Stochastik
(BSc D-MAVT / BSc D-MATH / BSc D-MATL)

Schreibe für die Aufgaben 2-4 stets alle Zwischenschritte und -rechnungen sowie Begründungen auf. Aufgabe 1 ist eine Multiple Choice Aufgabe (keine Begründungen notwendig).

Die für die Aufgaben benötigten Tabellen (Verwerfungsbereiche für den Wilcoxon-Test, Normalverteilung, Quantile der t-Verteilung) wurden mit der Prüfung ausgeteilt.

1. (10 Punkte) Bei den folgenden 10 Fragen ist jeweils genau eine Antwort richtig. Kreise die richtige Antwort ein.

Beachte: Jede richtige Antwort gibt einen Punkt. Es gibt keinen Abzug für falsche Antworten. Wir empfehlen daher alle Fragen zu beantworten.

a) Welche Aussage ist richtig?

1. Falls $P(A|B) = 0.6$ und $P(A \cap B) = 0.2$, dann muss $P(B) = 0.4$ sein.
2. Falls $P(A) = P(B) = 0.3$ mit A und B unabhängig, dann gilt $P(A|B) = 0.3$.
3. Falls $P(A \cup B) = 0.5$ und $P(A) = P(B) = 0.5$, dann ist $P(A|B) = 0.5$.

b) Welche Aussage ist richtig?

1. Sei $X \sim \mathcal{N}(\mu, \sigma^2)$ mit $\mu = \sigma^2 = 2$. Dann gilt $Var(3X + 3) = 21$
2. Sei $Var(X) = 2$ und $Var(Y) = 5$, und seien X und Y unabhängig. Dann gilt $Var(5X + 2Y) = 20$.
3. Sei $E[X] = 1$ und $E[Y] = -3$, und seien X und Y unkorreliert. Dann gilt $E[3X + XY] = 0$.

c) Sei $X_1 \sim Exp(\lambda_1)$ und $X_2 \sim Exp(\lambda_2)$ und seien X_1 und X_2 unabhängig. Welche Aussage ist richtig?

1. $X_1 + X_2 \sim Exp(\lambda_1 + \lambda_2)$.
2. $P(X_1 = 5) = \lambda_1 e^{-5\lambda_1}$.
3. $E[X_1 X_2] = \frac{1}{\lambda_1 \lambda_2}$.

d) Die stetige Zufallsvariable X hat eine Dichte, die symmetrisch um den Wert -1 ist. Welche Aussage ist richtig?

1. Der Median von $Y = 2X + 2$ ist gleich 0.
2. Der Erwartungswert von $Z = X^2$ ist gleich 1.
3. Für die Zufallsvariable $V = X + 2$ gilt $P(V \leq -3) = P(X \leq -1)$.

e) Seien $y_1, \ldots, y_n$ Realisierungen von $Y_1, \ldots, Y_n \sim Uni[0, \vartheta]$ (i.i.d.), wobei ϑ unbekannt ist. Welche Aussage ist richtig?

1. Der MLE für ϑ ist $\hat{\vartheta} = \max\{y_1, \ldots, y_n\}$.
2. Der Schätzer $\hat{\vartheta} = \bar{y}/2$ ist ein Momentenschätzer für ϑ.
3. Der Schätzer $\hat{\vartheta} = \bar{y}/2$ ist erwartungstreu.

f) Seien $X_1, ..., X_n$ i.i.d. $\mathcal{N}(\mu, \sigma^2)$-verteilten Zufallsvariablen, mit μ unbekannt und σ^2 bekannt. Mittels Simulation generieren wir Realisierungen $x_1, \ldots, x_n$ von $X_1, \ldots, X_n$. Basierend auf den Punktschätzer $\hat{\mu} = \bar{x}$ konstruieren wir ein 95%-Vertrauensintervall für μ. Wir wiederholen das Ganze 100 Mal, so dass wir am Ende 100 Vertrauensintervalle haben. Welche Aussage ist richtig?

1. 95 der konstruierten Vertrauensintervalle enthalten den wahren μ.
2. Die konstruierten Vertrauensintervalle sind alle gleich breit.
3. Etwa 95 der Vertrauensintervalle enthalten deren jeweiligen Punktschätzer.

g) Wir führen einen Test durch, mit Hypothesen $H_0 : \mu = 0$ und $H_A : \mu \neq 0$, zum Signifikanzniveau von 5%. Wir bekommen den P-Wert 0.02. Welche Aussage ist richtig?

1. $P(\text{“Fehler 1. Art”}) \leq 0.02$.
2. $P(H_0 \text{ stimmt}) = 0.02$.
3. Das 95%-Vertrauensintervall für μ enthält 0 nicht.

h) Bei der Berechnung des P-Wertes eines Tests

1. muss man wissen, ob der Test einseitig oder zweiseitig ist.
2. braucht man die konkreten Daten nicht.
3. muss man das Signifikanzniveau α kennen.

i) Wir führen einen Wilcoxon-Test zum Signifikanzniveau von 5% durch. Es gilt $n = 20$; $H_0 : \mu = 2$; $H_A : \mu > 2$. Welche Aussage ist richtig?

1. Wenn H_0 nicht verworfen wird, dann gilt $\mu = 2$.
2. Die Macht unter $H_A : \mu = 5$ ist grösser als die Macht unter $H_A : \mu = 4$.
3. Wenn die beobachtete Teststatistik 155 ist, dann wird H_0 nicht verworfen.

j) Wir möchten testen, ob das erstgeborene Kind in einer Familie im Allgemeinen einen höheren IQ hat, als das letztgeborene Kind. Dazu wählen wir zufällig 150 Familien mit mindestens zwei Kindern aus, und nehmen IQ Tests bei den erstgeborenen und letztgeborenen Kindern ab. Welcher statistischer Test ist am besten geeignet?

1. Einseitig, gepaart.
2. Einseitig, ungepaart.
3. Zweiseitig, gepaart.
4. Zweiseitig, ungepaart.

2. (7 Punkte) Zwei Fabriken stellen elektrische Widerstände her. Fabrik A stellt 70 Prozent aller Widerstände her und 20 Prozent der in Fabrik A hergestellten Widerstände sind fehlerhaft. Fabrik B stellt 30 Prozent aller Widerstände her und 10 Prozent aus Fabrik B stammende Widerstände sind fehlerhaft.

Zu einem zufällig ausgewählten Widerstand, definieren wir die Zufallsvariable

$$X := \begin{cases} 1 & \text{falls der Widerstand fehlerhaft ist} \\ 0 & \text{sonst.} \end{cases}$$

a) Zeige, dass $E[X] = 0.17$. Wie ist X verteilt? Berechne $Var(X)$.

b) Wie wahrscheinlich ist es, dass ein zufällig ausgewählter fehlerhafter Widerstand aus Fabrik A stammt?

Einem Elektronikunternehmen werden jeden Tag 1'000 zufällig ausgewählte Widerstände zugeliefert. Die Produktion des Unternehmens muss an einem Tag stillstehen, wenn mindestens ein Fünftel der an diesem Tag zugelieferten Widerstände fehlerhaft ist.

c) Sei Z die Anzahl fehlerhafter Widerstände in einer Tageslieferung. Wie ist Z verteilt? Gib auch die Parameterwerte an.

d) Wie gross (annähernd) ist die Wahrscheinlichkeit, dass die Produktion des Unternehmens an einem Tag stillsteht?

e) Wie gross (annähernd) ist die Wahrscheinlichkeit, dass die Produktion des Unternehmens in einer Woche (5 Arbeitstage) mindestens an einem Tag stillsteht?

3. **(8 Punkte)** Ein Versicherer modelliert die Schadenforderungen in Millionen CHF, die in einem bestimmten Jahr wegen Skiunfälle zustande gekommen sind, mit der Zufallsvariable X. Dabei wird auch eine zweite Zufallsvariable Y benutzt, die der Wetterlage im Skigebiet Rechnung trägt. Wir nehmen an, dass die gemeinsame Verteilung von X und Y wie folgt aussieht: $Y \sim Uni[1,3]$ und $X|Y = y \sim Exp(y)$.

Der Versicherer hat 3 Millionen CHF Zweckreserven. Der Versicherer muss also die ganzen Zweckreserven verbrauchen, wenn die Schadenforderungen grösser oder gleich 3 Millionen CHF sind.

a) Was ist die Wahrscheinlichkeit, dass der Versicherer die ganzen Zweckreserven verbrauchen muss, wenn $Y = 1.5$?

b) Bestimme die gemeinsame Dichte von X und Y.

c) Berechne die Wahrscheinlichkeit, dass der Versicherer die ganzen Zweckreserven verbrauchen muss.

Tipp: Integriere zuerst nach x.

Damit der Versicherer nicht gleich insolvent ist sobald die Zweckreserven ausgehen, kauft er bei einem Rückversicherer die folgende Versicherung. Der Rückversicherer übernimmt dem Versicherer in einem Jahr die ganzen Schadenforderungen, wenn sie in diesem Jahr mindestens 3 Millionen Franken betragen. Die Schadenforderungen, für die der Rückversicherer aufkommt (in jenen Jahren, wo er sie übernehmen muss) modellieren wir mit der Zufallsvarialbe Z. Wir nehmen an, dass der Wetterfaktor über mehrere Jahre konstant bleibt; das heisst $Y = \lambda$, wobei λ fix aber unbekannt ist. Dann hat Z die Dichte

$$f_Z(z) = \lambda e^{-\lambda z} 1_{[3,\infty)}(z) c.$$

wobei c eine Konstante ist, die von λ abhängt.

d) Zeige, dass $f_Z(z) = \lambda e^{\lambda(3-z)} 1_{[3,\infty)}(z)$.

Der Rückversicherer versucht λ aus Daten aus den Vorjahren zu schätzen. Es stehen sechs Daten zur Verfügung: z_1, z_2,..., z_6, die als unabhängige Realisierungen von Z betrachtet werden können. Es gilt $z_i > 3$ für jedes i.

e) Berechne den MLE $\hat{\lambda}$ als Funktion von z_1, z_2,..., z_6 .

4. (10 Punkte) Ein Fahrradteam möchte eine neue Fahrradbekleidung testen. Das Team hat vor, folgendes Experiment durchzuführen: Zuerst fahren alle Fahrradfahrer des Teams eine bestimmte Strecke mit der alten Bekleidung, dann, nach hinreichender Erholungszeit fahren sie dieselbe Strecke mit der neuen Bekleidung.

a) Optimiere die Versuchsplanung: Gib dem Team einen Verbesserungsvorschlag. Begründe ihn.

Nach Durchführung des verbesserten Experiments stehen folgende Daten zur Verfügung (alle Angaben in Minuten):

Fahrradfahrer Nr.	1	2	3	4	5	6
alte Bekleidung (x_i)	52.6	47.1	48.1	50.7	55.5	48.0
neue Bekleidung (y_i)	48.1	47.3	48.6	47.1	49.2	46.7

Aus diesen Daten berechnet man folgende Werte: $\bar{x} = 50.333$; $\bar{y} = 47.833$; $s_X^2 = 10.571$; $s_Y^2 = 0.927$; $s_{X-Y}^2 = 7.476$.

Das Team möchte statistisch überprüfen, ob die neue Bekleidung eine Verbesserung der Fahrleistung bewirkt. Zu diesem Zweck wird ein gepaarter Test durchgeführt. Die Daten $x_i - y_i$, $i = 1, ..., 6$ können als Realisierung von i.i.d. Zufallsvariablen U_i, $i = 1, ..., 6$, aufgefasst werden. Als Signifikanzniveau wählen wir 5%.

b) Beantworte die folgenden Fragen:

(i) Warum ist der gepaarte Test in diesem Fall die geeignete Wahl?

(ii) Ist der Test einseitig oder zweiseitig?

c) Vor der Auswertung des QQ-Plots vermutet das Team, dass die U_i eine schiefe Verteilung haben.

(i) Welchen Test kann man unter diesen Umständen durchführen? Begründe deine Antwort.

(ii) Wie ist die Nullhypothese und die Alternativhypothese?

(iii) Führe diesen Test durch, mittels Berechnung des P-Werts.

d) Nach Auswertung des QQ-Plots stellt man fast, dass die U_i eine symmetrische Verteilung haben, die sogar gut mit der Normalverteilung vereinbar ist.

(i) Welcher Test ist jetzt am meisten geeignet? Begründe deine Antwort.

(ii) Wie ist die Nullhypothese und die Alternativhypothese?

(iii) Führe diesen Test durch, mittels Berechnung des Verwerfungsbereichs.