

# Lista 6

Mikołaj Słupiński

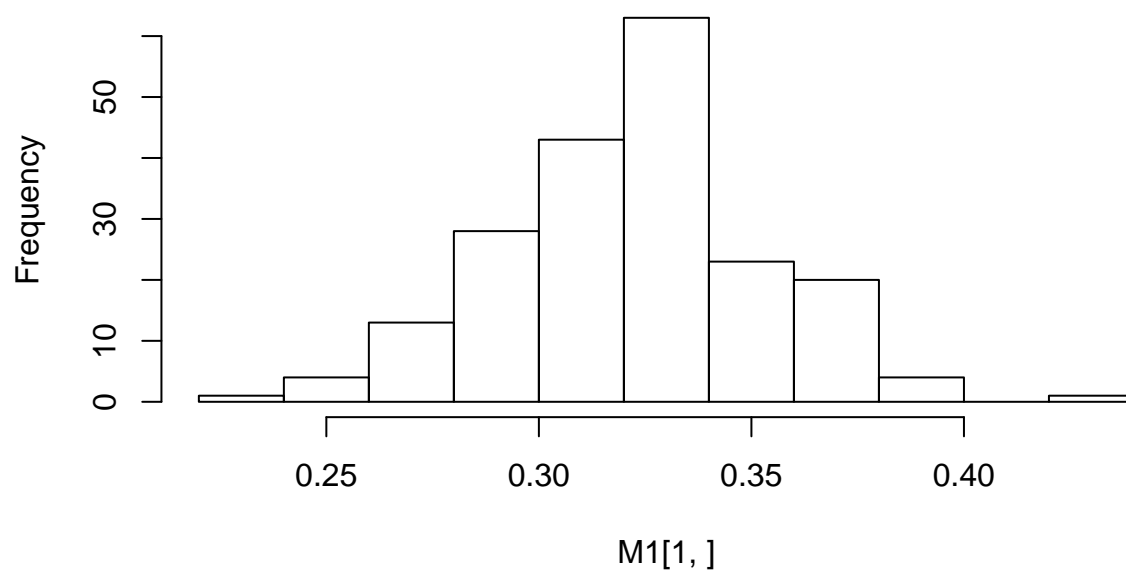
26 kwietnia 2018

## Zadanie 1

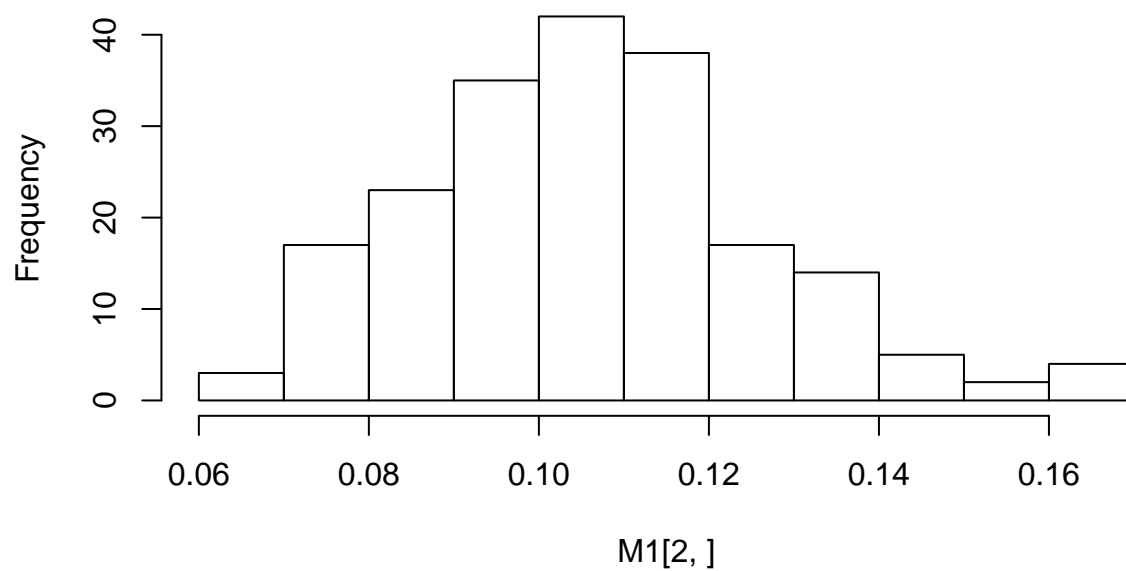
```
tab = read.table("http://www.math.uni.wroc.pl/~mbogdan/Podstawy/Dane/individuals.dat")
n = dim(tab)[1]
p.W = table(tab[[3]])["3"]/n
p.S = table(tab[[3]])["6"]/n
paste("p_W:", p.W)
paste("p_S:", p.S)
set.seed(213)
k = 200
M = sapply(1:k, function(i) sample(n,k))
z = qnorm(0.975)
M1 = apply(M, 2, function(v) {
  p.w = (table(tab[v,3])["3"] + z^2*0.5)/(k + z^2)
  p.s = (table(tab[v,3])["6"] + z^2*0.5)/(k + z^2)
  s.w = sqrt(p.w*(1-p.w)/(k+z^2))
  s.s = sqrt(p.s*(1-p.s)/(k+z^2))
  w = tab[v,3] == "3"
  i.s = p.s-s.s*z <= p.S & p.s+s.s*z >= p.S
  w = tab[v,3] == "6"
  i.w = p.w-s.w*z <= p.W & p.w+s.w*z >= p.W
  c(p.w, p.s, s.w, s.s, i.w, i.s)
})
paste("Estymator p_W:", M1[1,1])
paste("Estymator p_S:", M1[2,1])
paste("Czy przedział ufności zawiera p_W:", M1[5,1])
paste("Czy przedział ufności zawiera p_S:", M1[6,1])
paste("Średnia długość przedziału ufności p_W:", 2*mean(M1[3,])*z)
paste("Średnia długość przedziału ufności p_S:", 2*mean(M1[4,])*z)
paste("Prawdopodobieństwo zawierania p_W: ", mean(M1[5,]))
paste("Prawdopodobieństwo zawierania p_S: ", mean(M1[6,]))
hist(M1[1,], main = expression('Histogram estymatora p'["W"]))
hist(M1[2,], main = expression('Histogram estymatora p'["S"]))

## [1] "p_W: 0.316714073597023"
## [1] "p_S: 0.100198572425267"
## [1] "Estymator p_W: 0.328297932116022"
## [1] "Estymator p_S: 0.11244390391902"
## [1] "Czy przedział ufności zawiera p_W: 1"
## [1] "Czy przedział ufności zawiera p_S: 1"
## [1] "Średnia długość przedziału ufności p_W: 0.12798874155567"
## [1] "Średnia długość przedziału ufności p_S: 0.0841791434554746"
## [1] "Prawdopodobieństwo zawierania p_W: 0.965"
## [1] "Prawdopodobieństwo zawierania p_S: 0.96"
```

### Histogram estymatora $p_W$



### Histogram estymatora $p_S$



Widzimy, że wraz ze wzrostem liczby niezależnych zmiennych losowych o rozkładzie wykładniczym ich suma zbiega do rozkładu normalnego o odpowiednich parametrach.

## Zadanie 2

a)

```
n1 = 5
n2 = 10
A = matrix(rnorm(n1*1000), n1, 1000)
B = matrix(rnorm(n2*1000, 20, 10), n2, 1000)
s1 = apply(A, 2, sd)
s2 = apply(B, 2, sd)
se1 = s1/sqrt(n1)
se2 = s2/sqrt(n2)
uz = qt(0.975, n1+n2-2)
nse = sqrt(se1^2 + se2^2)
sc = sqrt((s1^2*n1 + s2^2*n2)/(n1+n2-2))
use = sc*sqrt(1/n1 + 1/n2)
df = (se1^2 + se2^2)^2/(se1^4/(n1-1) + se2^4/(n2-1))
C = rbind(A, B)
dmu = apply(C, 2, function(v) {
  mean(v[(n1+1):(n1+n2)]) - mean(v[1:n1])
})
ui = dmu - uz*use <= 20 & dmu + uz*use >= 20
nz = sapply(df, function(k) qt(0.975, k))
ni = dmu - nz*nse <= 20 & dmu + nz*nse >= 20
paste("Prawdopodobieństwo pokrycia przedziału ufności w metodzie uśrednionej:", mean(ui))
paste("Średnia długość przedziału w metodzie uśrednionej:", mean(2*uz*use))
paste("Prawdopodobieństwo pokrycia przedziału ufności w metodzie nieuśrednionej:", mean(ni))
paste("Średnia długość przedziału w metodzie uśrednionej:", mean(2*nz*nse))

## [1] "Prawdopodobieństwo pokrycia przedziału ufności w metodzie uśrednionej: 0.994"
## [1] "Średnia długość przedziału w metodzie uśrednionej: 20.4018131387693"
## [1] "Prawdopodobieństwo pokrycia przedziału ufności w metodzie nieuśrednionej: 0.958"
## [1] "Średnia długość przedziału w metodzie uśrednionej: 14.0910928627187"
```

Widzimy, że dla  $p = 0.5$  Rozkład normalny niemal idealnie aproksymuje rozkład dwumianowy.

b)

```
n1 = 5
n2 = 10
A = matrix(rnorm(n1*1000), n1, 1000)
B = matrix(rnorm(n2*1000, 20, 1), n2, 1000)
s1 = apply(A, 2, sd)
s2 = apply(B, 2, sd)
se1 = s1/sqrt(n1)
se2 = s2/sqrt(n2)
uz = qt(0.975, n1+n2-2)
nse = sqrt(se1^2 + se2^2)
sc = sqrt((s1^2*n1 + s2^2*n2)/(n1+n2-2))
use = sc*sqrt(1/n1 + 1/n2)
df = (se1^2 + se2^2)^2/(se1^4/(n1-1) + se2^4/(n2-1))
C = rbind(A, B)
dmu = apply(C, 2, function(v) {
```

```

    mean(v[(n1+1):(n1+n2)]) - mean(v[1:n1])
  })
  ui = dmu - uz*use <= 20 & dmu + uz*use >= 20
  nz = sapply(df, function(k) qt(0.975, k))
  ni = dmu - nz*nse <= 20 & dmu + nz*nse >= 20
  paste("Prawdopodobieństwo pokrycia przedziału ufności w metodzie uśrednionej:", mean(ui))
  paste("Średnia długość przedziału w metodzie uśrednionej:", mean(2*uz*use))
  paste("Prawdopodobieństwo pokrycia przedziału ufności w metodzie nieuśrednionej:", mean(ni))
  paste("Średnia długość przedziału w metodzie uśrednionej:", mean(2*nz*nse))

## [1] "Prawdopodobieństwo pokrycia przedziału ufności w metodzie uśrednionej: 0.956"
## [1] "Średnia długość przedziału w metodzie uśrednionej: 2.49531617847656"
## [1] "Prawdopodobieństwo pokrycia przedziału ufności w metodzie nieuśrednionej: 0.951"
## [1] "Średnia długość przedziału w metodzie uśrednionej: 2.50573544775496"

```

c)

```

n1 = 10
n2 = 10
A = matrix(rnorm(n1*1000), n1, 1000)
B = matrix(rnorm(n2*1000, 20, 10), n2, 1000)
s1 = apply(A, 2, sd)
s2 = apply(B, 2, sd)
se1 = s1/sqrt(n1)
se2 = s2/sqrt(n2)
uz = qt(0.975, n1+n2-2)
nse = sqrt(se1^2 + se2^2)
sc = sqrt((s1^2*n1 + s2^2*n2)/(n1+n2-2))
use = sc*sqrt(1/n1 + 1/n2)
df = (se1^2 + se2^2)^2/(se1^4/(n1-1) + se2^4/(n2-1))
C = rbind(A, B)
dmu = apply(C, 2, function(v) {
  mean(v[(n1+1):(n1+n2)]) - mean(v[1:n1])
})
  ui = dmu - uz*use <= 20 & dmu + uz*use >= 20
  nz = sapply(df, function(k) qt(0.975, k))
  ni = dmu - nz*nse <= 20 & dmu + nz*nse >= 20
  paste("Prawdopodobieństwo pokrycia przedziału ufności w metodzie uśrednionej:", mean(ui))
  paste("Średnia długość przedziału w metodzie uśrednionej:", mean(2*uz*use))
  paste("Prawdopodobieństwo pokrycia przedziału ufności w metodzie nieuśrednionej:", mean(ni))
  paste("Średnia długość przedziału w metodzie uśrednionej:", mean(2*nz*nse))

## [1] "Prawdopodobieństwo pokrycia przedziału ufności w metodzie uśrednionej: 0.946"
## [1] "Średnia długość przedziału w metodzie uśrednionej: 13.7106191343263"
## [1] "Prawdopodobieństwo pokrycia przedziału ufności w metodzie nieuśrednionej: 0.948"
## [1] "Średnia długość przedziału w metodzie uśrednionej: 13.9597278046728"

```

Widzimy, że około 95% obserwacji wpada do przedziału ufności, co zgadza się z teoretycznym poziomem ufności 95%. Można również zaobserwować, że wraz ze wzrostem liczby obserwacji, przedział się zawęża.