

# Exercise 6 - Selected distributions of a continuous random variable

Martina Litschmannová, Adéla Vrtková, Michal Béréš

## Overview of distributions and their functions

### Introduction: Probability Density, Distribution Function and Quantile Function

#### Probability density

- starts with the letter **d**:  $p = d \dots(x, \dots)$

#### Distribution function

- starts with the letter **p**:  $p = P(X < x): p = p \dots(x, \dots)$

#### Quantile function

- starts with the letter **q**: find  $x$  for the given  $p$ :  $p = F(x) \rightarrow x = F^{-1}(p): x = q \dots(p, \dots)$

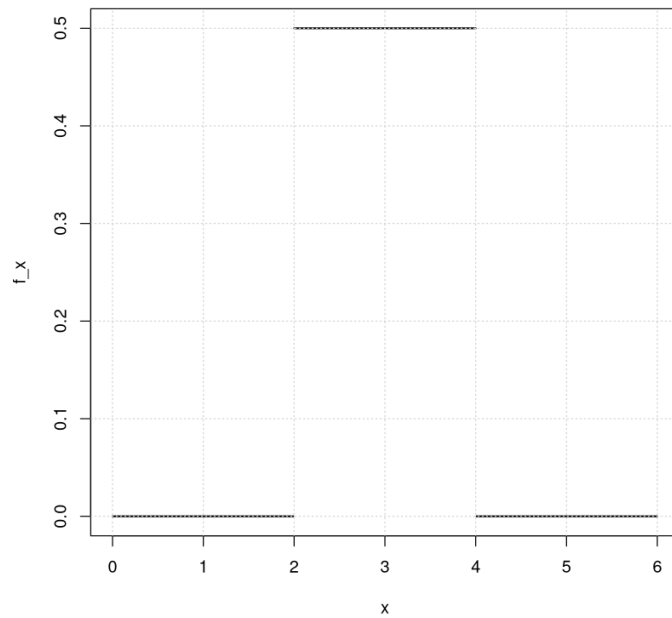
### Uniform distribution: $X \sim U(a, b)$

- the random variable takes on only values greater than  $a$  and less than  $b$
- all values have the same occurrence density  $\rightarrow$  the probability density is constant between  $a$  and  $b$ , elsewhere zero

```
In [1]: # Probability density f(x)
a = 2   # from where
b = 4   # up to
x = 3
dunif(x, a, b)
```

0.5

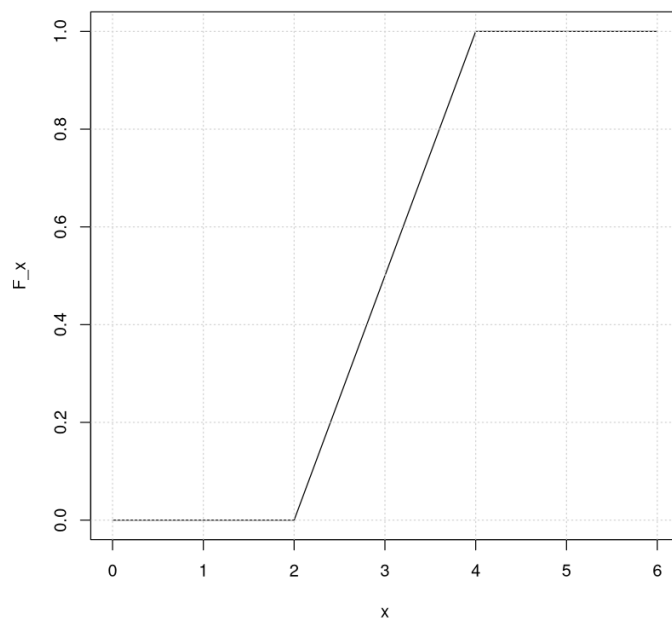
```
In [2]: # we plot the probability density
x = seq(from = 0, to = 6, by = 0.01)
f_x = dunif(x, a, b)
plot(x, f_x, cex = 0.1) # cex is the size of the markers
grid()
```



```
In [3]: # Distribution function  $F(x)=P(X<x)$ 
a = 2 # from where
b = 4 # where
x = 3
punif(x, a, b)
```

0.5

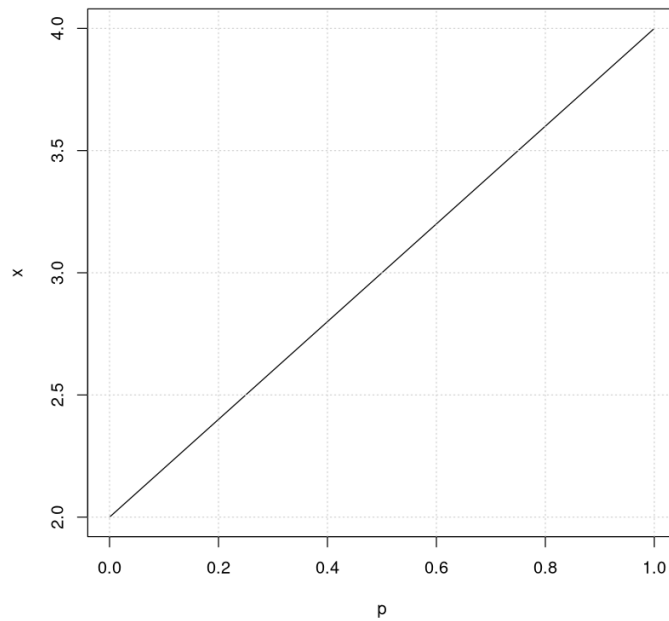
```
In [4]: # let's plot the Distribution function
x = seq(from = 0, to = 6, by = 0.01)
F_x = punif(x, a, b)
plot(x, F_x, type = 'l')
grid()
```



```
In [5]: # quantile function  $F^{-1}(p)=x: P(X<x)=p$ 
a = 2 # from where
b = 4 # where
p = 0.75
qunif(p, a, b)
```

3.5

```
In [6]: # plot - quantile function  $F^{-1}(p)=x$ 
p = seq(from=0, to=1, by=0.01)
x = qunif(p, a, b)
plot(p, x, type = 'l')
grid()
```



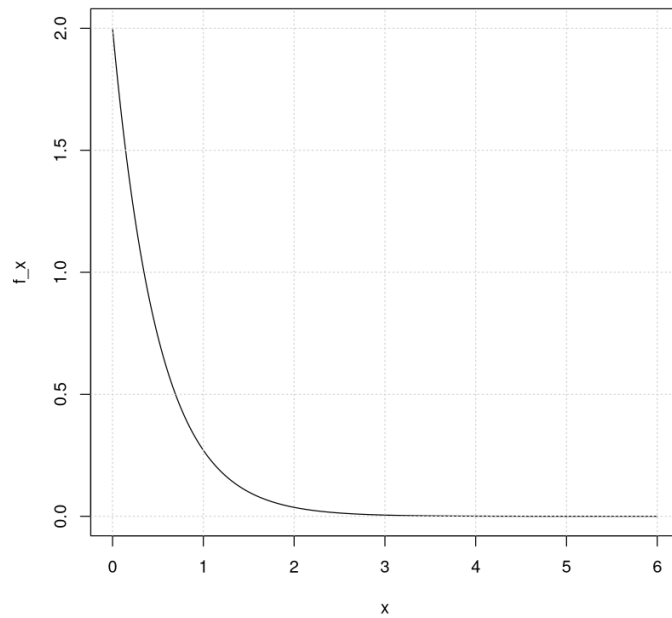
## Exponential distribution: $X \sim \text{Exp}(\lambda)$

- time to 1st event, time between events(only in the period of stable life - Poisson process)
- the parameter  $\lambda$  is the same as in the Poisson distribution
- the mean value is:  $E(X) = 1/\lambda$

```
In [7]: # Probability density  $f(x)$ 
lambda = 2
x = 1
dexp(x, lambda)
```

0.270670566473225

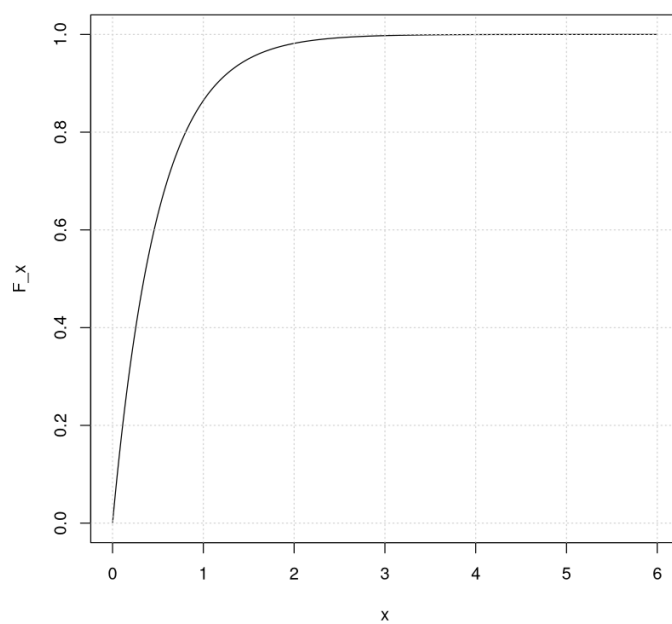
```
In [8]: # we plot the probability density
x = seq(from = 0, to = 6, by = 0.01)
f_x = dexp(x, lambda)
plot(x, f_x, type='l')
grid()
```



```
In [9]: # Distribution function  $F(x)=P(X<x)$ 
lambda = 2
x = 1
pexp(x, lambda)
```

0.864664716763387

```
In [10]: # let's draw the Distribution function
x = seq(from = 0, to = 6, by = 0.01)
F_x = pexp(x, lambda)
plot(x, F_x, type = 'l')
grid()
```

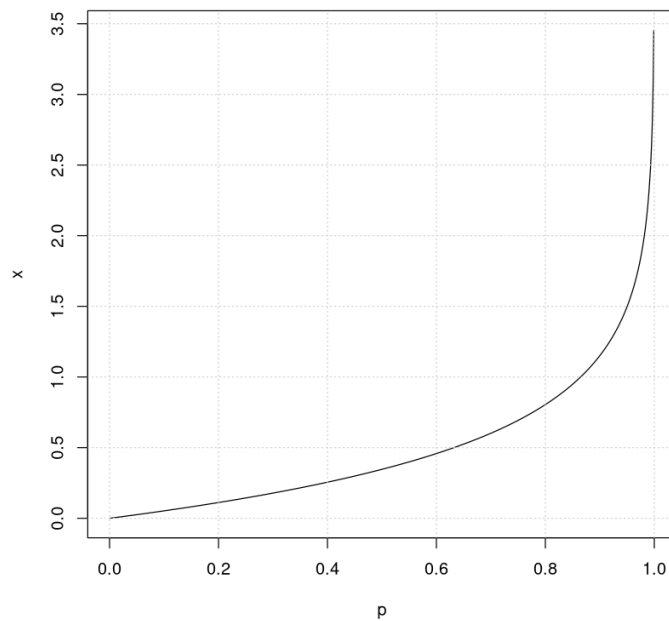


```
In [11]: # quantile function  $F^{-1}(p)=x: P(X<x)=p$ 
lambda = 2
p = 0.5
qexp(p, lambda)
```

0.346573590279973

```
In [12]: # plot- quantile function  $F^{-1}(p)=x$ 
```

```
p = seq(from=0, to=1, by=0.001)
x = qexp(p, lambda)
plot(p, x, type = 'l')
grid()
```



```
In [13]: qexp(0.9, lambda)
```

1.15129254649702

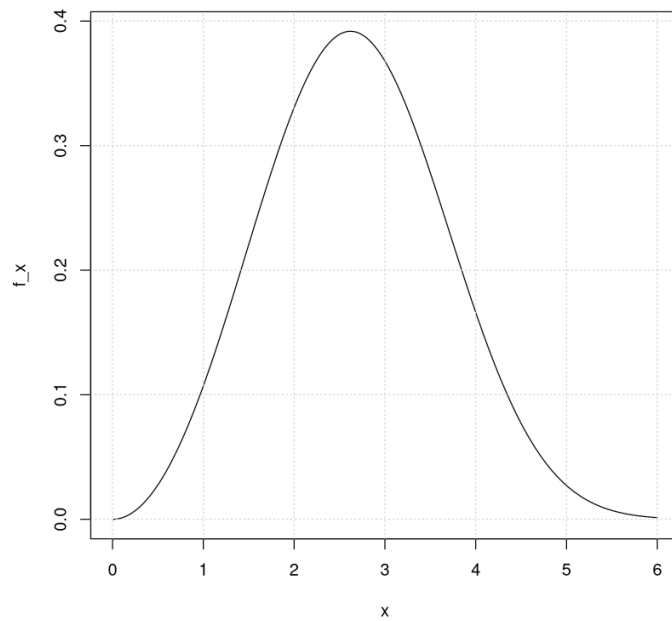
Weibull distribution:  $X \sim W(\theta, \beta)$

- time until the 1st event(faults)(suitable choice of  $\beta$  allows use in any period of fault intensity)
- extension of exponential distribution  $\text{Exp}(\lambda)=W(\Theta=1/\lambda, \beta=1)$

```
In [16]: # Probability density f(x)
theta = 3 # equivalent of 1/lambda in exp. distribution
beta = 3 # beta=1 ->exponential distribution
x = 5
dweibull(x,shape=beta, scale=theta)
```

0.0271065906811606

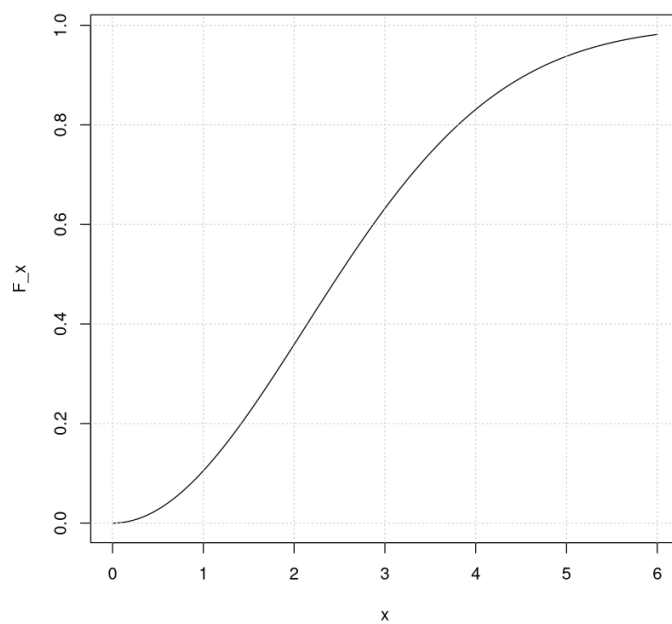
```
In [17]: # we plot the probability density
x = seq(from = 0, to = 6, by = 0.01)
f_x = dweibull(x,shape=beta, scale=theta)
plot(x, f_x, type='l')
grid()
```



```
In [18]: # Distribution function  $F(x)=P(X<x)$ 
theta = 3 # equivalent of  $1/\lambda$  in exp. distribution
beta = 2 #  $\beta=1 \rightarrow$  exponential distribution
x = 5
pweibull(x,shape=beta, scale=theta)
```

0.937823475977884

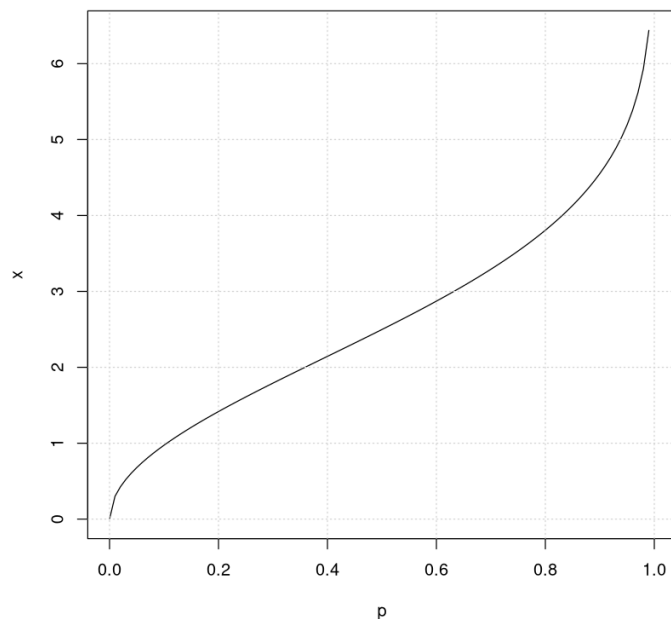
```
In [19]: # we draw the Distribution function
x = seq(from = 0, to = 6, by = 0.01)
F_x = pweibull(x,shape=beta, scale=theta)
plot(x, F_x, type = 'l')
grid()
```



```
In [20]: # quantile function  $F^{-1}(p)=x: P(X<x)=p$ 
theta = 3 # equivalent of  $1/\lambda$  in exp. distribution
beta = 2 #  $\beta=1 \rightarrow$  exponential distribution
p = 0.5
qweibull(p,shape=beta, scale=theta)
```

2.49766383347309

```
In [21]: # plot - quantile function  $F^{-1}(p)=x$ 
p = seq(from=0, to=1, by=0.01)
x = qweibull(p, shape=beta, scale=theta)
plot(p, x, type = 'l')
grid()
```



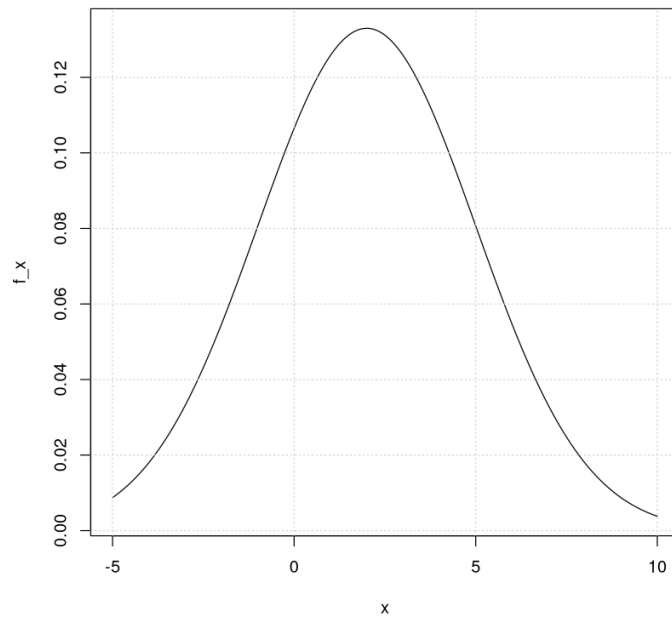
## Normal distribution: $X \sim N(\mu, \sigma^2)$

- distribution modeling eg. measurement errors, sum/average behavior of many other random variables
- see Central limit theorem
- $\mu$  is the average distribution value:  $E(X) = \mu$
- $\sigma$  is the directly standard deviation of the distribution:  $D(X) = \sigma^2$
- with parameters  $\mu = 0, \sigma = 1$  is called normalized Normal distribution

```
In [22]: # Probability density f(x)
mu = 2
sigma = 3
x = 4
dnorm(x, mean=mu, sd=sigma)
```

0.106482668507451

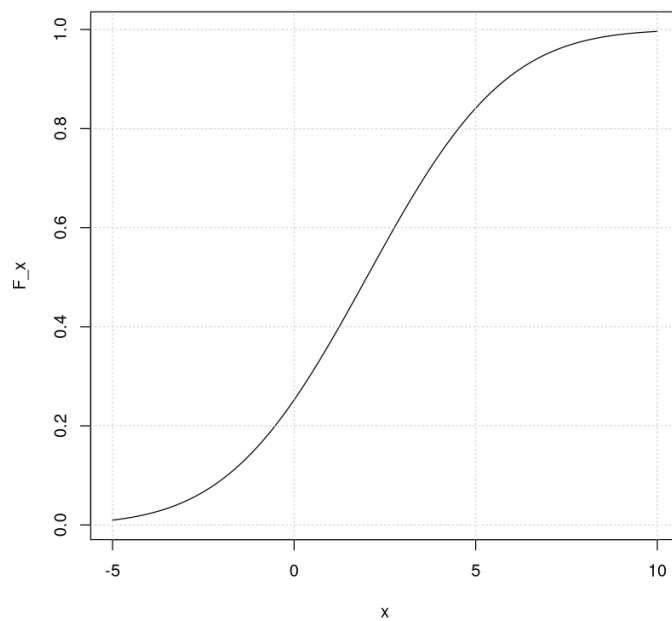
```
In [23]: # we plot the probability density
x = seq(from = -5, to = 10, by = 0.01)
f_x = dnorm(x, mean=mu, sd=sigma)
plot(x, f_x, type='l')
grid()
```



```
In [24]: # Distribution function  $F(x)=P(X<x)$ 
mu = 2
sigma = 3
x = 4
pnorm(x, mean=mu, sd=sigma)
```

0.747507462453077

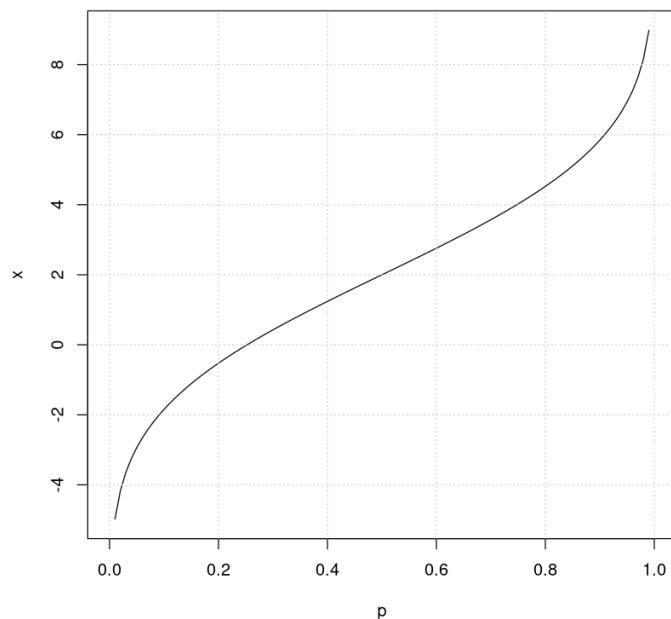
```
In [25]: # let's draw the Distribution function
x = seq(from = -5, to = 10, by = 0.01)
F_x = pnorm(x, mean=mu, sd=sigma)
plot(x, F_x, type = 'l')
grid()
```



```
In [26]: # quantile function  $F^{-1}(p)=x: P(X<x)=p$ 
mu = 2
sigma = 3
p = 0.5
qnorm(p, mean=mu, sd=sigma)
```



```
In [27]: # plot - quantile function  $F^{-1}(p)=x$ 
p = seq(from=0, to=1, by=0.01)
x = qnorm(p, mean=mu, sd=sigma)
plot(p, x, type = 'l')
grid()
```



## Examples

### Example 1.

Height in the population of boys aged 3.5-4 years has a normal distribution with a mean value of 102 cm and a standard deviation of 4.5 cm. Determine what percentage of boys at that age are less than or equal to 93 cm in height.

```
In [28]: # X... height of boys aged 3.5 to 4 years(cm)
# X~N(mu=102, sd=4.5)

mu = 102
sigma = 4.5

# P(X<=93)=F(93)
pnorm(93, mean=mu, sd=sigma)
```

0.0227501319481792

### Example 2.

The average life of a machine part is 30,000 hours. Assume that the component is in a period of stable life. Specify:

```
In [29]: # X... component life(h)
# X~Exp(lambda), where E(X)=1/lambda

lambda = 1/30000
```

a)

probability that the component will not last more than 2,000 hours,

```
In [30]: # a) P(X<2000)=F(2000)
pexp(2000, lambda)
```

0.0644930149683823

b)

probability that the component will last more than 35,000 hours,

```
In [31]: # b)  $P(X > 35000) = 1 - F(35000)$   
1 - pexp(35000, lambda)
```

0.311403223914598

c)

time until 95% of the components fail.

```
In [32]: # c)  $P(X < t) = 0.95 \rightarrow F(t) = 0.95 \rightarrow t \dots$  95% quantile  
qexp(0.95, lambda)
```

89871.9682066197

## Example 3.

The production facility fails on average once every 2000 hours. The quantity Y representing the time before the fault has an exponential distribution. Determine the time T0 so that the probability that the device will run longer than T0 is 0.99.

```
In [33]: # X... fault waiting time(h)  
#  $X \sim \text{Exp}(\lambda)$ , where  $E(X) = 1/\lambda$   
lambda = 1/2000  
  
#  $P(X > t) = 0.99 \rightarrow 1 - F(t) = 0.99 \rightarrow F(t) = 0.01 \rightarrow t \dots$  1% quant.  
qexp(0.01, lambda)
```

20.1006717070029

## Example 4.

The measurement results are corrupted only with a normally distributed error with zero mean and a standard deviation of 3 mm. What is the probability that there will be an error in the interval(0 mm; 2.4 mm) at least once in 3 measurements?

```
In [34]: # Y measurement error size(mm)  
#  $Y \sim N(\mu=0, \sigma=3)$   
  
mu = 0  
sigma = 3  
  
# pp... true that the measurement error will be in int. 0.0-2.4mm  
pp = pnorm(2.4, mean=mu, sd=sigma) - pnorm(0, mean=mu, sd=sigma)  
pp
```

0.288144601416603

```
In [35]: # X... number of measurement errors in int. 0 mm -2.4 mm in 3 measures.  
#  $X \sim \text{Bi}(n=3, p=pp)$   
n = 3  
p = pp  
  
#  $P(X \geq 1) = 1 - P(X=0)$   
1 - dbinom(0, n, p)
```

0.639275741801798

## Example 5.

An average of 25 users per hour log on to a large computer network. Determine the probability that:

a)

no one logs in during 14:30 - 14:36,

```
In [36]: # X... number of users logged in in 6 minutes
# X~Po(lt=2.5)

lambda = 25/60
t = 6
lt = lambda*t

# P(X=0)
dpois(0, lt)
```

0.0820849986238988

b)

2-3 minutes will elapse before the next login.

```
In [37]: # Y... time until next login
# Y~Exp(lambda=25/60), where E(X)=1/lambda
lambda = 25/60

# P(2<Y<3)=F(3) - F(2)
pexp(3, lambda) - pexp(2, lambda)
```

0.148093411646888

```
In [38]: lambda = 25

# P(2<Y<3)=F(3) - F(2)
pexp(3/60, lambda) - pexp(2/60, lambda)
```

0.148093411646888

c)

Specify the maximum length of the time interval so that the probability that no one will log in is at least 0.90.

```
In [39]: # P(Y>t)=0.90 -> 1-F(t)=0.90 -> F(t)=0.10 -> t... 10% kv.

qexp(0.10, lambda)*60
```

0.252865237578783

## Example 6.

The random variable X has a normal distribution  $N(\mu; \sigma)$ . Specify:

a)

$P(\mu - 2\sigma < X < \mu + 2\sigma)$ ,

```
In [40]: # P(μ - 2σ < X < μ + 2σ) = F(μ + 2σ) - F(μ - 2σ)
# X~N(μ, σ)
# it doesn't matter what values we choose
mu = -105.5447
sigma = 2.654

pnorm(mu + 2*sigma, mean=mu, sd=sigma) -
pnorm(mu - 2*sigma, mean=mu, sd=sigma)
```

0.954499736103641

b)

smallest  $k \in \mathbb{Z}$ , so that  $P(\mu - k\sigma < X < \mu + k\sigma) > 0.99$ .

```
In [41]: # normal distribution is symmetric
# P(μ - kσ < X < μ + kσ) =
```

```
# 1 - (P(X<μ - kσ) + P(X>μ + kσ))=
# 1 - 2 * P(X>μ + kσ)=0.99 ->P(X>μ + kσ)=0.005
# ->P(X<μ + kσ)=0.995

# x=μ + kσ
x = qnorm(0.995, mean=mu, sd=sigma)
(x - mu)/sigma
```

2.5758293035489

```
In [42]: for(k in 1:5){
          p = pnorm(mu + k*sigma, mean=mu, sd=sigma) -
            pnorm(mu - k*sigma, mean=mu, sd=sigma)
          print(paste0(k,":",p))
        }
```

```
[1] "1:0.682689492137085"
[1] "2:0.954499736103641"
[1] "3:0.99730020393674"
[1] "4:0.999936657516334"
[1] "5:0.999999426696856"
```

## Example 7.

An accompanying film about the life of the author of the exhibited works is screened on a tour of the exhibition. His screening begins every 20 minutes. Determine the probability that if you come to the screening room at random time (and you have no clue about screening shedule),

```
In [43]: # Y... time until the start of the next projection
          # Y~Ro(a=0, b=20)

          a = 0
          b = 20
```

a)

you won't have to wait more than 5 minutes for the movie to start,

```
In [44]: # P(X<5)
          punif(5, a, b)
```

0.25

b)

you will wait between 5 and 10 minutes,

```
In [45]: # P(5<X<10)
          punif(10, a, b) - punif(5, a, b)
```

0.25

c)

mean and standard deviation of the waiting time for the beginning of the movie.

```
In [46]: E_X = (a + b)/2
          E_X

          D_X = (a - b)^2/12
          D_X

          sigma_X = sqrt(D_X)
          sigma_X
```

10

33.3333333333333

5.77350269189626

## Example 8.

During quality control, we only accept the part if its dimension is in the range of 26-27 mm. The dimensions of the components have a normal distribution with a mean value of 26.4 mm and a standard deviation of 0.2 mm. What is the probability that the size of the part randomly selected for inspection will be within the required limits?

```
In [47]: # X... part size(mm)
# X~N(mu=26.4, sigma=0.2)

mu = 26.4
sigma = 0.2

# P(26<X<27)=F(27) -F(26)
pnorm(27, mean=mu, sd=sigma) - pnorm(26, mean=mu, sd=sigma)
```

0.97589997002019

## Example 9.

The length of jumps of the athlete Jakub measured in cm has a normal distribution  $N(\mu_1; \sigma_1)$ , where  $\mu_1=690$  and  $\sigma_1=10$ . The length of jumps of the athlete Aleš measured in cm also has a normal distribution  $N(\mu_2; \sigma_2)$ , where  $\mu_2=705$  and  $\sigma_2=15$ . The one who jumps more than 700 cm from two jumps at least once qualifies for the races.

```
In [48]: # SJ... length of Jakub's jump
# SJ~N(mu=690, sigma=10)
mu_J = 690
sigma_J = 10

# SA Aleš's jump length
# SA~N(mu=705, sigma=15)
mu_A = 705
sigma_A = 15

# J... Jakub's jump is successful(longer than 700 cm)
# And... Aleš's jump is successful(longer than 700 cm)

# P(J)=P(SJ>700)=1-F(700)
P.J = 1-pnorm(700, mean=mu_J, sd=sigma_J)
P.J

# P(A)=P(SA>700)=1-F(700)
P.A = 1-pnorm(700, mean=mu_A, sd=sigma_A)
P.A
```

0.158655253931457

0.630558659818236

```
In [49]: # KJ... Jakub qualifies for races,
# P(KJ)=1-(1-P(J))(1-P(J))
P.KJ=1-(1-P.J)*(1-P.J)
P.KJ

# KA... Aleš qualifies for races,
# P(KA)=1-(1-P(A))(1-P(A))
P.KA=1-(1-P.A)*(1-P.A)
P.KA
```

0.292139018262859

0.863513096164702

a)

How likely are they both to qualify for the race?

```
In [50]: # ada)
P.KJ*P.KA
```

0.252265868170678

b)

With what probability does Aleš qualify, but Jakub does not?

In [51]:

```
# adb)  
(1-P.KJ)*P.KA
```

0.611247227994024