$

Probability density function (PDF) There are two parameterizations of **gamma distribution**. The first one is via the **rate parameter** $\beta > 0$. The second one is via the **scale parameter** $\theta = \frac{1}{\beta} > 0$.

$$f_X(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} = \frac{1}{\theta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-\frac{x}{\theta}}, x > 0, \text{ where}$$
$$\Gamma(\alpha) = \int_0^{+\infty} x^{\alpha-1} e^{-x} \, \mathrm{d}\, x, \Gamma(n) = (n-1)!$$

and $a > 0$ is called **shape parameter**.



Mean

$$\mu = \frac{\alpha}{\beta}$$

Variance

$$\sigma^2 = \frac{\alpha}{\beta^2}$$

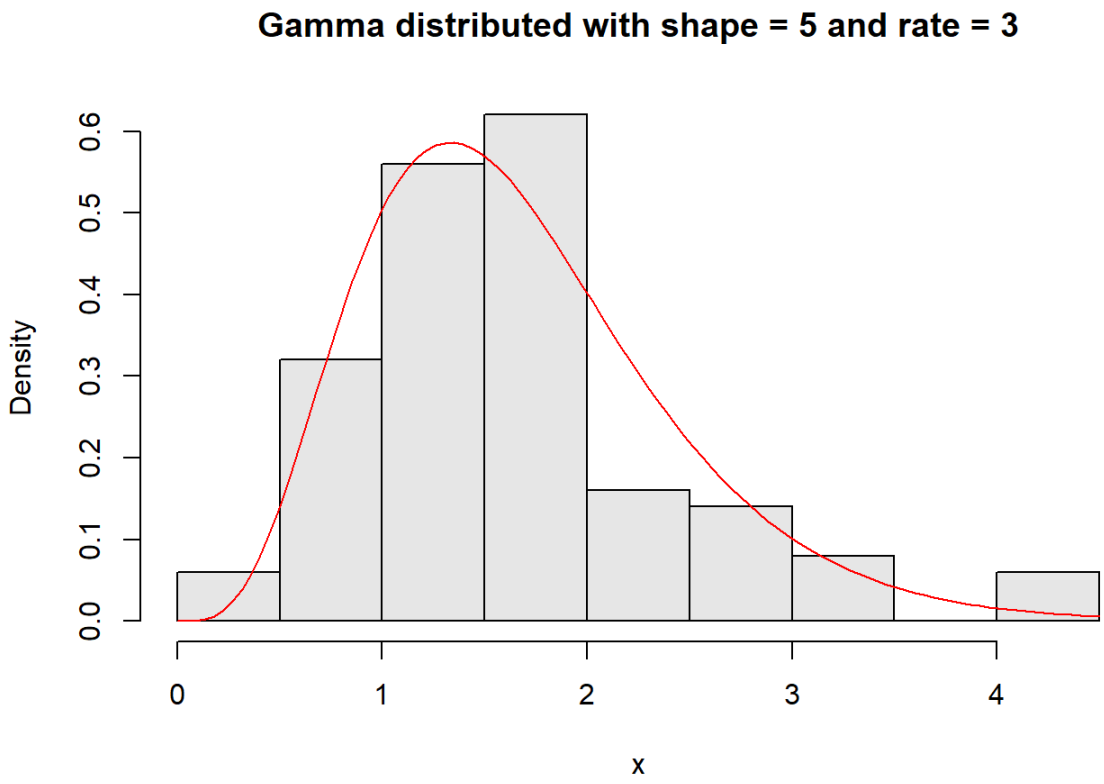
We can use the functions dgamma, rgamma, pgamma and qgamma to find the theoretical probability density function, to generate random numbers, to find the cumulative distribution function and quantiles of gamma distributed random variable.

Example

The average life in years of a bulb is gamma distributed with shape parameter 5 and rate parameter 3. Simulate the length of the life of 100 bulbs and summarize the results. Compute the expected length of the life and its variance.

$$X \in \Gamma(5,3)$$

```
> x <- rgamma(100, shape = 5, rate = 3)
> summary(x)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 0.3742  1.0860  1.5278  1.6234  1.8757  4.4014
> hist(x, probability = TRUE, col = gray(.9), main = "Gamma distributed with shape = 5
and rate = 3")
> curve(dgamma(x, shape = 5, rate = 3), add = TRUE, col = "red")
```



$$\mathbb{E}[X] = \frac{\alpha}{\beta} = \frac{5}{3} \approx 1.67$$
$$\mathbb{D}[X] = \frac{\alpha}{\beta^2} = \frac{5}{3^2} \approx 0.56$$

```
> var(x)
[1] 0.6956873
```

We can calculate the skewness and kurtosis of the sample.

```
> skewness(x)
[1] 1.193981
> kurtosis(x)
[1] 1.705705
```

As we can see the skewness is positive, because the gamma distribution $\Gamma(5,3)$ is right-skewed. As far as the kurtosis is positive, the distribution is leptokurtic.

The theoretical probability a bulb to live less than or equal to 1.5 years could be found by using the pgamma function.

$$\mathbb{P}(X \leq 1.5)$$

```
> pgamma(1.5, shape = 5, rate = 3)
[1] 0.4678964
```

A bulb to live more than 1.5 years

$$\mathbb{P}(X > 1.5)$$

```
> pgamma(1.5, shape = 5, rate = 3, lower.tail = FALSE)
[1] 0.5321036
```

$$\mathbb{P}(X > 1.5) = 1 - \mathbb{P}(X \leq 1.5)$$

```
> 1 - pgamma(1.5, shape = 5, rate = 3)
[1] 0.5321036
```

What is the 0.6 quantile of this distribution or with other words what is the length of the life that will be reached by 60% of the bulbs.

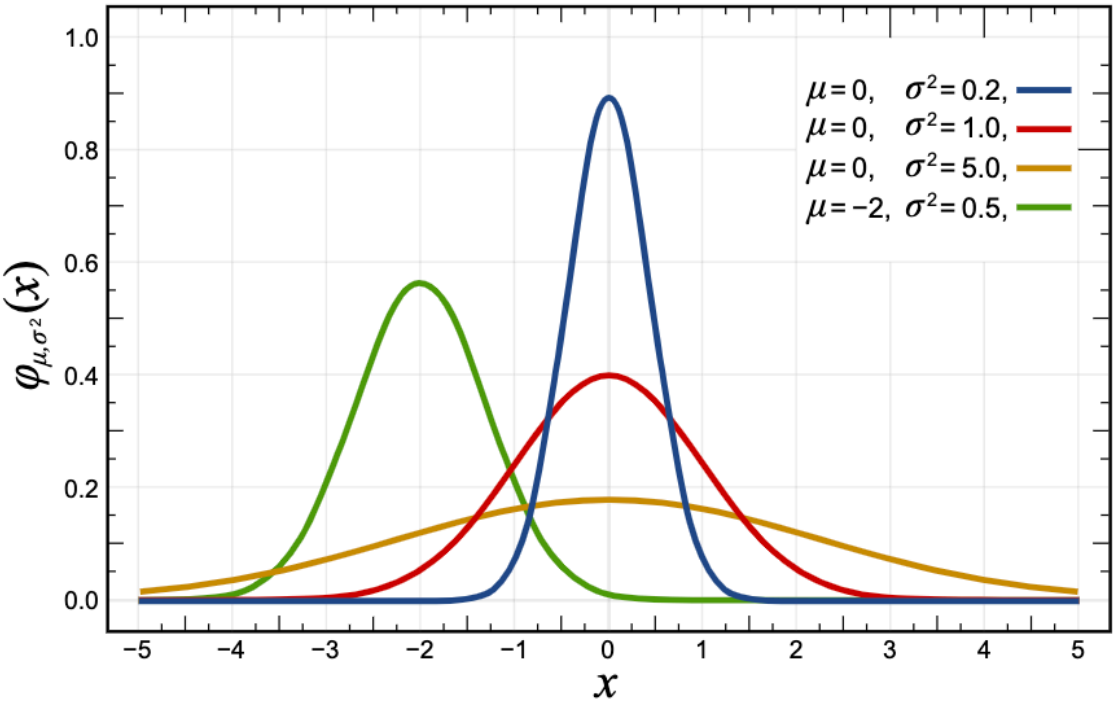
$$X_{0.6} = \min\{x \in \mathbb{R} : \mathbb{P}(X \leq x) \geq 0.6\}$$

```
> qgamma(0.6, shape = 5, rate = 3)
[1] 1.745539
```

Normal distribution /Нормално разпределение/

The normal distribution is defined by the **Probability density function (PMF)** equation

$$f_X(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, x \in \mathbb{R}$$



$X \in \mathcal{N}(\mu, \sigma^2)$ or similarly $X \in N(\mu, \sigma^2)$

Mean

$$\mu = \mathbb{E}[X]$$

Variance

$$\sigma^2 = \mathbb{D}[X]$$

We can use the functions dnorm, rnorm, pnorm and qnorm to find the theoretical probability density function, to generate random numbers, to find the cumulative distribution function and quantiles of normal distributed variable.
Let's generate a realization of $N(100,\ 16^2)$ random variable.

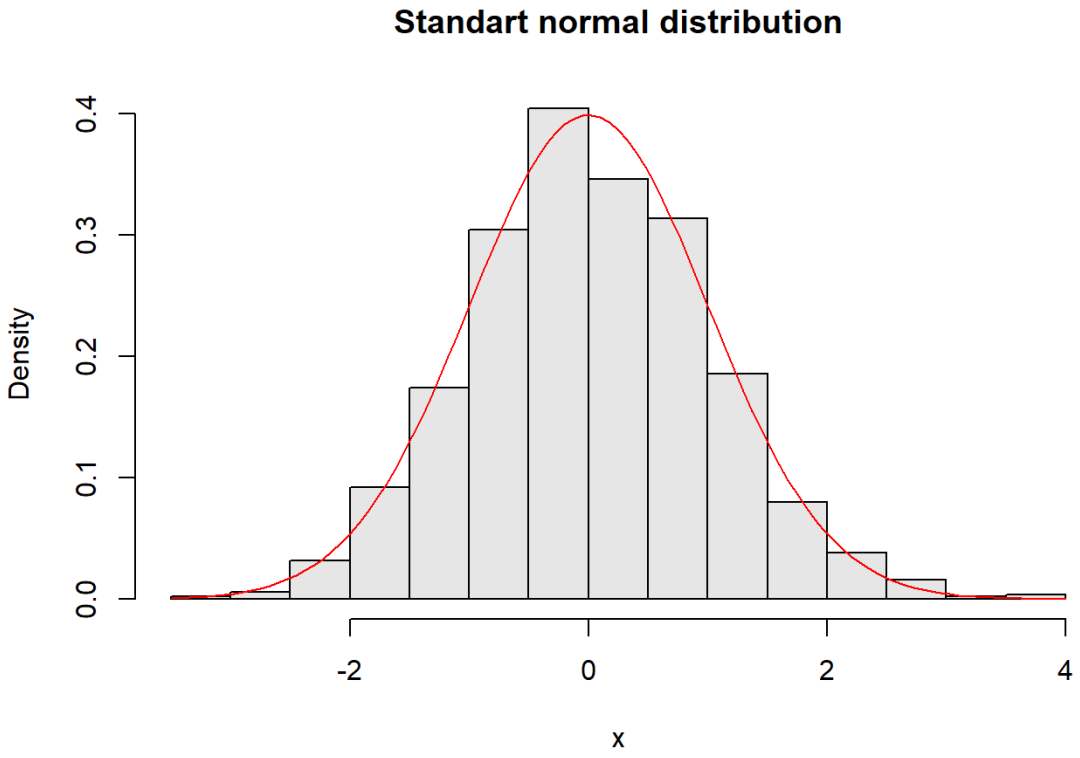
```
> rnorm(1, mean = 100, sd = 16)
[1] 82.76315
```

Let's generate a realization of $N(100,\ 10^2)$ random variable.

```
> rnorm(1, mean = 280, sd = 10)
[1] 287.8831
```

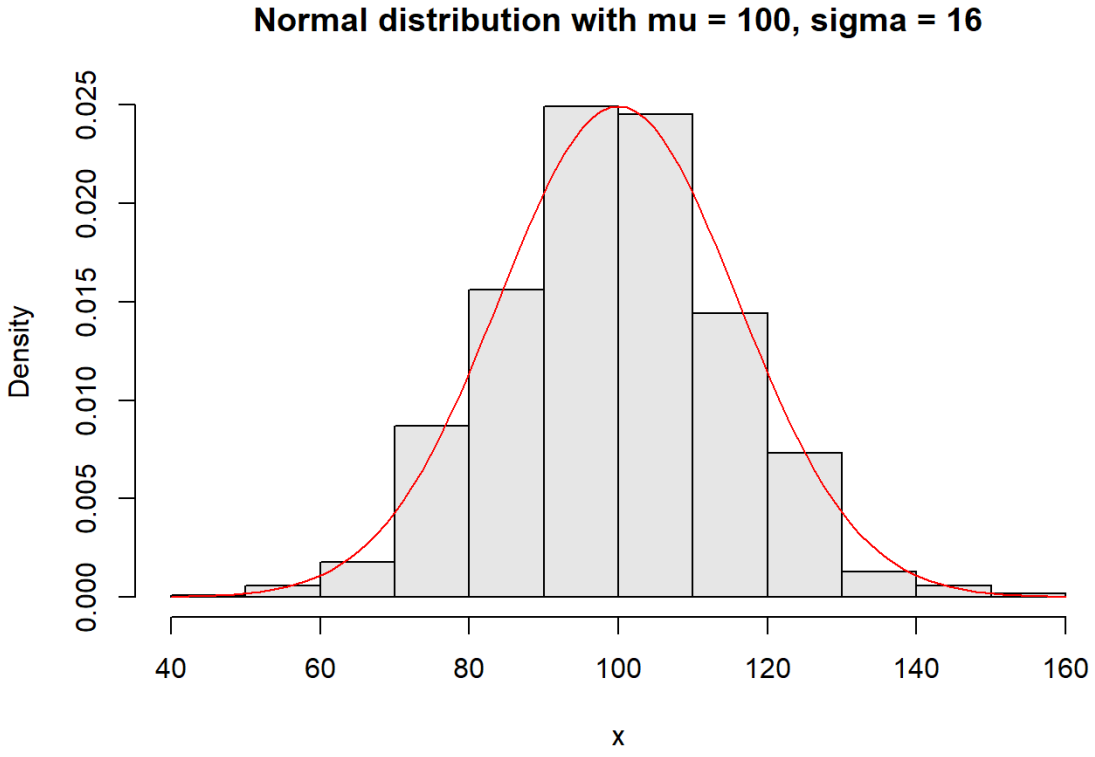
Let's generate 1000 realizations of $N(0,\ 1^2)$ random variable.

```
> x <- rnorm(1000)
> hist(x, probability = TRUE, col = gray(.9), main = "Standart normal distribution")
> curve(dnorm(x), add = TRUE, col = "red")
```



Let's generate 1000 realizations of $N(100,\ 16^2)$ random variable.

```
> x <- rnorm(1000, mean = 100, sd = 16)
> hist(x, probability = TRUE, col = gray(.9), main = "Normal distribution with mu = 100, sigma = 16")
> curve(dnorm(x, mean = 100, sd = 16), add = TRUE, col = "red")
```



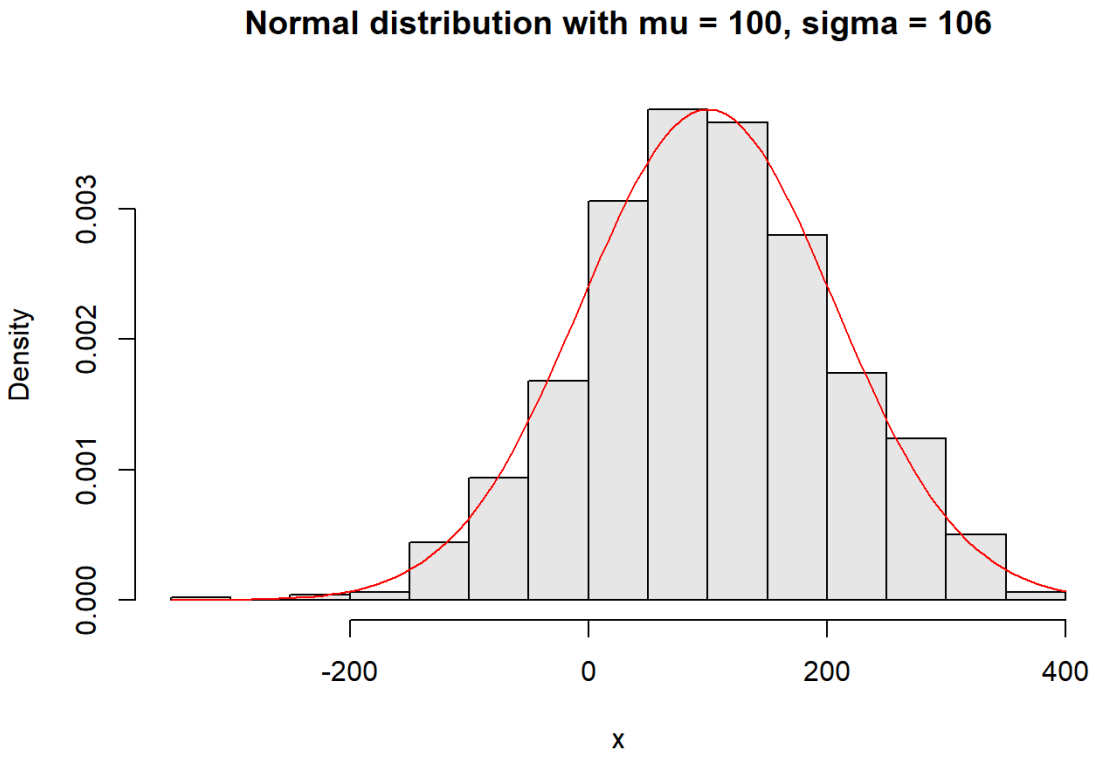
We can calculate the skewness and kurtosis of the sample.

```
> skewness(x)
[1] 0.03686954
> kurtosis(x)
[1] 0.1771157
```

As we can see the skewness is approximately 0, because the normal distribution is symmetric. As far as the kurtosis is almost 0, the distribution is mesokurtic.

Let's generate 1000 realizations of $N(100,\ 106^2)$ random variable.

```
> x <- rnorm(1000, mean = 100, sd = 106)
> hist(x, probability = TRUE, col = gray(.9), main = "Normal distribution with mu = 100, sigma = 106")
> curve(dnorm(x, mean = 100, sd = 106), add = TRUE, col = "red")
```



z-scores

$$Z\text{-}score(x_i) = \frac{x_i - \mu}{\sigma}$$

```
> mean.x <- 100; sd.x <- 16
> x <- rnorm(5, mean.x, sd.x); x
[1] 109.58559 105.78376  89.16187 127.06364 111.72651
> z.score = (x - mean.x) / sd.x; z.score
[1]  0.5990995  0.3614851 -0.6773831  1.6914773  0
```

Chi-squared distribution /Хи квадрат разпределение/

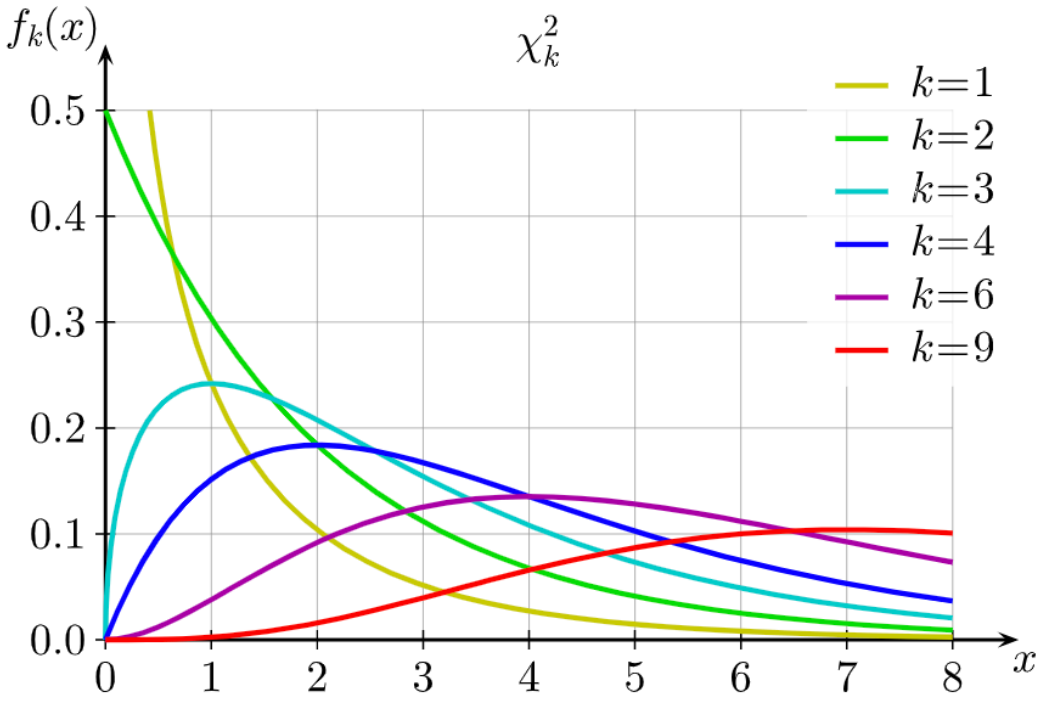
$$X \in \mathcal{X}^2(n)$$

Chi-squared distribution is a particular case of gamma distribution. More precisely

$$\Gamma\left(\frac{n}{2}, \frac{1}{2}\right) \equiv \mathcal{X}^2(n)$$

$\mathcal{X}^2$ distribution with n degrees of freedom is the distribution of a sum of the squares of n independent standard normal variables

$X_1, X_2, \ldots, X_n \in N(0,1)\ iid \Rightarrow X_1^2 + X_2^2 + \ldots + X_n^2 \in \mathcal{X}^2(n)$



Mean

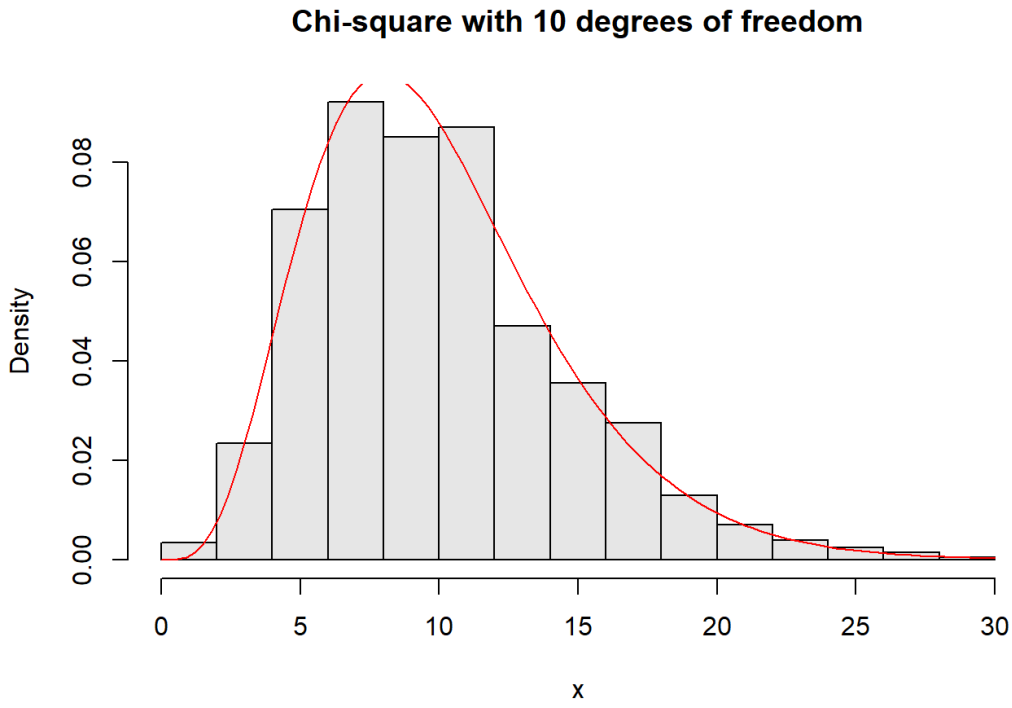
$$\mu = n$$

Variance

$$\sigma^2 = 2n$$

We can use the functions dchisq, rchisq, pchisq and qchisq to find the theoretical probability density function, to generate random numbers, to find the cumulative distribution function and quantiles of chi-square distributed variable.

```
> x <- rchisq(1000, 10)
> hist(x, probability = TRUE, col = gray(.9), main = "Chi-square with 10 degrees of freedom")
> curve(dchisq(x, 10), add = TRUE, col = "red")
```



We can calculate the skewness and kurtosis of the sample.

```
> skewness(x)
[1] 0.8620565
> kurtosis(x)
[1] 0.9117713
```

As we can see the skewness is positive, because the $\mathcal{X}^2$ distribution is right-skewed. As far as the kurtosis is positive, the distribution is leptokurtic.

Example:

```
> pchisq(0.7, df = 5)
[1] 0.01703132
> pchisq(0.7, df = 5, lower.tail = FALSE)
[1] 0.9829687
> qchisq(0.75, df = 5)
[1] 6.62568
```

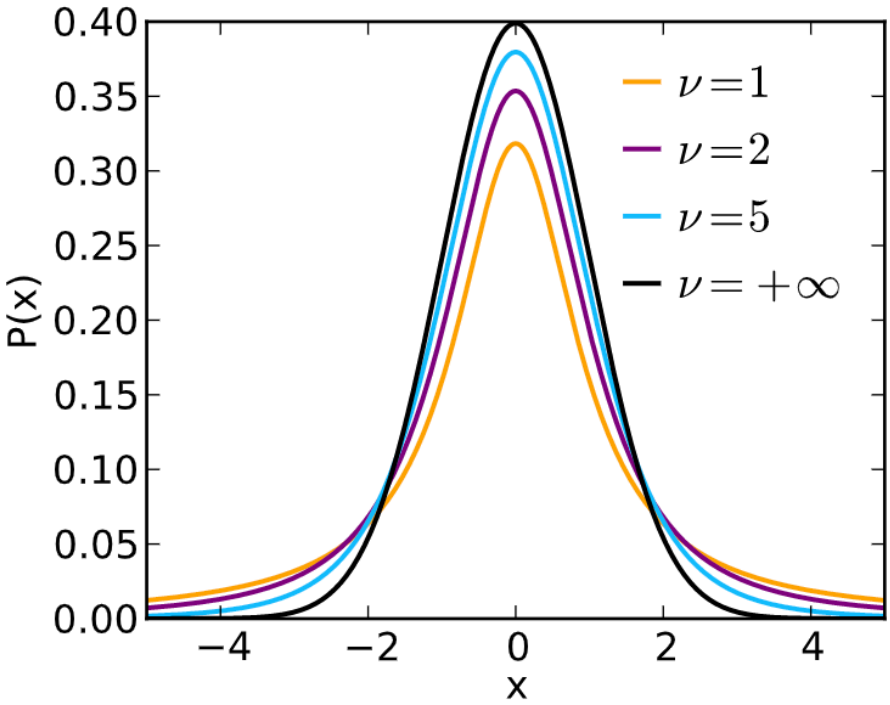
Student's t-distribution /t разпределение/

Student's t-distribution was developed by **William Sealy Gosset** under the pseudonym Student. It can be used when the population proportion is normal and the sample size is under 30.

$$X \in T(\nu)$$

Where ν = sample size $-$ 1 and is called **degrees of freedom**.

Probability density function (PDF)



Mean

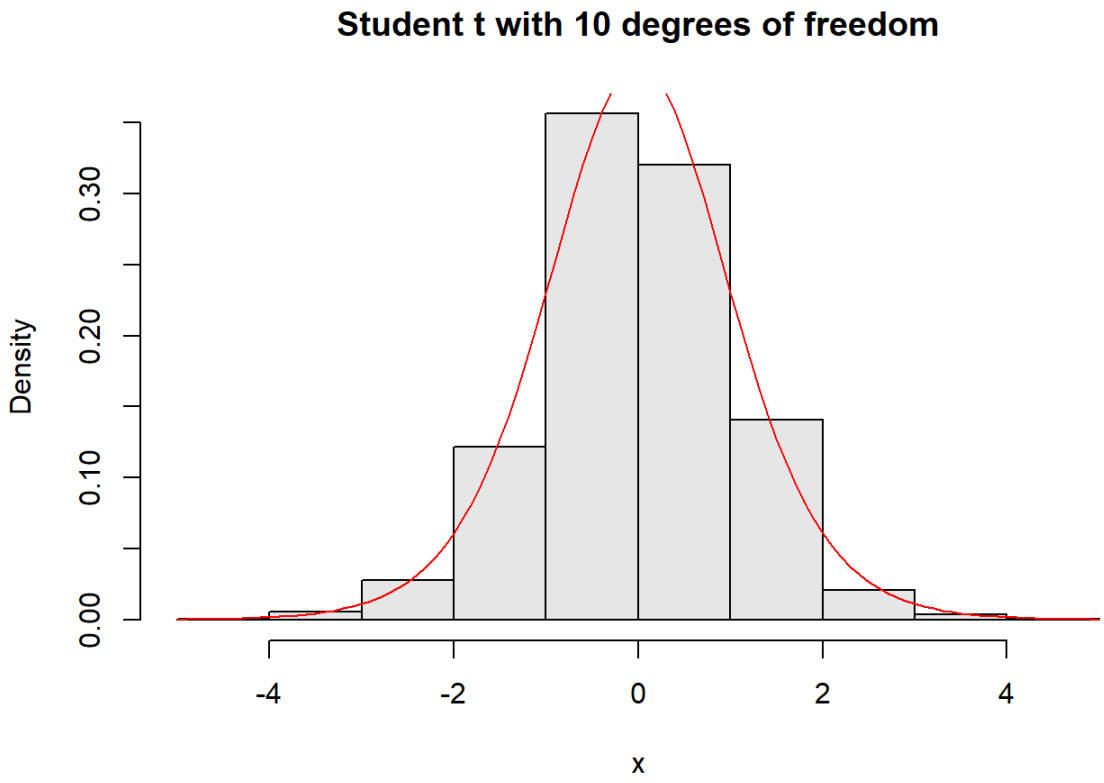
$$\mu = 0$$

Variance

$$\sigma^2 = \frac{\nu}{\nu - 2}$$

We can use the functions dt, rt, pt and qt to find the theoretical probability density function, to generate random numbers, to find the cumulative distribution function and quantiles of student-t distributed variable.

```
> x <- rt(1000, 10)
> hist(x, probability = TRUE, col = gray(.9), main = "Student t with 10 degrees of freedom")
> curve(dt(x, 10), add = TRUE, col = "red")
```



We can calculate the skewness and kurtosis of the sample.

```
> skewness(x)
[1] -0.05814847
> kurtosis(x)
[1] 0.9696638
```

As we can see the skewness is approximately 0, because the student-t distribution is symmetric. As far as the kurtosis is positive, the distribution is leptokurtic.

Sources

[1] Monika Petkova's notes on R programming language @ FMI, Sofia University