Data Science 2 (Wintersemester 23/24)
Prof. Dr. Mark Trede
Fabian Apostel B. Sc.
Daniel Stroth B. Sc.

Lösung Hausaufgabe 3

Aufgabe 1

Die folgende Aufgabe beschreibt ein Ein-Faktor-Modell für den Aktienmarkt:
Ein Portfolio bestehe aus $K \in \mathbb{N}$ Wertpapieren. Die Rendite R_i des Wertpapiers i für das kommende Jahr ist eine Zufallsvariable, $i = 1, \ldots, K$. Betrachten Sie das folgende einfache multivariate Renditemodell. Die Rendite R_i setzt sich additiv aus einer Marktkomponente Y (oft auch Faktor oder Marktfaktor genannt) und einer individuellen Komponente X_i zusammen, also

$$R_i = Y + X_i$$

mit $Y \sim N\left(\mu_Y, \sigma_Y^2\right)$ und $X_i \sim N\left(0, \sigma_X^2\right)$ für $i = 1, \ldots, K$. Die individuellen Komponenten $X_1, \ldots, X_K$ sind paarweise unabhängig und auch unabhängig von Y, sie haben alle die gleiche Varianz σ_X^2.

a) Berechnen Sie die Varianz von R_i (für ein beliebiges i).
b) Berechnen Sie Kovarianz von R_i und R_j für $i \neq j$.
c) Berechnen Sie den Korrelationskoeffizienten von R_i und R_j für $i \neq j$.

Lösung zu Aufgabe 1

a)

$$\begin{aligned}
Var(R_i) &= Var(Y + X_i) \\
&= Var(Y) + Var(X_i) + 2 \cdot Cov(Y, X_i) \\
&= \sigma_Y^2 + \sigma_X^2 + 2 \cdot 0 \\
&= \sigma_Y^2 + \sigma_X^2
\end{aligned}$$

b) Sei $i \neq j$:

$$\begin{aligned}
Cov(R_i, R_j) &= E(R_i \cdot R_j) - E(R_i)E(R_j) \\
&= E((Y + X_i) \cdot (Y + X_j)) - E(Y + X_i) \cdot E(Y + X_j) \\
&= E(Y^2 + Y \cdot X_i + Y \cdot X_j + X_i \cdot X_j) + (E(Y) + \underbrace{E(X_i)}_{=0}) \cdot (E(Y) + \underbrace{E(X_j)}_{=0}) \\
&= E(Y^2) + E(Y) \cdot \underbrace{E(X_i)}_{=0} + E(Y) \cdot \underbrace{E(X_j)}_{=0} + \underbrace{E(X_i) \cdot E(X_j)}_{=0} - E(Y)^2 \\
&= Var(Y) + E(Y)^2 - E(Y)^2 \\
&= \sigma_Y^2
\end{aligned}$$

c) Sei $i \neq j$

$$\begin{aligned}
Cor(R_i, R_j) &= \frac{Cov(R_i, R_j)}{\sqrt{Var(R_i)} \cdot \sqrt{Var(R_j)}} \\
&= \frac{\sigma_Y^2}{\sqrt{\sigma_Y^2 + \sigma_X^2} \cdot \sqrt{\sigma_Y^2 + \sigma_X^2}} \\
&= \frac{\sigma_Y^2}{\sigma_Y^2 + \sigma_X^2}
\end{aligned}$$

Aufgabe 2

Seien X_1 und X_2 zwei unabhängige diskrete Zufallsvariable mit Verteilung

$$P(X_i = k) = (1-p)^k \cdot p$$

für $k = 0, 1, 2, \ldots$ und $0 \leq p \leq 1$. Sei $Z = \max(X_1, X_2)$.

a) Zeigen Sie $P(X_2 \leq k) = 1 - (1-p)^k$ für beliebiges $k \in \mathbb{N}_0$. *Hinweis: Geometrische Reihe.*

b) Bestimmen Sie die gemeinsame Verteilung von Z und X_1.

c) Bestimmen Sie die Verteilung von Z.
Bemerkung: Die Verteilung der Zufallsvariablen X_1 und X_2 ist die sogenannte geometrische Verteilung. Man schreibt kurz: $X_1, X_2 \sim Geo(p)$.

Lösung zu Aufgabe 2

a)

$$
\begin{aligned}
P(X_2 \leq k) &= \sum_{i=0}^{k} P(X_2 = i) \\
&= \sum_{i=0}^{k} p \cdot (1-p)^i \\
&= p \cdot \sum_{i=0}^{k} (1-p)^i \\
&= p \cdot \frac{1-(1-p)^{k+1}}{1-(1-p)} \\
&= \cancel{p} \cdot \frac{1-(1-p)^{k+1}}{\cancel{p}} \\
&= 1-(1-p)^{k+1}
\end{aligned}
$$

b) Für $i, k \in \mathbb{N}_0$:

$$P(Z = i, X_1 = k) = P(\max(X_1, X_2) = i, X_1 = k)$$

Sei nun $i > k$:

$$
\begin{aligned}
P(\max(X_1, X_2) = i, X_1 = k) &= P(X_2 = i, X_1 = k) \\
&= P(X_2 = i)P(X_1 = k) \\
&= p \cdot (1-p)^i \cdot p \cdot (1-p)^k \\
&= p^2 \cdot (1-p)^{i+k}
\end{aligned}
$$

Für $i < k$ gilt $P(\max(X_1, X_2) = i, X_1 = k) = 0$. Und für $i = k$ gilt:

$$
\begin{aligned}
P(\max(X_1, X_2) = i, X_1 = k) &= P(X_2 \leq k, X_1 = k) \\
&= P(X_2 \leq k) \cdot P(X_1 = k) \\
&= (1-(1-p)^{k+1})p(1-p)^k
\end{aligned}
$$

c)

$$
\begin{aligned}
P(Z=k) &= \sum_{i=0}^{k} P(Z=k, X_1=i) \\
&= P(Z=k, X_1=k) + \sum_{i=0}^{k-1} P(Z=k, X_1=i) \\
&= (1-(1-p)^{k+1})p(1-p)^k + \sum_{i=0}^{k-1} p^2 \cdot (1-p)^{k+i} \\
&= (1-(1-p)^{k+1})p(1-p)^k + p^2 \cdot (1-p)^k \cdot \sum_{i=0}^{k-1} (1-p)^i \\
&= (1-(1-p)^{k+1})p(1-p)^k + p^{\cancel{2}} \cdot (1-p)^k \cdot \frac{1-(1-p)^k}{\cancel{p}} \\
&= p(1-p)^k \cdot \big((1-(1-p)^{k+1}) + (1-(1-p)^k)\big) \\
&= p(1-p)^k \cdot \big(2-(1-p)^{k+1}-(1-p)^k\big) \\
&= p(1-p)^k \cdot \big(2-(1-p)^k \cdot (1-p-1)\big) \\
&= p(1-p)^k \cdot \big(2+(1-p)^k \cdot p\big)
\end{aligned}
$$

Aufgabe 3

Seien X und Y gemeinsam stetig verteilte Zufallsvariablen mit der gemeinsamen Verteilungsfunktion

$$F_{X,Y}(x,y) = \left(x^{-2}+y^{-2}-1\right)^{-1/2}$$

für $0 < x, y \leq 1$. Für Werte von x und y außerhalb dieses Bereichs geht man wie folgt vor: Falls $x > 1$ setzt man $x = 1$ und analog für y. Für $x, y \leq 0$ ist $F_{XY}(x,y) = 0$.

a) Bestimmen Sie die Randverteilungsfunktion von X.

b) Sind X und Y stochastisch unabhängig?

c) Bestimmen Sie die Randdichten von X und Y.

d) Bestimmen Sie die gemeinsame Dichtefunktion von X und Y.

e) Bestimmen Sie die bedingte Dichtefunktion von X unter der Bedingung $Y = y$.

Lösung zu Aufgabe 3

a) Setzt $y = 1$ dann gilt $F_X(x) = F_{X,Y}(x,1) = (x^{-2})^{-\frac{1}{2}} = x$ für $0 < x \leq 1$ und $F_X(x) = 0$ für $x \leq 0$ und $F_X(x) = 1$ für $x \geq 1$.

b) Aufgrund der Symmetrie gilt $F_Y(y) = y$ für $0 < y \leq 1$ und $F_Y(y) = 0$ für $y \leq 0$ und $F_Y(y) = 1$ für $y \geq 1$. Wären X und Y unabhängig so würde gelten $F_X(x) \cdot F_Y(y) = F_{XY}(x,y)$. Es gilt jedoch, dass $x \cdot y \neq F_{XY}(x,y) = (x^{-2}+y^{-2}-1)^{-1/2}$ für $0 < x, y \leq 1$. Somit sind diese Zufallsvariablen nicht stochastisch unabhängig.

c) Es gilt $f_X(x) = \frac{\mathrm{d}}{\mathrm{d}\,x} F_X(x)$. Also ist $f_X(x) = 1 = f_Y(y)$ für $0 < x, y \leq 1$ und $f_X(x) = f_Y(y) = 0$ für alle anderen x, y.

d) Es gilt $\frac{\partial^2}{\partial x \partial y} F_{X,Y}(x,y) = f_{X,Y}(x,y)$.
Seien $x, y \in (0,1]$. Erst nach x ableiten:

$$\begin{aligned}
\frac{\partial}{\partial x} F_{X,Y}(x,y) &= \frac{\mathrm{d}}{\mathrm{d}\,x}\left(x^{-2} + y^{-2} - 1\right) \cdot \frac{-1}{2 \cdot (x^{-2} + y^{-2} - 1)^{\frac{3}{2}}} \\
&= -2x^{-3} \cdot \frac{-1}{2 \cdot (x^{-2} + y^{-2} - 1)^{\frac{3}{2}}} \\
&= \frac{1}{(x^{-2} + y^{-2} - 1)^{\frac{3}{2}}\, x^3}
\end{aligned}$$

Ableitung nach y mit einem ähnlichen Argument gibt:

$$\frac{\partial^2}{\partial x \partial y} F_{X,Y}(x,y) = \frac{3}{x^3 \cdot (y^{-2} + x^{-2} - 1)^{\frac{5}{2}}\, y^3}$$

Für $x, y \notin (0,1]$ haben wir, dass $f_{XY}(x,y) = 0$.

e) Weil für die Randdichten für $0 < x, y \leq 1$ gilt, dass $f_X(x) = f_Y(y) = 1$, gilt für die bedingte Dichtefunktion $f_{X|Y=y}(x) = f_{X,Y}(x,y)$.

Aufgabe 4

Zeigen Sie, dass die folgenden Gleichungen gelten (die erste Gleichung ist natürlich eine Definitionsgleichung):

$$\begin{aligned}
Cov(X,Y) &= E\left[(X - E(X))\,(Y - E(Y))\right] \\
&= E(XY) - E(X)E(Y) \\
&= E\left[(X - E(X))\,(Y - a)\right] \\
&= E\left[(X - a)\,(Y - E(Y))\right] \\
&= E\left[(X - E(X))\,Y\right] \\
&= E\left[X\,(Y - E(Y))\right]
\end{aligned}$$

für $a \in \mathbb{R}$.

Lösung zu Aufgabe 4

$$\begin{aligned}
Cov(X,Y) &= E\left[(X - E(X))\,(Y - E(Y))\right] \\
&= E\left[X \cdot Y - X \cdot E(Y) - E(X) \cdot Y + E(X) \cdot E(Y)\right] \\
&= E(X \cdot Y) - E(X \cdot E(Y)) - E(E(X) \cdot Y) + E(E(X) \cdot E(Y)) \\
&= E(X \cdot Y) - E(X) \cdot E(Y) - E(X) \cdot E(Y) + E(X) \cdot E(Y) \\
&= E(XY) - E(X)E(Y)
\end{aligned}$$

Es gilt nun:

$$
\begin{aligned}
E\left[(X-E(X))\,(Y-a)\right] &= E\left[X\cdot Y - E(X)\cdot Y - X\cdot a + E(X)\cdot a\right]\\
&= E(XY) - E(E(X)\cdot Y) - E(X\cdot a) + E(E(X)\cdot a)\\
&= E(XY) - E(X)E(Y) - E(X)\cdot a + E(X)\cdot a\\
&= E(XY) - E(X)E(Y)
\end{aligned}
$$

Analog für $E\left[(X-a)\,(Y-E(Y))\right]$. Es gilt nun

$$
\begin{aligned}
E\left[(X-E(X))\cdot Y\right] &= E\left[X\cdot Y - E(X)\cdot Y\right]\\
&= E(X\cdot Y) - E(E(X)\cdot Y)\\
&= E(XY) - E(X)E(Y)
\end{aligned}
$$

Analog für $E\left[X\,(Y-E(Y))\right]$.

Aufgabe 5

Seien $X_1,\dots,X_n$ positive, unabhängige und identisch verteilte Zufallsvariablen. Zeigen Sie, dass $E(\frac{X_1}{X_1+\dots+X_n}) = \frac{1}{n}$ ist.

Lösung zu Aufgabe 5

$$
\begin{aligned}
1 &= \frac{X_1+\dots+X_n}{X_1+\dots X_n}\\
&= E\left[\frac{X_1+\dots+X_n}{X_1+\dots X_n}\right]\\
&= E\left[\frac{X_1}{X_1+\dots+X_n}\right]+\dots+E\left[\frac{X_n}{X_1+\dots+X_n}\right]\\
&\overset{\text{Identisch verteilt}}{=} E\left[\frac{X_1}{X_1+\dots+X_n}\right]+\dots+E\left[\frac{X_1}{X_1+\dots+X_n}\right]\\
&= n\cdot E\left[\frac{X_1}{X_1+\dots+X_n}\right]
\end{aligned}
$$

Das ist äquivalent zu der zu zeigenden Aussage.