Prof. Dr. M. Schweizer ETH Zürich Frühling 2014

Wahrscheinlichkeit und Statistik
BSc D-INFK

Name:	
Vorname:	
Stud. Nr.:	

Das Folgende bitte nicht ausfüllen!

Aufg.	Summe	Kontr.	Pkte.-Max.
1			10
2			10
3			10
4			10

Punktetotal:	
Vollständigkeit:	

Bitte wenden!

Hinweise zur Prüfung

Prüfungsdauer: 2 Stunden.

Hilfsmittel: 10 A4-Seiten resp. 5 Blätter Zusammenfassung. Kein Taschenrechner!

Bitte beachten Sie folgende Punkte:

- Legen Sie Ihre Legi offen auf den Tisch.
- Tragen Sie Ihre Daten in dieses Deckblatt ein und schreiben Sie auf jedes Blatt Ihren Namen.
- Beginnen Sie jede Aufgabe **auf einem neuen Blatt**.
- Schreiben Sie nicht mit Bleistift, rotem oder grünem Kugelschreiber.
- Um die volle Punktzahl zu erreichen, begründen Sie alle Resultate durch Zwischenschritte und -rechnungen (ausser Aufgabe 1) und vereinfachen Sie die Resultate so weit wie möglich.
- Lesen Sie alle Aufgaben durch, bevor Sie beginnen. Für eine genügende Note wird nicht erwartet, dass Sie alle Aufgaben in der Ihnen zur Verfügung stehenden Zeit lösen können.
- Es dürfen sich nur erlaubte Hilfsmittel auf dem Tisch befinden, d.h. 10 A4-Seiten resp. 5 Blätter Zusammenfassung. Kein Taschenrechner!

Siehe nächstes Blatt!

Aufgaben

---

1. (10 Punkte)

Bei den folgenden 10 Fragen ist jeweils genau eine Antwort richtig. Für jede richtig beantwortete Frage erhält man 1 Punkt, für jede falsch beantwortete Frage erhält man $-1/2$ Punkte, und für jede nicht beantwortete Frage erhält man 0 Punkte. Ingesamt erhält man mindestens 0 Punkte für die gesamte Aufgabe. **Bitte benutzen Sie zur Beantwortung der Fragen das beiliegende Antwortblatt.**

a) In einer Gruppe von 100 Informatikstudierenden besitzen 20% einen Tablet-Computer; die Hälfte der Tablet-Computer-Besitzer nutzt ausserdem den Internetdienst "Zwitscher". Ingesamt nutzen 30 der 100 Studierenden "Zwitscher". Gegeben, dass eine aus diesen 100 Studierenden zufällig ausgewählte Person "Zwitscher" nutzt, wie gross ist die Wahrscheinlichkeit, dass sie auch einen Tablet-Computer besitzt?

(i) $\frac{1}{3}$.
(ii) $\frac{1}{10}$.
(iii) Es sind nicht genügend Informationen vorhanden, um diese Wahrscheinlichkeit zu bestimmen.

b) Aus den 100 Studierenden von Teilaufgabe **a)** wird eine Person zufällig ausgewählt. Gegeben, dass sie keinen Tablet-Computer besitzt, wie gross ist die Wahrscheinlichkeit, dass sie "Zwitscher" nutzt?

(i) $\frac{1}{2}$.
(ii) $\frac{1}{3}$.
(iii) $\frac{1}{4}$.

c) Ein fairer Würfel wird solange geworfen, bis zum ersten Mal eine 6 erscheint. Wie gross ist die Wahrscheinlichkeit, dass man mindestens 3 Würfe benötigt, bis dieses Ereignis eintritt?

(i) $\left(\frac{5}{6}\right)^2$.
(ii) $\left(\frac{5}{6}\right)^2 \frac{1}{6}$.
(iii) $\binom{3}{2} \frac{1}{6} \left(\frac{5}{6}\right)^2$.

d) Seien X und Y zwei diskrete Zufallsvariablen mit folgender gemeinsamer Gewichtsfunktion $p(x,y) = P[X = x, Y = y]$:

$x \quad \backslash\, y$	0	1	2
-1	0	$\frac{1}{4}$	0
0	$\frac{1}{4}$	0	$\frac{1}{4}$
1	0	$\frac{1}{4}$	0

Welche der folgenden Aussagen trifft zu?

(i) X und Y sind weder unabhängig noch unkorreliert.
(ii) X und Y sind nicht unabhängig, aber unkorreliert.
(iii) X und Y sind unabhängig.

e) Sei $a \in [0, \infty)$ und die Funktion $f : \mathbb{R} \to \mathbb{R}$ gegeben durch

$$f(x) = \begin{cases} 1 - a + 2ax & \text{falls } x \in [0,1], \\ 0 & \text{sonst.} \end{cases}$$

Welche der folgenden Aussagen trifft zu?

(i) f ist für jedes $a \in [0, \infty)$ die Dichte einer Zufallsvariablen.
(ii) f ist für kein $a \in [0, \infty)$ die Dichte einer Zufallsvariablen.
(iii) f ist genau dann die Dichte einer Zufallsvariablen, wenn $a \in [0, 1]$.

f) Seien X, Y Zufallsvariablen mit gemeinsamer Dichte

$$f(x,y) = \begin{cases} 8x(1-y) & \text{falls } x, y \geq 0 \text{ und } x + y \leq 1, \\ 0 & \text{sonst.} \end{cases}$$

Berechnen Sie die Dichte $f_X(x)$ der Randverteilung von X.

(i) $f_X(x) = 4x$ für $x \in [0, 1]$.
(ii) $f_X(x) = 4x(1 - x^2)$ für $x \in [0, 1]$.
(iii) $f_X(x) = 4x(1 - (1 - x)^2)$ für $x \in [0, 1]$.

In den folgenden Teilaufgaben sei $\overline{X}_n := \frac{1}{n} \sum_{i=1}^n X_i$.

g) Seien $X_1, \ldots, X_4$ i.i.d. $\mathcal{N}(0, 1)$-verteilt. Bestimmen Sie mit Hilfe der beiliegenden Tabelle die Wahrscheinlichkeit $P[\overline{X}_4 > -\frac{1}{2}]$.

(i) $P[\overline{X}_4 > -\frac{1}{2}] = 0.6915$.
(ii) $P[\overline{X}_4 > -\frac{1}{2}] = 0.8413$.
(iii) $P[\overline{X}_4 > -\frac{1}{2}] = 0.97725$.

Siehe nächstes Blatt!

h) Seien $X_1, \dots, X_n$ $\mathcal{P}(\lambda)$-verteilt mit $\lambda > 0$. Gilt dann $P[|\overline{X}_n - \lambda| \geq c] \leq \frac{\lambda}{nc^2}$ für jedes $c > 0$?

(i) Die Aussage stimmt ohne weitere Voraussetzungen.

(ii) Die Aussage stimmt, falls $X_1, \dots, X_n$ unabhängig sind.

(iii) Die Aussage stimmt im Allgemeinen nicht, auch nicht wenn $X_1, \dots, X_n$ unabhängig sind.

i) Sei $X_1, X_2, \dots$ eine Folge von i.i.d. Zufallsvariablen mit $X_i \sim \mathcal{U}(0, \vartheta)$ für einen Parameter $\vartheta > 0$. Für jedes $n \in \mathbb{N}$ sei $T^{(n)} := 2\overline{X}_n = \frac{2}{n}\sum_{i=1}^{n} X_i$ ein Schätzer für ϑ. Welche der folgenden Aussagen trifft zu?

(i) Jeder Schätzer $T^{(n)}$ ist erwartungstreu für ϑ und die Folge $T^{(n)}$, $n \in \mathbb{N}$, ist konsistent für ϑ.

(ii) Jeder Schätzer $T^{(n)}$ ist erwartungstreu für ϑ, aber die Folge $T^{(n)}$, $n \in \mathbb{N}$, ist nicht konsistent für ϑ.

(iii) $T^{(n)}$ ist im Allgemeinen nicht erwartungstreu für ϑ.

j) Ein Softwareunternehmen S will einen neuen Algorithmus zur Lösung eines Optimierungsproblems (dessen Lösung von verschiedenen Parametern abhängt) auf den Markt bringen. Um Kunden zu gewinnen, möchte das Unternehmen statistisch nachweisen, dass die mittlere Laufzeit μ_S ihres Algorithmus kleiner ist als die mittlere Laufzeit μ_M des Algorithmus des Marktführers M. Hierfür werden zufällig 1000 Parametersätze des Optimierungsproblems ausgewählt und die Laufzeiten beider Algorithmen für jedes dieser Probleme notiert. Welcher der folgenden Tests eignet sich am besten für das Vorhaben des Unternehmens S?

(i) Ein zweiseitiger ungepaarter Zweistichproben-Test.

(ii) Ein einseitiger ungepaarter Zweistichproben-Test.

(iii) Ein einseitiger gepaarter Zweistichproben-Test.

2. (10 Punkte)

Stellen Sie sich vor, Sie haben den Auftrag bekommen, für ein Projekt geeignete Fachleute zu finden. Falls Sie mindestens einen Bewerber vorschlagen, der im Bewerbungsprozess überzeugt und eingestellt wird, so bekommen Sie eine Kommission; andernfalls gehen Sie leer aus. Sie schlagen nur Bewerber vor, die nach allen gängigen Bewertungskriterien einwandfrei abgeschnitten haben, aber Sie wissen, es kommt darüber hinaus auf Fähigkeit A an, ob der Bewerber eingestellt wird: Wenn Sie einen Bewerber mit Fähigkeit A vorschlagen, bekommt dieser den Job mit 80% Wahrscheinlichkeit; allerdings geht aus Statistiken vergangener Jahre eindeutig hervor, dass nur 10% der einwandfreien Bewerber Fähigkeit A besitzen. Schlagen Sie einen einwandfreien Bewerber vor, der Fähigkeit A nicht hat, wird dieser mit einer Wahrscheinlichkeit von $\frac{48}{90}$ nicht eingestellt.

a) Leider können Sie Fähigkeit A nicht ohne Weiteres feststellen und wählen einen einwandfreien Kandidaten zufällig aus. Wie gross ist die Wahrscheinlichkeit, dass der von Ihnen vorgeschlagene Kandidat sowohl Fähigkeit A besitzt als auch eingestellt wird?

b) Ein von Ihnen zufällig ausgewählter einwandfreier Kandidat wird nicht eingestellt. Gegeben diese Information, wie gross ist die Wahrscheinlichkeit, dass ihr Kandidat Fähigkeit A besitzt?

c) Ihr Auftraggeber macht es zur Bedingung, dass Sie nur drei Kandidaten vorschlagen. Sie wählen drei einwandfreie Kandidaten zufällig aus.
Bezeichne Z die Anzahl der Kandidaten mit Fähigkeit A unter den drei vorgeschlagenen Kandidaten. Benennen Sie die Verteilung von Z und geben Sie die Wahrscheinlichkeiten $P[Z = k]$ für $k \in \{0, 1, 2, 3\}$ an. (Eine numerische Auswertung wird nicht zwingend erwartet.)

d) Sie haben einen Test entwickelt, mit dem Sie nun Fähigkeit A unter allen einwandfreien Kandidaten wesentlich besser feststellen können. Wenn sie nur noch einwandfreie Kandidaten in Betracht ziehen, die Ihren Test bestanden haben, und einen zufällig auswählen, dann besitzt dieser die Fähigkeit A mit Wahrscheinlichkeit 80%. Der Test gilt als relativ hart, ist jedoch sehr zuverlässig in der Hinsicht, dass alle einwandfreien Kandidaten mit Fähigkeit A den Test sicher bestehen. Sie wenden Ihren Test nun auf einen beliebigen einwandfreien Kandidaten an. Wie hoch ist die Wahrscheinlichkeit, dass dieser besteht?

e) Nehmen Sie an, dass Sie die Anzahl aller einwandfreien Kandidaten auf 5 reduzieren konnten, von denen nur noch genau einer die Fähigkeit A nicht besitzt. Sie wählen drei dieser fünf Kandidaten zufällig aus. Bezeichne $\widetilde{Z}$ die Anzahl der Kandidaten mit Fähigkeit A unter den drei vorgeschlagenen Kandidaten. Benennen Sie die Verteilung von $\widetilde{Z}$ und geben Sie die Wahrscheinlichkeiten $P[\widetilde{Z} = k]$ für $k \in \{0, 1, 2, 3\}$ an. Berechnen Sie die Wahrscheinlichkeit, dass alle drei Kandidaten Fähigkeit A besitzen.

Siehe nächstes Blatt!

3. (10 Punkte)

Bert hat im Warenladen Kerzen vom Typ A und B gekauft. Er beobachtet, dass die Brenndauer der beiden Kerzentypen jeweils nicht konstant ist. Nach seinen Beobachtungen hat die Brenndauer X_A (in Stunden) der Kerzen vom Typ A folgende Dichte:

$$f_{X_A}(x) = \begin{cases} 0, & 0 \leq x < 1, \\ \frac{1}{2}(3 - x), & 1 \leq x < 3, \\ 0, & x \geq 3. \end{cases}$$

Bert beobachtet zudem, dass die Brenndauer X_B (in Stunden) der Kerzen vom Typ B auf dem Intervall $[2, 6]$ gleichverteilt ist.

a) Berechnen Sie den Erwartungswert von X_A.

b) Wie viele Kerzen vom Typ A muss Bert gleichzeitig anzünden, damit nach zwei Stunden im Mittel noch genau 5 Kerzen brennen? Wir nehmen an, dass die Brenndauer der einzelnen Kerzen unabhängig sind.

c) Bert zündet eine Kerze vom Typ A an. Nach eineinhalb Stunden brennt die Kerze immer noch. Gegeben diese Information, wie gross ist die Wahrscheinlichkeit, dass die Kerze noch mindestens eine weitere Stunde brennt?

Im Folgenden nehmen wir an, dass die Brenndauern der beiden Kerzentypen unabhängig sind.

d) Bert zündet eine Kerze vom Typ A und eine Kerze vom Typ B gleichzeitig an. Wie gross ist die Wahrscheinlichkeit, dass beide Kerzen jeweils mindestens zweieinhalb Stunden brennen?

e) Bert zündet eine Kerze vom Typ A und eine Kerze vom Typ B gleichzeitig an. Wie gross ist die Wahrscheinlichkeit, dass die Kerze vom Typ B zuerst abgebrannt ist?

4. (10 Punkte)

Um die Frage zu beantworten, ob L. Messi oder C. Ronaldo der bessere Fussballspieler ist, hat die FIFA ein System zum Bewerten von Fussballspielern entwickelt. Nach jedem Spiel wird mittels einer komplexen Formel jedem Spieler eine Wertung verliehen. Die Formel basiert auf vielen verschiedenen Aspekten der fussballerischen Leistung wie beispielsweise Tore, Torvorlagen, Dribblings, Passquote, zurückgelegte Laufstrecke, rote und gelbe Karten, Fouls, etc. Es ist bekannt, dass die Wertung beider Spieler in guter Näherung normalverteilt ist mit unbekannten Mittelwerten $\mu_{CR} > 0$ für C. Ronaldo und $\mu_{LM} > 0$ für L. Messi. Die bekannte Varianz ist für beide Spieler $\sigma^2 = 9$. Desweiteren sind die Wertungen für verschiedene Spiele unabhängig voneinander.

Ronaldo ist exzellent in die Saison gestartet. Seine Punktewertungen für die ersten neun Spiel der Saison sind in folgender Tabelle aufgelistet.

i	1	2	3	4	5	6	7	8	9
x_i	93	81	73	79	90	86	72	83	90

empirisches Stichprobenmittel: $\bar{x}_9 = 83$
empirische Stichprobenstandardabweichung: $s = 7.4833$

Er selbst ist jedoch nicht davon überzeugt, dass er bereits seine beste Leistung abgeliefert hat und behauptet, dass seine Wertung im Mittel über 84 liegen sollte. Ihre Aufgabe ist es zu überprüfen, ob Ronaldos Behauptung gerechtfertigt ist.

a) Welcher statistische Test eignet sich aufgrund der Modellannahmen um diese Behauptung zu überprüfen? Ist der Test einseitig oder zweiseitig?

b) Führen Sie einen passenden statistischen Test durch, um zu überprüfen, ob Ronaldos Behauptung auf dem 10%-Niveau gerechtfertigt ist. Geben Sie

(i) die sinnvolle Nullhypothese und die Alternative,
(ii) die Teststatistik,
(iii) den Verwerfungsbereich zum 10%-Niveau,
(iv) den beobachteten Wert der Teststatistik und den Testentscheid an.

c) Die pro-Real Zeitung *Marca* ist sogar noch stärker von den Fähigkeiten Ronaldos überzeugt. Sie behauptet, dass $\mu_{CR} = 85$. Berechnen Sie die Macht des obigen Tests für diese Wahl von μ_{CR}.

Hinweis: Falls sie **b)** nicht lösen konnten, dann nehmen Sie an, dass die Teststatistik das Stichprobenmittel $T = \overline{X}_9$ und der Verwerfungsbereich $K = [85.28, \infty)$ sei.

Siehe nächstes Blatt!

Nachdem Messi bereits die letzten vier "Ballon D'Or" Trophäen gewonnen hat, möchte er unbedingt vor seinem Rivalen bleiben. Trotzdem erwischt er einen schlechten Start in die Saison und glaubt, dass er mindestens eine Wertung von 95 in seinem nächsten Spiel benötigt, um Ronaldo wieder näherzukommen. Sei p die Wahrscheinlichkeit, dass Messi in einem Spiel eine Wertung von mindestens 95 erreicht. Um seine Chancen für eine solche Leistung zu schätzen, betrachtet er seine Leistungen der letzten fünf Saisons. Genauer sei G_i, $i = 1, \ldots, 5$, die Anzahl der Spiele in der i-ten Saison, die Messi benötigte, um seine erste Wertung von mindestens 95 zu erreichen.

d) Schreiben sie p in Abhängigkeit der Verteilungsfunktion Φ der Standardnormalverteilung und des unbekannten Parameters μ_{LM}.

e) Bestimmen Sie den Maximum-Likelihood-Schätzer für p auf Basis von $G_1, \ldots, G_5$.

Viel Erfolg!