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 bitte 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. Es gibt pro richtig beantwortete Frage 1 Punkt und pro falsche Antwort $1/2$ Punkt Abzug. Minimal erhält man für die gesamte Aufgabe 0 Punkte. **Bitte benützen Sie das beiliegende Antwortblatt**.

a) Die Licht AG wird von 3 Lieferanten A, B und C mit insgesamt 500 Sparlampen vom gleichen Typ beliefert. Dabei gilt

Lieferant	A	B	C
Anzahl	100	250	150
davon defekt	2	6	2

Wie gross ist die Wahrscheinlichkeit, dass eine zufällig entnommene nicht defekte Lampe vom Lieferant A stammt?

(i) $\frac{1}{6}$,
(ii) $\frac{1}{4}$,
(iii) $\frac{1}{5}$.

b) Wir betrachten nun den Lieferanten C aus der Teilaufgabe **a)**. Zu Kontrollzwecken werden aus seiner Lieferung 3 Sparlampen zufällig entnommen. Die Zufallsvariable X beschreibe die Anzahl der dabei enthaltenen defekten Sparlampen. Wie gross ist die Wahrscheinlichkeit, dass sich in so einer Stichprobe genau 2 defekte Sparlampen befinden?

(i) $\frac{2}{150}$,
(ii) $\frac{1}{25*149}$,
(iii) $\frac{2}{149*150}$.

c) Sei $X \sim \mathcal{N}(0,1)$ und damit Var$[X^2] = 2$. Wie gross ist $E[X^4]$?

(i) $E[X^4] = 3$,
(ii) $E[X^4] = 1$,
(iii) $E[X^4] = 12$.

Bitte wenden!

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

x \ y	-2	0	2
0	$\frac{1}{5}$	0	$\frac{1}{5}$
1	0	$\frac{1}{5}$	0
2	$\frac{1}{5}$	0	$\frac{1}{5}$

Welche der Aussagen stimmt?

(i) X und Y sind korreliert und abhängig.
(ii) X und Y sind unkorreliert und abhängig.
(iii) X und Y sind korreliert und unabhängig.

e) Sei X eine Zufallsvariable mit $\text{Var}[X] = 3$ und $E[X^2] = 7$. Dann gilt

(i) $P[|X - 2| \geq 3] \leq \frac{1}{2}$,
(ii) $P[|X - 2| \geq 3] \leq \frac{1}{3}$,
(iii) $P[|X - 2| \geq 3] \leq \frac{1}{7}$.

f) Sei X eine beliebige Zufallsvariable und $c \in (0, \infty)$. Wir betrachten die Funktion

$$F_X(x) = \begin{cases} (x-2)^2 - 4, & 0 \leq x \leq c, \\ 0, & \text{sonst.} \end{cases}$$

Welche der folgenden Aussagen stimmt?

(i) F_X ist für $c = 2 + \sqrt{5}$ eine Verteilungsfunktion von X,
(ii) F_X ist für $c = \sqrt{5}$ eine Verteilungsfunktion von X,
(iii) F_X ist keine Verteilungsfunktion von X.

g) Sei X exponentialverteilt mit Parameter $\lambda > 0$ und $Y = e^{\lambda X}$. Was ist die Verteilungsfunktion $F_Y(t)$ von Y?

(i) Die Verteilungsfunktion von Y kann nicht bestimmt werden, da wir nicht wissen, ob X und Y unabhängig sind.
(ii) $F_Y(t) = 1 - \frac{1}{t}$, für alle $t \geq 1$, sonst $F_Y(t) = 0$,
(iii) $F_Y(t) = 1 - \frac{1}{t}$, für alle $t > 0$, sonst $F_Y(t) = 0$.

Siehe nächstes Blatt!

h) Seien A, B und C Ereignisse mit $P[A|C] = P[C] = \frac{1}{2}$ und $P[B \cap C] = \frac{1}{8}$. Welche Aussage stimmt?

(i) $P[A \cup B|C] = \frac{3}{4}$, falls A und B auch noch disjunkte Ereignisse sind,

(ii) $P[A \cup B|C] = \frac{3}{4}$, falls A, B und C auch noch paarweise unabhängige Ereignisse sind,

(iii) $P[A \cup B|C] = \frac{3}{4}$, wobei A, B und C sonst beliebige Ereignisse sind.

i) Welche Aussage gilt für einen Test zum Niveau α?

i) Die Macht des Tests soll möglichst klein sein.

ii) Die Wahrscheinlichkeit für den Fehler 2. Art soll möglichst klein sein.

iii) Die Wahrscheinlichkeit für den Fehler 1. Art soll möglichst gross sein.

j) Welche Aussage ist richtig?

(i) Eine Folge von Schätzern $T^{(n)}$, $n \in \mathbb{N}$, heisst konsistent für ϑ, falls eine Vergrösserung der Stichprobe dazu führt, dass $T^{(n)}$ tendenziell näher bei ϑ liegt.

(ii) Die Vergrösserung der Stichprobe führt nie zur besseren Schätzung von ϑ.

(iii) Ist $T^{(n)}$ erwartungstreu für ϑ, so ist $T^{(n)}$ auch konsistent für ϑ.

Bitte wenden!

2. (10 Punkte)

Ein Kleinunternehmen interessiert sich für eine neue Antivirensoftware. Die Unternehmensführung kann zwischen einem kostenlosen open-source Produkt A und einem zum Preis von 500 Franken kommerziell erhältlichen Produkt B wählen. Um den unterschiedlichen Kosten, die durch die verschiedenen Arten von Schadprogrammen verursacht werden, Rechnung zu tragen, werden die Computerviren in zwei Kategorien V und W eingeteilt.

Tests haben gezeigt, dass Produkt A in 80% der Fälle eine Warnung ausgibt, falls ein Dokument Viren der Kategorie V enthält. Falls die Viren in Kategorie W fallen, schlägt Produkt A nur in 70% der Fälle Alarm. Falls die Dokumente virenfrei sind, gibt Produkt A in 0.1% der Fälle ebenfalls eine Warnung aus. Produkt B warnt zu 90%, falls es sich um Viren der Kategorie V handelt und zu 80%, wenn die Viren in Kategorie W fallen. Bei 1% der virenfreien Dokumente schlägt Produkt B ebenfalls Alarm. Aus Erfahrung weiss man, dass 1% der heruntergeladenen Dokumente Viren (V oder W) enthalten und dass es sich bei 10% davon um Kategorie W handelt. Ausserdem geht man davon aus, dass ein Dokument nie beide Arten von Viren enthalten kann.

*Bei **a) - c)** muss das Ergebnis nur als Bruch von zwei ganzen Zahlen geschrieben und nicht weiter vereinfacht werden.*

a) Wie gross ist die Wahrscheinlichkeit, dass die Antivirensoftware B bei einem heruntergeladenen Dokument eine Warnung ausgibt?

b) Wie gross ist die Wahrscheinlichkeit, dass ein heruntergeladenes Dokument Viren (V oder W) enthält, obwohl die Antivirensoftware A keine Warnung ausgibt?

c) Das Unternehmen schätzt, dass pro nicht erkanntem Virus V Kosten in der Höhe von 10 Franken entstehen. Falls ein Virus der Kategorie W nicht erkannt wird, belaufen sich die Kosten sogar auf 100 Franken. Pro Fehlalarm fallen Kosten in der Höhe von 1 Franken an.

(i) Vergleichen Sie die erwarteten Kosten pro heruntergeladenem Dokument für das Produkt A mit der Möglichkeit, auf eine Antivirensoftware gänzlich zu verzichten.

(ii) Ab wieviel heruntergeladenen Dokumenten ist es sinnvoll, das Produkt B statt A zu wählen?
*Hinweis: Falls Sie die Teilaufgabe **2 c)** (i) nicht gelöst haben, dann nehmen Sie für die erwarteten Kosten aus Aufgabe **2 c)** (i) den Wert 0.05. Weiter multiplizieren Sie vor dem Ausrechnen mit einer geeigneten Zehnerpotenz.*

Siehe nächstes Blatt!

*Bei **d)** müssen die Zahlenwerte in den Resultaten nicht vereinfacht werden. In den Ergebnissen sollen aber nur Zahlen und keine Variablen vorkommen.*

d) Das Unternehmen entscheidet sich nach reiflicher Überlegung für Produkt A. Um das Produkt noch einmal zu testen, wird ein Virenscan von 1000 Dokumenten veranlasst. Der Anteil der Viren beträgt weiterhin 1%, wobei 10% davon in Kategorie W fallen. Wie gross ist die Wahrscheinlichkeit, dass die Antivirensoftware weniger als 10 Warnungen abgibt?

3. (10 Punkte)

Die Firma Apfel GmbH produziert PC's und verkauft ihre Artikel online. Herr Egger und Frau Steiner arbeiten beide bei der Apfel GmbH und befassen sich mit der Kalibrierung der Sendezeit der Artikel. Sei X die Zeit (in 10^{-1} Wochen), die von der Sendung eines defekten Artikels vergeht, bis ihn ein Kunde zurücksendet. Wir nehmen an, dass die minimale Rücksendezeit $\mu = 3$ beträgt und bezeichnen mit Y die Überschreitung dieser minimalen Zeit $Y := X - \mu$. Die Verteilung von Y hat die Dichte

$$f_Y(t) = \begin{cases} \alpha t^{\alpha-1} e^{-t^\alpha} & t \geq 0, \\ 0 & t < 0. \end{cases} \tag{1}$$

Herr Egger ist an der Kalibrierung dieser Modellverteilung beteiligt.

a) Berechne die Verteilungsfunktion $F_Y(t)$ von Y.

b) Sei $\lambda(t)$ die Häufigkeit (nach der minimalen Rücksendezeit $\mu = 3$), mit der ein Artikel zurückkommt. Mathematisch kann man dies als

$$\lambda(t) = \frac{f_Y(t)}{1 - F_Y(t)}$$

definieren. Herr Egger hat herausgefunden, dass sich die Wahrscheinlichkeit des Zurücksendens erhöht, je mehr Zeit vergeht. Berechne α so, dass $\lambda(t) = 2t$ für alle t ist.

*Für die Teilaufgaben **c) - e)** setzen wir $\alpha = 2$.*

c) Wie gross ist die Wahrscheinlichkeit $P(X > 5)$?

d) Die Firma Apfel GmbH weigert sich, Rücksendungen nach einer gewissen Zeit t zu akzeptieren. Für welche Werte t werden nur 10% aller Rücksendungen nicht angenommen?
Hinweis: Sie müssen t nicht explizit ausrechnen, jedoch müssen Sie das Resultat für t soweit als möglich vereinfachen.

e) Nach der letzten Studie wurde beschlossen, dass Rücksendungen nach einer Woche nicht mehr möglich sind. Wie lautet die neue Verteilungsfunktion von Y, wenn nach einer Woche keine Rücksendungen mehr erlaubt sind.
Hinweis: Da die Zeiteinheit eine $\frac{1}{10}$ Woche ist, entspricht eine Woche $t = 10$. Deshalb gilt $\{X \leq 10\} = \{Y \leq 7\} \ldots$

Siehe nächstes Blatt!

Der Ansatz von Herrn Egger ist eigentlich eine Dichte der Form

$$f_Y(t) = \begin{cases} \frac{\alpha}{\beta^\alpha} t^{\alpha-1} e^{-\left(\frac{t}{\beta}\right)^\alpha} & t \geq 0, \\ 0 & t < 0. \end{cases} \tag{2}$$

Aus Erfahrung weiss Herr Egger, dass β den Wert 1 annimmt und rechnet deshalb in **a) - e)** mit der vereinfachten Form (1). Gleichzeitig mit Herrn Egger modelliert aber Frau Steiner die gemeinsame Verteilung des Gewichts des Artikels und der Zeit, bevor er zurückkommt. Die Zufallsvariable Z bezeichnet die Zeit, bevor der Artikel zurückkommt und die Zufallsvariable U das Gewicht des Artikels. Die gemeinsame Dichte der Zufallsvariablen Z und U hat die Form

$$f_{Z,U}(z,u) = \frac{z}{10\sigma^2} e^{-\frac{z^2}{2\sigma^2}}\ I_{[0,+\infty)\times[2,12]}(z,u),$$

wobei σ^2 eine positive Konstante ist und $I_A(z,u)$ die Indikatorfunktion, d.h.
$I_A(z,u) = 1$, falls $(z,u) \in A$, sonst $I_A(z,u) = 0$.

f) i) Berechnen Sie die Randdichte $f_Z(z)$ von Z.
ii) Für welche Werte α and β hat die Dichte $f_Z(z)$ die Form (2)?

g) Berechnen Sie die Randdichte $f_U(u)$ von U.

4. (10 Punkte)

Ein Systemadministrator erweitert sein Rechenzentrum. Händler A schlägt ihm ein modernes Computersystem vor, das eine höhere Rechengeschwindigkeit bringen soll. Um diese Aussage zu prüfen, bestimmt der Administrator die Laufzeiten eines standardisierten Programms sowohl auf den alten als auch auf den neuen Rechnern.

Die Laufzeit des Programms auf den im Rechenzentrum vorhandenen Computern bezeichnen wir mit der Zufallsvariable X, jene auf den Systemen des Händlervorschlags mit Y. Auf dem im Rechenzentrum vorhandenen Computersystem ergibt sich bei $n_X = 21$ Durchläufen eine mittlere Laufzeit von $\bar{x}_{n_X} = 114[\mathrm{s}]$ und eine Varianz von $s_X^2 = 280[\mathrm{s}^2]$. Auf dem System, das Händler A vorschlägt, ergibt sich bei $n_Y = 21$ Durchläufen $\bar{y}_{n_Y} = 104[\mathrm{s}]$ und $s_Y^2 = 245[\mathrm{s}^2]$. Hierbei sind die gemessenen Laufzeiten in Sekunden s angegeben.

a) Handelt es sich um eine gepaarte oder eine ungepaarte Stichprobe? Begründen Sie Ihre Antwort.

b) Ist ein einseitiger Test besser geeignet als ein zweiseitiger? Formulieren Sie die Null- und die Alternativ-Hypothese.

c) Es sei vorausgesetzt, dass die Varianz der Laufzeiten auf beiden Systemen gleich $\sigma^2 = 262.5[\mathrm{s}^2]$ ist. Setzen Sie zusätzlich voraus, dass die Beobachtungen normalverteilt sind.

i) Geben Sie eine geeignete Teststatistik T an. Welche Verteilung hat T?

ii) Geben Sie den Verwerfungsbereich des Tests an. Wird die Hypothese auf einem Signifikanzniveau von 5% verworfen?

d) Im letzten Unterpunkt hatten wir eine bekannte Varianz vorausgesetzt. Ein t-Test ermöglicht den Fall unbekannter, aber identischer Varianzen, wobei wir weiterhin die Normalverteilungsannahme treffen.

i) Geben Sie die Teststatistik des t-Tests an. Wie lautet der Parameter der verwendeten t-Verteilung?

ii) Geben Sie den Verwerfungsbereich des Tests an. Wie entscheidet der Test auf einem Signifikanzniveau von 5%?

Viel Erfolg!

Siehe nächstes Blatt!

Antwortblatt zur Aufgabe 1

Bitte benützen Sie dieses Blatt um die Aufgabe 1 zu lösen, indem Sie an der entsprechenden Stelle ein Kreuz machen. Falls es in einer Zeile kein oder mehr als ein Kreuz hat, wird dies als “keine Antwort” (k. A.) gewertet.

Nur hier ausfüllen

	Antw. (i)	Antw. (ii)	Antw. (iii)	k. A.
1a)				
1b)				
1c)				
1d)				
1e)				
1f)				
1g)				
1h)				
1i)				
1j)				

Bitte nicht ausfüllen

richtig	falsch	k. A.

Das folgende bitte nicht ausfüllen!

Aufgabe 1	Korr.	Kontr.
richtig		
falsch		
k. A.		
Punkte		