Stochastik

Serie 9

Aufgabe 9.1

a) Gegeben sind zwei unabhängige Zufallsvariable X und Y mit den Kennwerten $\mu_X = 40, \sigma_X = 15, \mu_Y = 85$ und $\sigma_Y = 18$. Berechnen Sie $\mathrm{E}(X + 2Y)$, $\mathrm{Var}(X + 2Y)$ und $\mathrm{E}(X^2)$.

b) Ein Werk produziert rechteckige Glasscheiben, deren Länge X und Breite Y (in mm gemessen) voneinander unabhängig produktionsbedingten Schwankungen unterliegen. Es gilt $\mu_X = 1000, \sigma_X = 0.02, \mu_Y = 500, \sigma_Y = 0.01$. Wie gross sind Erwartungswert und Standardabweichung des Umfangs U?

c) (**Zusatzaufgabe**) Bestimmen Sie die Wahrscheinlichkeitsdichte $f_X(x)$ der Zufallsvariablen $X = Z^2$, wobei $Z \sim \mathcal{N}(0, 1)$.

Aufgabe 9.2

Erzeugen Sie mit `rnorm` $m = 500$ Stichproben aus einer Standardnormalverteilung mit Umfang $n = 5$. Speichern Sie die Stichproben als $(n \times m)-$ Matrix ab.

R-Hinweis:

```
m <- ...
n <- ...
data.sim <- matrix(rnorm(n = m * n, 0, 1), ncol = m, nrow = 5)
```

a) Stellen Sie die Stichproben als Runs graphisch dar, und staunen Sie über die zahlreichen möglichen Verläufe.

```
plot(data.sim[, 1], type = "l", ylim = c(-5, 5),
     ylab = "Simulierte Werte")
for (i in 1:m) {
     points(data.sim[, i], type = "l", col = i)
}
```

b) Berechnen Sie nun für jede Stichprobe den Mittelwert $\bar{x}$, und erzeugen Sie ein Histogramm der Mittelwerte. Aus der Theorie wissen Sie, dass die Mittelwerte normalverteilt mit Parameter $\mu = 0$ und Standardabweichung $\sigma = \frac{1}{\sqrt{n}}$ sein müssten.

Überprüfen Sie dies graphisch mit Hilfe von Histogrammen und der theoretischen Wahrscheinlichkeitsdichtekurve

```
hist(..., freq = FALSE, ..)
curve(dnorm(x, mean = 0, sd = 1/sqrt(n)), from = -4,
    to = 4, add = TRUE)
```

c) Wiederholen Sie die Aufgabe für $n = 2, 10, 100$.

Aufgabe 9.3

a) Eine Elektronik-Firma stellt Widerstände her, die einen mittleren Widerstand von $100\,\Omega$ und eine Standardabweichung von $10\,\Omega$ haben. Die Widerstände sind normalverteilt. Bestimmen Sie die Wahrscheinlichkeit, dass sich für eine zufällige Stichprobe von $n = 25$ Widerständen ein mittlerer Widerstand unter $95\,\Omega$ ergibt.

b) Nehmen Sie an, dass eine Zufallsvariable X einer uniformen Verteilung folgt, und zwar $X \sim \text{Uniform}[4, 6]$. Wie lautet die Verteilung der Zufallsvariablen $\overline{X}_{40}$, also des Mittelwertes einer Stichprobe vom Umfang $n = 40$.

Aufgabe 9.4

Es seien $\{X_i\}_{1 \leq i \leq 50}$ unabhängige und normalverteilte Zufallsvariablen mit Erwartungswert $\mu = 1$ und Standardabweichung $\sigma = 2$. Darüber hinaus sind folgende Zufallsvariablen definiert:

$$S_n = X_1 + X_2 + \cdots + X_n$$

und

$$\bar{X}_n = \frac{1}{n}(X_1 + X_2 + \cdots + X_n) = \frac{S_n}{n}\,.$$

Dabei ist $n = 50$.

a) Bestimmen Sie die Parameter der Normalverteilung von S_n sowie $\bar{X}_n$.

b) Berechnen Sie die Wahrscheinlichkeit $P\left(\mathrm{E}[X_1] - 1 \leq X_1 \leq \mathrm{E}[X_1] + 1\right)$.

c) Berechnen Sie $P\left(\mathrm{E}[S_n] - 1 \leq S_n \leq \mathrm{E}[S_n] + 1\right)$.

d) Berechnen Sie $P\left(\mathrm{E}[\bar{X}_n] - 1 \leq \bar{X}_n \leq \mathrm{E}[\bar{X}_n] + 1\right)$.

Aufgabe 9.5

Wir betrachten einen Datensatz, bei welchem zwei Methoden zur Bestimmung der latenten Schmelzwärme von Eis verglichen werden. Wiederholte Messungen der freigesetzten Wärme beim Übergang von Eis bei $-0.7°C$ zu Wasser bei $0°C$ ergaben die folgenden Werte (in cal/g):

Methode A	79.98	80.04	80.02	80.04	80.03	80.03	80.04	79.97	80.05
Methode A	80.03	80.02	80.00	80.02					
Methode B	80.02	79.94	79.98	79.97	79.97	80.03	79.95	79.97	

a) Berechnen Sie die arithmetischen Mittelwerte der beiden Methoden und geben Sie für jede Methode den absoluten Fehler an.

b) Wie lauten die relativen Fehler?

Aufgabe 9.6

Wir betrachten einen Datensatz, bei dem die Auswirkung von Vitamin C auf das Zahnwachstum von Meerschweinchen untersucht wurde. Dabei wurde die Länge der Odontoblasten (Zähne) bei jeweils 10 Meerschweinchen für drei unterschiedliche Dosierungen von Vitamin C (0,5 , 1 und 2 mg) gemessen. Das Vitamin C wurde auf zwei verschieden Arten den Meerschweinchen verabreicht, entweder in Form von Orangensaft (OJ) oder Ascorbinsäure (VC).

a) Verschaffen Sie sich eine Übersicht über den Datensatz.

R-Hinweise:

```
df <- ToothGrowth
head(df)
summary(df)
```

```
coplot(len ~ dose | supp,
    data = ToothGrowth,
    panel = panel.smooth,
    xlab = "Laenge vs Dosis bei zwei Verabreichungsarten")
```

b) Nun möchten Sie für jede Dosis und jede Verabreichungsart den arithmetischen Mittelwert und den Standardfehler graphisch darstellen. Berechnen Sie den Standardfehler für die Zahnlänge in jeder Gruppe, die dieselbe Dosis mit der gleichen Verabreichungsart erhalten hat.

```
# Wir organisieren den Datensatz nach
# Varabreichungsart (supp) und Dosis (dose) und
# berechnen Mittelwert und Standardabweichung der
# Zahnlaenge (len) und Anzahl Datenpunkte fuer jede
# Gruppe
len.mean <- aggregate(df[c("len")], by = df[c("supp",
    "dose")], FUN = mean)

len.sd <- aggregate(df[c("len")], by = df[c("supp",
    "dose")], FUN = sd)

len.anzahl <- aggregate(df[c("len")], by = df[c("supp",
    "dose")], FUN = length)
```

```
# Berechnen Sie den Standardfehler fuer jede
# Verabreichungsart und Dosis sem: standard error
# of mean
sem <- len.sd[, "len"]/sqrt(len.anzahl[, "len"])
# Wir fassen die Groessen in einer Datenmatrix
# len.aggregate zusammen
len.aggregate <- cbind(len.mean[, 1:2], len.mean[,
    3], len.sd[, 3], sem)

names(len.aggregate) <- c("supp", "dose", "mean", "sd",
    "sem")
df.oj <- len.aggregate[len.aggregate[, "supp"] == "OJ",
    ]
df.vc <- len.aggregate[len.aggregate[, "supp"] == "VC",
    ]
```

```
# Nun stellen wir den
# Mittelwert und Standardfehler
# fuer jede Dosis graphisch dar
stripchart(len ~ dose, data = ToothGrowth,
    method = "jitter", vertical = T,
    subset = supp == "VC", col = "blue",
    main = "Guinea Pigs' Tooth Growth",
    xlab = "Vitamin C Dosis", ylab = "Zahnlaenge",
```

```
    jit = 0.05, pch = 16)

stripchart(len ~ dose, data = ToothGrowth,
    method = "jitter", vertical = T,
    subset = supp == "OJ", col = "red",
    jit = 0.05, pch = 16, add = T)

arrows(1:3, df.vc[, "mean"] + df.vc[,
    "sem"], 1:3, df.vc[, "mean"] -
    df.vc[, "sem"], angle = 90,
    code = 3, length = 0.3, col = "blue",
    cex = 5)

arrows(1:3, df.oj[, "mean"] + df.oj[,
    "sem"], 1:3, df.oj[, "mean"] -
    df.oj[, "sem"], angle = 90,
    code = 3, length = 0.3, col = "red",
    cex = 5)

lines(1:3, df.vc[, "mean"], pch = 4,
    type = "b", cex = 3, col = "blue")

lines(1:3, df.oj[, "mean"], pch = 4,
    type = "b", cex = 3, col = "red")
```

c) Wie interpretieren Sie die Graphik? Warum benützen wir den Standardfehler und nicht die Standardabweichung als Fehlerbalken?

Aufgabe 9.7

Aus der uniformen Verteilung $X \sim \text{Uniform}([0, 10])$ soll eine Stichprobe vom Umfang n gezogen werden.

a) Es sei $n = 60$. Bestimmen Sie ein symmetrisches Intervall

$$I = [\mu_X - e, \mu_X + e]$$

um den Erwartungswert μ_X so, dass sich das arithmetische Mittel der Stichprobe, also $\overline{X}_{60}$, mit der Wahrscheinlichkeit von 95% in I befindet. Ein solches Intervall heisst **Prognoseintervall**.

Hinweis: Standardisieren Sie das arithmetische Mittel $\overline{X}_n$ und benützen Sie den

Zentralen Grenzwertsatz und $P(\overline{X}_n \in I) = 0.95$.

b) Umgekehrt: Wie gross muss n gewählt werden, damit $e = 0.2$ wird?

c) Überprüfen Sie a) experimentell, d. h. mit R: ziehen Sie viele Stichproben (z. B. 200) und zählen Sie, wie viele ausserhalb von I liegen.

R-Hinweise:

```
n <- 60  # Anzahl Stichproben
# X_1,...,X_n simulieren und in einer n-spaltigen
# Matrix (mit 200 Zeilen) anordnen
sim <- matrix(runif(n * 200, min = 0, max = 10), ncol = n)
# In jeder Matrixzeile Mittelwert berechnen
sim.mean <- apply(sim, 1, "mean")
plot(sim.mean)
# Zeichnen Sie mit abline(h=...) die
# Intervallgrenzen des Prognoseintervalls in der
# obigen Graphik ein.
```

Aufgabe 9.8

Zusatzaufgabe

Die Auswertung eines Integrals

$$\int_a^b f(x)\,\mathrm{d}x$$

kann sehr oft nicht analytisch erfolgen. Der gebräuchlichste Ansatz in diesem Fall besteht darin, das Integral numerisch zu berechnen. Dazu existieren verschiedene Computerprogramme. Eine andere geläufige Methode, um ein solches Integral zu berechnen, ist die sogenannte **Monte Carlo Methode**. Man generiert dabei uniform verteilte Zufallsvariablen auf dem Intervall $[a, b]$, d. h., $X_1, X_2, \ldots, X_n$ i.i.d. $\sim$ Uniform$([a, b])$ und berechnet

$$\frac{b-a}{n} \sum_{i=1}^{n} f(X_i)\,.$$

Dass dieser Ausdruck in etwa $\int_a^b f(x)\,\mathrm{d}x$ ist, möchten wir im Folgenden verstehen. Aufgrund des *Gesetzes der grossen Zahlen* gilt für grosse n

$$(b-a) \cdot \frac{1}{n} \sum_{i=1}^{n} f(X_i) \approx (b-a) \cdot \mathrm{E}[f(X)]\,.$$

Der Erwartungswert von $f(X)$ für $X \sim \text{Uniform}([a,b])$ kann aber auch geschrieben werden als

$$\mathrm{E}[f(X)] = \int_a^b f(x) \frac{1}{b-a} \, \mathrm{d}x \, ,$$

wobei $\frac{1}{b-a}$ die Wahrscheinlichkeitsdichte der Wahrscheinlichkeitsverteilung $\text{Uniform}([a,b])$ ist. Somit gilt

$$\frac{b-a}{n} \sum_{i=1}^{n} f(X_i) \approx (b-a) \cdot \mathrm{E}[f(X)] = (b-a) \cdot \int_a^b f(x) \frac{1}{b-a} \, \mathrm{d}x = \int_a^b f(x) \, \mathrm{d}x \, .$$

Berechnen Sie folgendes Integral

$$I(f) = \frac{1}{\sqrt{2\pi}} \int_0^1 \mathrm{e}^{-x^2/2} \, \mathrm{d}x \, .$$

Berechnen Sie das Integral, indem Sie 1000 uniform über das Intervall $[0,1]$ verteilte Zahlen $X_1, \ldots, X_{1000}$ mit der `R`-Funktion `runif(...)` generieren. Berechnen Sie den genauen numerischen Wert des Integrals mit der `R`-Funktion `pnorm(...)`.

Kurzlösungen vereinzelter Aufgaben

A 9.1:

a) $f_X(x) = \frac{x^{-1/2}}{\sqrt{2\pi}} e^{-x/2}, \quad x \geq 0$

A 9.3:

a) 0.0062

b) $\overline{X}_{40} \sim \mathcal{N}(5, 1/120)$

A 9.5:

a) Methode A: (80.02 ± 0.01)cal/g
Methode B: (79.98 ± 0.01)cal/g

b) Methode A: 80.02cal/g $\pm 8 \cdot 10^{-3}\%$
Methode B: 79.98cal/g $\pm 1 \cdot 10^{-2}\%$

A 9.7:

a) $[4.27, 5.73]$

b) $n = 800$