11 Mehrdimensionale Zufallsvariablen

11.1 Begriffe

Def. 34 *Es seien X_i, $i = 1, \ldots, p$ reellwertige, zufällige Variablen auf dem Wahrscheinlichkeitsraum $(\Omega, \mathcal{E}, P)$. Dann heißt*

$$\mathbf{X} = (X_1, \ldots, X_p)^T \quad : \Omega \longrightarrow \mathbb{R}^p,$$

zufälliger Vektor.

Er transformiert den Wahrscheinlichkeitsraum $(\Omega, \mathcal{E}, P)$ in den Wahrscheinlichkeitsraum $(\mathbb{R}^p, \mathcal{B}^p, P_X)$, wobei $\mathcal{B}^p$ die σ–Algebra der p–dimensionalen Borelmengen ist.

Def. 35 *Die Funktion*

$$F_{\mathbf{X}}(x_1,\ldots,x_p) := P(\{\omega\colon X_1(\omega) < x_1,\ldots,X_p(\omega) < x_p\})$$

heißt <u>*Verteilungsfunktion*</u> *des zufälligen Vektors* **X**. *Sie wird auch mit* $\underline{F_{X_1,\ldots,X_p}(x_1,\ldots,x_p)}$ *bezeichnet.*

Es gilt:

$$F_{X_1,\ldots,X_p}(x_1,\ldots,x_p) = P\left(\bigcap_{i=1}^{p}\{\omega\in\Omega\colon X_i(\omega) < x_i\}\right).$$

Die Verteilungsfunktion für zufällige Vektoren besitzt folgende Eigenschaften:

1. Invarianz gegenüber Permutationen, d.h.

$$F_{X_1,\ldots,X_p}(x_1,\ldots,x_p) = F_{X_{i_1},\ldots,X_{i_p}}(x_{i_1},\ldots,x_{i_p})$$

2. $\lim\limits_{x_p\to\infty} F_{\mathbf{X}}(x_1,\ldots,x_p) = F_{X_1,\ldots,X_{p-1}}(x_1,\ldots,x_{p-1})$;

$$F_X(x_1,\ldots,x_p) =$$

$$P(\underbrace{\{\omega\colon X_1(\omega) < x_1\} \cap \ldots \cap \{\omega\colon X_{p-1}(\omega) < x_{p-1}\}}_{=:A} \cap$$

$$\underbrace{\{\omega\colon X_p(\omega) < x_p\}}_{\longrightarrow_{x_p\to\infty}\Omega}.$$

Damit gilt also:

$$\begin{aligned}\lim_{x_p\to\infty} F_{\mathbf{X}}(x_1,\ldots,x_p) &= P(A\cap\Omega) = P(A)\\ &= F_{X_1,\ldots,X_{p-1}}(x_1,\ldots,x_{p-1}).\end{aligned}$$

3. $\lim\limits_{x_p\to-\infty} F_{\mathbf{X}}(x_1,\ldots,x_p) = 0$;

   Bem.: Man kann wegen 1. auch jede beliebige Komponente wählen!

4. $\lim\limits_{(x_1,\ldots,x_p)\to(\infty,\ldots,\infty)} F_X(x_1,\ldots,x_p) = 1$;

5. $F_{\mathbf{X}}(x_1,\ldots,x_p)$ ist in jedem Argument monoton wachsend;

6. $F_{\mathbf{X}}(x_1,\ldots,x_p)$ ist in jedem Argument linksseitig stetig.

Def. 36 *Ein zufälliger Vektor* $\mathbf{X} = (X_1, \ldots, X_p)^T$ *heißt* *stetig verteilt, wenn seine Verteilungsfunktion charakterisiert ist durch:*

$$F(x_1, \ldots, x_p) = \int_{-\infty}^{x_1} \ldots \int_{-\infty}^{x_p} f(t_1, \ldots, t_p)\, dt_p \ldots dt_1,$$

wobei für die Funktion f *gilt:*

1. $f(x_1, \ldots, x_p) \geq 0, \ \forall x_1, \ldots, x_p$;
2. $\int_{\mathbb{R}^p} f(x_1, \ldots, x_p)\, dx_1 \ldots dx_p = 1.$

Die Funktion $f(x_1, \ldots, x_p)$ *heißt dann Dichtefunktion des zufälligen Vektors* $\mathbf{X}$.

Falls die Dichtefunktion $f(x_1, \ldots, x_p)$ stetig ist, so gilt:

$$f(x_1, \ldots, x_p) = \frac{\partial^p F_X(x_1, \ldots, x_p)}{\partial x_1 \ldots \partial x_p}.$$

Def. 37 *Ein zufälliger Vektor* $\mathbf{X} = (X_1, \ldots, X_p)^T$ *heißt*

- *diskret, falls jede Komponente von* $\mathbf{X}$ *diskret ist, d.h. jedes* X_i *besitzt höchstens abzählbar viele Argumente.*

- *mixed(gemischt), falls einige seiner Komponenten diskret, die restlichen dagegen stetig sind.*

- *stetig, falls alle Komponenten von* $\mathbf{X}$ *stetige Zufallsgrößen sind.*

Def. 38 *Es sei* $\mathbf{X} = (X_1, \ldots, X_p)^T$ *ein diskreter zufälliger Vektor. Für* $i = 1, \ldots, p$ *habe* X_i *den Wertevorrat* $\{x_{i1}, \ldots, x_{ik}, \ldots\}$. *Dann definieren wir:*

$$p_{j\ldots k} = P(X_1 = x_{1j}, \ldots, X_p = x_{pk}).$$

Für die Verteilungsfunktion des zufälligen Vektors $\mathbf{X}$ gilt:

$$\begin{aligned} F(x_1, \ldots, x_p) &= P\left(\bigcap_{i=1}^{p} \{\omega \in \Omega \colon X_i(\omega) < x_i\}\right) \\ &= \sum_{j\colon x_{1j} < x_1} \sum_{\ldots} \sum_{k\colon x_{pk} < x_p} p_{j\ldots k} \end{aligned}$$

Es sei nun $p = 2$ und wir betrachten den zufälligen Vektor $\mathbf{X} = (X_1, X_2)^T$. Wir untersuchen zwei Fälle:

– **X ist diskret.** Dann gilt zunächst:

$$X_1 : \begin{pmatrix} x_1 & x_2 & \ldots & x_n & \ldots \\ p_1 & p_2 & \ldots & p_n & \ldots \end{pmatrix}$$

$$X_2 : \begin{pmatrix} y_1 & y_2 & \ldots & y_n & \ldots \\ q_1 & q_2 & \ldots & q_n & \ldots \end{pmatrix}$$

Daraus folgt (Definition 38):

$$p_{ij} = P(X_1 = x_i, X_2 = y_j) = P(\mathbf{X} = (x_i, y_j)).$$

Weiterhin gilt:

$$\begin{aligned} P(X_1 \in \{x_i : i \in \mathbb{N}\}) &= 1 \\ P(X_2 \in \{y_j : j \in \mathbb{N}\}) &= 1 \end{aligned}$$

Wir bezeichnen:

$$\mathcal{X} := \{x_i : i \in \mathbb{N}\}, \qquad \mathcal{Y} := \{y_j : j \in \mathbb{N}\}.$$

Der zufällige Vektor $\mathbf{X}$ kann Werte der Form $(x_i, y_j) \in \mathcal{X} \times \mathcal{Y}$ annehmen. Folglich gilt:

$$P(\mathbf{X} \in \mathcal{X} \times \mathcal{Y}) = P(X_1 \in \mathcal{X}, X_2 \in \mathcal{Y}) = \sum_{i,j \in \mathbb{N}} p_{ij} = 1.$$

Es könnte evtl. interessant sein, die Wahrscheinlichkeit zu untersuchen, daß eine der zufälligen Variablen X_1

bzw. X_2 einen bestimmten Wert x_i bzw. y_j annimmt.

$$
\begin{aligned}
P(X_1 = x_i) &= P(\{X_1 = x_i\} \cap \Omega) \\
&= P(\{X_1 = x_i\} \cap \\
&\underbrace{\{(X_2 = y_1) \vee (X_2 = y_2) \vee \ldots \vee (X_2 = y_n) \vee \ldots\}}_{=\bigcup_{j \in \mathbb{N}} \{X_2 = y_j\} = \Omega}) \\
&= P\left(\{X_1 = x_i\} \cap \left(\bigcup_{j \in \mathbb{N}} \{X_2 = y_j\}\right)\right) \\
&= P\left(\bigcup_{j \in \mathbb{N}} \{(X_1 = x_i) \,\&\, (X_2 = y_j)\}\right) \\
&= \sum_{j \in \mathbb{N}} p_{ij} =: p_{i\cdot}
\end{aligned}
$$

Wir erhalten also:

$$p_{i\cdot} = P(X_1 = x_i).$$

Analog erhalten wir, wenn wir die Betrachtung mit der zufälligen Variable X_2 wiederholen:

$$p_{\cdot j} = P(X_2 = y_j).$$

Def. 39 *Die Wahrscheinlichkeiten $p_{i\cdot}$ bzw. $p_{\cdot j}$ ($i, j \in \mathbb{N}$) nennen wir die Randwahrscheinlichkeiten des zufälligen Vektors $X = (X_1, X_2)^T$.*

Das folgende Schema verdeutlicht noch einmal die Zusammenhänge zwischen den einzelnen Wahrscheinlichkeiten:

$X_1 \backslash X_2$	y_1	y_2	y_3	$\ldots$	y_j	$\ldots$	y_n	$\ldots$	$\sum$
x_1	p_{11}	p_{12}	p_{13}	$\ldots$	p_{1j}	$\ldots$	p_{1n}	$\ldots$	$p_{1\cdot}$
x_2	p_{21}	p_{22}	p_{23}	$\ldots$	p_{2j}	$\ldots$	p_{2n}	$\ldots$	$p_{2\cdot}$
x_3	p_{31}	p_{32}	p_{33}	$\ldots$	p_{3j}	$\ldots$	p_{3n}	$\ldots$	$p_{3\cdot}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$		$\vdots$
x_i	p_{i1}	p_{i2}	p_{i3}	$\ldots$	p_{ij}	$\ldots$	p_{in}	$\ldots$	$p_{i\cdot}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$		$\vdots$
x_n	p_{n1}	p_{n2}	p_{n3}	$\ldots$	p_{nj}	$\ldots$	p_{nn}	$\ldots$	$p_{n\cdot}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\vdots$		$\vdots$
$\sum$	$p_{\cdot 1}$	$p_{\cdot 2}$	$p_{\cdot 3}$	$\ldots$	$p_{\cdot j}$	$\ldots$	$p_{\cdot n}$	$\ldots$	1

Bsp. 64 *Es wird eine Umfrage zum Thema „Rauchen“ durchgeführt. Dabei werden Männer und Frauen darüber befragt, ob sie Raucher oder Nichtraucher sind. Das ergibt die beiden folgenden Zufallsvariablen:*

$$X_1 = \begin{cases} 1 & \textit{, falls weiblich} \\ 2 & \textit{, falls männlich} \end{cases}$$

$$X_2 = \begin{cases} 1 & \textit{, falls Raucher} \\ 2 & \textit{, falls Nichtraucher} \end{cases}$$

Folglich erhalten wir für den zufälligen Vektor

$X = (X_1, X_2)^T$ *das folgende Schema:*

$X_1 \backslash X_2$	1	2
1	p_{11}	p_{12}
2	p_{21}	p_{22}

Wir erstellen nun eine 2×2*–Kontingenztafel. Dabei wird für jedes der vier Felder anhand der statistischen Erhebung die Anzahl der Personen ermittelt, für die die beiden*

Eigenschaften zutreffen, die dieses Feld charakterisieren.

$X_1 \backslash X_2$	1	2	
1	n_{11}	n_{12}	$n_{1.}$
2	n_{21}	n_{22}	$n_{2.}$
	$n_{.1}$	$n_{.2}$	$n_{..}$

Dabei bedeuten:

n_{ij} – *die Anzahl der Personen mit dem Geschlecht* i *und dem Raucherverhalten* j*;*

$n_{.1}$ – *die Anzahl der Raucher;*

$n_{.2}$ – *die Anzahl der Nichtraucher;*

$n_{1\cdot}$ *– die Anzahl der Frauen;*

$n_{2\cdot}$ *– die Anzahl der Männer;*

$n_{\cdot\cdot}$ *– die Gesamtzahl der Befragten.*

Mit $\hat{p}_{ij} = \frac{n_{ij}}{n_{\cdot\cdot}}$ *ergibt sich nun eine Schätzung für die Wahrscheinlichkeit* p_{ij}*.*

Bsp. 65 *Werfen zweier Würfel. Wir betrachten den zufälligen Vektor* $X = (X_1, X_2)^T$*, wobei* X_1 *die Augenzahl des ersten Würfels ist und* X_2 *die des zweiten. Für die zufälligen Variablen* X_1 *und* X_2 *gilt:*

$$X_1 : \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \end{pmatrix}$$

$$X_2 : \begin{pmatrix} 1 & 2 & 3 & 4 & 5 & 6 \\ \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} & \frac{1}{6} \end{pmatrix}$$

Da die Würfel voneinander unabhängig sind, gilt (Definition 10):

$$\begin{aligned} p_{ij} &= P(X_1 = i, X_2 = j) \\ &= P(\{X_1 = i\} \cap \{X_2 = j\}) \\ &= \frac{1}{6} \cdot \frac{1}{6} = \frac{1}{36} \end{aligned}$$

Damit erhalten wir das folgende Schema:

$X_1 \backslash X_2$	1	2	3	4	5	6
1	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$
2	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$
3	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$
4	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$
5	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$
6	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$	$\frac{1}{36}$

Wir wollen nun wissen, wie groß die Wahrscheinlichkeit ist, daß der erste Würfel weniger als vier Augen zeigt und der zweite Würfel weniger als drei. Wir wissen bereits, daß die *beiden Würfel voneinander unabhängig sind. Es gilt:*

$$P(X_1 < 4, X_2 < 3) = \sum_{i<4;\, j<3} p_{ij} = \frac{6}{36} = \frac{1}{6}$$

Die hier addierten Wahrscheinlichkeiten sind in dem oben angegebenen Schema eingerahmt.

Die Aussagen zu zweidimensionalen zufälligen Vektoren, die wir bis hierher gemacht haben, gelten analog erweitert auch für höherdimensionale zufällige Vektoren.

– X **ist stetig.** In diesem Fall erhalten wir eine zweidimensionale Dichtefunktion $f(x, y)$, für die gilt (vgl. Definition 36):

1. $\int\limits_{-\infty}^{+\infty} \int\limits_{-\infty}^{+\infty} f(x, y)\, dx\, dy = 1;$

2. $f(x, y) \geq 0, \forall (x, y) \in \mathbb{R}^2.$

Die Verteilungsfunktion ergibt sich dann nach Definition 36 durch:

$$F(x, y) = \int\limits_{-\infty}^{x} \int\limits_{-\infty}^{y} f(u, v)\, du\, dv = P(X_1 < x, X_2 < y).$$

Da die Dichtefunktion $f(x, y)$ stetig ist, gilt weiterhin:

$$f(x, y) = \frac{\partial^2 F(x, y)}{\partial x\, \partial y}.$$

Offenbar

$$\lim_{y\to\infty} F(x,y) = F_{X_1}(x) = P(X_1 < x).$$

$$\lim_{x\to\infty} F(x,y) = F_{X_2}(y) = P(X_2 < y).$$

Def. 40 *Die Verteilungsfunktionen F_{X_1} und F_{X_2} bezeichnen wir als Randverteilungen von X_1 bzw. X_2.*

Integrieren wir die Dichtefunktion nur nach einer der beiden Variablen, so erhalten wir:

$$\int\limits_{-\infty}^{+\infty} f(x,y)\,dy = \frac{dF_{X_1}(x)}{dx} =: f_{X_1}(x) \qquad (2)$$

Analog gilt:

$$\int_{-\infty}^{+\infty} f(x,y)\,dx = \frac{dF_{X_2}(y)}{dy} =: f_{X_2}(y).$$

Def. 41 *Die Funktionen* f_{X_1} *und* f_{X_2} *bezeichnen wir als* *<u>Randdichten</u>* *von* X_1 *bzw.* X_2.

Offenbar,

$$F_{X_1}(x) = \int_{-\infty}^{x} f_{X_1}(t)\,dt$$

$$F_{X_2}(y) = \int_{-\infty}^{y} f_{X_2}(t)\,dt$$

Bsp. 66 *Zweidimensionale Normalverteilung*

`Descr_normal_2D`

11.2 Unabhängigkeit von Zufallsgrößen

Def. 42 *Es seien X_1 und X_2 zwei zufällige Variablen auf dem Wahrscheinlichkeitsraum $(\Omega, \mathcal{E}, P)$. Diese beiden zufälligen Variablen X_1 und X_2 heißen stochastisch unabhängig, wenn für alle $A, B \in \mathcal{B}^1$ gilt:*

- $P(X_1 \in A, X_2 \in B) = P(X_1 \in A) \cdot P(X_2 \in B)$;
- $P(A \cap B) = P(A) \cdot P(B)$.

Erinnerung:

$$P(X_1 \in A, X_2 \in B) = P(\{\omega\colon X_1(\omega) \in A\} \cap \{\omega\colon X_2(\omega) \in B\}).$$

Es sei $X = (X_1, X_2)^T$ ein zufälliger Vektor, für den gilt, daß die zufälligen Variablen X_1 und X_2 stochastisch unabhängig sind. Dann gilt:

$$
\begin{aligned}
F_{X_1,X_2}(x_1, x_2) &= P(X_1 < x_1, X_2 < x_2) \\
&= P(X_1 \in \underbrace{(-\infty, x_1)}_{A \in \mathcal{B}^1}, X_2 \in \underbrace{(-\infty, x_2)}_{B \in \mathcal{B}^1}) \\
&= P(X_1 \in (-\infty, x_1)) \cdot P(X_2 \in (-\infty, x_2)) \\
&= F_{X_1}(x_1) \cdot F_{X_2}(x_2)
\end{aligned}
$$

- Ist der zufällige Vektor $X = (X_1, X_2)^T$ außerdem stetig, so folgt:

$$\underbrace{f_{X_1,X_2}(x_1,x_2)}_{\text{zweidimensio– nale Dichte}} = \underbrace{f_{X_1}(x_1) \cdot f_{X_2}(x_2)}_{\text{Randdichten}}.$$

- Ist der zufällige Vektor $X = (X_1, X_2)^T$ außerdem diskret, so folgt für alle $i, j = 1, \ldots$:

$$p_{ij} = p_{i.} \cdot p_{.j}.$$

Bsp. 67 *In der folgenden Tabelle sind die Einzelwahrscheinlichkeiten p_{ij} einer diskreten zweidimensionalen Zufallsvariablen eingetragen. Die Komponenten X und Y seien unabhängig. Zunächst seien nur die fett eingezeichneten Einträge bekannt. Bestimmen Sie die restlichen Einträge! Zur Kontrolle sind sie blau eingetragen.*

$X \backslash Y$	1	2	3	$p_{i.}$
-1	0.02	**0.06**	0.12	0.20
0	**0.03**	0.09	0.18	0.30
1	0.05	0.15	0.3	**0.50**
$p_{.j}$	0.1	**0.30**	0.60	1

$$
\begin{aligned}
\mathbf{E}X &= -1\cdot 0.2 + 0\cdot 0.3 + 1\cdot 0.5 = 0.3\\
\mathbf{E}Y &= 1\cdot 0.1 + 2\cdot 0.3 + 3\cdot 0.6 = 2.5\\
\mathbf{E}(X\cdot Y) &= -0.02 - 2\cdot 0.06 - 3\cdot 0.12 + 0\cdot(\ldots)\\
&\quad +1\cdot 0.05 + 2\cdot 0.15 + 3\cdot 0.3 = 0.75\\
cov(X,Y) &= \mathbf{E}(X\cdot Y) - (\mathbf{E}X)(\mathbf{E}Y) = 0.75 - 0.75 = 0.
\end{aligned}
$$

Merkwürdig?

Satz 30 *Es seien* X_1 *und* X_2 *zwei zufällige Variablen.* φ *und* ψ *seien zwei beliebige* $\mathcal{B}^1$*–meßbare Transformationen dieser beiden Variablen,*

$$X_1' = \varphi(X_1), \quad X_2' = \psi(X_2).$$

Die zufälligen Variablen X_1 *und* X_2 *sind genau dann stochastisch unabhängig, wenn die Zufallsgrößen* X_1' *und* X_2'*, für alle Transformationen* φ *und* ψ*, unabhängig sind.*

Beweis: Die Funktionen φ und ψ seien auf der Menge $\mathbb{R}$ definiert und reellwertig. Dann gilt für die jeweilige

Umkehrfunktion genau dann

$$\begin{aligned}\varphi^{-1}(A) &= \{x\colon \varphi(x)\in A\}\in\mathcal{B}^1, \quad \forall A\in\mathcal{B}^1\\ \psi^{-1}(B) &= \{y\colon \psi(y)\in B\}\in\mathcal{B}^1, \quad \forall B\in\mathcal{B}^1,\end{aligned}$$

wenn φ und ψ $\mathcal{B}^1$–meßbar sind.

($\Longrightarrow$) Es seien die zufälligen Variablen X_1 und X_2 stochastisch unabhängig. Wir zeigen, daß die zufälligen Variablen $\varphi(X_1)$ und $\psi(X_2)$ unabhängig sind. Da die Funktionen φ und ψ $\mathcal{B}^1$–meßbar sind, gilt

$$
\begin{aligned}
&P(\varphi(X_1) \in A, \psi(X_2) \in B) \\
&\qquad = \quad P(X_1 \in \varphi^{-1}(A), X_2 \in \psi^{-1}(B)) \\
&\qquad = \quad P(X_1 \in \varphi^{-1}(A)) \cdot P(X_2 \in \psi^{-1}(B)) \\
&\qquad = \quad P(\varphi(X_1) \in A) \cdot P(\psi(X_2) \in B)
\end{aligned}
$$

D.h. die zufälligen Variablen $\varphi(X_1)$ und $\psi(X_2)$ sind unabhängig.

($\Longleftarrow$) Es gelte also, daß für alle $\mathcal{B}^1$–meßbaren Funktionen φ und ψ die zufälligen Variablen $\varphi(X_1)$ und $\psi(X_2)$ unabhängig sind. Insbesondere ist das dann auch der Fall für die Funktionen $\varphi(x) \equiv \psi(x) \equiv x$. Das heißt aber nichts anderes, als daß gilt:

$$X_1 = \varphi(X_1), \quad X_2 = \psi(X_2).$$

Folglich sind auch die zufälligen Variablen X_1 und X_2 unabhängig.

$\square$

Bsp. 68 *Sei* $X \sim \mathcal{N}(0,1)$.

- X *und* $Y = X^2$ *sind nicht unabhängig, sogar funktional abhängig*
- X *und* Y *sind unkorreliert, wegen*

  $$\mathbf{E}X = 0,\ \ cov(X,Y) = \mathbf{E}(X \cdot X^2) - \mathbf{E}X \cdot \mathbf{E}Y = \mathbf{E}X^3 = 0,$$

  da X *symmetrisch ist.*

Die Aussage gilt also für beliebige symmetrische Zufallsvariablen X *mit endlicher Varianz.*

11.3 Transformationssatz für Zufallsvektoren

Es sei $\mathbf{X} = (X_1, \ldots, X_p)^T$ ein zufälliger Vektor. Er habe die Dichtefunktion $f(x_1, \ldots, x_p)$. Es sei $g\colon \mathbb{R}^p \longrightarrow \mathbb{R}^p$ eine umkehrbar eindeutige Abbildung. Sie ordnet einem Vektor $\mathbf{x} = (x_1, \ldots, x_p)^T$ einen Vektor $\mathbf{y} = (y_1, \ldots, y_p)^T$ zu und besteht aus Teilabbildungen $g_1, \ldots, g_p$ mit $g_i\colon \mathbb{R}^p \longrightarrow \mathbb{R}$ (für alle $i = 1, \ldots, p$).

Bsp. 69 $\mathbf{y} = g(\mathbf{x}) = \mathbf{A} \cdot \mathbf{x}$, *wobei* $\mathbf{A}$ *eine reguläre* (p, p)*–Matrix ist.*

Die Umkehrabbildung $g^{-1}\colon \mathbb{R}^p \longrightarrow \mathbb{R}^p$ ist durch Funktionen ψ_i ($i = 1, \ldots, p$) definiert, die einem Vektor

$\mathbf{y} = (y_1, \ldots, y_p)^T$ eine Zahl x_i zuordnen, d.h. $x_i = \psi_i(y_1, \ldots, y_p)$ ($i = 1, \ldots, p$). Diese Funktionen ψ_i ($i = 1, \ldots, p$) existieren aufgrund der umkehrbaren Eindeutigkeit der Funktion g. Die Funktion g^{-1} ist dann folgendermaßen definiert:

$$g^{-1}(\mathbf{y}) \;=\; g^{-1}\begin{pmatrix} y_1 \\ \vdots \\ y_p \end{pmatrix} = \begin{pmatrix} \psi_1(y_1, \ldots, y_p) \\ \vdots \\ \psi_p(y_1, \ldots, y_p) \end{pmatrix} = \begin{pmatrix} x_1 \\ \vdots \\ x_p \end{pmatrix}$$

In Kurzform,

$$g^{-1}(\mathbf{y}) = (\psi_1(\mathbf{y}), \ldots, \psi_p(\mathbf{y}))^T = \mathbf{x}.$$

Wir definieren nun einen weiteren zufälligen Vektor $\mathbf{Y} = (Y_1, \ldots, Y_p)^T$ wie folgt:

$$\mathbf{Y} = g(\mathbf{X}) := (g_1(X_1, \ldots, X_p), \ldots, g_p(X_1, \ldots, X_p))^T.$$

Die Abbildung $g = (g_1, \ldots, g_p)$ bildet die Transformation des zufälligen Vektors $\mathbf{X}$ in den zufälligen Vektor $\mathbf{Y}$. Wir nehmen an, die g_i ($i = 1, \ldots, p$) besitzen stetige partielle Ableitungen nach allen Argumenten.

Für den zufälligen Vektor $\mathbf{X}$ gilt umgekehrt:

$$\begin{aligned} \mathbf{X} &= (X_1, \ldots, X_p)^T \\ &= (\psi_1(Y_1, \ldots, Y_p), \ldots, \psi_p(Y_1, \ldots, Y_p))^T \\ &= g^{-1}(Y_1, \ldots, Y_p) = g^{-1}(\mathbf{Y}). \end{aligned}$$

Wir suchen nun die Dichtefunktion h_Y des zufälligen Vektors $\mathbf{Y}$. Analog zur Transformation zufälliger Variablen (vgl. Satz 29 auf Seite 370) gilt der folgende Satz (ohne Beweis)

Satz 31 *Die Zufallsvariable $\mathbf{X}$ habe die Dichte f. Die Dichte der Zufallsvariablen $\mathbf{Y} = g(\mathbf{X})$ ist*

$$h_Y(y_1, \ldots, y_p) = f(\psi_1(y_1, \ldots, y_p), \ldots, \psi_p(y_1, \ldots, y_p)) \cdot |J|,$$

wobei

$$J = \det \left(\frac{\partial \psi_i(y_1, \ldots, y_p)}{\partial y_j} \right)_{i,j=1,\ldots,p}$$

die sogenannte Jacobi-Determinante ist.

J ist also die Determinante der Matrix der partiellen Ableitungen der Funktion g^{-1}.

Bsp. 70 *Es sei* $\mathbf{X} = (X_1, X_2)^T$ *ein zufälliger Vektor (*$p = 2$*), mit unabhängigen Komponenten* X_1 *und* X_2*. Die Dichte* f_{X_1,X_2} *von* $\mathbf{X}$ *sei*

$$f_{X_1,X_2}(x_1, x_2) = f_{X_1}(x_1) \cdot f_{X_2}(x_2).$$

Es sei $\mathbf{Y} = g(\mathbf{X})$ *ein weiterer zufälliger Vektor,*

$$\mathbf{Y} = \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} = \begin{pmatrix} g_1(X_1, X_2) \\ g_2(X_1, X_2) \end{pmatrix} = \begin{pmatrix} X_1 + X_2 \\ X_2 \end{pmatrix}.$$

Wir suchen nun die Dichte des zufälligen Vektors $\mathbf{Y}$*.*

Zunächst wissen wir, daß die Funktion g *aus zwei*

Teilfunktionen g_1 und g_2 besteht:

$$\begin{aligned} g_1(X_1, X_2) &= Y_1 = X_1 + X_2 \\ g_2(X_1, X_2) &= Y_2 = X_2 \end{aligned}$$

bzw.

$$\begin{aligned} g_1(x_1, x_2) &= y_1 = x_1 + x_2 \\ g_2(x_1, x_2) &= y_2 = x_2 \end{aligned}$$

Die Umkehrfunktion g^{-1} besteht aus den beiden Teilfunktionen:

$$\begin{aligned} \psi_1(y_1, y_2) &= x_1 = y_1 - y_2 \\ \psi_2(y_1, y_2) &= x_2 = y_2 \end{aligned}$$

Wir bestimmen nun die Zahl $|J|$:

$$|J| \;=\; \left|\det\begin{pmatrix} \frac{\partial x_1}{\partial y_1} & \frac{\partial x_1}{\partial y_2} \\ \frac{\partial x_2}{\partial y_1} & \frac{\partial x_2}{\partial y_2} \end{pmatrix}\right| = \left|\det\begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix}\right| = 1$$

Dichte des zufälligen Vektors $\mathbf{Y} = (X_1 + X_2, X_2)$:

$$\begin{aligned} h_{\mathbf{Y}}(y_1, y_2) &= f_{X_1,X_2}(\psi_1(y_1, y_2), \psi_2(y_1, y_2)) \cdot |1| \\ &= f_{X_1,X_2}(y_1 - y_2, y_2) \\ &= \underline{f_{X_1}(y_1 - y_2) \cdot f_{X_2}(y_2)} \end{aligned}$$

Randdichte für Y_1:

$$\begin{aligned} h_{\mathbf{Y}_1}(y_1) &= \int\limits_{-\infty}^{+\infty} h_{\mathbf{Y}}(y_1, y_2)\, dy_2 = \int\limits_{-\infty}^{+\infty} f_{X_1}(y_1 - y_2) \cdot f_{X_2}(y_2)\, dy_2 \\ &=: f_{X_1} * f_{X_2}(y) \end{aligned}$$

Def. 43 *Die Verknüpfung $f_{X_1} * f_{X_2}$ zweier Funktionen f_1 und f_2 heißt <u>Faltung</u> aus f_1 und f_2.*

Bem.: Die Dichte der Summe zweier unabhängiger Zufallsvariablen ist Faltung der beiden Einzeldichten.

Bsp. 71 *Es seien $X_1, X_2 \sim R(0,1)$, d.h.*

$$f_{X_i}(x) = \begin{cases} 1 & , \textit{falls } 0 \leq x < 1 \\ 0 & , \textit{sonst} \end{cases}$$

Sei Y der zufällige Vektor aus Beispiel 70. Für die Dichtefunktion der zufälligen Variablen $Y_1 = X_1 + X_2$ gilt *(mittels Faltung):*

$$
\begin{aligned}
h_{Y_1}(y) &= \int_0^1 f_{X_1}(y-x) \cdot \underbrace{f_{X_2}(x)}_{\equiv 1} \, dx \\
&= \int_0^1 f_{X_1}(y-x)\, dx
\end{aligned}
$$

Nun gilt: $0 \leq X_i < 1, \quad i = 1, 2.$, *d.h.*

$$0 \leq X_1 + X_2 = Y_1 < 2.$$

und für die Funktion f_{X_1}*:*

$$f_{X_1}(y-x) = \begin{cases} 1 & , \textit{falls } 0 \le y-x \le 1 \\ 0 & , \textit{sonst} \end{cases}$$

$$= \begin{cases} 1 & , \textit{falls } 0 \le y-1 \le x \le 1 < y \\ 1 & , \textit{falls } 0 \le x < y \le 1 \\ 0 & , \textit{falls } y-x \notin [0,1[\end{cases}$$

Randdichte Y_1 *von* $\mathbf{Y}$*:*

$$h_{Y_1}(y) = \int\limits_0^1 f_{X_1}(y-x)\,dx$$

$$= \begin{cases} \int\limits_{y-1}^{1} dx & \text{, falls } 1 < y < 2 \\ \int\limits_{0}^{y} dx & \text{, falls } 0 \le y \le 1 \\ 0 & \text{, falls } y \notin [0,2[\end{cases}$$

$$= \begin{cases} 2-y & \text{, falls } 1 < y < 2 \\ y & \text{, falls } 0 \le y \le 1 \\ 0 & \text{, falls } y \notin [0,2[\end{cases}$$

Wir addieren nun drei zufällige Variablen X_1, X_2, X_3,
$X_i \sim R(0,1)$,

$$Y_3 = (X_1 + X_2) + X_3.$$

Für die Dichtefunktion der Zufallsgröße Y_3 gilt dann nach der Faltungsformel:

$$\begin{aligned} h_{Y_3}(z) &= h_{Y_1} * f_{X_3}(z) = \int\limits_{-\infty}^{+\infty} h_{Y_1}(z-x) \cdot f_{X_3}(x)\,dx \\ &= \int\limits_{0}^{1} h_{Y_1}(z-x) \cdot f_{X_3}(x)\,dx = \int\limits_{0}^{1} h_{Y_1}(z-x)\,dx \end{aligned}$$

Wegen $0 \le X_i < 1$ *folgt dann:*

$$0 \le (X_1 + X_2) + X_3 = Y_1 + X_3 = Y_3 < 3.$$

Für die Dichte der Summe der drei Zufallsgrößen X_1*,* X_2 *und* X_3 *gilt also:*

$$h_{Y_3}(z) = \begin{cases} 0 & \text{, falls } z \notin [0,3[\\ \frac{z^2}{2} & \text{, falls } 0 \le z \le 1 \\ 1 - \frac{(z-1)^2}{2} - \frac{(2-z)^2}{2} & \text{, falls } 1 < z \le 2 \\ \frac{(3-z)^2}{2} & \text{, falls } 2 < z < 3 \end{cases}$$

`/sasuser/Stochastik/ZGWS.sas`

Vermutung: Die Summe unabhängiger Zufallsgrößen nähert sich bei wachsender Zahl der Zufallsgrößen einer Normalverteilung.

Diese Vermutung ist richtig. Sie gilt sogar (unter sehr allgemeinen Voraussetzungen) unabhängig davon, welche Verteilung diese Zufallsgrößen vorher hatten (Normal–, Gleich–, Exponentialverteilung oder diskret). Wir kommen später beim Zentralen Grenzwertsatz noch einmal darauf zurück.

Bsp. 72 (Box–Müller–Transformation) *Es seien U_1 und U_2 zwei unabhängige, über dem Intervall $[0,1[$ gleichverteilte Zufallsgrößen ($U_i \sim R(0,1)$, $i = 1,2$), $\mathbf{U} = (U_1, U_2)^T$ ein zufälliger Vektor. Wir betrachten den zufälligen Vektor $\mathbf{V} = g(\mathbf{U}) = (X, Y)^T$, wobei:*

$$\begin{aligned} X = g_1(U_1, U_2) &= \sqrt{-2\ln U_1} \cdot \cos 2\pi U_2 \\ Y = g_2(U_1, U_2) &= \sqrt{-2\ln U_1} \cdot \sin 2\pi U_2 \end{aligned}$$

Wir suchen die Dichtefunktionen für die zufälligen Variablen X und Y. Wir bestimmen zunächst die Umkehrfunktion zur Abbildung g. Es gilt:

$$\mathbf{U} = g^{-1}(\mathbf{V}) = (\psi_1(X,Y), \psi_2(X,Y)).$$

Wir ermitteln die Teilfunktionen ψ_1 und ψ_2. Dazu berechnen wir

$$\begin{aligned} X^2+Y^2 &= (-2\ln U_1 \cdot \cos^2(2\pi U_2)) + (-2\ln U_1 \cdot \sin^2(2\pi U_2)) \\ &= (-2\ln U_1) \cdot (\cos^2(2\pi U_2) + \sin^2(2\pi U_2)) \\ &= -2\ln U_1 \end{aligned}$$

Durch Umstellen erhalten wir:

$$U_1 = \psi_1(X,Y) = e^{-\frac{1}{2}(X^2+Y^2)}.$$

Weiterhin gilt:

$$\frac{Y}{X} = \tan 2\pi U_2.$$

Daraus folgt:

$$U_2 = \psi_2(X,Y) = \frac{1}{2\pi}\arctan\left(\frac{Y}{X}\right).$$

Bestimmung von $|J|$. Wir wissen bisher:

$$\begin{aligned}
\psi_1(x,y) &= e^{-\frac{1}{2}(x^2+y^2)} = u_1 \\
\psi_2(x,y) &= \frac{1}{2\pi}\arctan\left(\frac{y}{x}\right) = u_2
\end{aligned}$$

Dann gilt für

$$|J| = \left\|\begin{matrix} \frac{\partial\psi_1}{\partial x} & \frac{\partial\psi_1}{\partial y} \\ \frac{\partial\psi_2}{\partial x} & \frac{\partial\psi_2}{\partial y} \end{matrix}\right\|$$

$$
\begin{aligned}
|J| &= \left\| \begin{matrix} -x \cdot \exp(-\frac{1}{2}(x^2+y^2)) & -y \cdot \exp(-\frac{1}{2}(x^2+y^2)) \\ \frac{1}{2\pi} \cdot \frac{-y}{\left(1+\frac{y^2}{x^2}\right)\cdot x^2} & \frac{1}{2\pi} \cdot \frac{1}{\left(1+\frac{y^2}{x^2}\right)\cdot x} \end{matrix} \right\| \\
&= \left| -\frac{1}{2\pi} \exp\left(-\frac{1}{2}(x^2+y^2)\right) \cdot \left(\frac{x^2}{x^2+y^2} + \frac{y^2}{x^2+y^2}\right) \right| \\
&= \frac{1}{2\pi} e^{-\frac{1}{2}(x^2+x^2)}
\end{aligned}
$$

Für die Dichtefunktion des zufälligen Vektors $\mathbf{V}$ gilt nach der Transformationsformel:

$$f_{\mathbf{V}}(x,y) = f_{\mathbf{U}}(\psi_1(x,y), \psi_2(x,y)) \cdot |J|.$$

Da die Zufallsgrößen U_1 und U_2 unabhängig sind, gilt:

$$f_{\mathbf{V}}(x,y) = f_{U_1}(\psi_1(x,y)) \cdot f_{U_2}(\psi_2(x,y)) \cdot |J|.$$

Nun sind $U_1, U_2 \sim R(0,1)$. Daraus folgt aber:

$$\begin{aligned} f_{\mathbf{V}}(x,y) &= |J| = \frac{1}{2\pi} e^{-\frac{1}{2}(x^2+y^2)} = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \cdot \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2} \\ &= f_X(x) \cdot f_Y(y). \end{aligned}$$

mit

$$\begin{aligned} f_X(x) &= \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2} \\ f_Y(y) &= \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}y^2} \end{aligned}$$

Das bedeutet, daß die Zufallsgrößen X und Y unabhängig und standardnormalverteilt sind:

$$X \sim \mathcal{N}(0,1), \quad Y \sim \mathcal{N}(0,1).$$

Bsp. 73 (Treffen einer Zielscheibe*) *Es seien folgende Bedingungen erfüllt*

- *V1: Die Randverteilungen von X und Y seien stetig*
- *V2: Die Dichte $h(x,y)$ von (X,Y) hängt nur vom Abstand $\sqrt{x^2+y^2}$ vom Nullpunkt ab (Radialsymmetrie)*
- *V3: Die Fehler in x- und y-Richtung sind unabhängig.*

Sei Z die zufällige Abweichung in beliebiger Richtung. Dann ist

$$Z \sim \mathcal{N}(0,\sigma^2).$$

Beweis: Seien $p(x)$ und $q(y)$ Randdichten von (X,Y). Aus

V2 und V3 folgt

$$p(x)q(y) = s(r), \quad r^2 = x^2 + y^2 \qquad (3)$$

Substitutionsmethode:

$x = 0$: $p(0)q(y) = s(y)$, $\quad p(0) \neq 0$

$y = 0$: $q(0)p(x) = s(x)$, $\quad q(0) \neq 0$ $x = y$: $p(x) = q(x)$ und damit $p(0)p(y) = s(y)$

Teilen obige Funktionalgleichung durch $p(0)^2$,

$$\frac{p(x)}{p(0)}\frac{p(y)}{p(0)} = \frac{s(r)}{p(0)^2} = \frac{p(r)}{p(0)}$$

Logarithmieren

$$\ln(\frac{p(x)}{p(0)}) + \ln(\frac{p(y)}{p(0)}) = \ln(\frac{p(r)}{p(0)})$$

Mit $f(x) := \ln(\frac{p(x)}{p(0)})$:

$$f(x) + f(y) = f(r), \quad r^2 = x^2 + y^2$$

$y = 0, x = -x_1$: $f(-x) = f(|x|)$ wegen $f(0) = 0$.

$x^2 = x_1^2 + x_2^2$:

$$f(r) = f(y) + f(x_1) + f(x_2), \quad r^2 = y^2 + x_1^2 + x_2^2$$

Wiederholtes Einsetzen:

$$f(r) = f(x_1) + f(x_2) + \ldots + f(x_k), \quad r^2 = \sum_{i=1}^{k} x_i^2$$

$k = n^2, x = x_1 = \ldots = x_k$:

$$f(nx) = n^2 f(x) \Rightarrow_{x=1} f(n) = n^2 f(1)$$

$x = \frac{m}{n}, m \in \mathbb{Z}$:

$$n^2 f(\frac{m}{n}) = f(n\frac{m}{n}) = f(m) = m^2 f(1)$$

$$\Rightarrow f(\tfrac{m}{n}) = f(1)\Big(\tfrac{m}{n}\Big)^2 \Rightarrow$$

$$f(x) = cx^2, \quad c = f(1)$$

für alle rationalen x. Wegen der Stetigkeit (V1) folgt diese Relation für alle $x \in \mathbb{R}$.

$$p(x) = p(0)e^{cx^2}$$

$p(x) > 0$ da Wkt.dichte, $c < 0$, $c := -\frac{1}{2\sigma^2}$.

$$1 = \int_{-\infty}^{\infty} p(x)\, dx = p(0) \int_{-\infty}^{\infty} e^{cx^2}\, dx = p(0)\sigma\sqrt{2\pi}$$

$$p(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{x^2}{2\sigma^2}}$$

Gemeinsame Dichte von (X, Y):

$$p(x)p(y) = \frac{1}{\sigma^2 2\pi} e^{-\frac{x^2+y^2}{2\sigma^2}}.$$

Fehler in einer beliebigen Richtung, θ, $0 \leq \theta \leq 2\pi$:

$Z = X\cos(\theta) + Y\sin(\theta)$

Variablentransformation

$$\begin{aligned} z &= x\cos(\theta) + y\sin(\theta) \\ u &= x\cos(\theta) - y\sin(\theta) \end{aligned}$$

Jacobi-Determinante J: $J = \left\| \begin{matrix} 1 & 0 \\ 0 & -1 \end{matrix} \right\| = |-1| = 1$

Quadrieren liefert

$$\begin{aligned} z^2 &= x^2\cos^2(\theta) + y^2\sin^2(\theta) + 2xy\cos(\theta)\sin(\theta) \\ u^2 &= x^2\cos^2(\theta) + y^2\sin^2(\theta) - 2xy\cos(\theta)\sin(\theta) \end{aligned}$$

Addition: $x^2 + y^2 = z^2 + u^2$ also gemeinsame Dichte von

(Z,U):

$$h_1(z,u) = \frac{1}{\sigma^2 2\pi} e^{-\frac{z^2+u^2}{2\sigma^2}} = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{z^2}{2\sigma^2}} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{u^2}{2\sigma^2}}$$

d.h. Z und U sind unabhängig, $h_1(z,u) = h_Z(z)h_U(u)$ und

$$h_Z(z) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{z^2}{2\sigma^2}}$$

$\square$

Satz 32 (Transformationssatz für EW)

Es sei $\mathbf{X} = (X_1, \ldots, X_p)^T$ *ein zufälliger Vektor und* $g\colon \mathbb{R}^p \longrightarrow \mathbb{R}$ *eine Abbildung.*

a) Es sei $\mathbf{X}$ *diskret mit Wkt.funktion (Zähldichte)* f. *Falls*

$$\sum_{\mathbf{x}} |g(\mathbf{x})| f(\mathbf{x}) < \infty \qquad \text{so gilt:} \qquad \mathbf{E}(g(\mathbf{X})) = \sum_{\mathbf{x}} g(\mathbf{x}) f(\mathbf{x}).$$

b) Es sei $\mathbf{X}$ *stetig.* $f\colon \mathbb{R}^p \longrightarrow \mathbb{R}$ *sei die Dichtefunktion des zufälligen Vektors* $\mathbf{X}$. *Falls*

$$\int_{\mathbb{R}^p} |g(x_1, \ldots, x_p)| \cdot f(x_1, \ldots, x_p)\, dx_1 \ldots dx_p < \infty$$

gilt, so:

$$\mathbf{E}g(X) = \int\limits_{\mathbb{R}^p} g(x_1, \ldots, x_p) \cdot f(x_1, \ldots, x_p)\, dx_1 \ldots dx_p.$$

Bem.: Der Beweis des Satzes erfordert Maßtheorie. Wir haben eine eingeschränkte Version bereits im Abschnitt Erwartungswerte gezeigt.

Bsp. 74 *Es sei* $\mathbf{X} = (X_1, X_2)^T$ *ein stetiger zufälliger Vektor mit Dichtefunktion* $f(X_1, x_2)$*. Wir definieren die Funktion* $g\colon \mathbb{R}^2 \longrightarrow \mathbb{R}$ *durch* $g(\mathbf{X}) := X_1 + X_2$*. Dann gilt:*

$$
\begin{aligned}
\mathbf{E}g(\mathbf{X}) &= \mathbf{E}(X_1 + X_2) \\
&= \int\limits_{\mathbb{R}^2} g(x_1, x_2) \cdot f(x_1, x_2)\, dx_1\, dx_2 \\
&= \int\limits_{-\infty}^{+\infty} \int\limits_{-\infty}^{+\infty} (x_1 + x_2) \cdot f(x_1, x_2)\, dx_1\, dx_2 \\
&= \int_{\mathbb{R}^2} x_1 f(x_1, x_2)\, dx_1\, dx_2 + \int_{\mathbb{R}^2} x_2 f(x_1, x_2)\, dx_1\, dx_2 \\
&= \int\limits_{-\infty}^{+\infty} x_1 \left(\int\limits_{-\infty}^{+\infty} f(x_1, x_2)\, dx_2 \right) dx_1 + \\
&\quad \int\limits_{-\infty}^{+\infty} x_2 \left(\int\limits_{-\infty}^{+\infty} f(x_1, x_2)\, dx_1 \right) dx_2
\end{aligned}
$$

$$\begin{aligned}\mathbf{E}g(X) &= \int\limits_{-\infty}^{+\infty} x_1 \cdot f_{X_1}(x_1)\,dx_1 + \int\limits_{-\infty}^{+\infty} x_2 \cdot f_{X_2}(x_2)\,dx_2 \\ &= \mathbf{E}X_1 + \mathbf{E}X_2\end{aligned}$$

Allgemeiner erhalten wir also für zwei stetige zufällige Variablen X_1 und X_2:

$$\mathbf{E}(c \cdot X_1 + d \cdot X_2) = c \cdot \mathbf{E}X_1 + d \cdot \mathbf{E}X_2.$$

Das ist ein Beweis der Aussage 5 in Satz 11 für stetige zufällige Variablen.

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(Raucher)*	*1(Nichtraucher)*	*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 $t, u \in \mathbb{R}$:

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

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:

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

Beides zusammen ergibt

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

Wir stellen etwas um und erhalten:

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

$\square$

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:

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

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

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

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

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: $\varrho(X_1, X_2) = \mathbf{E}\left(X_1^* \cdot X_2^*\right)$. 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

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

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

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

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

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.