

Probability and Statistics

FS 2018

Prüfung

13.08.2018

Dauer: 180 Minuten

Name: _____

Legi-Nummer: _____

- Diese Prüfung enthält 12 Seiten (zusammen mit dem Deckblatt) und 10 Aufgaben. Das Formelblatt wird separat verteilt.
- Begründen Sie ihre Lösungen sorgsam. Ergebnisse ohne Rechenwege und Begründungen geben KEINE Punkte.
- Beginnen Sie jede Aufgabe auf einem neuen Blatt.

Punktetabelle (wird von den Korrektoren beschriftet)

Aufgabe	Punkte	erreichte Punktzahl
1	10	
2	10	
3	10	
4	10	
5	10	
6	10	
7	10	
8	10	
9	10	
10 (*)	10	
Gesamt	90	

Die maximal erreichbare Punktzahl ist 90. Aufgabe 10 ist eine **Bonusaufgabe**.

Viel Erfolg

1. (10 Punkte) Die Zufallsvariable X beschreibt die tägliche Arbeitszeit eines Ingenieurs in Stunden und hat folgende Dichte:

$$f(x) = \begin{cases} c(x-7)^\alpha & : 7 \leq x \leq 10, \\ 0 & : \text{sonst,} \end{cases}$$

wobei $\alpha > -1$.

- a) Bestimme die Konstante c in Abhängigkeit von α so, dass $f(x)$ eine Wahrscheinlichkeitsdichte definiert.
- b) Wähle $c = \frac{1}{9}$ und $\alpha = 2$. Berechne die Verteilungsfunktion von X .
- c) Wähle $c = \frac{1}{9}$ und $\alpha = 2$. Berechne die Wahrscheinlichkeit, dass X einen Wert zwischen 8 und 9 Stunden einnimmt.

2. (10 Punkte) Student Christoph fährt jeden Tag mit dem Velo an die ETH. Dabei kommt er an genau zwei Kreuzungen mit Ampeln vorbei. Aus Erfahrung weiss Christoph, dass er an der ersten Ampel mit Wahrscheinlichkeit 25% warten muss. Auf Grund der Ampelsteuerung und Christophs Fahrgeschwindigkeit hängt die Wahrscheinlichkeit, dass Christoph an der zweiten Ampel warten muss davon ab, ob er an der ersten Ampel warten musste oder nicht. Falls er an der ersten Ampel nicht warten musste, ist die Wahrscheinlichkeit an der zweiten auch nicht warten zu müssen $\frac{2}{3}$. Hatte er an der ersten Ampel hingegen "rot", muss er mit Wahrscheinlichkeit $\frac{1}{7}$ auch an der zweiten Ampel warten. Christoph kommt genau dann zu spät zur Vorlesung, wenn er wenigstens an einer Ampel warten muss.

- a) Mit welcher Wahrscheinlichkeit muss Christoph an genau einer Ampel warten?
- b) Mit welcher Wahrscheinlichkeit kommt Christoph zu spät zur Vorlesung?
- c) Mit welcher Wahrscheinlichkeit hatte Christoph an der ersten Ampel "grün", gegeben, dass er an der zweiten Ampel nicht warten musste?

Hinweis: Falls Sie Teil b) nicht gelöst haben, benutzen Sie im Folgenden für die Wahrscheinlichkeit, dass Christoph zu spät zur Vorlesung kommt, den Wert $\frac{1}{2}$.

- d) Mit welcher Wahrscheinlichkeit musste Christoph an der ersten Ampel warten, gegeben, dass er zu spät zur Vorlesung kam? Sind die Ereignisse "Christoph kommt zu spät zur Vorlesung" und "Christoph muss an der ersten Ampel warten" unabhängig?
- e) Christoph fährt im Februar 10-mal an die ETH. Nehmen Sie an, die Ereignisse, dass er an einem der 10 Tage zu spät zur Vorlesung kommt, seien unabhängig. Wie gross ist die Wahrscheinlichkeit, dass er im Februar mindestens zweimal zu spät zur Vorlesung kommt?

Hinweis: Potenzen wie beispielsweise 2^{10} müssen nicht weiter vereinfacht werden.

3. (10 Punkte) Sei X geometrisch verteilt mit Parameter $p \in (0, 1)$, das heisst $\mathbb{P}(X = k) = (1 - p)^k p$, wobei $k \in \mathbb{N}_0 = \{0, 1, 2, \dots\}$. Die Wahrscheinlichkeitserzeugende Funktion ist definiert als

$$G_X(z) = \mathbb{E}[z^X] = \sum_{k=0}^{\infty} z^k \mathbb{P}(X = k).$$

- Berechne $G_X(z)$. Für welche $z \in \mathbb{R}$ ist $G_X(z)$ definiert?
- Was ist $G_X(0)$ und $G_X(1)$?
- Berechne $G'_X(z)|_{z=1}$. Was ist das?

4. (10 Punkte) a) Berechne mit Hilfe des zentralen Grenzwertsatzes

$$\lim_{n \rightarrow \infty} e^{-n} \sum_{k=0}^n \frac{n^k}{k!}.$$

Hinweis: Betrachte i.i.d. Zufallsvariablen X_1, X_2, \dots mit $X_i \sim \text{POIS}(1)$ und setze $S_n = X_1 + \dots + X_n$.

5. (10 Punkte) Sei X eine Zufallsvariable mit momenterzeugender Funktion $M(t) = \mathbb{E}[e^{tX}]$.

a) Zeigen Sie, dass

$$\mathbb{P}(X \geq a) \leq e^{-ta}M(t), \text{ für } t > 0$$

und

$$\mathbb{P}(X \leq a) \leq e^{-ta}M(t), \text{ für } t < 0$$

gilt.

b) Betrachten Sie eine normalverteilte Zufallsvariable $Z \sim \mathcal{N}(0, 1)$. Zeigen Sie, dass

$$\mathbb{P}(Z \geq a) \leq e^{-a^2/2}, \text{ für } a > 0$$

und

$$\mathbb{P}(Z \leq a) \leq e^{-a^2/2}, \text{ für } a < 0$$

gilt.

6. (10 Punkte) Sei $(\Omega, \mathcal{A}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und $(Z_n)_{n \in \mathbb{N}}$ eine Folge von Zufallsvariablen.

a) Zeige: Falls $Z_n \xrightarrow{\mathbb{P}} c \in \mathbb{R}$, dann gilt für alle beschränkten und stetigen Funktionen $f : \mathbb{R} \rightarrow \mathbb{R}$, dass

$$\mathbb{E}(f(Z_n)) \rightarrow f(c).$$

b) Zeige: Falls $Z_n \rightarrow c \in \mathbb{R}$ in Verteilung konvergiert, dann gilt $Z_n \xrightarrow{\mathbb{P}} c$.

7. (10 Punkte) Seien X und Y unabhängige, $\mathcal{N}(0, 1)$ - verteilte Zufallsvariablen. Wir betrachten die Zufallsvariable Z , welche definiert ist als

$$Z := \operatorname{sgn}(Y) \cdot X = \begin{cases} X & \text{für } Y > 0, \\ -X & \text{für } Y \leq 0. \end{cases}$$

- a) Bestimmen Sie die Verteilung von Z .
- b) Berechnen Sie die Korrelation von X und Z .
- c) Bestimmen Sie die Wahrscheinlichkeit $\mathbb{P}(X + Z = 0)$.
- d) Sind X und Z unabhängig? Begründen Sie Ihre Antwort durch ein präzises mathematisches Argument.

8. (10 Punkte) Wir betrachten i.i.d. Zufallsvariablen X_1, X_2, \dots , wobei X_i die uniforme Verteilung auf $[0, \vartheta]$ mit $\vartheta > 0$ besitze.
- a) Bestimme den Maximum-Likelihood Schätzer T_n für ϑ basierend auf den Beobachtungen (X_1, \dots, X_n) . Warum ist dieser nicht erwartungstreu? (Begründe ohne zu rechnen.)
- Vorsicht:** Die Likelihood-Funktion ist nicht differenzierbar.
- b) Berechne die Dichtefunktion für T_n . Benutze dies, um einen Faktor c_n zu bestimmen, so dass $c_n T_n$ ein erwartungstreuer Schätzer für ϑ ist.
- c) Wie verhält sich die Varianz von $c_n T_n$ für $n \rightarrow \infty$?
- d) Schätze nun, basierend auf den Beobachtungen (X_1, \dots, X_n) , ϑ mit der Momenten-Methode, d.h. finde denjenigen Parameter ϑ , für welchen der Erwartungswert gleich dem empirischen Mittel ist.

9. (10 Punkte) Einer Gruppe von 20 Tennisspielern mittleren Niveaus werden je zwei Tennisschläger zum Testen ausgehändigt. Einer der Schläger ist jeweils mit einer Nylon-Saite bespannt, der andere mit einer synthetischen Darm-Saite. Nach einigen Wochen Testzeit wird jeder Spieler gefragt, ob er Nylon- oder Darm-Saiten bevorzugt. Es sei p der Anteil aller Tennisspieler mittleren Niveaus, die Darm-Saiten bevorzugen und X sei die Anzahl der Spieler unter den 20 Testspielern, die Darm-Saiten bevorzugen. Da Darm-Saiten teurer sind als Nylon-Saiten, betrachten wir die Nullhypothese, dass höchstens die Hälfte der Spieler Darm-Saiten bevorzugt. Wir vereinfachen dies zu $H_0 : p = 0.5$ und werden H_0 nur ablehnen, falls der Versuchsausgang eindeutig Darm-Saiten bevorzugt.
- Welcher der beiden Verwerfungsbereiche $\{15, 16, 17, 18, 19, 20\}$ oder $\{0, 1, 2, 3, 4, 5\}$ ist eher für den Test geeignet?
 - Was ist die Wahrscheinlichkeit eines Fehlers erster Art für den gewählten Bereich in a)? Erhält man mit diesem Bereich einen Test zum Level $\alpha = 0.05$? Ist dies der beste Test zum Niveau α ?
 - Berechnen Sie die Wahrscheinlichkeit eines Fehlers zweiter Art für den Verwerfungsbereich, den Sie in a) gewählt haben, falls $p = p_1 = 0.6$ und falls $p = p_2 = 0.8$.
 - Berechnen Sie den P-Wert, falls $X = 13$. Wird H_0 auf einem Signifikanzniveau von $\alpha = 0.10$ abgelehnt?

Unten finden Sie eine Tabelle der *kumulierten* Binomialverteilung für $n = 20$; Für Werte $p \geq 0.50$ (über dem grau hinterlegten Bereich) ist abzulesen: $\mathbb{P}[X \leq l] = 1 -$ abgelesener Wert.

n	k	p	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50	l	n
20	0		0.1216	0.0388	0.0115	0.0032	0.0008	0.0001	0.0000	0.0000	0.0000	19	20
	1		0.3917	0.1756	0.0692	0.0243	0.0076	0.0021	0.0005	0.0001	0.0000	18	
	2		0.6769	0.4049	0.2061	0.0913	0.0355	0.0121	0.0036	0.0009	0.0002	17	
	3		0.8670	0.6477	0.4114	0.2252	0.1071	0.0444	0.0160	0.0049	0.0013	16	
	4		0.9568	0.8298	0.6296	0.4148	0.2375	0.1182	0.0510	0.0189	0.0059	15	
	5		0.9887	0.9327	0.8042	0.6172	0.4164	0.2454	0.1256	0.0553	0.0207	14	
	6		0.9976	0.9781	0.9133	0.7858	0.6080	0.4166	0.2500	0.1299	0.0577	13	
	7		0.9996	0.9941	0.9679	0.8982	0.7723	0.6010	0.4159	0.2520	0.1316	12	
	8		0.9999	0.9987	0.9900	0.9591	0.8867	0.7624	0.5956	0.4143	0.2517	11	
	9		1.0000	0.9998	0.9974	0.9861	0.9520	0.8782	0.7553	0.5914	0.4119	10	
	10		1.0000	1.0000	0.9994	0.9961	0.9829	0.9468	0.8725	0.7507	0.5881	9	
	11		1.0000	1.0000	0.9999	0.9991	0.9949	0.9804	0.9435	0.8692	0.7483	8	
	12		1.0000	1.0000	1.0000	0.9998	0.9987	0.9940	0.9790	0.9420	0.8684	7	
	13		1.0000	1.0000	1.0000	1.0000	0.9997	0.9985	0.9935	0.9786	0.9423	6	
	14		1.0000	1.0000	1.0000	1.0000	1.0000	0.9997	0.9984	0.9936	0.9793	5	
	15		1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9997	0.9985	0.9941	4	
	16		1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9997	0.9987	3	
	17		1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	0.9998	2	
	18		1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1	
n	k	p	0.90	0.85	0.80	0.75	0.37	0.65	0.60	0.55	0.50	l	n

10. (10 Punkte) (*)

- a) Sei $(A_n)_{n \geq 1}$ eine Folge von Ereignissen in einem Wahrscheinlichkeitsraum. Nimm an, dass gilt

$$\sum_{n=1}^{\infty} \mathbb{P}(A_n) < \infty.$$

Beweise, dass

$$\mathbb{P}\left(\bigcap_{k \in \mathbb{N}} \bigcup_{n \geq k} A_n\right) = 0.$$

- b) Es bezeichne \mathcal{A} eine nicht leere Menge von Teilmengen von $\Omega \neq \emptyset$. Zeige, dass

$$\sigma(\mathcal{A}) := \bigcap_{\substack{\mathcal{B} \supseteq \mathcal{A} \\ \mathcal{B} \text{ ist } \sigma\text{-Algebra}}} \mathcal{B}$$

die kleinste σ -Algebra ist, die \mathcal{A} enthält.

Bemerkung: Diese σ -Algebra heisst die von \mathcal{A} erzeugte σ -Algebra. “Kleinste σ -Algebra” bedeutet hier, dass jede σ -Algebra, die die Mengen von \mathcal{A} enthält, auch die von $\sigma(\mathcal{A})$ enthalten muss.