Wahrscheinlichkeitstheorie und Statistik

Serie 12 - Lösungen

MC 12-1. Sei $X_1, X_2, \ldots$ eine Folge von unabhängigen, identisch verteilten Zufallsvariablen mit $\mathbb{E}[X_1^2] < \infty$. Sei Z eine standardnormalverteilte Zufallsvariable. Wir definieren $\mu := \mathbb{E}[X_1]$ und $\sigma^2 := \mathrm{Var}[X_1]$. Welche der folgenden Aussagen ist korrekt? (Genau eine Antwort ist richtig.)

(a) $\lim_{n\to\infty} \mathbb{P}\left[\frac{1}{\sigma^2 n}\sum_{i=1}^n X_i \le a\right] = \mathbb{P}[Z \le a],\ a \in \mathbb{R}.$

(b) $\lim_{n\to\infty} \mathbb{P}\left[\frac{1}{\sigma^2 n}\sum_{i=1}^n (X_i - \mu) \le a\right] = \mathbb{P}[Z \le a],\ a \in \mathbb{R}.$

(c) $\lim_{n\to\infty} \mathbb{P}\left[\frac{1}{\sqrt{\sigma^2 n}}\sum_{i=1}^n (X_i - \mu) \le a\right] = \mathbb{P}[Z \le a],\ a \in \mathbb{R}.$

(d) $\lim_{n\to\infty} \mathbb{P}\left[\frac{1}{\sqrt{\sigma^2 n}}\sum_{i=1}^n X_i \le a\right] = \mathbb{P}[Z \le a],\ a \in \mathbb{R}.$

Lösung: (c) ist die korrekte Form des zentralen Grenzwertsatzes.

MC 12-2. Gilt die richtige Antwort von MC 12-1 auch für endliche n exakt (also wenn man $\lim_{n\to\infty}$ weglassen würde)? (Die Anzahl der möglichen richtigen Antworten kann zwischen 0 und 4 liegen.)

(a) Die richtige Antwort von MC 12-1 gilt auch, wenn man $\lim_{n\to\infty}$ weglässt, für alle Verteilungen die $\mathbb{E}[X_1^2] < \infty$ erfüllen.

(b) Die richtige Antwort von MC 12-1 gilt auch, wenn man $\lim_{n\to\infty}$ weglässt, falls die X_i normalverteilt sind.

(c) Die richtige Antwort von MC 12-1 gilt auch, wenn man $\lim_{n\to\infty}$ weglässt, falls $\mu = 0$ und $\sigma = 1$.

(d) Die richtige Antwort von MC 12-1 gilt nie, wenn man $\lim_{n\to\infty}$ weglässt.

Lösung: (b) ist korrekt. Wenn $X_i \sim \mathcal{N}(\mu, \sigma^2)$, dann gilt wegen Unabhängigkeit $\sum_{i=1}^n X_i \sim \mathcal{N}(n\mu, n\sigma^2)$. Folglich ist

$$\frac{1}{\sqrt{n\sigma^2}}\left(\sum_{i=1}^n X_i - n\mu\right) = \frac{1}{\sqrt{n\sigma^2}}\sum_{i=1}^n (X_i - \mu) \sim \mathcal{N}(0,1).$$

Keine der anderen Optionen ist korrekt.

Aufgabe 12-3. Es kostet 1 Dollar, an einem bestimmten Spielautomaten in Las Vegas zu spielen. Der Automat ist vom Casino so eingestellt, dass er mit einer Wahrscheinlichkeit von 0.45 2 Dollar auszahlt (der

Spieler gewinnt) und mit einer Wahrscheinlichkeit von 0.55 nichts auszahlt (das Casino gewinnt). Sei X_i der Nettogewinn des Casinos beim i-ten Spiel des Automaten. Sei $S_n := \sum_{i=1}^n X_i$ der Gewinn des Casinos nach n Spielrunden des Automaten. Unter der Annahme, dass aufeinanderfolgende Spiele unabhängig sind, finden Sie:

(a) $\mathbb{E}[S_n]$ und $\mathrm{Var}[S_n]$;

(b) die ungefähre Wahrscheinlichkeit, dass nach $10'000$ Runden des Automaten der Gewinn des Casinos zwischen 800 und 1100 Dollar liegt.

Nehmen Sie an, dass wir wissen, dass der Gewinn des Casinos nach $10'000$ Runden 1200 Dollar beträgt.

(c) Können wir sagen basierend auf unseren Beobachtungen, dass die Wahrscheinlichkeit, dass der Automat gewinnt, grösser ist als der angegebene Wert 0.55? Nehmen Sie das Niveau $\alpha = 0.05$.

Lösung:

(a) Die Gewinne des Casinos bei jedem Spiel sind unabhängig und haben die Werte

$$X_i = \begin{cases} 1 & \text{mit Wahrscheinlichkeit } p = 0.55, \\ -1 & \text{mit Wahrscheinlichkeit } 1 - p = 0.45. \end{cases}$$

Dies ergibt $\mathbb{E}[X_i] = 1 \times 0.55 + (-1) \times 0.45 = 0.1$ und $\mathrm{Var}[X_i] = \mathbb{E}[X_i^2] - (\mathbb{E}[X_i])^2 = 1 - 0.1^2 = 0.99$. Somit gilt

$$\mathbb{E}[S_n] = \sum_{i=1}^n \mathbb{E}[X_i] = n\mathbb{E}[X_1] = 0.1n \quad \text{und} \quad \mathrm{Var}[S_n] = \sum_{i=1}^n \mathrm{Var}[X_i] = n\mathrm{Var}[X_1] = 0.99n.$$

Alternativ kann man beobachten, dass

$$B_i := \frac{X_i + 1}{2}, \quad i \in \{1, \dots, n\},$$

unabhängige Bernoulli(0.55)-verteilte Zufallsvariablen sind. Ferner ist,

$$S_n = \sum_{i=1}^n X_i = \sum_{i=1}^n (2B_i - 1) = 2\sum_{i=1}^n B_i - n = 2\tilde{S}_n - n,$$

und $\tilde{S}$ eine $\mathrm{Binom}(n, 0.55)$-Verteilung hat. Also ist $\mathbb{E}[\tilde{S}_n] = 0.55n$, $\mathrm{Var}[\tilde{S}_n] = 0.55 \times 0.45n$ und

$$\mathbb{E}[S_n] = 2\mathbb{E}[\tilde{S}_n] - n = 0.1n \;\text{ und }\; \mathrm{Var}[S_n] = 4\mathrm{Var}[\tilde{S}_n] = 0.99n.$$

(b) Da die X_n i.i.d. sind, können wir mit Hilfe des zentralen Grenzwertsatzes eine gute Gauss-Approximation für diese Wahrscheinlichkeit finden. Alternativ erhalten wir das auch aus dem Satz von de Moivre–Laplace. Für grosse n gilt ungefähr

$$\frac{S_n - 0.1n}{\sqrt{0.99n}} \sim \mathcal{N}(0, 1)$$

und insbesondere ungefähr

$$\frac{S_{10'000} - 1000}{\sqrt{9900}} \sim \mathcal{N}(0, 1).$$

Daher erhalten wir

$$\begin{aligned}\mathbb{P}[800 \leq S_{10'000} \leq 1100] &= \mathbb{P}\left[-\frac{200}{\sqrt{9900}} \leq \frac{S_{10'000}-1000}{\sqrt{9900}} \leq \frac{100}{\sqrt{9900}}\right] \\ &\approx \Phi\left(\frac{100}{\sqrt{9900}}\right) - \Phi\left(-\frac{200}{\sqrt{9900}}\right) \\ &= \Phi(1.005) - \Phi(-2.010) \\ &\approx 0.82.\end{aligned}$$

(c) Wir formulieren

$$H_0' : p \leq 0.55, \quad H_1 : p > 0.55.$$

Ähnlich wie in Aufgabe 11-4 nehmen wir aber $H_0 : p = 0.55$. Unter H_0 haben wir bereits gesehen, dass ungefähr gilt

$$\frac{S_{10'000}-1000}{\sqrt{9900}} \sim \mathcal{N}(0,1).$$

Wir haben also

$$\mathbb{P}_{0.55}[S_{10'000} \geq c] = \mathbb{P}_{0.55}\left[\frac{S_{10'000}-1000}{\sqrt{9900}} \geq \frac{c-1000}{\sqrt{9900}}\right] \approx 1 - \Phi\left(\frac{c-1000}{\sqrt{9900}}\right).$$

Ausserdem gilt

$$\begin{aligned}1 - \Phi\left(\frac{c-1000}{\sqrt{9900}}\right) = 0.05 &\iff \frac{c-1000}{\sqrt{9900}} = \Phi^{-1}(0.95) \\ &\iff \frac{c-1000}{\sqrt{9900}} = 1.6449 \\ &\iff c = 1000 + 1.6449\sqrt{9900} \approx 1163.67.\end{aligned}$$

Wir erhalten also den approximativen kritischen Bereich $[1164, \infty)$. Wegen $1200 \in [1164, \infty)$ können wir also H_0 verwerfen und zuversichtlich behaupten, dass die Wahrscheinlichkeit, dass das Casino gewinnt, grösser als 0.55 ist.

Aufgabe 12-4. Ein Team von drei Personen wird zufällig aus einer Gruppe von sechs Personen ausgewählt. Von diesen sechs Personen sind drei Frauen (Anna, Elsa und Helga) und drei Männer (Franz, Mario und Tobias). Sei X die Anzahl der Frauen und Y die Anzahl der Männer im ausgewählten Team.

(a) Was ist die bedingte Wahrscheinlichkeit, dass das Team nur aus Frauen besteht, gegeben, dass das Team mindestens eine Frau hat?

(b) Was ist die bedingte Wahrscheinlichkeit, dass das Team nur aus Frauen besteht, gegeben, dass Helga im Team ist?

(c) Finden Sie die gemeinsame Verteilung von (X, Y). Berechnen Sie $\mathbb{E}[X]$ und $\mathbb{E}[Y]$.

(d) Berechen Sie $\mathrm{Var}[X]$, $\mathrm{Var}[X+Y]$, $\mathrm{Cov}(X,Y)$ und $\mathrm{Corr}(X,Y)$.

Lösung:

(a) Für $x = 0, 1, 2, 3$ gibt es

$$\binom{3}{x}\binom{3}{3-x}, \tag{1}$$

mögliche Teams mit x Frauen und $y = 3 - x$ Männern. Es gibt also

$$\sum_{x=1}^{3}\binom{3}{x}\binom{3}{3-x} = \binom{3}{1}\binom{3}{2} + \binom{3}{2}\binom{3}{1} + \binom{3}{3}\binom{3}{0} = 3 \times 3 + 3 \times 3 + 1 = 19$$

mögliche Teams mit mindestens einer Frau. Genau eines von diesen besteht nur aus Frauen. Also ist die bedingte Wahrscheinlichkeit, dass das Team nur aus Frauen besteht, gegeben, dass das Team eine Frau hat, $\frac{1}{19}$.

Alternativ: Die Anzahl aller möglichen Teams beträgt

$$\binom{6}{3} = 20.$$

Also ist

$$\begin{aligned}
&\mathbb{P}[\text{“das Team besteht nur aus Frauen”}|\text{“das Team hat eine Frau”}]\\
&= \frac{\mathbb{P}[\text{“das Team besteht nur aus Frauen und das Team hat eine Frau”}]}{\mathbb{P}[\text{“das Team hat eine Frau”}]}\\
&= \frac{\mathbb{P}[\text{“das Team besteht nur aus Frauen”}]}{\mathbb{P}[\text{“das Team hat eine Frau”}]}\\
&= \frac{1/20}{19/20} = \frac{1}{19}.
\end{aligned}$$

(b) Für $x = 0, 1, 2$ gibt es

$$\binom{2}{x}\binom{3}{2-x},$$

mögliche Teams zu denen Helga, x zusätzliche Frauen und $2 - x$ Männer gehören. Also ist

$$\sum_{x=0}^{2}\binom{2}{x}\binom{3}{2-x} = \binom{2}{2}\binom{3}{0} + \binom{2}{1}\binom{3}{1} + \binom{2}{0}\binom{3}{2} = 1 + 2 \times 3 + 3 = 10$$

die Anzahl der Teams, zu denen Helga gehört. Genau eines von diesen besteht nur aus Frauen. Also ist die bedingte Wahrscheinlichkeit, dass das Team nur aus Frauen besteht, gegeben, dass Helga im Team ist, $1/10$.

Alternativ ist

$$\begin{aligned}
&\mathbb{P}[\text{“das Team besteht nur aus Frauen”}|\text{“Helga ist im Team”}]\\
&= \frac{\mathbb{P}[\text{“das Team besteht nur aus Frauen und Helga ist im Team”}]}{\mathbb{P}[\text{“Helga ist im Team”}]}\\
&= \frac{\mathbb{P}[\text{“das Team besteht nur aus Frauen”}]}{\mathbb{P}[\text{“Helga ist im Team”}]}\\
&= \frac{1/20}{10/20} = \frac{1}{10}.
\end{aligned}$$

(c) Selbverständlich ist $\mathbb{P}[X = x, Y = y] = 0$, wenn $y \neq 3 - x$. Von (1) sehen wir, dass für $y = 3 - x$ gilt

$$\mathbb{P}\left[X = x, Y = y\right] = \frac{\binom{3}{x}\binom{3}{y}}{\sum_{k=0}^{3}\binom{3}{k}\binom{3}{3-k}} = \frac{\binom{3}{x}\binom{3}{y}}{\binom{6}{3}} = \frac{\binom{3}{x}\binom{3}{y}}{20}.$$

Also ist

$$\begin{aligned} \mathbb{P}[X = 0, Y = 3] &= \frac{1}{20}, \qquad \mathbb{P}[X = 1, Y = 2] = \frac{9}{20}, \\ \mathbb{P}[X = 2, Y = 1] &= \frac{9}{20}, \qquad \mathbb{P}[X = 3, Y = 0] = \frac{1}{20}. \end{aligned} \tag{2}$$

Wir haben

$$\mathbb{E}[X] = \sum_{x=0}^{3} x\mathbb{P}[X = x] = \sum_{x=1}^{3} x\mathbb{P}[X = x, Y = 3 - x] = \frac{9}{20} + 2 \times \frac{9}{20} + 3 \times \frac{1}{20} = \frac{30}{20} = \frac{3}{2}.$$

Da X und Y offensichtlich die gleiche Verteilung haben (Symmetrie), ist auch $\mathbb{E}[Y] = \frac{3}{2}$. Alternativ ist $Y = 3 - X$. Also gilt $\mathbb{E}[Y] = 3 - \mathbb{E}[X] = 3 - \frac{3}{2} = \frac{3}{2}$.

(d) Wir haben

$$\mathbb{E}[X^2] = \sum_{x=0}^{3} x^2\mathbb{P}[X = x] = \sum_{x=1}^{3} x^2\mathbb{P}[X = x, Y = 3 - x] = \frac{9}{20} + 4 \times \frac{9}{20} + 9 \times \frac{1}{20} = \frac{54}{20}.$$

Also ist

$$\mathrm{Var}[X] = \frac{54}{20} - \left(\frac{3}{2}\right)^2 = \frac{9}{20}.$$

Offensichtlich haben wir immer $X + Y = 3$, und deshalb ist

$$\mathrm{Var}[X + Y] = \mathrm{Var}[3] = 0.$$

Schliesslich sehen wir aus der Formel

$$\mathrm{Var}[X + Y] = \mathrm{Var}[X] + 2\mathrm{Cov}(X, Y) + \mathrm{Var}[Y],$$

dass

$$\mathrm{Cov}(X, Y) = \frac{\mathrm{Var}[X + Y] - \mathrm{Var}[X] - \mathrm{Var}[Y]}{2} = -\frac{\mathrm{Var}[X] + \mathrm{Var}[Y]}{2}.$$

Da X und Y dieselbe Verteilung haben, ist

$$\mathrm{Var}[Y] = \mathrm{Var}[X] = \frac{9}{20} \quad \text{und also} \quad \mathrm{Cov}(X, Y) = -\mathrm{Var}[X] = -\frac{9}{20}.$$

Alternativ haben wir mit Hilfe von (2)

$$\begin{aligned} \mathbb{E}[XY] &= \sum_{x=0}^{3} x(3 - x)\mathbb{P}[X = x, Y = 3 - x] = 0 \times 3 \times \frac{1}{20} + 1 \times 2 \times \frac{9}{20} + 2 \times 1 \times \frac{9}{20} + 3 \times 0 \times \frac{1}{20} \\ &= 4 \times \frac{9}{20} = \frac{9}{5}. \end{aligned}$$

Folglich ist

$$\operatorname{Cov}(X,Y) = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y] = \frac{9}{5} - \frac{3}{2} \times \frac{3}{2} = -\frac{9}{20} = -\operatorname{Var}[X].$$

Schliesslich ist

$$\operatorname{Corr}(X,Y) = \frac{\operatorname{Cov}(X,Y)}{\sqrt{\operatorname{Var}[X]}\sqrt{\operatorname{Var}[Y]}} = -\frac{\operatorname{Var}[X]}{\operatorname{Var}[X]} = -1.$$

Mit anderen Worten sind X und Y perfekt (negativ) korreliert. Dies kann bereits an der Formel $X = 3 - Y$ gesehen werden.

Aufgabe 12-5. Seien X und Y unabhängige Zufallsvariablen mit X uniform verteilt auf $[0,1]$ und Y exponentialverteilt mit Parameter 1. Seien $U = X + Y$ und $V = XY$.

(a) Berechnen Sie $E\left[\frac{V}{X^2+1}\right]$.

(b) Bestimmen Sie die Verteilungsfunktion sowie die Dichtefunktion von U.

(c) Berechnen Sie die beste lineare Prognose von V durch U.

Lösung:

(a) Wir haben wegen der Unabhängigkeit und $\mathbb{E}[Y] = 1$, dass

$$\mathbb{E}\left[\frac{V}{X^2+1}\right] = \mathbb{E}\left[Y\frac{X}{X^2+1}\right] = \mathbb{E}[Y]\mathbb{E}\left[\frac{X}{X^2+1}\right] = \int_0^1 \frac{x}{x^2+1}dx = \frac{1}{2}\log(x^2+1)\Big|_{x=0}^1 = \frac{\log 2}{2}.$$

(b) Wegen der Unabhängigkeit ist die gemeinsame Dichte von (X,Y)

$$f_{X,Y}(x,y) = f_X(x)f_Y(y) = \mathbf{1}_{[0,1]}(x)e^{-y}\mathbf{1}_{[0,\infty)}(y).$$

Es gilt $\mathbb{P}[U < u] = 0$ für $u < 0$. Für $u \geq 0$ gilt

$$\begin{aligned}
\mathbb{P}[U \leq u] &= \iint f_{X,Y}(x,y)\mathbf{1}_{(-\infty,u]}(x+y)dxdy \\
&= \int_0^{\min\{u,1\}} \int_0^{u-x} e^{-y}dydx \\
&= \int_0^{\min\{u,1\}} (1 - e^{-(u-x)})dx \\
&= \int_0^{\min\{u,1\}} (1 - e^x e^{-u})dx.
\end{aligned}$$

Für $u > 1$ gilt

$$\mathbb{P}[U \leq u] = \int_0^1 1 - e^x e^{-u}dx = 1 - e^{-u}(e-1).$$

Für $0 \leq u \leq 1$ gilt

$$\mathbb{P}[U \leq u] = \int_0^u 1 - e^x e^{-u}dx = u - e^{-u}\left(e^u - 1\right).$$

Also ist

$$\mathbb{P}[U \le u] = \begin{cases} 0, & \text{falls } u < 0, \\ u + e^{-u} - 1, & \text{falls } 0 \le u < 1, \\ 1 - e^{-u}(e-1), & \text{sonst.} \end{cases}$$

In kompakter Form ist das

$$\mathbb{P}[U \le u] = \mathbf{1}_{[0,\infty)}(u)(\min(1,u) - e^{-u}(e^{\min(1,u)} - 1)).$$

Wir können die Dichte f_U durch Differentiation finden:

$$f_U(u) = \begin{cases} 0, & \text{falls } u < 0, \\ 1 - e^{-u}, & \text{falls } 0 \le u < 1, \\ e^{-u}(e-1), & \text{sonst.} \end{cases}$$

(c) Die beste lineare Prognose von V durch U ist gegeben durch

$$\begin{aligned} Z &= \frac{\mathrm{Cov}(V,U)}{\mathrm{Var}[U]}(U - \mathbb{E}[U]) + \mathbb{E}[V] \\ &= \frac{\mathrm{Cov}(V,U)}{\mathrm{Var}[X+Y]}(U - \mathbb{E}[X+Y]) + \mathbb{E}[XY] \\ &= \frac{\mathrm{Cov}(V,U)}{\mathrm{Var}[X] + \mathrm{Var}[Y]}(U - \mathbb{E}[X] - \mathbb{E}[Y]) + \mathbb{E}[X]\mathbb{E}[Y], \end{aligned}$$

wobei wir die Unabhängigkeit von X und Y benutzt haben. Es gilt $\mathbb{E}[X] = 1/2$, $\mathrm{Var}[X] = 1/12$ und $\mathbb{E}[Y] = \mathrm{Var}[Y] = 1$. Ferner ist wegen Unabhängigkeit

$$\begin{aligned} \mathrm{Cov}(V,U) &= \mathbb{E}[VU] - \mathbb{E}[V]\mathbb{E}[U] \\ &= \mathbb{E}[X^2Y + XY^2] - \mathbb{E}[XY]\mathbb{E}[X+Y] \\ &= \mathbb{E}[X^2Y] + \mathbb{E}[XY^2] - \mathbb{E}[XY]\mathbb{E}[X] - \mathbb{E}[XY]\mathbb{E}[Y] \\ &= \mathbb{E}[X^2]\mathbb{E}[Y] + \mathbb{E}[X]\mathbb{E}[Y^2] - \mathbb{E}[Y](\mathbb{E}[X])^2 - \mathbb{E}[X](\mathbb{E}[Y])^2 \\ &= \mathbb{E}[Y]\big(\mathbb{E}[X^2] - (\mathbb{E}[X])^2\big) + \mathbb{E}[X]\big(\mathbb{E}[Y^2] - (\mathbb{E}[Y])^2\big) \\ &= \mathbb{E}[Y]\mathrm{Var}[X] + \mathbb{E}[X]\mathrm{Var}[Y]. \end{aligned}$$

Also erhalten wir

$$\begin{aligned} Z &= \frac{\mathbb{E}[Y]\mathrm{Var}[X] + \mathbb{E}[X]\mathrm{Var}[Y]}{\mathrm{Var}[X] + \mathrm{Var}[Y]}(U - \mathbb{E}[X] - \mathbb{E}[Y]) + \mathbb{E}[X]\mathbb{E}[Y] \\ &= \frac{1/12 + 1/2}{1/12 + 1}\Big(U - \frac{3}{2}\Big) + \frac{1}{2} \\ &= \frac{7}{13}\Big(U - \frac{3}{2}\Big) + \frac{1}{2} \\ &= \frac{7}{13}U - \frac{4}{13}. \end{aligned}$$

Quantiltabelle der $\mathcal{N}(0,1)$-Verteilung

0.5	0.75	0.9	0.95	0.975	0.99	0.995	0.999
0	0.6745	1.2816	1.6449	1.9600	2.3263	2.5758	3.0902

Zum Beispiel ist $\Phi^{-1}(0.9) = 1.2816$.