Zorizko aldagaiak

Praktika honetan, batetik R lengoiako aspektu oinarrizko batzuk gogoratuko ditugu eta bestetik zorizko aldagaien inguruko kontzeptuak errepasatu.

Zorizko aldagaien definizioa: dentsitate (probabilitate-masa) eta banaketa funtzioak

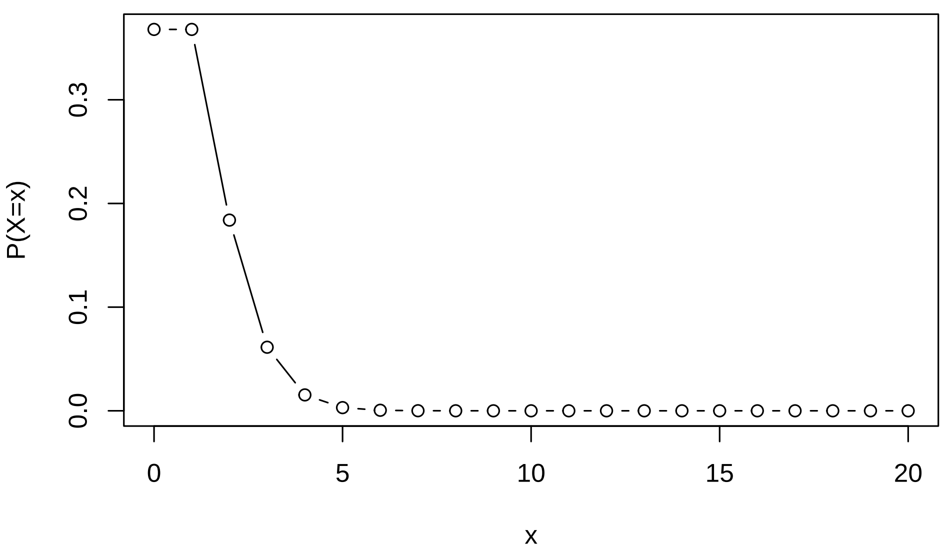
Teknologia enpresa batek X zorizko aldagaia erabiltzen du, astero, sare bat zenbat aldiz erortzen den modelatzeko. Aldagai honen probabilitate-masa funtzioa honakoa da:

$$f(x) = P(X = x) = \frac{e^{-1}}{x!}, \qquad x = 0, 1, 2, ...$$

Hasteko definitu dezagun Rn funtzio hau eta irudikatu dezagun:

```r
#Definizioa
MasaFuntzioa <- function(x){
        return(exp(-1)/factorial(x))
}

#Irudikapena
x <- 0:20
plot(x, MasaFuntzioa(x), type="b", ylab="P(X=x)")
```



Ariketa: Kalkulatu ‘MasaFuntzioa‘ funtzioa erabiliz $P(X = 1)$, $P(X > 3)$ eta $P(X <= 11)$.

```r
# P(X=1)
MasaFuntzioa(1)
```

```
## [1] 0.3678794
```

```r
# P(X>3)
1 - sum(MasaFuntzioa(0:3))
```

```
## [1] 0.01898816
```

```r
# P(X<=11)
sum(MasaFuntzioa(0:11))
```

```
## [1] 1
```

Ariketa: Definitu ‘BanaketaFuntzioa‘ funtzioa X aldagaiaren probabilitate metatua ematen duena $F(x) = P(X \leq x)$ eta kalkulatu funtzio hori erabiliz $P(X > 3)$ eta $P(X <= 11)$. Saiatu funtzioa definitzean ‘sapply‘ funtzioa erabiltzen ‘for‘ ziklo bat beharrean.

```r
BanaketaFuntzioa <- function(x) {
  return(sum(sapply(0:x, MasaFuntzioa)))
}

# P(X>3)
1 - BanaketaFuntzioa(3)
```

```
## [1] 0.01898816
```

```r
# P(X<=11)
BanaketaFuntzioa(11)
```

```
## [1] 1
```

Banaketa ezagunak aztertzeko Rko funtzioak

Apur bat fijatuz gero, erraz ikus daiteke, X aldagaia $\lambda = 1$ parametroko Poisson banaketa bat dela. Hau horrela izanik, Rk definituta dauzka hainbat funtzio banaketa ezagunekin lan egitea asko erraztuko digutenak.

Hasteko, dentsitate edo probabilitate masa funtzioa `dbanaketa` motako funtzioetan dago definituta, gure kasuan `dpois` funtzioan, hain zuzen ere:

```r
dpois(x=1, lambda=1) #P(X=1)
```

```
## [1] 0.3678794
```

Ariketa: X aldagaiari jarraituz, zein da aste batean sarea 2 aldiz erortzeko probabilitatea?

```r
dpois(x=2, lambda = 1)
```

```
## [1] 0.1839397
```

Gainera, probabilitate metatuaren funtzioa `pbanaketa` motako funtzioetan aurki dezakegu:

```r
ppois(q=2, lambda=1, lower.tail=TRUE) #P(X<=2)
```

```
## [1] 0.9196986
```

Ariketa: Erabili Rren laguntza ‘ppois‘ funtzioa zuzenean erabiliz $P(X > 4)$ nola kalkulatzen den ikasteko.

```r
ppois(q=4, lambda = 1, lower.tail = FALSE)
```

```
## [1] 0.003659847
```

Hirugarrenik, banaketaren kuantilak eta pertzentilak `qbanaketa` motako funtzioekin lor ditzakegu:

```r
qpois(p=0.5, lambda=1, lower.tail=TRUE) #P(X<=a)=0.5
```

```
## [1] 1
```

Ariketa: Zer esan nahi du aurreko kode zatian lortutako zenbakiak? Nola deitzen da balio hori?

Balio horren ezkerrean banaketaren erdia dagoela, horri mediana deitzen zaio

Azkenik, edozein banaketa ezaguneko ausazko lagin bat lortzeko `rbanaketa` motako funtzioak erabil ditzakegu. Adibidez Poisson(1) banaketatik 10 aleko lagin bat honela lor dezakegu:

```r
rpois(10, lambda=1)
```

```
##  [1] 1 1 0 0 2 1 1 2 0 3
```

Ariketa: Aurreko kodean lortutako lagina X aldagaiaren lagin bat bada, zer esan nahi dute balioek? Suposatu orain X $Poisson(4)$ dela. Lortu ausazko lagin bat eta esan nola aldatu diren balioak.

Balioak esperimentu baten hamar lagin dira

```r
rpois(10, lambda=4)
```

```
##  [1] 4 6 3 2 2 3 4 6 3 3
```

Laginak handiagoak dira

Ariketak

1. ariketa X zorizko aldagaiak ondorengo dentsitate-funtzioa du:

$$f(x) = \left\{ \begin{array}{cc} 1 - |1-x|, & 0 < x < 2 \\ 0, & \text{bestela} \end{array} \right.$$

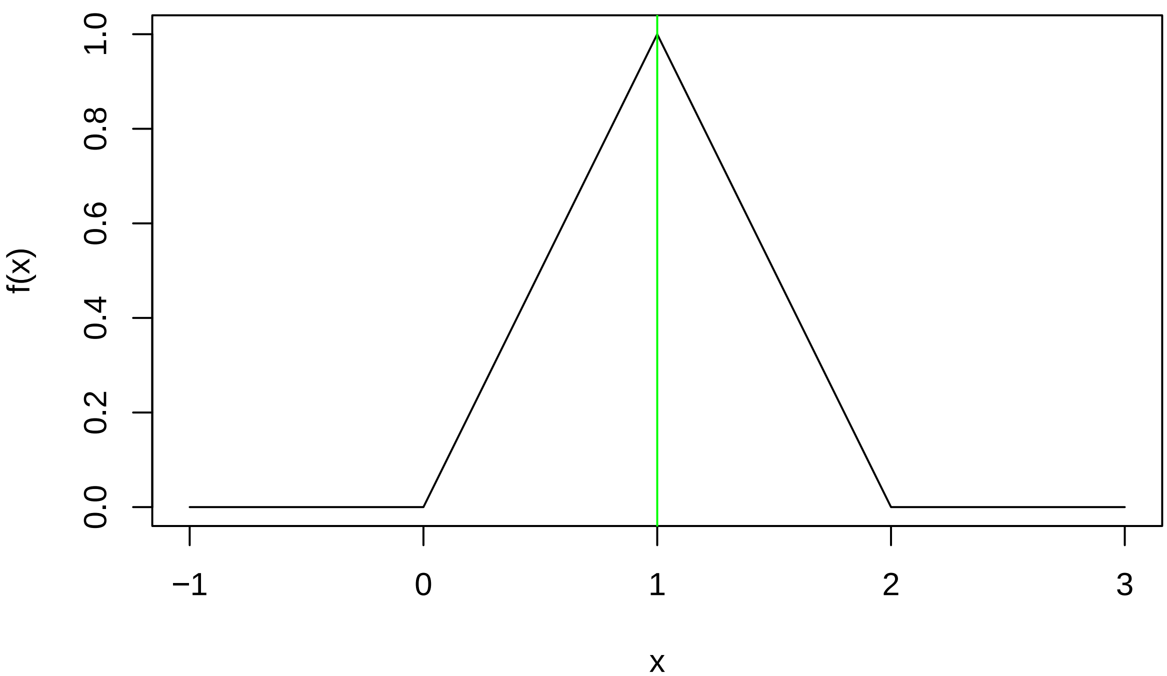
- Definitu eta irudikatu dentsitate-funtzioa. Horren arabera, zein da EX? Adierazi itxaropena aurreko grafikoan lerro bertikal berde bat gehituz.

```
f <- function(x) {
  if (x <= 0 | x >= 2) return(0)
  return(1-abs(1-x))
}

x <- seq(-1, 3, 0.1)
f <- Vectorize(f)
plot(x, f(x), type = "l")

#EX = 1 da

abline(v = 1, col = "green")
```



- Kalkulatu dagokion banaketa-funtzioa eskuz eta marraztu. Horren arabera, zein da Q_3?

$$F(x) = \int_{-\infty}^{x} f(t)\,dt = \int_{0}^{x} f(t)\,dt = \begin{cases} 0 & : x < 0 \\ \int_0^x t\,dt & : 0 \leq x \leq 1 \\ \int_0^1 t\,dt + \int_1^x (2-t)\,dt & : 1 < x \leq 2 \\ 1 & : 2 < x \end{cases} = \begin{cases} 0 & : x < 0 \\ \frac{x^2}{2} & : 0 \leq x \leq 1 \\ 2x - \frac{x^2}{2} - 1 & : 1 < x \leq 2 \\ 1 & : 2 < x \end{cases}$$

Honen arabera:

$$P(X \le a) = \frac{3}{4} \implies F(a) = \frac{3}{4} \implies 2a - \frac{a^2}{2} - 1 = \frac{3}{4} \implies a = 2 - \frac{1}{\sqrt{2}} \simeq 1.2929$$

- Begiratu Rko laguntzan `integrate` funtzioari buruzko informazioa. Erabili funtzio hau EX kalkulatzeko eta banaketa-funtzioa zuzenean lortzeko (eskuz kalkulatu gabe).

```
#EX
integrate(function(x){x * f(x)}, 0, 2)
```

```
## 1 with absolute error < 1.1e-14
```

```
banaketaF <- function(y){
  integrate(f, 0, y)
}

#F(Q_3)
Q_3 = 2 - 1/sqrt(2)
banaketaF(Q_3)
```

```
## 0.7499994 with absolute error < 6.8e-06
```

2. ariketa Zorizko aldagai jarraituen hainbat **banaketa ezagunekin** lan egingo dugu. Kasu bakoitzean:

Banaketak		
$X \sim \mathcal{U}(0,1)$	$X \sim Beta(1,1)$	$X \sim Beta(15,1)$
$X \sim Beta(15,15)$	$X \sim Exp(5)$	$X \sim Gamma(5,10)$
$X \sim N(5,10)$	$X \sim N(5,1)$	$X \sim t_2$
$X \sim t_{30}$	$X \sim \chi^2_1$	$X \sim \chi^2_6$

- Marraztu dentsitate-funtzioa. Horren arabera ondorioztatu gutxi-gora-behera EXren balioa. Egiaztatu teoriako emaitzekin.
- Kalkulatu x_0 non $P(X \le x_0) = 0.9$ den.
- Kalkulatu $P(X < a)$, zuk aukeratutako $a < EX$ balio baterako.

$$X \sim \mathcal{U}(0,1)$$

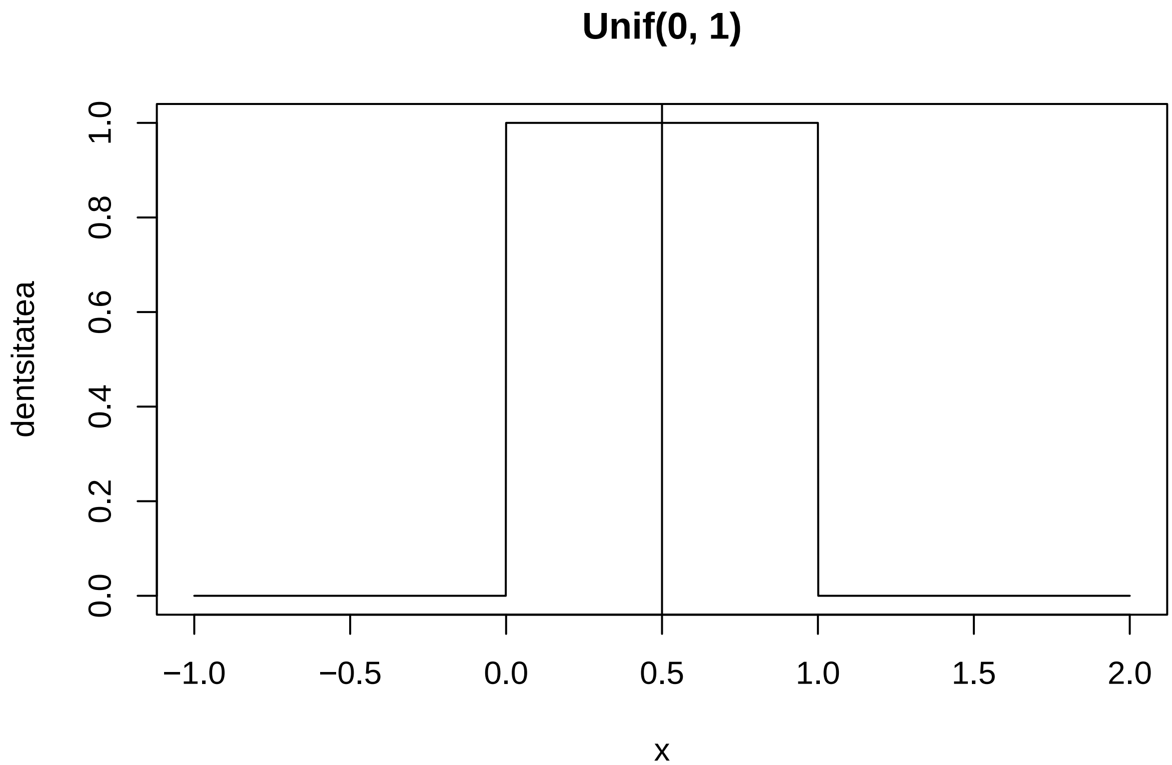
$$f(x) = \begin{cases} \frac{1}{b-a} & , \quad a \le x \le b \\ 0 & , \quad \text{bestela} \end{cases}$$

```
f <- function(x){
  if (x < 0 || x > 1) return(0)
  1
}

f <- Vectorize(f)
```

```r
x <- seq(-1, 2, length.out = 3001)
dentsitatea <- f(x)

plot(x, dentsitatea, t="l", main = "Unif(0, 1)")
abline(v=0.5)
```




```r
# x_0 | P(X<= x_0) = 0.9
qunif(0.9, 0, 1)
```

```
## [1] 0.9
```

```r
# P(X < a) | a < EX
# a = 0.3
punif(0.3)
```

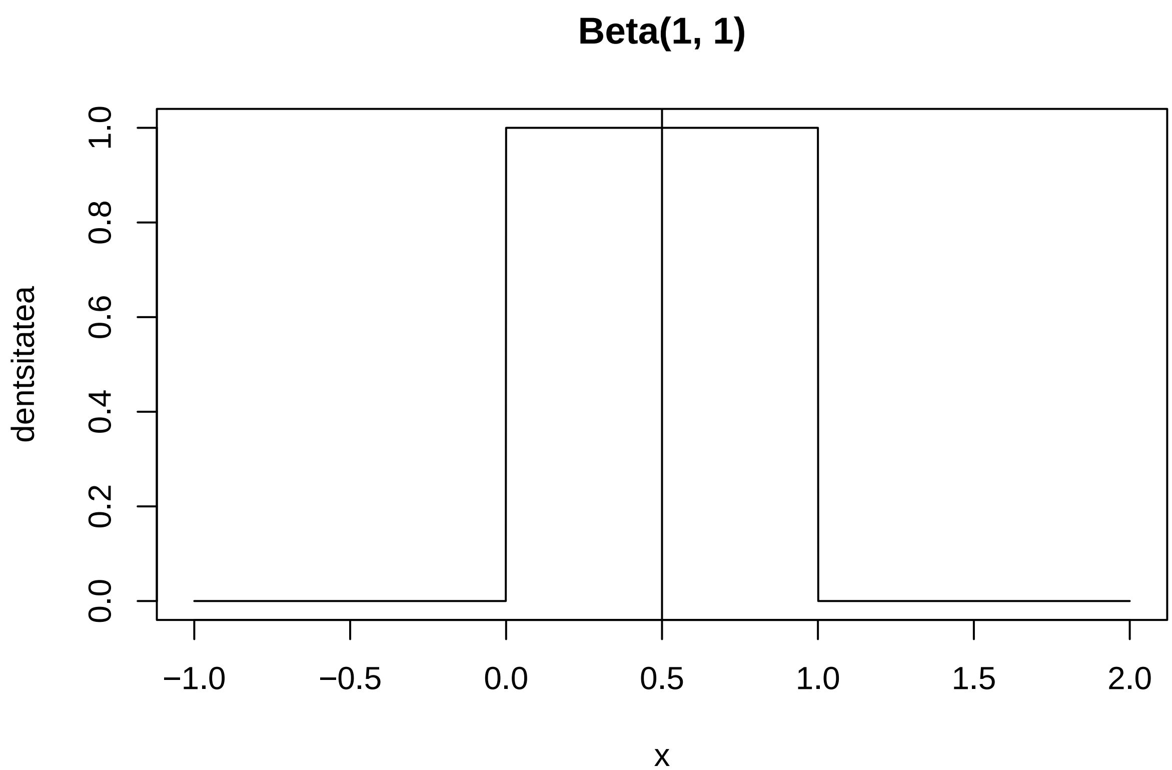
```
## [1] 0.3
```

$EX = \frac{a+b}{2}$ eta $VAR(X) = \frac{(b-a)^2}{12}$

$$X \sim Beta(1, 1)$$

```r
x <- seq(-1, 2, length.out = 3001)
dentsitatea <- dbeta(x, 1, 1)
```

```r
plot(x, dentsitatea, t="l", main = "Beta(1, 1)")
abline(v=0.5)
```




```r
# x_0 | P(X<= x_0) = 0.9
qbeta(0.9, 1, 1)
```

```
## [1] 0.9
```

```r
# P(X < a) | a < EX
# a = 0.3
pbeta(0.3, 1, 1)
```

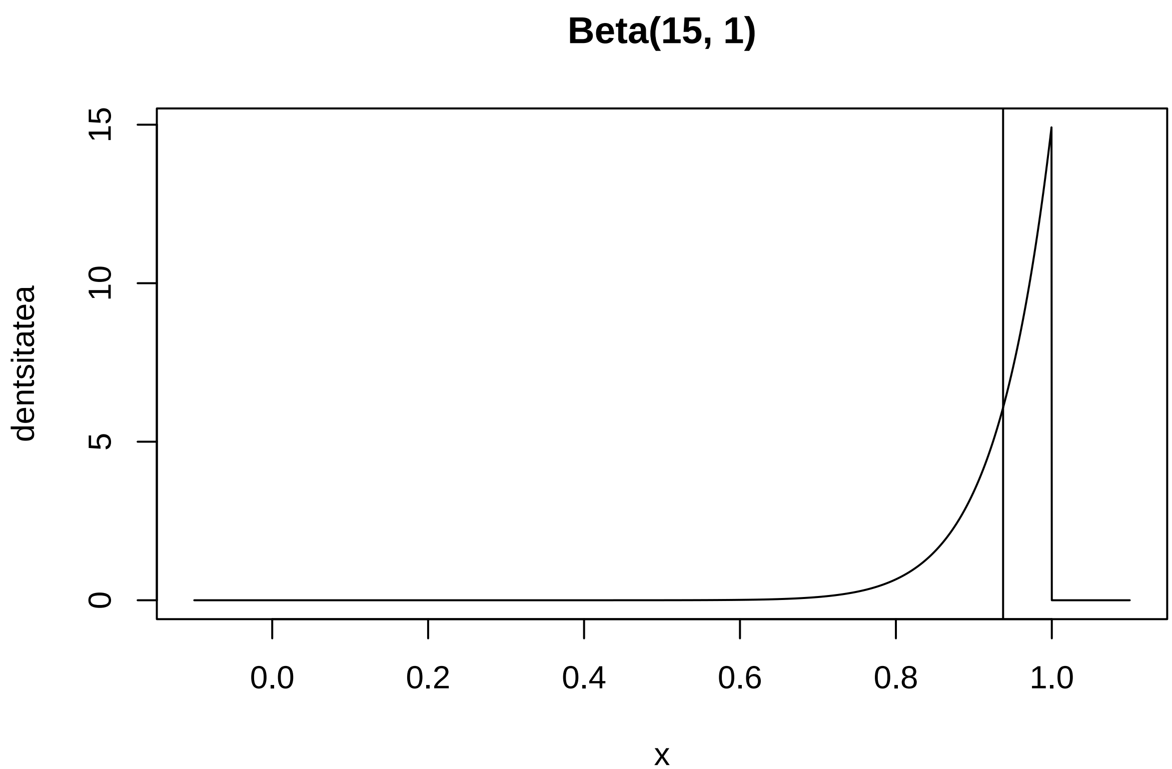
```
## [1] 0.3
```

$EX = \frac{\alpha}{\alpha+\beta}$ eta $VAR(X) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$

$$X \sim Beta(15, 1)$$

```r
x <- seq(-0.1, 1.1, length.out = 3001)
dentsitatea <- dbeta(x, 15, 1)

plot(x, dentsitatea, t="l", main = "Beta(15, 1)")
abline(v=15/16)
```




```
# x_0 | P(X<= x_0) = 0.9
qbeta(0.9, 15, 1)
```

```
## [1] 0.9930006
```

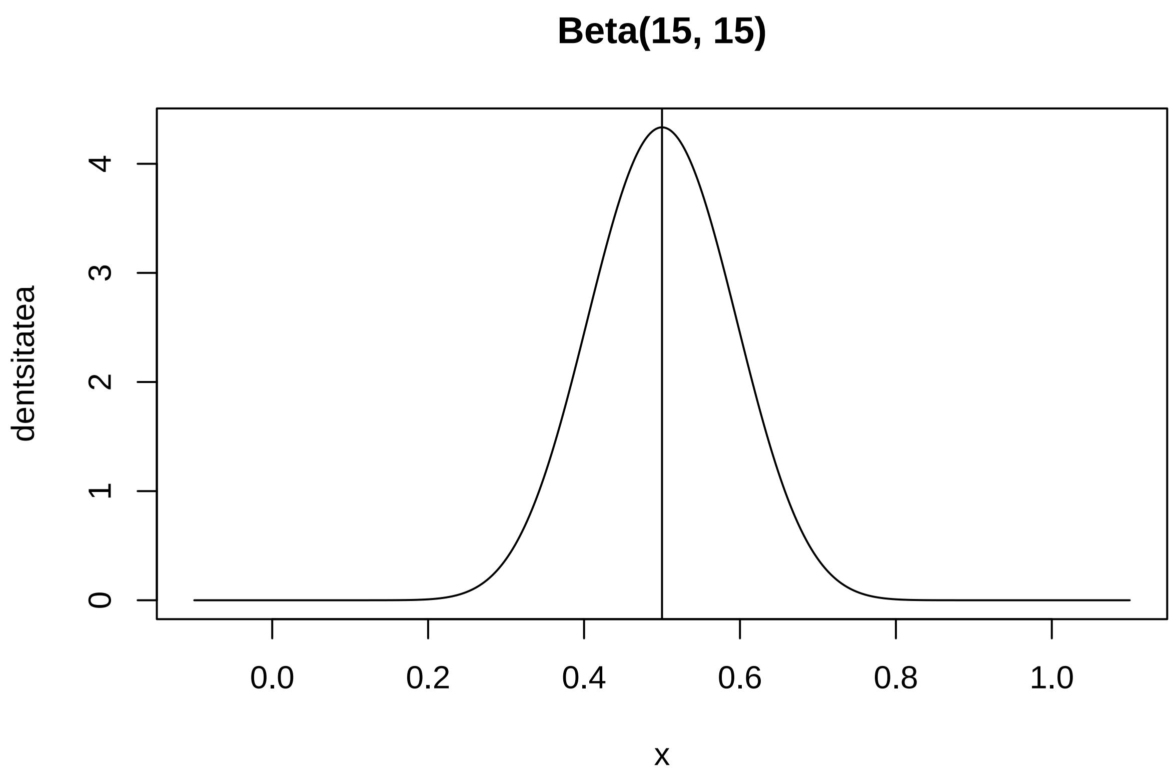
```
# P(X < a) | a < EX
# a = 14/16
pbeta(14/16, 15, 1)
```

```
## [1] 0.1349338
```

$$X \sim Beta(15, 15)$$

```
x <- seq(-0.1, 1.1, length.out = 3001)
dentsitatea <- dbeta(x, 15, 15)

plot(x, dentsitatea, t="l", main = "Beta(15, 15)")
abline(v=15/30)
```




```r
# x_0 | P(X<= x_0) = 0.9
qbeta(0.9, 15, 15)
```

```
## [1] 0.6163407
```

```r
# P(X < a) | a < EX
# a = 0.3
pbeta(0.3, 15, 15)
```

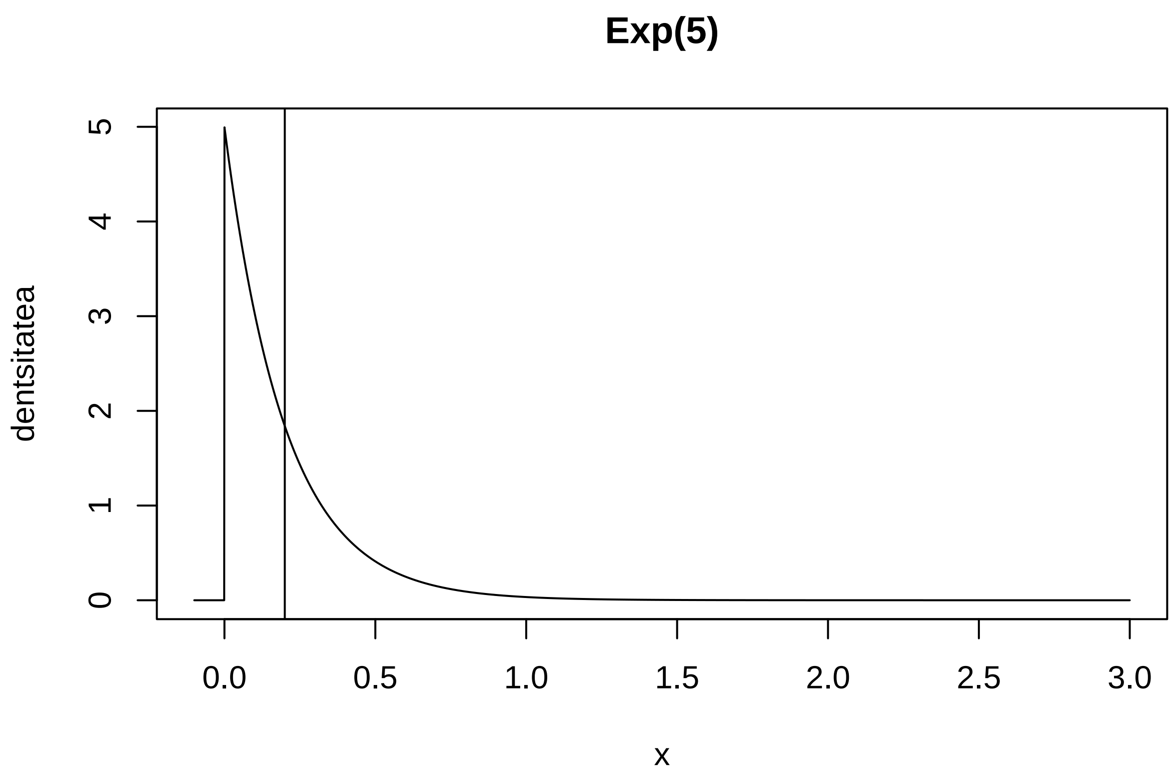
```
## [1] 0.01165383
```

$$X \sim Exp(5)$$

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & , \quad x \geq 0 \\ 0 & , \quad \text{bestela} \end{cases}$$

```r
x <- seq(-0.1, 3, length.out = 3001)
dentsitatea <- dexp(x, 5)

plot(x, dentsitatea, t="l", main = "Exp(5)")
abline(v=1/5)
```




```r
# x_0 | P(X<= x_0) = 0.9
qexp(0.9, 5)
```

```
## [1] 0.460517
```

```r
# P(X < a) | a < EX
# a = 1/6
pexp(1/6, 5)
```

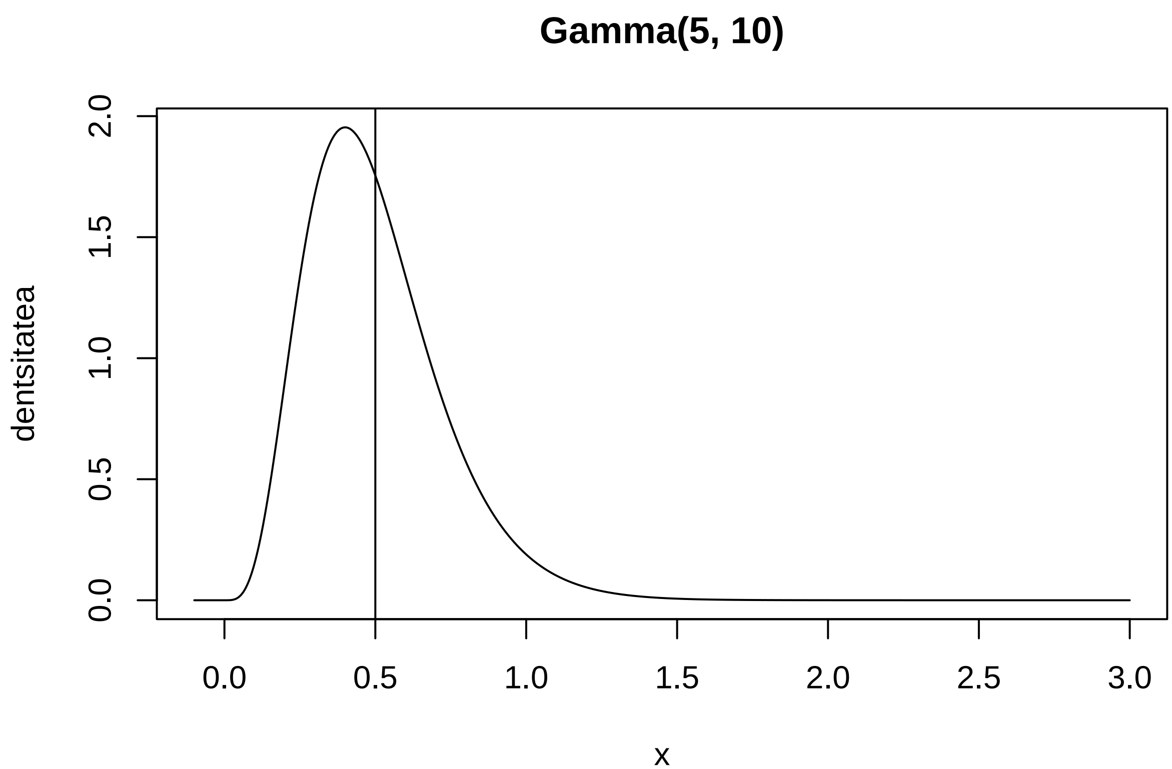
```
## [1] 0.5654018
```

$EX = \frac{1}{\lambda}$ eta $VAR(X) = \frac{1}{\lambda^2}$

$$X \sim Gamma(5, 10)$$

```r
x <- seq(-0.1, 3, length.out = 3001)
dentsitatea <- dgamma(x, 5, 10)

plot(x, dentsitatea, t="l", main = "Gamma(5, 10)")
abline(v=1/2)
```




```r
# x_0 | P(X<= x_0) = 0.9
qgamma(0.9, 5, 10)
```

```
## [1] 0.799359
```

```r
# P(X < a) | a < EX
# a = 0.3
pgamma(0.3, 5, 10)
```

```
## [1] 0.1847368
```

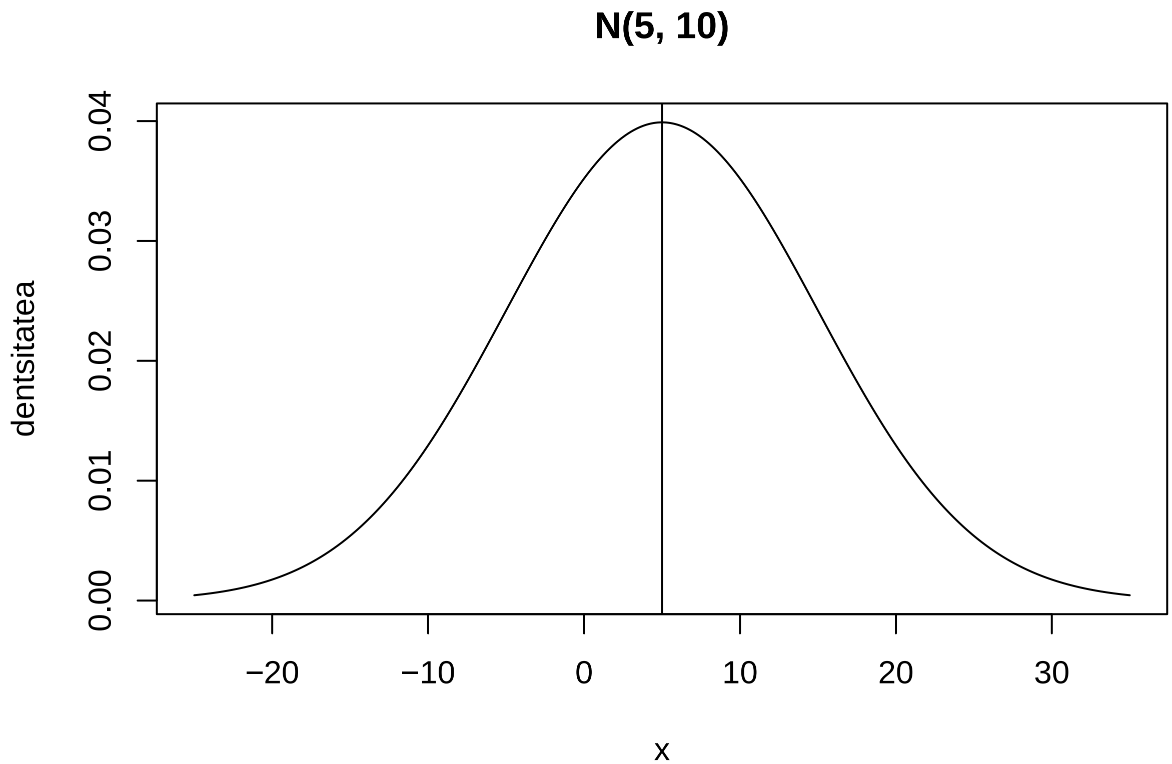
$EX = \frac{\alpha}{\beta}$ eta $VAR(X) = \frac{\alpha}{\beta^2}$

$$X \sim N(5, 10)$$

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

```r
x <- seq(-25, 35, length.out = 3001)
dentsitatea <- dnorm(x, 5, 10)

plot(x, dentsitatea, t="l", main = "N(5, 10)")
abline(v=5)
```




```r
# x_0 | P(X<= x_0) = 0.9
qnorm(0.9, 5, 10)
```

```
## [1] 17.81552
```

```r
# P(X < a) | a < EX
# a = 4
pnorm(4, 5, 10)
```

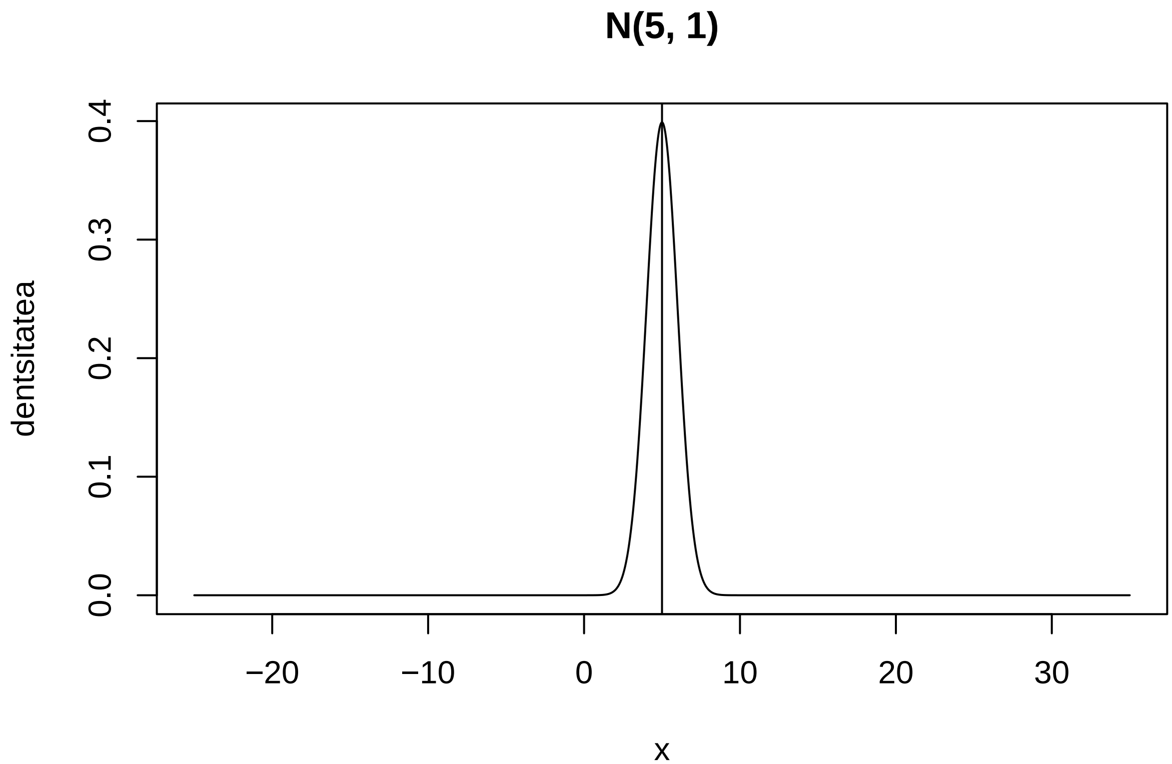
```
## [1] 0.4601722
```

$EX = \mu$ eta $VAR(X) = \sigma^2$

$$X \sim N(5, 1)$$

```r
x <- seq(-25, 35, length.out = 3001)
dentsitatea <- dnorm(x, 5, 1)

plot(x, dentsitatea, t="l", main = "N(5, 1)")
abline(v=5)
```




```r
# x_0 | P(X<= x_0) = 0.9
qnorm(0.9, 5, 1)
```

```
## [1] 6.281552
```

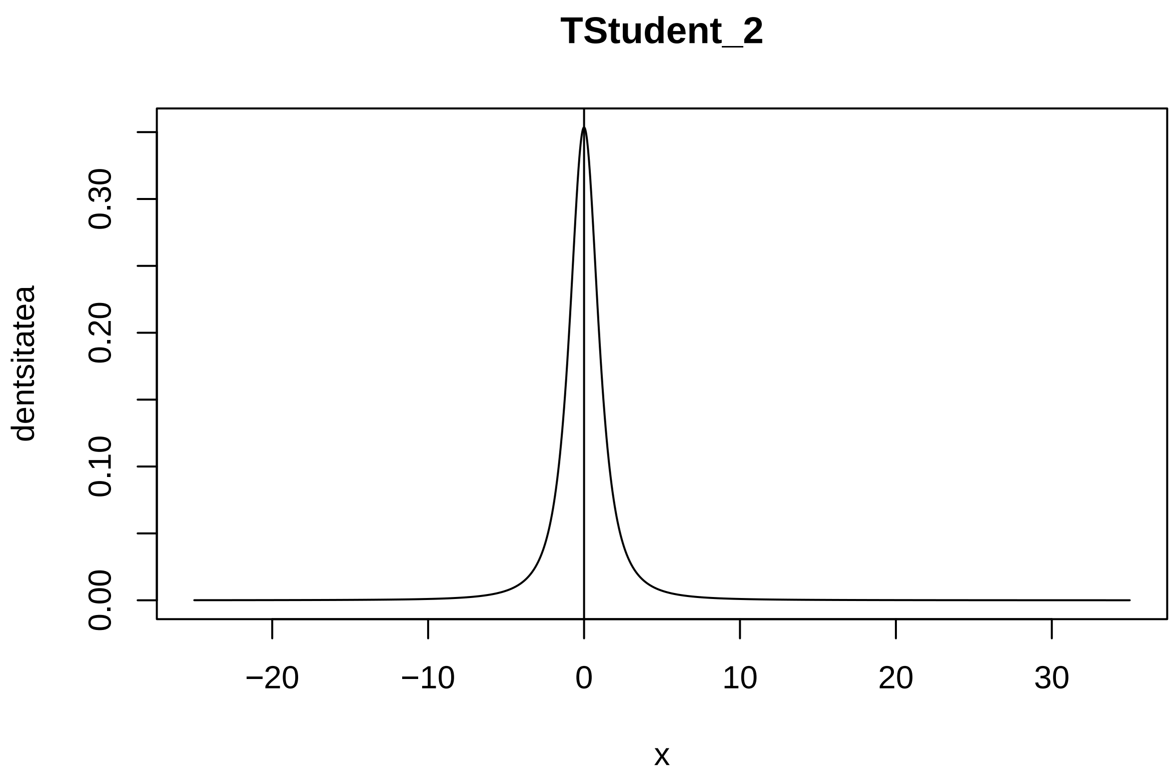
```r
# P(X < a) | a < EX
# a = 4
pnorm(4, 5, 1)
```

```
## [1] 0.1586553
```

$$X \sim t_2$$

```r
x <- seq(-25, 35, length.out = 3001)
dentsitatea <- dt(x, 2)

plot(x, dentsitatea, t="l", main = "TStudent_2")
abline(v=0)
```




```r
# x_0 | P(X<= x_0) = 0.9
qt(0.9, 2)
```

```
## [1] 1.885618
```

```r
# P(X < a) | a < EX
# a = -1
pt(-1, 2)
```

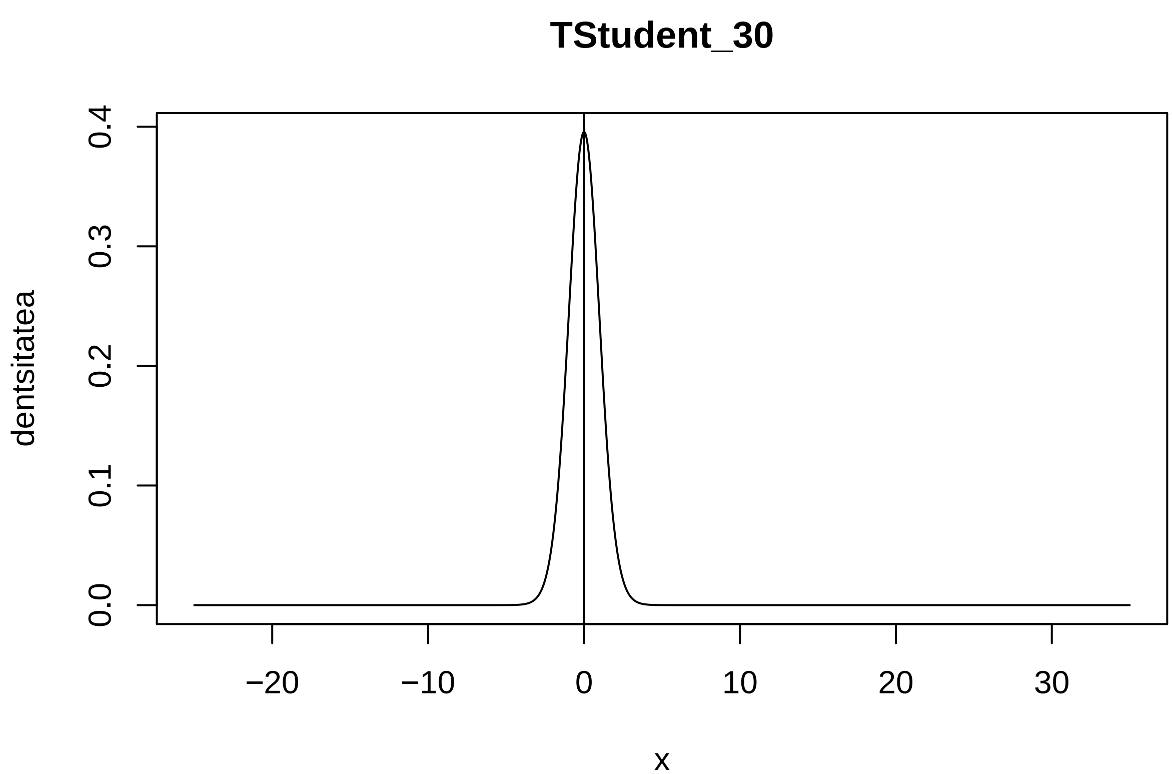
```
## [1] 0.2113249
```

$EX = 0$ eta $VAR(X) = \frac{n}{n-2}$

$X \sim t_{30}$

```r
x <- seq(-25, 35, length.out = 3001)
dentsitatea <- dt(x, 30)

plot(x, dentsitatea, t="l", main = "TStudent_30")
abline(v=0)
```




```r
# x_0 | P(X<= x_0) = 0.9
qt(0.9, 30)
```

```
## [1] 1.310415
```

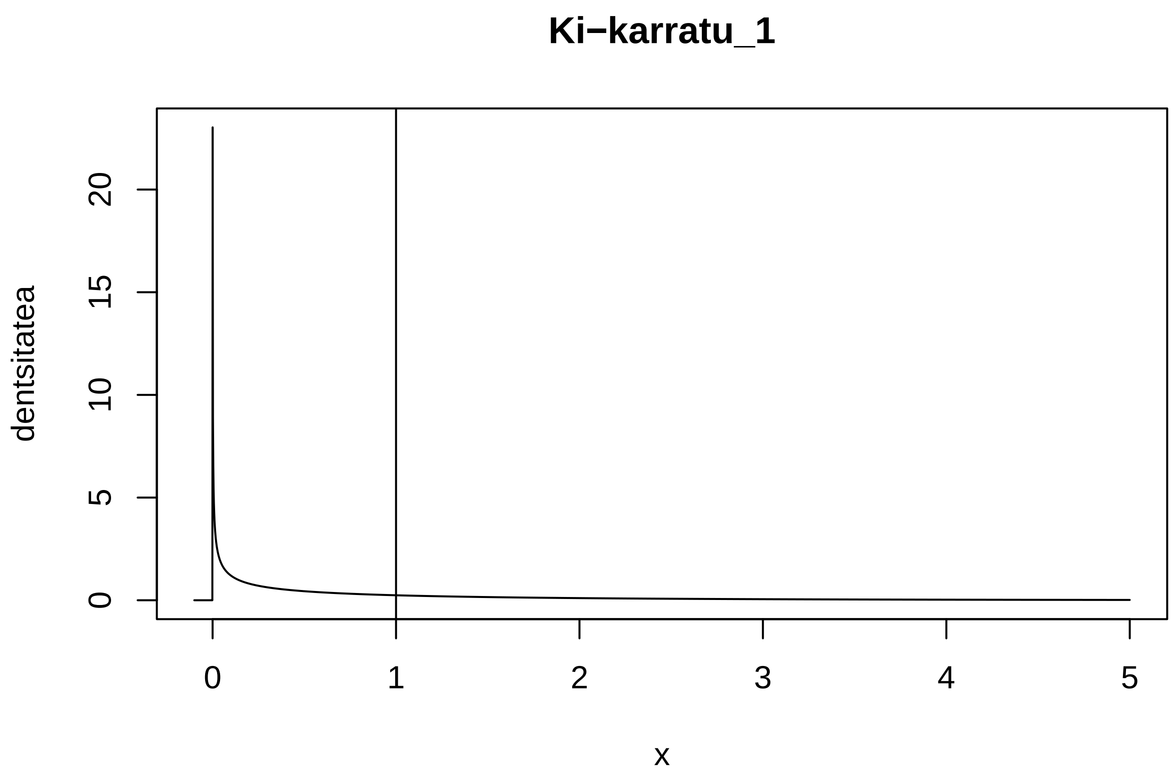
```r
# P(X < a) | a < EX
# a = -1
pt(-1, 30)
```

```
## [1] 0.1626543
```

$$X \sim \chi^2_1$$

```r
x <- seq(-0.1, 5, length.out = 3001)
dentsitatea <- dchisq(x, 1)

plot(x, dentsitatea, t="l", main = "Ki-karratu_1")
abline(v=1)
```




```r
# x_0 | P(X<= x_0) = 0.9
qchisq(0.9, 1)
```

```
## [1] 2.705543
```

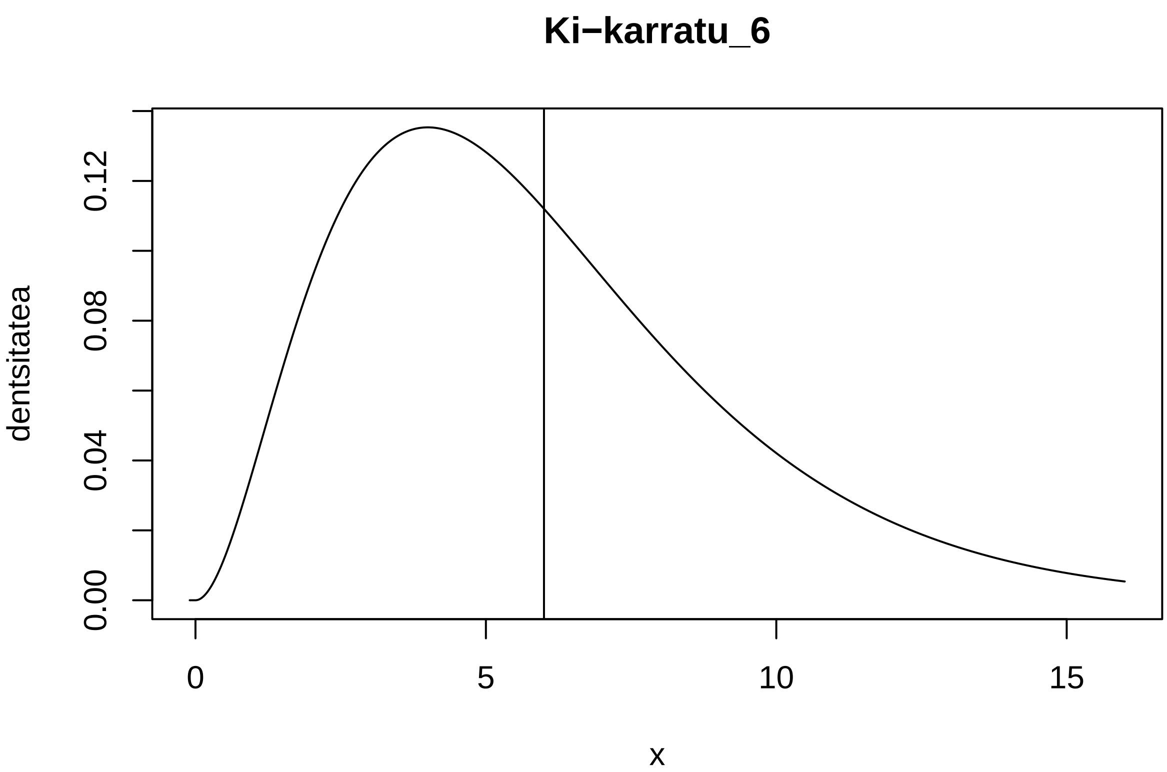
```r
# P(X < a) | a < EX
# a = 0.1
pchisq(0.1, 1)
```

```
## [1] 0.2481704
```

$$X \sim \chi^2_6$$

```r
x <- seq(-0.1, 16, length.out = 3001)
dentsitatea <- dchisq(x, 6)

plot(x, dentsitatea, t="l", main = "Ki-karratu_6")
abline(v=6)
```




```r
# x_0 | P(X<= x_0) = 0.9
qchisq(0.9, 6)
```

```
## [1] 10.64464
```

```r
# P(X < a) | a < EX
# a = 5
pchisq(5, 6)
```

```
## [1] 0.4561869
```

$EX = n$ eta $VAR(X) = 2n$