Prof. P. Cheridito ETH Zürich 29. August 2022

D-MATL D-MAVT Neur. Syst. Comp.

Stochastik (401-0603-00L)

Prüfung

1. (9 Punkte)

Carolin und Jan möchten sich zum Tennisspielen verabreden. Dafür möchten sie gerne wissen, ob es am nächsten Tag regnen wird. In ihrer Region liegt die für diese Jahreszeit typische Regenwahrscheinlichkeit bei 10%. Für genauere Vorhersagen benützt jeder der beiden eine andere Wetterapp: Carolin benützt App A und Jan benützt App B. Wir bezeichnen mit

- R das Ereignis, dass es am nächsten Tag regnet, mit
- KR das Ereignis, dass es am nächsten Tag keinen Regen gibt, mit
- AR das Ereignis, dass App A Regen vorhergesagt hat, und mit
- BR das Ereignis, dass App B Regen vorhergesagt hat.

Die Wahrscheinlichkeit, dass App A Regen vorhergesagt hat, obwohl es keinen Regen gibt, liegt bei 20%, d.h. $\mathbb{P}(\mathrm{AR} \mid \mathrm{KR}) = 0.2$.

Analog ist $\mathbb{P}(\mathrm{BR} \mid \mathrm{KR}) = 0.3$.

Ausserdem wissen wir, dass

$$\mathbb{P}(\mathrm{BR} \mid \mathrm{KR} \cap \mathrm{AR}) = 0.7, \text{ und } \mathbb{P}(\mathrm{AR} \cap \mathrm{BR} \mid \mathrm{R}) = 0.8,$$

d.h. bedingt darauf, dass es keinen Regen gibt und dass App A Regen vorhergesagt hat, ist die Wahrscheinlichkeit, dass auch App B fälschlicherweise Regen vorhergesagt hat 70%, und die Wahrscheinlichkeit, dass App A und App B korrekterweise Regen vorhersagen liegt bei 80%.

a) **(2 Punkte)** Bestimmen Sie die Wahrscheinlichkeit, dass sowohl App A als auch App B Regen vorhergesagt haben, bedingt darauf, dass es keinen Regen gibt.

b) **(1 Punkt)** Weissen Sie nach, ob, bedingt auf KR, die Ereignisse AR und BR unabhängig sind oder nicht.

c) **(2 Punkte)** Bestimmen Sie die Wahrscheinlichkeit, dass es am nächsten Tag regnet, gegeben, dass sowohl App A als auch App B Regen vorhergesagt haben.

Carolin und Jan entschliessen sich dazu am nächsten Tag Tennis spielen zu gehen, wenn zumindest eine der Apps keinen Regen vorhersagt.

Begründen Sie Ihre Antworten zu den folgenden zwei Fragen.

d) **(2 Punkte)** Wie gross ist die Wahrscheinlichkeit, dass die beiden Tennis spielen gehen?

e) **(2 Punkte)** Wie gross ist die Wahrscheinlichkeit, dass es am nächsten Tag regnet, gegeben dass Carolin und Jan Tennis spielen gehen?

2. (11 Punkte) Beurteilen Sie die folgenden 11 Aussagen auf ihre Richtigkeit. **Kreuzen Sie Ihre Antworten auf dem zusätzlich angehängten Antwortblatt an**. Nur das Antwortblatt wird ausgewertet. Pro korrekter Antwort gibt es einen Punkt. Es gibt **keinen** Punkteabzug für falsche Antworten.

a) (1 Punkt) Gegeben sind die Ereignisse A und B. Wenn $\mathbb{P}(A) > 0$ und $A \subseteq B$, dann gilt, dass $\mathbb{P}(B|A) = 1$.

b) (1 Punkt) Gegeben ist eine stetige, reellwertige Zufallsvariable X mit Verteilungsfunktion F auf $\mathbb{R}$. Für ein beliebiges $c \in \mathbb{R}$ gilt dann, dass $\mathbb{P}(X = c) = 0$.

c) (1 Punkt) Für alle $X \sim \mathrm{Bin(m, p)}$ und $Y \sim \mathrm{Bin(n, p)}$ gilt, dass $X + Y \sim \mathrm{Bin(m + n, p)}$.

d) (1 Punkt) Für eine stetige, auf (0, 1) uniform verteilte Zufallsvariable U, sprich $U \sim \mathrm{Uni}(0, 1)$, gilt, dass $\mathrm{Var}(U) = \frac{1}{3}$.

e) (1 Punkt) Gegeben ist die Zufallsvariable X. Die Grösse $\mathbb{E}[(X - c)^2]$ ist minimal, wenn $c = \mathbb{E}[X]$.

f) (1 Punkt) $X \sim \mathcal{N}(\mu, \sigma^2)$ genau dann, wenn $X = \mu + \sigma Z$, wobei $Z \sim \mathcal{N}(0, 1)$.

g) (1 Punkt) Gegeben sind die Zufallsvariablen X und Y. Wenn gilt, dass

$$\mathbb{E}[XY] = \mathbb{E}[X]\mathbb{E}[Y],$$

dann sind X und Y unabhängig.

h) (1 Punkt) Gegeben ist eine Zufallsvariable X mit $\mathbb{E}[|X|] < \infty$. Es gilt, dass

$$\mathbb{E}[|X|] \leq |\mathbb{E}[X]|.$$

i) (1 Punkt) Für zwei reellwertige Zufallsvariablen X und Y gilt, dass

$$\mathrm{Var}(X + Y) + \mathrm{Var}(X - Y) = 2(\mathrm{Var}(X) + \mathrm{Var}(Y)).$$

j) (1 Punkt) Für zwei reellwertige Zufallsvariablen X und Y mit $\mathrm{Var}(X) = \mathrm{Var}(Y)$ gilt, dass $X + Y$ und $X - Y$ korreliert sind.

k) (1 Punkt) Gegeben ist die stetige Zufallsvariable X. Der Median der zugehörigen Verteilung ist jener Wert x, für den gilt, dass $\mathbb{P}(X < x) = \frac{1}{2} = \mathbb{P}(X > x)$.

3. (9 Punkte) Alle θ Minuten schliesst die Bahnschranke an der S-Bahnstation Zürich Binz. Jeden Morgen muss Sophie die Gleise an dieser Bahnschranke überqueren. Wenn sie dort ankommt, ist die Bahnschranke oft geschlossen. Die Dauer X, die Sophie dann warten muss, bis sich die Bahnschranke wieder öffnet, ist stetig uniform-verteilt auf dem Intervall $[0, \theta]$, d.h.

$$X \sim \text{Uni}(0, \theta).$$

Sie sammelt so eine Stichprobe von n Wartedauern $X_1, \ldots, X_n \overset{i.i.d.}{\sim} \text{Uni}(0, \theta)$.

a) (1 Punkt) Bestimmen Sie die Verteilungsfunktion von $\max_{i=1,\ldots,n} X_i$.

Wir möchten nun den unbekannten Parameter $\theta > 0$ schätzen.

b) (1 Punkt) Zeigen Sie, dass der Maximum-Likelihood-Schätzer $\hat{\theta}_{MLE}$ für θ gegeben ist als

$$\hat{\theta}_{MLE} = \max_{i=1,\ldots,n} X_i.$$

c) (2 Punkte) Ist der MLE $\hat{\theta}_{MLE}$ erwartungstreu? Begründen Sie Ihre Antwort.

d) (1 Punkt) Zeigen Sie, dass der Momentenschätzer $\hat{\theta}_{MoM}$ für θ gegeben ist als

$$\hat{\theta}_{MoM} = \frac{2}{n} \sum_{i=1}^{n} X_i.$$

Bewerten Sie folgende Aussagen auf ihre Richtigkeit und begründen Sie Ihre Antworten!

e) (2 Punkte) Die Varianz des Momentenschätzers halbiert sich, wenn Sophie eine doppelt so grosse Stichprobe nimmt.

f) (2 Punkte) Der realisierte Wert des Momentenschätzers $\hat{\theta}_{MoM}$ is immer grösser als der realisierte Wert des Maximum-Likelihood-Schätzers $\hat{\theta}_{MLE}$, unabhängig von den Realisierungen $x_1, \ldots, x_n$.

4. (11 Punkte)

a) Eine Gruppe von Wissenschaftlern möchte das mittlere Alter einer Tierpopulation bestimmen. Für eine Stichprobe bestehend aus 30 Tieren wird das mittlere Alter von 23 Jahren festgestellt. Die Wissenschaftler vermuten, dass der tatsächliche Mittelwert grösser oder kleiner 25 ist. Sie nehmen an, dass die Population approximativ normalverteilt ist, mit bekannter Varianz $\sigma^2 = 18$, und möchten statistisch belegen, ob sich der tatsächliche Mittelwert von 25 unterscheidet.

- **i) (1 Punkt)** Nennen Sie die Modellannahmen.
- **ii) (1 Punkt)** Formulieren Sie geeignete Null- und Alternativhypothesen um die Vermutung der Wissenschaftler statistisch zu testen.
- **iii) (3 Punkte)** Führen Sie einen geeigneten statistischen Test zum Niveau $\alpha = 0.05$ durch: nennen Sie die Teststatistik, bestimmen Sie den realisierten Wert der Teststatistik, berechnen Sie den Verwerfungsbereich und nennen Sie den Testentscheid.

b) In einer Flugzeugbaufirma werden zwei Gruppen von Arbeitern damit beauftragt, die Effektivität einer neuen Technologie zu testen. Gruppe A, bestehend aus 50 Arbeitern, wendet die alte Technologie an, während Gruppe B, bestehend aus 72 Arbeitern, die neue Technologie anwendet.

Um einen speziellen Bauteil herzustellen braucht ein Arbeiter der Gruppe A im Durchschnitt 80 Stunden, mit einer empirischen Standardabweichung von 10 Stunden. Die Mittlere Zeit, die ein Arbeiter der Gruppe B zur Herstellung desselben Bauteils benötigt, liegt bei 75 Stunden, mit einer empirischen Standardabweichung von 8 Stunden.

Wir nehmen an, dass die Zeit, die ein Arbeiter zur Herstellung des Bauteils benötigt, in beiden Gruppen normalverteilt ist mit der selben Varianz σ^2.

- **i) (1 Punkt)** Nennen Sie die Modellannahmen.
- **ii) (1 Punkt)** Formulieren Sie geeignete Null- und Alternativhypothesen um zu testen, ob die Herstellung des Bauteils durch die neue Technologie effizienter wird.
- **iii) (3 Punkte)** Führen Sie einen geeigneten statistischen Test zum Niveau $\alpha = 0.05$ durch: nennen Sie die Teststatistik, bestimmen Sie den realisierten Wert der Teststatistik, berechnen Sie den Verwerfungsbereich und nennen Sie den Testentscheid.
- **iv) (1 Punkt)** Handelt es sich um einen gepaarten oder ungepaarten Vergleich? Begründen Sie Ihre Antwort.

Stochastik - Tabellen

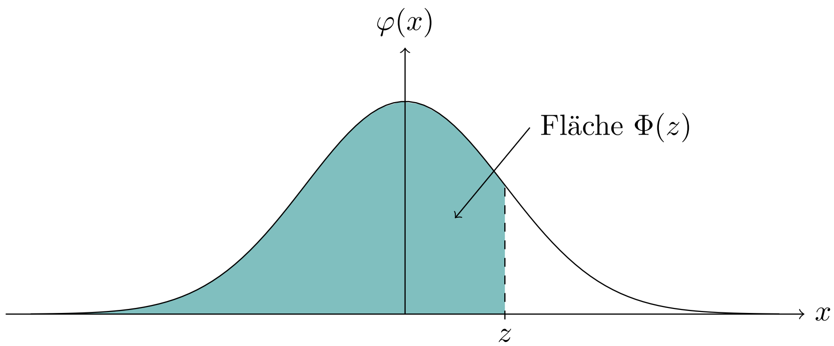
Kritische Grenzen beim Wilcoxon-Test für das 5%-Niveau:

	zweiseitig		einseitig	
n	l	u	l	u
6	0	21	2	19
7	2	26	3	25
8	3	33	5	31
9	5	40	8	37
10	8	47	10	45
11	10	56	13	53
12	13	65	17	61
13	17	74	21	70
14	21	84	25	80
15	25	95	30	90
16	29	107	35	101
17	34	119	41	112
18	40	131	47	124
19	46	144	53	137
20	52	158	60	150
21	58	173	67	164
22	65	188	75	178
23	73	203	83	193
24	81	219	91	209
25	89	236	100	225
26	98	253	110	241
27	107	271	119	259
28	116	290	130	276
29	126	309	140	295
30	137	328	151	314

Für den zweiseitigen Test ist der Verwerfungsbereich gegeben durch $K = \{W \leq l\} \cup \{W \geq u\}$.
Bei einem einseitigen Test verwendet man die entsprechenden Werte in der Spalte “einseitig”.

Tabelle der Standardnormalverteilung:

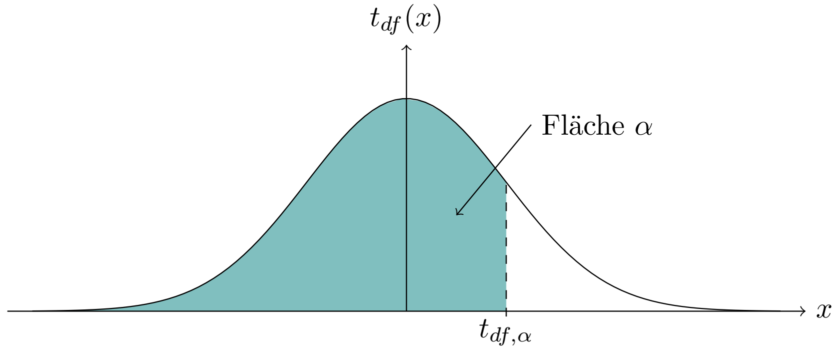
$$\Phi(z) = \mathbb{P}\left(Z \leq z\right), \; Z \sim \mathcal{N}\left(0,1\right)$$



Lesebeispiel Tabelle: $\mathbb{P}\left(Z \leq 1.96\right) = 0.975$

z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998

Quantile der t-Verteilung: (df bezeichnet den Freiheitsgrad)



Lesebeispiel Tabelle: $t_{9,\,0.975} = 2.262$

$df \setminus \alpha$	0.60	0.70	0.80	0.90	0.95	0.975	0.99	0.995
1	0.325	0.727	1.376	3.078	6.314	12.706	31.821	63.657
2	0.289	0.617	1.061	1.886	2.920	4.303	6.965	9.925
3	0.277	0.584	0.978	1.638	2.353	3.182	4.541	5.841
4	0.271	0.569	0.941	1.533	2.132	2.776	3.747	4.604
5	0.267	0.559	0.920	1.476	2.015	2.571	3.365	4.032
6	0.265	0.553	0.906	1.440	1.943	2.447	3.143	3.707
7	0.263	0.549	0.896	1.415	1.895	2.365	2.998	3.499
8	0.262	0.546	0.889	1.397	1.860	2.306	2.896	3.355
9	0.261	0.543	0.883	1.383	1.833	2.262	2.821	3.250
10	0.260	0.542	0.879	1.372	1.812	2.228	2.764	3.169
11	0.260	0.540	0.876	1.363	1.796	2.201	2.718	3.106
12	0.259	0.539	0.873	1.356	1.782	2.179	2.681	3.055
13	0.259	0.538	0.870	1.350	1.771	2.160	2.650	3.012
14	0.258	0.537	0.868	1.345	1.761	2.145	2.624	2.977
15	0.258	0.536	0.866	1.341	1.753	2.131	2.602	2.947
16	0.258	0.535	0.865	1.337	1.746	2.120	2.583	2.921
17	0.257	0.534	0.863	1.333	1.740	2.110	2.567	2.898
18	0.257	0.534	0.862	1.330	1.734	2.101	2.552	2.878
19	0.257	0.533	0.861	1.328	1.729	2.093	2.539	2.861
20	0.257	0.533	0.860	1.325	1.725	2.086	2.528	2.845
21	0.257	0.532	0.859	1.323	1.721	2.080	2.518	2.831
22	0.256	0.532	0.858	1.321	1.717	2.074	2.508	2.819
23	0.256	0.532	0.858	1.319	1.714	2.069	2.500	2.807
24	0.256	0.531	0.857	1.318	1.711	2.064	2.492	2.797
25	0.256	0.531	0.856	1.316	1.708	2.060	2.485	2.787
26	0.256	0.531	0.856	1.315	1.706	2.056	2.479	2.779
27	0.256	0.531	0.855	1.314	1.703	2.052	2.473	2.771
28	0.256	0.530	0.855	1.313	1.701	2.048	2.467	2.763
29	0.256	0.530	0.854	1.311	1.699	2.045	2.462	2.756
30	0.256	0.530	0.854	1.310	1.697	2.042	2.457	2.750
40	0.255	0.529	0.851	1.303	1.684	2.021	2.423	2.704
60	0.254	0.527	0.848	1.296	1.671	2.000	2.390	2.660
90	0.254	0.526	0.846	1.291	1.662	1.987	2.368	2.632
120	0.254	0.526	0.845	1.289	1.658	1.980	2.358	2.617
∞	0.253	0.524	0.842	1.282	1.645	1.960	2.326	2.576