

## Aufgabe 1

(10 Punkte 20 Minuten)

1. Mit der Tschebyscheff-Ungleichung können keine Aussagen über den Erwartungswert einer Gesamtheit gemacht werden.

*Lösung:* NEIN.

Mit Hilfe der Tschebyscheff-Ungleichung

$$P(|X - E(X)| < k\sigma) \geq 1 - \frac{1}{k^2}$$

kann sowohl ein  $(1 - \frac{1}{k^2})$ -Prognoseintervall aufgestellt werden

$$X \in [E(X) - k\sigma; E(X) + k\sigma]$$

als auch ein  $(1 - \frac{1}{k^2})$ -Konfidenzintervall

$$E(X) \in [x - k\sigma; x + k\sigma].$$

Eine Konfidenzaussage über den Erwartungswert einer Gesamtheit ist also möglich.

2. Wird beim Lotto 6 aus 49 zwei Jahre (104 Ziehungen) lang die Kugel mit der Zahl 13 nicht gezogen aber jedesmal die Kugel mit der Zahl 42, dann ist es aufgrund des Starken Gesetzes der Großen Zahlen wahrscheinlicher die "13" im nächsten Spiel zu ziehen als die "42".

*Lösung:* NEIN.

Das Gesetz der großen Zahlen gilt nur in Wahrscheinlichkeit und die einzelnen Ziehungen sind von einander unabhängig und damit ist die Wahrscheinlichkeit für eine "13" nicht höher als für eine "42".

3. Die Kleinste-Quadrate-Schätzung im linearen Regressionsmodell ist eine ML-Schätzung, wenn die Störterme unabhängig identisch normalverteilt sind.

*Lösung:* JA.

Sind die Fehler normalverteilt, dann ist die Likelihoodfunktion für das Regressionsproblem

$$l(\beta_0, \beta_1 | x_1, \dots, x_n) = e^{-\left(\frac{\sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i)^2}{2\sigma}\right)}.$$

Maximiert dieser Funktion ist gleichbedeutend mit der Minimierung der Klammer im Exponenten, also mit der KQ-Methode.

4. Ein approximatives Konfidenzintervall für den Parameter  $\pi$  einer  $B(n, \pi)$ -verteilten Zufallsvariable ist nicht symmetrisch zum Schätzwert  $\hat{\pi}$ , wenn die Varianz  $\pi(1 - \pi)$  nicht geschätzt wird.

*Lösung:* JA.

Das approximative Prognoseintervall hat die Form

$$n\pi - \tau_{1-\alpha} \sqrt{\frac{\pi(1-\pi)}{n}} \leq p \leq n\pi + \tau_{1-\alpha} \sqrt{\frac{\pi(1-\pi)}{n}}.$$

Auflösen nach  $\pi$  ergibt (siehe Skript 7.LATEX, S. 217, (4.129)).

5. Es seien zehn Bodenproben mit jeweils dem gleichen Schadstoffgehalt  $\mu$  gegeben. Für eine der zehn Bodenproben werden fünf Messwerte  $x_1, \dots, x_5$  erhoben. Dabei seien die Messungen Realisationen von unabhängig identisch normalverteilten Zufallsvariablen mit Mittelwert  $\mu$  und bekannter Varianz  $\sigma^2$ .

Dann garantiert das 90%-Konfidenzintervall  $\left[ \bar{x} - \tau_{0.95} \frac{\sigma}{\sqrt{5}}; \bar{x} + \tau_{0.95} \frac{\sigma}{\sqrt{5}} \right]$  für den Schadstoffgehalt  $\mu$  basierend auf den 5 Messung der einen Bodenprobe, daß man mit der Aussage "der Schadstoffgehalt liegt im Bereich des 90%-Konfidenzintervalls" bei mindestens 9 von 10 Bodenproben richtig liegt.

*Lösung:* NEIN.

Basierend auf den fünf Messwerten erhält man ein Konfidenzintervall  $K(x_1, \dots, x_5)$ . Darin liegt der Schadstoffgehalt  $\mu$  der betrachteten Bodenprobe oder er liegt nicht drin. Da alle anderen Bodenproben den gleichen Schadstoffgehalt haben, sind also entweder der Schadstoffgehalt aller zehn Bodenproben im Konfidenzintervall oder keine.

## Aufgabe 2

(10 Punkte 15 Minuten)

In einer Wetterstation werden die Phänomene Frühnebel und Regen beobachtet. Langjährige Untersuchungen haben ergeben, daß es im Schnitt alle drei Tage regnet und jeden zweiten Tag Frühnebel zu beobachten ist. Alle fünf Tage ist im Schnitt sowohl Frühnebel als auch Regen zu beobachten.

1. Berechnen Sie die Wahrscheinlichkeit, daß mindestens eins der beiden Wetterphänomene an einem Tag zu beobachten ist.

*Lösung:*  $F$  entspreche dem Ereignis Frühnebel,  $R$  dem Ereignis Regen

$$\begin{aligned} P(F) &= \frac{1}{2}, \\ P(R) &= \frac{1}{3}, \end{aligned}$$

$$P(F \cap R) = \frac{1}{5},$$

$$\begin{aligned} P(F \cup R) &= P(F) + P(R) - P(F \cap R) \\ &= \frac{1}{2} + \frac{1}{3} - \frac{1}{5} \\ &= \frac{15 + 10 - 6}{30} \\ &= \frac{19}{30} \approx 0.6\bar{3}. \end{aligned}$$

2. Berechnen Sie die Wahrscheinlichkeit, daß es an einem Tag mit Frühnebel regnet.

*Lösung:* Zu berechnen ist die bedingte Wahrscheinlichkeit für Regen unter der Bedingung Frühnebel.

$$\begin{aligned} P(R|F) &= \frac{P(R \cap F)}{P(F)} \\ &= \frac{\frac{1}{5}}{\frac{1}{2}} \\ &= 0.4. \end{aligned}$$

Die Wahrscheinlichkeit, daß es an einem Tag mit Frühnebel regnet, ist 40 Prozent.

3. Berechnen Sie die Wahrscheinlichkeit, daß es an einem Tag ohne Frühnebel regnet.

*Lösung:* Zu berechnen ist die bedingte Wahrscheinlichkeit für Regen unter der Bedingung kein Frühnebel.

$$\begin{aligned} P(R|\bar{F}) &= \frac{P(R \cap \bar{F})}{P(\bar{F})} = (*) \\ P(R \cap \bar{F}) &= P(R) - P(R \cap F) \\ &= \frac{1}{3} - \frac{1}{5} \\ &= \frac{2}{15} \\ (*) &= \frac{\frac{2}{15}}{\frac{1}{2}} \\ &= \frac{4}{15} \approx 0.2\bar{6}. \end{aligned}$$

Die Wahrscheinlichkeit, daß es an einem Tag ohne Frühnebel regnet, beträgt 26.7 Prozent.

**Aufgabe 3***(20 Punkte 20 Minuten)*

Schießt der Bogenschütze Gutfreud auf eine Zielscheibe, so ist der Abstand des Treffpunktes vom Scheibenmittelpunkt als stetige reelle Zufallsvariable mit der Wahrscheinlichkeitsfunktion

$$f(x) = \begin{cases} c(1-x) & \text{für } 0 \leq x \leq 1 \\ 0 & \text{sonst} \end{cases}$$

aufzufaßen, wobei  $c$  eine geeignet zu wählende Konstante ist.

1. Berechnen Sie die Konstante  $c$ .

(Hinweis: Sollten Sie  $c$  nicht berechnen können, dürfen Sie in den weiteren Aufgabenteilen  $c$  beibehalten)

*Lösung:* Die Fläche unter der Dichtefunktion muß Eins betragen.

$$\begin{aligned} \int_0^1 c(1-x) dx &= 1 \\ \Leftrightarrow c(x - \frac{1}{2}x^2) \Big|_0^1 &= 1 \\ \Leftrightarrow c(1 - \frac{1}{2}) &= 1 \\ \Leftrightarrow c &= 2. \end{aligned}$$

Die Dichtefunktion lautet also

$$f(x) = \begin{cases} 2(1-x) & \text{für } 0 \leq x \leq 1 \\ 0 & \text{sonst} \end{cases}$$

und die Verteilungsfunktion

$$F(x) = 2x - x^2, \quad 0 \leq x \leq 1$$

2. Wie groß ist der Erwartungswert und wie groß ist die Varianz des Abstandes?

*Lösung:* Berechnung des Erwartungswertes

$$\begin{aligned} \int_0^1 2x(1-x) dx &= \int_0^1 2x - 2x^2 dx \\ &= x^2 - \frac{2}{3}x^3 \Big|_0^1 \\ &= 1 - \frac{2}{3} \\ E(X) &= \frac{1}{3}. \end{aligned}$$

Berechnung der Varianz

$$\begin{aligned}\int_0^1 2x^2(1-x) dx &= \int_0^1 2x^2 - 2x^3 dx \\ &= \left. \frac{2}{3}x^3 - \frac{1}{2}x^4 \right|_0^1 \\ &= \frac{2}{3} - \frac{1}{2} \\ E(X^2) &= \frac{1}{6}\end{aligned}$$

Damit ist die Varianz  $\text{Var}(X) = E(X^2) - E(X)^2 = \frac{1}{6} - \frac{1}{9} = \frac{1}{18}$  und der Erwartungswert  $E(X) = \frac{1}{3}$ .

3. Wie groß ist die Wahrscheinlichkeit, daß der Abstand

(a) kleiner als 0,2 ist?

*Lösung:*  $P(X < 0.2) = F(0.2) = 0.36$

(b) zwischen 0,1 und 0,2 beträgt?

*Lösung:*  $P(0.1 \leq X \leq 0.2) = F(0.2) - F(0.1) = 0.36 - 0.19 = 0.17$

4. Wie groß ist annähernd die Wahrscheinlichkeit, daß von 1000 voneinander unabhängigen Schüssen Guttreffs wenigstens 360 einen Abstand von weniger als 0,2 vom Scheibenmittelpunkt haben.

*Lösung:* Betrachte

$$Y_i = \begin{cases} 1 & \text{Treffer mit weniger Abstand als 0.2} \\ 0 & \text{sonst} \end{cases}.$$

Dann ist  $Y = \sum Y_i$  binomialverteilt mit  $n = 1000$  und  $p$ . Die Bestimmung von  $p$  erfolgt über die Verteilungsfunktion und ist identisch mit 3a, also  $p = 0.36$ .

Näherungsweise ist  $Y$  dann normalverteilt mit Erwartungswert  $\mu = 360$  und Varianz  $\sigma^2 = np(1-p) = 259.2$ , da  $np > 5$  und  $n(1-p) > 5$ . Somit läßt sich die Wahrscheinlichkeit, daß von 1000 Schuß wenigstens 360 einen Abstand von weniger als 0.2 vom Scheibenmittelpunkt haben, berechnen als

$$\begin{aligned}P(Y > 360) &= 1 - P(Y \leq 360) \\ &= P\left(\frac{Y - \mu}{\sigma} \leq \frac{360 - 360}{\sqrt{259.2}}\right) \\ &= P(Y^* \leq 0) \\ &= \Phi(0) \\ &= \frac{1}{2}.\end{aligned}$$

**Aufgabe 4***(10 Punkte 20 Minuten)*

Um den Benzinverbrauch des Autos der Familie Meyer gibt es Ärger. Vater Meyer behauptet, daß durch die rasante Fahrweise seines Sohnes der Verbrauch pro Kilometer zunimmt. Sohn Meyer ist ganz anderer Meinung und schlägt vor, ein paar Testfahrten zu machen. Die allwöchentliche Fahrt zur Großmutter, die 300 Kilometer entfernt wohnt, wird dazu genutzt. Drei Wochen hintereinander wird die Strecke mit einer Durchschnittsgeschwindigkeit von 100, 110 und 120 km/h zurückgelegt und der Benzinverbrauch festgestellt. Das Resultat ist

	1. Woche	2. Woche	3. Woche	
Durchschnittsgeschwindigkeit	100	110	120	km/h
Benzinverbrauch	49.6	50.2	51.4	Litern pro Fahrt
	8.2	8.3	8.5	Litern pro 100 km

Der Sohn, der sich in statistischen Schlußweisen ein wenig auskennt, paßt folgendes Regressionsmodell an

Benzinverbrauch pro Fahrt = Konstante + Koeffizient · Durchschnittsgeschwindigkeit + Störterm .

Beide einigen sich darauf, daß die Störterme als unabhängig normalverteilt angesehen werden können. Zusätzlich setzen sie für Ihre Testentscheidungen ein Signifikanzniveau von  $\alpha = 10\%$  fest, obwohl sie nicht recht wissen, was das zu bedeuten hat.

Sie geben die Werte der Testfahrten in das Programmpaket R ein und erhält folgendes Ergebnis:

```

Residual Standard Error    0.2449
R-Square                    0.9643
F-statistic (df=1, 1)      27    p-value    0.121

              Estimate  Std.Err  t-value  Pr(>|t|)
Intercept    40.50    1.9105  21.1987   0.030
X             0.09    0.0173   5.1962   0.121

```

1. Erklären Sie Vater und Sohn, was ein Test zum Signifikanzniveau  $\alpha$  ist.

*Lösung:* Siehe Skript S. 237, (5.7)

2. Der Vater will nun mittels eines statistischen Test die Behauptung seines Sohnes, daß der Benzinverbrauch nicht mit höherer Durchschnittsgeschwindigkeit steigt, widerlegen.

(a) Stellen Sie die Nullhypothese und die Alternative auf.

*Lösung:* Der Sohn behauptet, daß der Benzinverbrauch mit höherer Gewindigkeit nicht steigt. Also lautet die Nullhypothese  $H_0 : \beta_1 \leq 0$  und die Alternative  $H_1 : \beta_1 > 0$ .

(b) Geben Sie den Wert der Prüfgröße und den Ablehnbereich an.

*Lösung:* Der Wert der Prüfgröße (5.1962) kann aus den R-Ergebnisse abgelesen werden. Ebenso kann er berechnet werden als

$$\frac{\hat{\beta}_1 - 0}{\hat{\sigma}_{(\beta_1)}} = \frac{0.09}{0.0173} = 5.2023.$$

Den Ablehnbereich kann man dann aus der Tabelle für die t-Verteilung ablesen. Dabei ist die Anzahl der Freiheitsgrad  $\#(\text{Beobachtungen}) - \#(\text{zu schätzender Parameter}) = 3 - 2 = 1$ . Das  $(1 - \alpha)$ -Quantile der t-Verteilung mit einem Freiheitsgrad für  $\alpha = 0.1$  ist 3.0777. Der Ablehnbereich ist also  $[3.0777; \infty]$ .

(c) Wie lautet die Testentscheidung?

*Lösung:* Der Wert der Prüfgröße liegt im Ablehnbereich. Die Nullhypothese ist also zu einem Signifikanzniveau von 90 Prozent abzulehnen. Die Behauptung, daß der Benzinverbrauch mit steigender Durchschnittsgeschwindigkeit nicht wächst, ist also angesichts der Testfahren nicht haltbar.

## Aufgabe 5

(10 Punkte 15 Minuten)

Sei eine Zufallsvariable  $X$  gleichverteilt auf dem Intervall  $[a; b]$ , d.h. die Dichtefunktion von  $X$  ist

$$f(x) = \begin{cases} \frac{1}{b-a} & a \leq x \leq b \\ 0 & \text{sonst} \end{cases},$$

wobei  $a$  und  $b$  unbekannt mit  $a < b$ .

1. Stellen Sie die Likelihoodfunktion für eine Zufallsstichprobe der Länge  $n$  auf.

*Lösung:* Die gemeinsame Dichtefunktion von  $X_1, \dots, X_n$  lautet

$$f(x_1, \dots, x_n | a, b) = \prod_{i=1}^n \frac{1}{b-a} \mathbb{I}_{[a;b]}(x_i).$$

Damit ist die Likelihood gegeben durch

$$\begin{aligned}l(a, b|x_1, \dots, x_n) &= \prod_{i=1}^n \frac{1}{b-a} \mathbb{I}_{[a;b]}(x_i) \\ &= \frac{1}{b-a} \prod_{i=1}^n \mathbb{I}_{[a;b]}(x_i).\end{aligned}$$

2. Bestimmen Sie den ML-Schätzer für  $a$  und  $b$ .

(HINWEIS: Versuche Sie nicht die Likelihoodfunktion abzuleiten!)

*Lösung:* Die Likelihood ist also immer Null, wenn eine Beobachtung nicht zwischen  $a$  und  $b$  liegt. Damit muß gelten  $\hat{a}_{ML} \leq \min\{x_1, \dots, x_n\}$  und  $\hat{b}_{ML} \leq \max\{x_1, \dots, x_n\}$ . Die Likelihoodfunktion hängt dann nur noch von  $\frac{1}{b-a}$  ab und wird maximal, wenn  $\hat{a}_{ML} = \min\{x_1, \dots, x_n\}$  und  $\hat{b}_{ML} = \max\{x_1, \dots, x_n\}$ , da dann die Differenz im Nenner minimal wird.

3. Wie lautet für die Stichproben 3.2; 4.3; 1.3; 4.5; 2.2 die ML-Schätzung für  $a$  und  $b$ ?

*Lösung:* Minimaler Wert der Stichproben ist 1.3 und somit die ML-Schätzung für  $a$ . Maximaler Wert der Stichprobe ist 4.5 und somit  $\hat{b}_{ML} = 4.5$ .