Basisprüfung D-INFK

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) Seien A und B zwei Ereignisse mit $P[A] = 1/2$, $P[B] = 1/3$ und $P[A|B] = 1/4$. Wie gross ist dann $P[B|A]$?

1) 1/24 2) 1/6 3) 3/4

b) Sei A ein Ereignis mit $P[A] > 0$. Es gilt:

1) $P[B|A] \geq P[B]$ für jedes Ereignis B,
2) $P[B|A] \leq P[B]$ für jedes Ereignis B,
3) weder 1) noch 2) ist richtig.

c) Sei X eine Zufallsvariable mit Varianz $\mathrm{Var}(X)$. Die Aussage $\mathrm{Var}(X) = 0$ gilt

1) genau dann, wenn $P[X = 0] = 1$,
2) genau dann, wenn ein $r \in \mathbb{R}$ existiert mit $P[X = r] = 1$,
3) nie.

d) Seien X, Y und Z Zufallsvariablen mit endlichen Varianzen. Es bezeichne $\mathrm{Cov}(X, Z)$ die Kovarianz von X und Z. Die Aussage $\mathrm{Cov}(X + Y, Z) = \mathrm{Cov}(X, Z) + \mathrm{Cov}(Y, Z)$ gilt

1) genau dann, wenn X und Y unkorreliert sind,
2) genau dann, wenn X, Y und Z unkorreliert sind,
3) immer.

e) Sei X eine normalverteilte Zufallsvariable mit Erwartungswert 1 und Varianz 1. Dann gilt

1) $E\big[X^2\big] = 1$ 2) $E\big[X^2\big] = 2$ 3) $E\big[X^2\big] = 4$

f) Sei X eine normalverteilte Zufallsvariable mit Erwartungswert 0 und Varianz 1. Weiter sei U eine von X unabhängige Zufallsvariable mit $P[U = 1] = P[U = -1] = 1/2$. Dann ist $Z = UX$

1) normalverteilt,
2) uniformverteilt,
3) exponentialverteilt.

g) Sei $X_1, X_2, \ldots$ eine Folge von unabhängigen Zufallsvariablen, die alle den gleichen Erwartungswert $E[X_i] = \mu$ und die gleiche Varianz $\mathrm{Var}(X_i) = \sigma^2$ haben. Dann gilt:

1) $\frac{1}{\sqrt{n}}\Sigma_{i=1}^n X_i$ konvergiert für $n \to \infty$ gegen μ fast sicher,
2) $\frac{1}{n}\Sigma_{i=1}^n X_i$ konvergiert für $n \to \infty$ gegen μ fast sicher,
3) die Verteilung von $\frac{1}{n}\Sigma_{i=1}^n X_i$ konvergiert für $n \to \infty$ gegen eine Normalverteilung mit Erwartungswert μ und Varianz σ^2.

Bitte wenden!

h) Seien $\vartheta \in [0,1]$ ein zu schätzender Parameter und $X_1, \ldots, X_n$ eine Stichprobe von stetigen Zufallsvariablen mit gemeinsamer Dichte f_ϑ. Dann gilt:

1) $\int_{\mathbb{R}} \cdots \int_{\mathbb{R}} f_\vartheta(x_1, \ldots, x_n)\, dx_1 \cdots dx_n = 1$, aber nicht immer $\int_{[0,1]} f_\vartheta(x_1, \ldots, x_n)\, d\vartheta = 1$,
2) $\int_{[0,1]} f_\vartheta(x_1, \ldots, x_n)\, d\vartheta = 1$, aber nicht immer $\int_{\mathbb{R}} \cdots \int_{\mathbb{R}} f_\vartheta(x_1, \ldots, x_n)\, dx_1 \cdots dx_n = 1$,
3) sowohl $\int_{[0,1]} f_\vartheta(x_1, \ldots, x_n)\, d\vartheta = 1$ als auch $\int_{\mathbb{R}} \cdots \int_{\mathbb{R}} f_\vartheta(x_1, \ldots, x_n)\, dx_1 \cdots dx_n = 1$.

i) Bei einem statistischen Test bedeutet ein Signifikanzniveau $\alpha = 0.05$, dass

1) unter der Alternativhypothese die Teststatistik mit einer Wahrscheinlichkeit von höchstens 5 % im Verwerfungsbereich liegt,
2) unter der Alternativhypothese die Teststatistik mit einer Wahrscheinlichkeit von mindestens 95 % im Verwerfungsbereich liegt,
3) unter der Nullhypothese die Teststatistik mit einer Wahrscheinlichkeit von höchstens 5 % im Verwerfungsbereich liegt.

j) Nehmen Sie an, dass Student Markus bei dieser Multiple-Choice-Aufgabe rein zufällig und unabhängig bei jeder der zehn Teilaufgaben eine der drei möglichen Antworten angekreuzt hat. Wie gross ist die Wahrscheinlichkeit, dass Markus bei mindestens neun direkt aufeinanderfolgenden Teilaufgaben die richtige Antwort angekreuzt hat?

1) $\frac{5}{3^{10}}$ 2) $\frac{1}{3^9}$ 3) $\frac{7}{3^9}$

Siehe nächstes Blatt!

2. (12 Punkte)

Die Unternehmung Mail AG möchte einen neuen Spam-Filter zur Kontrolle der eingehenden Mails anschaffen. Zur Verfügung stehen drei mögliche Spam-Filter A, B und C. Die folgende Tabelle gibt die Wahrscheinlichkeiten an, dass eine eingehende Nachricht als 'Spam' klassiert wird.

	A	B	C
Spam	90 %*	95 %	98 %
Nicht-Spam	1 %**	2 %	3 %

* Dies bedeutet, dass 90 % der Mails, die Spam sind, vom Filter A (korrekterweise) als 'Spam' klassiert werden.

** Dies bedeutet, dass 1 % der Mails, die nicht Spam sind, vom Filter A (fälschlicherweise) als 'Spam' klassiert werden.

Nehmen Sie an, dass 90 % aller eingehenden Mails Spam sind.

a) Wie gross ist der Anteil der Spam-Mails an denjenigen Mails, die durch den Filter A als 'Nicht-Spam' klassiert werden?

b) Die Mail AG schätzt, dass ihr pro nicht erkannte Spam-Mail Kosten von 10 Rappen entstehen und pro Nicht-Spam-Mail, die fälschlicherweise als 'Spam' klassiert wird, Kosten von 1 Franken entstehen.

i) Für welchen Filter soll sich die Mail AG unter dieser Annahme entscheiden?

ii) Bei welchem Spam-Anteil an den eingegangenen Mails wären die durchschnittlichen Kosten bei den Filtern A und B gleich hoch?

c) Nehmen Sie für diesen Teil an, dass die Mail AG den Filter A einsetzt. Mitarbeiter Hans schaltet entnervt seinen Computer aus, wenn er das 10. Spam-Mail erhalten hat. Es bezeichne X die Anzahl der Mails (Spam und Nicht-Spam), die Hans erhalten hat, wenn er den Computer ausschaltet. *Bei den folgenden beiden Fragen ist keine Begründung nötig.*

i) Welche Verteilung hat X? Geben Sie auch die Parameter der Verteilung an.

ii) Wie gross ist $E[X]$?

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

d) Da die Filter A und B auf verschiedenen Technologien basieren, können wir annehmen, dass die Erkennung von Spam und Nicht-Spam beim Filter A unabhängig vom Filter B erfolgt. (Genauer heisst das, dass für jede Nachricht die Entscheidung bei den beiden Filtern unabhängig voneinander passiert.) Die Mail AG setzt ein zweistufiges Verfahren ein. Zuerst werden die eingehenden Mails vom Filter A kontrolliert. Wenn sie als 'Nicht-Spam' klassiert werden, werden sie zusätzlich vom Filter B kontrolliert. Nur Mails, die sowohl von A als auch B als 'Nicht-Spam' klassiert wurden, gelangen zu den Mitarbeitenden.

i) Wie gross ist der Spam-Anteil an den Mails, die am Ende als 'Nicht-Spam' klassiert werden?

ii) Wie viele Prozent der Nicht-Spam-Mails werden durch das Verfahren als 'Spam' klassiert?

Bitte wenden!

3. (12 Punkte)

Um an die ETH zu gelangen, fährt Rochus jeden Morgen mit dem Regionalbus zum Bahnhof und nimmt dort den Zug nach Zürich. Von der fahrplanmässigen Ankunft des Busses am Bahnhof bis zur Abfahrt des Zuges vergehen 4 Minuten. Wir nehmen an, dass Rochus keine Zeit fürs Umsteigen braucht. Der Zug fährt jeweils rechtzeitig ab; auf dem Weg kann es jedoch zu Verspätungen kommen. Der nächste Zug fährt erst wieder eine Stunde später. Um rechtzeitig an der ETH zu sein, darf der Zug in Zürich maximal 10 Minuten Verspätung haben. Dank seiner Pendler-Erfahrung weiss Rochus, dass die Verspätung B (in Minuten) des Busses gemäss der *Dichtefunktion*

$$f_B(x) = \begin{cases} c\,(4+x)\,, & x \in (-4, 0], \\ c\frac{1}{2}\,(8-x)\,, & x \in (0, 8), \\ 0, & \text{sonst} \end{cases}$$

verteilt ist. Ein negativer Wert von B bedeutet, dass der Bus vor der fahrplanmässigen Ankunft eintrifft. Die Verspätung Z (in Minuten) des Zuges hat folgende *Verteilungsfunktion*

$$F_Z(x) = \begin{cases} 0, & x < -4, \\ \frac{1}{10}\,(x+4)\,, & -4 \leq x < 4, \\ \frac{1}{60}\,(x+44)\,, & 4 \leq x \leq 16, \\ 1, & \text{sonst.} \end{cases}$$

Die Verspätung des Busses und diejenige des Zuges sind unabhängig.

a) Wie muss c gewählt werden, damit f_B eine Wahrscheinlichkeitsdichte darstellt?

b) Bestimmen Sie $E[B]$.

c) Wie gross ist die Wahrscheinlichkeit, dass Rochus den Zug verpasst?
Falls diese Teilaufgabe nicht gelöst wird, kann man mit dem Wert 1/3 fortfahren.

d) Wie gross ist die Wahrscheinlichkeit, dass Rochus zu spät an der ETH ankommt?

e) Bestimmen Sie die Wahrscheinlichkeit, dass mindestens eines der beiden Verkehrsmittel mehr als 4 Minuten Verspätung hat (unabhängig davon, ob Rochus im Zug sitzt).

Rochus hat Pech. Am Tag seiner Prüfung schneit es und sein Bus erreicht den Bahnhof erst nach Abfahrt des Zuges. Um doch noch rechtzeitig in Zürich zu sein, läuft er zur nächsten vielbefahrenen ($n = 50$ Autos pro 10 Minuten) Strasse und versucht per Autostopp nach Zürich zu gelangen. Die Bereitschaft eines Autolenkers, einen Autostopper mitzunehmen, beträgt $p = 0.03$ und ist unabhängig von den anderen Autolenkern.

f) i) Wie gross ist die Wahrscheinlichkeit, dass in 10 Minuten genau 3 Autolenker vorbeifahren, die bereit sind, einen Autostopper mitzunehmen.

ii) Mit Hilfe von welcher Verteilung kann diese Wahrscheinlichkeit geeignet approximiert werden? (Keine Normalverteilung!) Geben Sie den Parameter dieser Verteilung an.

Die Resultate müssen nicht numerisch ausgerechnet werden.

4. (12 Punkte)

Beim Skispringen hängen die gesprungenen Weiten eines Springers von vielen äusseren Faktoren ab und können daher als Realisierungen von unabhängigen identisch normalverteilten Zufallsvariablen angenommen werden. Der Erwartungswert dieser Zufallsvariablen ist von Schanze zu Schanze verschieden, die Varianz der gesprungenen Weiten von Skispringer Simon ist immer $4m^2$. Simon hat auf seiner Heimschanze, der Säntisschanze in Unterwasser, bei seinen letzten neun Trainingssprüngen die folgenden Weiten erzielt:

i	1	2	3	4	5	6	7	8	9
x_i	65.2	62.7	66.3	66.6	63.7	68.5	68.4	65.9	66.7

Kennzahlen: $\bar{x} = 66$, $s_x \approx 1.93$, $s_x^2 \approx 3.72$

Stolz verkündet er hierauf einem Zeitungsreporter, dass der Erwartungswert seiner Sprünge bei mehr als $65m$ liege.

a) Versuchen Sie Simons Aussage zu überprüfen, indem Sie anhand der Weiten der Trainingssprünge einen geeigneten statistischen Test auf dem 5%-Niveau durchführen. Interpretieren Sie das Testergebnis.

b) Simons Trainer ist davon überzeugt, dass der Erwartungswert von Simons Sprüngen bei $66m$ liege. Berechnen Sie die Macht des Tests aus Aufgabenteil **a)** unter dieser Alternative.
*Falls Sie Aufgabenteil **a)** nicht gelöst haben, dürfen Sie den kritischen Bereich $\tilde{K} = (-\infty, -1.96]$ für die Teststatistik T verwenden.*

An den drei Trainingstagen auf der Grossschanze in Oberstdorf ist Simon etwas aufgeregt und kann sich daher nur die Anzahl der Sprünge merken, die er an jedem der Trainingstage benötigt, um das erste Mal über $120m$ zu springen. Über den Erwartungswert μ seiner Sprünge auf dieser Schanze ist er sich im Unklaren. Die Zufallsvariablen N_1, N_2 und N_3 bezeichnen die Anzahl der Sprünge am ersten, zweiten und dritten Trainingstag, die Simon benötigt, um das erste Mal über $120m$ zu springen.

c) Geben Sie die Verteilung von N_1, N_2 und N_3 und den dazugehörigen Parameter p in Abhängigkeit von μ mittels der kumulativen Verteilungsfunktion Φ der Standard-Normalverteilung an.

d) Berechnen Sie den Maximum-Likelihood-Schätzer $\widehat{p}_{MLE}$ von p.

Viel Erfolg!