Institut für Stochastik

Prof. Dr. N. Henze · Dipl.-Math. V. Riess

Name:

Vorname:

Matrikelnummer:

Lösungsvorschlag zur Klausur der Vorlesung

Grundlagen der Wahrscheinlichkeitstheorie und Statistik für Studierende der Informatik

Datum: 07. Februar 2013

Dauer: 90 Minuten

Aufgabe 1				Aufgabe 2			Aufgabe 3				
a	b	c	d	a	b	c	a	b	c	d	e
4	2	1.5	2.5	2.5	2.5	2	4	2	1.5	1.5	3

Aufgabe 4			Aufgabe 5			$\sum$	Note
a	b	c	a	b	c		
5	2	2	2.5	3.5	6	50	

ACHTUNG:
Bei dieser Klausur werden **nur** diejenigen Ergebnisse gewertet, die in die vorgesehenen Kästchen eingetragen sind! Eine Begründung bzw. Herleitung der Ergebnisse wird nicht verlangt, es sei denn, dies wird explizit gefordert!
Vereinfachen Sie die Ergebnisse soweit wie möglich. Geben Sie die Ergebnisse so exakt wie möglich an (z.B. als Bruch), runden Sie ggf. auf 3 Nachkommastellen.

Als **Hilfsmittel** sind zugelassen:
Skriptum zur Vorlesung, Vorlesungsmitschrift, Taschenrechner (nicht vernetzbar), Wörterbuch

Hinreichend zum Bestehen der Klausur sind 20 Punkte.

VIEL ERFOLG!

Verteilungsfunktion $\Phi(x)$ der Standard – Normalverteilung $\mathcal{N}(0,1)$

x	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
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	0.7703	0.7734	0.7764	0.7793	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	0.9996	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
3.5	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998
3.6	0.9998	0.9998	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.7	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.8	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.9	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

p-Quantile $\chi^2_{k;p}$ der χ^2 – Verteilung mit k Freiheitsgraden

	p					
k	0.900	0.950	0.975	0.990	0.995	0.999
1	2.71	3.84	5.02	6.63	7.88	10.83
2	4.61	5.99	7.38	9.21	10.60	13.82
3	6.25	7.81	9.35	11.34	12.84	16.27
4	7.78	9.49	11.14	13.28	14.86	18.47
5	9.24	11.07	12.83	15.09	16.75	20.51
6	10.64	12.59	14.45	16.81	18.55	22.46
7	12.02	14.07	16.01	18.48	20.28	24.32
8	13.36	15.51	17.53	20.09	21.95	26.12
9	14.68	16.92	19.02	21.67	23.59	27.88
10	15.99	18.31	20.48	23.21	25.19	29.59
11	17.28	19.68	21.92	24.73	26.76	31.26
12	18.55	21.03	23.34	26.22	28.30	32.91
13	19.81	22.36	24.74	27.69	29.82	34.53
14	21.06	23.68	26.12	29.14	31.32	36.12
15	22.31	25.00	27.49	30.58	32.80	37.70
16	23.54	26.30	28.85	32.00	34.27	39.25
17	24.77	27.59	30.19	33.41	35.72	40.79
18	25.99	28.87	31.53	34.81	37.16	42.31
19	27.20	30.14	32.85	36.19	38.58	43.82
20	28.41	31.41	34.17	37.57	40.00	45.31
21	29.62	32.67	35.48	38.93	41.40	46.80
22	30.81	33.92	36.78	40.29	42.80	48.27
23	32.01	35.17	38.08	41.64	44.18	49.73
24	33.20	36.41	39.36	42.98	45.56	51.18
25	34.38	37.65	40.65	44.31	46.93	52.62
26	35.56	38.89	41.92	45.64	48.29	54.05
27	36.74	40.11	43.19	46.96	49.64	55.48
28	37.92	41.34	44.46	48.28	50.99	56.89
29	39.09	42.56	45.72	49.59	52.34	58.30
30	40.26	43.77	46.98	50.89	53.67	59.70
40	51.81	55.76	59.34	63.69	66.77	73.40
50	63.17	67.50	71.42	76.15	79.49	86.66
60	74.40	79.08	83.30	88.38	91.95	99.61
80	96.58	101.88	106.63	112.33	116.32	124.84
100	118.50	124.34	129.56	135.81	140.17	149.45

p-Quantile der t-Verteilung mit s Freiheitsgraden

In der Zeile zu $s = \infty$ stehen die Quantile $\Phi^{-1}(p)$ der $\mathcal{N}(0,1)$-Verteilung.

	p					
s	0.900	0.950	0.975	0.990	0.995	0.999
1	3.078	6.314	12.706	31.820	63.657	318.309
2	1.886	2.920	4.303	6.965	9.925	22.327
3	1.638	2.353	3.182	4.541	5.841	10.214
4	1.533	2.132	2.776	3.747	4.604	7.173
5	1.476	2.015	2.571	3.365	4.032	5.893
6	1.440	1.943	2.447	3.143	3.707	5.208
7	1.415	1.895	2.365	2.998	3.499	4.785
8	1.397	1.860	2.306	2.896	3.355	4.501
9	1.383	1.833	2.262	2.821	3.250	4.297
10	1.372	1.812	2.228	2.764	3.169	4.144
11	1.363	1.796	2.201	2.718	3.106	4.025
12	1.356	1.782	2.179	2.681	3.055	3.930
13	1.350	1.771	2.160	2.650	3.012	3.852
14	1.345	1.761	2.145	2.625	2.977	3.787
15	1.341	1.753	2.131	2.602	2.947	3.733
16	1.337	1.746	2.120	2.584	2.921	3.686
17	1.333	1.740	2.110	2.567	2.898	3.646
18	1.330	1.734	2.101	2.552	2.878	3.610
19	1.328	1.729	2.093	2.539	2.861	3.579
20	1.325	1.725	2.086	2.528	2.845	3.552
22	1.321	1.717	2.074	2.508	2.819	3.505
23	1.319	1.714	2.069	2.500	2.807	3.485
24	1.318	1.711	2.064	2.492	2.797	3.467
26	1.315	1.706	2.056	2.479	2.779	3.435
28	1.313	1.701	2.048	2.467	2.763	3.408
30	1.310	1.697	2.042	2.457	2.750	3.385
40	1.303	1.684	2.021	2.423	2.704	3.307
50	1.299	1.676	2.009	2.403	2.678	3.261
60	1.296	1.671	2.000	2.390	2.660	3.232
80	1.292	1.664	1.990	2.374	2.639	3.195
100	1.290	1.660	1.984	2.364	2.626	3.174
∞	1.282	1.645	1.960	2.326	2.576	3.090

Aufgabe 1 (10 Punkte)

Gegeben sei eine Urliste mit den folgenden Paaren $(x_1, y_1), \ldots, (x_{10}, y_{10})$:

j	1	2	3	4	5	6	7	8	9	10
x_j	1.8	2.1	2.2	2.3	2.7	3.2	3.2	3.3	3.5	4.2
y_j	2.6	2.4	2.4	2.3	2.0	1.6	1.6	1.5	1.4	0.9

a) Berechnen Sie die Stichprobenmittel $\bar{x}$, $\bar{y}$, die Stichproben-Standardabweichungen s_x, s_y und den empirischen Korrelationskoeffizienten r_{xy}.
Hinweis:

$$\sum_{j=1}^{10} x_j = 28.5, \quad \sum_{j=1}^{10} y_j = 18.7, \quad \sum_{j=1}^{10} x_j^2 = 86.33, \quad \sum_{j=1}^{10} y_j^2 = 37.71, \quad \sum_{j=1}^{10} x_j y_j = 49.56.$$

b) Bestimmen Sie die zugehörige Regressionsgerade $y = a^* + b^* \cdot x$ von y auf x.

c) Bestimmen Sie das 0.22-getrimmte Stichprobenmittel von $(y_1, \ldots, y_{10})$.

d) Bestimmen Sie den Median, das untere und obere Quartil sowie den Quartilsabstand von $(x_1, \ldots, x_{10})$.

Lösungsvorschlag:

a) Es ist (mit den Hinweisen und $n = 10$)

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{10}\sum_{i=1}^{10} x_i = \frac{1}{10} \cdot 28.5 = 2.850,$$

$$\bar{y} = \frac{1}{n}\sum_{i=1}^{n} y_i = \frac{1}{10}\sum_{i=1}^{10} y_i = \frac{1}{10} \cdot 18.7 = 1.870,$$

$$s_x = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2} = \sqrt{\frac{1}{n-1}\left(\sum_{i=1}^{n} x_i^2 - n\bar{x}^2\right)} = \sqrt{\frac{1}{9}\left(86.33 - 10 \cdot 2.85^2\right)} \approx 0.753,$$

$$s_y = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(y_i - \bar{y})^2} = \sqrt{\frac{1}{n-1}\left(\sum_{i=1}^{n} y_i^2 - n\bar{y}^2\right)} = \sqrt{\frac{1}{9}\left(37.71 - 10 \cdot 1.87^2\right)} \approx 0.552,$$

$$r_{xy} = \frac{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})(y - \bar{y}))}{s_x s_y} = \frac{\frac{1}{n-1}\left(\sum_{i=1}^{n} x_i y_i - n\bar{x}\bar{y}\right)}{s_x s_y} = \frac{\frac{1}{9}\left(49.56 - 10 \cdot 2.85 \cdot 1.87\right)}{0.753 \cdot 0.552} \approx -0.998.$$

b) Es gilt

$$b^* = r_{xy}\frac{s_y}{s_x} = -0.998 \cdot \frac{0.552}{0.753} \approx -0.732,$$
$$a^* = \bar{y} - b^*\bar{x} = 1.87 - (-0.732) \cdot 2.85 \approx 3.956.$$

c) Die geordnete Stichprobe ist $y_{()} = (0.9, 1.4, 1.5, 1.6, 1.6, 2.0, 2.3, 2.4, 2.4, 2.6)$ und $k = [n\alpha] = [10 \cdot 0.22] = [2.2] = 2$. Damit ist das 0.22-getrimmte Mittel

$$\bar{y}_{0.22} = \frac{1}{6}(1.5 + 1.6 + 1.6 + 2.0 + 2.3 + 2.4) = 1.9.$$

d) Der Median ist das 0.5-Quantil. Da $k = [n\alpha] = [5] = 5$ und $n\alpha \in \mathbb{N}$ gilt

$$\tilde{x}_{0.5} = \frac{1}{2}(x_{(5)} + x_{(6)}) = 2.95.$$

Da $0.75 \cdot 10$ und $0.25 \cdot 10 \notin \mathbb{N}$, berechnen sich das obere und untere Quartil zu

$$\tilde{x}_{0.75} = x_{(8)} = 3.3, \quad \tilde{x}_{0.25} = x_{(3)} = 2.2,$$

und der Quartilsabstand ist

$$\tilde{x}_{0.75} - \tilde{x}_{0.25} = 3.3 - 2.2 = 1.1.$$

Aufgabe 2 (7 Punkte)

Tom will im Karlsruher Zoo zur Eisbärenanlage. Er kann hierfür nach rechts (richtiger Weg) oder nach links (falscher Weg) gehen. Fragt er einen Besucher des Zoos mit Tageskarte nach dem Weg dorthin, so erhält er mit Wahrscheinlichkeit 2/3 die richtige Antwort und mit Wahrscheinlichkeit 1/3 eine falsche Antwort. Fragt er einen Dauerkartenbesitzer nach dem Weg dorthin, so erhält er stets die richtige Antwort. Antworten und Eintrittskarten von verschiedenen Personen sind unabhängig. Die Wahrscheinlichkeit, dass eine zufällig angesprochene Person eine Dauerkarte besitzt, sei 1/10.

Hinweis:
Modellieren Sie die Situation als zweistufiges Experiment (evtl. könnte ein Baumdiagramm nützlich sein).

a) Tom fragt einen Besucher B1 nach dem Weg zur Eisbärenanlage. Mit welcher Wahrscheinlichkeit erhält er eine richtige Antwort?

b) Wie groß ist die bedingte Wahrscheinlichkeit, dass der Besucher B1 eine Dauerkarte besitzt, wenn er die richtige Antwort gegeben hat?

c) Tom fragt einen weiteren Besucher B2 nach dem Weg zur Eisbärenanlage. Mit welcher Wahrscheinlichkeit geben B1 und B2 dieselbe Antwort?

Lösungsvorschlag:

Es seien die Ereignisse R, F, T, D wie folgt definiert:

$$\begin{aligned}
R &:= \{\text{Tom erhält die richtige Antwort}\}\\
F &:= \{\text{Tom erhält die falsche Antwort}\}\\
D &:= \{\text{befragte Person ist Dauerkartenbesitzer}\}\\
T &:= \{\text{befragte Person besitzt eine Tageskarte}\}
\end{aligned}$$

Die bekannten Wahrscheinlichkeiten sind dann:

$$\begin{aligned}
&\mathbb{P}(R|T) = 2/3, \qquad \mathbb{P}(F|T) = 1/3,\\
&\mathbb{P}(R|D) = 1, \qquad \mathbb{P}(F|D) = 0,\\
&\mathbb{P}(D) = 1/10, \qquad \mathbb{P}(T) = 9/10.
\end{aligned}$$

a) Gesucht ist $\mathbb{P}(R)$. Mit der Formel von der totalen Wahrscheinlichkeit erhält man

$$\mathbb{P}(R) = \mathbb{P}(R|T)\mathbb{P}(T) + \mathbb{P}(R|D)\mathbb{P}(D) = \frac{2}{3}\cdot\frac{9}{10} + 1\cdot\frac{1}{10} = \frac{7}{10}.$$

b) Gesucht ist $\mathbb{P}(D|R)$. Mit der Bayes-Formel erhält man

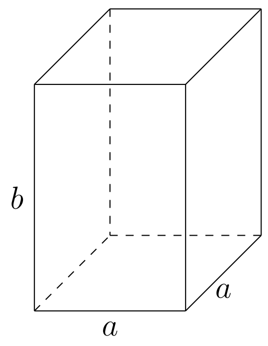
$$\mathbb{P}(D|R) = \frac{\mathbb{P}(R|D)\mathbb{P}(D)}{\mathbb{P}(R)} = \frac{1\cdot 1/10}{7/10} = \frac{1}{7}$$

c) Seien R_1, R_2, F_1, F_2 die Ereignisse, dass B_1 bzw. B_2 die richtige bzw. falsche Antwort gibt. Dann ist

$$\begin{aligned}
\mathbb{P}(\text{gleiche Antwort}) &= \mathbb{P}(R_1\cap R_2) + \mathbb{P}(F_1\cap F_2) = (\mathbb{P}(R))^2 + (\mathbb{P}(F))^2\\
&= \left(\frac{7}{10}\right)^2 + \left(\frac{3}{10}\right)^2 = \frac{29}{50}
\end{aligned}$$

Aufgabe 3 (14 Punkte)

Ein Quader mit quadratischer Grundfläche (siehe Bild) hat die Solllängen $a = 50$ und $b = 70$.



Man weiß, dass die bei der Herstellung entstehenden zufälligen Abweichungen X, Y zwischen den tatsächlichen Längen $A := a + X$, $B := b + Y$ und den Solllängen hinreichend genau normalverteilt sind, nämlich $X \sim \mathcal{N}(0, 1)$ und $Y \sim \mathcal{N}(0, 4)$. Die zufälligen tatsächlichen Längen A und B sind dabei unabhängig.

a) Bestimmen Sie die Verteilung von A, B und der Gesamtlänge $L := 8A + 4B$ aller Kanten.

b) Berechnen Sie die Wahrscheinlichkeit, dass die Gesamtlänge L um mehr als deren Standardabweichung von der Solllänge 680 abweicht.

c) Berechnen Sie die Kovarianz $C(2A, A + 2B)$.

d) Die zufällige Oberfläche des Quaders ist $O := 2A^2 + 4AB = 2A(A+2B)$. Berechnen Sie den Erwartungswert $\mathbb{E}[O]$.

e) Frau Bern hat leider die Parameter der Verteilung der Abweichung von b vergessen. Sie weiß nur, dass $Y \sim \mathcal{N}(\mu, \sigma^2)$, $\mu \in \mathbb{R}$ und $\sigma^2 > 0$ unbekannt, gilt. Um Näheres über den Erwartungswert herauszufinden, misst sie 10 Längen verschiedener (unabhängig voneinander hergestellter) Quader und erhält als Stichprobe der Abweichungen
$y = (-2.25, 0.68, 0.20, -1.09, -1.90, -0.19, 0.69, 1.67, 1.19, 2.53)$.

Hinweis: $\sum_{i=1}^{10} y_i = 1.530$ und $\sum_{i=1}^{10} (y_i - \bar{y})^2 \approx 21.247$.

i) Geben Sie für diese Situation ein zweiseitiges $(1-\alpha)-$Konfidenzintervall für den Erwartungswert von Y an.

ii) Helfen Sie Frau Bern, indem Sie das 0.95-Konfidenzintervall aus i) für die Stichprobe y berechnen.

Lösungsvorschlag:

a) Es ist

$$A \sim \mathcal{N}(a, 1),\ B \sim \mathcal{N}(b, 4),\ L \sim \mathcal{N}(8a + 4b, 64 \cdot 1 + 16 \cdot 4),$$

also

$$A \sim \mathcal{N}(50, 1),\ B \sim \mathcal{N}(70, 4),\ L \sim \mathcal{N}(680, 128).$$

b) Es ist

$$\begin{aligned}\mathbb{P}(|L - 680| > \sqrt{128}) &= \mathbb{P}(|L - \mathbb{E}L| > 1 \cdot \sqrt{V(L)}) = 1 - \mathbb{P}(|L - \mathbb{E}L| \le 1 \cdot \sqrt{V(L)}) \\ &= 1 - (2\Phi(1) - 1) = 2 - 2\Phi(1) \approx 2 - 2 \cdot 0.8413 = 0.3174.\end{aligned}$$

c) Es ist

$$C(2A, A+2B) = 2C(A,A) + 4C(A,B) = 2V(A) + 0 = 2.$$

d) Mit $C(X,Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] \implies \mathbb{E}[XY] = C(X,Y) + \mathbb{E}[X]\mathbb{E}[Y]$ ergibt sich

$$\begin{aligned}\mathbb{E}[O] &= \mathbb{E}[2A(A+2B)] = C(2A, A+2B) + \mathbb{E}[2A] \cdot \mathbb{E}[A+2B] \\ &= 2 + 2\mathbb{E}[A](\mathbb{E}[A] + 2\mathbb{E}[B]) \\ &= 2 + 2 \cdot 50(50 + 2 \cdot 70) = 19002\end{aligned}$$

<u>ALTERNATIV:</u>

$$\begin{aligned}\mathbb{E}[O] &= 2\mathbb{E}[A^2] + 4\mathbb{E}[A]\mathbb{E}[B] = 2(V(A) + \mathbb{E}[A]^2) + 4\mathbb{E}[A]\mathbb{E}[B] \\ &= 2 \cdot (1 + 50^2) + 4 \cdot 50 \cdot 70 = 19002\end{aligned}$$

e) Wir sind hier in der Situation $Y_1, \ldots, Y_n$ unabhängig und $Y_i \sim \mathcal{N}(\mu, \sigma^2)$, wobei die Parameter unbekannt sind.

i) Dann ist laut Vorlesung ein $(1-\alpha)-$Konfidenzintervall für μ gegeben durch

$$C(Y) = \left[\bar{Y}_n - t_{n-1;1-\alpha/2} \cdot \frac{\hat{\sigma}_n}{\sqrt{n}}, \bar{Y}_n + t_{n-1;1-\alpha/2} \cdot \frac{\hat{\sigma}_n}{\sqrt{n}}\right],$$

wobei $t_{n-1;1-\alpha/2}$ das $(1-\alpha/2)-$Quantil der $t-$Verteilung mit $n-1$ Freiheitsgraden ist.

ii) In der Situation von Frau Bern ist nun

$$\bar{y}_{10} = 0.153, \;\; \hat{\sigma}_{10} \approx 1.536 \;\; \text{und } t_{9;0.975} = 2.262.$$

Damit ergibt sich

$$C(y) = [-0.946; 1.252].$$

Aufgabe 4 (10 Punkte)

Eine Zufallsvariable X habe die Dichte

$$x \mapsto f_\vartheta(x) := \begin{cases} 0, & x \leq 0, \\ \frac{x}{\vartheta^2} \cdot e^{-\frac{x}{\vartheta}}, & x > 0, \end{cases}$$

wobei $\vartheta > 0$ ein unbekannter Parameter ist.
<u>*Hinweis:*</u> Sie können ohne Nachweis verwenden, dass $\mathbb{E}_\vartheta[X] = 2\vartheta$ und $V_\vartheta(X) = 2\vartheta^2$ gilt.

a) Bestimmen Sie für die unabhängigen Stichprobenvariablen $X_1, \ldots, X_n$ mit der Dichte f_ϑ einen Maximum-Likelihood-Schätzer für ϑ zur Stichprobe $x = (x_1, \ldots, x_n)$, wobei $x_i > 0$ für alle $i = 1, \ldots, n$.
Dazu gehen Sie wie folgt vor:

   i) Geben Sie die zur Stichprobe $x = (x_1, \ldots, x_n)$ gehörende Likelihood-Funktion $L_x(\vartheta)$ an.

   ii) Geben Sie die zur Stichprobe $x = (x_1, \ldots, x_n)$ gehörende Loglikelihood-Funktion $M_x(\vartheta)$ an und vereinfachen Sie diese.

   iii) Berechnen Sie die Ableitung $M_x'(\vartheta)$ der Loglikelihood-Funktion.

   iv) Bestimmen Sie einen Maximum-Likelihood-Schätzwert $\hat{\vartheta}(x)$ für ϑ zur Stichprobe x.

b) Ist der Schätzer $\hat{\vartheta}$ aus Teilaufgabe a) erwartungstreu für ϑ? Begründen Sie Ihre Antwort!

c) Ist die Schätzfolge $(\hat{\vartheta}_n)$ konsistent für ϑ? Begründen Sie Ihre Antwort!

Lösungsvorschlag:

a) Es werden Teil i)- iv) gleichzeitig gelöst:
Zur Bestimmung des Maximum-Likelihood-Schätzers stellen wir zunächst die Likelihood-Funktion auf. Es ist

$$\begin{aligned} L_x(\vartheta) &= \prod_{i=1}^n f_\vartheta(x_i) = \prod_{i=1}^n \frac{x_i}{\vartheta^2} \cdot e^{-\frac{x_i}{\vartheta}} \\ &= \frac{\prod_{i=1}^n x_i}{\vartheta^{2n}} e^{-\frac{1}{\vartheta}\sum_{i=1}^n x_i}. \end{aligned}$$

Als nächstes wird die Log-Likelihood-Funktion bestimmt.

$$M_x(\vartheta) = \log(L_x(\vartheta)) = \sum_{i=1}^n \log(x_i) - 2n \log \vartheta - \frac{1}{\vartheta} \sum_{i=1}^n x_i.$$

Diese Funktion wird nun maximiert. Hierzu fordern wir

$$M_x'(\vartheta) = -\frac{2n}{\vartheta} + \frac{1}{\vartheta^2} \sum_{i=1}^n x_i \stackrel{!}{=} 0.$$

Damit ist der Kandidat für den Maximum-Likelihood-Schätzer

$$\hat{\vartheta}(X_1, \ldots, X_n) = \frac{1}{2n} \sum_{x=1}^n X_i.$$

Da

$$M_x'(\vartheta) = \frac{1}{\vartheta^2} \left(-2n\vartheta + \sum_{i=1}^n x_i \right) \gtreqless 0 \iff \frac{1}{2n} \sum_{i=1}^n x_i \gtreqless \vartheta,$$

ist $\hat{\vartheta}$ wirklich eine Maximumsstelle und $\hat{\vartheta}$ der Maximum-Likelihood-Schätzer für ϑ.

b) Es ist

$$\begin{aligned}\mathbb{E}_\vartheta[\hat{\vartheta}(X_1,\dots,X_n)] &= \mathbb{E}_\vartheta\left[\frac{1}{2n}\sum_{x=1}^{n} X_i\right] = \frac{1}{2n}\sum_{i=1}^{n}\mathbb{E}_\vartheta[X_i] \\ &= \frac{1}{2}\mathbb{E}_\vartheta[X_1] = \frac{1}{2}\cdot 2\vartheta = \vartheta.\end{aligned}$$

Damit ist der Schätzer erwartungstreu für ϑ.

	ja	nein
erwartungstreu:	$\times$	

c) Die Schätzfolge ist konsistent, denn es gilt mit dem Gesetz großer Zahlen

$$\lim_{n\to\infty}\mathbb{P}_\vartheta\left(\left|\frac{1}{2n}\sum_{i=1}^{n} X_i-\vartheta\right|\geq\varepsilon\right) = \lim_{n\to\infty}\mathbb{P}_\vartheta\left(\left|\frac{1}{n}\sum_{i=1}^{n}\frac{X_i}{2}-\mathbb{E}_\vartheta\left[\frac{X_1}{2}\right]\right|\geq\varepsilon\right) = 0.$$

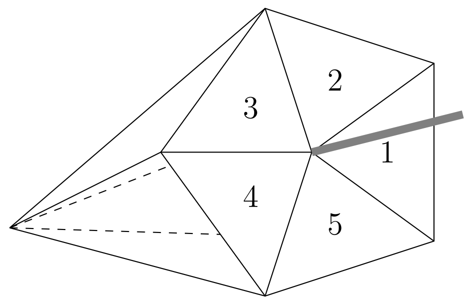
Alternativ lässt sich die Konsistenz mit Hilfe der Tschebyschev-Ungleichung zeigen:

$$\begin{aligned}\mathbb{P}_\vartheta\left(\left|\frac{1}{2n}\sum_{i=1}^{n} X_i-\vartheta\right|\geq\varepsilon\right) &= \mathbb{P}_\vartheta\left(\left|\frac{1}{2n}\sum_{i=1}^{n} X_i-\mathbb{E}_\vartheta\left[\frac{1}{2n}\sum_{i=1}^{n} X_i\right]\right|\geq\varepsilon\right) \\ &\leq \frac{1}{\varepsilon^2}V_\vartheta\left(\frac{1}{2n}\sum_{i=1}^{n} X_i\right) = \frac{1}{\varepsilon^2}\frac{1}{4n^2}\sum_{i=1}^{n} V_\vartheta\left(X_i\right) \\ &= \frac{V_\vartheta(X_1)}{4n\varepsilon^2} = \frac{2\vartheta^2}{4n\varepsilon^2}\overset{n\to\infty}{\longrightarrow} 0.\end{aligned}$$

	ja	nein
konsistent:	$\times$	

Aufgabe 5 (13 Punkte)

Ein Kreisel mit fünfeckiger „Grundfläche“ (siehe Bild) wird n mal in unabhängiger Folge gedreht ($n \in \mathbb{N}$).



Dabei bezeichne X_i die Zahl, die bei der $i-$ten Drehung unten ist. Die Zufallsvariable

$$Y_n := \sum_{i=1}^{n} X_i$$

bezeichne die nach n Drehungen erzielte Zahlen-Summe.
Nehmen Sie zunächst an, dass die Grundfläche ein regelmäßiges Fünfeck ist.

a) Berechnen Sie den Erwartungswert und die Varianz von Y_n.

b) Approximieren Sie mit Hilfe des Zentralen Grenzwertsatzes von Lindeberg-Lévy die Wahrscheinlichkeit

$$\mathbb{P}(275 < Y_{100} < 325).$$

c) Herr Franz glaubt nicht an die mit dem regelmäßigen Fünfeck verbundene Laplace-Annahme und führt deshalb ein Experiment durch. Er dreht den Kreisel 150 mal in unabhängiger Folge. Dabei treten die Zahlen $1, \dots, 5$ in den Häufigkeiten $22, 36, 40, 24, 28$ auf. Nun möchte Herr Franz einen Test durchführen, um seine Behauptung, der Kreisel habe KEIN regelmäßiges Fünfeck als Grundseite, statistisch nachzuweisen.

   i) Wie wählt Herr Franz die Hypothese bzw. Alternative? Welchen aus der Vorlesung bekannten Test muss Herr Franz nun anwenden?

   ii) Wie lautet die Testgröße? Berechnen Sie die Testgröße aus den Daten.

   iii) Wie lautet der Testentscheid zum Niveau α?

   iv) Kann Herr Franz die Hypothese der Regelmäßigkeit des Fünfecks verwerfen, wenn er eine Wahrscheinlichkeit von 0.05 für den Fehler 1. Art toleriert?

Lösungsvorschlag:

Da wir zunächst annehmen, dass das Fünfeck regelmäßig ist, gilt

$$\mathbb{P}(X_i = k) = \frac{1}{5}, \quad \text{für } k = 1, \dots, 5 \text{ und } i \in \mathbb{N}.$$

a) Es ist

$$\begin{aligned}\mathbb{E}[Y_n] = \mathbb{E}\left[\sum_{i=1}^{n} X_i\right] &= \sum_{i=1}^{n} \mathbb{E}[X_i] = n \cdot \mathbb{E}[X_1] \\ &= n \cdot \frac{1}{5} \cdot (1+2+3+4+5) = 3n\end{aligned}$$

und da die X_i unabhängig sind

$$\begin{aligned} V(Y_n) &= V\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} V[X_i] = n \cdot V(X_1) \\ &= n \cdot (\mathbb{E}(X_1^2) - \mathbb{E}[X_1]^2) = n \cdot (11 - 3^2) = 2n. \end{aligned}$$

b) Es gilt

$$\begin{aligned} \mathbb{P}(275 < Y_{100} < 325) &= \mathbb{P}\left(\frac{275-300}{\sqrt{200}} < \frac{Y_{100}-300}{\sqrt{200}} < \frac{325-300}{\sqrt{200}}\right) \\ &\approx \Phi\left(\frac{25}{10\sqrt{2}}\right) - \Phi\left(-\frac{25}{10\sqrt{2}}\right) \approx \Phi(1.768) - \Phi(-1.768) \\ &= \Phi(1.768) - (1 - \Phi(1.768)) = 2\Phi(1.768) - 1 = 2 \cdot 0.9616 - 1 = 0.9232 \end{aligned}$$

c) i) Wir sind hier in der Situation eines Chi-Quadrat-Anpassungstests, der aus der Vorlesung bekannt ist. Da Herr Franz die Behauptung $\mathbb{P}(X_i = k) \neq \frac{1}{5}$ statistisch nachweisen möchte, ist die Hypothese bzw. Alternative

$$H_0 : p_j = \frac{1}{5} \qquad\qquad H_1 : p_j \neq \frac{1}{5}.$$

ii) Die Testgröße ist

$$\chi_n^2(k_1, \ldots, k_s) = \sum_{i=1}^{s} \frac{(k_i - n \cdot \pi_i)^2}{n \cdot \pi_i}.$$

Mit $n = 150$, $s = 5$, $(k_1, \ldots, k_5) = (22, 36, 40, 24, 28)$ und $\pi_i = \frac{1}{5}$ für $1 \leq i \leq 5$ ergibt sich

$$\chi_{150}^2(22, 36, 40, 24, 28) = \sum_{i=1}^{5} \frac{(k_i - 30)^2}{30} = 8.$$

iii) Der Test zum Niveau α lautet

$$\begin{aligned} H_0 \text{ wird verworfen, } &\text{ falls } \chi_n^2(k_1, \ldots, k_s) \geq \chi_{s-1;1-\alpha}^2 \\ \text{kein Widerspruch zu } H_0, &\text{ falls } \chi_n^2(k_1, \ldots, k_s) < \chi_{s-1;1-\alpha}^2 \end{aligned}$$

wobei $\chi_{s-1;1-\alpha}^2$ das $(1-\alpha)-$Quantil der χ^2-Verteilung mit $s-1$ Freiheitsgraden ist.

iv) Es ist

$$\chi_{4;0.95}^2 = 9.49 > 8 = \chi_{150}^2(22, 36, 40, 24, 28).$$

Damit stehen die Daten nicht im Widerspruch zu H_0.