FoS_HW_06_Group_II

RNR

2025-11-24

1.

1 Order statistics

a) $F_Y(y) = \mathbb{P}(Y \leq y) = \mathbb{P}(\max\{X_1, \dots, X_n\} \leq y) = \mathbb{P}(X_1 \leq y \wedge \dots \wedge X_n \leq y)$

$= \mathbb{P}(X_1 \leq y) \cdot \dots \cdot \mathbb{P}(X_n \leq y) = \mathbb{P}(X_1 \leq y)^n = F(y)^n \blacksquare$

X_i are iid

$f_Y(y) = \frac{d}{dy} F_Y(y) = \frac{d}{dy} F(y)^n = n \cdot F(y)^{n-1} \cdot f(y)$

b) $F_Z(z) = \mathbb{P}(Z \leq z) = \mathbb{P}(\min\{X_1, \dots, X_n\} \leq z) = \mathbb{P}(X_1 \leq z \vee \dots \vee \mathbb{P}(X_n \leq z)$

$= 1 - \mathbb{P}(X_1 > z \wedge \dots X_n > z) = 1 - [\prod_{i=1}^{n} \mathbb{P}(X_i > z)]$

$= 1 - [\mathbb{P}(X_1 > z)]^n = 1 - [1 - \mathbb{P}(X_1 \leq z)]^n = 1 - [1 - F(z)]^n \blacksquare$

$f_Z(z) = \frac{d}{dz} F_Z(z) = \frac{d}{dz} 1 - [1 - F(z)]^n = n \cdot [1 - F(z)]^{n-1} \cdot f(z)$

c) $F_{Z,Y}(z,y) = \mathbb{P}(Z \leq z, Y \leq y) = 0 \quad$ if $y < z$

if the max of all X_i's is smaller than y, then the min of them cannot be bigger

$F_{Y,Z}(y,z) = \mathbb{P}(Y < y, Z < z) = \mathbb{P}(Y < y \wedge Z < z)$

$= \mathbb{P}(\max\{X_1, \dots, X_n\} < y \wedge \min\{X_1, \dots, X_n\} < z)$

$= \mathbb{P}(\max\{X_1, \dots, X_n\} < y) - \mathbb{P}(\text{"all } X_i \in (z, y]\text{"})$

$= F(y)^n - \mathbb{P}(z < X_i \leq y \;\; \forall i \in \{1, \dots, n\})$

$= F(y)^n - (F(y) - F(z))^n$

$$\Rightarrow F_{Y,Z}(y,z) = \begin{cases} F(y)^n - (F(y) - F(z))^n & \text{if } z < y \\ 0 & \text{otherwise} \end{cases}$$

d) $f_{Y,Z}(y,z) = \dfrac{\partial^2 F_{Y,Z}(y,z)}{\partial y \, \partial z} = \dfrac{\partial}{\partial y}\left[\dfrac{\partial}{\partial z}\left(F(y)^n - \left(F(y) - F(z)\right)^n\right)\right]$

$= \dfrac{\partial}{\partial y} \; n \cdot \left(F(y) - F(z)\right)^{n-1} \cdot f(z) = n \cdot (n-1) \cdot \left(F(y) - F(z)\right)^{n-2} \cdot f(z) \cdot f(y)$

Z and Y are not independent since $f_{Y,Z}(y,z) \neq f_Y(y) \cdot f_Z(z)$. ∎

e) Convergence in distribution: $\lim_{n\to\infty} F_n(x) = F(x)$

n is just a scalar

$\lim_{n\to\infty} T_n(z) = \lim_{n\to\infty} n \cdot \min\{X_1(z), \dots, X_n(z)\} = \lim_{n\to\infty} \min\left\{X_1\left(\frac{z}{n}\right), \dots, X_n\left(\frac{z}{n}\right)\right\}$

part a) →

$= \lim_{n\to\infty} 1 - \left[1 - F\left(\frac{z}{n}\right)\right]^n \approx \lim_{n\to\infty} 1 - \left[1 - \frac{\lambda z}{n}\right]^n \xrightarrow{n\to\infty} 1 - e^{-\lambda z}$

And $\lim_{n\to\infty} T_n(z) = 0$ if $z < 0$ since X_i's are positive R.V's

Therefore $\lim_{n\to\infty} T_n \sim \text{Exp}(\lambda)$ ∎

2.

Nr 2 $\quad Y = \max\{\cancel{X_1, \dots, X_n}\ U, V\}$ $\quad \boxed{F(y) = y}$

a)

$$f_Y(y) = \frac{d}{dy} F_Y(y) \overset{1a}{=} \frac{d}{dy} F(y)^2 = \underbrace{2F(y)}_{2y}$$

$Z = \min\{U, V\}$

$$f_Z(z) = \frac{d}{dz} F_Z(z) \overset{1b}{=} \frac{d}{dz} 1 - [1 - F(z)]^2 = 1 - 1 + 2F(z) - F(z)^2$$

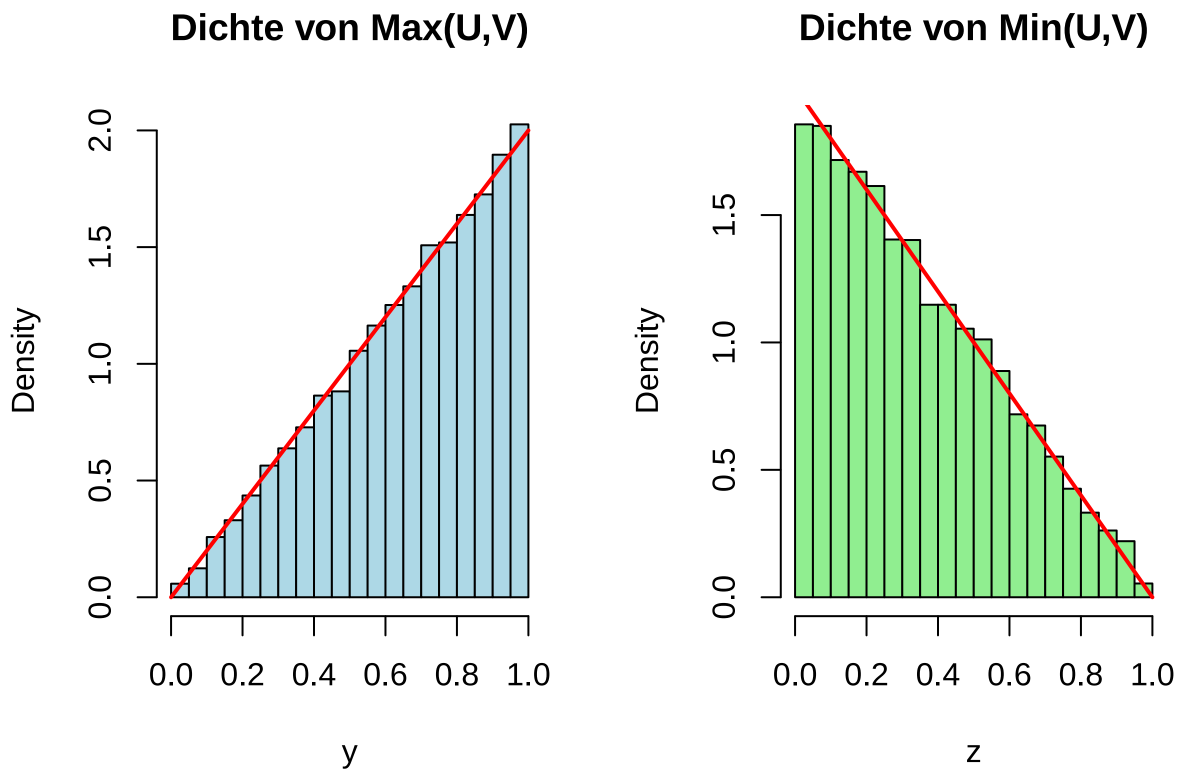
$$= 2 - 2F(z) = 2 - 2z$$

```r
n <- 10000
u <- runif(n)
v <- runif(n)

y_max <- pmax(u, v)
z_min <- pmin(u, v)


par(mfrow=c(1,2))
hist(y_max, freq=FALSE, main="Dichte von Max(U,V)", xlab="y", col="lightblue")
curve(2*x, add=TRUE, col="red", lwd=2)

hist(z_min, freq=FALSE, main="Dichte von Min(U,V)", xlab="z", col="lightgreen")
curve(2-2*x, add=TRUE, col="red", lwd=2)
```



Dichte von Max(U,V)
Density
0.0
0.5
1.0
1.5
2.0
0.0
0.2
0.4
0.6
0.8
1.0
y
Dichte von Min(U,V)
Density
0.0
0.5
1.0
1.5
0.0
0.2
0.4
0.6
0.8
1.0
z

code to a

b

b) $I_{\{U \le p\}} I_{\{V \le p\}} = 1 \Leftrightarrow U \le p$ and $V \le p$

$\Leftrightarrow \max\{U, V\} \le p$

$$\mathbb{P}(\max\{U, V\} \le p) \overset{1a)}{=} F_Y(p) = p^2$$

$$\Rightarrow I_{\{U \le p\}} I_{\{V \le p\}} \sim Ber(p^2)$$

$I_{\{U > p\}} I_{\{V > p\}} \Leftrightarrow \min\{U, V\} > p$

$$\Rightarrow \mathbb{P}(\min\{U, V\} > p) = 1 - F_Z(p) \overset{1b)}{=}$$

$$= 1 - 1 + [1 - p]^2 = (1-p)^2$$

$$I_{\{U > p\}} I_{\{V > p\}} \sim Ber((1-p)^2)$$

c) $F(x) = \frac{x-a}{b-a}$ ← Destribution Of Unif(a, b)

$\Rightarrow$

$F_Y(y) \overset{1a)}{=} \left(\frac{y-a}{b-a}\right)^n$

$F_Z(z) = 1 - \left(1 - \frac{z-a}{b-a}\right)^n$

$= 1 - \left(\frac{b-a}{b-a} - \frac{z-a}{b-a}\right)^n = \underline{\underline{1 - \left(\frac{b-z}{b-a}\right)^n}}$

$n \to \infty \quad F_Y(y) \to ?$

<u>Case 1</u>

$y < b \Rightarrow \frac{y-a}{b-a} < 1 \Rightarrow F_Y(y) \to \underline{\underline{0}}$

<u>Case 2</u>

$y = b \Rightarrow \frac{y-a}{b-a} = 1 \Rightarrow F_Y(y) \to \underline{\underline{1}}$

$n \to \infty \quad F_Z(z) \to ?$

<u>Case 1</u> $z > a \Rightarrow \frac{b-z}{b-a} < 1 \Rightarrow F_Z(z) \to 1$

<u>Case 2</u> $z = a \Rightarrow \frac{b-z}{b-a} = 1 \Rightarrow F_Z(z) \to 0$

c

$$d)\ F(x) = 1 - e^{-\lambda x}$$

$\Rightarrow$	$F_Y(y) = (1 - e^{-\lambda y})^n$	$1-[1-(1-e^{-\lambda z})]^n$ $F_Z(z) = 1 - (e^{-\lambda z})^n = 1 - e^{-\lambda z n}$
$n \to \infty$! $n > 0$	Case $y = 0 \Rightarrow F_Y(y) \to 0$ Case $y > 0 \Rightarrow 0 < e^{-\lambda y} < 1$ $\Rightarrow F_Y(y) \to 0$	~~$z \to 0$~~ Case $z = 0$ $F_Z(z) \to 1$ Case $z > 0$ $\Rightarrow e^{-\lambda z} < 1 \Rightarrow F_Z(z) \to 1$

d

3.

a)

$$P(|X - \mu| \geq a) \leq \frac{Var(X)}{a^2}, \forall a > 0$$

```r
library("UsingR")
```

```
## Lade nötiges Paket: MASS
```

```
## Lade nötiges Paket: HistData
```

```
## Lade nötiges Paket: Hmisc
```

```
##
## Attache Paket: 'Hmisc'
```

```
## Die folgenden Objekte sind maskiert von 'package:base':
##
##     format.pval, units
```

```r
var(exec.pay)/(3*sd(exec.pay))**2
```

```
## [1] 0.1111111
```

```r
mu <- mean(exec.pay)
sd <- sd(exec.pay)

lb <- mu - 3*sd
ub <- mu + 3*sd

1-(pnorm(ub, mean = mu, sd = sd)-pnorm(lb, mean = mu, sd = sd))
```

b)

```
## [1] 0.002699796
```

c)

$$z_i := \frac{1}{s}(x_i - \bar{x}), i = 1, ..., n$$

$$sz_i = (x_i - \bar{x}), i = 1, ..., n$$

For $z_i = 3$ is the deviation from the mean, three times the standard deviation.

```r
# install.packages("UsingR")
# require("UsingR")
summary(exec.pay)
```

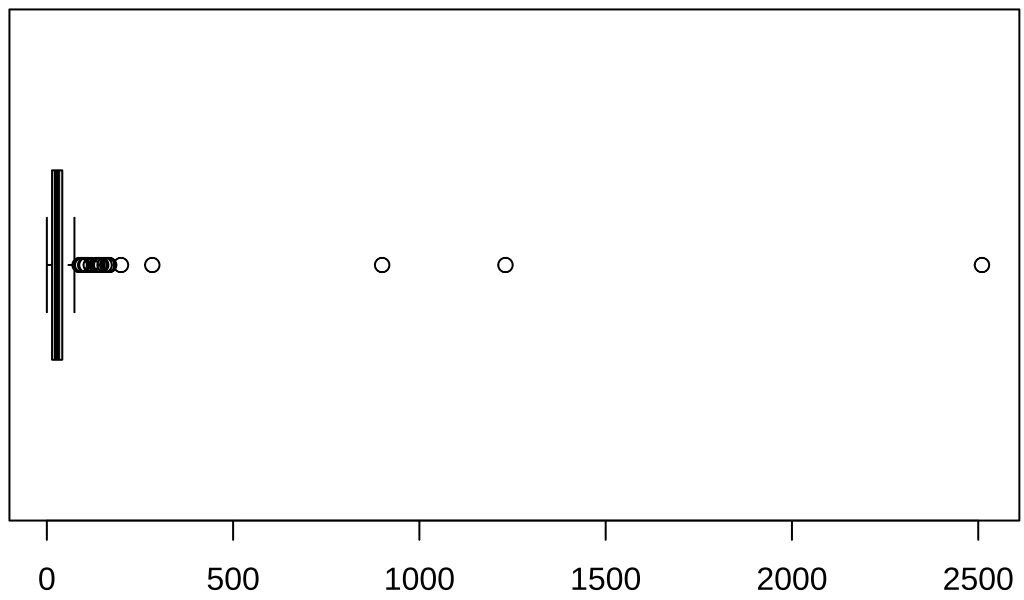
d)

```
##    Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##    0.00   14.00   27.00   59.89   41.50 2510.00
```

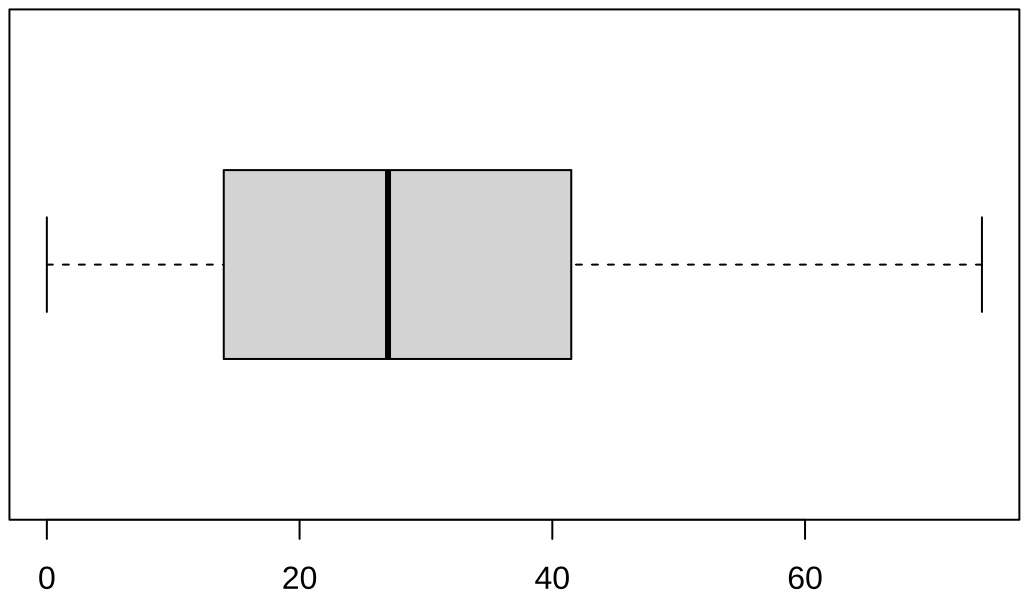
```r
head(exec.pay)
```

```
## [1] 136  74   8  38  46  43
```

```r
boxplot(exec.pay, horizontal = TRUE)
```




```r
boxplot(exec.pay, horizontal = TRUE, outline = FALSE)
```



From the second boxplot it is clear that the outliers are after the second whisker(about 76).

```
mean <- mean(exec.pay)
sd <- sd(exec.pay)
z_score <- (exec.pay-mean)/sd
z_score   # or just < - scale(exec.pay)
```

e)

```
##   [1]  0.367606505  0.068152586 -0.250620941 -0.105723883 -0.067084668
##   [6] -0.081574374 -0.245791039 -0.245791039 -0.231301333 -0.236131235
##  [11] -0.192662118 -0.245791039  0.169580526 -0.125043491 -0.255450843
##  [16] -0.221641530 -0.100893981 -0.231301333 -0.149193001 -0.187832216
##  [21]  0.000533959 -0.120213589 -0.207151824 -0.115383687 -0.149193001
##  [26]  0.493183955  0.135771213 -0.139533197 -0.260280745  0.362776603
##  [31] -0.226471431 -0.192662118 -0.245791039 -0.221641530 -0.154022903
##  [36] -0.086404276 -0.240961137 -0.120213589 -0.279600353 -0.211981726
##  [41] -0.154022903 -0.086404276  0.396585917 -0.129873393  0.357946701
##  [46] -0.178172412 -0.125043491 -0.211981726 -0.226471431  0.517333465
##  [51] -0.245791039 -0.183002314 -0.100893981 -0.154022903 -0.144363099
##  [56] -0.183002314 -0.221641530 -0.245791039 -0.168512608  0.222709448
##  [61] -0.134703295 -0.144363099  0.140601115  0.140601115 -0.062254766
##  [66] -0.207151824 -0.163682706  5.656349112 -0.260280745  0.208219742
```

```
##  [71] -0.057424864 -0.173342510 -0.236131235 -0.197492020 -0.226471431
##  [76] -0.149193001 -0.192662118 -0.071914570 -0.274770451 -0.129873393
##  [81] -0.091234178 -0.255450843 -0.236131235 -0.240961137 -0.183002314
##  [86] -0.115383687 -0.255450843 -0.197492020 -0.091234178 -0.096064079
##  [91] -0.240961137 -0.216811628  0.159920723  0.034343272 -0.149193001
##  [96] -0.168512608  0.150260919 -0.105723883 11.833793673 -0.265110647
## [101] -0.134703295  0.024683469 -0.289260156 -0.226471431 -0.158852804
## [106] -0.211981726 -0.187832216 -0.260280745 -0.289260156 -0.154022903
## [111] -0.250620941 -0.226471431  0.053662880 -0.115383687 -0.236131235
## [116]  0.222709448 -0.110553785 -0.091234178 -0.226471431  4.057651575
## [121] -0.105723883 -0.173342510 -0.216811628 -0.158852804 -0.231301333
## [126] -0.231301333 -0.183002314 -0.096064079 -0.052594962 -0.183002314
## [131]  0.280668271 -0.057424864 -0.240961137 -0.284430255 -0.115383687
## [136]  0.459374642 -0.245791039 -0.125043491 -0.149193001 -0.231301333
## [141] -0.289260156 -0.154022903 -0.187832216 -0.134703295 -0.202321922
## [146] -0.038105256 -0.149193001 -0.226471431  0.671890326 -0.096064079
## [151] -0.236131235 -0.042935158 -0.071914570 -0.081574374 -0.139533197
## [156] -0.265110647 -0.202321922 -0.216811628 -0.168512608 -0.245791039
## [161] -0.202321922 -0.226471431 -0.009125845 -0.183002314 -0.096064079
## [166] -0.125043491 -0.211981726 -0.139533197 -0.158852804 -0.216811628
## [171] -0.178172412 -0.052594962  0.000533959 -0.154022903  0.068152586
## [176] -0.086404276 -0.173342510 -0.207151824 -0.245791039  0.005363861
## [181] -0.192662118 -0.178172412 -0.163682706 -0.139533197  0.517333465
## [186] -0.197492020 -0.221641530 -0.226471431  0.415905524  1.077602088
## [191] -0.231301333 -0.033275354 -0.163682706 -0.211981726 -0.149193001
## [196] -0.042935158 -0.216811628 -0.183002314 -0.158852804
```

```r
propb_3sd <- mean(abs(z_score) > 3)
propb_3sd
```

```
## [1] 0.01507538
```

```r
cat("Comparison: normal distribution 0.27% < z-score 1.51% << Chebyshevs inequality 11%")
```

```
## Comparison: normal distribution 0.27% < z-score 1.51% << Chebyshevs inequality 11%
```

f) definition of z-score

$$z_i = \frac{1}{s}(x_i - \bar{x})$$

where s is the sample standard deviation with denominator n:

$$s = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

From the z-score definition:

$$x_i - \bar{x} = s \cdot z_i$$

Substitute into the numerator

$$\sum_{i=1}^{n}(x_i-\bar{x})^3=\sum_{i=1}^{n}(s\cdot z_i)^3=s^3\sum_{i=1}^{n}z_i^3$$

Substitute into the denominator

$$\sum_{i=1}^{n}(x_i-\bar{x})^2=\sum_{i=1}^{n}(s\cdot z_i)^2=s^2\sum_{i=1}^{n}z_i^2$$

Since $s^2=\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2$, we have:

$$\sum_{i=1}^{n}(x_i-\bar{x})^2=ns^2$$

Therefore:

$$s^2\sum_{i=1}^{n}z_i^2=ns^2$$

This implies:

$$\sum_{i=1}^{n}z_i^2=n$$

Simplify the denominator term

$$\left[\sum_{i=1}^{n}(x_i-\bar{x})^2\right]^{3/2}=\left[s^2\sum_{i=1}^{n}z_i^2\right]^{3/2}=[s^2\cdot n]^{3/2}=s^3\cdot n^{3/2}$$

everything combined

$$\sqrt{n}\cdot\frac{\sum_{i=1}^{n}(x_i-\bar{x})^3}{[\sum_{i=1}^{n}(x_i-\bar{x})^2]^{3/2}}=\sqrt{n}\cdot\frac{s^3\sum_{i=1}^{n}z_i^3}{s^3\cdot n^{3/2}}$$

$$=\sqrt{n}\cdot\frac{\sum_{i=1}^{n}z_i^3}{n^{3/2}}=\frac{n^{\frac{1}{2}}}{n^{3/2}}\sum_{i=1}^{n}z_i^3$$

$$=\frac{1}{n}\sum_{i=1}^{n}z_i^3$$

g) From f);

```
sum(z_score**3)/length(z_score)
```

```
## [1] 9.578542
```

h) definition of z-score

$$z_i = \frac{1}{s}(x_i - \bar{x})$$

where s is the sample standard deviation with denominator n:

$$s = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

From the z-score definition:

$$x_i - \bar{x} = s \cdot z_i$$

Substitute into the numerator

$$\sum_{i=1}^{n}(x_i - \bar{x})^4 = \sum_{i=1}^{n}(s \cdot z_i)^4 = s^4 \sum_{i=1}^{n} z_i^4$$

Substitute into the denominator

$$\sum_{i=1}^{n}(x_i - \bar{x})^2 = \sum_{i=1}^{n}(s \cdot z_i)^2 = s^2 \sum_{i=1}^{n} z_i^2$$

Recall from f) that $\sum_{i=1}^{n} z_i^2 = n$

Simplify the denominator term

$$\left[\sum_{i=1}^{n}(x_i - \bar{x})^2\right]^2 = \left[s^2 \sum_{i=1}^{n} z_i^2\right]^2 = [s^2 \cdot n]^2 = s^4 \cdot n^2$$

everything combined

$$n \cdot \frac{\sum_{i=1}^{n}(x_i - \bar{x})^4}{\left[\sum_{i=1}^{n}(x_i - \bar{x})^2\right]^2} - 3 = n \cdot \frac{s^4 \sum_{i=1}^{n} z_i^4}{s^4 \cdot n^2} - 3$$
$$= n \cdot \frac{\sum_{i=1}^{n} z_i^4}{n^2} - 3 = \frac{1}{n}\sum_{i=1}^{n} z_i^4 - 3$$

Why subtract 3?

The subtraction of 3 is a normalization convention. For a standard normal distribution, the fourth moment $E[Z^4] = 3$ (since for $Z \sim N(0,1)$, we have $E[Z^4] = 3 \cdot (E[Z^2])^2 = 3 \cdot 1^2 = 3$).

By subtracting 3, we get:

- **Kurtosis $= 0$** for normal distributions (mesokurtic)
- **Kurtosis > 0** for distributions with heavy tails (leptokurtic)
- **Kurtosis < 0** for distributions with light tails (platykurtic)

i) From g);

```r
sum(z_score**4-3)/length(z_score)
```

```
## [1] 102.064
```

```r
#(sum(z_score**4)-3*length(z_score))/length(z_score)
```

4.

Nr. 4 *uniform Distr., $a = -0,5$ $b = 0,5$

a) $E[X_i] = \frac{a+b}{2} = 0$ $\quad Var(X_i) \overset{*}{=} \frac{(b-a)^2}{12} = \frac{1}{12}$

b) $Y = \sum_{i=1}^{100} X_i$ $\quad E[Y] = \sum_{i=1}^{100} E[X_i] = 0$

$Var(Y) = \sum_{i=1}^{100} Var(X_i) = \frac{100}{12}$

= equal, Cheb.

$\Rightarrow p = P(|Y - 0| \geq 10) \leq \frac{Var(Y)}{100} = \frac{1}{12}$ □

c) $|X_i| = \begin{cases} X_i & \text{if } X_i \geq 0 \\ -X_i & \text{else} \end{cases} \Rightarrow |X_i| \sim Unif(0, 0.5)$

$\Rightarrow E(|X_i|) = 0,25$

LLN $\Rightarrow \frac{1}{n} \sum_{i=1}^{n} |X_i| \xrightarrow[n \to \infty]{LLN} E[|X_i|] = 0,25$

d) $E[Y] = 0 < \infty$ $\quad Var(Y) = \frac{100}{12} \begin{matrix} > 0 \\ < \infty \end{matrix}$

why???

$Z_{100} = \frac{Y}{\sqrt{\frac{100}{12} \cdot 100}} = \frac{Y\sqrt{12}}{100} \neq \frac{\sqrt{n}}{\sigma}(\bar{X}_n - \mu)$

* slide 32 Theorem 3

$\Rightarrow$

$\mathbb{P}(|Y| > 10) = 1 - \mathbb{P}(-10 \leq Y \leq 10)$

$\approx 1 - \mathbb{P}\left(-\frac{\sqrt{12}}{10} \leq Z_{100} \leq \frac{\sqrt{12}}{10}\right) = 1 - \Phi\left(\frac{\sqrt{12}}{10}\right) + \overbrace{\Phi\left(-\frac{\sqrt{12}}{10}\right)}^{1 - \Phi\left(\frac{\sqrt{12}}{10}\right)}$

$= 2 - 2\Phi\left(\frac{\sqrt{12}}{10}\right) \approx 2 - 2\Phi(0,3464) \approx 2 - 2 \cdot 0,6368$

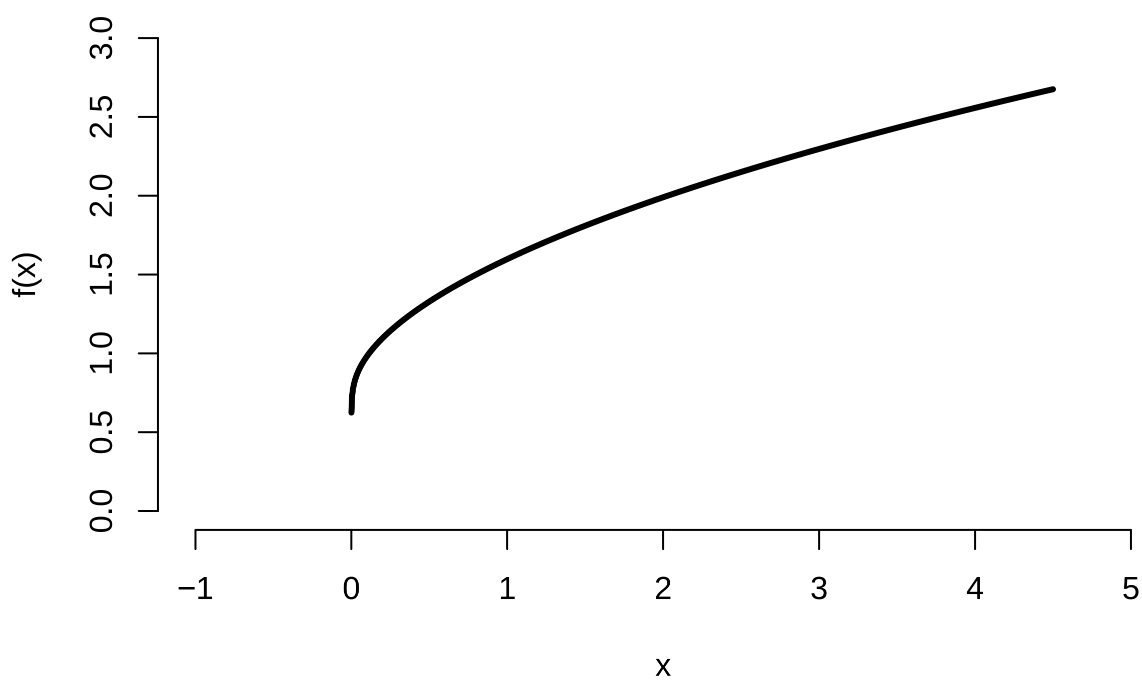
* $\Rightarrow 2 - 2\Phi(\sqrt{12}) \approx 2 - 2\Phi(3,46) \approx 2 - 2 \cdot 0,9997299 = 0,0005462$

5.

In this exercise we just have to modify the code from the lecture slides

```
n <- 10e4
f<-function(x){sqrt(x+sqrt(x+sqrt(x+sqrt(x))))}
x<-seq(-0.5,4.5,length=1000)
plot(x,f(x),type="l",xlim=c(-1,5), ylim=c(0,3),bty="n",lwd=3)
```

```
## Warning in sqrt(x): NaNs wurden erzeugt
```




```
h<-0 # absolute frequency of getting points under the curve f
for (i in 1:n){
  x<-runif(1,0,4)
  y<-runif(1,0,4)
  if (y<f(x)){
    h<-h+1
  }
  else {
  }
}

16*h/n # approximation of the grey area
```

```
## [1] 7.71088
```