

–Klausur vom 22.7.23 – Lösungen –

1. Bitte füllen Sie die folgenden Felder aus

Name: (bitte Druckbuchstaben
Vorname: verwenden)
Matrikelnummer:
Fach:

2. Bitte nehmen Sie die folgenden Regeln für den Wissensteil zur Kenntnis

- Jede richtige Aufgabe ergibt 2 Punkte
- Wird die Aufgabe unvollständig oder in Teilen fehlerhaft gelöst, so erhalten Sie für die korrekt gelösten Teile anteilig Punkte
- Nicht gelöste Aufgaben geben weder einen Punkt noch einen Punktabzug

3. Ihre Klausurnummer

Merken Sie sich Ihre Nummer

Aus Datenschutzgründen dürfen wir die Klausurergebnisse nicht in Verbindung mit Ihrer Matrikelnummer im Internet veröffentlichen. Bitte notieren Sie sich daher Ihre **Klausurnummer**, mit deren Hilfe Sie Ihr Klausurergebnis auf ISIS erfahren können.

Das Vorlesungsteam wünscht Ihnen eine erfolgreiche Klausur!

Platz für Notizen und Nebenrechnungen

A. Wahrscheinlichkeitsrechnung (20 Punkte)

A.a. Wissensteil $5 \times 2 = 10$ Punkte

1. Wie groß ist die Wahrscheinlichkeit, dass bei einem dreifachen Wurf mit einem fairen Würfel die Sequenz

$$1 - 1 - 5$$

gewürfelt wird, wobei hier die Reihenfolge der Ergebnisse zu beachten ist. Begründen Sie kurz!

Die Wahrscheinlichkeit beträgt $1/216 = 0.00463$.

Begründung: $\Omega = \{1, 2, \dots, 6\}^3$ und $P(\omega) = 1/216$ für alle $\omega \in \Omega$ oder

$$P(1 - 1 - 5) = \frac{1}{6} \cdot \frac{1}{6} \cdot \frac{1}{6} \text{ (Würfel fair und Würfe unabhängig).}$$

Punkteschema: 1 Punkt für die richtige Lösung und ein Punkt für die Begründung. Ist vorhanden aber unvollständig, dann nur einen halben Punkt für die Begründung.

2. Wie groß ist die Wahrscheinlichkeit, beim Wurf mit zwei fairen Würfeln (Würfel Eins und Würfel Zwei sind unterscheidbar), dass die Augenzahl im von Würfel Nummer Eins genau um eins höher ist, als die von Würfel Nummer Zwei? Begründen Sie kurz!

Die Wahrscheinlichkeit beträgt $\frac{5}{6}$, dass mit Würfel Eins eine der Ziffern 2,3,4,5,6 gewürfelt wird. Haben wir mit Würfel Eins eine eins gewürfelt, so kann Würfel zwei kein kleine um ein kleinere Augenzahl aufweisen.

Falls Würfel Eins eine der Ziffern $w \in \{2, \dots, 6\}$ zeigt, dann muss der zweite Würfel die Augenzahl $w - 1$ annehmen, damit das gefragte Ereignis eintritt. Hierfür ist die Wahrscheinlichkeit $\frac{1}{6}$, da Würfel Zwei 6 Möglichkeiten hat.

Da die beiden Würfel unabhängig sind, erhalten wir für die W.-keit des gefragten Ereignisses

$$p = \frac{5}{6} \cdot \frac{1}{6} = \frac{5}{36} \approx 13.9\%.$$

Punkteschema: 1 Punkt für die richtige Lösung und ein Punkt für die Begründung. Ist vorhanden aber unvollständig, dann nur einen halben Punkt für die Begründung.

3. Für zwei Ereignisse A und B werden die folgenden Angaben zur Wahrscheinlichkeit gemacht:

$$P(A) = 0.3 \qquad P(B) = 0.2 \qquad P(A \cup B) = 0.7$$

Ist das möglich? Falls nein, finden Sie den Fehler und geben Sie eine mathematische Begründung. Falls ja, geben Sie ein Beispiel.

Es muss $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ gelten. Hier

$$0.7 = 0.3 + 0.2 - P(A \cap B) \quad \Leftrightarrow \quad P(A \cap B) = -0.2.$$

Widerspruch zu $0 \leq P(\cdot) \leq 1$.

Punkteschema: Wenn ein valider Widerspruch zu den Eigenschaften eines Wahrscheinlichkeitsmaßes abgeleitet wird, gibt es 2 Punkte. Sonst gibt es 0 Punkte. Achtung: Es gibt viele Möglichkeiten, den Widerspruch zu zeigen. Die Musterlösung zeigt eine Möglichkeit.

4. In einer Wissensshow im Fernsehen werden der/dem Kandidat*in völlig willkürlich zu 50% schwere und zu 50% leichte Aufgaben gestellt. Die leichten Aufgaben beantwortet sie/er zu 90% richtig, die schweren nur zu 30%. Wie groß ist die Wahrscheinlichkeit, dass die/der Kandidat*in die nächste Frage richtig beantwortet?

Bezeichne R das Ereignis, dass die Frage richtig beantwortet wird, und S das Ereignis, dass die nächste Frage schwer wird. Dann gilt

$$P(R) = P(R|S)P(S) + P(R|S^c)P(S^c) = 0.3 \cdot 0.5 + 0.9 \cdot 0.5 = 0.6.$$

Punkteschema: Ein Punkt für den Ansatz über den Satz für die totale Wahrscheinlichkeit, ein halber Punkt für das richtige Einsetzen der Angaben aus der Aufgabe und ein halber Punkt für das richtige Ergebnis.

5. Als Anti-Dopingexpert*in im Radsport gehen Sie davon aus, dass 80% der Spitzenfahrer gedopt sind. Da die Dopingverfahren jedoch sehr fortgeschritten sind, wird ein gedopter Spitzenfahrer nur in 5% der Fälle durch den Test überführt. Bei einem unschuldigen Fahrer zeigt der Test in 1% der Fälle ein falsch positives Resultat. Beim Test des führenden Fahrers bei der Tour de France erhalten Sie ein negatives Testresultat. Wie groß ist dann ihrer Meinung nach die Wahrscheinlichkeit, dass der Fahrer tatsächlich sauber ist?

Sei G das Ereignis "gedopt" und T das Ereignis "Test positiv". Dann gilt $P(G^c) = 0.2$, $P(T|G) = 0.05$ und $P(T|G^c) = 0.01$. Es folgt $P(T^c|G) = 0.95$ und $P(T^c|G^c) = 0.99$. Nach dem Satz von Bayes gilt

$$\begin{aligned} P(G^c|T^c) &= \frac{P(T^c|G^c)P(G^c)}{P(T^c)} = \frac{P(T^c|G^c)P(G^c)}{P(T^c|G)P(G) + P(T^c|G^c)P(G^c)} \\ &= \frac{0.99 \cdot 0.2}{0.95 \cdot 0.8 + 0.99 \cdot 0.2} \\ &= 0.20668. \end{aligned}$$

Punkteschema: Ein Punkt für den Ansatz über den Satz von Bayes, ein halber Punkt für das richtige Einsetzen der Angaben aus der Aufgabe und ein halber Punkt für das richtige Ergebnis.

A.b.: Textaufgabe $4 \times 2.5 = 10$ Punkte

Der Bau einer Büroimmobilie in Berlin ist geplant. Die Kalkulation geht von monatlichen Mieten von 30 Euro pro Quadratmeter aus, was dem gegenwärtigen mittleren Mietpreis für Büroräume in Berlin entspricht. Werden dagegen nur 25 Euro pro Quadratmeter erzielt, ist das Baukonsortium pleite. Ebenfalls können Verzögerungen beim Bau nicht ausgeschlossen werden. Bei zwei Jahren Verzögerung bricht die Finanzierung zusammen, auch wenn die Miete bei 30 Euro/m² bleibt. Bei einer Verzögerung um ein Jahr können Mietpreise nur bis 28.5 Euro weggesteckt werden. Im Risikomanagement des Bauträgers werden folgende Szenarien betrachtet:

- 1) Bauverzögerung um ein Jahr wegen Mangel an kompetenten Firmen: Wahrscheinlichkeit 30%.
- 2) Bauverzögerung um ein Jahr, wenn sich Pfusch am Bau ereignet: Wahrscheinlichkeit unter der Bedingung, dass Risiko 1) eintritt, 30% und unter der Bedingung, dass Risiko 1) nicht eintritt, 10%.
- 3) Die weitere Verbreitung von Homeoffice schlägt auf die Büromieten durch und senkt das Mietniveau um 3 Euro mit Wahrscheinlichkeit 50% bzw. um 6 Euro mit Wahrscheinlichkeit 20%. Zu 30% wird kein Effekt von der Verbreitung von Homeoffice auf die Büroimmobilien erwartet. Diese Wahrscheinlichkeiten sind unabhängig von den anderen Ereignissen.

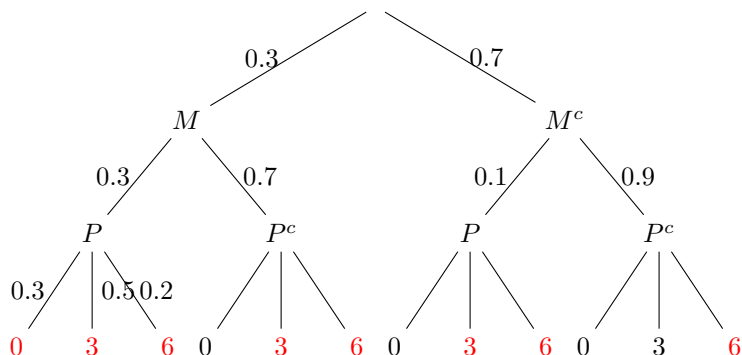
Bitte lösen Sie die folgenden Aufgaben:

- a) Zeichnen Sie den Wahrscheinlichkeitsbaum mit den (bedingten) Wahrscheinlichkeiten für die 'Zweige', der den Risikoszenarien des Baukonsortiums entspricht.
- b) Mit welcher Wahrscheinlichkeit geht das Konsortium pleite?
- c) Wie groß ist die Wahrscheinlichkeit, dass eine Bauverzögerung um zwei Jahre eintritt?
- d) Falls das Konsortium pleite geht - mit welcher Wahrscheinlichkeit (aus heutiger Sicht) hat es auch an der zweimaligen Bauverzögerung gelegen?

Schreiben Sie Ihre Lösung auf die Seiten 5 und 6.

Teil A–Aufgabe	1	2	3	4	5	Textaufgabe	Σ
Punkte							

- a) Bezeichne M den “Mangel an kompetenten Firmen” und P den “Pfusch am Bau”. Die Pleiteszenarien sind rot markiert.



Punkteschema: Ein halber Punkt für den Ansatz Wahrscheinlichkeitsbaum, ein Punkt für die Korrektheit des Wahrscheinlichkeitsbaums, ein Punkt für die Korrekte angebe aller Astwahrscheinlichkeiten (bei nur einem Flüchtigkeitsfehler 0.5 Pkt Abzug).

- b) Es gilt

$$\begin{aligned}
 P(\text{“Pleite”}) &= 0.3 \cdot 0.3 \cdot (0.3 + 0.5 + 0.2) + 0.3 \cdot 0.7 \cdot (0.5 + 0.2) \\
 &\quad + 0.7 \cdot 0.1 \cdot (0.5 + 0.2) + 0.7 \cdot 0.9 \cdot 0.2 \\
 &= 0.412.
 \end{aligned}$$

Punkteschema: Ein Punkt für die korrekte Identifikation aller Pleiteszenarien, ein Punkt für die Korrektheit der Einzelwahrscheinlichkeiten, ein halber Punkt für die korrekte Gesamtlösung).

- c) Es gilt

$$P(\text{“Bauverzögerung um zwei Jahre”}) = 0.3 \cdot 0.3 = 0.09.$$

Punkteschema: Ein Punkt für die korrekte Identifikation des Szenarios, ein Punkt für die Korrektheit der Einzelwahrscheinlichkeiten, ein halber Punkt für die korrekte Gesamtlösung).

- d) Es gilt

$$\begin{aligned}
 &P(\text{“Bauverzögerung um zwei Jahre”} \mid \text{“Pleite”}) \\
 &= \frac{P(\text{“Bauverzögerung um zwei Jahre”} \cap \text{“Pleite”})}{P(\text{“Pleite”})} \\
 &= \frac{P(\text{“Bauverzögerung um zwei Jahre”})}{P(\text{“Pleite”})} = \frac{0.09}{0.412} = 0.2184.
 \end{aligned}$$

Punkteschema: Ein Punkt für den Ansatz nach mit Bed. W.-keit/Satz v. Bayes, ein Punkt für die Korrektheit der Einzelwahrscheinlichkeiten, ein halber Punkt für die korrekte Gesamtlösung). Folgefehler aus aufgabe d) führen nicht zum Punktabzug.

B. Zufallsvariablen, Verteilungen und Grenzwertsätze (20 Punkte)

B.a. Wissensteil $5 \times 2 = 10$ Punkte

1. Nennen Sie zwei kontinuierliche Verteilungen und geben Sie die Wahrscheinlichkeitsdichten an!

- Exponentialverteilung mit Parameter $\lambda > 0$ hat Dichte

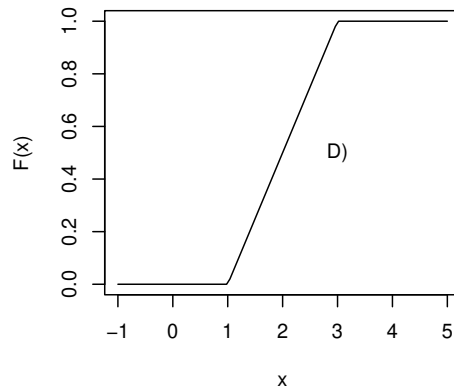
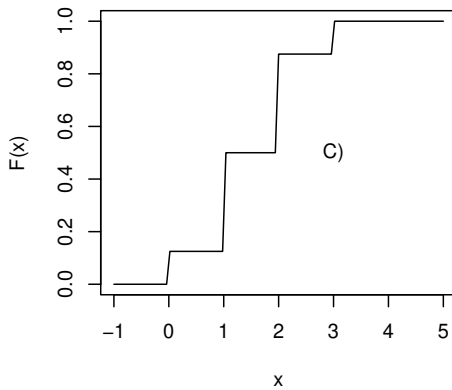
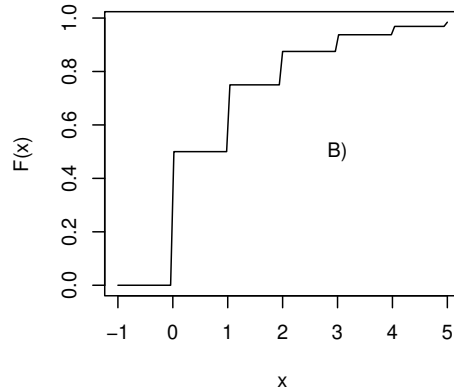
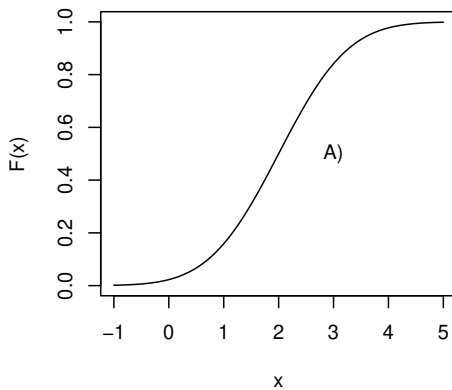
$$f(x) = \begin{cases} 0, & x < 0, \\ \lambda e^{-\lambda x}, & x \geq 0. \end{cases}$$

- Normalverteilung mit Parametern $\mu \in \mathbb{R}$, $\sigma^2 > 0$ hat Dichte

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}.$$

Punkteschema: Je ein halber Punkt für eine korrekte Verteilung und eine korrekte Dichte. Passt die Dichte zur Verteilung, obwohl die Verteilung nicht kontinuierlich ist, geben wir einen halben Punkt für die Dichte.

2. Die untenstehende Grafik zeigt vier Verteilungsfunktionen.



Ordnen Sie diese Verteilungsfunktionen den folgenden vier Verteilungen zu, indem sie den entsprechenden Buchstaben hinter die richtige Verteilung schreiben.

- Binomialverteilung $B(p = 0.5, n = 3)$ C
 - Geometrische Verteilung $G(p = 0.5)$ B
 - Gleichverteilung, $U(a = 1, b = 3)$ D
 - Normalverteilung $N(\mu = 2, \sigma^2 = 1)$ A
- Punkteschema:** 0.5 pro richtige Zuordnung

3. Eine Zufallsvariable X hat die folgende, kontinuierliche Verteilung ("Dreiecksverteilung")

$$f_X(x) = \begin{cases} 2x & x \in [0, 1] \\ 0 & \text{sonst} \end{cases}$$

Berechnen Sie den Erwartungswert $\mathbb{E}[X]$ und die Varianz $\mathbb{V}[X]$ von X !

Hinweis: Verwenden Sie, dass $\mathbb{V}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2$!

Es gilt

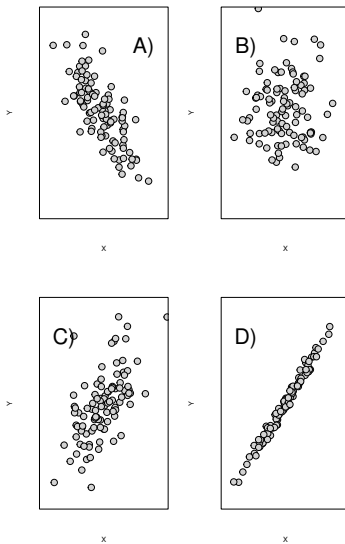
$$\mathbb{E}[X] = \int_0^1 x f_X(x) dx = \int_0^1 2x^2 dx = \left[\frac{2x^3}{3} \right]_0^1 = \frac{2}{3},$$

$$\mathbb{E}[X^2] = \int_0^1 x^2 f_X(x) dx = \int_0^1 2x^3 dx = \left[\frac{x^4}{2} \right]_0^1 = \frac{1}{2},$$

$$\mathbb{V}[X] = \mathbb{E}[X^2] - \mathbb{E}[X]^2 = \frac{1}{2} - \left(\frac{2}{3} \right)^2 = \frac{1}{18}.$$

Punkteschema: Ein halber Punkt für mindestens einen richtigen Ansatz für $\mathbb{E}[X]$ oder $\mathbb{E}[X^2]$ und jeweils ein halber Punkt für das richtige Ergebnis der Integration. Ein halber Punkt für das korrekte Ergebnis für die Varianz.

4. Welche der Streuplots in den folgenden Diagrammen stammen von Zufallsvariablen mit eindeutig negativer Korrelation? Woran erkennt man das und in welchem Fall kann man nicht gut entscheiden, ob die Korrelation positiv oder negativ ist?



Streuplot A) stammt von Zufallsvariablen mit eindeutig negativer Korrelation, da hier ein negativer linearer Zusammenhang zwischen den betrachteten Merkmalen besteht.

Für Streuplot B) ist schwer zu entscheiden, ob die Korrelation positiv oder negativ ist.

Punkteschema: Je

5. Sind die Zufallszahlen mit der folgenden gemeinsamen Verteilung unabhängig? Begründen Sie durch eine Rechnung!

X/Y	$Y = 0$	$Y = 1$	$Y = 2$	$Y = 3$
$X = 0$	$P(X = 0 \wedge Y = 0) = 0.05$	$P(X = 0 \wedge Y = 1) = 0.15$	$P(X = 0 \wedge Y = 2) = 0.25$	$P(X = 0 \wedge Y = 3) = 0.05$
$X = 1$	$P(X = 1 \wedge Y = 0) = 0.1$	$P(X = 1 \wedge Y = 1) = 0.1$	$P(X = 1 \wedge Y = 2) = 0.2$	$P(X = 1 \wedge Y = 3) = 0.1$

\wedge steht hier für die logische **and**-Verknüpfung, d.h. $P(X = 1 \wedge Y = 2)$ gibt die Wahrscheinlichkeit an, dass gleichzeitig $X = 1$ und $Y = 2$ beobachtet wird.

Es gilt

$$P(X = 0) = 0.5,$$

$$P(X = 1) = 0.5,$$

$$P(Y = 0) = 0.15,$$

$$P(Y = 1) = 0.25,$$

$$P(Y = 2) = 0.45,$$

$$P(Y = 3) = 0.15.$$

Die Zufallsvariablen sind nicht unabhängig. Beispielsweise ist

$$P(X = 0 \wedge Y = 0) = 0.05 \neq 0.075 = 0.5 \cdot 0.15 = P(X = 0) \cdot P(Y = 0).$$

Punkteschema: Einen halben Punkt für die richtige Antwort, einen Punkt für den richtigen Ansatz zur Begründung und einen halben Punkt für die korrekte Rechnung in der Begründung.

B.b.: Textaufgabe $5 \times 2 = 10$ Punkte

In einem alten Auto zeigt die Tankuhr $\mu = 5$ Liter an. Die tatsächliche Tankfüllung schwankt jedoch mit einer Standardabweichung von $\sigma = 0.25$ Liter um den gemessenen Wert. Die tatsächliche Tankfüllung X sei als normalverteilt vorausgesetzt $X \sim N(\mu, \sigma^2)$. Das Auto verbraucht genau 10 Liter/100 km.

- a) Welche Verteilung hat die zufällige Reichweite des Autos?
- b) Skizzieren Sie die kontinuierliche Dichte der **zufälligen Reichweite des Autos** mit korrekter Achsenbeschriftung und zeichnen Sie den Erwartungswert ein.
- c) Geben Sie einen symmetrisch um die erwartete Reichweite gelegenen Bereich an, in dem das Auto zu ca. 95.45% aus Spritmangel zum Stehen kommt.
- d) Geben Sie eine Entfernung an, sodass das Auto diese Entfernung zu 95% ohne nachtanken schafft.
- e) Die nächste Autobahntankstelle ist 48 km entfernt – mit welcher Wahrscheinlichkeit erreicht das Auto diese noch, ohne vorher zu tanken?

Hinweis: Die Werte der Verteilungsfunktion der Standardnormalverteilung sowie deren Quantile entnehmen Sie bitte den Tabellen am Ende der Klausur. Beachten Sie auch die Relation $z_\alpha = -z_{1-\alpha}$ für die Quantile der Standardnormalverteilung.

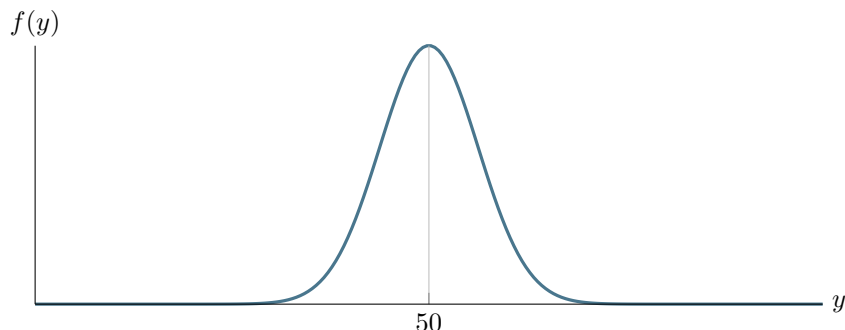
Schreiben Sie Ihre Lösung auf die Seiten 9–11.

Teil B–Aufgabe	1	2	3	4	5	Textaufgabe	Σ
Punkte							

- a) Sei Y die zufällige Reichweite des Autos. Dann

$$Y = 10X \sim N(10 \cdot 5, (10 \cdot 0.25)^2) = N(50, 2.5^2).$$

- b)



Punkteschema: Einen Punkt für eine gelungene Skizze der Glockenkurve, (ein halber Punkt Abzug, wenn die Form überhaupt nicht getroffen wurde). Ein halber Punkt für die Achsen-Beschriftung und den eingezeichneten Erwartungswert. Folgefehler aus Teil a) werden nicht geahndet.

- c) Gesucht ist die 2σ -Umgebung um μ , d.h.

$$[\mu - 2\sigma, \mu + 2\sigma] = [50 - 5, 50 + 5] = [45, 55].$$

Punkteschema: Ein Punkt für den Ansatz, ein halber Punkt für's richtige Einsetzen und ein halber Punkt für das richtige Ergebnis. Die Alternativlösung mit Standardisieren und der Verteilungsfunktion ist statthaft.

- d) Nach Tabelle ist das 5%-Quantil der Standardnormalverteilung $z_{0.05} = -z_{0.95} = -1.64$. Damit schafft das Auto eine Entfernung von

$$\mu + z_{0.05} \cdot \sigma = 50 - 1.64 \cdot 2.5 = 45.9 \quad [\text{in km}]$$

zu 95% ohne nachzutanken. **Punkteschema:** Ein halber Punkt für die Korrekte bestimmung von $z_{0.05}$, ein halber Punkt für den richtigen Ansatz, ein halber Punkt für's richtige Einsetzen und ein halber Punkt für das korrekte Endergebnis.

- e) Es gilt

$$\begin{aligned} P(Y \geq 48) &= 1 - P(Y < 48) = 1 - P(Y \leq 48) = 1 - P\left(\frac{Y - 50}{2.5} \leq \frac{48 - 50}{2.5}\right) \\ &= 1 - \Phi(-0.8) = 1 - (1 - \Phi(0.8)) = 1 - (1 - 0.7881) = 0.7881. \end{aligned}$$

D.h. die Tankstelle wird mit einer Wahrscheinlichkeit von 78.81% erreicht.

Punkteschema: Ein halber Punkt für den richtigen Ansatz als Schwellen-W.-keit, ein halber Punkt für das Standardisieren als Ansatz, ein halber Punkt für das korrekte berechnen des z -Wertes der Standardisierung, ein halber Punkt für die korrekte Verwendung der Verteilungsfunktion und des Ergebnisses.

Lösung zu Aufgabe B.b.1

C. Statistik (nur für 9LP-Modul Stochastik für Informatik) (20 Punkte)

C.a. Wissensteil $5 \times 2 = 10$ Punkte

1. Geben Sie für Parameterschätzer die Definition der Erwartungstreue und der Konsistenz wieder!

Sei $(x_1, \dots, x_n) \in \mathbb{R}^n$ ein Vektor von Daten, welche als Realisierungen von identisch verteilten Zufallsvariablen X_1, \dots, X_n mit Parameter θ auf einem Wahrscheinlichkeitsraum (Ω, P_θ) gelten. Sei θ_n eine Schätzfunktion zur Stichprobengröße n . Dann ist θ_n ein erwartungstreuer Schätzer für θ , falls

$$\mathbb{E}_\theta[\theta_n(X_1, \dots, X_n)] = \theta.$$

Weiter ist θ_n ein konsistenter Schätzer für θ , falls

$$\lim_{n \rightarrow \infty} \theta_n(X_1, \dots, X_n) = \theta.$$

Punkteschema: Je ein Punkt pro korrekter Antwort.

2. Stellen Sie die Maximum-Likelihood-Gleichungen für den Parameter p der Geometrischen Verteilung auf und lösen Sie diese! Setzen Sie hierbei die unabhängigen Beobachtungen $x_1, \dots, x_n \in \mathbb{N}_0$ voraus. Verwenden Sie möglichst die Konvention für die Dichte $p(X = x) = p(1-p)^x$, $x \in \mathbb{N}_0$. Wir erkennen aber auch die alternative Definition aus dem Buch von N. Klein an.

Die Likelihood-Funktion ist

$$L(x_1, \dots, x_n; p) = \prod_{i=1}^n p(1-p)^{x_i} = p^n (1-p)^{\sum_{i=1}^n x_i}.$$

Die Log-Likelihood-Funktion ist

$$l(x_1, \dots, x_n; p) = n \cdot \ln(p) + \left(\sum_{i=1}^n x_i \right) \ln(1-p).$$

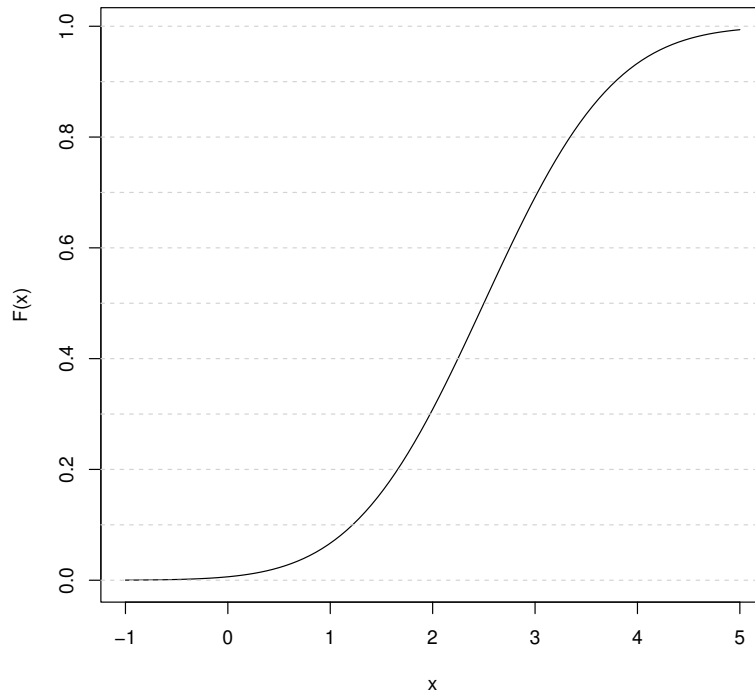
Damit

$$\begin{aligned} \frac{d}{dp} l(x_1, \dots, x_n; p) = \frac{n}{p} - \frac{\sum_{i=1}^n x_i}{1-p} = 0 &\Leftrightarrow \frac{n}{p} = \frac{\sum_{i=1}^n x_i}{1-p} \\ &\Leftrightarrow \frac{n}{p} - n = \sum_{i=1}^n x_i \\ &\Leftrightarrow p = \frac{n}{\sum_{i=1}^n x_i + n} = \frac{1}{\bar{X}_n + 1}. \end{aligned}$$

Für die alternative Definition aus dem Buch von N. Klein gilt $p^* = \frac{n}{\sum_{i=1}^n X_i} = \frac{1}{\bar{X}_n}$.

Punkteschema: Ein halber Punkt für die Aufstellung der Likelihood, ein halber Punkt für die Korrekte logarithmierung, ein halber Punkt für die korrekte ableitung nach p und ein halber Punkt für die korrekte Auflösung.

3. Gegeben ist die folgende Verteilungsfunktion (siehe Grafik). Bestimmen Sie durch Ablesen die Quantile zu den Wahrscheinlichkeiten $p = 0.1$ und $p = 0.9$!



Ungefähr $x_{0,1} = 1.25$ und $x_{0,9} = 3.75$.

Punkteschema: 1 Punkt pro richtige Antwort.

4. In einer Stichprobe aus unabhängigen, normalverteilten und identisch verteilten Daten vom Umfang $n = 5$ beträgt das arithmetische Mittel $\bar{x} = 7.4$ und die empirische Standardabweichung $\hat{\sigma} = 0.32$. Berechnen Sie ein zweiseitiges Konfidenzintervall für den Erwartungswert zum Konfidenzniveau $1 - \alpha = 0.9$.

Es gilt

$$\left[\bar{x} - \frac{\hat{\sigma}}{\sqrt{n}} t_{n-1, 1-\alpha/2}, \bar{x} + \frac{\hat{\sigma}}{\sqrt{n}} t_{n-1, 1-\alpha/2} \right] = \left[7.4 - \frac{0.32}{\sqrt{5}} 2.132, 7.4 + \frac{0.32}{\sqrt{5}} 2.132 \right]$$

$$= [7.09489, 7.70511]$$

mit $t_{4,0.95} = 2.132$.

Punkteschema: 1 Punkt für die richtige Intervallart (beidseitig, t-basiert), ein halber Punkt für die richtige Formel, ein halber Punkt für das richtige Einsetzen incl. dem richtigen t-Quantil und ein halber Punkt für das richtige Ergebnis.

5. Was ist der p -Wert in einem Statistischen Test? Erläutern Sie kurz, indem Sie auch auf den mathematischen Hintergrund eingehen!

Unter der Nullhypothese H_0 beschreibt der p -Wert die Wahrscheinlichkeit, dass eine Teststatistik Y Werte gleich oder extremer als y annimmt. D.h. der rechts- bzw. linksseitige p -Wert für $y \in Y(\Omega)$ ist

$$p(y) = P(Y \geq y | H_0 \text{ ist wahr}) \quad \text{bzw.} \quad P(Y \leq y | H_0 \text{ ist wahr}).$$

Punkteschema: 1 Punkt für die sachlich grob richtige Antwort, 1/2 Punkt für die Erwähnung der Null-Hypothese, 1/2 Punkt für die Erwähnung der Teststatistik.

C.b.: Textaufgabe 10 Punkte (4×2.5 Punkte)

1. Eine Studie untersucht den Zusammenhang zwischen Rauchen und dem Auftreten von Diabeteserkrankungen vom Typ 2 zwischen 60 und 70 Jahren bei einer Stichprobe von 100 Männern. Die folgende Tabelle fasst die Ergebnisse zusammen:

Diabetes Typ 2	Nicht-Raucher	Ex-Raucher	Raucher
Nein	26	20	15
Ja	10	9	20

- Stellen Sie die Kontingenztabelle der relativen Häufigkeiten auf.
- Entscheiden Sie, ob Sie zur Beantwortung der oben stehenden Frage einen χ^2 - Homogenitätstest oder einen χ^2 -Anpassungstest durchführen?
- Berechnen Sie den Wert der χ^2 -Teststatistik (im Skript auch K^2 genannt)!
- Führen Sie die Testentscheidung zur Nullhypothese H_0 : „Rauchen hängt nicht mit Typ 2 Diabetes im Alter zusammen“ gegen H_1 : „Es gibt einen Zusammenhang“ zum Signifikanzniveau $\alpha = 1\%$ durch!

Hinweis: Die Quantile der χ^2 -Verteilung finden Sie in der Tabelle im Anhang an diese Klausur!

a)

Diabetes Typ 2	Nicht-Raucher	Ex-Raucher	Raucher	Gesamt
Nein	13/50	1/5	3/20	61/100
Ja	1/10	9/100	1/5	39/100
Gesamt	9/25	29/100	7/20	1

Punkteschema: Ein Punkt halber für die richtige Struktur der Tabelle, ein halber Punkt für relative Häufigkeiten, ein Punkt für die richtigen inneren Felder (pro Fehler 0.5 pkt Abzug, max 1 Pkt Abzug), ein halber Punkt für die korrekten Marginalverteilungen.

b) Wie führen einen χ^2 -Homogenitätstest durch.

Punkteschema: 1,5 Punkte für den Chi-Quadrat Test, ein Punkt für Homogenitätstest.

c) Es gilt

$$\chi^2 = n \sum_i \sum_j \frac{(h_{i,j} - h_i^X h_j^Y)^2}{h_i^X h_j^Y} = 7.52187.$$

Punkteschema: Ein Punkt für die richtige Formel, 1 Punkt für das korrekte Einsetzen und 0,5 punkte für das korrekte Ergebnis. Wurde mit programmiertem TR gerechnet, dann 1,5 punkte für richtiges Ergebnis.

d) Mit $\chi_{1-\alpha, (r-1)(s-1)}^2 = \chi_{0.99, 2}^2 = 9.210$ gilt $\chi^2 \leq \chi_{1-\alpha, 2}^2$, d.h. H_0 wird nicht verworfen.

Punkteschema: Ein halber Punkt für die korrekten Freiheitsgrade $(r - 1)(s - 1) = 2$, ein halber Punkt für das korrekte $1 - \alpha$ -Quantil ein halber Punkt für korrektes Ablesen des Quantils, ein halber Punkt für das richtige Testkriterium und ein halber Punkt für die richtige Testentscheidung. Folgefehler (zB aus c)) werden nicht bestraft.

Lösung zu Aufgabe C.b.1

\tilde{x}_p -Quantile der Student- t_n -Verteilung							
n	p						
	0,90	0,95	0,975	0,99	0,995	0,999	0,9995
1	3,078	6,314	12,706	31,821	63,656	318,289	636,578
2	1,886	2,920	4,303	6,965	9,925	22,328	31,600
3	1,638	2,353	3,182	4,541	5,841	10,214	12,924
4	1,533	2,132	2,776	3,747	4,604	7,173	8,610
5	1,476	2,015	2,571	3,365	4,032	5,894	6,869
6	1,440	1,943	2,447	3,143	3,707	5,208	5,959
7	1,415	1,895	2,365	2,998	3,499	4,785	5,408
8	1,397	1,860	2,306	2,896	3,355	4,501	5,041
9	1,383	1,833	2,262	2,821	3,250	4,297	4,781
10	1,372	1,812	2,228	2,764	3,169	4,144	4,587
11	1,363	1,796	2,201	2,718	3,106	4,025	4,437
12	1,356	1,782	2,179	2,681	3,055	3,930	4,318
13	1,350	1,771	2,160	2,650	3,012	3,852	4,221
14	1,345	1,761	2,145	2,624	2,977	3,787	4,140
15	1,341	1,753	2,131	2,602	2,947	3,733	4,073
16	1,337	1,746	2,120	2,583	2,921	3,686	4,015
17	1,333	1,740	2,110	2,567	2,898	3,646	3,965
18	1,330	1,734	2,101	2,552	2,878	3,610	3,922
19	1,328	1,729	2,093	2,539	2,861	3,579	3,883
20	1,325	1,725	2,086	2,528	2,845	3,552	3,850
21	1,323	1,721	2,080	2,518	2,831	3,527	3,819
22	1,321	1,717	2,074	2,508	2,819	3,505	3,792
23	1,319	1,714	2,069	2,500	2,807	3,485	3,768
24	1,318	1,711	2,064	2,492	2,797	3,467	3,745
25	1,316	1,708	2,060	2,485	2,787	3,450	3,725
26	1,315	1,706	2,056	2,479	2,779	3,435	3,707
27	1,314	1,703	2,052	2,473	2,771	3,421	3,689
28	1,313	1,701	2,048	2,467	2,763	3,408	3,674
29	1,311	1,699	2,045	2,462	2,756	3,396	3,660
∞	1,282	1,645	1,960	2,327	2,577	3,092	3,293

Die Letzte Zeile " ∞ " enthält die Quantile der Standard-Normalverteilung und gilt in guter Näherung für die t_n -Verteilung mit $n \geq 30$

Beispiel: Das $p = 99\%$ -Quantil der t -Verteilung mit $n = 5$ Freiheitsgraden beträgt

$$t_{0,99}(5) = 3.365$$

\tilde{x}_p -Quantile der χ_n^2 -Verteilung									
n	p								
	0,005	0,01	0,025	0,05	0,9	0,95	0,975	0,99	0,995
1	0,000	0,000	0,001	0,004	2,706	3,841	5,024	6,635	7,879
2	0,010	0,020	0,051	0,103	4,605	5,991	7,378	9,210	10,597
3	0,072	0,115	0,216	0,352	6,251	7,815	9,348	11,345	12,838
4	0,207	0,297	0,484	0,711	7,779	9,488	11,143	13,277	14,860
5	0,412	0,554	0,831	1,145	9,236	11,070	12,832	15,086	16,750
6	0,676	0,872	1,237	1,635	10,645	12,592	14,449	16,812	18,548
7	0,989	1,239	1,690	2,167	12,017	14,067	16,013	18,475	20,278
8	1,344	1,647	2,180	2,733	13,362	15,507	17,535	20,090	21,955
9	1,735	2,088	2,700	3,325	14,684	16,919	19,023	21,666	23,589
10	2,156	2,558	3,247	3,940	15,987	18,307	20,483	23,209	25,188
11	2,603	3,054	3,816	4,575	17,275	19,675	21,920	24,725	26,757
12	3,074	3,571	4,404	5,226	18,549	21,026	23,337	26,217	28,300
13	3,565	4,107	5,009	5,892	19,812	22,362	24,736	27,688	29,819
14	4,075	4,660	5,629	6,571	21,064	23,685	26,119	29,141	31,319
15	4,601	5,229	6,262	7,261	22,307	24,996	27,488	30,578	32,801
16	5,142	5,812	6,908	7,962	23,542	26,296	28,845	32,000	34,267
17	5,697	6,408	7,564	8,672	24,769	27,587	30,191	33,409	35,718
18	6,265	7,015	8,231	9,390	25,989	28,869	31,526	34,805	37,156
19	6,844	7,633	8,907	10,117	27,204	30,144	32,852	36,191	38,582
20	7,434	8,260	9,591	10,851	28,412	31,410	34,170	37,566	39,997
21	8,034	8,897	10,283	11,591	29,615	32,671	35,479	38,932	41,401
22	8,643	9,542	10,982	12,338	30,813	33,924	36,781	40,289	42,796
23	9,260	10,196	11,689	13,091	32,007	35,172	38,076	41,638	44,181
24	9,886	10,856	12,401	13,848	33,196	36,415	39,364	42,980	45,558

Hinweis: n steht hier in der Tabelle für den Wert "(Zeilen-1)" \times "(Spalten-1)" und nicht für den Stichprobenumfang!

Teil	A	B	C	Σ	Note
Punkte					