

Im Folgenden Beschränkung auf den diskreten Fall und zweidimensionale Zufallsvariablen.

Vorstellung: Auswerten eines mehrdimensionalen Merkmals

$$\begin{pmatrix} \tilde{X} \\ \tilde{Y} \end{pmatrix}$$

also z.B.  $\omega \in \tilde{\Omega}$ , zufällig gezogene Person und damit  $\tilde{X}(\omega)$  und  $\tilde{Y}(\omega)$   
Auswertung der Merkmale jeweils an *derselben* Person.

⇒ zweidimensionale Zufallsvariable  $\begin{pmatrix} X \\ Y \end{pmatrix}$  (wie bei Zusammenhangsanalyse in Statistik I)

Das Hauptinteresse gilt (entsprechend der Kontingenztafel in Statistik I) der gemeinsamen Verteilung

$$P(\{X = x_i\} \cap \{Y = y_j\})$$



Betrachtet werden zwei eindimensionale diskrete Zufallselemente  $X$  und  $Y$  (zu demselben Zufallsexperiment). Die Wahrscheinlichkeit

$$P(X = x_i, Y = y_j) := P(\{X = x_i\} \cap \{Y = y_j\})$$

in Abhängigkeit von  $x_i$  und  $y_j$  heißt *gemeinsame Verteilung* der mehrdimensionalen Zufallsvariable  $\begin{pmatrix} X \\ Y \end{pmatrix}$  bzw. der Variablen  $X$  und  $Y$ .  
Randwahrscheinlichkeiten:

$$p_{i\bullet} = P(X = x_i) = \sum_{j=1}^m P(X = x_i, Y = y_j)$$

$$p_{\bullet j} = P(Y = y_j) = \sum_{i=1}^k P(X = x_i, Y = y_j)$$



## Bedingte Verteilungen

$$P(X = x_i | Y = y_j) = \frac{P(X = x_i, Y = y_j)}{P(Y = y_j)}$$

$$P(Y = y_j | X = x_i) = \frac{P(X = x_i, Y = y_j)}{P(X = x_i)}$$

**Stetiger Fall:** Zufallsvariable mit zweidimensionaler Dichtefunktion  $f(x, y)$ :

$$P(a \leq X \leq b, c \leq Y \leq d) = \int_a^b \left( \int_c^d f(x, y) dy \right) dx$$



## Kovarianz

### Definition

Seien  $X$  und  $Y$  zwei Zufallsvariablen. Dann heißt

$$\sigma_{X,Y} := \text{Cov}(X, Y) = \mathbb{E}((X - \mathbb{E}(X))(Y - \mathbb{E}(Y)))$$

Kovarianz von  $X$  und  $Y$ .



- $\text{Cov}(X, X) = \text{Var}(X)$
- $\text{Cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X) \cdot \mathbb{E}(Y)$
- $\text{Cov}(X, Y) = \text{Cov}(Y, X)$
- Mit  $\tilde{X} = a_X X + b_X$  und  $\tilde{Y} = a_Y Y + b_Y$  ist
 
$$\text{Cov}(\tilde{X}, \tilde{Y}) = a_X \cdot a_Y \cdot \text{Cov}(X, Y)$$
- $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + 2 \cdot \text{Cov}(X, Y)$

## Definition

Zwei Zufallsvariablen  $X$  und  $Y$  mit  $\text{Cov}(X, Y) = 0$  heißen unkorreliert.

- Stochastisch unabhängige Zufallsvariablen sind unkorreliert. Die Umkehrung gilt jedoch im allgemeinen nicht.
- Vergleiche Statistik I: Kovarianz misst nur lineare Zusammenhänge.



## Korrelationskoeffizient

## Definition

Gegeben seien zwei Zufallsvariablen  $X$  und  $Y$ . Dann heißt

$$\rho(X, Y) = \frac{\text{Cov}(X, Y)}{\sqrt{\text{Var}(X)}\sqrt{\text{Var}(Y)}}$$

Korrelationskoeffizient von  $X$  und  $Y$ .



## Eigenschaften des Korrelationskoeffizienten

- Mit  $\tilde{X} = a_X X + b_X$  und  $\tilde{Y} = a_Y Y + b_Y$  ist
 
$$|\rho(\tilde{X}, \tilde{Y})| = |\rho(X, Y)|.$$
- $-1 \leq \rho(X, Y) \leq 1$ .
- $|\rho(X, Y)| = 1 \iff Y = aX + b$
- Sind  $\text{Var}(X) > 0$  und  $\text{Var}(Y) > 0$ , so gilt  $\rho(X, Y) = 0$  genau dann, wenn  $\text{Cov}(X, Y) = 0$ .



## Beispiel: Chuck a Luck

$X_1$  Gewinn, wenn beim ersten Wurf ein Einsatz auf 1 gesetzt wird.

$X_6$  Gewinn, wenn beim ersten Wurf ein Einsatz auf 6 gesetzt wird.

Kovarianz zwischen  $X_1$  und  $X_6$ :

$(x_1, x_6)$	$P(X_1 = x_1, X_6 = x_6)$	$(x_1, x_6)$	$P(X_1 = x_1, X_6 = x_6)$
$(-1, -1)$	$\frac{64}{216}$	$(-1, 3)$	$\frac{1}{216}$
$(-1, 1)$	$\frac{48}{216}$	$(3, -1)$	$\frac{1}{216}$
$(1, -1)$	$\frac{48}{216}$	$(1, 1)$	$\frac{24}{216}$
$(-1, 2)$	$\frac{12}{216}$	$(1, 2)$	$\frac{3}{216}$
$(2, -1)$	$\frac{12}{216}$	$(1, 2)$	$\frac{3}{216}$

## Berechnungen

$$\Rightarrow E(X_1 \cdot X_6) = -50/216 = -0.23148$$

$$\text{Cov}(X_1, X_6) = -0.23148 - (-0.0787) \cdot (-0.0787) = -0.23768$$

$X_1$  und  $X_6$  sind negativ korreliert.