Statistik 1 für WInf, WI(MB), WI(ET), WI(BI)
Übung 6, Lösungsvorschlag

Gruppenübung

G 18 Maximum-Likelihood-Methode

Aus Erfahrung sei bekannt, dass die störungsfreie Betriebsdauer (in 1000 Stunden) eines bestimmten Systems durch eine stetig verteilte Zufallsvariable X mit der Dichte

$$f(x) = \begin{cases} 2\theta x e^{-\theta \cdot x^2} & \text{für } x > 0 \\ 0 & \text{sonst} \end{cases}$$

mit einem unbekannten Parameter $\theta > 0$ beschrieben werden kann. Bestimmen Sie aus den folgenden 20 Betriebsdauern (in 1000 Stunden) den Maximum-Likelihood-Schätzwert für $\tau(\theta) = \theta$.

1.530	1.173	1.832	1.075	1.539
0.998	2.083	0.693	2.529	1.693
1.325	1.487	1.298	1.743	1.432
1.369	0.987	2.222	1.818	1.505

Likelihood-Funktion zu $x_1, x_2, \ldots, x_{20}$:

$$L(\theta;\, x_1, x_2, \ldots, x_{20}) \;=\; \prod_{i=1}^{20} f_\theta(x_i) \;=\; (2\,\theta)^{20} \cdot \left(\prod_{i=1}^{20} x_i\right) \cdot e^{-\theta \cdot \sum_{i=1}^{20} x_i^2}$$

Log-Likelihood-Funktion zu $x_1, x_2, \ldots, x_{20}$:

$$g(\theta) := \ln L(\theta;\, x_1, x_2, \ldots, x_{20}) = 20 \cdot \ln(2\theta) + \sum_{i=1}^{20} \ln(x_i) - \theta \cdot \sum_{i=1}^{20} x_i^2$$

Bestimmung der Maximalstelle von $g(\theta)$*:*

$$g'(\theta) = \frac{20}{\theta} - \sum_{i=1}^{20} x_i^2 \overset{!}{=} 0$$

Hieraus folgt

$$\theta = \frac{20}{\sum_{i=1}^{20} x_i^2}.$$

Wegen $g''(\theta) = -\frac{20}{\theta^2} < 0$ *handelt es sich dabei tatsächlich um eine Maximalstelle.*

Also ist der ML-Schätzer gleich

$$\hat{\theta} = \frac{20}{\sum_{i=1}^{20} x_i^2}.$$

Aus den Daten ergibt sich $\hat{\theta} = \frac{20}{49.77} \approx 0.4018$.

G 19 Konfidenzintervalle

In einer Stadt liegen für 161 Jahre Niederschlagsmessungen im Monat April vor. Die Meßreihe $x_1, x_2, \ldots, x_{161}$ (x_i = Niederschlagshöhe in mm im i-ten Jahr) hat das arithmetische Mittel $\bar{x}_{(161)} = 53.68$ und die empirische Standardabweichung $s_{(161)} = 6.13$. Es wird angenommen, dass die Werte $x_1, x_2, \ldots, x_{161}$ eine Realisierung von 161 unabhängigen, identisch $N(\mu, \sigma^2)$ - verteilten Zufallsvariablen sind. Bestimmen Sie mit Konfidenzschätzverfahren zum Niveau $1 - \alpha = 0.95$ je ein konkretes Schätzintervall

a) für μ,

Konfidenzintervall (Grenzen sind Zufallsvariablen!) für μ bei unbekannter Varianz:

$$I(X_1, \ldots, X_{161}) = \left[\bar{X}_{(161)} - t_{160;\,0.975} \cdot \sqrt{\frac{S^2_{(161)}}{161}}\,;\, \bar{X}_{(161)} + t_{160;\,0.975} \cdot \sqrt{\frac{S^2_{(161)}}{161}}\right]$$

Mit dem Hinweis folgt für das konkrete Schätzintervall (basierend auf der Realisierung $x_1, \ldots, x_{161}$)

$$\begin{aligned} I(x_1, \ldots, x_{161}) &= \left[53.68 - 1.975 \cdot \sqrt{\frac{6.13^2}{161}}\,;\, 53.68 + 1.975 \cdot \sqrt{\frac{6.13^2}{161}}\right] \\ &\approx [52.7259\,;\, 54.6341]. \end{aligned}$$

Hinweis: Sollte eine entsprechende Quantiltabelle nicht vorhanden sein, kann man für große n auch die Näherung $t_{n,p} \approx u_p$ mit Hilfe der Quantile u_p der Standardnormalverteilung verwenden.

b) für σ^2,

Konfidenzintervall für σ^2:

$$I(X_1, \ldots, X_{161}) = \left[\frac{(n-1)S^2_{(161)}}{\chi^2_{160;\,0.975}}\,;\, \frac{(n-1)S^2_{(161)}}{\chi^2_{160;\,0.025}}\right]$$

Mit $\chi^2_{n;p} \approx n + u_p \cdot \sqrt{2n}$ für große n folgt für das konkrete Schätzintervall

$$I(x_1, \ldots, x_{161}) \approx \left[\frac{160 \cdot 6.13^2}{195.062}\,;\, \frac{160 \cdot 6.13^2}{124.938}\right] \approx [30.8226\,;\, 48.1221].$$

c) für μ unter der Voraussetzung $\sigma^2 = 6.13^2$.

Konfidenzintervall für μ bei bekannter Varianz:

$$I\left(X_1,\ldots,X_{161}\right) = \left[\bar{X}_{(161)} - u_{0.975}\cdot\frac{\sigma}{\sqrt{161}}\,;\,\bar{X}_{(161)} + u_{0.975}\cdot\frac{\sigma}{\sqrt{161}}\right]$$

Als konkretes Schätzintervall ergibt sich hier ein etwas kleineres als in a) (dank der Information über σ^2), nämlich

$$\begin{aligned} I\left(x_1,\ldots,x_{161}\right) &= \left[53.68 - 1.96\cdot\frac{6.13}{\sqrt{161}}\,;\,53.68 + 1.96\cdot\frac{6.13}{\sqrt{161}}\right] \\ &\approx [52.7331\,;\,54.6269]. \end{aligned}$$

G 20 Bisher ist der Betreiber des öffentlichen Verkehrsnetzes in einer Großstadt davon ausgegangen, dass 35% der Fahrgäste Zeitkarteninhaber sind. Bei einer Fahrgastbefragung geben 112 der insgesamt 350 Befragten an, dass sie eine Zeitkarte benutzen. Testen Sie zum Niveau $\alpha = 0.05$, ob sich der Anteil der Zeitkarteninhaber verändert hat. Formulieren Sie die Fragestellung zunächst als statistisches Testproblem.

Das statistische Testproblem lautet hier

$$H_0:\ p = p_0 = 0.35 \qquad \textit{vs.} \qquad H_1:\ p \neq p_0 = 0.35.$$

Es handelt sich also um einen Test auf den unbekannten Anteil in der Grundgesamtheit. Da der Stichprobenumfang sehr groß ist, kann der approximative Binomialtest angewendet werden. Damit lautet die Prüfgröße

$$T(X_1,\ldots,X_n) = \frac{S_n - np_0}{\sqrt{np_0(1-np_0)}}.$$

Dabei wird H_0 zum Niveau $\alpha = 0.05$ verworfen, falls $|T| > z_{1-\frac{\alpha}{2}} = z_{0.975} = 1.96$ gilt.
Mit $n = 350$, $p_0 = 0.35$ und $\sum_{i=1}^{n} x_i = 112$ ergibt sich

$$|T(x_1,\ldots,x_n)| = \left|\frac{112 - 122.5}{\sqrt{350\cdot 0.35\cdot 0.65}}\right| = |-1.177| = 1.177\,.$$

Da $1.177 \not> 1.96$, kann H_0 zm Niveau $\alpha = 0.05$ nicht verworfen werden, d. h. die Beobachtung von 112 Zeitkarteninhabern spricht nicht dafür, dass sich der Anteil an Zeitkarteninhabern verändert hat.

G 21 Es wird angenommen, dass die vorliegende Messreihe $x_1, \ldots, x_{25}$ (s. Tabelle) eine Realisierung von 25 unabhängigen $N(\mu, \sigma^2)$-verteilten Zufallsvariablen ist.

i	1	2	3	4	5	6	7	8	9	10	11	12	13
x_i	704	695	703	693	698	694	701	706	705	697	701	705	693
i	14	15	16	17	18	19	20	21	22	23	24	25	
x_i	702	686	696	703	693	712	706	693	702	699	693	695	

Für die Messreihe gilt $\sum_{i=1}^{25} x_i = 17475$ und $\sum_{i=1}^{25} x_i^2 = 12\,215\,867$.

a) Überprüfen Sie die Hypothese $H_0 : \mu \geq 700$ gegen $H_1 : \mu < 700$ mit einem geeigneten Test zum Niveau $\alpha = 0.05$.

geeigneter Test: t-Test mit $\alpha = 0.05$

- *Nullhypothese:* $H_0 :\ \mu \geq 700$ *vs.* $H_1 : \mu < 700$ *(einseitig)!*
- *Testgröße:*

$$T(X_1, \ldots, X_n) = \sqrt{n}\frac{\bar{X}_{(n)} - \mu_0}{\sqrt{S^2_{(n)}}}$$

- *Realisierung der Testgröße:*

$$T(x_1, \ldots, x_{25}) \; = \; \sqrt{25} \cdot \frac{699 - 700}{\sqrt{35.0833}} \; \approx \; -0.8441$$

denn

$$\bar{x} \; = \; \frac{1}{25} \sum_{i=1}^{25} x_i \; = \; \frac{17475}{25} \; = \; 699$$

und

$$s^2 \; = \; \frac{1}{24} \left(\sum_{i=1}^{25} x_i^2 - 25 \cdot \bar{x}^2 \right) \; = \; \frac{842}{24} \; \approx \; 35.0833$$

- *Kritischer Bereich:*

$$K = \{(x_1, \ldots, x_{25}) \mid T(x_1, \ldots, x_{25}) < t_{24;0.05}\}$$

- *Entscheidung: Wegen* $-0.8441 \geq t_{24;0.05} = -1.7109$ *wird* H_0 *nicht verworfen.*

b) Es sei nun vorausgesetzt, dass für obige Messwerte $\sigma^2 = 25$ bekannt ist. Überprüfen Sie nun mit einem anderen Test dieselbe Nullhypothese wie in Aufgabenteil a) zum Niveau $\alpha = 0.05$.

geeigneter Test: Gauß-Test mit $\alpha = 0.05$

- *Nullhypothese:* $H_0: \ \mu \geq 700$ *vs.* $H_1: \mu < 700$ *(einseitig)!*
- *Testgröße:*

$$T(X_1, \dots, X_n) = \sqrt{n}\frac{\bar{X}_{(n)} - \mu_0}{\sigma_0}$$

- *Realisierung der Testgröße:*

$$T(x_1, \dots, x_{25}) \ = \ \sqrt{25} \cdot \frac{699 - 700}{25} \ = \ -1$$

- *Kritischer Bereich:*

$$K = \{(x_1, \dots, x_{25}) \mid T(x_1, \dots, x_{25}) < u_{0.05}\}$$

- *Entscheidung: Wegen* $-1 \geq -1.645 = u_{0.05}$ *wird* H_0 *nicht verworfen.*

Hausübung

H 23 Maximum-Likelihood-Methode

Ein Tierpark besitzt sieben Exemplare einer seltenen Tierart. In einem biologischen Forschungsinstitut wurde eine bisher unbekannte, nicht ansteckende Krankheit an Tieren dieser Rasse entdeckt. Um herauszufinden, wie viele seiner Tiere von der Krankheit befallen sind, lässt der Leiter des Tierparks an drei aufeinander folgenden Tagen je ein Tier fangen und auf die Krankheit hin untersuchen, um es anschließend wieder ins Gehege zu entlassen. Die Zufallsvariable X_i nehme den Wert 1 an, falls das am i-ten Tag gefangene Tier für krank befunden wird und sonst den Wert 0 ($i = 1, \dots, 3$). Die unbekannte Anzahl der kranken Tiere sei θ.

a) Bestimmen Sie die Verteilung der Zufallsvariablen X_1, indem Sie die Wahrscheinlichkeiten $P_\theta(X_1 = 1)$ und $P_\theta(X_1 = 0)$ angeben.

Da von den 7 Tieren genau θ Tiere die Krankheit haben, gilt

$$P_\theta(X_1 = 1) = \frac{\theta}{7} \qquad \textit{und} \qquad P_\theta(X_1 = 0) = \frac{7 - \theta}{7}.$$

b) Die Krankheit wurde lediglich bei den an den ersten beiden Tagen untersuchten Tieren diagnostiziert. Geben Sie auf der Basis dieser Beobachtung an, welche Werte für θ in Frage kommen und welcher dieser Werte am plausibelsten für die Anzahl an kranken Tieren ist.

Für θ kommen alle natürliche Zahlen zwischen 1 und 6 in Frage, da ja mindestens ein gesundes und ein krankes Tier im Gehege leben.
Die Frage nach dem plausibelsten Wert führt uns zur Maximum-Likelihood-Methode. Diese Methode liefert denjenigen Wert θ, bei dem die Wahrscheinlichkeit für die gemachte Beobachtung am größten ist. Diese Wahrscheinlichkeit entspricht im diskreten Fall gerade der sogenannten Likelihoodfunktion L, die wir nun für das konkrete Beispiel angeben wollen.

Da die gefangenen Tiere nach der Untersuchung wieder ins Gehege entlassen werden und man davon ausgehen kann, dass sich die Zahl der kranken Tiere über die Tage nicht ändert (Krankheit ist nicht ansteckend), besitzen die Zufallsvariablen X_2 und X_3 die gleiche Verteilung wie X_1. Die in der Aufgabenstellung beschriebene Beobachtung kann als Realisierung der unabhängigen Zufallsvariablen X_1, X_2 und X_3 interpretiert werden, die konkret

$$(x_1, x_2, x_3) = (1, 1, 0)$$

lautet. Als Likelihoodfunktion

$$L(\theta; x_1, x_2, x_3) = P_\theta(X_1 = x_1) \cdot P_\theta(X_2 = x_2) \cdot P_\theta(X_3 = x_3)$$

zu dieser Beobachtung ergibt sich damit

$$\begin{aligned} L(\theta; 1, 1, 0) &= P_\theta(X_1 = 1) \cdot P_\theta(X_2 = 1) \cdot P_\theta(X_3 = 0) \\ &= \frac{\theta}{7} \cdot \frac{\theta}{7} \cdot \frac{7-\theta}{7} = \frac{\theta^2(7-\theta)}{7^3} \; . \end{aligned}$$

Als konkreten Maximum-Likelihood-Schätzwert wählen wir nun dasjenige $\hat{\theta}$, das für die aufgetretene Realisierung am plausibelsten erscheint. Dies ist gleichzeitig derjenige Parameter, für den die Likelihoodfunktion ihren größten Wert annimmt. Wir berechnen nun die verschiedenen Funktionswerte von $L(\theta; 1, 1, 0)$ für diese Werte θ:

θ	$L(\theta; 1, 1, 0)$	
1	0.0175	
2	0.0583	
3	0.1050	
4	0.1399	
5	0.1458	$\leftarrow$
6	0.1050	

Bei $\hat{\theta} = 5$ nimmt die Likelihoodfunktion ihren maximalen Wert an. Deshalb schätzen wir, dass fünf kranke Tiere im Gehege leben.

H 24 Maximum-Likelihood-Schätzer, verschobene Exponentialverteilung

Die Zufallsvariablen $X_1, \ldots, X_n$ seien unabhängig und identisch verteilt mit der Verteilungsfunktion

$$F_\theta(x) = \begin{cases} 1 - e^{-(x-\theta)} & \text{falls } x \geq \theta \\ 0 & \text{sonst.} \end{cases}$$

a) Bestimmen Sie den Maximum-Likelihood-Schätzer T_n für $\tau(\theta) = \theta$.

b) Geben Sie die Verteilungsfunktion von T_n an.

c) Bestimmen Sie den Bias und den mittleren quadratischen Fehler von T_n.

d) Ist die Folge der Schätzer $T_1, T_2, \ldots$ asymptotisch erwartungstreu für $\tau(\theta) = \theta$?

a) *Für Realisierungen $x_1, \ldots, x_n$ der Zufallsvariablen $X_1, \ldots, X_n$ ist*

$$L(\theta) \;=\; \prod_{i=1}^{n} f(x_i|\theta) \;=\; \begin{cases} \prod_{i=1}^{n} e^{-(x_i-\theta)} & \text{falls } x_i \geq \theta \quad \forall i, \\ 0 & \text{sonst.} \end{cases}$$

die zu maximierende Likelihood-Funktion, d.h.

$$\ln L(\theta) \;=\; -\sum_{i=1}^{n}(x_i - \theta) \;=\; n(\theta - \bar{x}), \qquad \text{für } x_i \geq \theta \quad \forall i.$$

Die Log-Likelihoodfunktion ist monoton wachsend auf $(-\infty, \min_{1\leq i\leq n} x_i]$ und fällt unter alle Schranken $(-\infty)$ auf $(\min_{1\leq i\leq n} x_i, \infty)$. Daher ist der ML-Schätzer gleich

$$T_n(X_1, \ldots, X_n) \;=\; \min\{X_1, \ldots, X_n\}.$$

b) *Verteilungsfunktion von $T_n(X_1, \ldots, X_n) \;=\; \min\{X_1, \ldots, X_n\}$:*
Die Verteilungsfunktion $G_\theta(x)$ von T_n ist gegeben durch

$$G_\theta(x) \;=\; 1 - (1 - F_\theta(x))^n \;=\; \begin{cases} 1 - e^{-n(x-\theta)} & \text{falls } x \geq \theta, \\ 0 & \text{sonst.} \end{cases}$$

c) *Die Zufallsvariable $Y_n := T_n - \theta$ ist wegen*

$$P(Y_n \leq x) \;=\; P(T_n \leq x + \theta) \;=\; F_\theta(x + \theta) \;=\; \begin{cases} 1 - e^{-nx} & \text{falls } x \geq 0, \\ 0 & \text{sonst,} \end{cases}$$

exponentialverteilt mit dem Parameter $\lambda = n$, woraus $E_\theta(Y_n)\frac{1}{n}$ und $Var_\theta(Y_n) = \frac{1}{n^2}$ folgen.
Damit gelten

$$\begin{aligned} E_\theta(T_n) &= E_\theta(Y_n + \theta) \;=\; \frac{1}{n} + \theta, \\ Var_\theta(T_n) &= Var_\theta(Y_n + \theta) \;=\; Var_\theta(Y_n) \;=\; \frac{1}{n^2}. \end{aligned}$$

Für den mittleren quadratischen Fehler erhält man

$$E_\theta((T_n - \theta)^2) \;=\; (E_\theta(T_n) - \theta)^2 + Var_\theta(T_n) \;=\; \frac{1}{n^2} + \frac{1}{n^2} \;=\; \frac{2}{n^2}.$$

d) *Wegen*

$$\lim_{n\to\infty} E_\theta(T_n) = \lim_{n\to\infty}\left(\frac{1}{n} + \theta\right) \;=\; \theta$$

ist die Schätzerfolge $T_1, T_2, \ldots$ asymptotisch erwartungstreu.

H 25 Konfidenzintervalle

Im Andenhochland wurde eine neue Sorte Tomaten gefunden. Um deren Ertrag zu bestimmen, wurde bei 70 Pflanzen der Ernteertrag in *kg* gemessen. Es ergab sich das arithmetische Mittel $\bar{x} = 3.73$ und die Stichprobenvarianz $s^2 = 1.27$.

a) Unter geeigneten Annahmen bestimme man ein Konfidenzintervall zum Niveau 0.99 für den Erwartungswert m und eines für die Varianz σ^2.

Ein Teil der Pflanzen ist von einer Krankheit befallen, was die Tomaten ungenießbar macht. Bei den 70 untersuchten Tomatenstöcken waren 14 von dieser Krankheit befallen.

b) Unter geeigneten Annahmen bestimme man ein (approximatives) Konfidenzintervall zum Niveau 0.95 für die Wahrscheinlichkeit p, dass eine Pflanze von dieser Krankheit befallen ist.

a) *Für die Ernteerträge $X_1, X_2, \dots, X_{70}$ der einzelnen Tomatenpflanzen kann man annehmen, dass diese unabhängig voneinander und identisch $N(m, \sigma^2)$-verteilt sind.*
Als konkretes Konfidenzintervall für m bei unbekannter Varianz σ^2 zum Niveau $1-\alpha = 0.99$ erhält man dann

$$\begin{aligned} & \left[\bar{x} - t_{n-1,1-\frac{\alpha}{2}} \frac{s}{\sqrt{n}}, \bar{x} + t_{n-1,1-\frac{\alpha}{2}} \frac{s}{\sqrt{n}}\right] \\ = & \left[3.73 - \underbrace{t_{69,0.995}}_{=2.6490} \frac{1.27}{\sqrt{70}}, 3.73 + t_{69,0.995} \frac{1.27}{\sqrt{70}}\right] \\ \approx & \ [3.73 - 0.36, 3.73 + 0.36] \ = \ [3.37, 4.09] \end{aligned}$$

Mit einer Wahrscheinlichkeit von 0.99 liegt also der unbekannte Erwartungswert m zwischen 3.37 und 4.09 kg. Für das Konfidenzintervall für σ^2 zum Niveau $1-\alpha = 0.99$ erhält man

$$\begin{aligned} \left[\frac{(n-1)s^2}{\chi^2_{n-1,1-\frac{\alpha}{2}}}, \frac{(n-1)s^2}{\chi^2_{n-1,\frac{\alpha}{2}}}\right] &= \left[\frac{69 \cdot 1.27}{\chi^2_{69,0.995}}, \frac{69 \cdot 1.27}{\chi^2_{69,0.005}}\right] \\ &= \left[\frac{69 \cdot 1.27}{102.996}, \frac{69 \cdot 1.27}{42.494}\right] \approx [0.851, 2.062] \end{aligned}$$

Mit einer Wahrscheinlichkeit von 0.99 liegt die unbekannte Varianz σ^2 zwischen 0.851 und 2.062 (kg^2).

b) *Wir gehen davon aus, dass jede Pflanze unabhängig voneinander von der Krankheit befallen ist oder nicht und zwar mit einer (unbekannten) Wahrscheinlichkeit p. Die Gesamtzahl der kranken Pflanzen ist dann binomialverteilt, $Bin(70, p)$. Das approximative Konfidenzintervall für p zum Niveau $1 - \alpha = 0.95$ hat dann mit $\bar{x} = 14/70 = 0.2$ die Form*

$$\begin{aligned}
&\left[\bar{x} - z_{1-\frac{\alpha}{2}}\sqrt{\frac{\bar{x}(1-\bar{x})}{n}}, \bar{x} + z_{1-\frac{\alpha}{2}}\sqrt{\frac{\bar{x}(1-\bar{x})}{n}}\right] \\
&= \left[0.2 - \underbrace{z_{0.975}}_{=1.960}\sqrt{\frac{0.2(1-0.2)}{70}}, 0.2 + z_{0.975}\sqrt{\frac{0.2(1-0.2)}{70}}\right] \\
&\approx [0.2 - 0.094, 0.2 + 0.094] = [0.106, 0.294]
\end{aligned}$$

Mit einer Wahrscheinlichkeit von ca. 0.95 liegt die Erkrankungrate p zwischen 0.106 und 0.294.

H 26 a) Ein Hersteller von Präzisionswaagen gibt an, dass die Messgenauigkeit einer neu entwickelten, besonders preisgünstigen Waage nicht schlechter sei als die des bisher gelieferten und bewährten Modells. Zur Überprüfung dieser Angabe führt ein potentieller Käufer mit einer dieser neuen Waagen 30 Messungen eines Gewichtes durch. Aus der entstandenen Messreihe $x_1, \ldots, x_{30}$ wurde die empirische Varianz $s^2 = 2.07 \cdot 10^{-11}$ berechnet. Aufgrund langjähriger Praxis ist bekannt, dass sich die Wiegeergebnisse bei der bisher verwendeten Waage durch eine Zufallsvariable mit der Varianz $\sigma_0^2 = 1.5 \cdot 10^{-11}$ beschreiben lassen. Versuchen Sie, die Angabe des Herstellers mit einem geeigneten Test zum Niveau 0.05 zu widerlegen. Nehmen Sie dabei an, dass $x_1, \ldots, x_{30}$ Realisierungen unabhängiger identisch normalverteilter Zufallsvariablen sind. Als Maß für die Messgenauigkeit ist die Varianz zu nehmen.

b) Wie würde die Überprüfung ausfallen, wenn sich die empirische Varianz $s^2 = 2.07 \cdot 10^{-11}$ auf der Grundlage von 100 Messwerten $x_1, \ldots, x_{100}$ ergeben hätte?

Eventuell benötigte Quantile : $\chi^2_{29;0.95} = 42.557 \quad \chi^2_{99;0.95} = 123.225$

a) *verwendetes Testverfahren: χ^2-Streuungstest*

$H_0 : \sigma^2 \leq \sigma_0^2 = 1.5 \cdot 10^{-11}$ *vs.* $H_1 : \sigma^2 > \sigma_0^2$

Wert der Testgröße: $T = \dfrac{2.9}{1.5 \cdot 10^{-11}} \cdot 2.07 \cdot 10^{-11} = 40.02$

Da $T = 40.02 < 42.557 = \chi^2_{29;\,0.95}$ ist, sprechen die Daten nicht gegen die Hypothese, dass die Messgenauigkeit der neuen Waage nicht schlechter ist.

b) *Wert der Testgröße:* $T = \dfrac{99}{1.5 \cdot 10^{-11}} \cdot 2.07 \cdot 10^{-11} = 136.62$

Da $\chi^2_{99;\,0.95} = 123.225 < 136.62 = T$ *ist, ist in diesem Fall die Hypothese zu verwerfen.*

FAZIT: Es hängt entscheidend vom Stichprobenumfang ab, ob der Schätzwert S^2 *eine signifikante Abweichung der Nullhypothese signalisiert.*

H 27 Gegeben sei die folgende Messreihe $x_1, \ldots, x_{16}$:

0.534	0.545	0.283	0.445	0.519	0.513	0.464	0.499
0.605	0.526	0.568	0.427	0.546	0.527	0.566	0.597

mit $\sum_{i=1}^{16} x_i = 8.164$ und $\sum_{i=1}^{16} x_i^2 = 4.257386$.
Dabei wird angenommen, dass $x_1, \ldots, x_{16}$ Realisierungen von 16 unabhängigen, identisch $N(\mu, 0.01)$-verteilten Zufallsvariablen sind.

a) Überprüfen Sie die Hypothese $H_0 : \mu_0 = 0.5$ gegen die Alternative $H_1 : \mu \neq 0.5$ mit einem geeigneten Test zum Signifikanzniveau $\alpha = 0.05$.

geeigneter Test: Gauß-Test mit $\alpha = 0.05$

- *Nullhypothese:* $H_0 :\ \mu = 0.5$ *vs.* $H_1 : \mu \neq 0.5$
- *Testgröße:*

$$T(X_1, \ldots, X_n) = \sqrt{n}\frac{\bar{X}_{(n)} - \mu_0}{\sigma_0}$$

- *Realisierung der Testgröße:*

$$T(x_1, \ldots, x_{16}) = \frac{\sqrt{16}}{0.1} \cdot \left(\frac{8.164}{16} - 0.5 \right) = 0.41$$

- *Kritischer Bereich:*

$$K = \{(x_1, \ldots, x_{16}) \ :\ |T(x_1, \ldots, x_{16})| > u_{1-\frac{\alpha}{2}}\}$$

- *Entscheidung: Wegen* $|0.41| \leq 1.96 = u_{0.975}$ *wird* H_0 *nicht verworfen.*

b) Betrachten Sie nun die Hypothese $H_0 : \mu \geq 0.5$ gegen die Alternative $H_1 : \mu < 0.5$. Für welche Werte von μ wird die Wahrscheinlichkeit für einen Fehler 2. Art kleiner als 10 % auf gleichem Signifikanzniveau $\alpha = 0.05$?

geeigneter Test: Gauß-Test mit $\alpha = 0.05$

- *Nullhypothese:* $H_0 :\ \mu \geq 0.5$ *vs.* $H_1 : \mu < 0.5$ *(einseitig!)*
- *Testgröße: wie in Aufgabenteil a)*

– *Kritischer Bereich:*

$$K = \{(x_1, \ldots, x_{16}) \mid T(x_1, \ldots, x_{16}) < u_{0.05}\}$$

Fehler 2. Art: H_0 *wird fälschlicherweise nicht verworfen*
gesucht: $\mu_1 < 0.5$ *(unter* H_1*) mit* $P_{\mu=\mu_1}(T(X_1, \ldots, X_{16}) \notin K) \ \leq \ 0.1$
Es gilt: $\bar{X}_{(16)} \sim N\left(\mu_1, \frac{0.1^2}{16}\right)$ *unter der Annahme* $\mu = \mu_1$
Damit folgt:

$$\begin{aligned} P_{\mu=\mu_1}(T(X_1, \ldots, X_{16}) \notin K) &= P_{\mu=\mu_1}(\frac{\sqrt{16}}{0.1} \cdot \left(\bar{X}_{(16)} - 0.5\right) \geq -1.64) \\ &= P_{\mu=\mu_1}(40\left(\bar{X}_{(16)} - \mu_1\right) \geq -1.64 + 40 \cdot 0.5 - 40\mu_1) \\ &= P_{\mu=\mu_1}(\underbrace{40\left(\bar{X}_{(16)} - \mu_1\right)}_{\sim N(0,1)} \geq 18.36 - 40\mu_1) \\ &= 1 - \Phi(18.36 - 40\mu_1) \end{aligned}$$

gesucht: $\mu_1 < 0.5$ *mit*

$$\begin{aligned} & 1 - \Phi(18.36 - 40\mu_1) \ \leq 0.1 \\ \Leftrightarrow \quad & \Phi(18.36 - 40\mu_1) \ \geq \ 0.9 \\ \Leftrightarrow \quad & 18.36 - 40\mu_1 \ \geq \ u_{0.9} = 1.282 \\ \Leftrightarrow \quad & \mu_1 \leq 0.42695 \end{aligned}$$