Wahrscheinlichkeitstheorie und Statistik

Serie 10 - Lösungen

MC 10-1. Seien X und Y i.i.d. Zufallsvariablen mit Varianz $\sigma^2 > 0$. Was ist die Standardabweichung der Zufallsvariablen $Z := X + X - Y$? (Genau eine Antwort ist richtig.)

(a) σ.

(b) 3σ.

(c) 5σ.

(d) $\sqrt{3}\,\sigma$.

(e) $\sqrt{5}\,\sigma$.

Lösung: (e) ist korrekt. Wenn X und Y unabhängig sind, so sind auch $2X$ und $-Y$ unabhängig. Es gilt also

$$\mathrm{Var}[2X - Y] = \mathrm{Var}[2X] + \mathrm{Var}[-Y] = 2^2\mathrm{Var}[X] + \mathrm{Var}[Y] = 4\sigma^2 + \sigma^2 = 5\sigma^2.$$

Die Standardabweichung ist somit

$$\sqrt{5\sigma^2} = \sqrt{5}\,\sigma.$$

MC 10-2. Sei X eine Zufallsvariable mit wohldefinierter positiver Varianz. Finden Sie $a \in \mathbb{R}$ und $b > 0$ so, dass für $Z := (X - a)/b$ gilt $\mathbb{E}[Z] = 0$ und $\mathrm{Var}[Z] = 1$. (Genau eine Antwort ist richtig.)

(a) $a = \mathbb{E}[X]$, $b = \mathrm{Var}[X]$.

(b) $a = \mathbb{E}[X]$, $b = \sqrt{\mathrm{Var}[X]}$.

(c) $a = \mathbb{E}[X^2]$, $b = \mathrm{Var}[X]$.

(d) $a = (\mathbb{E}[X])^2$, $b = \mathrm{Var}[X]$.

Lösung: (b) ist korrekt. Wir berechnen

$$\mathbb{E}[Z] = \frac{1}{b}(\mathbb{E}[X] - a) \quad \text{und} \quad \mathrm{Var}[Z] = \frac{1}{b^2}\mathrm{Var}[X - a] = \frac{1}{b^2}\mathrm{Var}[X].$$

Es gilt $\frac{1}{b^2}\mathrm{Var}[X] = 1$ genau dann, wenn $b = \pm\sqrt{\mathrm{Var}[X]}$. Da wir $b > 0$ voraussetzen, erhalten wir $b = \sqrt{\mathrm{Var}[X]}$. Ferner ist $\frac{\mathbb{E}[X]-a}{b} = 0$ genau dann, wenn $a = \mathbb{E}[X]$. Insgesamt ergibt sich

$$Z = \frac{X - \mathbb{E}[X]}{\sqrt{\mathrm{Var}[X]}}.$$

MC 10-3. Seien X und Y Zufallsvariablen. Unter welchen Vorraussetzungen ist die gemeinsame Verteilungsfunktion $F_{X,Y} : \mathbb{R}^2 \to [0,1], (x,y) \mapsto F_{X,Y}(x,y) = \mathbb{P}[X \le x, Y \le y]$ durch die Verteilungsfunktionen F_X und F_Y von X und Y eindeutig definiert? (Die Anzahl der möglichen richtigen Antworten kann zwischen 0 und 4 liegen.)

(a) Immer. Sobald man die Verteilung von X und die von Y kennt, kennt man auch die gemeinsame Verteilung.

(b) Unter der Voraussetzung, dass X und Y unabhängig sind.

(c) Unter der Voraussetzung, dass X und Y identisch verteilt sind ($F_X = F_Y$).

(d) Unter der Voraussetzung, dass X und Y diskret sind.

(e) Unter der Voraussetzung, dass X und Y Bernoulli-verteilt sind.

(f) Unter der Voraussetzung, dass X und Y absolutstetig sind.

(g) Unter der Voraussetzung, dass X und Y normalverteilt sind.

Lösung: Nur (b) ist korrekt. Wenn X und Y unabhängig sind, dann gilt $F_{X,Y}(x,y) = F_X(x)F_Y(y)$ für $x, y \in \mathbb{R}$. Keine der anderen Antworten ist korrekt.

Aufgabe 10-4. Sei $X_1, \ldots, X_n$ eine Zufallsstichprobe aus einer Poisson-Verteilung mit Parameter $\lambda > 0$, der unbekannt ist. Bestimmen Sie den Maximum-Likelihood-Schätzer für λ.

Lösung: Die Gewichtsfunktion der Poisson-Verteilung ist

$$\mathbb{P}[X_i = k] = e^{-\lambda}\frac{\lambda^k}{k!}, \quad k \in \mathbb{N}_0.$$

Wegen Unabhängigkeit ist die Likelihood-Funktion

$$L(x_1, \ldots, x_n; \lambda) = \prod_{i=1}^n \mathbb{P}[X_i = x_i] = \prod_{i=1}^n e^{-\lambda}\frac{\lambda^{x_i}}{x_i!} = \frac{e^{-n\lambda}\prod_{i=1}^n \lambda^{x_i}}{x_1! \cdots x_n!}.$$

Wir berechnen

$$\begin{aligned}\log L(x_1, \ldots, x_n; \lambda) &= \log \frac{e^{-n\lambda}\prod_{i=1}^n \lambda^{x_i}}{x_1! \ldots x_n!} = \log(e^{-n\lambda}) + \log\Big(\prod_{i=1}^n \lambda^{x_i}\Big) - \log(x_1! \cdots x_n!) \\ &= -n\lambda + \sum_{i=1}^n \log(\lambda^{x_i}) + \log(x_1! \cdots x_n!) \\ &= -n\lambda + \log\lambda \sum_{i=1}^n x_i + \log(x_1! \cdots x_n!).\end{aligned}$$

Wir wollen das über λ maximieren und berechnen deshalb

$$\frac{\partial}{\partial\lambda}\log L(x_1, \ldots, x_n; \lambda) = -n + \frac{1}{\lambda}\sum_{i=1}^n x_i.$$

Dies ist gleich 0 genau dann, wenn $\lambda = \frac{1}{n}\sum_{i=1}^n x_i$. Ferner gilt

$$\frac{\partial^2}{\partial\lambda^2}\log L(x_1, \ldots, x_n; \lambda) = -\frac{1}{\lambda^2}\sum_{i=1}^n x_i \leq 0,$$

weil $x_i \geq 0$, $i \in \{1, \dots, n\}$, mit Wahrscheinlichkeit 1 ist. Somit ist die Funktion $(0, \infty) \to \mathbb{R}$, $\lambda \mapsto \log L(x_1, \dots, x_n; \lambda)$ konkav und $\widehat{\lambda}(x_1, \dots, x_n) = \frac{1}{n} \sum_{i=1}^n x_i$ ist der Maximierer.

Beachten Sie, dass $\lambda = \mathbb{E}[X^i]$, $i \in \{1, \dots, n\}$. Es ist daher nicht überraschend, dass der Maximum-Likelihood-Schätzer der Durchschnitt der beobachteten Werte ist.

Aufgabe 10-5. Seien $X_1, \dots, X_n$ unabhängige Zufallsvariablen mit $X_i \sim \mathcal{N}(\theta\alpha_i, 1)$, wobei $\alpha_i \neq 0$ bekannte Parameter sind, aber $\theta \in \mathbb{R}$ unbekannt ist. Bestimmen Sie den Maximum-Likelihood-Schätzer für θ.

Lösung: Die Randdichten sind $f_{X_i}(x_i) = \frac{1}{\sqrt{2\pi}} \exp(-\frac{1}{2}(x_i - \theta\alpha_i)^2)$. Wegen Unabhängigkeit ist die Likelihood-Funktion

$$L(x_1, \dots, x_n; \theta) = \prod_{i=1}^n f_{X_i}(x_i) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}} \exp\Big(-\frac{1}{2}(x_i - \theta\alpha_i)^2\Big).$$

Wir berechnen

$$\log L(x_1, \dots, x_n; \theta) = \sum_{i=1}^n \Big(-\frac{1}{2}\log(2\pi) - \frac{1}{2}(x_i - \theta\alpha_i)^2\Big).$$

Wir wollen das über θ maximieren und berechnen deshalb

$$\frac{\partial}{\partial\theta} \log L(x_1, \dots, x_n; \theta) = -\sum_{i=1}^n (x_i - \theta\alpha_i)(-\alpha_i) = \sum_{i=1}^n (x_i - \theta\alpha_i)\alpha_i.$$

Wegen $\sum_{i=1}^n \alpha_i^2 > 0$ gilt

$$\sum_{i=1}^n (x_i - \theta\alpha_i)\alpha_i = 0 \iff \sum_{i=1}^n x_i\alpha_i - \theta \sum_{i=1}^n \alpha_i^2 = 0 \iff \theta = \frac{\sum_{i=1}^n x_i\alpha_i}{\sum_{i=1}^n \alpha_i^2}.$$

Ferner gilt

$$\frac{\partial^2}{\partial\theta^2} \log L(x_1, \dots, x_n; \theta) = -\sum_{i=1}^n \alpha_i^2 \leq 0.$$

Somit ist die Funktion $\mathbb{R} \to \mathbb{R}$, $\theta \mapsto \log L(x_1, \dots, x_n; \theta)$ konkav und

$$\widehat{\theta}(x_1, \dots, x_n) = \frac{\sum_{i=1}^n x_i\alpha_i}{\sum_{i=1}^n \alpha_i^2}$$

ist der Maximierer.

Aufgabe 10-6. Das amerikanische Roulette hat 38 mögliche Ergebnisse. Es gibt 18 schwarze Zahlen, 18 rote Zahlen und zwei grüne Zahlen (0 und 00). Daher sind die theoretischen Wahrscheinlichkeiten

$$p_1 = \mathbb{P}[\text{“schwarze Zahl”}] = \frac{18}{38} = \frac{9}{19}, \quad p_2 = \mathbb{P}[\text{“rote Zahl”}] = \frac{9}{19}, \quad p_3 = \mathbb{P}[\text{“grüne Zahl”}] = \frac{2}{38} = \frac{1}{19}. \quad (1)$$

Wir beobachten 100 Spiele in einem Casino, und die Ergebnisse lauten wie folgt:

Schwarz	Rot	Grün
51	40	9

Wir möchten testen, ob das Casino betrügt oder nicht. Das heisst, wir möchten testen, ob die Ergebnisse den theoretischen Wahrscheinlichkeiten entsprechen.

(a) Schlagen Sie einen geeigneten statistischen Test vor.

(b) Formulieren Sie die Nullhypothese und die Alternative.

(c) Finden Sie die Teststatistik und den kritischen Bereich auf dem approximativen Niveau 5%.

(d) Bestimmen Sie das Ergebnis des Tests und ziehen Sie eine Schlussfolgerung.

Lösung:

(a) Wir verwenden den χ^2-Anpassungstest.

(b)

$$\begin{aligned} &\mathrm{H}_0 : \text{Die Wahrscheinlichkeiten entsprechen (1).} \\ &\mathrm{H}_1 : \text{Die Wahrscheinlichkeiten unterscheiden sich von (1).} \end{aligned}$$

(c) Sind n_i die beobachteten Anzahlen und p_i die theoretischen Wahrscheinlichkeiten, so ist der Wert der Teststatistik des χ^2-Anpassungstests

$$t = \sum_{i=1}^{K} \frac{(n_i - Np_i)^2}{Np_i}.$$

Wir wissen, dass $T = \sum_{i=1}^{K} \frac{(N_i - Np_i)^2}{Np_i}$ unter der Nullhypothese approximativ χ^2_{K-1}-verteilt ist. Daher ist der approximative kritische Bereich $(t_{0.95}(\chi^2_{K-1}), \infty)$, wobei $t_{0.95}(\chi^2_{K-1})$ das 0.95-Quantil der χ^2_{K-1}-Verteilung ist, d.h. derjenige Wert c, für den $\int_c^\infty f_{K-1}(x)dx = \alpha$ gilt.

(d) Wir haben $K = 3$, $N = 100$, $n_1 = 51$, $n_2 = 40$, $n_3 = 9$. Wir berechnen daher

$$\begin{aligned} t &= \sum_{i=1}^{3} \frac{(n_i - Np_i)^2}{Np_i} \\ &= \frac{(51 - 100 \times \frac{9}{19})^2}{100 \times \frac{9}{19}} + \frac{(40 - 100 \times \frac{9}{19})^2}{100 \times \frac{9}{19}} + \frac{(9 - 100 \times \frac{1}{19})^2}{100 \times \frac{1}{19}} \\ &= \frac{347}{90} \approx 4.08. \end{aligned}$$

Ferner finden wir in der Tabelle $t_{0.95}(\chi^2_2) = 5.991 > 3.787$. Daher können wir die Nullhypothese nicht ablehnen und finden keinen Hinweis darauf, ob das Casino betrügt.

Quantiltabelle der χ^2_ν-Verteilung

ν	0.6	0.667	0.75	0.80	0.875	0.9	0.95	0.975	0.99	0.995	0.999
1	0.708	0.936	1.323	1.642	2.354	2.706	3.841	5.024	6.635	7.879	10.828
2	1.833	2.197	2.773	3.219	4.159	4.605	5.991	7.378	9.210	10.597	13.816
3	2.946	3.405	4.108	4.642	5.739	6.251	7.815	9.348	11.345	12.838	16.266
4	4.045	4.579	5.385	5.989	7.214	7.779	9.488	11.143	13.277	14.860	18.467
5	5.132	5.730	6.626	7.289	8.625	9.236	11.070	12.833	15.086	16.750	20.515
6	6.211	6.867	7.841	8.558	9.992	10.645	12.592	14.449	16.812	18.548	22.458
7	7.283	7.992	9.037	9.803	11.326	12.017	14.067	16.013	18.475	20.278	24.322
8	8.351	9.107	10.219	11.030	12.636	13.362	15.507	17.535	20.090	21.955	26.125
9	9.414	10.215	11.389	12.242	13.926	14.684	16.919	19.023	21.666	23.589	27.877
10	10.473	11.317	12.549	13.442	15.198	15.987	18.307	20.483	23.209	25.188	29.588
11	11.530	12.414	13.701	14.631	16.457	17.275	19.675	21.920	24.725	26.757	31.264
12	12.584	13.506	14.845	15.812	17.703	18.549	21.026	23.337	26.217	28.300	32.910
13	13.636	14.595	15.984	16.985	18.939	19.812	22.362	24.736	27.688	29.819	34.528
14	14.685	15.680	17.117	18.151	20.166	21.064	23.685	26.119	29.141	31.319	36.123
15	15.733	16.761	18.245	19.311	21.384	22.307	24.996	27.488	30.578	32.801	37.697
16	16.780	17.840	19.369	20.465	22.595	23.542	26.296	28.845	32.000	34.267	39.252
17	17.824	18.917	20.489	21.615	23.799	24.769	27.587	30.191	33.409	35.718	40.790
18	18.868	19.991	21.605	22.760	24.997	25.989	28.869	31.526	34.805	37.156	42.312
19	19.910	21.063	22.718	23.900	26.189	27.204	30.144	32.852	36.191	38.582	43.820
20	20.951	22.133	23.828	25.038	27.376	28.412	31.410	34.170	37.566	39.997	45.315
21	21.991	23.201	24.935	26.171	28.559	29.615	32.671	35.479	38.932	41.401	46.797
22	23.031	24.268	26.039	27.301	29.737	30.813	33.924	36.781	40.289	42.796	48.268
23	24.069	25.333	27.141	28.429	30.911	32.007	35.172	38.076	41.638	44.181	49.728
24	25.106	26.397	28.241	29.553	32.081	33.196	36.415	39.364	42.980	45.559	51.179
25	26.143	27.459	29.339	30.675	33.247	34.382	37.652	40.646	44.314	46.928	52.620
26	27.179	28.520	30.435	31.795	34.410	35.563	38.885	41.923	45.642	48.290	54.052
27	28.214	29.580	31.528	32.912	35.570	36.741	40.113	43.195	46.963	49.645	55.476
28	29.249	30.639	32.620	34.027	36.727	37.916	41.337	44.461	48.278	50.993	56.892
29	30.283	31.697	33.711	35.139	37.881	39.087	42.557	45.722	49.588	52.336	58.301
30	31.316	32.754	34.800	36.250	39.033	40.256	43.773	46.979	50.892	53.672	59.703
35	36.475	38.024	40.223	41.778	44.753	46.059	49.802	53.203	57.342	60.275	66.619
40	41.622	43.275	45.616	47.269	50.424	51.805	55.758	59.342	63.691	66.766	73.402
45	46.761	48.510	50.985	52.729	56.052	57.505	61.656	65.410	69.957	73.166	80.077
50	51.892	53.733	56.334	58.164	61.647	63.167	67.505	71.420	76.154	79.490	86.661
55	57.016	58.945	61.665	63.577	67.211	68.796	73.311	77.380	82.292	85.749	93.168
60	62.135	64.147	66.981	68.972	72.751	74.397	79.082	83.298	88.379	91.952	99.607