Data Science 2

Prof. Dr. Mark Trede

Institut für Ökonometrie und Wirtschaftsstatistik

September 2023

Kovarianz

Die Kovarianz zwischen X und Y ist definiert als

$$Cov(X,Y) = E[(X - E(X))(Y - E(Y))]$$

Messung des Zusammenhangs zwischen X und Y

Notation häufig auch σ_{XY}

Kovarianz

Definition

Nützliche Formel $Cov(X,Y) = E(XY) - E(X)E(Y)$

Kovarianz

Definition

Nützliche Formel $Cov(X,Y) = E(XY) - E(X)E(Y)$

$$
\begin{aligned}
&Cov(X,Y) \\
&= E((X - E(X))(Y - E(Y))) \\
&= E(XY - XE(Y) - E(X)Y + E(X)E(Y)) \\
&= E(XY) - E(XE(Y)) - E(E(X)Y) + E(E(X)E(Y)) \\
&= E(XY) - E(X)E(Y) - E(X)E(Y) + E(X)E(Y) \\
&= E(XY) - E(X)E(Y)
\end{aligned}
$$

Kovarianz

Definition

- Gemeinsam diskret verteilte Zufallsvariable

$$Cov(X,Y) = \sum_{j=1}^{J}\sum_{k=1}^{K}(x_j - E(X))(y_k - E(Y))p_{jk}$$

- Gemeinsam stetig verteilte Zufallsvariable

$$Cov(X,Y) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}(x - E(X))(y - E(Y))f(x,y)\,dx\,dy$$

Eigenschaften der Kovarianz:

- Symmetrie: $Cov(X, Y) = Cov(Y, X)$
- Lineare Transformation:

$$Cov(aX + b, cY + d) = a \cdot c \cdot Cov(X, Y)$$

- Wenn X und Y unabhängig sind, gilt $Cov(X, Y) = 0$
- $Cov(X, X) = Var(X)$
- Die Kovarianz ist schlecht zu interpretieren

Erwartungswert und Varianz einer Summe:

- Seien X und Y zwei Zufallsvariablen
- $X + Y$ ist ebenfalls eine Zufallsvariable
- Erwartungswert der Summe

$$E(X + Y) = E(X) + E(Y)$$

Kovarianz

Summe zweier Zufallsvariablen

- Varianz der Summe

$$Var(X+Y) = Var(X) + Var(Y) + 2Cov(X,Y)$$

- Bei Unkorreliertheit (insb. Unabhängigkeit)

$$Var(X+Y) = Var(X) + Var(Y)$$

Kovarianz

Summe zweier Zufallsvariablen

Herleitung $Var(X + Y) = ...$

Herleitung $Var(X+Y) = \ldots$

$$\begin{aligned}
\ldots &= E([(X+Y) - E(X+Y)]^2) \\
&= E([X+Y - E(X) - E(Y))]^2) \\
&= E([(X - E(X)) + (Y - E(Y))]^2) \\
&= E((X - E(X))^2 + (Y - E(Y))^2 \\
&\quad + 2(X - E(X))(Y - E(Y))) \\
&= E((X - E(X))^2) + E((Y - E(Y))^2) \\
&\quad + 2E((X - E(X))(Y - E(Y))) \\
&= Var(X) + Var(Y) + 2Cov(X,Y)
\end{aligned}$$

Beispiel: Varianz einer Summe

Eine Person wird zufällig aus der Bevölkerung ausgewählt, sei X das monatliche Arbeitseinkommen und Y das monatliche Vermögenseinkommen; gegeben sei

$$\begin{aligned} E(X) &= 2500 \qquad & Var(X) &= 4000000 \\ E(Y) &= 300 \qquad & Var(Y) &= 160000 \\ Cov(X,Y) &= 240000 \end{aligned}$$

Kovarianz

Summe zweier Zufallsvariablen

Beispiel: Varianz einer Summe

$E(X) = 2500, Var(X) = 4000000, E(Y) = 300,$
$Var(Y) = 160000, Cov(X, Y) = 240000$

Erwartungswert und Varianz des Gesamteinkommens

Beispiel: Varianz einer Summe

$E(X) = 2500, Var(X) = 4000000, E(Y) = 300,$
$Var(Y) = 160000, Cov(X, Y) = 240000$

Erwartungswert und Varianz des Gesamteinkommens

$$\begin{aligned} E(X+Y) &= E(X) + E(Y) \\ &= 2\,500 + 300 \\ &= 2\,800 \\ Var(X+Y) &= Var(X) + Var(Y) + 2Cov(X, Y) \\ &= 4\,000\,000 + 160\,000 + 2 \cdot 240\,000 \\ &= 4\,640\,000 \end{aligned}$$

Erwartungswert und Varianz einer gewichteten Summe

- Seien X und Y zwei Zufallsvariablen, dann ist $aX + bY$ ebenfalls eine Zufallsvariable
- Erwartungswert und Varianz

$$\begin{aligned} E(aX + bY) &= aE(X) + bE(Y) \\ Var(aX + bY) &= a^2 Var(X) + b^2 Var(Y) \\ &\quad + 2ab \cdot Cov(X, Y) \end{aligned}$$

Beispiel: Risikominimierung

X und Y seien Jahresrenditen von zwei Wertpapieren mit

$$\begin{aligned} E(X) &= 0.02 & E(Y) &= 0.06 \\ Var(X) &= 0.01 & Var(Y) &= 0.04 \\ Cov(X,Y) &= -0.002 \end{aligned}$$

Wie hoch soll der Anteil a des Vermögens sein, der in Papier A investiert wird, wenn das Gesamtrisiko minimiert werden soll?

Forts. Beispiel: Risikominimierung

Portfolio-Rendite

$$R = aX + (1-a)Y$$

mit Erwartungswert und Varianz

$$\begin{aligned}
E(R) &= aE(X) + (1-a)E(Y) \\
&= a \cdot 0.02 + (1-a) \cdot 0.06 \\
Var(R) &= a^2 Var(X) + (1-a)^2 Var(Y) \\
&\quad + 2a(1-a)Cov(X,Y) \\
&= a^2 \cdot 0.01 + (1-a)^2 \cdot 0.04 + 2a(1-a) \cdot (-0.002)
\end{aligned}$$

Forts. Beispiel: Risikominimierung

Minimiere

$$Var(R) = 0.054 \cdot a^2 - 0.084 \cdot a + 0.04$$

bzgl. a. Ableitung:

$$\frac{\partial Var(R)}{\partial a} = 0.108 \cdot a - 0.084$$

Nullsetzen und Auflösen nach a:

$$a = \frac{0.084}{0.108} \approx 0.778.$$

Forts. Beispiel: Risikominimierung

Das Risiko des Gesamtportfolios wird minimal, wenn 77.8 Prozent des Vermögens in Papier A und 22.2 Prozent in Papier B investiert werden. Dann gilt

$$E(R) = 0.02888$$
$$Var(R) = 0.00733.$$

Es wäre also unklug, sein ganzes Vermögen vollständig in dem sichereren Wertpapier A anzulegen!

Verallgemeinerung auf n Zufallsvariablen

- Seien $X_1, \dots, X_n$ Zufallsvariablen
- Erwartungswert der Summe

$$E\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} E(X_i)$$

Kovarianz

Summe vieler Zufallsvariablen

- Varianz der Summe

$$
\begin{aligned}
&Var\left(\sum_{i=1}^{n} X_i\right) \\
&= \sum_{i=1}^{n}\sum_{j=1}^{n} Cov(X_i, X_j) \\
&= Cov(X_1, X_1) + Cov(X_1, X_2) + ... + Cov(X_1, X_n) \\
&+ Cov(X_2, X_1) + Cov(X_2, X_2) + ... + Cov(X_2, X_n) \\
&\vdots \\
&+ Cov(X_n, X_1) + Cov(X_n, X_2) + ... + Cov(X_n, X_n)
\end{aligned}
$$

- Varianz der Summe

$$Var\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} Var(X_i) + 2\sum_{i=1}^{n-1}\sum_{j=i+1}^{n} Cov(X_i, X_j)$$

- Bei Unabhängigkeit

$$Var\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} Var(X_i)$$

Erwartungswert und Varianz einer gewichteten Summe

- Seien $X_1, \dots, X_n$ Zufallsvariablen und $a_1, \dots, a_n$ reelle Zahlen
- Erwartungswert und Varianz der gewichteten Summe

$$
\begin{aligned}
E\left(\sum_{i=1}^{n} a_i X_i\right) &= \sum_{i=1}^{n} a_i E(X_i) \\
Var\left(\sum_{i=1}^{n} a_i X_i\right) &= \sum_{i=1}^{n} a_i^2 Var(X_i) \\
&\quad + \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} a_i a_j Cov(X_i, X_j)
\end{aligned}
$$

Korrelationskoeffizient

Der Korrelationskoeffizient ist definiert als

$$\rho_{XY} = \frac{Cov(X,Y)}{\sqrt{Var(X)}\sqrt{Var(Y)}} = \frac{\sigma_{XY}}{\sigma_X \cdot \sigma_Y}$$

Der Korrelationskoeffizient ist die normierte Kovarianz

Eigenschaften des Korrelationskoeffizienten

- Der Korrelationskoeffizient ist dimensionslos
- Symmetrie: Es gilt $\rho_{XY} = \rho_{YX}$
- Normierung: Es gilt $-1 \leq \rho_{XY} \leq 1$
- Invarianz bzgl. linearer Transformationen
- Falls X und Y unabhängig sind, dann ist $\rho_{XY} = 0$
- $Cov(X, Y) = \rho_{XY} \cdot \sigma_X \cdot \sigma_Y$