```r
#################################################################
# L2.1. DISCRETE PROBABILITY DISTRIBUTIONS
#################################################################

##---------------------------------------------------------------
## L2.1.1. Discrete uniform distribution
##---------------------------------------------------------------
## generate n=10 experiments with a rolling die
n=10
ceiling(6*runif(n))

##---------------------------------------------------------------
## L2.1.2. Binomial distribution
##---------------------------------------------------------------
## Assuming that the probability of a side effect for a patient
## is 0.1. What is the prob. to get 0, 1, etc. side effects in a
## group of 5 patients?
dbinom(x = 0:5, size = 5, prob = 0.1)

barplot( dbinom(x = 0:5, size = 5, prob = 0.1), names.arg=0:5)

## What is the probability that not more then 1 get a side effect
sum(dbinom(x = c(0,1), size = 5,prob = 0.1))
pbinom(q = 1, size = 5,prob = 0.1)

## What is the expected number of side effects in the group?
5*0.1 = 0.5

##---------------------------------------------------------------
## L2.1.3. Hypergeometric distribution
##---------------------------------------------------------------
## There are 12 mice, of which 5 have an early brain tumor.
## A researcher randomly selects 3 of 12.
barplot( dhyper(x=0:3, k=3, m=5, n=12-5), names.arg=0:3)

## What is the probability that none of these 3 has a tumor?
dhyper(x=0, k=3, m=5, n=12-5)

## What is the probability that more then 1 have a tumor?
sum(dhyper(x=c(2,3), k=3, m=5, n=12-5))

##---------------------------------------------------------------
## L2.1.4. Poisson distribution
##---------------------------------------------------------------
## An ichthyologist studying the spoonhead sculpin catches
## specimens in a large bag seine that she trolls through the lake.
## She knows from many years experience that on averages she will
## catch 2 fish per trolling.
m = 2
## Draw distribution
barplot( dpois(c(0:10),lambda=m), names.arg=0:10)

## Find the probabilities of catching:  No fish;
dpois(0,lambda=m)

## Find the probabilities of catching:  less then 4 fishes;
sum(dpois(c(0,1,2,3),lambda=m))
```

```r
## Find the probabilities of catching:  more then 1 fish;
1-sum(dpois(c(0,1),lambda=m))

##-----------------------------------------------------------------
## L2.1.5. Negative binomial distribution
##-----------------------------------------------------------------
## How many fruitless calls (x) to random respondents should you
## make in order to complete  n=3 surveys? Probability that a
## random respondent will communicate  p=0.2


n = 3
p = 0.2
## Draw distribution
barplot( dnbinom(c(0:50),size=n,prob = p), names.arg=0:50, col="#880044")

#>>>>>>>>>>>>>>>>>>>>>>>>>
#> please, do Task L2.1
#>>>>>>>>>>>>>>>>>>>>>>>>>

###################################################################
# L2.2. CONTINUOUS PROBABILITY DISTRIBUTIONS
###################################################################

##-----------------------------------------------------------------
## L2.2.1. Uniform distribution
##-----------------------------------------------------------------
## Values from uniform distribution b/w 1 and 2:
runif(n=10,min=1, max=2)

##-----------------------------------------------------------------
## L2.2.2. Normal distriburion
##-----------------------------------------------------------------
## The volume of a liquid in bottles of one company is distributed
## normally with an expected value of 0.33 liter, and standard
## deviation of 0.01.

## Draw probability density function
x11()
x=seq(0.29,0.37,by=0.0001)
plot(x,dnorm(x,mean=0.33,sd=0.01),type="l",lwd=2)
abline(v=0.31)
## Estimate the probability to buy a bottle with > 0.31 liters of
## the liquid
pnorm(q=0.31,mean=0.33, sd=0.01)

## Estimate the interval, in which the volumes of 95% bottles are
## lying from 5% percentile to +Inf
c(qnorm(p=0.05,mean=0.33, sd=0.01),Inf)
## from 0 to 95% percentile
c(0, qnorm(p=0.95,mean=0.33, sd=0.01))
## (optimal) b/w 2.5% and 97.5% percentiles
c(qnorm(p=0.025,mean=0.33, sd=0.01),
    qnorm(p=0.975,mean=0.33, sd=0.01))

##-----------------------------------------------------------------
```

```r
## L2.2.3. Exponential distribution
##----------------------------------------------------------------
## An ichthyologist studying the spoonhead sculpin catches
## specimens in a large bag seine that she trolls through the lake.
## She knows from many years  experience that on averages she will
## catch 2 fishes per trolling. Each trolling take ~30 minutes.

## Draw probability density function
x11()
x=0:100
plot(x,dexp(x,rate=1/15),type="l",lwd=2)

## Find the probability of catching no fish in the next hour
## if pexp(60,rate=1/15) - prob to cach a fishe before 60 minutes,
## then
1-pexp(60,rate=1/15)

#>>>>>>>>>>>>>>>>>>>>>>>>>>
#> please, do Task L2.2
#>>>>>>>>>>>>>>>>>>>>>>>>>>

#################################################################
# L2.3. SAMPLING AND SAMPLING DISTRIBUTION
#################################################################

## load the data
Mice=read.table("http://edu.sablab.net/data/txt/mice.txt",
								header=T,sep="\t")
str(Mice)

sample(Mice$Ending.weight,size=20)

## run 5 times. see the variability in m and s
idx = sample(1:nrow(Mice), size=20)
mean(Mice$Ending.weight[idx])
sd(Mice$Ending.weight[idx])

#>>>>>>>>>>>>>>>>>>>>>>>>>>
#> please, do Task L2.4
#>>>>>>>>>>>>>>>>>>>>>>>>>>

#################################################################
# L2.4. INTERVAL ESTIMATION
#################################################################
Mice=read.table("http://edu.sablab.net/data/txt/mice.txt",
								header=T,sep="\t")
str(Mice)

x = Mice$Bleeding.time
t.test(x,conf.level=0.99)$conf.int

n = sum(!is.na(x))
st.error = sd(x,na.rm=TRUE) / n
c(mean(x,na.rm=TRUE) - qt(0.05/2,n-1)*st.error ,
  mean(x,na.rm=TRUE) + qt(0.05/2,n-1)*st.error )

##----------------------------------------------------------------
```

```
## L2.4.2. Interval estimation for proportion
##-----------------------------------------------------------------

Pan=read.table("http://edu.sablab.net/data/txt/pancreatitis.txt",
                            header=T, sep="\t")
## this is not completely correct as we pool control and
## experimental group.
## try to avoid on practice
x=(Pan$Smoking == "Never") & (Pan$Disease == "other")
n=sum(Pan$Disease == "other")#length(x)
p=sum(x)/n
sp = sqrt(p*(1-p)/n)
E=-qnorm(0.025)*sp


##-----------------------------------------------------------------
## L2.4.3. Interval estimation for mean
##-----------------------------------------------------------------
m=22.73
s=8.84
n=20
sm=s/sqrt(20)
E=-qt(0.025,n-1)*sm

##-----------------------------------------------------------------
## L2.4.4. Interval estimation for variance
##-----------------------------------------------------------------
n=36
s=0.18
a=0.05
## limits (in Excel values are inverted)
sqrt((n-1)*s^2 / qchisq(1-a/2, n-1))
sqrt((n-1)*s^2 / qchisq(a/2, n-1))

##-----------------------------------------------------------------
## L2.4.5. Interval estimation for correlation
##-----------------------------------------------------------------
n=10
x = 1:n + rnorm(n)
y = 1:n + rnorm(n)
r=cor(x,y)

Z=0.5 * log( (1+r) / (1-r))
sZ = 1/sqrt(n-3)

Z.min  = Z - 1.96*sZ
Z.max  = Z + 1.96*sZ

r.min = (exp(2*Z.min)-1)/(exp(2*Z.min)+1)
r.max = (exp(2*Z.max)-1)/(exp(2*Z.max)+1)

print(x)
print(y)
cat("cor(x,y) =",r," and 95% C.I. is (",r.min,",",r.max,")\n")

#>>>>>>>>>>>>>>>>>>>>>>>>>>
#> please, do Task L2.5
```

```
#>>>>>>>>>>>>>>>>>>>>>>>>>>

####################################################################
# L2.6. INTERVAL ESTIMATION FOR A RANDOM FUNCTION
####################################################################
n=10^5
## simultate mean X for n times
X = rnorm(n,mean=226.2, sd=9.57)
## simultate mean Y for n times
Y = rnorm(n,mean=76.2, sd=5.03)
## estimte confidence intervals by quantiles
plot(density(X/Y))
q95 = quantile(X/Y,prob=c(0.025,0.975))
abline(v=q95,lty=2)
lq95 = exp(quantile(log(X/Y),prob=c(0.025,0.975)))
print(q95)

## can we use analytical solution with log?
m = mean(log(X)) - mean(log(Y))
s = sqrt(var(log(X)) + var(log(Y)))
lim = c(m - s*1.96, m + s*1.96)
print(exp(lim))
abline(v=exp(lim),col=2,lty=2)
```