11.4 Korrelation

Def. 44 *Es seien X_1 und X_2 zwei zufällige Variablen, für die gilt:* $0 < \sigma_{X_1}, \sigma_{X_2} < +\infty$. *Dann heißt der Quotient*

$$\varrho(X_1, X_2) = \frac{\operatorname{cov}(X_1, X_2)}{\sigma_{X_1} \cdot \sigma_{X_2}}$$

Korrelationskoeffizient der Zufallsgrößen X_1 und X_2.

Ist $\operatorname{cov}(X_1, X_2) = 0$ *dann heißen die beiden Zufallsgrößen unkorreliert.*

Bem.: X_1 und X_2 unabhängig $\Rightarrow \operatorname{cov}(X_1, X_2) = 0$.

Die Umkehrung der Aussage gilt im allgemeinen nicht.

Bsp. 75 (2x2 Tafel)

Y *X*	*0(Sportler)*	*1(Nichtsportler)*	*Summe*
0(w)	p_{11}	p_{12}	$p_{1.}$
1(m)	p_{21}	p_{22}	$p_{2.}$
Summe	$p_{.1}$	$p_{.2}$	*1*

$$X \sim Bi(1, p_{.2}) \qquad Y \sim Bi(1, p_{2.})$$

$$\begin{aligned} \mathbf{E}(X) &= p_{.2} \quad var(X) = p_{.2}(1 - p_{.2}) = p_{.2}p_{.1} \\ \mathbf{E}(Y) &= p_{2.} \quad var(Y) = p_{2.}(1 - p_{2.}) = p_{2.}p_{1.} \end{aligned}$$

$$cov(X, Y) = \mathbf{E}(X \cdot Y) - \mathbf{E}(X)\mathbf{E}(Y) = p_{22} - p_{.2}p_{2.}$$

Korrelationskoeffizient:

$$\rho = \frac{p_{22} - p_{.2}p_{2.}}{\sqrt{p_{.2}p_{1.}p_{2.}p_{.1}}} = \frac{p_{11}p_{22} - p_{12}p_{21}}{\sqrt{p_{.2}p_{2.}p_{1.}p_{.1}}}$$

$$\begin{aligned}
p_{22} - p_{2.}p_{.2} &= p_{22} - (p_{21} + p_{22})(p_{12} + p_{22}) \\
&= p_{22} - (p_{21}p_{12} + p_{22}p_{12} + p_{21}p_{22} + p_{22}^2) \\
&= p_{22}(1 - p_{12} - p_{21} - p_{22}) - p_{21}p_{12} \\
&= p_{22}p_{11} - p_{21}p_{12}
\end{aligned}$$

Satz 33 *Es seien* X_1 *und* X_2 *zwei Zufallsgrößen, für die* $\sigma_{X_1}, \sigma_{X_2} > 0$ *ist. Dann gilt für den Korrelationskoeffizienten dieser beiden zufälligen Variablen:*

$$-1 \leq \varrho(X_1, X_2) \leq 1.$$

Beweis: Wir definieren eine Funktion A wie folgt:

$$A(t, u) := \mathbf{E}[t \cdot (X_1 - \mathbf{E}X_1) + u \cdot (X_2 - \mathbf{E}X_2)]^2 \geq 0.$$

Nun gilt für alle $t, u \in \mathbb{R}$:

$$
\begin{aligned}
A(t,u) &= \mathbf{E}(t^2(X_1 - \mathbf{E}X_1)^2 + u^2(X_2 - \mathbf{E}X_2)^2) \\
&\quad +2tu\mathbf{E}(X_1 - \mathbf{E}X_1)(X_2 - \mathbf{E}X_2) \\
&= t^2\mathbf{E}(X_1 - \mathbf{E}X_1)^2 + u^2\mathbf{E}(X_2 - \mathbf{E}X_2)^2 \\
&\quad +2tu\mathbf{E}((X_1 - \mathbf{E}X_1)(X_2 - \mathbf{E}X_2)) \\
&= t^2 \cdot \mathrm{Var}\, X_1 + 2 \cdot t \cdot u \cdot \mathrm{cov}\,(X_1, X_2) + u^2 \cdot \mathrm{Var}\, X_2 \\
&\geq 0
\end{aligned}
$$

Wir setzen $t := \sigma_{X_2}$, $u := \sigma_{X_1}$ und dividieren durch

$\sigma_{X_1} \cdot \sigma_{X_2}$:

$$\frac{A(\sigma_{X_2}, \sigma_{X_1})}{\sigma_{X_1} \cdot \sigma_{X_2}} = \frac{\sigma_{X_2}^2 \cdot \sigma_{X_1}^2 + 2 \cdot \sigma_{X_1} \cdot \sigma_{X_2} \cdot \operatorname{cov}(X_1, X_2) + \sigma_{X_1}^2 \cdot \sigma_{X_2}^2}{\sigma_{X_1} \cdot \sigma_{X_2}}$$

$$= \sigma_{X_1} \cdot \sigma_{X_2} + 2 \cdot \operatorname{cov}(X_1, X_2) + \sigma_{X_1} \cdot \sigma_{X_2}$$

$$= 2 \cdot \sigma_{X_1} \cdot \sigma_{X_2} + 2 \cdot \operatorname{cov}(X_1, X_2) \geq 0$$

Also:

$$\underline{\sigma_{X_1} \cdot \sigma_{X_2} + \operatorname{cov}(X_1, X_2) \geq 0}.$$

Andererseits gilt aber auch mit $t := -\sigma_{X_2}$ und $u := \sigma_{X_1}$ sowie derselben Herleitung wie oben:

$$\frac{A(-\sigma_{X_2}, \sigma_{X_1})}{\sigma_{X_1} \cdot \sigma_{X_2}} = \frac{\sigma_{X_2}^2 \cdot \sigma_{X_1}^2 - 2 \cdot \sigma_{X_1} \cdot \sigma_{X_2} \cdot \operatorname{cov}(X_1, X_2) + \sigma_{X_1}^2 \cdot \sigma_{X_2}^2}{\sigma_{X_1} \cdot \sigma_{X_2}}$$

$$= 2 \cdot \sigma_{X_1} \cdot \sigma_{X_2} - 2 \cdot \operatorname{cov}(X_1, X_2) \geq 0$$

Also:

$$\underline{\sigma_{X_1} \cdot \sigma_{X_2} - \operatorname{cov}(X_1, X_2) \geq 0}.$$

Beides zusammen ergibt

$$-\sigma_{X_1} \cdot \sigma_{X_2} \leq \operatorname{cov}(X_1, X_2) \leq \sigma_{X_1} \cdot \sigma_{X_2}.$$

Wir stellen etwas um und erhalten:

$$-1 \leq \frac{\operatorname{cov}(X_1, X_2)}{\sigma_{X_1} \cdot \sigma_{X_2}} = \varrho(X_1, X_2) \leq 1.$$

$\square$

Bem. 16 *Die Ungleichung kann auch direkt aus der Cauchy-Schwarz'schen Ungleichung hergeleitet werden.*

Satz 34 *Es seien X_1 und X_2 zwei Zufallsgrößen, für die $\sigma_{X_1}, \sigma_{X_2} > 0$ ist. Dann gilt $|\varrho(X_1, X_2)| = 1$ genau dann, wenn es Zahlen $a, b \in \mathbb{R}$ ($a \neq 0$) gibt, so daß gilt:*
$P(X_1 = a \cdot X_2 + b) = 1.$

Beweis:

($\Longleftarrow$) Es seien die Zahlen $a, b \in \mathbb{R}$ so gewählt, daß gilt

$$P(X_1 = a \cdot X_2 + b) = 1.$$

Für Erwartungswert und Varianz der Zufallsgröße X_1

gilt dann:

$$\mathbf{E}X_1 = \mathbf{E}(a \cdot X_2 + b) = a \cdot \mathbf{E}X_2 + b,$$

$$\sigma^2_{X_1} = \sigma^2_{a \cdot X_2 + b} = \sigma^2_{a \cdot X_2} = a^2 \cdot \sigma^2_{X_2}.$$

Damit gilt für den Korrelationskoeffizienten der

Zufallsgrößen X_1 und X_2:

$$\begin{aligned}
\varrho(X_1, X_2) &= \frac{\operatorname{cov}(X_1,X_2)}{\sigma_{X_1} \cdot \sigma_{X_2}} = \frac{\mathbf{E}((X_1 - \mathbf{E}X_1) \cdot (X_2 - \mathbf{E}X_2))}{|a| \cdot \sigma_{X_2} \cdot \sigma_{X_2}} \\
&= \frac{\mathbf{E}([(a \cdot X_2 + b) - (a \cdot \mathbf{E}X_2 + b)] \cdot (X_2 - \mathbf{E}X_2))}{|a| \cdot \sigma^2_{X_2}} \\
&= \frac{a \cdot \mathbf{E}(X_2 - \mathbf{E}X_2)^2}{|a| \cdot \sigma^2_{X_2}} = \frac{a \cdot \sigma^2_{X_2}}{|a| \cdot \sigma^2_{X_2}} \\
&= \begin{cases} 1 & \text{, falls } a > 0 \\ -1 & \text{, falls } a < 0 \end{cases}
\end{aligned}$$

Das bedeutet: $|\varrho(X_1, X_2)| = 1$.

($\Longrightarrow$) Es gelte $|\varrho(X_1, X_2)| = 1$. Nun gilt:

$$\begin{aligned}
\varrho(X_1, X_2) &= \frac{\text{cov}\,(X_1,X_2)}{\sigma_{X_1}\cdot\sigma_{X_2}} \\
&= \frac{\mathbf{E}((X_1-\mathbf{E}X_1)\cdot(X_2-\mathbf{E}X_2))}{\sigma_{X_1}\cdot\sigma_{X_2}} \\
&= \mathbf{E}\left(\frac{X_1-\mathbf{E}X_1}{\sigma_{X_1}}\cdot\frac{X_2-\mathbf{E}X_2}{\sigma_{X_2}}\right)
\end{aligned}$$

Wir definieren zwei Zufallsgrößen:

$$X_1^* := \frac{X_1-\mathbf{E}X_1}{\sigma_{X_1}}, \qquad X_2^* := \frac{X_2-\mathbf{E}X_2}{\sigma_{X_2}}.$$

Für die Varianz dieser Zufallsgrößen X_i^* ($i = 1, 2$) gilt dann:

$$
\begin{aligned}
\sigma^2_{X_i^*} &= \mathbf{E}\left(X_i^* - \mathbf{E}X_i^*\right)^2 \\
&= \mathbf{E}\left(X_i^*\right)^2 - \left(\mathbf{E}X_i^*\right)^2 \\
&= \mathbf{E}\left(\frac{X_i - \mathbf{E}X_i}{\sigma_{X_i}}\right)^2 - \left(E\left(\frac{X_i - \mathbf{E}X_i}{\sigma_{X_i}}\right)\right)^2 \\
&= \frac{1}{\sigma^2_{X_i}} \cdot \Big(\mathbf{E}(X_i - \mathbf{E}X_i)^2 - (\mathbf{E}(X_i - \mathbf{E}X_i))^2\Big) \\
&= \frac{1}{\sigma^2_{X_i}} \cdot \sigma^2_{X_i - \mathbf{E}X_i} \\
&= \frac{1}{\sigma^2_{X_i}} \cdot \sigma^2_{X_i} \\
&= 1
\end{aligned}
$$

Wir ermitteln jetzt die Erwartungswerte ($i = 1, 2$):

$$\begin{aligned}
\mathbf{E}X_i^* &= \mathbf{E}\left(\frac{X_i - \mathbf{E}X_i}{\sigma_{X_i}}\right) \\
&= \frac{1}{\sigma_{X_i}} \cdot (\mathbf{E}X_i - \mathbf{E}(\mathbf{E}X_i)) \\
&= \frac{1}{\sigma_{X_i}} \cdot (\mathbf{E}X_i - \mathbf{E}X_i) \\
&= 0
\end{aligned}$$

Daraus folgt: $\varrho(X_1, X_2) = \mathbf{E}\,(X_1^* \cdot X_2^*)$. Wir unterscheiden zwei Fälle:

$\underline{\varrho(X_1, X_2) = 1}$**:** Wir untersuchen die Varianz der

Zufallsgröße $X_1^* - X_2^*$:

$$\begin{aligned} \sigma^2_{X_1^*-X_2^*} &= \mathbf{E}\left((X_1^* - X_2^*) - \mathbf{E}\left(X_1^* - X_2^*\right)\right)^2 \\ &= \mathbf{E}\left((X_1^* - X_2^*) - \mathbf{E}X_1^* + \mathbf{E}X_2^*\right)^2 \\ &= \mathbf{E}\left(X_1^* - X_2^*\right)^2 \\ &= \mathbf{E}\left(X_1^*\right)^2 - 2 \cdot \mathbf{E}\left(X_1^* \cdot X_2^*\right) + \mathbf{E}\left(X_2^*\right)^2 \end{aligned}$$

Für $i = 1, 2$ gilt nun:

$$\mathbf{E}\left(X_i^*\right)^2 = \mathbf{E}\left(X_i^* - \mathbf{E}X_i^*\right)^2 = \sigma^2_{X_i^*} = 1.$$

Damit erhalten wir dann:

$$\begin{aligned}\sigma^2_{X_1^*-X_2^*} &= \mathbf{E}\left(X_1^*\right)^2 - 2\cdot\mathbf{E}\left(X_1^*\cdot X_2^*\right) + \mathbf{E}\left(X_2^*\right)^2 \\ &= 2 - 2\cdot\varrho(X_1, X_2) \\ &= 0\end{aligned}$$

Nun gilt aber $\sigma^2_{X_1^*-X_2^*} = 0$ genau dann, wenn es ein $c \in \mathbb{R}$ gibt, so daß $P\left(X_1^* - X_2^* = c\right) = 1$ ist. Das bedeutet aber, daß gilt: $\mathbf{E}\left(X_1^* - X_2^*\right) = c$. Wegen $EX_1^* = EX_2^* = 0$ ist $c = 0$, woraus folgt

$$P\left(X_1^* = X_2^*\right) = 1.$$

Dann gilt:

$$\begin{aligned}
1 &= P\left(X_1^* = X_2^*\right) \\
&= P\left(\frac{X_1 - \mathbf{E}X_1}{\sigma_{X_1}} = \frac{X_2 - \mathbf{E}X_2}{\sigma_{X_2}}\right) \\
&= P\left(X_1 = \frac{\sigma_{X_1} \cdot X_2 - \sigma_{X_1} \cdot EX_2}{\sigma_{X_2}} + \mathbf{E}X_1\right) \\
&= P\left(X_1 = \frac{\sigma_{X_1}}{\sigma_{X_2}} \cdot X_2 - \frac{\sigma_{X_1}}{\sigma_{X_2}} \cdot \mathbf{E}X_2 + \mathbf{E}X_1\right)
\end{aligned}$$

Wir definieren $a := \frac{\sigma_{X_1}}{\sigma_{X_2}} > 0$ und $b := \frac{\sigma_{X_1}}{\sigma_{X_2}} \cdot \mathbf{E}X_2 + EX_1$, und die Aussage ist für diesen Fall gezeigt.

$\underline{\varrho(X_1, X_2) = -1}$**:** Hier untersucht man die Varianz der Zufallsgröße $X_1^* + X_2^*$ und zeigt, daß sie ebenfalls gleich Null ist. Danach verläuft der Beweis völlig analog zum Fall $\varrho(X_1, X_2) = 1$.

$\square$

Bem. 17 *Eine Zufallsgröße, deren Erwartungswert gleich Null und deren Varianz gleich Eins sind, heißt standardisierte Zufallsgröße.*

Seien $X, Y \sim (0,1)$.

$$
\begin{aligned}
X^* &= X \\
Y^* &= \rho X + \sqrt{1-\rho^2} Y
\end{aligned}
$$

Offenbar

$$
\begin{aligned}
var X^* &= var Y^* = 1 \\
cov(X, Y) &= \rho.
\end{aligned}
$$

Seien $X, Y \sim \mathcal{N}(0,1)$, unabhängig, d.h. die gemeinsame Dichte ist

$$f(x,y) = \phi(x) \cdot \phi(y) = \frac{1}{2\pi} e^{-\frac{1}{2}(x^2+y^2)}$$

$$\begin{aligned} X^* &= X \\ Y^* &= \rho X + \sqrt{1-\rho^2}\, Y \end{aligned}$$

Wir suchen die gemeinsame Verteilung von (X^*, Y^*).

Transformation:

$$\begin{aligned} g_1(x,y) &= x \\ g_2(x,y) &= \rho x + \sqrt{1-\rho^2}\, y \end{aligned}$$

Inverse Transformation:

$$\begin{aligned}\psi_1(x^*, y^*) &= x^* \\ \psi_2(x^*, y^*) &= \frac{y^* - \rho x^*}{\sqrt{1-\rho^2}}\end{aligned}$$

Jacobi-Determinate

$$detJ = det \begin{pmatrix} 1 & 0 \\ \frac{-\rho}{\sqrt{1-\rho^2}} & \frac{1}{\sqrt{1-\rho^2}} \end{pmatrix} = \frac{1}{\sqrt{1-\rho^2}}$$

$$
\begin{aligned}
h(x^*, y^*) &= f(\psi_1(x^*, y^*), \psi_2(x^*, y^*)) \cdot |det(J)| \\
&= f(x^*, \frac{y^* - \rho x^*}{\sqrt{1-\rho^2}}) \cdot \frac{1}{\sqrt{1-\rho^2}} \\
&= \frac{1}{2\pi\sqrt{1-\rho^2}} e^{-\frac{1}{2}(x^{*2} + (\frac{y^* - \rho x^*}{\sqrt{1-\rho^2}})^2)} \\
&= \frac{1}{2\pi\sqrt{1-\rho^2}} e^{-\frac{1}{2(1-\rho^2)}(x^{*2} - 2\rho x^* y^* + y^{*2})}
\end{aligned}
$$

da der Exponent

$$
x^{*2} + (\frac{y^* - \rho x^*}{\sqrt{1-\rho^2}})^2 = \frac{1}{\sqrt{1-\rho^2}^2}\Big((1-\rho^2)x^{*2} + (y^* - \rho x^*)^2\Big)
$$

$h(x^*, y^*)$ ist Dichte der zweidimensionalen Normalverteilung.

Frohe Weihnachten

nächste Vorlesung: Mo., 7.1.08