Statistik II

Paul Strimtu 3898312
Jakob Striegel 4351490

Abgabe 4: 29 April 2024

1 Zweidimensionale Verteilung

```
fXY <- matrix(c(0.0,0.1,0.4,
                0.3,0.2,0.0), ncol = 3, byrow = TRUE)

fX <- apply(fXY, 1, sum); fX

## [1] 0.5 0.5

fY <- apply(fXY, 2, sum); fY

## [1] 0.3 0.3 0.4

EX <- sum((0:1)*fX); EX

## [1] 0.5

EY <- sum((1:3)*fY); EY

## [1] 2.1

products <- (0:1)%*%t(1:3)
gewProducts <- products*fXY

EXY <- sum(gewProducts); EXY

## [1] 0.7

covXY <- EXY-(EX*EY); covXY

## [1] -0.35

varX <- sum((0:1)^2*fX)-EX^2; varX

## [1] 0.25

varY <- sum((1:3)^2*fY)-EY^2; varY

## [1] 0.69

rho <- covXY / (sqrt(varX)*sqrt(varY)); rho

## [1] -0.842701
```

Es besteht eine hohe negative Korrelation zwischen den diskreten Variablen X und Y. Diese beträgt ca. $\rho = -0.84$.

2 Normalverteilung

a)

```
SP500 <- read.csv("SP500.csv")

sortedSP500 <- SP500[order(SP500$Rendite), ]

sortedSP500[1:2,]

##           Datum    Rendite
## 4001 2020-03-16 -11.984055
## 1129 2008-10-15  -9.734978
```

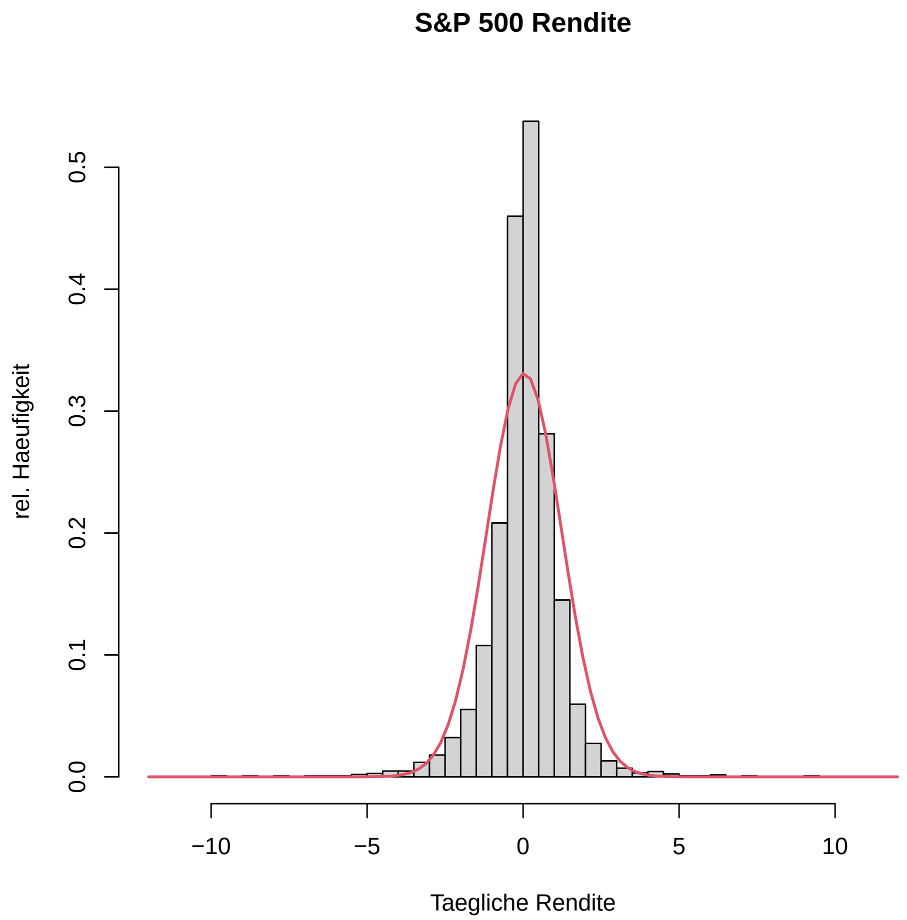
Die zwei Tage, an denen die höchsten prozentualen Verluste auftraten, sind der 16.03.2020 - zu diesem Zeitpunkt brach Corona seit einigen Tagen in Europa aus.
Am 15.10.2008 bzw. Sep.-Okt. 2008 fand der Höhepunkt der Weltwirtschaftskrise statt, kurz nachdem die Investmentbank Lehman Brothers Insolvenz melden musste.

b)

```
hist(SP500$Rendite,breaks=50,col="lightgrey",main="S&P 500 Rendite",
     xlab="Taegliche Rendite", ylab="rel. Haeufigkeit",prob=TRUE,ylim=c(0,0.55))

mu <- mean(SP500$Rendite); sigma <- sd(SP500$Rendite)

curve(dnorm(x,mu,sigma),col=2,lwd=2,add=TRUE)
```



Die gewählte Normalverteilung spiegelt die Daten nicht optimal. Der Erwartungswert wird zum Beispiel viel niedriger angesetzt.

c)

(i)

```
1-pnorm(1,mu,sigma)

## [1] 0.2120047
```

Die Wahrscheinlichkeit dafür, dass der S&P 500 am nächsten Tag, falls normal verteilt, um 1 % oder mehr steigt, beträgt ca. 21.2 %.

(ii)

```
pnorm(-2,mu,sigma)

## [1] 0.04555864
```

Die Wahrscheinlichkeit dafür, dass der S&P 500 am nächsten Tag, falls normal verteilt, um 2 % oder mehr fällt, beträgt ca. 4.56 %.

(iii)

```
current <- 5000
upperBound <- (5050/current-1)*100
lowerBound <- (4950/current-1)*100

pnorm(upperBound,mu,sigma)-pnorm(lowerBound,mu,sigma)

## [1] 0.5930624
```

Die Wahrscheinlichkeit dafür, dass der S&P 500 bei einem Ausgangsindex von 5000 Punkten am nächsten Tag, falls normal verteilt, bei einem Index zwischen 4950 und 5050 Punkten liegt, beträgt ca. 59.31 %.

3 Monte-Carlo-Simulation

a)

```
simDayReturn <- rnorm(1, mu, sigma); simDayReturn

## [1] -1.656439

newStockPrice <- 5000 * (1 + simDayReturn/100); newStockPrice

## [1] 4917.178
```

Der Wert von *simDayReturn* gibt, nach der passenden Normalverteilung, eine ßufällige" prozentuale Kursveränderung. In *newStockPrice* wird der neue Indexwert für den Tag gespeichert.

b)

```
stockPrice <- numeric(250)
startPrice <- 5000
previousPrice <- startPrice

for (i in 1:250) {
  dayReturn <- rnorm(1,mu,sigma)
  currentPrice <- previousPrice * (1 + dayReturn/100) #calc price for today
  stockPrice[i] <- currentPrice #append to current list
  previousPrice <- currentPrice #update previousPrice (for new day)
}

head(round(stockPrice,2))

## [1] 5073.49 5033.71 5001.72 4986.89 4849.08 4835.37

round(stockPrice[250],2)

## [1] 5570.03
```

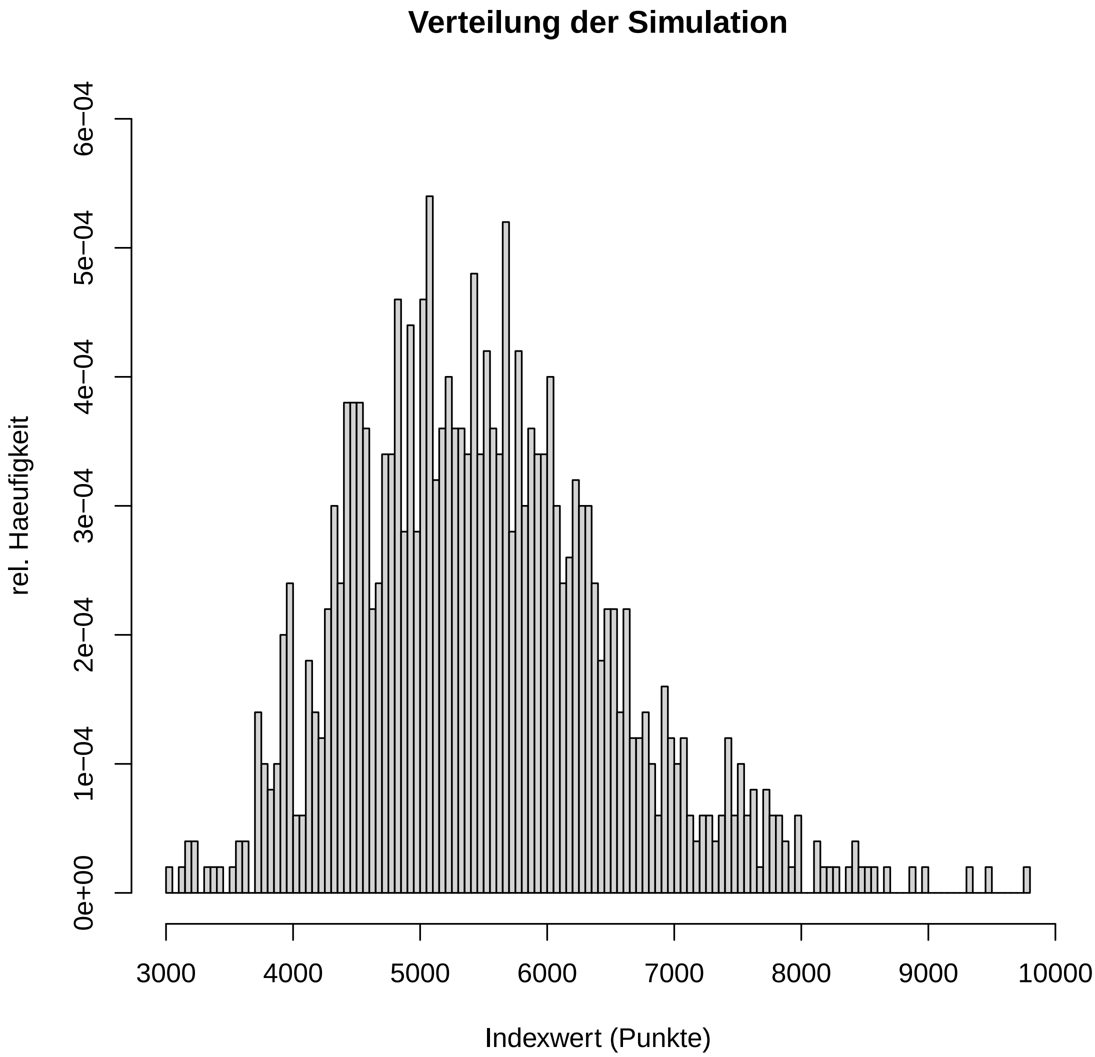
In der ersten Ausgabe werden die ersten 6 Werte der Simulation angezeigt, also die ersten 6 Tage eines Bösrenjahres. Die zweite Ausgabe zeigt den letzten Wert der Simulation und somit den Index des letzten Tages des Börsenjahres.

c)

```
endPrice_2 <- numeric(1000)

for (j in 1:1000) {
  stockPrice_2 <- numeric(250)
  startPrice_2 <- 5000
  previousPrice_2 <- startPrice_2
  for (i in 1:250) {
    dayReturn_2 <- rnorm(1,mu,sigma)
    currentPrice_2 <- previousPrice_2 * (1 + dayReturn_2/100) #calc price for today
    stockPrice_2[i] <- currentPrice_2 #append to current list
    previousPrice_2 <- currentPrice_2 #update previousPrice (for new day)
  }
  endPrice_2[j] <- stockPrice_2[250]
}
my_data <- data.frame(x = 1:1000, value = endPrice_2)

hist(my_data$value,breaks=100,col="lightgrey",main="Verteilung der Simulation",
     xlab="Indexwert (Punkte)",ylab="rel. Haeufigkeit", prob=TRUE,ylim=c(0,0.0006))
```



In der oberen Simulation wurde ein Börsenjahr 1000 mal simuliert und jeweils der Wert des letzten Tages berüksichtigt. Die Werte zeigen mit Hilfe des Histogramms eine ungefähr normalverteilte Verteilung mit einem Erwartungswert (peak) zwischen 5000 und 6000 Punkten.

d)

Die Simulation ist möglicherweise für einen "ruhigen" Jahresverlauf zumindest grundsätzlich geeignet. Es werden jedoch keine unvorhersehbaren Ereignisse (z.B. eine Weltwirtschaftskrise) realistisch dargestellt, somit ist eine realitätsgetreue Simulation eher schwer realisierbar, solange man nicht mehrere Faktoren berüksichtigt.

4 ZGWS

a)

Sei X die Gesamtdauer, bis alle 20 Übungsaaufgaben durchgerechnet wurden. X ist $Exp(0.04)$ verteilt mit $\lambda = 0.04$. Damit gilt:

$$\mathbb{E}(X) = \frac{1}{\lambda} = \frac{1}{0.04} = 25$$
$$Var(X) = \frac{1}{\lambda^2} = \frac{1}{0.04^2} = 625$$

Nach ZWGS $\Rightarrow X_1 + ... + X_{20} \overset{approx}{\sim} N(\underbrace{20 \cdot 25}_{n \cdot \mu}, \underbrace{20 \cdot 625}_{n \cdot \sigma^2})$.

b)

Zu berechnen ist die Wahrscheinlichkeit, dass das Bearbeiten der Aufgaben weniger als 10 h, also 600 min dauert: $P(X_1 + ... + X_{20} \leq 600)$.

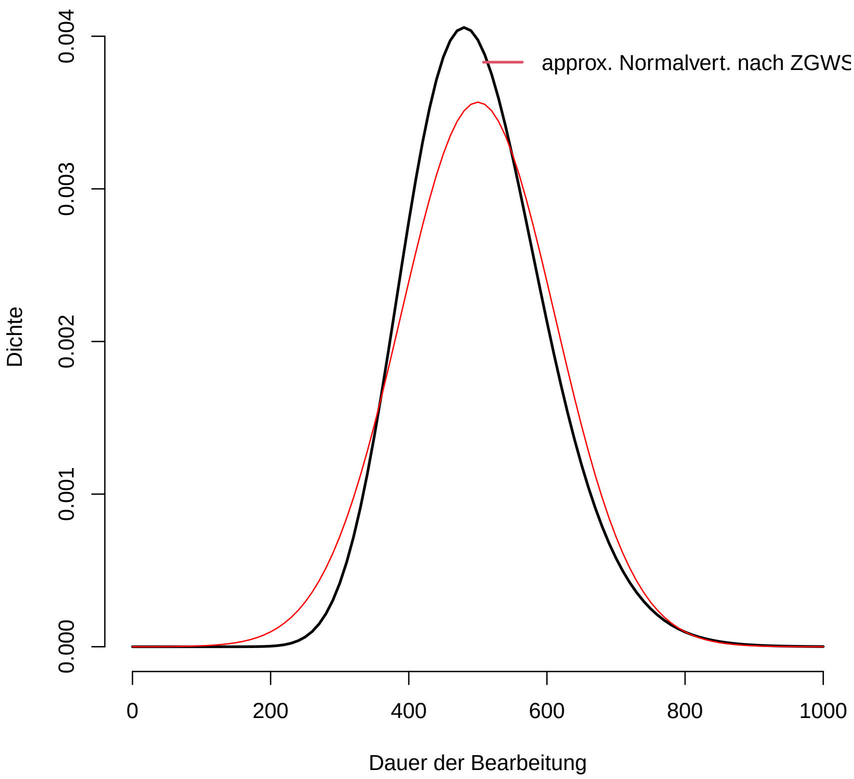
```
pnorm(600,500,sqrt(12500))-pnorm(0,500,sqrt(12500))

## [1] 0.8144494
```

Die Wahrscheinlichkeit beträgt also ca. 81.45 %.

c)

```
curve(dgamma(x,25,1/20),xlab="Dauer der Bearbeitung",from=0,to=1000,bty="n",lwd=2,ylab="Dichte")
curve(dnorm(x,500,sqrt(12500)),col="red", from = 0,  to = 1000, add=TRUE)
legend(480,0.004,"approx. Normalvert. nach ZGWS",lwd=2,col=2,bty="n")
```



Die Approximation aus b) liegt unterhalb der Gammaverteilung und ist somit zu niedrig.

5 Schäzfunktion

a)

$$
\begin{aligned}
\hat{\mu}_1 &= \frac{1}{2}(X_1 + X_2) \\
\mathbb{E}(\hat{\mu}_1) &= \frac{1}{2}(\mathbb{E}(X_1) + \mathbb{E}(X_2)) = \frac{2\mu}{2} = \mu \\
Bias(\hat{\mu}_1) &= \mathbb{E}(\hat{\mu}_1) - \mu = \mu - \mu = 0 \\
Var(\hat{\mu}_1) &= \frac{1}{2^2}(Var(X_1) + Var(X_2)) = \frac{2\sigma^2}{4} = \frac{\sigma^2}{2}
\end{aligned}
$$

$$
\begin{aligned}
\hat{\mu}_2 &= \frac{1}{6}(X_1 + X_2 + X_3) + \frac{1}{2}X_4 \\
\mathbb{E}(\hat{\mu}_2) &= \frac{1}{6}(\mathbb{E}(X_1) + \mathbb{E}(X_2) + \mathbb{E}(X_3)) + \frac{1}{2}\mathbb{E}(X_4) = \frac{3\mu}{6} + \frac{\mu}{2} = \mu \\
Bias(\hat{\mu}_2) &= \mathbb{E}(\hat{\mu}_2) - \mu = \mu - \mu = 0 \\
Var(\hat{\mu}_2) &= \frac{1}{6^2}(Var(X_1) + Var(X_2) + Var(X_3)) + \frac{1}{2^2}Var(X_4) = \frac{3\sigma^2}{36} + \frac{\sigma^2}{4} = \frac{\sigma^2}{3}
\end{aligned}
$$

$$
\begin{aligned}
\hat{\mu}_3 &= \frac{1}{5}(X_1 + X_3) + \frac{1}{6}(X_2 + X_4) + \frac{1}{4}X_5 \\
\mathbb{E}(\hat{\mu}_3) &= \frac{1}{5}(\mathbb{E}(X_1) + \mathbb{E}(X_3)) + \frac{1}{6}(\mathbb{E}(X_2) + \mathbb{E}(X_4)) + \frac{1}{4}\mathbb{E}(X_5) = \frac{2\mu}{5} + \frac{2\mu}{6} + \frac{\mu}{4} = \frac{59}{60}\mu \\
Bias(\hat{\mu}_3) &= \mathbb{E}(\hat{\mu}_3) - \mu = \frac{59}{60}\mu - \mu = \frac{-1}{60}\mu \\
Var(\hat{\mu}_3) &= \frac{1}{5^2}(Var(X_1) + Var(X_3)) + \frac{1}{6^2}(Var(X_2) + Var(X_4)) + \frac{1}{4^2}Var(X_5) \\
&= \frac{2\sigma^2}{25} + \frac{2\sigma^2}{36} + \frac{\sigma^2}{16} \\
&= \frac{713}{3600}\sigma^2
\end{aligned}
$$

Um einen erwartungstreuen Schätzer zu wählen, muss der Bias 0 sein. Dies ist der Fall bei $\hat{\mu}_1$ und $\hat{\mu}_2$. Da $Var(\hat{\mu}_2) < Var(\hat{\mu}_1)$, kann letzendlich $\hat{\mu}_2$ als geeignetster Schätzer gewählt werden, dieser ist also effizient.

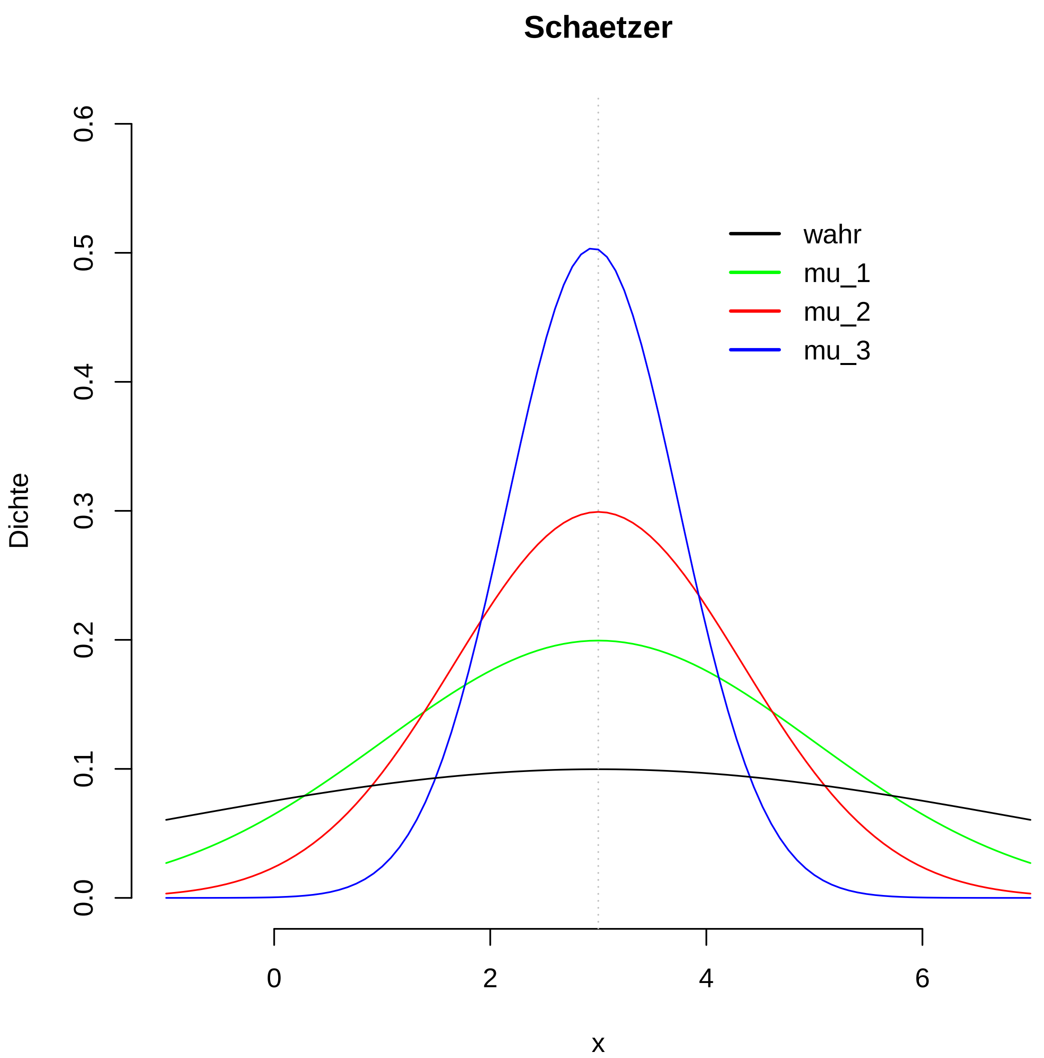
b)

```r
mu <- 3
sigma_sq <- 4

curve(dnorm(x,mu,sigma_sq/2),from=-1,to=7,bty="n",ylim=c(0,0.6),col="green",
      ylab="Dichte", main="Schaetzer")
curve(dnorm(x,mu,sigma_sq/3),from=-1,to=7,bty="n",col="red",add=TRUE)
curve(dnorm(x,(59/60)*mu,(713/3600)*sigma_sq),from=-1,to=7,bty="n",col="blue",add=TRUE)
curve(dnorm(x,mu,sigma_sq),from=-1,to=7,bty="n",add=TRUE)

legend(4,0.51,"mu_1",lwd=2,col="green",bty="n")
legend(4,0.48,"mu_2",lwd=2,col="red",bty="n")
legend(4,0.45,"mu_3",lwd=2,col="blue",bty="n")
legend(4,0.54,"wahr",lwd=2,col="black",bty="n")

abline(v = mu, col="gray", lty="dotted")
```



Schätzer $\hat{\mu}_1$ ist grafisch zu wählen, da dieser am meisten der wahren Verteilung (schwarzer Graph) entspricht.