Technische Universität Berlin
SoSe 2025

Prof. Dr. Carlos Améndola
Janike Oldekop

Stochastik für Informatik (9 LP)

– Klausur vom 07.10.2025 –

Füllen Sie bitte zuerst dieses Deckblatt vollständig und leserlich aus. Schreiben Sie auf jedes von Ihnen benutzte Blatt Ihren Namen und Ihre Matrikelnummer. Mit Bleistift geschriebene Klausuren können nicht gewertet werden.

Bei der Klausur sind 100 Punkte zu erreichen. Als Hilfsmittel dürfen ein beidseitig handbeschriebenes DIN A4-Blatt und ein nicht-programmierbarer Taschenrechner benutzt werden. Es sind keine weiteren Hilfsmittel zugelassen.

Verteilungstabellen befinden sich am Ende der Klausur. Runden Sie Ihre Zwischenergebnisse auf vier Nachkommastellen und Ihre Endergebnisse auf zwei Nachkommastellen. Versuchen Sie, alle Teile einer Aufgabe zu lösen, da Teilpunkte vergeben werden können, insbesondere bei Folgefehlern.

Geben Sie immer vollständige Rechenwege bzw. Begründungen an. In Aufgabe 1 sind keine Begründungen notwendig. Die Bearbeitungszeit beträgt 90 Minuten.

Das Vorlesungsteam wünscht Ihnen eine erfolgreiche Klausur!

Aufgabe 1 $2 \times 10 = 20$ *Punkte*

Entscheiden Sie, ob die folgenden Aussagen wahr oder falsch sind.

		Wahr	Falsch
a)	In einem Laplace-Raum ist jedes Ereignis gleich wahrscheinlich.	○	○
b)	Für alle Ereignisse A, B aus einem Wahrscheinlichkeitsraum gilt $\mathbb{P}(A \cup B) \leq \mathbb{P}(A) + \mathbb{P}(B)$.	○	○
c)	Der Erwartungswert einer diskreten Zufallsvariable kann niemals negativ sein.	○	○
d)	Die Varianz einer stetigen Zufallsvariable kann niemals null sein.	○	○
e)	Bei einer stetigen Zufallsvariable ist die Wahrscheinlichkeit, dass sie genau einen bestimmen Wert annimmt, gleich null.	○	○
f)	Für $X \sim \mathrm{Ber}(p)$ gilt $\mathbb{E}[X] = \mathbb{E}[X^2]$.	○	○
g)	Der Maximum-Likelihood-Schätzer ist immer erwartungstreu.	○	○
h)	Die Varianz eines Schätzers sollte möglichst groß sein, um zuverlässige Schätzungen zu erhalten.	○	○
i)	Ein Fehler 1. Art bedeutet, dass man die Nullhypothese beibehält, obwohl sie falsch ist.	○	○
j)	Ist der p-Wert größer als das Signifikanzniveau, so wird die Nullhypothese verworfen.	○	○

Aufgabe 2 $3 + 3 + 3 + 6 = 15$ *Punkte*

Angenommen, 20 Prozent der Raucher*innen und ein Prozent der Nichtraucher*innen erkranken an Lungenkrebs. In einer Bevölkerung gibt es gleich viele Männer und Frauen, jedoch sind 30 Prozent der Männer und 20 Prozent der Frauen Raucher*innen. Der Einfachheit halber nehmen wir an, dass jede Person in der Bevölkerung entweder männlich oder weiblich ist. Eine Person wurde zufällig aus der Bevölkerung ausgewählt.

a) Wie groß ist die Wahrscheinlichkeit, dass die Person raucht und weiblich ist?

b) Wie groß ist die Wahrscheinlichkeit, dass die Person nicht raucht und Lungenkrebs hat?

c) Wie groß ist die Wahrscheinlichkeit, dass die ausgewählte Person Lungenkrebs hat?

d) Wenn die ausgewählte Person Lungenkrebs hat, wie groß ist die bedingte Wahrscheinlichkeit, dass es sich um einen Nichtraucher (männlich) handelt?

Aufgabe 3 $3 + 2 + 4 + 3 + 5 = 17$ *Punkte*

Die Flugzeit (in Minuten) eines Wildvogels zu seinem Nestplatz wird durch eine Zufallsvariable X mit Erwartungswert $\mu = 45$ und Standardabweichung $\sigma = 8$ beschrieben. Es wird eine Stichprobe von 64 Vögeln beobachtet.

a) Bestimmen Sie eine obere Schranke für die Wahrscheinlichkeit, dass die Flugzeit eines einzelnen Vogels um mehr als 20 Minuten vom Erwartungswert abweicht.

b) Bestimmen Sie den Erwartungswert und die Standardabweichung des Stichprobenmittelwertes.

c) Bestimmen Sie unter Verwendung der Normalverteilung die Wahrscheinlichkeit, dass der Stichprobenmittelwert zwischen 43 und 47 Minuten liegt.

d) Bestimmen Sie ein 95%-Konfidenzintervall für den Mittelwert der Flugzeit, wobei nun angenommen wird, dass der Erwartungswert unbekannt ist und ein mittlerer Flugzeitwert von $\bar{\mu} = 45$ gemessen wurde.

e) Ein Vogelschutzgebiet möchte eine Ruhezone einrichten, in der die Flugzeit im Durchschnitt höchstens d Minuten beträgt, sodass 90% der beobachteten Mittelwerte der 64 Vögel darunter liegen. Berechnen Sie d.

Aufgabe 4 $3 + 6 + 4 + 2 + 4 = 19$ *Punkte*

Ein Softwareunternehmen untersucht bei einem Entwicklerteam den Zusammenhang zwischen täglichem Kaffeekonsum (in Tassen) und der Codequalität (Score 0-100). Bei fünf Entwickler:innen wurden die folgenden Daten erhoben:

Entwickler	**Kaffee (x)**	**Codequalität (y)**
A	0	60
B	1	65
C	2	72
D	3	75
E	4	80

Es wird angenommen, dass ein linearer Zusammenhang vorliegt.

a) Visualisieren Sie die Daten in einem Punktplot. Wirkt ein linearer Zusammenhang plausibel?

b) Bestimmen Sie die empirische Varianz von x und die empirische Kovarianz von x und y.

c) Berechnen Sie die Regressionsgerade $y = ax + b$.

d) Schätzen Sie die Codequalität bei fünf Tassen Kaffee pro Tag.

e) Berechnen Sie den empirischen Korrelationskoeffizienten und interpretieren Sie dessen Wert hinsichtlich der Stärke und Richtung des Zusammenhangs.

Aufgabe 5 $2 + 3 + 2 + 3 + 5 = 15$ *Punkte*

Ein Softwareunternehmen möchte überprüfen, ob die neue Version ihres Algorithmus im Durchschnitt schneller ist, als die alte Version. Für die alte Version ist bekannt, dass die durchschnittliche Laufzeit zehn Sekunden beträgt. Es werden zwölf Messungen mit der neuen Version durchgeführt, die folgende Laufzeiten (in Sekunden) ergeben:

9,1	8,7	9,5	9,3	8,9	9,8	9,0	8,6	9,2	8,8	9,4	9,1

Hinweis: Es gilt: $9,1 + 8,7 + \cdots + 9,1 = 109,4$ und $9,1^2 + 8,7^2 + \cdots + 9,1^2 = 998,7$.
Die Analyse soll auf dem Signifikanzniveau $\alpha = 0,05$ durchgeführt werden.

a) Berechnen Sie den empirischen Mittelwert und die empirische Standardabweichung der Laufzeiten.

b) Welcher statistische Test ist geeignet, um zu überprüfen, ob die neue Version im Durchschnitt schneller ist, als die alte Version? Nennen Sie die Vorraussetzungen, die für die Anwendung dieses Tests erfüllt sein müssen.

c) Formulieren Sie die Null- und Alternativhypothese für den Test, den Sie in Teilaufgabe b) vorgeschlagen haben.

d) Berechnen Sie den Wert der Teststatistik der Stichprobe.

e) Entscheiden Sie, ob die Nullhypothese verworfen werden kann. Interpretieren Sie das Ergebnis im Kontext des Problems.

Aufgabe 6 $4 + 2 + 3 + 3 + 2 = 14$ *Punkte*

In einer Universität steht ein Getränkeautomat mit drei möglichen Getränken: Kaffee, Kakao und Mangoshake (neu eingeführt vom tropikaffinen Institutsleiter). Im Rahmen eines Studierendenprojekte wird das Trinkverhalten einer großen Anzahl an Personen über mehrere Tage beobachtet. Dabei zeigt sich folgendes Muster:

- Wer heute Kaffee trinkt, entscheidet sich am nächsten Tag mit 70% Wahrscheinlichkeit für Kakao, mit 10% für Mangoshake und mit 20% wieder für Kaffee.
- Wer heute Kakao trinkt, wählt am nächsten Tag zu 60% erneut Kakao, zu 30% Mangoshake und zu 10% Kaffee.
- Wer heute Mangoshake trinkt, entscheidet sich am nächsten Tag mit 30% für Kaffee, mit 20% für Kakao und mit 50% wieder für einen Mangoshake.

Es wird angenommen, dass sich dieses Trinkverhalten durch eine Markov-Kette modellieren lässt.

a) Geben Sie die Übergangsmatrix P der Markov-Kette an und stellen Sie ihren Übergangsgraphen dar.

b) Ist die Markov-Kette irreduzibel? Ist sie aperiodisch? Begründen Sie Ihre Entscheidung.

c) Berechnen Sie die Wahrscheinlichkeit, dass eine Person in genau zwei Tagen von Kaffee auf Mangoshake wechselt.

d) Zeigen Sie, dass $\pi^T = (\pi_{\text{Kaffee}}, \pi_{\text{Kakao}}, \pi_{\text{Mango}}) = \frac{1}{76}(14, 37, 25)$ eine invariante Verteilung der Markov-Kette ist. Ist diese eindeutig?

e) Wie groß ist der langfristige Anteil der Personen, die Mangoshake trinken?