Най-често използвани вероятностни разпределения. Генериране на резултати от наблюдения от дадено разпределение.

От математическа гледна точка, две случайни величини са еднакво разпределени, ако имат една и съща функция на разпределени. Да припомним, че теоретичната функция на разпределение, това е

$$F_\xi(x) = P(\xi < x).$$

В случаите на дискретни случайни величини обикновено се използва, че функцията на разпределение се задава от реда на разпределение

$$P(\xi = x_i), \quad i = 1, 2, \dots, k.$$

В случая на абсолютно непрекъснати случайни величини, се използва, че плътността на разпределение $P_\xi(x)$ задава закона на разпределение. По-точно

$$P(a \le \xi < b) = F_\xi(b) - F_\xi(a) = \int_a^b P_\xi(x)\,dx.$$

Т.е. тази вероятност е равна на лицето под кривата на плътността и над абсцисната ос, когато първите координати на точките са в интервала [a,b].

Да припомним, че когато съществува, плътността на разпределение $P_\xi(x)$, тя е онази неотрицателна функция, за която можем да кажем, че

$$F_\xi(x_0) = \int_{-\infty}^{x0} P_\xi(x)\,dx.$$

Тази дефиниция означава, че когато можем да диференцираме горния израз, плътността на разпределение е производната на функцията на разпределение.

От приложна гледна точка това, че две случайни величини са еднакво разпределени означава, че те имат едно и също поведение. Това изобщо не означава, че те трябва да съвпадат. Например ако искаме да симулираме число, което се е паднало при едно подхвърляне на симетричен зар, това означава, че трябва да симулираме реализация на случайна величина, която има ред на разпределение:

k	1	2	3	4	5	6	Общо:
$P(\omega: \xi(\omega) = k)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	1

Със средствата на R, за да симулираме такива реализации можем да използваме функцията ***sample***. Тя генерира реализации на случайна величина, която е дискретно равномерно разпределена върху зададеното в първия параметър множество.

Например: Резултатите от 20 подхвърляния на симетричен зар могат да бъдат генерирани с

```
> sample(1:6, 20, replace = T)
[1] 5 1 5 3 2 5 2 3 6 3 5 4 6 6 2 4 6 2 3 6
```

Можем да напишем функция, която да симулира резултатите от n подхвърляния на зар.

```
> RollDie = function(n)
        {
        sample(1:6, n, replace = T)
        }
```

Функцията се вика чрез своето име и задаване на входните параметри. Например ако искаме с тази функция да симулираме 5 подхвърляния на зар, това ще стане с

```
> RollDie(5)
[1] 3 6 1 2 2
```

Всъщност R може да генерира реализации на случайни величини с много разнообразно поведение. На най-често използваните разпределения учените са задали имена. С някои от тези типове разпределения ще се запознаем по-долу.

Най-често използваните в практиката абсолютно непрекъснати разпределения са равномерното и нормалното.

Непрекъснато равномерно разпределение.

ξ е равномерно разпределена случайна величина върху интервала [a, b], (накратко $\xi \sim U(a, b)$), ако плътността на разпределение на ξ има вида

$$P_\xi(x) = \begin{cases} 0 & x \notin [a,b] \\ \dfrac{1}{b-a} & x \in [a,b] \end{cases}.$$

Тъй като тази плътност е постоянна върху целия интервал [a, b], то шансът случайната величина да попадне, в който и да е подинтервал с фиксирана дължина, на интервала [a, b], е един и същ.

За по-добро интуитивно разбиране на равномерното разпределение е полезен следният резултат: Ако върху отсечка с дължина 1 по случаен начин се избира точка и ξ е разстоянието от левия край на отсечката до избраната точка, тогава $\xi \sim U(0, 1)$.

Свойства: Ако $\xi \sim U(a, b)$, то

1. $E\xi = \dfrac{a+b}{2}$ - средата на интервала;

2. $D\xi = \dfrac{(b-a)^2}{12}$. (В числителя е квадрата на дължината на интервала.)

3. За всяка абсолютно-непрекъсната случайна величина η е вярно, че ако в нейната функция на разпределение заместим променливата със същата случайна величина η, получаваме равномерно разпределена случайна величина върху интервала [0, 1], т.е.

$$F\eta(\eta) \sim U(0, 1).$$

4. За всяка абсолютно-непрекъсната случайна величина η и за всяка $\xi \sim U(0, 1)$ е вярно, че ако в обратната на функцията на разпределение на η (тя се нарича още квантил функцията на η), заместим променливата с равномерно разпределена върху интервала [0, 1], то новата случайна величина, $F\eta^{\leftarrow}(\xi)$ има функция на разпределение, която съвпада с $F\eta(x)$. Т.е. $F\eta^{\leftarrow}(\xi)$ и η са еднакво разпределени случайни величини.

Със средствата на R, например резултатите от 10 реализации на случайна величина, която е равномерно разпределена върху ивнервала [2, 7] могат да бъдат получени с

```
> runif(10, 2, 7)
[1] 5.550283  3.113392  4.282246  6.159806  6.977515  2.564523  6.480799  5.460151
[9] 2.706110  4.812207
```

Общият вид на функцията е

runif(n, min, max)

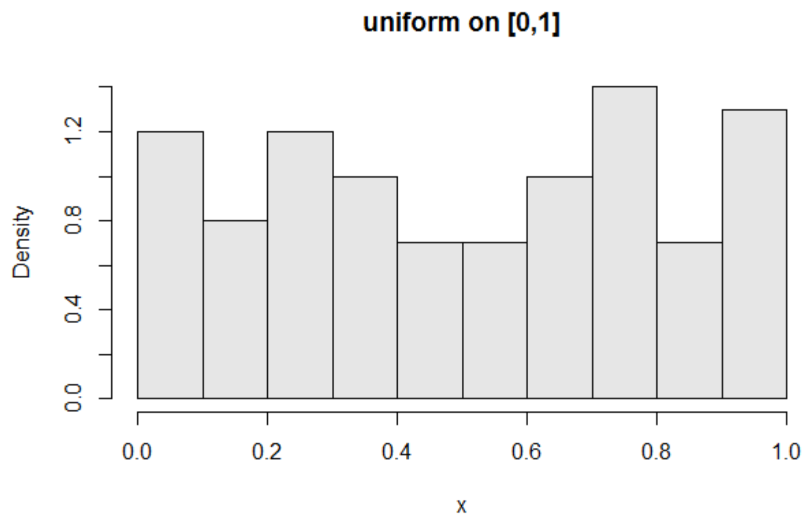
Той позволява да изберете реализациите на колко равномерно разпределени случайни величини да симулирате (n) и в какъв интервал да са те ([min, max]). По-принцип в R, всички функции, които се използват за симулиране на случайни величини с дадено разпределение започват с r и след това следва съкращение на името на съответното разпределение. По подразбиране интервалът е [0, 1].

Например: На следващия ред са симулирани 100 реализации от наблюдения върху равномерно разпределена върху интервала [0, 1] случайна величина.

```
> x = runif(100)
```

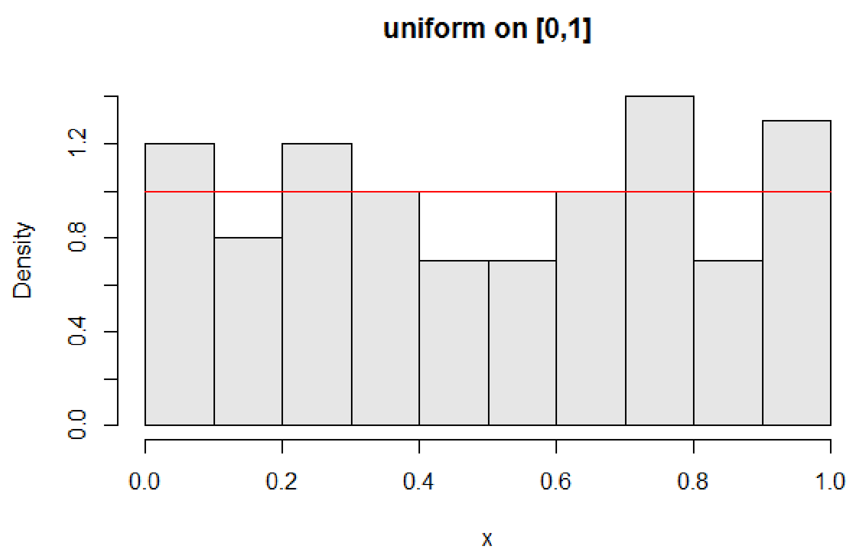
Хистограмата на данните е оценка на плътността и тя може да бъде построена например с

```
> hist(x, probability = TRUE, col = gray(.9), main = "uniform on [0,1]")
```



Съответната крива на плътността може да бъде изчертана с помощта на функциите ***curve*** и ***dunif***. По-долу тя е начертана с червен цвят.

```
> curve(dunif(x, 0, 1), add = T, col = 2)
```



Функцията

dunif(x, min, max)

пресмята плътността на разпределение на равномерно разпределена случайна величина върху интервала [min, max], в точката x. Т.е. Плътността на разпределение на равномерно разпределена случайна величина върху интервала [0, 1] в 0.5 е

```
> dunif(0.5, 0, 1)
[1] 1
```

По подразбиране интервалът е [0, 1].

По-принцип в R, всички функции, които се използват за определяне на плътността на случайни величини с дадено разпределение започват с d и след това следва съкращение на името на съответното разпределение.

Всички функции, които се използват за определяне на функцията на разпределение на случайни величини, с дадено разпределение, започват с p и след това следва съкращение на името на съответното разпределение.

Например: Функцията на разпределение на равномерно разпределена случайна величина върху интервала [0, 1] и в 0.2, т.е. $P(X < 0.2)$ може да се намери с

```
> punif(0.2, 0, 1)
[1] 0.2
```

Тук имаме възможност да пресметнем $P(X > 0.2)$ като на параметъра lower.tail зададем стойност FALSE

```
> punif(0.2, 0, 1, lower.tail = FALSE)
[1] 0.8
```

Четвъртият вид функция, която се свързва с равномерното разпределение е функцията ***qunif***. Тя връща квантилите на равномерното разпределение. Например ако min = 0 и max = 1, тя има вида

```
> qunif(0.6, 0, 1)
[1] 0.6.
```

Отново с помощта на параметъра lower.tail можем да зададем дали лицето преди този квантил да е $p \in (0, 1)$

qunif(p, min = 0, max = 1, lower.tail = TRUE)

или лицето след този квантил да е p

qunif(p, min = 0, max = 1, lower.tail = FALSE)

Бернулиевото разпределение

Случайната величина

$$I_У(\omega) = \xi(\omega) = \begin{cases} 0 & , \quad \omega = \overline{У} \\ 1 & , \quad \omega = У \end{cases}$$

се нарича Бернулиева или индикатор на събитието У.

Свойства: Ако ξ е Бернулиева случайна величина, то

1.

k	0	1	Общо:
$P(\omega: \xi(\omega) = k)$	p	q	1

2. $E\xi = p$.

3. $D\xi = pq$.

Биномно разпределение.

Бернулиевото разпределение е частен случай на Биномното разпределение. То се дефинира по средния тачин.

Нека N пъти да се повтаря един и същ опит и резултатите от всеки опит да са независими един от друг. С p означаваме вероятността да се осъществи събитието У, в резултат от провеждането на един от тези опити, а с μ_N броят на сбъдванията на събитието

У при всичките n опита, тогава μ_n е биномно разпределена случайна величина с параметри n и p, накратко $\mu_n \sim Bi(n, p)$.

Свойства: Ако $\mu_n \sim Bi(n, p)$.

1. Редът на разпределение на μ_n е

$$P(\mu_n = k) = C_n^k p^k (1-p)^{n-k},$$ където k = 0, 1, 2, … , n,

$C_n^k = \binom{n}{k} = \dfrac{n(n-1)...(n-k+1)}{1.2...k}$, при k = 1, 2, … , n, е броят на ненаредените k-елементни подмножества на крайно множество, съдържащо n елемента, а $C_n^0 = 1$. Символът $\binom{n}{k}$ е известен още като Нютонов бином. По дефиниция 0! = 1.

2. Ако означим с m = (n+1)p и ако то е цяло число, то случайната величина μ_n има две моди (най-вероятни значения) m и m-1. Ако m не е цяло число, mode μ_n = [m].

3. $E\mu_n = np$, т.е. математическото очакване е равно на произведението от параметрите.

4. $D\mu_n = np(1-p)$.

С какво може да ни бъде полезен R в случая:

Ако $\xi \sim Bi(n, p)$ функцията

pbinom(q, n, p, lower.tail = …)

пресмята функцията на разпределение $P(\xi \le q)$ ако параметърът lower.tail = TRUE и $P(\xi > q)$ ако параметърът lower.tail = FALSE.

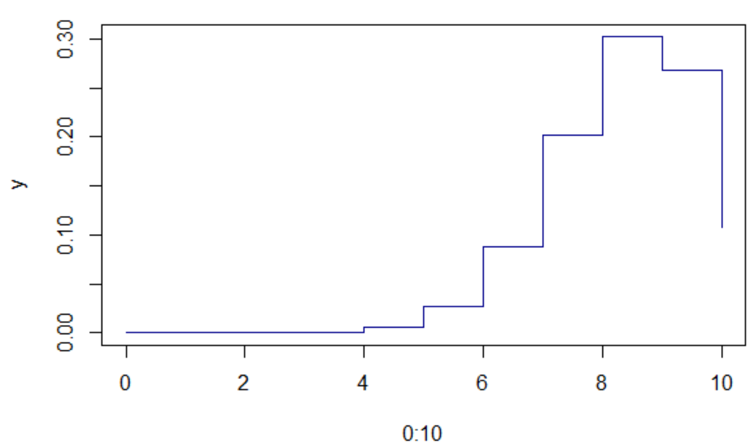
Функцията

dbinom(k, n, p)

пресмята $P(\xi = k)$.

Например: Теоретичната хистограма на $\xi \sim Bi(10, 0.8)$ може да бъде получена по следния начин

```
> y = dbinom(seq(from = 0, to = 10, by = 1), 10, 0.8)
> plot(0:10, y, type = "s", col = "darkblue")       # параметърът s означава, че ще изчертаваме
                                                    # графика на стъпала.
```



Третият вид функция, която се свързва с Биномното разпределение е функцията ***qbinom.*** Тя връща квантилите на Биномното разпределение. Т.е.

$$qbinom(\alpha, N, p, lower.tail = \ldots)$$

връща най-малкото x такова, че $P(\xi \leq x) \geq \alpha$, ако параметърът lower.tail = TRUE и същата функция връща най-малкото x такова, че $P(\xi \geq x) \leq \alpha$, ако параметърът lower.tail = FALSE.

Пример: Намерете третия квартил на Биномно разпределена случайна величина с параметри 5 и 0.3 може да бъде получен с

```
> qbinom(0.75, 5, 0.3)
[1] 2
```

Можем да намерим x такова, че $P(\xi \geq x) \leq 0.25$,

```
> qbinom (0.25, 5, 0.3, lower.tail = FALSE)
[1] 2
```

Да обърнем внимание, че т.к. това разпределение е дискретно, тази функция НЕ е точно е обратна на функцията ***pbinom.*** Например в случая

```
> pbinom (2, 5, 0.3)
 [1] 0.83692
```

Функцията

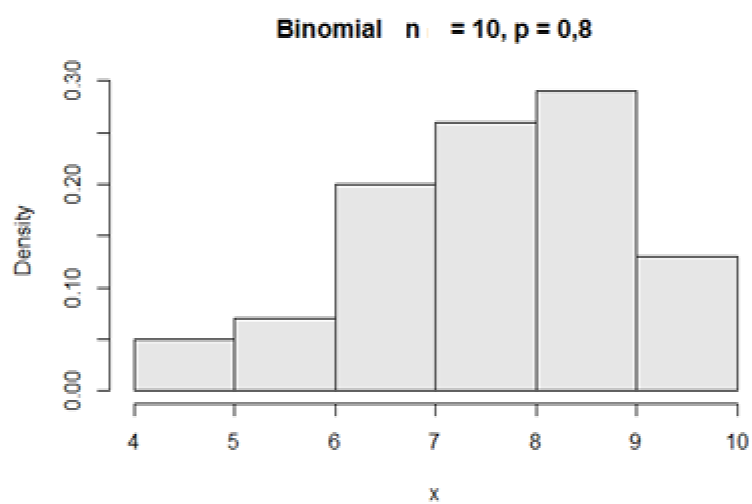
$$rbinom(m, n, p)$$

връща m реализации на тази Биномно разпределена случайна величина с параметри n и p.

```
> x = rbinom(100, 10, 0.8); x            # Брой успехи при 10 повторения на независими
                                         # опити, с вероятност за успех 0,8.
  [1]  9  8 10  9 10  7  7  9  8  7  4  7  8 10  7  8  8  9  7  8 10  9  9  7  6  9
 [27]  9  9  7  8  8  7  9  8  8  6  8  7  8  9  8  9  9  9  7  9  8  8  9  7  9 10
 [53]  5  9  9  8  9  7  8  6  9  7  6 10 10  6  5  7 10  9  9  7 10  8  7  8  6 10
 [79]  7  6  8  9  7  4  5 10 10  7  9  8  8  9  9  9  8  8 10  8  8  9
```

Да начертаем (емпиричната) хистограма и да сравним с горната теоретична графика.

```
> hist(x, probability = TRUE, col = gray(.9), main = "Binomial n = 10, p = 0,8")
```



Геометрично разпределение.

Нека имаме независими повторения на един и същ опит, докато се сбъдне събитието У = “успех”. Нека p е вероятността да се осъществи събитието У, в резултат от провеждането на един от тези опити. Случайната величина μ, която показва броя на неуспехите до I – вия “успех” се нарича геометрично разпределена случайна величина с вероятност за успех p. Накратко μ ~ Ge(p).

Свойства: Ако μ ~ Ge(p), то

1. Редът на разпределение на μ има вида

$$P(\mu = k) = (1-p)^k p, \text{ където } k = 0, 1, 2, \ldots.$$

2. mod μ = 0, $E\mu = \dfrac{1-p}{p}$, a $D\mu = \dfrac{1-p}{p^2}$.

Отрицателно биномно разпределение.

Отрицателното биномно разпределение е обобщение на геометричното разпределение.

Случайната величина ξ, която показва броя на “неуспехите” до n – тия “успех” се нарича отрицателно биномно разпределена случайна величина, с вероятност за “успех” p. Накратко ξ ~ NBi(n; p).

Свойства:

1. NBi(1; p) съвпада с Ge(p).
2. Ако ξ ~ NBi(n; p), то редът на разпределение на ξ има вида

$$P(\xi = k) = C_{n+k-1}^{k}(1-p)^k p^n, \text{ където } k = 0, 1, 2, \ldots.$$

2. mod μ = 0, $E\mu = \dfrac{n(1-p)}{p}$, a $D\mu = \dfrac{n(1-p)}{p^2}$.

С какво може да ни бъде полезен R в случая:

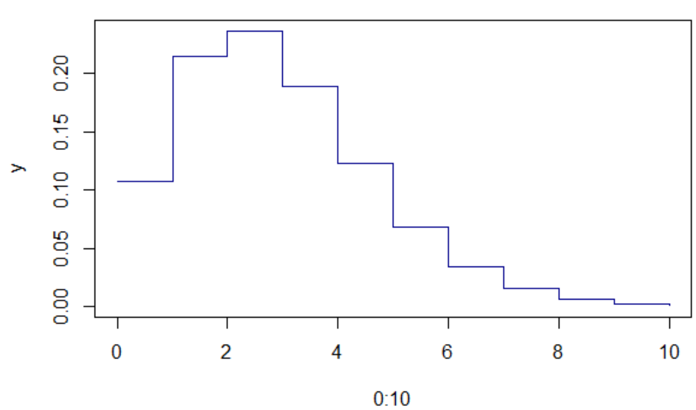
Ако ξ ~ NBi(n; p) функцията

dnegbinom(k, n, p)

пресмята P(ξ = k).

Например: Теоретичната хистограма на ξ ~ NBi(10, 0.8) може да бъде получена по следния начин

```
> y = dnbinom(seq(from = 0, to = 10, by = 1), 10, 0.8)
> plot(0:10, y,  type = "s", col = "darkblue")
```



Функцията

pnbinom(q, n, p, lower.tail = …)

пресмята функцията на разпределение P(ξ ≤ q) ако параметърът lower.tail = TRUE и същата функция пресмята P(ξ > q), ако параметърът lower.tail = FALSE.

Третият вид функция, която се свързва с Отрицателното биномно разпределение е функцията ***qnbinom.*** Тя връща квантилите на Отрицателното биномно разпределение. Т.е.

qnbinom(α, n, p, lower.tail = …)

връща най-малкото x такова, че P(ξ ≤ x) ≥ α, ако параметърът lower.tail = TRUE и същата функция връща най-малкото x такова, че P(ξ ≥ x) ≤ α, ако параметърът lower.tail = FALSE.
Пример: Намерете третият квартил на Отрицателно биномно разпределена случайна величина с параметри 5 и 0.3.

```
> qnbinom(0.75, 5, 0.3)
[1] 15
```

Намерете x такова, че P(ξ ≥ x) ≤ 0.25,

```
> qnbinom (0.25, 5, 0.3, lower.tail = FALSE)
[1] 15
```

Да обърнем внимание, че т.к. това разпределение е дискретно, тази функция НЕ е точно е обратна на функцията ***pnbinom.*** Например в случая

```
> pnbinom (15, 5, 0.3)
 [1] 0.7624922
```

Функцията

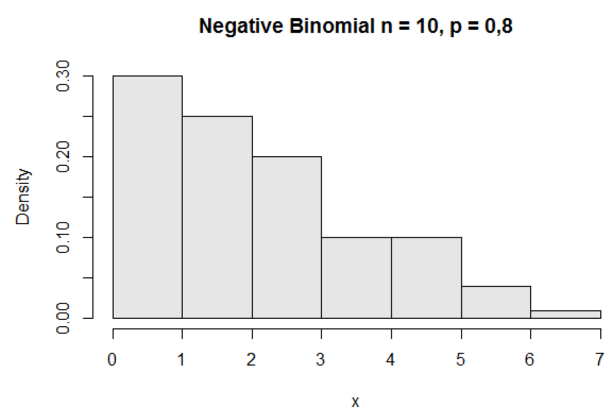
rbinom(m, n, p)

връща m реализации на тази Отрицателно биномно разпределена случайна величина с параметри n и p.

```
> x = rnbinom(100, 10, 0.8)
> x
  [1] 0 2 4 4 2 5 3 3 2 3 8 4 5 3 1 0 4 2 2 2 4 2 2 1 6 3 2 2 1 3 6 4 3 1 2 5 4 6 3 2
 [41] 0 6 4 3 1 3 4 1 1 4 0 1 2 2 0 0 4 1 3 2 2 9 6 4 4 1 4 4 4 9 3 2 4 3 6 2 4 3 3 2
 [81] 2 2 1 3 6 1 7 5 2 4 3 1 1 2 0 2 0 2 3 2
```

Да начертаем хистограмата и да сравним с горната теоретична графика.

```
> hist(x, probability = TRUE, col = gray(.9), main = "Negative Binomial n = 10, p = 0,8")
```



Хипергеометрично разпределение.

Много често се прави избор на част от елементите на множество без връщане. В такъв случай се стига до хипергеометричното разпределение.

Ако разполагаме с a елемента от един вид и с b елемента от друг вид. Условно да ги наречем a бели топки и b черни топки. По случаен начин, без връщане избираме n от тях, където $n \leq a + b$. Нека ξ е броят на извадените елементи от първия вид, т.е. извадените бели топки. Дискретната случайна величина ξ е хипергеометрично разпределена с параметри n, a и b, накратко

$$\xi \sim HG(n; a, b).$$

Възможните значения на ξ са целите числа в интервала [max(0, n-b), min(n, a)].

Свойства: Ако $\xi \sim HG(n; a, b)$.

1. ξ има следния ред на разпределение

$$P(\xi = k) = \frac{C_a^k C_b^{N-k}}{C_N^{a+b}},$$

където k = max(0, n-b), max(0, n-b)+1, … , min(n, a).

2. Нека (с цел улесняване на записа) $m = \frac{(a+1)(n+1)}{a+b+2}$. Ако m е цяло число, то ξ има две моди m и m-1. Ако m не е цяло число, mode ξ = [m].

3. Вярно е, че $E\xi = \frac{na}{a+b}$, $D\xi = \frac{nab(a+b-N)}{(a+b)^2(a+b-1)}$.

С какво може да ни бъде полезен R в случая:

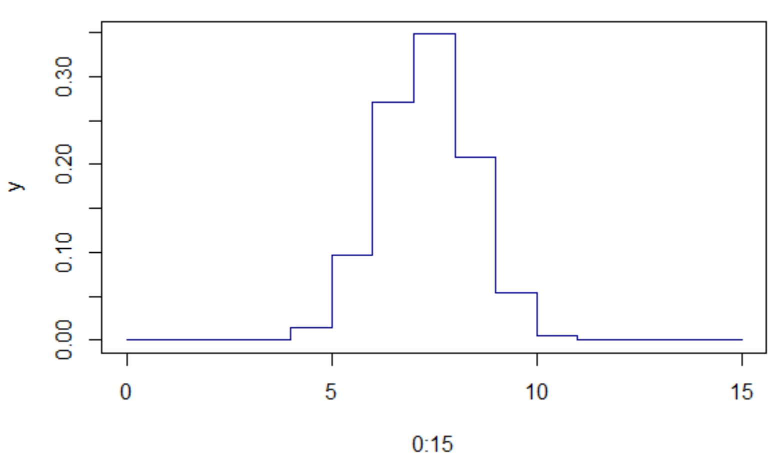
Ако ξ ~ HG(n; a, b) функцията

dhyper(k, a, b, N)

пресмята P(ξ = k).

Например: Теоретичната хистограма на ξ ~ HG(15; 10, 12) може да бъде получена по следния начин

```
> y = dhyper(seq(from = 0, to=15,by=1), 10, 12, 15)
> plot(0:15,y, type = "s", col = "darkblue")
```



Функцията

phyper(q, a, b, n, lower.tail = …)

пресмята функцията на разпределение P(ξ ≤ q) ако параметърът lower.tail = TRUE и същата функция пресмята P(ξ > q) ако параметърът lower.tail = FALSE.

Третият вид функция, която се свързва с Хипер-геометричното разпределение е функцията ***qhyper.*** Тя връща квантилите на Хипер-геометричното разпределение. Т.е.

qhyper(α, a, b, n, lower.tail = …)

връща най-малкото x такова, че P(ξ ≤ x) ≥ α, ако параметърът lower.tail = TRUE и същата функция връща най-малкото x такова, че P(ξ ≥ x) ≤ α, ако параметърът lower.tail = FALSE.

Пример: Намерете третият квартил на Хипер-геометрична разпределена случайна величина с параметри 5, 6 и 8.

```
> qhyper(0.75, 6, 8, 5)
[1] 3
```

Намерете x такова, че P(ξ ≥ x) ≤ 0.25,

```
> qhyper(0.25, 6, 8, 5, lower.tail = FALSE)
[1] 3
```

Да обърнем внимание, че т.к. това разпределение е дискретно, тази функция НЕ е точно е обратна на функцията ***phyper.*** Например в случая

```
> phyper(3, 6, 8, 5)
 [1] 0.9370629
```

Функцията

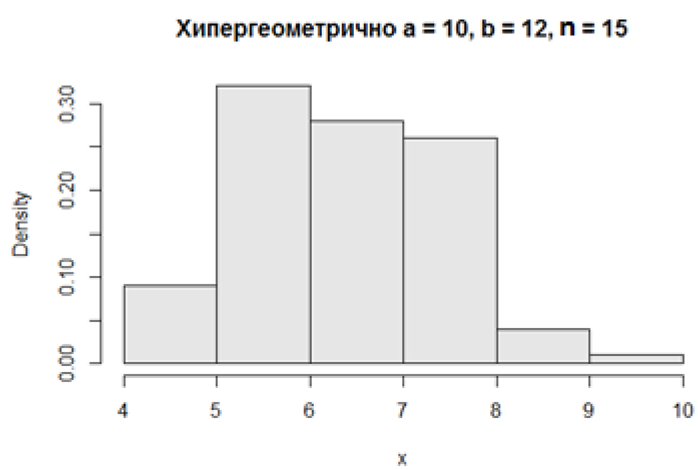
rhyper(m, a, b, n)

връща m реализации на тази Хипер-геометрично разпределена случайна величина с параметри n, a и b.

```
> x = rhyper(100, 10, 12, 15)
> x
  [1]  4  8  7  6  8  6  6  6  7  8  6  6  7  6  7  6  7  8  7  7  6  6  6  7  7  6  5
 [28]  8  7  9  6  8  6  6  5  6  5  8  6  8  7  5  7  9  7  7  8  6  8  7  7  9  6  6
 [55]  7  6  8  7  6  8  8  8  6  6  6  8  8  6  4  6  6  7  7  9  8  8  7  6  7  7  7
 [82]  5  8  6  8  7  7  8  8  6  8  8  7 10  5  5  6  8  7  8
```

Да начертаем хистограмата и да сравним с горната графика.

```
> hist(x, probability = TRUE, col=gray(.9), main = "Хипергеометрично a = 10, b = 12, n = 15")
```



Поасоново разпределение

η е разпределена по закона на Поасон с параметър $\lambda > 0$, накратко $\eta \sim P_0(\lambda)$, ако

$$P(\eta = k) = \frac{\lambda^k}{k!} e^{-\lambda}, \quad \text{където } k = 0, 1, 2, \ldots$$

Това разпределение е важно от теоретична гледна точка, защото с него обикновено се моделира броя на клиентите, които пристигат в дадена система за единица време. Тогава параметърът му λ е средният брой на клиентите, които пристигат в системата за единица време.

Свойства: Ако $\eta \sim P_0(\lambda)$, то

1. ако λ е цяло число, то η има две моди λ и λ-1. Ако λ не е цяло число, mode $\eta = [\lambda]$.
2. $E\eta = \lambda$,
3. $D\eta = \lambda$.

С какво може да ни бъде полезен R в случая:

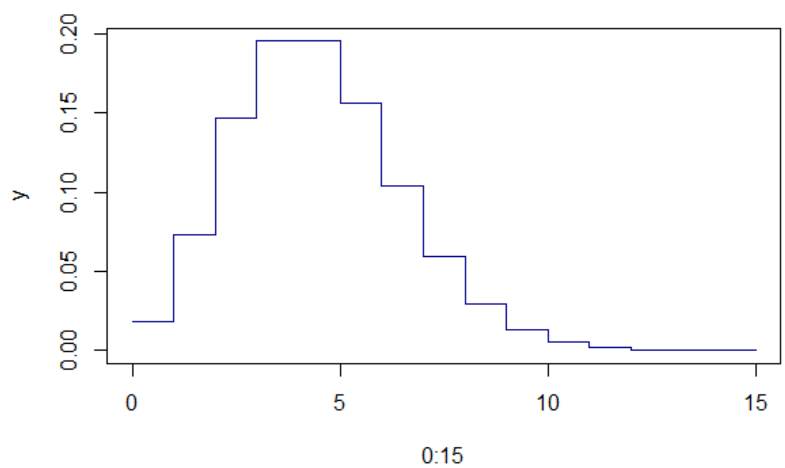
Ако $\xi \sim Po(\lambda)$ функцията

dpois(k, lambda)

пресмята $P(\xi = k)$.

Например: Теоретичната хистограма на $\xi \sim Po(4\lambda)$ може да бъде получена по следния начин

```
> y = dpois(seq(from = 0, to = 15, by = 1), 4)
> plot(0:15, y,  type = "s", col = "darkblue")
```



Функцията

ppois(q, lambda, lower.tail = …)

пресмята функцията на разпределение $P(\xi \le q)$ ако параметърът lower.tail = TRUE и същата функция пресмята $P(\xi > q)$ ако параметърът lower.tail = FALSE.

Например: Ако средно 12 коли пресичат мост в минута, намерете вероятността най-много 16 коли да пресекат моста в следващата минута.

```
> ppois(16, lambda = 12)
[1] 0.89871
```

Намерете вероятността 17 и повече коли да пресекат моста в следващата минута.

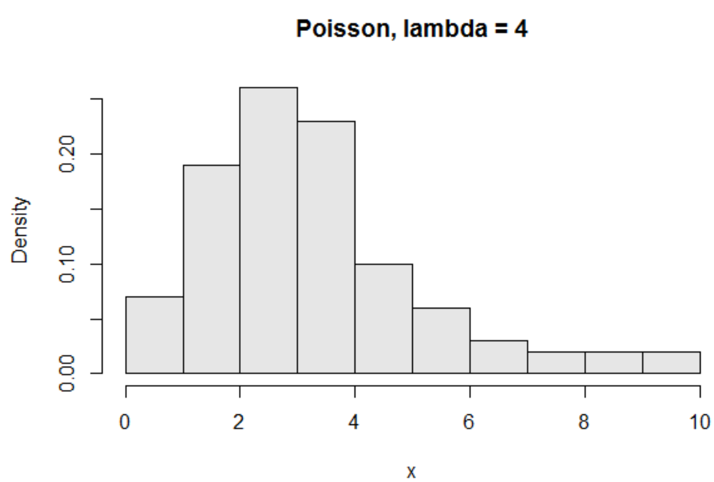
```
> ppois(16, lambda = 12, lower = FALSE)
[1] 0.10129
```
Третият вид функция, която се свързва с Поасоновото разпределение е функцията qpois. Тя връща квантилите на Поасоновото разпределение. Т.е.

qpois(α, lambda, lower.tail = …)

връща най-малкото x такова, че $P(\xi \le x) \ge \alpha$, ако параметърът lower.tail = TRUE и същата функция връща най-малкото x такова, че $P(\xi \ge x) \le \alpha$, ако параметърът lower.tail = FALSE.

Пример: Намерете третият квартил на Поасоново разпределена случайна величина с параметър 5.
```
> qpois(0.75, 5)
[1] 6
```
Намерете x такова, че $P(\xi \ge x) \le 0.25$,
```
> qpois (0.25, 5, lower.tail = FALSE)
[1] 6
```
Да обърнем внимание, че т.к. това разпределение е дискретно, тази функция НЕ е точно е обратна на функцията ***ppois.*** Например в случая
```
> ppois (6, 5)
[1] 0.7621835
```
Функцията

rpois(m, lambda)

връща m реализации на тази Поасоново разпределена случайна величина с параметър lambda.
```
> x = rpois(100, 4)
> x
  [1]  2  3  5  2  4  4  4  4  6  5  5  6  3  2  7  6  1 10  0  7  4  5  2  4  3  4  3
 [28]  3  7  4  3  3  2  3  4  9  4  3  4  3  4  3  4  6  2  6  3  3  5  3  2  4  4  2
 [55] 10  4  3  2  3  0  3  2  2  3  2  1  5  5  2  4  5  3  3  1  1  1  3  3  4  3  4
 [82]  8  3  2  2  8  2  6  3  2  5  4  9  2  5  4  3  4  4  2
```
Да начертаем хистограмата и да сравним с горната графика.
```
> hist(x, probability = TRUE, col = gray(.9), main = "Poisson, lambda = 4")
```


Сега ще преминем на най-често използваните абсолютно непрекъснати разпределения.

Нормално (Гаусово разпределение)

ξ е нормално (Гаусово) разпределена случайна величина с параметри а $\in$ R и $\sigma > 0$, накратко $\xi \sim N(a, \sigma^2)$, ако за всяко реално число х, плътността на разпределение на ξ има вида

$$P_\xi(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-a)^2}{2\sigma^2}}, x \in R.$$

Т.е. нормалното разпределение има два параметъра а и σ^2.
Ще казваме, че ξ е стандартно Гаусово разпределена случайна величина ако $\xi \sim N(0, 1)$.
Смисълът на параметрите се вижда от следните свойства.

Свойства: Ако $\xi \sim N(a, \sigma^2)$, то

1. $E\eta = a$,

2. $D\eta = \sigma^2$.

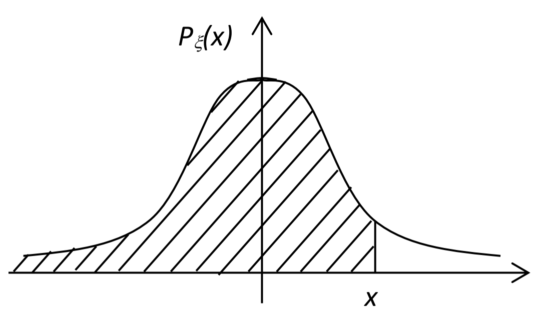
3. Нормалното разпределение винаги може да бъде стандартизирано до стандартно нормално. По-точно, ако $\xi \sim N(a, \sigma^2)$, то $\frac{\xi - a}{\sigma} \sim N(0,\ 1)$.

4. Ако случайните величини $\xi_k \sim N(a_k, \sigma_k^2)$ при $k = 1, 2,\ldots, n$, са независими, то тяхната сума

$$\xi_1 +\ldots+ \xi_n \sim N(a_1 +\ldots+ a_n, \sigma_1^2 +\ldots+ \sigma_n^2).$$

3. Ако $\xi \sim N(a, \sigma^2)$ и k и b са константи, то $k\xi + b \sim N(ka+b, k^2\sigma^2)$.

На следващата фигура е дадена геометричната интерпретация на връзката между квантилите, функцията на разпределение $\Phi(x)$ и плътността на разпределение на нормално разпределена случайна величина. $\Phi(x)$ е лицето на защрихованата част. Лицето на незащрихованата фигура между кривата на плътността на стандартното Гаусово разпределение и абсцисната ос е $1-\Phi(x)$.



Лицето на цялата фигура, получена под кривата на плътността и над абсцисната ос винаги е 1.

При $x > 3$, $\Phi(x)$ е почти 1, а когато x е отрицателно число, стойностите на $\Phi(x)$ могат да се определят като се използва равенството $\Phi(-x) = 1 - \Phi(x)$.

5. Ако $\xi \sim N(0, 1)$ и $\alpha \in [0, 1]$ и $P(-z_\alpha < \xi < z_\alpha) = 2(1 - \Phi(z_\alpha)) = 2\alpha$.

Нормалното разпределение е важно, т.к. от Централната гранична теорема, средното аритметично на всяка проста извадка[1] с обем n, от популация със средно μ и дисперсия $\sigma^2 < \infty$ е приблизително нормално разпределена случайна величина с средно μ и дисперсия σ^2/n, когато n клони към безкрайност. По-точно ако центрираме и нормираме тази величина получаваме, че

$$\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \approx N(0, 1)$$

Т.е. е центрираното и нормирано следно аритметично е приблизително стандартно нормално разпределена случайна величина при големи n.

С какво може да ни бъде полезен R в случая:

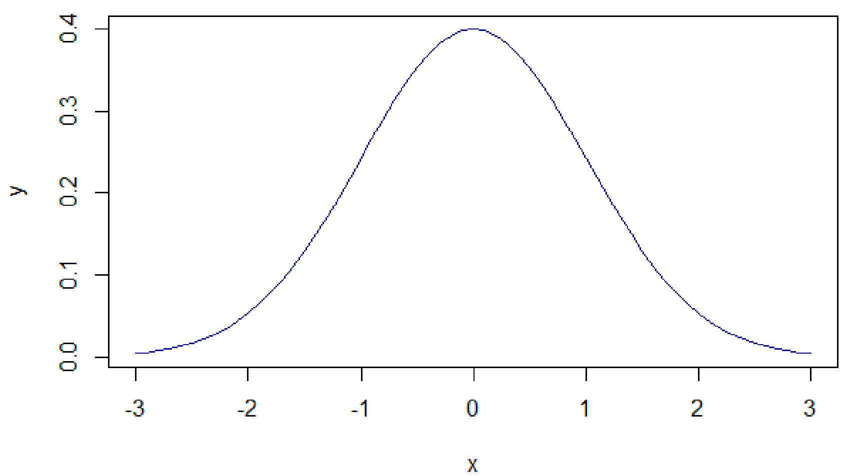
Ако $\xi \sim N(a, \sigma^2)$ функцията

dnorm(x, mean, standard deviation)

пресмята $P_\xi(x)$.

Например: Плътността (теоретичното приближение на хистограмата) на $\xi \sim N(0, 1)$, която се намира по формулата от дефиницията за стандартно нормално разпределена случайна величина, при $\mu = 0$ и $\sigma^2 = 1$, може да бъде получена по следния начин

```
> x = seq(from = -3, to = 3, by = 0.1)
> y = dnorm(x, mean = 0, sd = 1)
> plot(x, y, type  = "l", col = "darkblue")
```



Функцията

pnorm(q, mean, standard deviation, lower.tail = …)

пресмята функцията на разпределение $P(\xi \leq q)$ ако параметърът lower.tail = TRUE и същата функция пресмята $P(\xi > q)$, ако параметърът lower.tail = FALSE.

Например: Ако $\xi \sim N(0, 1)$, стойностите на $\Phi(x)$ в x = 2, т.е. лицето на защрихованата част на по-предната графика при x = 2 може да бъде получено с

```
> pnorm(2, mean = 0, sd = 1)
```

---

[1] Проста извадка е извадка от независими наблюдения върху една и съща случайна величина.

[1] 0.9772499

Лицето на незащрихованата част на същата графика се явява горна опашка на разпределението и за това може да бъде получено с

```
> pnorm(2, mean = 0, sd = 1, lower = FALSE)
[1] 0.02275013
```

Да припомним, че лицето под всяка крива на плътността винаги е 1.

Третият вид функция, която се свързва с нормалното разпределение е функцията ***qnorm.*** Тя връща квантилите на Нормалното разпределение. Т.е.

qnorm(α, mean, standard deviation, lower.tail = …)

връща най-малкото x такова, че $P(\xi \leq x) \geq \alpha$, ако параметърът lower.tail = TRUE и същата функция връща най-малкото x такова, че $P(\xi \geq x) \leq \alpha$, ако параметърът lower.tail = FALSE.

Пример: Намерете третият квартил на стандартното нормално разпределение.

```
> qnorm(0.75, 0, 1)
[1] 0.6744898
```

Намерете x такова, че $P(\xi \geq x) \leq 0.25$,

```
> qnorm(0.25, 0, 1, lower.tail = FALSE)
[1] 0.6744898
```

Да обърнем внимание, че тъй като нормалното разпределение е абсолютно непрекъснато, тази функция е обратна на ***pnorm.*** Например в случая

```
> pnorm(0.6744898, 0, 1)
[1] 0.75
```

Функцията

rnorm(m, mean, sd)

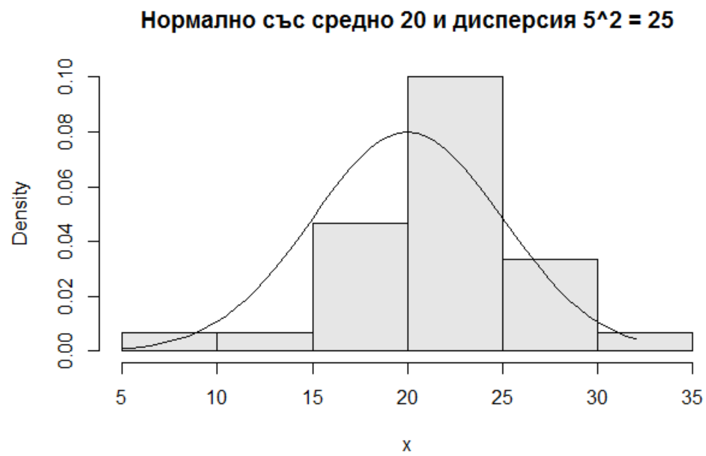
връща m реализации на $\xi \sim N(mean, sd^2)$.

Например: Ако е известно, че разпределението на заредените количества бензин на клиентите от бензиностанция са $\xi \sim N(20, 25)$. Можем да симулираме приблизителните заредени количества от следващите 30 клиента чрез

```
> x = rnorm(30, mean = 20, sd = 5); x
[1] 15.911418  13.161052  21.596320 27.628101 27.090141 23.560447   9.981635 23.881215
 [9] 22.096000 18.723552  20.108449 18.062732 20.413937 19.921584 20.277872 22.081947
[17] 21.116695 22.738274 16.733473 20.160969 22.922462 18.522323 21.346882 18.592223
[25] 26.794038 34.004732 25.554833 22.690287 21.471080 27.428638
```

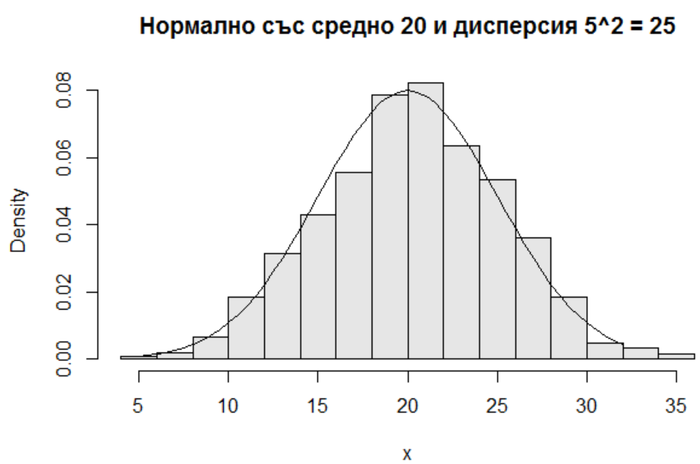
Да начертаем хистограмата и да сравним с горната графика, но без да центрираме и нормираме.

```
> hist(x, probability = TRUE, col = gray(.9), main = "Нормално със средно 20 и дисперсия 5^2 = 25")
>  y = dnorm(seq(from = 5, to = 32, by = 0.5), mean = 20, sd = 5)
>  lines(seq(from = 5, to = 32, by = 0.5), y, type = "l")
```



Оказва, че се че при по-голям брой наблюдения това приближение ще е по-добро. Това се дължи на централната гранична теорема, която ще разгледаме пак по-долу, т.е. наблюдаваното разпределение не е точно нормално, а само асимптотично нормално. Нека сега да направим 1000 симулации на същата случайна величина

```
>  x = rnorm(1000, mean = 20, sd = 5);
>  hist(x, probability = TRUE, col = gray(.9), main = "Нормално със средно 20 и дисперсия 5^2
= 25")
>  x1 = seq(from = 5, to = 32, by = 0.5)
>  y = dnorm(x1, mean = 20, sd = 5)
>  lines(x1, y, type = "l")
```



Виждаме, че хистограмата много повече се доближава до съответната стандартна нормална крива. Т.е. ако искаме да оценим типа на дадено разпределение е добре да разполагаме с достатъчно голям брой наблюдения.

Пример: Ако резултатите от тест на входен изпит в колеж са със средно 72 и стандартно отклонение на средното аритметично 15.2, какъв е очакваният процент на студентите от цялата генерална съвкупност, които имат поне 84 точки?

Отг. От Централна гранична теорема, при големи извадки средното е N(72, 15.2^2). Т.к. търсим оценка на вероятността то да е по-голямо от 84 ние се нуждаем от оценка на upper tail на това разпределение.

```
> pnorm(84, mean = 72, sd = 15.2, lower.tail = FALSE)
[1] 0.21492
```

Т.е. търсеният процент е приблизително 21.5%.

Симулирайте приблизителните резултати от 5 такива наблюдения.

```
> rnorm(5, mean = 72, sd = 15.2)
[1] 66.54061 66.84594 40.37060 64.94596 69.92551
```

Пример: Ако резултатите от IQ тест са нормално разпределени със средно 100 и стандартно отклонение 16. Обичайно ли е човек да има над 150 точки? Симулирайте 10 резултата от подобно наблюдение

Отг. Определяме вероятността за това

```
> pnorm(150, mean=100, sd=16, lower.tail=FALSE)
[1] 0.0008890253
```

Тя е много малка, следователно това не е обичайно.

Сега да симулираме приблизителните резултати от 10 такива наблюдения.

```
> rnorm(10, mean=100, sd=16)
[1]  88.21164 110.80687  88.12782 118.88787 109.50535 101.32563
[7]  118.08886  88.49568 106.85910 107.03174
```

Пример: Ако дължината на бебе на 10 дни е нормално разпределени със средно 52 см и стандартно отклонение 9 см. Обичайно ли е да има 10 дневно дете с дължина над 60 см? Симулирайте 10 резултата от подобно наблюдение.

Отг. Определяме вероятността за това

```
> pnorm(60, mean = 52, sd = 9, lower.tail = FALSE)
[1] 0.1870314
```

Тя не е малка, следователно това е обичайно.

Сега да симулираме приблизителните резултати от 10 такива наблюдения.

```
> rnorm(10, mean = 52, sd = 9)
[1]  46.70253 68.24710 48.17501 61.09274 72.32679 48.69047 44.31309 45.75081
[9] 60.43029 55.28738
```

Сега да отговорим на въпроса защо нормалното разпределение е толкова важно и толкова често срещано.

Да припомним Централна гранична теорема, приложена за суми от независими и еднакво разпределени случайни величини с математически очаквания μ и равни дисперсии σ^2, при голям брой опити n

$$\frac{S_n - n\mu}{\sigma\sqrt{n}} \sim N(0, 1).$$

Биномното разпределение описваше броя на успехите ν_n при независими повторения на един и същи опит. То е сума от независими и еднакво разпределени индикатори на събитието „Успех“. Този брой е разбира се дискретен, но от горната Централна гранична теорема, броят на успехите става приблизително нормален при големи n. Като вземем в предвид, че математическото очакване и дисперсията на Бернулиево разпределените случайни величини (които са индикатори и теоретично описват пропорциите) са съответно $\mu = p$ и $\sigma^2 = p(1-p)$, от Централната гранична теорема, ако $\nu_n \sim Bi(n, p)$, при големи n,

$$\frac{\nu_n - np}{\sqrt{np(1-p)}} \sim N(0, 1).$$

Тази сума ν_n може да бъде представена като n пъти средното аритметично на тези индикатори. Да разделим в горното равенство числителя и знаменателя на броя на индикаторите. Виждаме, че средното аритметично е обобщение на пропорциите. Т.е. ако разделим в горната теорема числителя и знаменателя на n или пък от Централната гранична теорема, приложена за средното аритметично, т.е. от

$$\frac{\bar{X}-\mu}{\frac{\sigma}{\sqrt{n}}} \approx N(0, 1)$$

при големи n, получаваме, че при $\bar{X} = \frac{\vartheta_n}{n}$,

$$\frac{\bar{X}-p}{\frac{\sqrt{p(1-p)}}{\sqrt{n}}} \approx N(0, 1)$$

т.е. при големи n

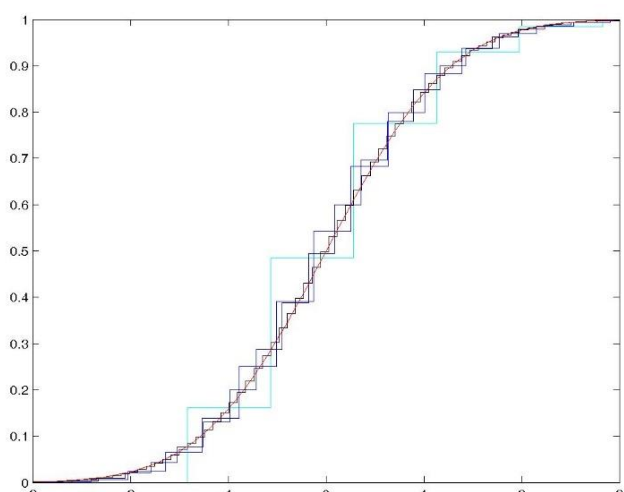
$$\frac{\sqrt{n}(\bar{X}-p)}{\sqrt{p(1-p)}} \approx N(0, 1)$$

Това ще наричаме Централна гранична теорема, приложена за пропорции.

Например. Нека разгледаме броя ν_n на падналите се шестици, при n подхвърляния на симетричен зар. На следващата фигура с все по-тъмен цвят са изобразени функциите на разпределение на

$$\frac{\nu_n - n\frac{1}{6}}{\sqrt{n\frac{1}{6}\frac{5}{6}}} = \frac{\sqrt{n}\left(\frac{\nu_n}{n} - \frac{1}{6}\right)}{\sqrt{\frac{1}{6}\frac{5}{6}}} = \frac{\sqrt{n}\left(\bar{X} - \frac{1}{6}\right)}{\sqrt{\frac{1}{6}\frac{5}{6}}}$$

където $\nu_n \sim Bi(n; \frac{1}{6})$, съответно за n = 10, 50, 100 и 1000. С червен цвят е изобразена функцията на разпределение на стандартно нормално разпределена случайна величина. Очевидно, при увеличаване на обема на извадката, биномното разпределение все повече и повече се доближава до Нормалното.



По подобен начин можем да сравним реда на разпределение на

$$\frac{\nu_n - n\frac{1}{6}}{\sqrt{n\frac{1}{6}\frac{5}{6}}} = \frac{\sqrt{n}\left(\frac{\nu_n}{n} - \frac{1}{6}\right)}{\sqrt{\frac{1}{6}\frac{5}{6}}} = \frac{\sqrt{n}\left(\overline{X} - \frac{1}{6}\right)}{\sqrt{\frac{1}{6}\frac{5}{6}}}$$

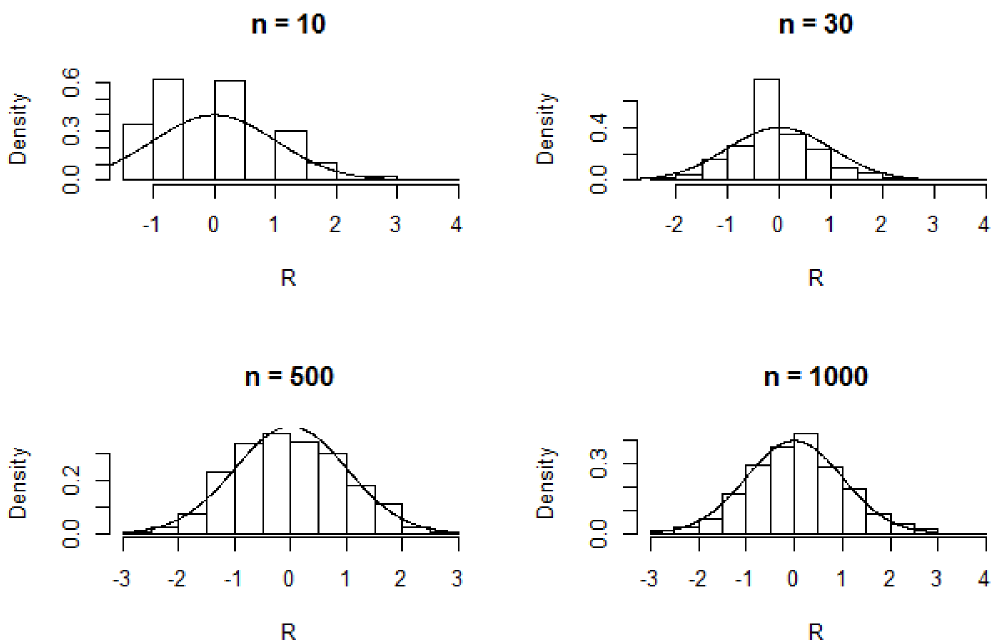
и плътността на разпределение на стандартното нормално разпределение.

Нека използваме R и да генерираме по 1000 от тези случайни числа, за всяко n. След това да начертаем хистограмата на получената извадка и да я сравним със стандартната нормална крива.

```
> par(mfrow = c(2, 2))
> p = 1/6;
> xvals = seq(-3, 3, .01)
> n = 10;
> Nu = rbinom(1000, n, p)
> R = (Nu - n*p)/sqrt(n*p*(1-p))
> hist(R, probability = TRUE, main = "n = 10")
> points(xvals, dnorm(xvals, 0, 1), type="l")
> n = 30;
> Nu = rbinom(1000, n, p)
> R = (Nu - n*p)/sqrt(n*p*(1-p))
> hist(R, probability = TRUE, main = "n = 30")
> points(xvals, dnorm(xvals, 0, 1), type = "l")
> n = 500;
> Nu = rbinom(1000, n, p)
> R = (Nu - n*p)/sqrt(n*p*(1-p))
> hist(R, probability = TRUE, main = "n = 500")
> points(xvals, dnorm(xvals, 0, 1), type = "l")
> n = 1000;
> Nu = rbinom(1000, n, p)
>  R = (Nu - n*p)/sqrt(n*p*(1-p))
>  hist(R, probability = TRUE, main = "n = 1000")
> points(xvals, dnorm(xvals, 0, 1), type = "l")
```

Получаваме следващата графика.

Забележка: Навсякъде в този пример вместо функцията ***points*** можем да използваме ***lines***. При това е достатъчно само да сменим имената им.



Вече видяхте, че повторихме аналогични действия няколко пъти. По тази причина същите графики можем да получим и с помощта на цикъл. Обърнете внимание на употребата на функцията ***paste*** за вмъкване на параметър в заглавието.

```
> par(mfrow = c(2, 2))
> p = 1/6
> n = c(10, 30, 500, 1000)
> xvals = seq(-3, 3, .01)
> for (i in 1:4) {
            Nu =  rbinom(1000, n[i], p)
            R = (Nu- n[i]*p)/sqrt(n[i]*p*(1-p))
            hist(R, probability = TRUE, main = paste("n =", n[i]))
            points(xvals, dnorm(xvals, 0, 1), type = "l")
            }
```

Отново наблюдаваме, че при увеличаване на броя на опитите, т.е. n, хистограмата на разпределението на случайната величина

$$\frac{\nu_n - n\frac{1}{6}}{\sqrt{n\frac{1}{6}\frac{5}{6}}} = \frac{\sqrt{n}\left(\frac{\nu_n}{n} - \frac{1}{6}\right)}{\sqrt{\frac{1}{6}\frac{5}{6}}} = \frac{\sqrt{n}\left(\overline{X} - \frac{1}{6}\right)}{\sqrt{\frac{1}{6}\frac{5}{6}}}$$

все повече и повече се доближава до стандартната нормална крива.

Гама разпределение

ξ е гама разпределена случайна величина с параметри $\alpha > 0$ и $\beta > 0$, накратко $\xi \sim \Gamma(\alpha, \beta)$, ако плътността на разпределение на ξ има вида

$$P_\xi(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} \; x^{\alpha-1}e^{-\beta x}, \quad x > 0,$$

където $\Gamma(\alpha) = \int_0^\infty x^{\alpha-1}e^{-x}dx$ е гама функцията с параметър $\alpha > 0$.

Свойства: Ако $\eta \sim \Gamma(\alpha, \beta)$, то

1. $E\eta = \frac{\alpha}{\beta}$,

2. $D\eta = \frac{\alpha}{\beta^2}$.

Забележка: Когато $\alpha \in N$, това разпределение се нарича Ерлангово. Частен случай на Ерланговото разпределение (при n = 1) е експоненциалното разпределение.

Няма да се спираме подробно на общия случай на това разпределение, но с него се работи по аналогичен начин на предходните и с помощта на функциите

dgamma(x, shape, rate)
pgamma(q, shape, rate, lower.tail = …)
qgamma(p, shape, rate, lower.tail = TRUE)
rgamma(n, shape, rate)

Параметърът shape е равен на α, а параметърът rate е равен на β.

Експоненциалното разпределение съвпада с $\Gamma(1, \lambda)$.

ξ е експоненциално разпределена случайна величина с параметър $\lambda > 0$, накратко $\xi \sim Exp(\lambda)$, ако плътността на разпределение на ξ има вида

$$P_\xi(x) = \lambda \; e^{-\lambda x}, \quad x > 0,$$

и нула иначе.

Експоненциалното разпределение е важно от теоретична гледна точка, защото с него обикновено се моделират интервалите между пристиганията на клиенти в дадена система, когато техният брой до момента t е Поасоново разпределен.

Свойства: Ако $\eta \sim Exp(\lambda)$, то

1. $E\eta = \frac{1}{\lambda}$.

2. $D\eta = \frac{1}{\lambda^2}$.

3. $P(\eta > x + y | \eta > x) = P(\eta > y)$. Това свойство се нарича липса на последействие (памет). Това е така защото ако с η е моделирана "продължителността на живот" на даден уред или система, това свойство означава, че ако знаем, че "устройството" е "живяло" вече време x, шансът да "живее" още време y е същия както ако не знаем колко е "живяло" това "устройство". Тогава първото свойство означава, че $1/\lambda$ е средната продължителност на живот.

4. Ако $\xi_1, \xi_2, \ldots, \xi_n$ са независими, еднакво експоненциално разпределени случайни величини с параметър λ, то сумата $\xi_1 + \xi_2 + \ldots + \xi_n$ е разпределена по закона на Ерланг с параметри n и λ, а това разпределение съвпада с $\Gamma(n, \lambda)$.

С какво може да ни бъде полезен R в случая:

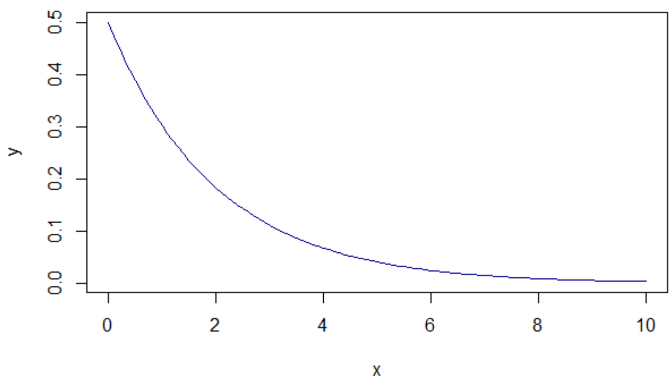
Ако $\xi \sim Exp(\lambda)$, функцията

dexp(x, rate = 1/λ)

пресмята $P_\xi(x)$.

Например: Плътността (теоретичното приближение на хистограмата), която се намира по формулата от дефиницията на експоненциалното разпределение, при λ = 2 може да бъде получена по следния начин

```
> x = seq(from = 0, to=10,by=0.1)
> y = dexp(x, rate =1/2)
> plot(x,y, type = "l", col = "darkblue")
```



Функцията

pexp(q, rate = 1/λ, lower.tail = …)

пресмята функцията на разпределение P(ξ ≤ q) ако параметърът lower.tail = TRUE и същата функция пресмята P(ξ > q) ако параметърът lower.tail = FALSE.

Например: Ако ξ ~ Exp(2), стойността на функцията на разпределение на ξ, при x = 1, може да бъде получена с

```
> pexp(1, rate =1/2)
[1] 0.3934693
```

Горната опашка на това разпределение може да бъде получена с

```
> pexp(2, rate =1/2, lower=FALSE)
[1] 0.3678794
```

Да припомним, че при едно и също x сумата от двете опашки винаги е равна на 1, т.к. тя е равна на лицето под кривата на плътността, а то е винаги 1.

Третият вид функция, която се свързва с Експоненцилното разпределение е функцията ***qexp.*** Тя връща квантилите на Експоненциалното разпределение. Т.е.

qexp (α,rate = 1/λ, lower.tail = …)

връща най-малкото x такова, че P(ξ ≤ x) ≥ α, ако параметърът lower.tail = TRUE и същата функция връща най-малкото x такова, че P(ξ ≥ x) ≤ α, ако параметърът lower.tail = FALSE.

Пример: Намерете третият квартил на експоненциалното разпределение с параметър 4.

```
> qexp (0.75, 1/4)
[1] 5.545177
```

Намерете x такова, че P(ξ ≥ x) ≤ 0.25,

```
> qexp(0.25, 1/4, lower.tail = FALSE)
[1] 5.545177
```

Да обърнем внимание, че когато разпределението е абсолютно непрекъснато, тази функция е обратна на ***pexp.*** Например в случая

```
> pexp (5.545177, 1/4)
[1] 0.75
```

Функцията

$$rexp(m, rate = 1/\lambda)$$

връща m реализации на $\xi \sim Exp(\lambda)$.

Пример: Ако времената на пристигане на клиент в система са Експоненциално разпределени, със средното аритметично 15.2 мин, какъв е очакваният процент от интервалите между пристиганията на клиентите от цялата генерална съвкупност, които ще са по-дълги от 20 мин.?

Отг. При експоненциалното разпределение, от първото свойство, параметърът λ се оценява с реципрочното на средното аритметично, т.е. в нашия случай тази оценка е 1 / 15.2 = 0.06578947. Търсим оценка на вероятността на събитието, интервалите между пристиганията на клиентите от цялата генерална съвкупност, да са по-дълги от 20 мин. Т.е. нуждаем се от оценка на upper tail на това разпределение.

```
> pexp(20, rate = 1 / 15.2, lower.tail = FALSE)
[1] 0.2682625
```

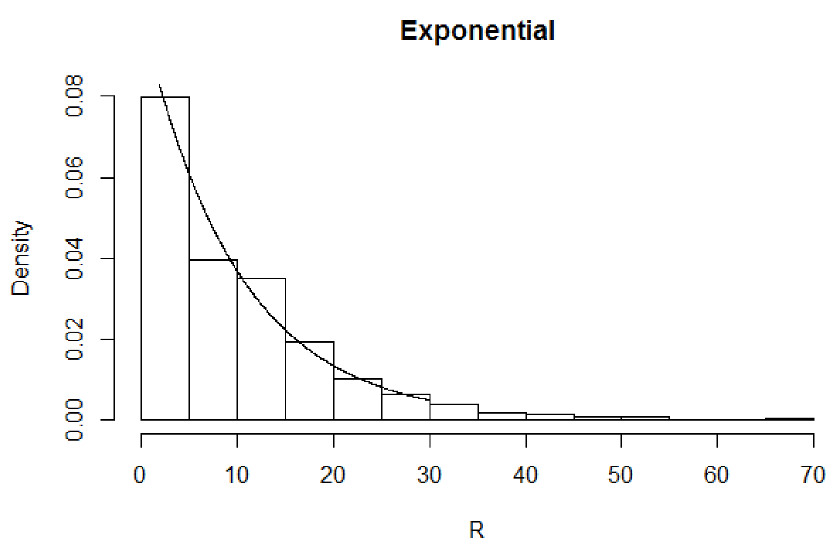
Т.е. търсеният процент е приблизително 26.83%.

Симулирайте приблизителните резултати от 5 такива наблюдения.

```
> rexp(5, rate = 1 / 15.2)
[1] 6.871076 45.010494 17.985595  2.982161 20.150225
```

Пример: Симулирайте приблизителните резултати от 500 наблюдения върху Exp(10). Начертайте хистограмата на данните и сравнете с теоретичната плътност на наблюдаваната величина.

```
> R = rexp(500, rate = 1 / 10)
>  x = seq(0, 30, .01)
>  hist(R, probability = TRUE, main = "Exponential")
> lines(x, dexp(x, 1 / 10), type = "l")
```



χ2 (хи – квадрат) разпределение

ξ е $\chi 2$ разпределена случайна величина, с n степени на свобода, накратко $\xi \sim \chi 2(n)$, ако тя съвпада по разпределение с $\Gamma(\frac{n}{2}, \frac{1}{2})$.

n е естествено число.

Свойства: Ако $\eta \sim \chi 2(n)$, то

1. $E\eta = n$,
2. $D\eta = 2n$.
3. Ако $\xi_1, \xi_2, \ldots, \xi_n$ са независими, еднакво стандартно нормално разпределени случайни величини, то сумата от квадратите им $\xi_1^2 + \xi_2^2 + \ldots + \xi_n^2 \sim \chi 2(n)$.

С какво може да ни бъде полезен R в случая.

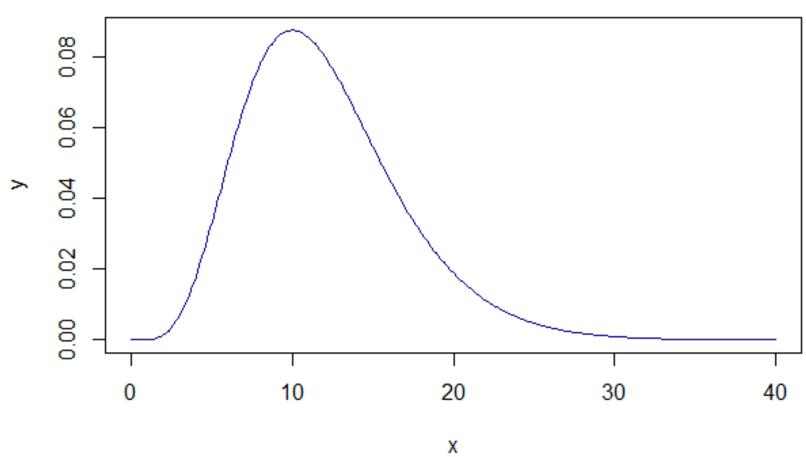
Ако $\xi \sim \chi 2(n)$, функцията

dchisq(x, df = n)

пресмята $P_\xi(x)$.

Например: Плътността (теоретичното приближение на хистограмата) на $\chi 2(12)$ може да бъде получена по следния начин

```
> x = seq(from = 0, to = 40, by = 0.1)
> y = dchisq(x, df =12)
> plot(x, y, type = "l", col = "darkblue")
```



Функцията

pchisq (q, df, lower.tail = …)

пресмята функцията на разпределение $P(\xi \leq q)$ ако параметърът lower.tail = TRUE и същата функция пресмята $P(\xi > q)$, ако параметърът lower.tail = FALSE.

Например: Ако $\xi \sim \chi 2$ (20), стойността на функцията на разпределение на ξ, при аргумент x = 10, може да бъде получена с

```
> pchisq (10, df = 20)
[1] 0.03182806
```

Горната опашка на това разпределение може да бъде получена с

```
> pchisq (10, df = 20, lower = FALSE)
```

[1] 0.9681719

Да припомним, че при едно и също x сумата от двете опашки винаги е равна на 1, т.к. тя е равна на лицето под кривата на плътността, а то е винаги 1.

Третият вид функция, която се свързва с $\chi 2$ разпределението е функцията ***qchisq.*** Тя връща квантилите на $\chi 2$ разпределението. Т.е.

qchisq (α, df, lower.tail = …)

връща най-малкото x такова, че $P(\xi \leq x) \geq \alpha$, ако параметърът lower.tail = TRUE и същата функция връща най-малкото x такова, че $P(\xi \geq x) \leq \alpha$, ако параметърът lower.tail = FALSE.

Пример: Намерете третият квартил на $\chi 2$ разпределението с 4 степени на свобода.

```
> qchisq (0.75, 4)
[1] 5.385269
```

Намерете x такова, че $P(\xi \geq x) \leq 0.25$,

```
> qchisq (0.75, 4, lower.tail = FALSE)
[1] 5.385269
```

Да обърнем внимание, че това разпределение е абсолютно непрекъснато, по тази причина, тази функция е обратна на ***pchisq.*** Например в случая

```
> pchisq (5.385269, 4)
[1] 0.75
```

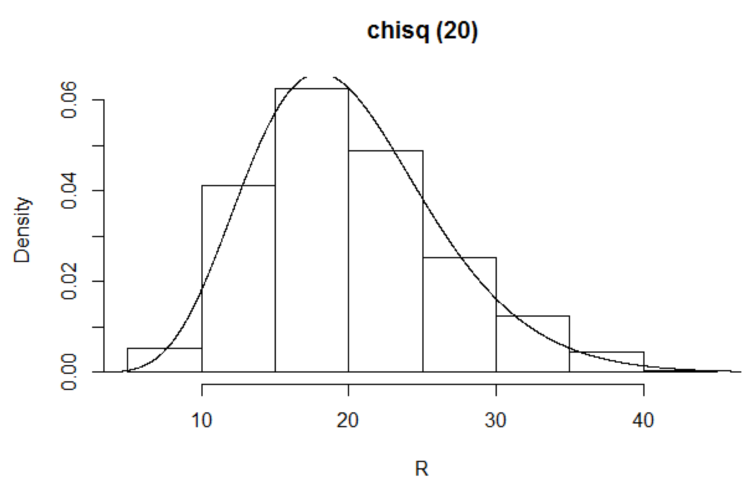
Функцията

rchisq (m, df = n)

връща m реализации на $\xi \sim \chi 2(n)$.

Пример: Симулирайте приблизителните резултати от 500 наблюдения върху $\chi 2$ (20). Начертайте хистограмата на данните и я сравнете с теоретичната плътност на наблюдаваната величина.

```
> R = rchisq (500, df = 20)
>  x = seq(0, 60, .01)
>  hist(R, probability = TRUE, main= "chisq (20)")
> lines(x, dchisq (x, 20), type="l")
```



t разпределение

ξ е t разпределена случайна величина с n степени на свобода, накратко $\xi \sim t(n)$, ако плътността на разпределение на ξ има вида

$$P_\xi(x) = \frac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n}{2})\Gamma(\frac{1}{2})\sqrt{n}(1+\frac{x^2}{n})^{\frac{n+1}{2}}}, \qquad x \in R.$$

Теорема: Ако $\xi \sim N(0, 1)$, $\eta \sim \chi2(n)$ и ако ξ и η са независими, то

$$\frac{\xi}{\sqrt{\frac{\eta}{n}}} \sim t(n).$$

Разпределението t(1) се нарича още разпределение на Коши. За него е характерно, че то няма математическо очакване.

С какво може да ни бъде полезен R в случая.

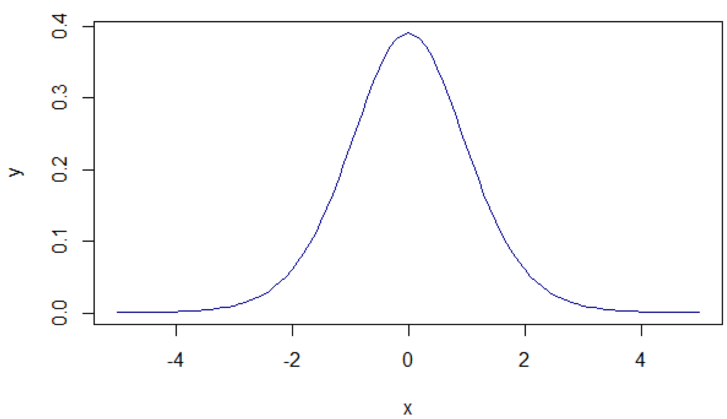
Ако $\xi \sim t(n)$, функцията

dt(x, df = n)

пресмята $P_\xi(x)$.

Например: Плътността (теоретичното приближение на хистограмата) на t(12) може да бъде получена по следния начин

```
> x = seq(from = -5, to = 5, by = 0.1)
> y = dt(x, df  = 12)
> plot(x, y, type = "l", col = "darkblue")
```



Функцията

pt(q, , df, lower.tail = …)

пресмята функцията на разпределение $P(\xi \le q)$ ако параметърът lower.tail = TRUE и същата функция връща $P(\xi > q)$, ако параметърът lower.tail = FALSE.

Например: Ако $\xi \sim t(20)$, стойността на функцията на разпределение на ξ, при x = 1, може да бъде получена с

```
> pt (1, df = 20)
[1] 0.8353717
```

Горната опашка на това разпределение може да бъде получена с

```
> pt (1, df = 20, lower=FALSE)
[1] 0.1646283
```

Да припомним, че при едно и също x сумата от двете опашки винаги е равна на 1, т.к. тя е равна на лицето под кривата на плътността, а то е винаги 1.

Третият вид функция, която се свързва с t разпределението е функцията ***qt.*** Тя връща квантилите на t разпределението. Т.е.

qt (α, df, lower.tail = …)

връща най-малкото x такова, че $P(\xi \leq x) \geq \alpha$, ако параметърът lower.tail = TRUE и същата функция връща най-малкото x такова, че $P(\xi \geq x) \leq \alpha$, ако параметърът lower.tail = FALSE.

Пример: Намерете третият квартил на t разпределението с 4 степени на свобода.

```
> qt (0.75, 4)
[1] 0.7406971
```

Намерете x такова, че $P(\xi \geq x) \leq 0.25$,

```
> qt (0.75, 4, lower.tail = FALSE)
[1] 0.7406971
```

Да обърнем внимание, че когато разпределението е абсолютно непрекъснато, тази функция е обратна на ***pt.*** Например в случая

```
> pt (0.7406971, 4)
[1] 0.75
```

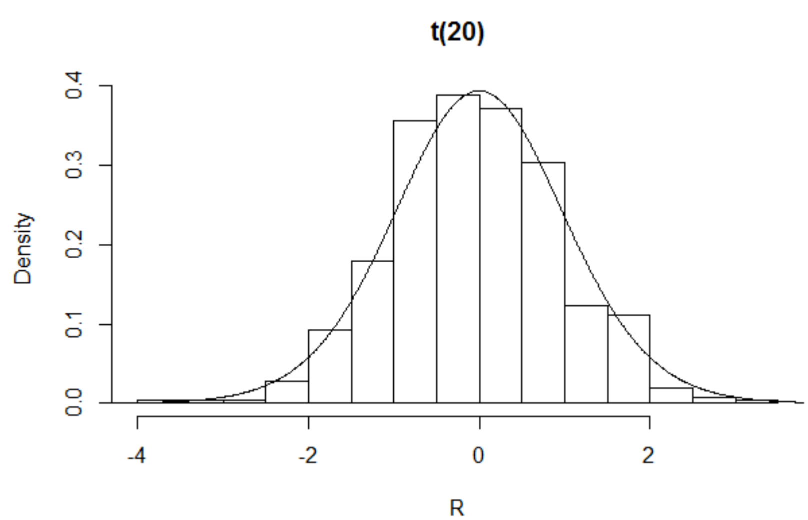
Функцията

rt (m, df = n)

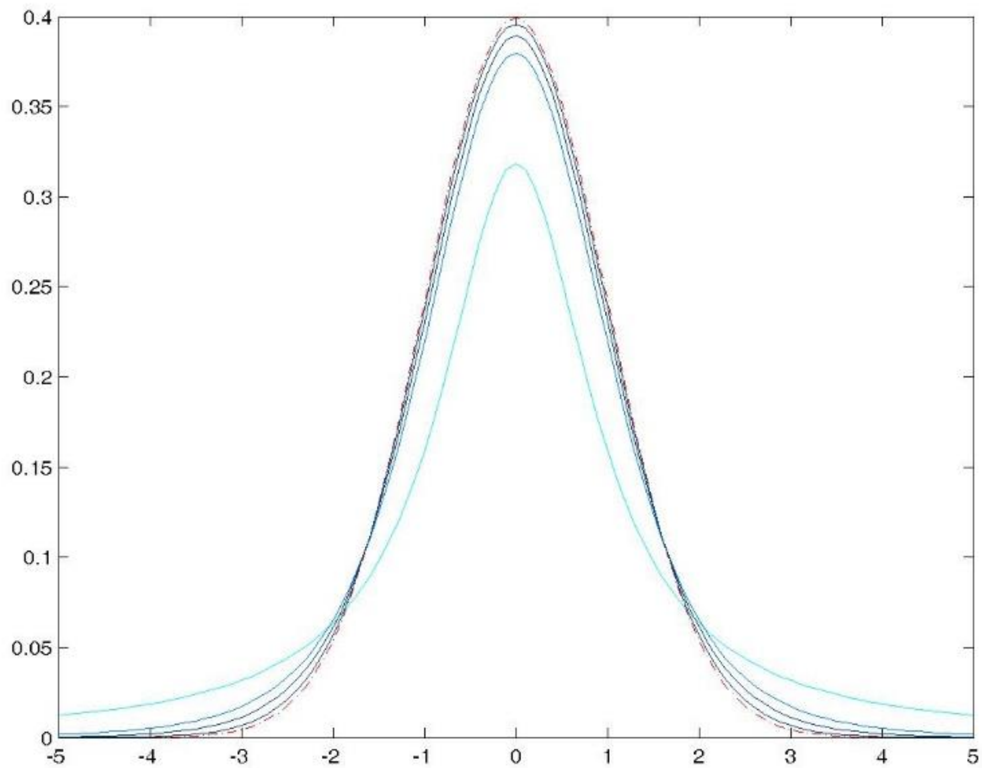
връща m реализации на $\xi \sim t(n)$.

Пример: Симулирайте приблизителните резултати от 500 наблюдения върху t(20). Начертайте хистограмата на данните и сравнете с теоретичната плътност на наблюдаваната величина.

```
> R = rt (500, df = 20)
>  x = seq(-5, 5, .01)
>  hist(R, probability = TRUE, main = "t(20)")
> lines(x, dt (x, 20), type = "l")
```



Много често, от гледна точка на статистическите методи при повече от 30 степени на свобода, вместо t(n) се използва нормалното разпределение. Причината за това е, че то е по-удобно за работа и практически не се различава от t(n). На следващата фигура с все по-тъмен син цвят са показани графиките на плътността на t(n) при увеличаване на степените на свобода. С черна пунктирна линия е показана графиката на плътността на стандартно нормално разпределена случайна величина.



F разпределена случайна величина (разпределение на Fisher)

ξ е F разпределена случайна величина с m степени на свобода на числителя и n степени на свобода на знаменателя, накратко ξ ~ F(m, n), ако плътността на разпределение на ξ има вида

$$P_{\xi}(x) = \frac{\Gamma(\frac{m+n}{2})}{\Gamma(\frac{n}{2})\Gamma(\frac{m}{2})}\left(\frac{m}{n}\right)^{\frac{m}{2}} x^{\frac{m}{2}-1} \frac{1}{(1+\frac{m}{n}x)^{\frac{n+m}{2}}}, \qquad x \in R.$$

Теорема:

Ако ξ ~ χ2(m), η ~ χ2(n) и ако ξ и η са независими, то $\dfrac{\frac{\xi}{m}}{\frac{\eta}{n}}$ ~ F(m, n).

Свойство: Fn,m (x) = 1 - Fm,n (1/x), където с Fn,m (x) сме означили функцията на разпределение на Фишър с n степени на свобода на числителя и m степени на свобода на знаменателя.

С какво може да ни бъде полезен R в случая:

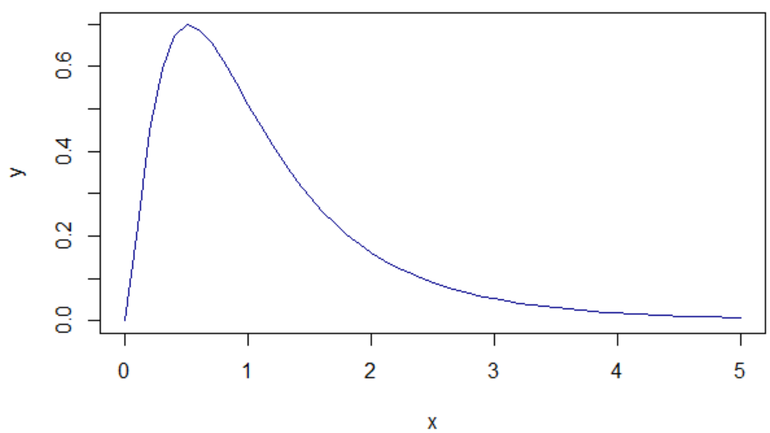
Ако ξ ~ F(m, n), функцията

df(x, df1 = m, df2 = n)

пресмята $P_\xi(x)$.

Например плътността (теоретичното приближение на хистограмата), при m = 5 и n = 12 може да бъде получена по следния начин

```
> x = seq(from = 0, to = 5, by = 0.1)
> y = df(x, df1 = 5, df2 = 12)
> plot(x, y, type = "l", col = "darkblue")
```



Функцията

pf(q, , df1 = m, df2 = n, lower.tail = …)

пресмята функцията на разпределение P(ξ ≤ q) ако параметърът lower.tail = TRUE и същата функция пресмята P(ξ > q), ако параметърът lower.tail = FALSE.

Например: Ако ξ ~ F(5, 12), стойността на функцията на разпределение на ξ, при x = 1, може да бъде получена с

```
> pf (1, df1 = 5, df2 = 12)
[1] 0.5418033
```

Горната опашка на това разпределение може да бъде получена с

```
> pf (1, df1 = 5, df2 = 12, lower = FALSE)
[1] 0.4581967
```

Да припомним, че при едно и също x сумата от двете опашки винаги е равна на 1, т.к. тя е равна на лицето под кривата на плътността, а то е винаги 1.

Третият вид функция, която се свързва с F разпределението е функцията ***qt.*** Тя връща квантилите на t разпределението. Т.е.

qt (α,df1 = m, df2 = n, lower.tail = …)

връща най-малкото x такова, че P(ξ ≤ x) ≥ α, ако параметърът lower.tail = TRUE и същата функция връща най-малкото x такова, че P(ξ ≥ x) ≤ α, ако параметърът lower.tail = FALSE.

Пример: Намерете третият квартил на F разпределението с 4 степени на свобода на числителя и 10 степени на свобода на знаменателя.

```
> qf(0.75, 4, 10)
[1] 1.594866
```

Намерете x такова, че P(ξ ≥ x) ≤ 0.25,

```
> qf(0.75, 4, 10, lower.tail = FALSE)
[1] 1.594866
```

Да обърнем внимание, че когато разпределението е абсолютно непрекъснато, тази функция е обратна на ***pf.*** Например в случая

```
> pf(1.594866, 4, 10)
[1] 0.75
```

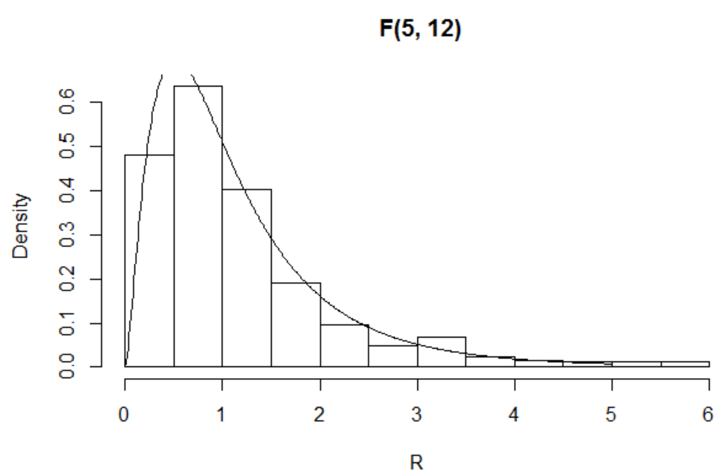
Функцията

rt (k, df1 = m, df2 = n)

връща k реализации на ξ ~ F(m, n).

Пример: Симулирайте приблизителните резултати от 500 наблюдения върху F(5, 12). Начертайте хистограмата на данните и сравнете с теоретичната плътност на наблюдаваната величина.

```
> R = rf (500, df1 = 5, df2 = 12)
>  x = seq(0, 5, .01)
>  hist(R, probability = TRUE, main = " F(5, 12)")
> lines(x, df (x, df1 = 5, df2 = 12), type="l")
```



Извадки с и без връщане.

R има възможности да прави извадки с или без връщане. За целта се използва функцията ***sample***. Тя прави случаен избор на определен брой елементи от няколко. По подразбиране извадките са без връщане.

Например при тото 6 от 49 се избира 6 числа от 49 без връщане.

```
> sample(1:49, 6)
[1] 49  5 24  3 21 29
```

Пример: Симулирайте резултатите от 5 случайни избора на карта от колода от 52 карти, без връщане.

Отг. Тук ще използваме и функциите ***paste*** и ***repeat*** за да си зададем колодата.

```
> cards = paste(rep(c("A", 2:10, "J", "Q", "K"), 4), c("H", "D", "S", "C"))
> cards
 [1] "A H"  "2 D"  "3 S"  "4 C"  "5 H"  "6 D"  "7 S"  "8 C"  "9 H"  "10 D" "J S"
[12] "Q C"  "K H"  "A D"  "2 S"  "3 C"  "4 H"  "5 D"  "6 S"  "7 C"  "8 H"  "9 D"
```

```
[23] "10 S" "J C"  "Q H"  "K D"  "A S"  "2 C"  "3 H"  "4 D"  "5 S"  "6 C"  "7 H"
[34] "8 D"  "9 S"  "10 C" "J H"  "Q D"  "K S"  "A C"  "2 H"  "3 D"  "4 S"  "5 C"
[45] "6 H"  "7 D"  "8 S"  "9 C"  "10 H" "J D"  "Q S"  "K C"
> sample(cards, 5)
[1] "J D" "5 C" "A S" "2 D" "J H"
```

При 10 подхвърляния на симетричен зар се прави случаен избор на 6 числа, с връщане. Ако искате да направите извадка с връщане трябва да зададете на параметъра ***replace*** стойност TRUE.

Пример: Симулирайте резултатите от 10 подхвърляния на симетричен зар.

Отг.

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

Пример: Симулирайте резултатите от 10 подхвърляния на симетрична монета.

Отг.

```
> sample(c("H","T"), 10, replace = TRUE)
[1] "H" "H" "T" "T" "T" "T" "H" "H" "T" "T"
```

Пример: Симулирайте резултатите от 5 пъти по 2 подхвърляния на симетричен зар.

Отг. Можем да зададем пространството от елементарни изходи при един опит. За целта ще използваме функцията ***outer.*** Например ако искаме да умножим всеки елемент на X с всеки елемент на Y тази функция ще има вида.

outer(X, Y, FUN = "*", ...)

В нашия случай долепяме символи за това функцията ще е ***paste***.

```
>  Qmega = outer(1:6, 1:6, paste)
     [,1]  [,2]  [,3]  [,4]  [,5]  [,6]
[1,] "1 1" "1 2" "1 3" "1 4" "1 5" "1 6"
[2,] "2 1" "2 2" "2 3" "2 4" "2 5" "2 6"
[3,] "3 1" "3 2" "3 3" "3 4" "3 5" "3 6"
[4,] "4 1" "4 2" "4 3" "4 4" "4 5" "4 6"
[5,] "5 1" "5 2" "5 3" "5 4" "5 5" "5 6"
[6,] "6 1" "6 2" "6 3" "6 4" "6 5" "6 6"
```

Сега правим това пространство вектор

```
> Omega = as.vector(Omega)
> Omega
 [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" "1 3"
[14] "2 3" "3 3" "4 3" "5 3" "6 3" "1 4" "2 4" "3 4" "4 4" "5 4" "6 4" "1 5" "2 5"
[27] "3 5" "4 5" "5 5" "6 5" "1 6" "2 6" "3 6" "4 6" "5 6" "6 6"
```

Вече сме готови да си изберем 5 пъти елементарен изход, т.к. можем да имаме един и същ резултат изборът трябва да е с връщане, т.е.

```
> sample(dice, 5 , replace = TRUE)
[1] "1 1" "4 1" "6 3" "4 4" "2 6"
```

Тези примери показват колко много възможности има R при случаен избор на елементи. В много случаи са полезни функциите:

- ***paste*** за да вмъкваме заедно стрингове,

- ***rep*** за повтаряне на елементи и
- ***outer*** за генериране на всички възможни двойки от елементи вектори и извършване на определена функция с тях.

По подразбиране извадките са без връщане и всеки елемент има еднакъв шанс да се появи. Може да зададете и специфични вероятности на елементарните си изходи. Например ако зарът е направен от кибритена кутия.

bootstrap извадки с и без връщане.

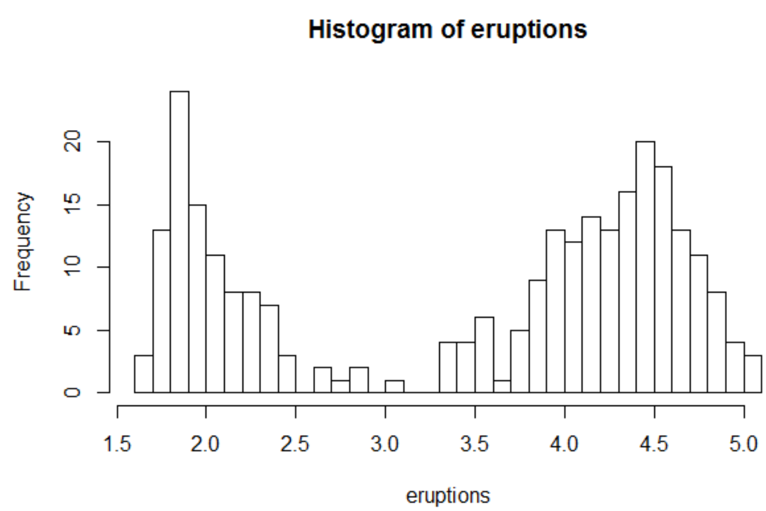
Bootstrap техниката е метод, при който чрез правене на много извадки от една и съща съвкупност се правят статистически оценки или статистически заключения. С този метод е удачно и да се проверяват заключенията. Ето една проста илюстрация на получаване на извадка.

Пример: Да разгледаме данните ***faithful*** (верен, предан, точен) от библиотеката ***datasets*** в R. Те съдържат 272 реда и две колони с данни за интервалите между изригванията (***eruptions***) и тяхната продължителност (***waiting***) за гейзерът Old Faithful в Национален парк Yellowstone в щата Wyoming(Уайоминг), USA.

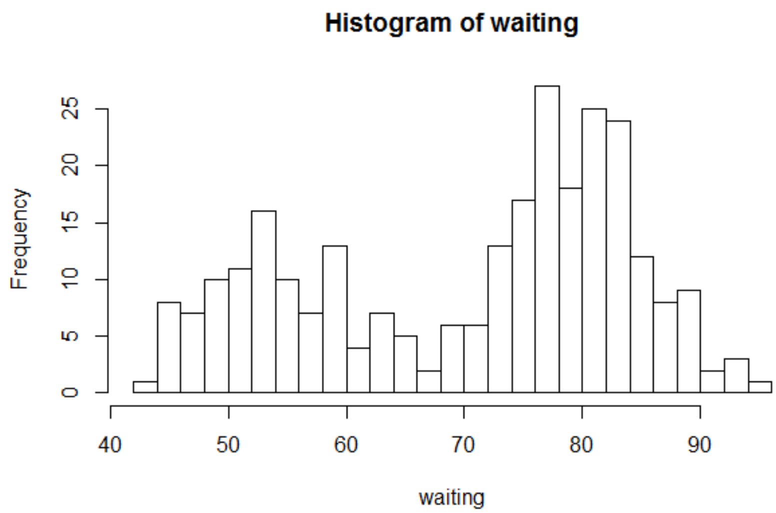
```
> data(faithful)
> head(faithful)
  eruptions waiting
1     3.600      79
2     1.800      54
3     3.333      74
4     2.283      62
5     4.533      85
6     2.883      55
```

За да моделираме разпределенията на наблюдаваните величини можем да използваме bootstrap техниката, например с повторение(избор с връщане). Нека отделим двата вектора от наблюдения във вектори със същите имена. Вместо това можем да използваме ***attach***(faithful) и ***detach***(faithful)

```
> eruptions = faithful[['eruptions']]
> hist(eruptions, breaks = 25)
```

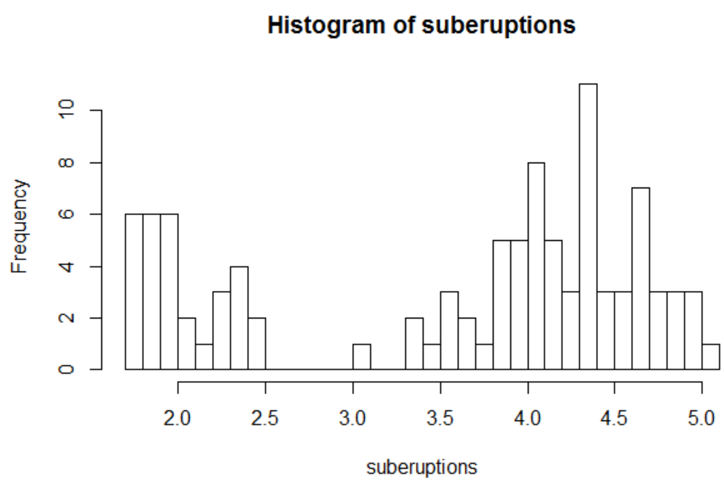


```
> waiting = faithful[['waiting']]
> hist(waiting, breaks = 25)
```



Сега да направим извадка от 100 наблюдения с възможни повторения и да начертаем нейната хистограма.

```
> suberuptions = sample(eruptions, 100, replace = TRUE)
> hist(suberuptions, breaks = 25)
```



Забелязва се, че тази хистограма прилича на по-предходната, но не е съвсем същата, т.к. е построена от по-малко данни при това с възможни повторения.

Стандартизиране със z scores

Вече знаем, че да се стандартизира една случайна величина означава да се извади от нея нейното средно и да се раздели на стандартното отклонение. Т.е.

$$Z = \frac{X-\mu}{\sigma}.$$

За целта е нужно да имаме информация за нейното средно и стандартно отклонение.

По аналогичен начин може да се стандартизира извадка. Резултатната извадка ще има средно 0 и стандартно отклонение 1. Това е полезно, когато се сравняват поведенията на случайни величини, измерени на различни скали.

Ако наблюдаваната величина е нормално разпределена, стандартизираната величина е стандартно нормално разпределена.

Например: Симулирайте 5 наблюдения върху $\xi \sim N(100, 16^2)$ и определете техните z-scores.

```
> x = rnorm(5,100,16)
> x
[1] 93.45616 83.20455 64.07261 90.85523 63.55869
> z = (x-100)/16
> z
[1] -0.4089897 -1.0497155 -2.2454620 -0.5715479 -2.2775819
```

Тогава вероятността стандартна нормално разпределена случайна величина да е в ляво от своите z-score е същата както изходната случайна величина $\xi \sim N(100, 16^2)$ да е по-малка от съответната не трансформирана със z-score. Т.е.

```
> pnorm(z)
[1] 0.34127360 0.14692447 0.01236925 0.28381416 0.01137575
> pnorm(x,100,16)
[1] 0.34127360 0.14692447 0.01236925 0.28381416 0.01137575
```

Във втория случай, да обърнем внимание, че трябва да знаем параметрите на разпределението.

Това може да се използва, да се провери типа на съответната наблюдавана случайна величина. Например за по-горе симулираната x така можем да проверим дали действително е $N(100, 16^2)$.

Задачи:

1. Генерирайте 100 наблюдения върху нормално разпределена случайна величина със средно 100 и стандартно отклонение 10. Колко е средното +- 2 стандартни отклонения. Постройте хистограмата на разпределението. Сравнете я с теоретичната плътност на разпределение на нормално разпределена случайна величина със средно 100 и стандартно отклонение 10. Какъв процент от наблюденията ви излизат извън интервала средното +- 2 стандартни отклонения? Каква е вероятността нормално разпределена случайна величина със средно 100 и стандартно отклонение 10 да излезе извън интервала средното +- 2 стандартни отклонения?

2. Симулирайте 100 подхвърляния на симетрична монета. Определете броя на езитата до момента на всяко подхвърляне. Начертайте графика, като по абсцисната ос сложите номера на подхвърлянето, а по ординатната пропорцията на езитата до момента. Какво наблюдавате?

3. Симулирайте 1000 подхвърляния на симетричен зар. Определете броя на шестиците до момента на всяко подхвърляне. Начертайте графика, като по абсцисната ос сложите номера на подхвърлянето, а по ординатната пропорцията на шестиците до момента. Какво наблюдавате?

4. Изберете по случаен начин 6 числа от 45 без връщане.

5. Ако $\xi \sim N(0,1)$, намерете числото x, такова, че

а) $P(\xi < x) = 0.05$ (използвайте qnorm);

б) $P(\xi \geq x) = 0.05$ (използвайте qnorm);

в) $P(- x < \xi < x) = 0.35$ (използвайте qnorm);

6. Определете площта под стандартната нормална крива, която е над абсцисната ос и

а) в ляво от правата $x = 1,5$;

б) в дясно от правата $x = 1,5$;

в) в дясно от правата $x = -1,5$;

г) между правите $x = -1,5$ и $x = 1,5$.

7. Определете площта под нормалната крива с параметри 0 и 4, която е над абсцисната ос и

а) в ляво от правата $x = 1,5$;

б) в дясно от правата $x = 1,5$;

в) в дясно от правата $x = -1,5$;

г) между правите $x = -1,5$ и $x = 1,5$.