Einführung in die Wahrscheinlichkeitstheorie und Statistik
Wintersemenster 2022/23
Dozent: Prof. Dr. Rainer Dahlhaus
Assistenz: Marilena Müller

UNIVERSITÄT HEIDELBERG
ZUKUNFT SEIT 1386

Lösungsskizzen zu Übungsblatt 7

Aufgabe 1

Sei $(\Omega, \mathcal{A}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und $X, Y : \Omega \to \mathbb{R}$ zwei stetig verteilte Zufallsvariablen, deren gemeinsame Dichte durch

$$f_{X,Y}(x,y) = c \cdot xy \cdot \mathbb{1}_{\{0 \leq x \leq y \leq 2\}} = \begin{cases} c \cdot xy, & 0 \leq x \leq y \leq 2, \\ 0, & \text{sonst} \end{cases}$$

gegeben ist.

(i) Bestimmen Sie $c > 0$, so dass $f_{X,Y}$ eine Wahrscheinlichkeitsdichte ist.

(ii) Berechnen Sie die Randdichten f_X und f_Y von X bzw. Y.

(iii) Berechnen Sie die Wahrscheinlichkeiten $\mathbb{P}(X + Y \leq 1)$ und $\mathbb{P}(2X \geq Y)$.

(iv) Bestimmen Sie $\mathbb{E}[X \cdot Y]$.

(v) Sind X, Y stochastisch unabhängig?

(1+1+2+1+1 Punkte)

Beweis. (i) $f_{X,Y} \geq 0$ ist offensichtlich erfüllt. Es bleibt die Normierungskonstante zu berechnen..

$$\begin{aligned} 1 &= \int_{\mathbb{R}^2} f_{X,Y}(x,y) \mathrm{d}(x,y) = c \cdot \int_{\mathbb{R}} \int_{\mathbb{R}} xy \cdot \mathbb{1}_{\{0 \leq x \leq y \leq 2\}} \mathrm{d}y \mathrm{d}x \\ &= c \cdot \int_0^2 \int_x^2 xy \mathrm{d}y \mathrm{d}x = c \cdot \int_0^2 x \cdot \left[\frac{1}{2} y^2\right]_x^2 \mathrm{d}x = c \cdot \int_0^2 2x - \frac{1}{2} x^3 \mathrm{d}x \\ &= c \cdot \left[x^2 - \frac{1}{8} x^4\right]_0^2 = c \cdot (2 - 0) = 2c. \end{aligned}$$

Es muss also $c = \frac{1}{2}$ gelten.

(ii) Für die Berechnung der Randdichten müssen wir jeweils nach der anderen Variable integrieren.

$$\begin{aligned}
f_X(x) &= \int_{\mathbb{R}} f_{X,Y}(x,y)\mathrm{d}y = \frac{1}{2}\int_{\mathbb{R}} xy \cdot \mathbb{1}_{\{0\le x\le y\le 2\}}(x,y)\mathrm{d}y \\
&= \frac{1}{2}\int_{\mathbb{R}} xy \cdot \mathbb{1}_{\{0\le x\le 2\}}(x) \cdot \mathbb{1}_{\{x\le y\le 2\}}(x,y)\mathrm{d}y \\
&= \frac{1}{2}\mathbb{1}_{\{0\le x\le 2\}}(x) \cdot x \cdot \int_x^2 y\mathrm{d}y = \frac{1}{2}\mathbb{1}_{\{0\le x\le 2\}}(x) \cdot x \cdot \left(2 - \frac{1}{2}x^2\right) \\
&= \mathbb{1}_{\{0\le x\le 2\}}(x)\left(x - \frac{1}{4}x^3\right), \\
f_Y(y) &= \int_{\mathbb{R}} f_{X,Y}(x,y)\mathrm{d}x = \frac{1}{2}\int_{\mathbb{R}} xy \cdot \mathbb{1}_{\{0\le x\le y\le 2\}}(x,y)\mathrm{d}x \\
&= \frac{1}{2}\int_{\mathbb{R}} xy \cdot \mathbb{1}_{\{0\le y\le 2\}}(y) \cdot \mathbb{1}_{\{0\le x\le y\}}(x,y)\mathrm{d}x \\
&= \frac{1}{2}\mathbb{1}_{\{0\le y\le 2\}}(y) \cdot y \cdot \int_0^y x\mathrm{d}x \\
&= \mathbb{1}_{\{0\le y\le 2\}}(y) \cdot \frac{1}{4}y^3.
\end{aligned}$$

(iii) Wir rechnen

$$\mathbb{P}(X+Y\le 1) = \mathbb{E}[\mathbb{1}_{\{X+Y\le 1\}}] = \int_{\mathbb{R}^2} \mathbb{1}_{\{x+y\le 1\}} \cdot f_{X,Y}(x,y)\mathrm{d}(x,y).$$

Wir bemerken, das ist das gleiche wie folgender Ansatz.

$$\mathbb{P}(X+Y\le 1) = \mathbb{P}((X,Y) \in \{(x,y)\in\mathbb{R}^2 : x+y\le 1\}) = \int_{\{(x,y)\in\mathbb{R}^2: x+y\le 1\}} f_{X,Y}(x,y)\mathrm{d}(x,y).$$

Wir bemerken, dass für festes $x \in [0,2]$ folgende Bedingung erfüllt sein muss.

$$x \le y \le 2, y \le 1-x \quad \Leftrightarrow \quad x\le y\le 1-x$$

Die Bedingung ist also nur nicht leer, wenn $x \le \frac{1}{2}$. Daher wird tatsächlich über $x \in [0, \frac{1}{2}]$ integriert. Es folgt

$$\begin{aligned}
\mathbb{P}(X+Y\le 1) &= \frac{1}{2}\int_{\mathbb{R}}\int_{\mathbb{R}} x\cdot y \cdot \mathbb{1}_{\{0\le x\le y\le 2\}}(x,y)\mathbb{1}_{\{x+y\le 1\}}(x,y)\mathrm{d}y\mathrm{d}x \\
&= \frac{1}{2}\int_0^{\frac{1}{2}} x \cdot \int_x^{1-x} y\mathrm{d}y\mathrm{d}x = \frac{1}{4}\int_0^{\frac{1}{2}} x\cdot((1-x)^2 - x^2)\mathrm{d}x \\
&= \frac{1}{4}\int_0^{\frac{1}{2}} x - 2x^2\mathrm{d}x = \frac{1}{4}\left[\frac{1}{2}x^2 - \frac{2}{3}x^3\right]_0^{\frac{1}{2}} = \frac{1}{96}.
\end{aligned}$$

Gleiches Vorgehen bei der zweiten Wahrscheinlichkeit liefert die äquivalenten Bedingungen $0\le y\le 2$ und $\frac{y}{2}\le x\le y$. Daher

$$\mathbb{P}(2X\ge Y) = \frac{1}{2}\int_0^2 y\cdot\int_{\frac{y}{2}}^y x\mathrm{d}x\mathrm{d}y = \frac{1}{4}\int_0^2 y\cdot(y^2 - \frac{1}{4}y^2)\mathrm{d}y = \frac{3}{16}\int_0^2 y^3\mathrm{d}y = \frac{3}{4}.$$

(iv) Es ist

$$\begin{aligned}
\mathbb{E}[X\cdot Y] &= \int_{\mathbb{R}^2} xyf_{X,Y}(x,y)\mathrm{d}(x,y) = \frac{1}{2}x^2\int_0^2\int_x^2 y^2\mathrm{d}y\mathrm{d}x = \frac{1}{6}\int_0^2 x^2\cdot(8-x^3)\mathrm{d}x \\
&= \frac{1}{6}\left[\frac{8}{3}x^3 - \frac{1}{6}x^6\right]_0^2 = \frac{16}{9}.
\end{aligned}$$

(v) X, Y sind nicht stochastisch unabhängig. Wir betrachten, ob $f_{X,Y}(x,y) = f_X(x) \cdot f_Y(y)$ überall außer auf einer Lebesgue-Nullmenge gilt. Konkret ist $f_{X,Y}$ auf dem gesamten Bereich $\{(x,y) \in [0,2]^2 : x > y\}$ Null, die rechte Seite $f_X \cdot f_Y$ aber nicht. Man kann zum Beispiel

$$\mathbb{P}(X \in [1,2], Y \in [0,1]) \neq \mathbb{P}(X \in [1,2]) \cdot \mathbb{P}(Y \in [0,1])$$

zeigen, da

$$\mathbb{P}(X \in [1,2], Y \in [0,1]) \leq \mathbb{P}(X \geq Y) = 0,$$

aber

$$\mathbb{P}(X \in [1,2]) = \int_1^2 f_X(x)\mathrm{d}x \neq 0, \quad \mathbb{P}(Y \in [0,1]) = \int_0^1 f_Y(y)\mathrm{d}y \neq 0.$$

$\square$

Aufgabe 2

Sei $(\Omega, \mathcal{A}, \mathbb{P})$ ein Wahrscheinlichkeitsraum und $X, Y : \Omega \to \mathbb{R}$ zwei stochastisch unabhängige und stetig verteilte Zufallsvariablen.

(i) Sei $Z := X + Y$. Zeigen Sie, dass Z ebenfalls stetig verteilt ist mit Wahrscheinlichkeitsdichte

$$f_Z(z) = \int_{\mathbb{R}} f_X(z-y) f_Y(y)\mathrm{d}y.$$

Z wird auch als *Faltung* von X, Y bezeichnet.

(ii) Seien $X, Y \sim \mathcal{E}(\lambda)$ exponentialverteilt mit Parameter $\lambda > 0$. Berechnen Sie die Verteilung von $X + Y$.

(iii) Seien nun $X \sim \mathcal{N}(\mu_1, \sigma_1^2)$ und $Y \sim \mathcal{N}(\mu_2, \sigma_2^2)$ mit $\mu_1, \mu_2 \in \mathbb{R}$ und $\sigma_1, \sigma_2 > 0$. Zeigen Sie, dass

$$X + Y \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2)$$

(iv) Es bezeichne nun f_X bzw. f_Y die Wahrscheinlichkeitsdichten von X bzw. Y. Zeigen Sie, dass die Dichte von $Z := \frac{X}{Y}$ gegeben ist durch

$$f_Z(z) = \int_{\mathbb{R}} |y| \cdot f_X(z \cdot y) \cdot f_Y(y)\mathrm{d}y.$$

Sei nun $X, Y \sim \mathcal{N}(0,1)$ standardnormalverteilt. Berechnen Sie die Dichte von Z.

(2+1+2+2 Punkte)

Beweis. (i) Da X, Y unabhängig sind, gilt für alle $x, y \in \mathbb{R}$: $f_{X,Y}(x,y) = f_X(x) \cdot f_Y(y)$. Es gilt für alle $z \in \mathbb{R}$

$$\begin{aligned} F_Z(z) &= \mathbb{P}(Z \leq z) = \mathbb{P}(X + Y \leq z) = \mathbb{P}((X,Y) \in \{(x,y) \in \mathbb{R}^2 : x + y \leq z\}) \\ &= \int_{\{(x,y)\in\mathbb{R}^2 : x+y \leq z\}} f_{X,Y}(x,y)\mathrm{d}(x,y) = \int_{\mathbb{R}} \int_{\mathbb{R}} \mathbb{1}_{\{x+y\leq z\}}(x,y,z) f_X(x) f_Y(y)\mathrm{d}y\mathrm{d}x \\ &= \int_{-\infty}^{z} \int_{\mathbb{R}} f_X(v-y) f_Y(y)\mathrm{d}y\mathrm{d}v, \end{aligned}$$

wobei hier eine Substitution mit $v = x + y$ durchgeführt wurde. Nach der Definition einer Wahrscheinlichkeitsdichte muss daher

$$f_Z(z) = \int_{\mathbb{R}} f_X(z-y) f_Y(y)\mathrm{d}y$$

gelten.

(ii) Wir nutzen die Formel für die Dichte einer Faltung aus (i) und definieren $Z := X+Y$.

$$\begin{aligned} f_Z(z) &= \int_{\mathbb{R}} f_X(z-y) f_Y(y) \mathrm{d}y = \int_{\mathbb{R}} \lambda e^{-\lambda(z-y)} \mathbb{1}_{\{z-y\geq 0\}}(z,y) \cdot \lambda e^{-\lambda y} \mathbb{1}_{\{y \geq 0\}}(y) \mathrm{d}y \\ &= \lambda^2 \cdot \int_0^z e^{-\lambda z}(z) \mathrm{d}y \cdot \mathbb{1}_{\{z\geq 0\}}(z) = \lambda^2 z e^{-\lambda z} \cdot \mathbb{1}_{\{z \geq 0\}}(z). \end{aligned}$$

(iii) Nach dem Satz über die lineare Transformation von Zufallsvariablen und die konkrete Anwendung in Bemerkung 6.15 (Spezialfall) gilt für eine normalverteilte Zufallsvariable $W \sim N(\mu, \sigma^2)$: $aW + b \sim N(a\mu + b, a^2\sigma^2)$. Betrachte die Darstellung

$$X + Y = \sigma_1 \left(\frac{X - \mu_1}{\sigma_1} + \frac{Y - \mu_2}{\sigma_1} \right) + (\mu_1 + \mu_2).$$

Die Zufallsvariablen $X' := \frac{X-\mu_1}{\sigma_1} \sim N(0,1)$, $Y' := \frac{Y-\mu_2}{\sigma_1} \sim N(0, \frac{\sigma_2^2}{\sigma_1^2})$ sind laut Satz 8.15 (Funktionen von unabhängigen Zufallsvariablen sind immer noch unabhängig) immer noch unabhängig. Haben wir nun

$$X' + Y' \sim N\left(0, 1 + \frac{\sigma_2^2}{\sigma_1^2}\right) \tag{1}$$

gezeigt, so folgt wieder mit der linearen Transformation 6.15, dass

$$X + Y = \sigma_1(X' + Y') + (\mu_1 + \mu_2) \sim N(\mu_1 + \mu_2, \sigma_1^2 + \sigma_2^2).$$

Zum Beweis von (1) nutzen wir die Formel aus (i) und die Abkürzung $\sigma := \frac{\sigma_2}{\sigma_1}$: Hier ist $f_{X'}(x) = \frac{1}{\sqrt{2\pi}} \exp(-x^2/2)$ und $f_Y(y) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-x^2/(2\sigma^2))$, also für $Z = X' + Y'$

$$f_Z(z) = \int_{\mathbb{R}} f_X(z-y) f_Y(y) \mathrm{d}y = \frac{1}{2\pi\sigma} \int_{\mathbb{R}} \exp\left(-\frac{(z-y)^2}{2}\right) \cdot \exp\left(-\frac{y^2}{2\sigma^2}\right) \mathrm{d}y.$$

Beachte

$$\begin{aligned} \frac{y^2}{\sigma^2} + (z-y)^2 &= \frac{y^2}{\sigma^2} + z^2 - 2yz + y^2 \\ &= \left(1 + \frac{1}{\sigma^2}\right) \cdot \left(y^2 - 2y \cdot \frac{z}{1 + \frac{1}{\sigma^2}} + \left(\frac{z}{1+\frac{1}{\sigma^2}}\right)^2\right) + z^2 \left(1 - \frac{1}{1 + \frac{1}{\sigma^2}}\right) \\ &= \left(1 + \frac{1}{\sigma^2}\right) \cdot \left(y - \frac{z}{1 + \frac{1}{\sigma^2}}\right)^2 + \frac{z^2}{1 + \sigma^2}. \end{aligned}$$

Dadurch erhalten wir

$$\begin{aligned} f_Z(z) = \frac{(1 + \frac{1}{\sigma^2})^{-1}}{\sqrt{2\pi\sigma^2}} &\underbrace{\int_{\mathbb{R}} \frac{1}{\sqrt{2\pi(1+\frac{1}{\sigma^2})^{-2}}} \exp\left(-\frac{1}{2(1+\frac{1}{\sigma^2})^{-1}} \left(y - \frac{z}{1+\frac{1}{\sigma^2}}\right)^2\right) \mathrm{d}y}_{=1} \\ &\cdot \exp\left(-\frac{z^2}{2(1+\sigma^2)}\right) \\ = &\frac{1}{\sqrt{2\pi(1+\sigma^2)}} \exp\left(-\frac{z^2}{1+\sigma^2}\right). \end{aligned}$$

Das ist die Dichte einer $N(0, 1 + \sigma^2)$-Verteilung.

(iv) Wir berechnen zunächst die Verteilungsfunktion von Z.

$$F_Z(z) = \mathbb{P}(Z \le z) = \mathbb{P}\left(\frac{X}{Y} \le z\right) = \int_{\mathbb{R}^2} \mathbb{1}_{\{\frac{x}{y} \le z\}}(x,y,z) \cdot f_{X,Y}(x,y)\mathrm{d}(x,y)$$
$$= \int_{\mathbb{R}} \int_{\mathbb{R}} \mathbb{1}_{\{\frac{x}{y} \le z\}}(x,y,z) f_X(x) \cdot f_Y(y)\mathrm{d}x\mathrm{d}y$$

Wir wollen $v = v(x) := \frac{x}{y}$ substituieren, d.h. $\frac{dv}{dx} = v'(x) = \frac{1}{y}$ genau dann, wenn $ydv = dx$. Diese müssen für $y > 0$ und $y < 0$ separat durchgeführt werden.

$$\begin{aligned} F_Z(z) = &\int_0^\infty \int_{-\infty}^\infty \mathbb{1}_{\{v \le z\}}(v,z) y f_X(vy) \cdot f_Y(y)\mathrm{d}v\mathrm{d}y \\ &+ \int_{-\infty}^0 \int_\infty^{-\infty} y\mathbb{1}_{\{v \le z\}}(v,z) y f_X(vy) \cdot f_Y(y)\mathrm{d}v\mathrm{d}y \\ = &\int_{-\infty}^z \int_{\mathbb{R}} |y| f_X(vy) f_Y(y)\mathrm{d}y\mathrm{d}v. \end{aligned}$$

Das zeigt, dass die Dichte f_Z von Z gegeben ist durch

$$f_Z(z) = \int_{\mathbb{R}} |y| f_X(zy) f_Y(y)\mathrm{d}y.$$

Für $X, Y \sim N(0,1)$ ist nun $f_X(x) = f_Y(x) = \frac{1}{\sqrt{2\pi}} \exp(-x^2/2)$. Wir erhalten

$$\begin{aligned} f_Z(z) &= \frac{1}{2\pi} \int_{\mathbb{R}} |y| \exp(-(zy)^2/2) \exp(-y^2/2)\mathrm{d}y \\ &= \frac{1}{\pi} \int_0^\infty y \exp\left(-\frac{y^2}{2} \cdot (1+z^2)\right) \mathrm{d}y \\ &\overset{w:=\frac{y^2}{2}}{=} \frac{1}{\pi} \int_0^\infty \exp(-w \cdot (1+z^2))\mathrm{d}w \\ &= \frac{1}{\pi}\Big[-\frac{1}{1+z^2} e^{-w\cdot(1+z^2)}\Big]_0^\infty = \frac{1}{\pi}\frac{1}{1+z^2}. \end{aligned}$$

Dies ist die Dichte einer sogenannten Cauchy-Verteilung.

□

Aufgabe 3

Ein sechsseitiger, fairer Würfel wird zweimal (unabhängig voneinander) geworfen. Seien X, Y Zufallsvariablen, welche die zwei Ergebnisse der Würfelwürfe angeben.

(i) Geben Sie die gemeinsame Zähldichte $p_{X,Y}(x,y)$ für die Zufallsvariablen X, Y an.

(ii) Wir definieren $M := \max\{X, Y\}$. Berechnen Sie die gemeinsame Zähldichte $p_{X,M}(x,m)$ von X, M.

(iii) Berechnen Sie die Randdichte $p_M(m)$ von M.

(iv) Berechnen Sie die Kovarianz $\mathrm{Cov}(X,M) := \mathbb{E}[X \cdot M] - \mathbb{E}[X] \cdot \mathbb{E}[M]$.

(2+1+1+1 Punkte)

Beweis. (i) Da X_1, X_2 unabhängig sind, gilt

$$p_{X,Y}(x,y) = \mathbb{P}(X = x, Y = y) = \mathbb{P}(X = x) \cdot \mathbb{P}(Y = y).$$

Aus der Aufgabe wissen wir, dass mit $A := \{1,2,3,4,5,6\}$ für $x \in \mathbb{N}_0$ gilt:

$$\mathbb{P}(X = x) = \mathbb{P}(Y = x) = \frac{1}{6} \cdot \mathbb{1}_A(x).$$

Damit erhalten wir:

$$p_{X,Y}(x,y) = \frac{1}{36} \cdot \mathbb{1}_A(x) \cdot \mathbb{1}_A(y).$$

(ii) Wir berechnen für $x, m \in \mathbb{N}_0$:

$$\begin{aligned}
&p_{X,M}(x,m) \\
&= \mathbb{P}(X = x, M = m) = \mathbb{P}(X = x, \max(X,Y) = m) \\
&= \mathbb{1}_{\{x=m\}}(x,m) \cdot \mathbb{P}(X = x, Y \le m) + \mathbb{1}_{\{x<m\}}(x,m) \cdot \mathbb{P}(X = x, Y = m) \\
&= \mathbb{1}_{\{x=m\}}(x,m) \cdot \underbrace{\mathbb{P}(X = x) \cdot \mathbb{P}(Y \le m)}_{=\frac{m}{36}\mathbb{1}_A(x)\mathbb{1}_A(m)} + \mathbb{1}_{\{x<m\}}(x,m) \cdot \underbrace{\mathbb{P}(X = x) \cdot \mathbb{P}(Y = m)}_{=\frac{1}{36}\mathbb{1}_A(x)\mathbb{1}_A(m)} \\
&= \frac{1}{36} \cdot \Big(\mathbb{1}_{\{x=m\}}(x,m) \cdot m + \mathbb{1}_{\{x<m\}}(x,m)\Big) \cdot \mathbb{1}_A(x) \cdot \mathbb{1}_A(m).
\end{aligned}$$

(iii) Die Randdichte p_M erhalten wir aus der gemeinsamen Dichte. Für $m \in A$ gilt:

$$\begin{aligned}
p_M(m) &= \sum_{x=0}^{\infty} p_{X,M}(x,m) \\
&= \mathbb{1}_A(m) \cdot \frac{1}{36} \cdot \left(m \cdot \sum_{x=1}^{6} \mathbb{1}_{\{x=m\}} + \sum_{x=1}^{6} \mathbb{1}_{\{x<m\}}\right) \\
&= \mathbb{1}_A(m) \cdot \frac{1}{36} \cdot \Big(m + (m-1)\Big) \\
&= \mathbb{1}_A(m) \cdot \frac{2m-1}{36}.
\end{aligned}$$

(iv) Offensichtlich gilt

$$\begin{aligned}
\mathbb{E}[X] &= \sum_{x=0}^{\infty} x \cdot p_X(x) = \frac{1}{6}\sum_{x=1}^{6} 1 = \frac{21}{6} = \frac{7}{2} = 3.5, \\
\mathbb{E}[M] &= \sum_{m=0}^{\infty} m \cdot p_M(m) \overset{(c)}{=} \frac{1}{36}\sum_{m=1}^{6} m \cdot (2m-1) = \frac{161}{36} \approx 4.47.
\end{aligned}$$

Es bleibt $\mathbb{E}[X \cdot M]$ zu berechnen.

$$\begin{aligned}
\mathbb{E}[X \cdot M] &= \sum_{x,m=0}^{\infty} x \cdot m \cdot p_{X,M}(x,m) \\
&= \frac{1}{36}\left(\sum_{x,m=1}^{6} x \cdot m^2 \cdot \mathbb{1}_{\{x=m\}} + \sum_{x,m=1}^{6} x \cdot m \cdot \mathbb{1}_{\{x<m\}}\right) \\
&= \frac{1}{36}\left(\sum_{m=1}^{6} m^3 + \sum_{m=1}^{6} m \cdot \sum_{x=1}^{m-1} x\right) \\
&= \frac{1}{36}\sum_{m=1}^{6}\left(m^3 + \frac{1}{2}m^2(m-1)\right) \\
&= \frac{616}{36} = \frac{154}{9} \approx 17.11.
\end{aligned}$$

Damit haben wir insgesamt für die Kovarianz

$$\mathrm{Kov}(X, M) = \mathbb{E}[XM] - \mathbb{E}X\mathbb{E}M \approx 17.11 - 4.47 \cdot 3.5 \approx 1.47.$$

$\square$

Aufgabe 4

Sei $(\Omega, \mathcal{A}, \mathbb{P})$ ein Wahrscheinlichkeitsraum, $n \in \mathbb{N}$ und $X_1, ..., X_n : \Omega \to \mathbb{R}$ unabhängig und identisch verteilte Zufallsvariablen mit $\mathbb{E}[X_1] = \mu$, $\mathrm{Var}(X_1) = \sigma^2$ und Verteilungsfunktion F. Seien für $t \in \mathbb{R}$

$$\overline{X_n} := \frac{1}{n}\sum_{i=1}^n X_i, \qquad S_n^2 := \frac{1}{n-1}\sum_{i=1}^n (X_i - \overline{X_n})^2, \qquad \hat{F}_n(t) := \frac{1}{n}\sum_{i=1}^n \mathbb{1}_{\{X_i \le t\}}$$

der Stichprobenmittelwert, die Stichprobenvarianz und die empirische Verteilungsfunktion.

(i) Zeigen Sie, dass $S_n^2 = \frac{n}{n-1}\left(\left(\frac{1}{n}\sum_{i=1}^n X_i^2\right) - \overline{X_n}^2\right)$. Berechnen Sie $\mathbb{E}[S_n^2]$ in Termen von n, μ, σ^2.

(ii) Berechnen Sie die Verteilung von $n \cdot \hat{F}_n(t)$ und geben Sie dann $\mathbb{E}[\hat{F}_n(t)]$ und $\mathrm{Var}(\hat{F}_n(t))$ in Termen von $n, F(t)$ an.

(2+2 Punkte)

Beweis. (i) Wir berechnen

$$\begin{aligned} S_n^2 &= \frac{1}{n-1}\sum_{i=1}^n \left(X_i^2 - 2X_i\overline{X_n} + \overline{X_n}^2\right) \\ &= \frac{1}{n-1}\left(\sum_{i=1}^n X_i^2 - 2\overline{X_n} \cdot \sum_{i=1}^n X_i + \sum_{i=1}^n \overline{X_n}^2\right) \\ &= \frac{n}{n-1}\left(\frac{1}{n}\sum_{i=1}^n X_i^2 - 2\overline{X_n}^2 + \overline{X_n}^2\right) \\ &= \frac{n}{n-1}\left(\frac{1}{n}\sum_{i=1}^n X_i^2 - \overline{X_n}^2\right). \end{aligned}$$

Wir berechnen nun den Erwartungswert der einzelnen Summanden. Es ist wegen $\mathbb{E}[X_i^2] = \mathrm{Var}(X_i) + \mathbb{E}[X_i]^2 = \sigma^2 + \mu^2$

$$\mathbb{E}\left[\frac{1}{n}\sum_{i=1}^n X_i^2\right] = \frac{1}{n}\sum_{i=1}^n \mathbb{E}[X_i^2] = \frac{1}{n}\sum_{i=1}^n (\sigma^2 + \mu^2) = \sigma^2 + \mu^2$$

und

$$\begin{aligned} \mathbb{E}\left[\overline{X_n}^2\right] = \mathbb{E}\left[\frac{1}{n^2}\sum_{i,j=1}^n X_iX_j\right] &= \frac{1}{n^2}\left[\sum_{i=1}^n \mathbb{E}[X_i^2] + \sum_{\substack{i,j=1 \\ i \neq j}}^n \underbrace{\mathbb{E}[X_iX_j]}_{=\mu^2}\right] \\ &= \frac{1}{n^2}\Big[n \cdot (\sigma^2 + \mu^2) + n \cdot (n-1) \cdot \mu^2\Big] \\ &= \frac{1}{n}\Big[(\sigma^2 + \mu^2) + (n-1) \cdot \mu^2\Big]. \end{aligned}$$

Eingesetzt in die Darstellung für S_n^2 ergibt sich:

$$\begin{aligned}\mathbb{E}[S_n^2] &= \frac{n}{n-1}\Big((\sigma^2+\mu^2) - \frac{1}{n}\Big((\sigma^2+\mu^2) + (n-1)\mu^2\Big)\Big) \\ &= \frac{n}{n-1}\Big(\frac{n-1}{n}(\sigma^2+\mu^2) - \frac{n-1}{n}\mu^2\Big) = \sigma^2.\end{aligned}$$

Anmerkung: Alternativ könnte man auch mit der Kovarianz und der initialen Darstellung von S_n^2 arbeiten.
Es ist wegen $\mathbb{E}[X_i - \overline{X_n}] = \mu - \mu = 0$ und $\mathrm{Kov}(X_i, \overline{X_n}) = \frac{1}{n}\sum_{j=1}^n \mathrm{Kov}(X_i, X_j) = \frac{\sigma^2}{n}$

$$\begin{aligned}\mathbb{E}[S_n^2] &= \frac{1}{n-1}\sum_{i=1}^n \mathbb{E}[(X_i - \overline{X_n})^2] = \frac{1}{n-1}\sum_{i=1}^n \mathrm{Var}(X_i - \overline{X_n}) \\ &= \frac{1}{n-1}\sum_{i=1}^n \Big[\mathrm{Var}(X_i) - 2\mathrm{Kov}(X_i, \overline{X_n}) + \mathrm{Var}(\overline{X_n})\Big] \\ &= \frac{1}{n-1}\sum_{i=1}^n \Big[\sigma^2 - \frac{\sigma^2}{n}\Big] = \sigma^2.\end{aligned}$$

(ii) Es gilt $\mathbb{P}(\mathbb{1}_{\{X_i \le t\}} = 1) = \mathbb{P}(X_i \le t) = F(t)$. Daher ist $\mathbb{1}_{\{X_i \le t\}} \sim \mathcal{B}(1, F(t))$. Da die X_i für $i = 1, ..., n$ unabhängig sind, sind auch $\mathbb{1}_{\{X_i \le t\}}$ für $i = 1, ..., n$ unabhängig. Es folgt nach der Vorlesung, dass

$$n \cdot \hat{F}_n(t) = \sum_{i=1}^n \mathbb{1}_{\{X_i \le t\}} \sim \mathcal{B}(n, F(t)).$$

Aus den bekannten Formeln für Erwartungswert und Varianz für eine Binomialverteilung erhalten wir $\mathbb{E}[n \cdot \hat{F}_n(t)] = n \cdot F(t)$ und $\mathrm{Var}(n \cdot \hat{F}_n(t)) = n \cdot F(t) \cdot (1 - F(t))$, womit

$$\mathbb{E}[\hat{F}_n(t)] = F(t), \quad \mathrm{Var}(\hat{F}_n(t)) = \frac{F(t) \cdot (1 - F(t))}{n}$$

folgt.

□