1 Multivariate Verteilungen

Ein p-dimensionaler Zufallsvektor ist ein Spaltenvektor $X = \begin{pmatrix} X_1 \\ \vdots \\ X_p \end{pmatrix}$.

Beispiel 1 *Sei $\Omega = \{a, b\}$, P das Wahrscheinlichkeitsmaß auf Ω mit $P(\{a\}) = P(\{b\}) = \frac{1}{2}$, $X_1, X_2 : \Omega \to \mathbb{R}$ seien Zufallsvariablen mit:*

$$i = 1, 2 : P(X_i = 0) = P(X_i = 1) = \frac{1}{2}$$

Dann ist $X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$ ein zweidimensionaler Zufallsvektor.

Lässt sich aus den gegebenen Information $P(X_1 + X_2 = 1)$ berechnen?

Nein, denn die eindimensionale Beschreibung (1) enthält nicht genügend Information über die Verteilung von X.

1. Falls $X_1(a) = X_2(a) = 0$ und $X_1(b) = X_2(b) = 1$, dann gilt (1) und $P(X_1 + X_2 = 1) = P(\emptyset) = 0$.
2. Falls $X_1(a) = X_2(b) = 1$ und $X_1(b) = X_2(a) = 1$, dann gilt (1) und $P(X_1 + X_2 = 1) = P(\Omega) = 1$.

Die Verteilung eines Zufallsvektors $X = \begin{pmatrix} X_1 \\ \vdots \\ X_p \end{pmatrix}$ lässt sich durch seine Verteilungsfunktion $F = F_X : \mathbb{R}^p \to \mathbb{R}$ beschreiben, wobei:

$$\begin{pmatrix} x_1 \\ \vdots \\ x_p \end{pmatrix} \in \mathbb{R}^p : F(x_1, \ldots, x_p) = P(X_1 \leqslant x_1, \ldots, X_p \leqslant x_p)$$

Beispiel 1.1

$$F(x_1, x_2) = P(X_1 \leqslant x_1, X_2 \leqslant x_2) X_1 \equiv X_2 : F(x_1, x_2) = P(X_1 \leqslant x_1, X_1 \leqslant x_2) = P(X_1 \leqslant \min(x_1, x_2))$$

$$= \begin{cases} 0 & \min(x_1, x_2) < 0 \\ P(X_1 = 0) = \frac{1}{2} & 0 \leqslant \min(x_1, x_2) < 1 \\ 1 & \min(x_1, x_2) \geqslant 1 \end{cases}$$

Ist $X = \begin{pmatrix} X_1 \\ \vdots \\ X_p \end{pmatrix}$ ein diskreter Zufallsvektor, d. h. nimmt X nur endlich oder abzählbar unendlich viele Werte an, dann lässt sich seine Verteilung durch seine Wahrscheinlichkeitsfunktion $f : \mathbb{R}^p \to \mathbb{R}$ beschreiben, wobei $f(x) = P(X = x), x \in \mathbb{R}^p$. Es gilt:

$$P(X \in D) = \sum_{j : x^{(j)} \in D} f(x^{(j)}), D \subset \mathbb{R}^p$$

Beispiel 1.2

$$x^{(1)} = \begin{pmatrix} 0 \\ 1 \end{pmatrix}, x^{(2)} = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$$

$$f(x^{(1)}) = P(X = x^{(1)}) = P(X_1 = 0, X_2 = 1) = P(\{a\}) = P(\{b\}) = f(x^{(2)}) = \frac{1}{2}$$

$$f(x) = 0, x \in \mathbb{R} \setminus \{x^{(1)}, x^{(2)}\}$$

Mit $D = \left\{ \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \in \mathbb{R}^2 : x_1 + x_2 = 1 \right\}$ gilt:

$$P(X_1 + X_2 = 1) = P(X \in D) = \sum_{j:x^{(j)} \in D} f(x^{(j)}) = \frac{1}{2} + \frac{1}{2} = 1$$

Ein Zufallsvektor $X = \begin{pmatrix} X_1 \\ \vdots \\ X_p \end{pmatrix}$ heißt stetig verteilt, wenn es eine Dichte $f : \mathbb{R}^p \to [0, \infty)$ gibt, sodass:

$$\forall \begin{pmatrix} x_1 \\ \vdots \\ x_p \end{pmatrix} \in \mathbb{R}^p : P(X_1 \leqslant x_1, \ldots, X_p \leqslant x_p) = \int_{-\infty}^{x_p} \ldots \int_{-\infty}^{x_1} f(u_1, \ldots, u_p) \, \mathrm{d}u_1 \ldots \mathrm{d}u_p$$

1.1 Mehrdimensionale Integrale

Zur Berechnung des zweidimensionalen Integrals $\int_c^d \int_a^b g(x, y) \, \mathrm{d}x \, \mathrm{d}y$ berechne zunächst für jedes festgehaltene $y \in [c, d]$ das innere Integral $G(y) = \int_a^b g(x, y) \, \mathrm{d}x$.

Beispiel 2 *Berechne*

$$\int_0^2 \int_0^1 xy \ \mathrm{d}x \, \mathrm{d}y$$

$$g(x, y) = xy$$

$$G(y) = \int_0^1 xy \ \mathrm{d}x = \left[\frac{x^2 y}{2} \right]_{x=0}^1 = \frac{y}{2} - \frac{0}{2} = \frac{y}{2}$$

$$\int_0^2 \int_0^1 g(x, y) \ \mathrm{d}x \, \mathrm{d}y = \int_0^2 G(y) dy = \int_0^2 \frac{y}{2} \, \mathrm{d}y = \left[\frac{y^2}{4} \right]_{y=0}^2 = 1$$

Für $D \subset \mathbb{R}^2$ und $g : \mathbb{R}^2 \to \mathbb{R}$ ist $\int\limits_D g(x)\,\mathrm{d}x = \int\limits_D g(x_1, x_2)\ \mathrm{d}x_1\,\mathrm{d}x_2$ definiert durch:

$$\int\limits_D g(x)\ \mathrm{d}x = \int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty} \mathbb{1}_D(x_1, x_2) \cdot g(x_1, x_2)\,\mathrm{d}x_1\,\mathrm{d}x_2, \mathbb{1}_D(x_1, x_2)$$

$$= \begin{cases} 1 & (x_1, x_2) \in D \\ 0 & (x_1, x_2) \notin D \end{cases}$$

Im Fall $D = \mathbb{R}^2$ schreibt man auch

$$\int g(x)\,\mathrm{d}x = \int g(x_1, x_2)\,\mathrm{d}x_1\,\mathrm{d}x_2 = \int\limits_{\mathbb{R}^2} g(x)\,\mathrm{d}x$$

$$= \int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty} g(x_1, x_2)\,\mathrm{d}x_1\,\mathrm{d}x_2$$

.

Ist D ein zweidimensionales Intervall der Form $D = (a_1, b_1) \times (a_2, b_2)$, so ist:

$$\int\limits_{(a_1,b_1)\times(a_2,b_2)} g(x)\,\mathrm{d}x = \int\limits_{-\infty}^{\infty}\int\limits_{-\infty}^{\infty} 1_D(x_1, x_2) \cdot g(x_1, x_2)\ \ \mathrm{d}x_1\,\mathrm{d}x_2$$

$$= \int\limits_{a_2}^{b_2}\int\limits_{a_1}^{b_1} \mathbb{1}_D(x_1, x_2) \cdot g(x_1, x_2)\,\mathrm{d}x_1\,\mathrm{d}x_2 = \int\limits_{a_2}^{b_2}\int\limits_{a_1}^{b_1} g(x_1, x_2)\,\mathrm{d}x_1\,\mathrm{d}x_2$$

1.2 Rechenregeln

- Linearität

  $$\forall\ \alpha, \beta \in \mathbb{R},\ f, g : \mathbb{R}^2 \to \mathbb{R},\ D \subset \mathbb{R}^2 : \int\limits_D (\alpha \cdot f(x) + \beta \cdot g(x)\,\mathrm{d}x) = \alpha \int\limits_D f(x)\,\mathrm{d}x + \beta \int\limits_D g(x)\,\mathrm{d}x$$

- Additivität bezüglich des Integrationsbereichs
  Für disjunkte Mengen $D_1, D_2 \subset \mathbb{R}^2, D_1 \cap D_2 = \emptyset$ gilt:

  $$\int\limits_{D_1 \cup D_2} g(x)\,\mathrm{d}x = \int\limits_{D_1} g(x)\,\mathrm{d}x + \int\limits_{D_2} g(x)\,\mathrm{d}x$$

1.3 Monotonie

$$\int_D f(x)\mathrm{d}x \leqslant \int_D g(x)\,\mathrm{d}x \quad \text{falls} \quad f(x) \leqslant g(x),\ \forall x \in D$$

$$\int_{D_1} g(x)\,\mathrm{d}x \leqslant \int_{D_2} g(x)\,\mathrm{d}x \quad \text{falls} \quad D_1 \subset D_2 \subset \mathbb{R}^2 \quad \text{und} \quad g(x) \geqslant 0,\ \forall x \in D_2$$

1.4 Vertauschbarkeit der Integrationsreihenfolge

$$\int_{a_2}^{b_2}\int_{a_1}^{b_1} g(x_1,x_2)\,\mathrm{d}x_1\,\mathrm{d}x_2 = \int_{a_1}^{b_1}\int_{a_2}^{b_2} g(x_1,x_2)\,\mathrm{d}x_2\,\mathrm{d}x_1,\ -\infty \leqslant a_1 \leqslant b_1 \leqslant \infty, -\infty \leqslant a_2 \leqslant b_2 \leqslant \infty$$

Für Integranden der Form $g(x_1,x_2) = h_1(x_1,x_2)h_2(x_2)$ gilt:

$$\int_{a_2}^{b_2}\int_{a_1}^{b_1} h_1(x_1,x_2)h_2(x_2)\,\mathrm{d}x_1\,\mathrm{d}x_2 = \int_{a_2}^{b_2} h_2(x_2)\int_{a_1}^{b_1} h_1(x_1,x_2)\,\mathrm{d}x_1\,\mathrm{d}x_2$$

Für Integranden der Form $h_1(x_1)h_2(x_2)$ gilt:

$$\int_{a_2}^{b_2}\int_{a_1}^{b_1} h_1(x_1)h_2(x_2)\,\mathrm{d}x_1\,\mathrm{d}x_2 = \int_{a_2}^{b_2} h_2(x_2)\int_{a_1}^{b_1} h_1(x_1)\,\mathrm{d}x_1\,\mathrm{d}x_2 = \int_{a_1}^{b_1} h_1(x_1)\,\mathrm{d}x_1\int_{a_2}^{b_2} h_2(x_2)\,\mathrm{d}x_2$$

Analoge Regeln und Bezeichnungen gelten für höherdimensionale Integrale. Zum Beispiel gilt für $g : \mathbb{R}^3 \to \mathbb{R}$ und $a_i \leqslant b_i, i = 1,2,3,\ldots$

$$\int_{a_1}^{b_1}\int_{a_2}^{b_2}\int_{a_3}^{b_3} g(x_1,x_2,x_3)\,\mathrm{d}x_1\,\mathrm{d}x_2\,\mathrm{d}x_3 = \int_{a_2}^{b_2}\int_{a_3}^{b_3}\int_{a_1}^{b_1} g(x_1,x_2,x_3)\,\mathrm{d}x_2\,\mathrm{d}x_3\,\mathrm{d}x_1$$

Die Integration kann in beliebiger Reihenfolge durchgeführt werden.

Für $h_1,\ldots,h_p : \mathbb{R}^1 \to \mathbb{R}^1$ gilt:

$$\int_{a_p}^{b_p}\cdots\int_{a_1}^{b_1} h_1(x_1)\cdots h_p(x_p)\,\mathrm{d}x_1\cdots\mathrm{d}x_p = \int_{a_1}^{b_1} h_1(x_1)\,\mathrm{d}x_1\cdots\int_{a_p}^{b_p} h_p(x_p)\,\mathrm{d}x_p$$

In Beispiel 2:

$$\int_0^2 \int_0^1 xy \, \mathrm{d}x \, \mathrm{d}y = \int_0^1 x \, \mathrm{d}x \cdot \int_0^2 y \, \mathrm{d}y = \frac{x^2}{2}\bigg|_0^1 \frac{y^2}{2}\bigg|_0^2 = \frac{1}{2} \cdot 2 = 1$$

Ist $X = (X_1, \ldots, X_p)^T$ ein Zufallsvektor mit Dichte $f : \mathbb{R}^p \to \mathbb{R}$, dann gilt:

$$P(X \in D) = \int_D f(x) \, \mathrm{d}x \quad \text{für} \quad D \subset R^p \quad \text{und} \quad \int f(x) \, \mathrm{d}x = 1$$

Beispiel 3 *: Sei $X = (X_1, X_2)^T$ Zufallsvektor mit Dichte*

$$f(x_1, x_2) = \begin{cases} 1 & \textit{falls} \quad 0 \leqslant x_1 \leqslant 1, \, 0 \leqslant x_2 \leqslant 1 \\ 0 & \textit{sonst} \end{cases}$$

also Gleichverteilung auf $[0,1]^2$*.*

Berechne $P(X_1 \leqslant X_2)$.

Setze

$$D := \{(X_1, X_2)^T \in R^2 : x_1 \leqslant x_2\}$$

Daher gilt

$$P(X_1 \leqslant X_2) = P(X \in D) = \int_D f(x) \, \mathrm{d}x = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mathbb{1}_D(x_1, x_2) f(x_1, x_2) \, \mathrm{d}x_1 \, \mathrm{d}x_2$$

($\mathbb{1}$ bezeichnet die Indikatorfunktion).

$$\int_0^1 \int_0^1 \mathbb{1}_D(x_1, x_2) f(x_1, x_2) \, \mathrm{d}x_1 \, \mathrm{d}x_2 = \int_0^1 x_2 dx_2 = \frac{x_2^2}{2}\bigg|_0^1 = \frac{1}{2}$$

Die Indikatorfunktion kann nur die Werte 1 oder 0 annehmen und ist gleich 1, falls $0 \leqslant x_1 \leqslant x_2 \leqslant 1$ und gleich 0 sonst.

Sei X ein p-dimensionaler Zufallsvektor mit Dichte $f : \mathbb{R}^p \to \mathbb{R}$. Betrachte den partitionierten Vektor $X = (x_1, x_2)^T$ wobei X_1 k-dimensional und X_2 $p-k$-dimensionale Zufallsvektoren sind, $1 \leqslant k \leqslant p-1$.

X_1 hat die sogenannte **Randdichte**

$$f_{X_1}(x_1) = \int f(x_1, x_2) \, \mathrm{d}x_2, \; x_1 \in R^k.$$

X p-dimensional. Dichte f.

$$X = (X_1, X_2)^T,$$

X_1 k-dimensional, X_2 p-k-dimensional. X_1 hat die Randdichte:

$$f_{X_1}(x_1) = \int f(x_1, x_y)\,\mathrm{d}x_2,\ x_1 \in \mathbb{R}^k.$$

X_2 hat die Randdichte

$$f_{X_2}(x_2) = \int f(x_1, x_2)\,\mathrm{d}x_1,\ x_2 \in \mathbb{R}^{p-k}.$$

f heisst dann auch *gemeinsame Dichte*. Die Verteilungen von X_1 und X_2 heißen *Randverteilungen* (*Marginalverteilungen*).

Beweis *f_{X_1} ist eine Dichte von X_1, für den Fall $p = 2$, $k = 1$. Für jedes $x_1 \in \mathbb{R}$ gilt:*

$$P(X_1 \leqslant x_1) = P(X_1 \leqslant x_1,\ X_2 \in \mathbb{R}) = P(X \in (-\infty, x_1] \times \mathbb{R}) = \int\limits_{(-\infty,x_1]\times\mathbb{R}} f(u)\,\mathrm{d}u$$

$$= \int\limits_{-\infty}^{x_1}\int\limits_{-\infty}^{\infty} f(u_1, u_2)\,\mathrm{d}u_2\,\mathrm{d}u_1 = \int\limits_{-\infty}^{x_1} f_{X_1}(u_1)\,\mathrm{d}u_1$$

Beispiel 4 *$X = (X_1, X_2)^T$ habe die Dichte f aus Beispiel 3. $Y = (Y_1, Y_2)^T$ habe die Dichte* $g(y_1, y_2) = \begin{cases} 1 + (2y_1 - 1)(2y_2 - 1),\ (y_1, y_2) \in v \in [0,1]^2 \\ 0 \quad \textit{sonst} \end{cases}.$

Bestimme Randdichten f_{X_i}, g_{y_i}

$$f_{X_1}(x_1) = \int\limits_{-\infty}^{\infty} f(x_1, x_2)\,\mathrm{d}x_2 = \begin{cases} 1,\ x_1 \in [0,1] \\ 0 \quad \text{sonst.} \end{cases}$$

$$f_{X_2}(x_2) = \begin{cases} 1,\ x_1 \in [0,1] \\ 0 \quad \text{sonst.} \end{cases}$$

$$g_{y_1}(y_1) = \int\limits_{-\infty}^{\infty} g(y_1, y_2)\,\mathrm{d}y_2$$

$$= \begin{cases} \int_0^1 1 + (2y_1 - 1)(2y_2 - 1)\,\mathrm{d}y_2 = 1 + (2y_1 - 1)[y_2^2 - y_2]_0^1 = 1,\ y_1 \in [0,1] \\ 0 \quad \text{sonst} \end{cases}$$

analog folgt

$$g_{y_2}(y_2) = \begin{cases} 1, & y_2 \in [0,1] \\ 0 & \text{sonst.} \end{cases}$$

$\rightarrow X_1, X_2, Y_1, Y_2$ haben dieselben Randverteilungen, aber X und Y haben verschiedene Verteilungen.

Sei wieder $X = (X_1, X_2)^T$ p-dimensionaler Zufallsvektor mit Dichte $f : \mathbb{R}^p \rightarrow \mathbb{R}$. X_1 k-dimensionaler und X_2 $p-k$-dimensionaler Zufallsvektor. Beschreibe die Verteilung von X_1 unter der Annahme, dass X_2 den festgehaltenen Wert x_2 annimmt - X bezeichnet hier eine Zufallsvariable, x ist deterministisch.

Für $x_2 \in \mathbb{R}^{p-k}$ ist die *bedingte Dichte* von X_1 gegeben $X_2 = x_2$ definiert durch

$$f_{X_1}(x_1 \,|\, X_2 = x_2) = \frac{f(x_1, x_2)}{f_{X_2}(x_2)}, \; x_1 \in \mathbb{R}^k, \quad \text{falls} \quad f_{X_2}(x_2) > 0.$$

Die bedingte Dichte von X_2 gegeben $X_1 = x_1$ ist

$$f_{X_2}(x_2 \,|\, X_1 = x_1) = \frac{f(x_1, x_2)}{f_{X_1}(x_1)}, \; x_2 \in \mathbb{R}^p - k, \quad \text{falls} \quad f_{X_1}(x_1) > 0.$$

X_1 und X_2 heißen *(stochastisch) unabhängig*, falls

$$f(x_1, x_2) = f_{X_1}(x_1) f_{X_2}(x_2), \; \forall x_1 \in \mathbb{R}^k, \; x_2 \in \mathbb{R}^{p-k}.$$

Falls X_1 und X_2 unabhängig sind, gilt für alle möglichen Werte die hier miteinander verglichen werden können, also für alle $x_1 \in \mathbb{R}^k$ mit der Eigenschaft, dass die bedingte Dichte definiert ist, also $f_{X_1}(x_1) > 0$:

$$f_{X_2}(x_2 \,|\, X_1 = x_1) = \frac{f(x_1, x_2)}{f_{X_1}(x_1)} = \frac{f_{X_1}(x_1) f_{X_2}(x_2)}{f_{X_1}(x_1)} = f_{X_2}(x_2)$$

Das heisst die bedingte Dichte von X_2 gegeben $X_1 = x_1$ hängt nicht von x_1 ab und ist gleich der Randdichte von X_2.

Sind $X_1, \ldots, X_p$ Zufallsvariablen und f eine Dichte von $(X_1, \ldots, X_p)^T$, dann sind $(X_1, \ldots, X_p)$ unabhängig, falls

$$f(x_1, \ldots, x_p) = f_{X_1}(x_1) \cdots f_{X_p}(x_p), \; \forall x_1, \ldots, x_p \in \mathbb{R}.$$

Beispiel 5 *$X = (X_1, X_2)^T$ sei Zufallsvektor mit Dichte*

$$f(x_1, x_2) = \begin{cases} x_1 + x_2, & \text{für} \quad 0 \leqslant x_1, x_2 \leqslant 1 \\ 0, & \text{sonst} \end{cases} .$$

Randdichte:

$$f_{X_1}(x_1) = \begin{cases} \int\limits_0^1 x_1 + x_2 \, \mathrm{d}x_2 = x_1 + [\frac{x_2^2}{2}]_0^1 = x_1 + \frac{1}{2}, \, x_1 \in [0,1] \\ 0, \, , \quad \text{sonst.} \end{cases}$$

$$f_{X_2}(x_2) = \begin{cases} x_2 + \frac{1}{2}, \, x_2 \in [0,1] \\ 0, \, , \quad \text{sonst.} \end{cases}$$

Es gilt also nicht $f(x_1, x_2) = f_{X_1}(x_1) f_{X_2}(x_2) \forall x_1, x_2 \in \mathbb{R}$. das heisst X_1 und X_2 sind nicht unabängig.

Bedingte Dichten: Für $x_2 \in [0,1]$ ist

$$f_{X_1}(x_1 \,|\, X_2 = x_2) = \frac{f(x_1, x_2)}{f_{X_2}(x_2)} = \begin{cases} \frac{x_1 + x_2}{0.5 + x_2}, \, x_1 \in [0,1] \\ 0 \quad \text{sonst.} \end{cases}$$

Für $x_1 \in [0,1]$ ist

$$f_{X_2}(x_2 \,|\, X_1 = x_1) = \begin{cases} \frac{x_1 + x_2}{x_1 + \frac{1}{2}}, \, x_2 \in [0,1] \\ 0, \quad \text{sonst.} \end{cases}$$

Für $0 \leqslant \alpha \leqslant \beta \leqslant 1$ und für $0 < \delta \leqslant \frac{1}{2}$ gilt:

$$P(\alpha \leqslant X_1 \leqslant \beta \,|\, \frac{1}{2} \leqslant X_2 \leqslant \frac{1}{2 + \delta}) = \frac{P(\alpha \leqslant X_1 \leqslant \beta, \, \frac{1}{2} \leqslant X_2 \leqslant \frac{1}{2} + \delta)}{P(0.5 \leqslant X_2 \leqslant 0.5 + \delta)} = \frac{\int\limits_\alpha^\beta \int\limits_{\frac{1}{2}}^{\frac{1}{2}+\delta} f(x_1, x_2) \, \mathrm{d}x_2 \, \mathrm{d}x_1}{\int\limits_{\frac{1}{2}}^{\frac{1}{2}+\delta} f_{X_2}(x_2) \, \mathrm{d}x_2}$$

$$= \frac{\int\limits_\alpha^\beta \int\limits_{\frac{1}{2}}^{\frac{1}{2}+\delta} x_1 + x_2 \, \mathrm{d}x_2 \, \mathrm{d}x_1}{\int\limits_{\frac{1}{2}}^{\frac{1}{2}+\delta} 0.5 + x_2 \, \mathrm{d}x_2} = \frac{\int\limits_\alpha^\beta \delta x_1 + [\frac{x_2^2}{2}]_{x_2=0.5}^{0.5+\delta} \, \mathrm{d}x_1}{\frac{\delta}{2} + [\frac{x_2^2}{2}]_{0.5}^{0.5+\delta}}$$

$$= \frac{\int\limits_\alpha^\beta \delta x_1 + \frac{\delta}{2} + \frac{\delta^2}{2} \, \mathrm{d}x_1}{\delta + \frac{\delta^2}{2}} = \frac{\int\limits_\alpha^\beta x_1 + \frac{1}{2} \, \mathrm{d}x_1 + \frac{\delta}{2}(\beta - \alpha)}{1 + \frac{\delta}{2}}$$

$$\to \lim_{\delta \to 0^+} P(\alpha \leqslant X_1 \leqslant \beta \,|\, 0.5 \leqslant X_2 \leqslant 0.5 + \delta) = \int\limits_\alpha^\beta x_1 + 0.5 \, \mathrm{d}x_1 = \int\limits_\alpha^\beta f_{X_1}(x_1 | X_2 = 0.5) \, \mathrm{d}x_1 \quad \blacksquare$$

Wichtige Parameter für univariate Verteilungen: Erwartungswert und Varianz. Entsprechende Parameter für multivariate Verteilungen ist der Vektor der Erwartungswerte und Kovarianzmatrix. Sei $X = (X_1, \dots X_p)^T$ p-dimensionaler Zufallsvektor mit Dichte f.

$$E(X_1) = \int_{-\infty}^{\infty} x_1 f_{X_1}(x_1)\,\mathrm{d}x_1 = \int_{-\infty}^{\infty} x_1 \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(x_1, \dots, x_p)\,\mathrm{d}x_2 \dots \mathrm{d}x_p\,\mathrm{d}x_1 = \int x_1 f(X)\,\mathrm{d}x.$$

Entsprechend ist der Erwartungswert von X_j

$$E(X_j) = \int_{-\infty}^{\infty} x_j f_{X_j}\,\mathrm{d}x_j = \int x_j f(x_j)\,\mathrm{d}x,\ j = 1, \dots, p$$

Erwartungswert(-vektor) ist definiert durch $E(X) = (E(X_1), \dots, E(X_p))^T$.

Für $g \cdot \mathbb{R}^p \to \mathbb{R}$ ist $E[g(X)] = \int g(x) f(x)\,\mathrm{d}x$. Hierbei sind die Komponenten Vektoren, also:

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} g(x_1, \dots, x_p) f(x_1, \dots, x_p)\,\mathrm{d}x_1 \dots \mathrm{d}x_p.$$

Allgemeiner: Für matrixwertige Funktionen $G : \mathbb{R}^p \to \mathbb{R}^{n \times m}$:

$$G(x) = \begin{pmatrix} g_{11}(x) & \dots & g_{1m}(x) \\ \vdots & & \vdots \\ g_{n1}(x) & \dots & g_{nm}(x) \end{pmatrix}$$

ist $E[G(x)]$ komponentenweise definiert.

$$E[G(X)] = \begin{pmatrix} E[g_{11}(X)] & \dots & E[g_{1m}(X)] \\ \vdots & & \vdots \\ E[g_{n1}(X)] & \dots & E[g_{nm}(X)] \end{pmatrix} = (E[g_{ij}(X)])_{i=1,\dots,n,\ j=1,\dots,m}$$

Für matrixwertige Funktion $H : \mathbb{R}^p \to \mathbb{R}^{n \times m}$, $H(x) = (h_{ij}(x))_{i,j}$ ist $\int H(x)\,\mathrm{d}x$ ebenfalls komponentenweise definiert.

$$\int H(x)\,\mathrm{d}x = \begin{pmatrix} \int h_{11}(x)\,\mathrm{d}x] & \dots & \int h_{1m}(x)\,\mathrm{d}x \\ \vdots & & \vdots \\ \int h_{n1}(x)\,\mathrm{d}x & \dots & \int h_{nm}(x)\,\mathrm{d}x \end{pmatrix}$$

Damit ergibt (G matrixwertig) sich

$$E[G(X)] = E[g_{ij}(X)]_{i,j} = \left(\int g_{ij}(x) f(x)\,\mathrm{d}x\right)_{i,j} = \int G(x) f(x)\,\mathrm{d}x$$

und insbesondere

$$E(X) = \int x f(x)\,\mathrm{d}x$$

Beispiel 6 *Berechne $E(X)$ für $X = (X_1, X_2)^T$ aus Beispiel 5.*

$$E(X_1) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x_1 f(x_1, x_2)\, dx_1\, dx_2 = \int_0^1 \int_0^1 x_1(x_1 + x_2)\, dx_1\, dx_2$$

$$= \int_0^1 x_1^2\, dx_1 + x_2 \int_0^1 x_1\, dx_1\, dx_2 = \int_0^1 \frac{1}{3} + x_2 \frac{1}{2}\, dx_2 = \frac{1}{3} + \frac{1}{2}\frac{1}{2} = \frac{7}{12}$$

Daher

$$E(X_2) = \cdots = \frac{7}{12}$$

$$E(X) = (E(X_1), E(X_2))^T = (\frac{7}{12}, \frac{7}{12})^T.$$

1.5 Rechenregeln

Für p-dimensionale Zufallsvektoren X, Y, $\alpha, \beta \in \mathbb{R}$, $g, h : \mathbb{R}^p \to \mathbb{R}^n$ gilt:

$$E[\alpha g(X) + \beta h(Y)] = \alpha E[g(X)] + \beta E[h(Y)]$$

$$E[AX + b] = AE(X) + b, \quad \text{wobei} \quad A \in \mathbb{R}^{n\times p},\, b \in \mathbb{R}^n$$

Ist $X = (X_1, X_2)^T$ mit X_1 k-dimensional und X_2 $p-k$-dimensional und ist $g : \mathbb{R}^k \to \mathbb{R}$, dann

$$E[g(X_1)] = \int g(X_1) f(x)\, dx = \int g(X_1) f_{X_1}(x_1)\, dx_1.$$

Sind X_1 und X_2 unabhängig und $g : \mathbb{R}^k \to \mathbb{R},\ h : \mathbb{R}^{p-k} \to \mathbb{R}$, dann

$$E[g(X_1)h(X_2)] = E[g(X_1)]E[h(X_2)]$$

Nachweis für den Fall $p = 2$, $k = 1$:

$$E[g(X_1)h(X_2)] = \int\int g(x_1)h(x_2) f(x_1, x_2)\, dx_1\, dx_2$$

$$\int\int g(x_1)h(x_2) f_{X_1}(x_1) f_{X_2}(x_2)\, dx_1\, dx_2 = \int g(x_1) f_{X_1}(x_1)\, dx_1 \int h(x_2) f_{X_2}(x_2)\, dx_2$$

$$= E[g(X_1)]E[h(X_2)] \blacksquare$$

Sind X, Y p-dimensionale unabhängige Zufallsvektoren, dann gilt

$$E(XY^T) = E(X)E(Y)^T \quad \text{und} \quad E(X^T Y) = E(X)^T E(Y)$$

denn mit $X = (X_1, \dots, X_p)^T$, $Y = (Y_1, \dots, Y_p)^T$ ist

$$E(XY^T) = \begin{pmatrix} (X_1) \\ \vdots \\ (X_p) \end{pmatrix} \begin{pmatrix} Y_1 & \dots & Y_n \end{pmatrix} = (E(X_j Y_k))_{j,k=1,\dots,p} = (E(X_j)E(Y_k))_{j,k=1,\dots,p}$$

$$\begin{pmatrix} E(X_1) \\ \vdots \\ E(X_p) \end{pmatrix} (E(X_1), \dots, E(Y_p)) = E(X)E(Y)^T$$

$$E(X^T Y) = E(\sum_{j=1}^{p} X_j Y_j) = \sum_{j=1}^{p} E(X_j)E(Y_j) = E(X)^T E(Y) \blacksquare$$

Einschub für Übungsblatt, Kovarianz:

$$\text{Cov}(X_j, X_k) := E[X_j - E(X_j)(X_k - E(X_k))]$$

Sei $X = (X_1, \dots, X_p)^T$ p-dimensionaler Zufallsvektor und $\mu = (\mu_1, \dots, \mu_p) = E(X)$. (Alle ps unten sind eigentlich müs μ)

Varianz von X_j:

$$\sigma_j^2 = Var(X_j) = E[(X_j - \mu_j)^2]$$

ist ein Maß für die Streuung von X_j.

Kovarianz zwischen X_j und X_k

$$\sigma_{jk}^2 = \text{Cov}(X_j, X_k) = \text{E}[(X_j - \mu_j)(X_k - \mu_k)]$$

ist ein Maß für den Zusammenhang zwischen X_j und X_k, $j \neq k$, $(\sigma_{jj} = \sigma_j^2)$.

$$\text{Cov}(X_j, X_k) \geqslant 0 :$$

X_j und X_k weisen tendenziell einen gleichsinnigen linearen Zusammenhang auf.

$$\text{Cov}(X_j, X_k) \leqslant 0 :$$

Sind X_j und X_k unabhängig gilt:

$$\text{Cov}(X_j, X_k) = \text{E}(X_j - \mu_j)\text{E}(X_k - \mu_k) = 0.$$

Aus $\text{Cov}(X_j, X_k) = 0$ folgt im allg. nicht, dass X_j und X_k unabhängig sind, siehe Aufgabe 2, Blatt 1.

Die **Kovarianzmatrix** Σ der Zufallsvariable X ist definiert durch

$$\Sigma = Var(X) = \mathrm{E}[(X-\mu)(X-\mu)^T]$$

$$\Sigma = (\mathrm{E}[(X_j-\mu_j)(X_k-\mu_k)^T])_{j,k=1,\ldots,p}$$

$$= \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \ldots & \sigma_{1p} \\ \sigma_{21} & \sigma_2^2 & \ldots & \sigma_{2p} \\ \vdots & & \ddots & \\ \sigma_{p1} & & \ldots & \sigma_p^2 \end{pmatrix}$$

Allgemeiner: Ist $X = (X_1, \ldots, X_p)^T$ ein p-dimensionaler und $Y = (Y_1, \ldots, Y_q)^T$ ein q-dimensionaler Zufallsvektor dann heißt

$$\text{Cov}(X, Y) = \mathrm{E}[(X - \mathrm{E}(x))(Y - \mathrm{E}(Y))^T]$$

die Kovarianzmatrix von X und Y. In Position (j, k) von $\text{Cov}(X, Y)$ steht

$$\mathrm{E}[(X_j - \mathrm{E}(X_j))(Y_k - \mathrm{E}(Y_k))] = \text{Cov}(X_j, X_k)$$

also

$$\text{Cov}(X, Y) = (\text{Cov}(X_j, Y_k))_{j=1,\ldots,p,\ k=1,\ldots,q}$$

Lemma 7 *Sei X ein p-dimensionaler, Y ein q-dimensionanler Zufallsvektor.*

1. $\text{Cov}(X, X) = \text{Var}(X)$

2. $\text{Cov}(X, Y) = \mathrm{E}(XY^T) - \mathrm{E}(X)[\mathrm{E}(Y)]^T$, $\text{Var}(X) = \mathrm{E}(XX^T) - \mathrm{E}(X)(\mathrm{E}(X))^T$

3. $\text{Var}(a^T X) = a^T \text{Var}(X) a$ für $a \in \mathbb{R}^p$

4. $\text{Var}(AX + b) = A\text{Var}(X)A^T$ für $a \in \mathbb{R}^{n \times p}$, $b \in \mathbb{R}^n$

5. Sei $p = q$. Dann gilt $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) + \text{Cov}(X, Y) + \text{Cov}(Y, X)$. Falls X und Y unabhängig sind, dann $\text{Cov}(X, Y) = \text{Cov}(Y, X) = 0$ und $\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y)$.

6. $\text{Var}(X)$ ist symmetrisch und nicht-negativ definit, d.h. $a^T \text{Var}(X) a \geqslant 0$, $\forall a \in \mathbb{R}^p$

Beweis *Setze* $\mu = \mathrm{E}(X)$, $\upsilon = \mathrm{E}(Y)$

1. $\mathrm{Cov}(X,Y) = \mathrm{E}[(X-\mu)(X-\mu)^T] = \mathrm{Var}(X)$

2. $\mathrm{Cov}(X,Y) = \mathrm{E}[(X-\mu)(Y-\upsilon)^T] = \mathrm{E}(XY^T - \mu Y^T - X\upsilon^T + \mu\upsilon^T) = \mathrm{E}(XY^T) - \mu\mathrm{E}(Y^T) - \mathrm{E}(X)\upsilon^T + \mu\upsilon^T = \mathrm{E}(XY^T - \mathrm{E}(X)(\mathrm{E}(Y))^T$. Daraus folgt mit (a), dass $\mathrm{Var}(X) = \mathrm{Cov}(X,X) = \mathrm{E}(XX^T) - \mathrm{E}(X)(\mathrm{E}(X))^T$

3. Spezialfall von (d) mit $A = a^T,\ b = 0$

4. $\mathrm{Var}(AX+b) = \mathrm{E}[(AX+b-\mathrm{E}(AX+b))(AX+b-\mathrm{E}(AX+b))^T] = A\mathrm{E}[(X-\mathrm{E}(X))(X-\mathrm{E}(X))^T A^T] = A\mathrm{Var}(X)A^T$

5. $\mathrm{Var}(X+Y) = \mathrm{E}[(X+Y)(X+Y)]^T - (\mathrm{E}(X+Y))^T = \mathrm{E}(XX^T) + \mathrm{E}(XY^T) + \mathrm{E}(YX^T) + \mathrm{E}(YY^T) - \mathrm{E}(X)(\mathrm{E}(X))^T - \mathrm{E}(X)(\mathrm{E}(Y))^T - \mathrm{E}(Y)(\mathrm{E}(X))^T - \mathrm{E}(Y)(\mathrm{E}(Y))^T = \mathrm{Var}(X) + \mathrm{Cov}(X,Y) + \mathrm{Cov}(Y,X) + \mathrm{Var}(Y)$. Sind $X = (X_1,\dots,X_p)^T$ und $Y = (Y_1,\dots,Y_p)^T$ unabhängig, dann gilt $\mathrm{E}(X_j,Y_k) = \mathrm{E}(X_j)\mathrm{E}(Y_k)$ also $\mathrm{Cov}(X_j,Y_k) = 0$. Daraus folgt $\mathrm{Cov}(X,Y) = 0,\ \mathrm{Cov}(Y,X) = 0$.

6. $\{\mathrm{Var}(X)\}_{jk} = \mathrm{Cov}(X_j,X_k) = \mathrm{Cov}(X_k,X_j) = \{\mathrm{Var}(X)\}_{kj}$ für alle $j,k = 1,\dots,p$. D.h. $\mathrm{Var}(X)$ ist symmetrisch. Für alle Vektoren $a^T \in \mathbb{R}^p$ gilt $a^T\mathrm{Var}(X)a = \mathrm{Var}(a^TX) \geqslant 0$ (wegen c).

Beispiel 8 *Sei* $X = (X_1,X_2)$ *wie in Bsp. 5. Berechne (i)* $\mathrm{Var}(X)$, *(ii)* $\mathrm{Var}(X_1 - 3X_2)$, *(iii)* $\mathrm{Var}(X_1+X_2+1, 2X_1-X_2-1)^T$.

1. $\mathrm{E}(X_1) = \mathrm{E}(X_2) = \frac{7}{12}$ nach Bsp. 6.

$$\mathrm{E}(X_1^2) = \int x_1^2 f_{X_1}(x_1)\,\mathrm{d}x = \int_0^1 x_1^2(x_1+\frac{1}{2})\,\mathrm{d}x = \left.\frac{x_1^4}{4}\right|_0^1 + \left.\frac{x_1^3}{6}\right|_0^1 = \frac{1}{4} + \frac{1}{6} = \frac{5}{12}$$

$$\Longrightarrow\ \mathrm{Var}(X_1) = \mathrm{E}(X_1^2 - (\mathrm{E}(X_1))^T = \frac{5}{12} - (\frac{5}{12})^T = \frac{11}{144} \quad \text{ebenso} \quad \mathrm{Var}X_2 = \frac{11}{144}.$$

$$\mathrm{E}(X_1X_2) = \iint x_1x_2 f(x_1,x_2)\,\mathrm{d}x_1\,\mathrm{d}x_2 = \int_0^1\int_0^1 x_1x_2 f(x_1,x_2)\,\mathrm{d}x_1\,\mathrm{d}x_2$$

$$= \int_0^1 x_1^2\,\mathrm{d}x \int_0^1 x_2\,\mathrm{d}x_2 + \int_0^1 x_1\,\mathrm{d}x_1 \int_0^1 x_2^2\,\mathrm{d}x_2 = \frac{1}{3}\frac{1}{2} + \frac{1}{2}\frac{1}{3} = \frac{1}{3}$$

$$\mathrm{Cov}(X_1,X_2) = \mathrm{E}(X_1X_2) - \mathrm{E}(X_1)\mathrm{E}(X_2) = \frac{1}{3} - \frac{7}{12}\cdot\frac{7}{12} = -\frac{1}{144} \Longrightarrow \mathrm{Var}(X)$$

$$\begin{pmatrix} \mathrm{Var}(X_1) & \mathrm{Cov}(X_1, X_2) \\ \mathrm{Cov}(X_2, X_1) & \mathrm{Var}(X_2) \end{pmatrix} = \begin{pmatrix} \frac{11}{144} & -\frac{11}{144} \\ -\frac{11}{144} & \frac{11}{144} \end{pmatrix}$$

2. Mit $a = (1, -3)^T$ ist $X_1 - 3X_3 = a^T X \implies \mathrm{Var}(X_1 - 3X_3) = \mathrm{Var}(a^T X) = a^T \mathrm{Var}(X) a$. Einsetzen liefert $\frac{29}{36}$.

3. Mit $A = \begin{pmatrix} 1 & 1 \\ 2 & -1 \end{pmatrix}$ und $b = (1, -1)^T$ ist $AX + b = \begin{pmatrix} X_1 + X_2 + 1 \\ 2X_1 - X_2 - 1 \end{pmatrix} \implies \mathrm{Var}(AX + b) = A\mathrm{Var}(X)A^T$

Übungsblatt 1

Aufgabe 1

a)

$$f_{X_1}(x_1) = \int\limits_0^1 f(x_1, x_2)\,\mathrm{d}x_2 = 4x_1 e^{-x_1^2} \int\limits_0^1 x_2\,\mathrm{d}x_2 = x_1 e^{-x_1^2}, \quad x_1 \geqslant 0.$$

$$f_{X_1} x_1 = 0, \quad x_1 < 0$$

b)

$$f_{X_1}(x_2) = \int\limits_0^\infty f(x_1, x_2)\,\mathrm{d}x_1 = 4x_2 \int\limits_0^\infty x_1 e^{-x_1^2}\,\mathrm{d}x_1 = -\frac{e^{-x_1^2}}{2}\Bigg|_0^\infty = 2x_2, \quad x_2 \in [0,1]$$

$$\Longrightarrow f_{X_1}(x_1 | X_2 = \frac{1}{2}) = \frac{f(x_1, \frac{1}{2})}{f_{X_2}(\frac{1}{2})} = 2x_1 e^{-x_1^2}, \quad x_1 \geqslant 0, x_1 < 0.$$

c)

$$\mathrm{E}(X_1) = \int\limits_0^\infty x_1 f_{X_1}(x_1)\,\mathrm{d}x_1 \underbrace{=}_{a)} 2\int\limits_0^\infty x_1^2 e^{-x_1^2}\,\mathrm{d}x_1$$

$$= \int\limits_{-\infty}^\infty x_1^2 e^{-x_1^2}\,\mathrm{d}x_1 \underbrace{=}_{x_1 = \frac{t^2}{2}} \int\limits_{-\infty}^\infty \frac{t^2}{2} e^{-\frac{t^2}{2}} \frac{1}{\sqrt{2}}\,\mathrm{d}t = \frac{\sqrt{\pi}}{2} \int\limits_{-\infty}^\infty t^2 \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}\,\mathrm{d}t = \frac{\sqrt{\pi}}{2}$$

$$\mathrm{E}(X_2) = \int\limits_0^1 x_2 f_{X_2}(x_2)\,\mathrm{d}x_2 \underbrace{=}_{b)} \int\limits_0^1 2x_2^2\,\mathrm{d}x_2 = \frac{2}{3} \Longrightarrow \mathrm{E}(X) = \left(\frac{\sqrt{\pi}}{2}, \frac{2}{3}\right)^T$$

d) Nach a), b) gilt

$$f_{X_1} f_{X_2} = f(x_1, x_2)\ \forall x_1, x_2 \in \mathbb{R} \quad \Longrightarrow X_1, X_2 \quad \text{sind unabhängig.} \quad \mathrm{Cov}(X_1, X_2) = 0.$$

e)

$$P(X_1 > X_2) = \int\limits_0^1 \int\limits_{x_2}^\infty f(x_1, x_2)\,\mathrm{d}x_1\,\mathrm{d}x_2 = \int\limits_0^1 4x_2 \int\limits_{x_2}^\infty x_1 e^{-x_1^2}\,\mathrm{d}x_1\,\mathrm{d}x_2$$

$$= \int\limits_0^1 4x_2 - \frac{e^{-x_1^2}}{2}\Bigg|_{x_1=x_2}^{\infty} \,\mathrm{d}x_2 = \int_0^1 4x_2 \frac{e^{-x_2^2}}{2}\,\mathrm{d}x_2 = -e^{-x_2^2}\Big|_0^1 = 1 - e^{-1}$$

Aufgabe 2

a)

$$1 = c\int_{-} 1^1 \int_{-} 1^1 x_1^2 x_2^2 \,\mathrm{d}x_1\,\mathrm{d}x_2 = \int_{-} 1^1 \frac{2}{3} + 2x_2^2\,\mathrm{d}x_2$$

$$= c(\frac{4}{3} + \frac{4}{3}) = \frac{8c}{3} \implies c = \frac{3}{8}$$

b)

Für $x_1 \in [-1, 1]$ gilt

$$f_{X_1}(x_1) = \int\limits_{-1}^{1} f(x_1, x_2)\,\mathrm{d}x_2 = c\int_{-1}^{1} x_1^2 + x_2^2\,\mathrm{d}x_2 = c(2x_1^2 + \frac{2}{3})$$

$$f_{X_2}(x_2|X_1 = x_1) = \frac{f(x_1, x_2)}{f_{X_1}(x_1)} = \frac{c(x_1^2 + x_2^2)}{c(2x_1^2 + \frac{2}{3})} = \frac{x_1^2 + x_2^2}{2x_1 + \frac{2}{3}} \quad \text{falls} \quad x_2 \in [-1, 1]$$

$$f_{X_2}(x_2|X_1 = x_1) = 0 \quad \text{für} \quad x_2 \in \mathbb{R}\backslash[-1, 1]$$

c)

$$\mathrm{E}(X_1) = \int\limits_{-1}^{1} x_1 f_{X_1}(x_1)\,\mathrm{d}x_1 \underbrace{=}_{b)} = c\int_{-1}^{1} 2x_1^3 + \frac{2x_1}{3}\,\mathrm{d}x_1 = 0 \quad \mathrm{E}(X_2) = 0 \quad \text{analog.}$$

$$\mathrm{Cov}(X_1, X_2) = \mathrm{E}(X_1X_2) = c\int\limits_{-1}^{1}\int\limits_{-1}^{1} x_1x_2(x_1^2 + x_2^2)\,\mathrm{d}x_1\,\mathrm{d}x_2$$

$$= c\int\limits_{-1}^{1}\int\limits_{-1}^{1} \underbrace{x_1^3x_2 + x_1x_2^3}_{\text{ungerade Funktion von} \quad x_1} \mathrm{d}x_1\,\mathrm{d}x_2 = 0$$

d) X_1 und X_2 sind nicht unabhängig, denn nach b) ist die bedingte Dichte der zweiten Variable $f_{X_2}(x_2|X_1 = x_1)$ nicht unabhängig von x_1.

Aufgabe 3

Für $j = 1, \dots, p$ und alle $x_j \in \mathbb{R}$ gilt:

$$\begin{aligned} f_{X_j}(x_j) &= \int \cdots \int f(x_1, \dots x_p)\, \mathrm{d}x_1 \dots \mathrm{d}x_{j-1}\, \mathrm{d}x_{j+1} \dots \mathrm{d}x_p \\ &= \int \cdots \int h_1(x_1) \cdots h_p(x_p)\, \mathrm{d}x_1 \dots \mathrm{d}x_{j-1}\, \mathrm{d}x_{j+1} \dots \mathrm{d}x_p \end{aligned}$$

$$= h_j(x_j) = \prod_{i \neq j} \int h_i(x_i)\, \mathrm{d}x_i \implies f_{X_j}(x_j) = c_j h_j(x_j) \quad \text{für eine Konstante} \quad c_j$$

Wegen $1 = \int f_{X_j}(x_j)\, \mathrm{d}x_j = c_j \int h_j(x_j)\, \mathrm{d}x_j$ ist $c_j = \frac{1}{\int h_j(x_j)\, \mathrm{d}x_j}$.

Für alle $x_1, \dots, x_p \in \mathbb{R}$ gilt:

$$f_{X_1}(x_1), \cdots f_{X_p}(x_p) = \frac{h_1(x_1)}{\int h_1(u_1)\, \mathrm{d}u_1} \cdots \frac{h_p(x_p)}{\int h_p(u_p)\, \mathrm{d}u_p}$$

$$= \frac{f(x_1, \dots, x_p)}{\int \cdots \int h_1(u_1) \cdots h_p(u_p)\, \mathrm{d}u_1 \dots \mathrm{d}u_p} = \frac{f(x_1, \dots, x_p)}{\int \cdots \int f(u_1, \dots, u_p)\, \mathrm{d}u_1 \dots \mathrm{d}u_p} = f(x_1, \dots, x_p)$$

Daher sind $X_1, \dots, X_p$ unabhängig.

$\mathrm{Cov}(X,Y)$ ist ein Maß für den linearen Zusammenhang zwischen den beiden Zufallsvariablen X und Y. Der Wert ist aber schwer interpretierbar, denn er ist maßstabsabhängig.

$$\mathrm{Cov}(aX, bY) = ab\mathrm{Cov}(X,Y) \quad \forall a, b \in \mathbb{R}$$

Man verwendet daher ein standardisiertes Zusammenhangsmaß, nämlich den Korrelationskoeffizienten. Der *Korrelationskoeffizient* von Zufallsvariablen X und Y ist definiert durch $\rho(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}}$, sofern $\mathrm{Var}(X), \mathrm{Var}(Y) > 0$.

$(\mathrm{Var}(X) = 0 \iff P(X = c) = 1 \quad$ für ein $\quad c \in \mathbb{R}$. ρ ist maßstabsunabhängig:

$$\rho(aX, bY) = \frac{ab\mathrm{Cov}(X,Y)}{\sqrt{a^2\mathrm{Var}(X)b^2\mathrm{Var}(Y)}} = \frac{a}{|a|}\frac{b}{|b|}\rho(X,Y), \quad a, b \neq 0$$

Vorzeichen können sich jedoch ändern.

Satz 9 *X und Y seien Zufallsvariablen mit* $\mathrm{Var}(X) > 0$ *und* $\mathrm{Var}(Y) > 0 \implies -1 \leqslant \rho(X,Y) \leqslant 1$.

$$\rho(X,Y) = 1 \iff \exists \alpha > 0, \beta \in \mathbb{R} \quad \text{so dass} \quad P(Y = \alpha X + \beta) = 1$$

$$\rho(X,Y) = -1 \iff \exists \alpha < 0, \beta \in \mathbb{R} \quad \text{so dass} \quad P(Y = \alpha X + \beta) = 1$$

Lemma 10 *Cauchy-Schwarz Ungleichung.*

Für Zufallsvariablen X und Y gilt $(\mathrm{E}(XY))^2 \leqslant \mathrm{E}(X^2)\mathrm{E}(Y^2)$ und

$$(\mathrm{E}(XY))^2 = \mathrm{E}(X^2)\mathrm{E}(Y^2) \iff \exists a, b \neq 0 \in \mathbb{R} : P(aX = bY) = 1$$

Beweis *Setze* $\alpha = \mathrm{E}(Y^2)\beta = \mathrm{E}(XY)$.

Behauptung klar, falls $\alpha = 0$, da dann $P(Y = 0) = 1$ und $(\mathrm{E}(XY)) = 0 = \mathrm{E}(X^2)\mathrm{E}(Y^2)$ und es gilt $P(aX = bY) = 1$ mit $a = 0, b = 1$.

Sei nun α nicht degeniert, also $\alpha \geqslant 0$.

$$0 \leqslant \mathrm{E}[(\alpha X - \beta Y)^2] = \alpha^2\mathrm{E}(X^2) - 2\alpha\beta \underbrace{\mathrm{E}(XY)}_{=\beta} + \beta^2 \underbrace{\mathrm{E}(Y^2)}_{=\alpha}$$

$$= \alpha^2\mathrm{E}(X^2) - \alpha\beta^2 = \alpha(\mathrm{E}(Y^2)\mathrm{E}(X^2) - (\mathrm{E}(XY))^2) \underbrace{\implies}_{\alpha>0} (\mathrm{E}(XY))^2 \leqslant \mathrm{E}(X^2)\mathrm{E}(Y^2)$$

Falls $(\mathrm{E}(XY))^2 = \mathrm{E}(X^2\mathrm{E}(Y^2))$, dann

$$\mathrm{E}[(\alpha X - \beta Y)^2] = 0 \implies P(\alpha X - \beta Y = 0) = 1 \quad \text{also} \quad P(aX = bY) = 1 \quad \text{mit} \quad a = \alpha \neq 0, b = \beta$$

Falls $P(aX = bY) = 1$ mit z.B. $a \neq 0$m dann $P(X = \frac{b}{a}Y) = 1$ und

$$(\mathrm{E}(XY))^2 = (\mathrm{E}(\frac{b}{a}Y^2))^2 = \frac{b^2}{a^2}\mathrm{E}(Y^2)\mathrm{E}(Y^2) = \mathrm{E}(X^2)\mathrm{E}(Y^2)$$

Erinnerung Satz 9 und Lemma 10:

$$-1 \leqslant \rho(X, Y) \leqslant 1$$

Cauchy Schwarz:

$$(\mathrm{E}(XY))^2 \leqslant \mathrm{E}(X^2)\mathrm{E}(Y^2)$$

X und Y seien Zufallsvariablen mit $\mathrm{Var}(X) > 0, \mathrm{Var}(Y) > 0$

$$(\rho(X, Y))^2 = \frac{\mathrm{Cov}(X, Y)^2}{\mathrm{Var}(X)\mathrm{Var}(Y)} = \frac{(\mathrm{E}[(X - \mathrm{E}(X))(Y - \mathrm{E}(Y))])^2}{\mathrm{Var}(X)\mathrm{Var}(Y)}$$

$$\underbrace{\leqslant}_{\text{Lemma 10}} \frac{\mathrm{E}[(X - \mathrm{E}(X))^2]\mathrm{E}[(Y - \mathrm{E}(Y))]^2}{\mathrm{Var}(X)\mathrm{Var}(Y)} = 1$$

$$|\rho(X, Y)| = 1 \underbrace{\Longrightarrow}_{\text{Lemma 10}} = \exists\, a, b \in \mathbb{R}, \quad \text{nicht beide} \quad = 0, \quad \text{mit}$$

$$(*)P(a - (X - \mathrm{E}(X)) = b(Y - \mathrm{E}(Y))) = 1$$

Hier muss zusätzlich $a \neq 0$ sein, denn sonst $P(Y = \mathrm{E}(Y)) = 1$ im Widerspruch zu $\mathrm{Var}(Y) > 0$. Ebenso muss $b \neq 0$ sein.

Also kann man $(*)$ schreiben als

$$P(Y = \underbrace{\frac{a}{b}}_{\alpha} X \underbrace{- \frac{a}{b}\mathrm{E}(X) + \mathrm{E}(Y))}_{\beta}$$

Damit gezeigt

$$|\rho(X, Y) = 1| \implies \exists \alpha \neq 0, \beta \in \mathbb{R} \quad \text{mit} \quad P(Y = \alpha X + \beta) = 1$$

Bzgl. Vorzeichen von α :

Falls $P(Y = \alpha X + \beta) = 1$ mit $\alpha \neq 0$, dann

$$\rho(X, Y) = \rho(X, \alpha X + \beta) = \rho(X, \alpha X) = \frac{\alpha}{|\alpha|}\frac{\mathrm{Cov}(X, X)}{\mathrm{Var}(X)} = \begin{cases} 1, & \alpha > 0 \\ -1, & \alpha < 0. \end{cases}$$

Beispiel 11 *Betrachte zweimaligen Wurf einer Münze mit $P(\text{'Kopf'}) = p \in (0,1)$.*

Sei X die Anzahl Würfe, in denen Kopf fällt und Y der Rest. Berechnen nun $\rho(X,Y)$. Betrachte Erwartungswerte:

$$\mathrm{E}(X) = 2p,\ \mathrm{Var}(X) = 2p(1-p),\ \mathrm{E}(Y) = 2(1-p),\ \mathrm{Var}(Y) = 2p(1-p),$$

$$\mathrm{E}(XY) = 1 \cdot P(X=1, Y=1) = 2p(1-p)$$

$$\mathrm{Cov}(X,Y) = \mathrm{E}(XY) - \mathrm{E}(X)\mathrm{E}(Y) = 2p(1-p) - 4p(1-p) = -2p(1-p)$$

$$\Longrightarrow\ \rho(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}} = \frac{-2p(1-p)}{2p(1-p)} = -1$$

Hier ist $P(X = \alpha X + \beta) = 1$ mit $\alpha = -1$ und $\beta = 2$.

Die *Korrelationsmatrix* eines Zufallsvektors $X(X_1, \dots, X_p)^T$ ist definiert durch die Matrix der einzelnen Korrelationskoeffizienten

$$P = (\rho_j k)_{j,k=1,\dots,p}, \rho_{jk} = \rho(X_j, X_k).$$

sofern $\rho_j = \sqrt{\mathrm{Var}(X_j)} > 0$ für alle $j = 1, \dots, p$.

P ist die Kovarianzmatrix des Vektors der standardisierten Zufallsvariablen $Z_j = \frac{X_j - \mathrm{E}(X_j)}{\sigma_j}$.

Mit $\Delta = \mathrm{diag}(\sigma_1, \dots, \sigma_p)$ gilt

$$P = \mathrm{Var}((Z_1, \dots, Z_p)^T) = \mathrm{Var}(\Delta^{-1}(X - \mathrm{E}(X))) \underbrace{=}_{\text{Lemma 7}} \Delta^{-1}\mathrm{Var}(X)\Delta^{-1}.$$

X_j und X_k heißen *unkorreliert*, falls $\rho(X_j, X_k) = 0$.

Satz 12 *Transformationssatz für Dichten.*

Sei X ein p-dimensionaler Zufallsvektor mit Dichte f_X (nicht Randdichte!).

Sei $T = \{x \in \mathbb{R}^p : f_X(x) > 0\}$ Träger von X und M sei eine offene Teilmenge von T mit $P(X \in M) = 1$.

Sei $h : M \to h(M) \subset \mathbb{R}^p$ eine bijektive Abbildung, stetig differenzierbar mit $J(x) = \det \mathrm{D}h(x) = \det \frac{\partial h_i(x)}{\partial x_j}\,_{i,j=1,\dots,p} \neq 0\ \forall x \in M$. D ist Differentialoperator.

Dann hat $Y = h(X)$ die Dichte

$$f_Y(y) = \begin{cases} f_X(h^{-1}(y))|\det(\mathrm{D}h^-1(y))| = \frac{f_X(h^-1(y))}{|J(h^-1(y))|}, & y \in h(M) \\ 0 \quad \text{sonst.} \end{cases}$$

Beispiel 13 *Log-Normalverteilung.*

Sei $X\tilde{N}(\mu, \sigma^2)$, also
$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\,(-\frac{(x-\mu)^2}{2\sigma^2}),\; x \in \mathbb{R}.$$

Hier ist $p = 1$, $T = \mathbb{R}$, $M = \mathbb{R}$, $h(x) = e^x$, $h(M) = (0, \infty)$, $h^-1(y) = \log y$, Funktionaldeterminante $J(x) = h'(x) = e^x \neq 0\; \forall x \in M$.

$\implies$ $Y = h(X)$ hat die Dichte:

$$f_Y(y) = \frac{f_X(h^{-1}(y))}{|J(h^-1(y))|} = \frac{f_X(\log y)}{|J(\log y)|} = \frac{1}{\sqrt{2\pi\sigma^2}} \exp(-\frac{(\log y - \mu)^2}{2\sigma^2})\frac{1}{y}$$

Beispiel 14 *Sei $X = (X_1, X_2)^T$ ein 2-dimensionaler Zufallsvektor mit Dichte f_X und es sei $T = \{x \in \mathbb{R}^2 : f_X(x) > 0\}$ offen.*

Bestimme Dichte von $Y = \begin{pmatrix} X_1 + X_2 \\ X_1 - X_2 \end{pmatrix}$.

Setze $M = T$ und

$$h(x) = \begin{pmatrix} h_1(x) \\ h_2(x) \end{pmatrix} = \begin{pmatrix} x_1 + x_2 \\ x_1 - x_2 \end{pmatrix} \quad \text{für} \quad x = (x_1, x_2)^T \in M$$

$$J(x) = \det(\frac{\partial h_i(x)}{\partial x_j})_i, j = 1, 2 = \det \begin{pmatrix} 1 & 1 \\ 1 & -1 \end{pmatrix} = 2 \neq 0$$

$$h(x) = (Y_1, Y_2)^T \iff \begin{matrix} x_1 + x_2 = y_1 \\ x_1 - x_2 = y_2 \end{matrix} \iff x_1 = \frac{y_1 + y_2}{2},\; x_2 = \frac{y_1 - y_2}{2}$$

Also
$$h^{-1}(y) = \begin{pmatrix} \frac{y_1 + y_2}{2} \\ \frac{y_1 - y_2}{2} \end{pmatrix}$$

Insbesondere ist $h : M \to h(M)$ bijektiv.

$\implies$ Y hat die Dichte

$$f_Y(y) = \begin{cases} \frac{f_X(h^-1(y))}{|J(h^-1(y))|} = \frac{1}{2} f_X(\frac{Y_1+Y_2}{2}, \frac{Y_1-Y_2}{2}),\; y \in h(M) \\ 0 \quad \text{sonst.} \end{cases}$$

Falls $y \in \mathbb{R}^2 \backslash h(M)$, dann $h^-1(y) \in \mathbb{R}^2 \backslash M$, also $f_X(h^-1(y)) = 0$.

$$f_Y(y) = \frac{1}{2} f_X\left(\frac{Y_1 + Y_2}{2}, \frac{Y_1 - Y_2}{2}\right),\; \forall\, y \in \mathbb{R}^2$$

Wichtiger Spezialfall: Sind X_1 und X_2 unabhängig, also

$$f_X(x) = f_{X_1}(x_1) f_{X_2}(x_2) \quad \text{für alle} \quad x \in \mathbb{R}$$

Dann hat $X_1 + X_2$ die Dichte

$$f_{X_1+X_2}(z) = f_{Y_1}(z) = \int\limits_{-\infty}^{\infty} f_Y(z, y_2)\,\mathrm{d}y_2 = \frac{1}{2} \int\limits_{-\infty}^{\infty} f_X\left(\frac{t+y_2}{2}, \frac{z-y^2}{2}\right)\mathrm{d}y_2$$

$$= \frac{1}{2} \int\limits_{-\infty}^{\infty} f_{X_1}\left(\frac{z-y^2}{2}\right) f_{X_2}\left(\frac{z-y^2}{2}\right)\mathrm{d}y_2 \underbrace{=}_{t=\frac{z+y_2}{2}} \frac{1}{2} \int\limits_{-\infty}^{\infty} f_{X_1}(t) f_{X_2}\left(\frac{z-(2t-z)}{2}\right) 2\,\mathrm{d}t$$

$$= \int\limits_{-\infty}^{\infty} f_{X_1}(t) f_{X_2}(z-t)\,\mathrm{d}t$$

Dies nennt man Faltung von f_{X_1} und f_{X_2}.
Beispiel 15 *Kunden erscheinen zu zufälligen Zeitpunkten in einem Laden.*

Der erste Kunde erscheint zur Zeit X_1. Die Zeitspanne zwischen dem Eintreffen von Kunde $i-1$ und Kunde i sei $X_i, i = 2, 3, \ldots$
Dabei seien $X_1, X_2, \ldots$ unabhängige, exponentialverteilte Zufallsvariablen mit Erwartungswert $\mathrm{E}[X_i] = \frac{1}{\lambda}, i = 1, 2, \ldots$
Bestimme die gemeinsame Dichte der Ankunftszeiten $T_1, \ldots, T_n$ der ersten n Kunden

$$T_i = X_1 + \cdots + X_i = \sum_i X_i.$$

X_i hat Dichte $f_{X_i}(x_i) = \lambda e^{-\lambda x_i}, x_i > 0$

$\Rightarrow X = \begin{pmatrix} X_1 \\ \vdots \\ X_n \end{pmatrix}$ hat Dichte

$$f_X(x_1, \ldots, x_n) = \prod_{i=1}^{n} f_{X_i}(x_i) = \lambda^n e^{-\lambda \sum_{i=1}^n x_i}, x_1, \ldots, x_n > 0$$

$M = \{x \in \mathbb{R}^n : f_X(x) > 0\}$ ist offen.

$$T = \begin{pmatrix} T_1 \\ \vdots \\ T_n \end{pmatrix} = h(X) \quad \text{mit} \quad h_i(x_1, \ldots, x_n) = x_1 + \ldots, x_n$$

$$h(M) = \{t = \begin{pmatrix} t_1 \\ \vdots \\ t_n \end{pmatrix} \in \mathbb{R}^n : 0 < t_1 < t_2 < \cdots < t_n\}$$

$$t = h(x) \Leftrightarrow t_j = \sum_{i=1}^{j} x_i, j = 1, \ldots, n \Leftrightarrow x_1 = t_1 \wedge x_j = t_j - t_{j-1}, j = 2, \ldots, n$$

Also

$$h^{-1}(t) = \begin{pmatrix} t_1 \\ t_2 - t_1 \\ \vdots \\ t_n - t_{n-1} \end{pmatrix}, J(x) = \det \left(\frac{\partial h_i(x)}{\partial x_j} \right)_{i,j=1,\ldots,n} = \det \begin{pmatrix} 1 & 0 & 0 & \ldots & 0 \\ 1 & 1 & 0 & \ldots & 0 \\ \vdots & & & \ddots & \vdots \\ 1 & 1 & 1 & \ldots & 1 \end{pmatrix} = 1 \neq 0$$

$$\Longrightarrow \begin{pmatrix} T_1 \\ \vdots \\ T_n \end{pmatrix}$$

hat die Dichte:

$$f_T(t_1, \ldots, t_n) = \frac{f_X(h^{-1}(t_1, \ldots, t_n))}{|J(h^{-1}(t_1, \ldots, t_n))|} = \lambda^n e^{-\lambda(t_1 + t_2 - t_1 + \cdots + t_n - t_{n-1})} = \lambda^n e^{-\lambda t_n}, 0 < t_1 < \cdots < t_n$$

Sei $X = \begin{pmatrix} X_1 \\ X_2 \end{pmatrix}$ ein p-dimensionaler Zufallsvektor, wobei X_1 k-dimensional und X_2 $(p-k)$-dimensional.

Der **bedingte Erwartungswert** von $g(X_1)$ gegeben $X_2 = x_2$ ist definiert durch:

$$\mathrm{E}[g(X_1)|X_2 = x_2] = \int g(x_1) f_{X_1}(x_1|X_2 = x_2)\, \mathrm{d}x_1$$

Analog berechnet man die **bedingte Kovarianzmatrix** $\mathrm{Var}(X_1|X_2 = x_2)$ unter Verwendung der bedingten Dichte $f_{X_1}(x_1|X_2 = x_2)$.

Übungsblatt 2

Aufgabe 4

Zeigen Sie: Ist X eine $m \times n$ Zufallsmatrix und sind $A \in \mathbb{R}^{k\times n}, B \in \mathbb{R}^{m\times l}, C \in \mathbb{R}^{k\times l}$, dann gilt:

$$\mathrm{E}[AXB + C] = A\mathrm{E}[X]B + C$$

Sei

$$A = (a_{ij})_{i=1,\dots,k;j=1,\dots,n},$$
$$X = (x_{ij})_{i=1,\dots,n;j=1,\dots,m}, B = (b_{ij})_{i=1,\dots,m;j=1,\dots,l}, C = (c_{ij})_{i=1,\dots,k;\, j=1,\dots,l}.$$

$$\forall\, i = 1,\dots,k; j = 1,\dots,l : \{\mathrm{E}[AXB + C]\}_{ij} = \mathrm{E}[\{AXB + C\}_{ij}]$$
$$= \mathrm{E}\left[\sum_{\mu=1}^{n}\sum_{\nu=1}^{m} a_{i\mu}X_{\mu\nu}B_{\nu j} + C_{ij}\right] = \sum_{\mu}\sum_{\nu} a_{i\mu}\mathrm{E}[X_{\mu\nu}]b_{\nu j} + c_{ij} = A\mathrm{E}[X]B + C_{ij}$$

Aufgabe 5

Es sei $X = (X_1, X_2, X_3)^T$ ein Zufallsvektor mit Kovarianzmatrix

$$\mathrm{Var}(X) = \begin{pmatrix} 4 & 0 & -2 \\ 0 & 3 & 0 \\ -2 & 0 & 3 \end{pmatrix}$$

Berechnen Sie für jedes $\alpha \in \mathbb{R}$ den Korrelationskoeffizienten

$$\rho(\alpha X_1 - 7X_2 + X_3 + 2, \alpha X_1 + 7X_2 - X_3 - 2)$$

Setze

$$Y_1 = \alpha X_1 - 7X_2 + X_3 + 2, Y_2 = \alpha X_1 + 7X_2 - X_3 - 2, Y = \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix}.$$

$$\mathrm{Var}(Y) = \mathrm{Var}\left(\begin{pmatrix} \alpha & -7 & 1 \\ \alpha & 7 & -1 \end{pmatrix} X + \begin{pmatrix} 2 \\ -2 \end{pmatrix}\right) = \begin{pmatrix} \alpha & -7 & 1 \\ \alpha & 7 & -1 \end{pmatrix} \mathrm{Var}(X) \begin{pmatrix} \alpha & \alpha \\ -7 & 7 \\ 1 & -1 \end{pmatrix} = \dots$$
$$= \begin{pmatrix} 4\alpha^2 - 4\alpha + 150 & 4\alpha^2 - 150 \\ & 4\alpha^2 + 4\alpha + 150 \end{pmatrix} = \begin{pmatrix} \mathrm{Var}(Y_1) & \mathrm{Cov}(Y_1, Y_2) \\ & \mathrm{Var}(Y_2) \end{pmatrix}$$

$$\rho(Y_1, Y_2) = \frac{\mathrm{Cov}(Y_1, Y_2)}{\sqrt{\mathrm{Var}(Y_1)\mathrm{Var}(Y_2)}}$$
$$= \frac{4\alpha^2 - 150}{\sqrt{(4\alpha^2 - 4\alpha + 150)(4\alpha^2 + 4\alpha + 150)}} = \frac{2\alpha^2 - 75}{\sqrt{4\alpha^4 + 296\alpha^2 + 5625}}$$

Aufgabe 6

Betrachte Sie n Ehepaare und nehmen Sie an, dass jede der $2n$ Personen mit Wahrscheinlichkeit $p \in (0,1)$ in 20 Jahren noch lebt (unabhängig von den anderen). Es sei X die Anzahl der in 20 Jahren noch lebenden Personen und Y sei die Anzahl der Ehepaare, bei denen in 20 Jahren sowohl die Frau als auch der Mann noch lebt. Berechnen Sie den Korrelationskoeffizienten von X und Y.

Seien $F_1, \dots, F_n, M_1, \dots, M_n$ unabhängige Bernoulli-Variablen.
$F_i = 1 \leftrightarrow$ Frau von Paar i lebt in 20 Jahren.
$M_i = 1 \leftrightarrow$ Mann von Paar i lebt in 20 Jahren.

$$P(F_i = 1) = P(M_i = 1) = p, i = 1, \dots, n$$

$$X = \sum_{i=1}^{n} F_i + \sum_{i=1}^{n} M_i, Y = \sum_{i=1}^{n} F_i M_i$$

$$X \sim \mathrm{BIN}(2n, p) \Rightarrow \mathrm{E}[X] = 2n \cdot p, \mathrm{Var}(X) = 2n \cdot p \cdot (1-p)$$

$$P(F_i M_i = 1) = P(F_i = 1) P(M_i = 1) = p^2$$

$$\Rightarrow Y \sim \mathrm{BIN}(n, p^2) \Rightarrow \mathrm{E}[Y] = n \cdot p^2, \mathrm{Var}(Y) = n \cdot p^2 \cdot (1-p^2)$$

$$\mathrm{E}[XY] = \mathrm{E}\left[\left(\sum_{i=1}^{n} F_i + \sum_{i=1}^{n} M_i\right)\left(\sum_{j=1}^{n} F_j M_j\right)\right] =$$

$$\sum_i \sum_j \mathrm{E}[F_i F_j M_j] + \mathrm{E}[M_i F_j M_j] = \sum_i \mathrm{E}[F_i^2 M_i] + \mathrm{E}[M_i^2 F_i] + \sum_i \sum_{j \neq i} \mathrm{E}[F_i F_j M_j] + \mathrm{E}[M_i F_j M_j]$$

$$= 2np^2 + 2n(n-1)p^3$$

Benutze dabei:

$$\mathrm{E}[F_i^2 M_i] = \mathrm{E}[F_i M_i] = \mathrm{E}[F_i]\mathrm{E}[M_i] = p^2 = \mathrm{E}[M_i^2 F_i]$$

$$\mathrm{E}[F_i F_j M_j] = \mathrm{E}[F_i]\mathrm{E}[F_j]\mathrm{E}[M_j] = p^3 = \mathrm{E}[M_i F_j M_j]$$

$$\Longrightarrow \rho(X,Y) = \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}} = \frac{\mathrm{E}[XY] - \mathrm{E}[X]\mathrm{E}[Y]}{\sqrt{2np(1-p)np^2(1-p)^2}}$$

$$= \frac{2np^2 + 2n^2p^3 - 2np^3 - 2n^2p^3}{np\sqrt{2p(1-p)(1-p^2)}} = \frac{2p - 2p^2}{(1-p)\sqrt{2p(1+p)}} = \sqrt{\frac{2p}{1+p}}$$

Aufgabe 7

Es seien $X_1, \dots, X_n$ unabhängig, identisch verteilte p-dimensionale Zufallsvektoren mit $\mathrm{E}[X_1] = \mu$ und $\mathrm{Var}(X_1) = \Sigma$. Es sei

$$\bar{X} = \frac{1}{n} \sum_{j=1}^{n} X_j$$

$$S = \frac{1}{n} \sum_{j=1}^{n} (X_j - \bar{X})(X_j - \bar{X})^T$$

Drücken Sie $\mathrm{E}[\bar{X}]$, $\mathrm{Var}(\bar{X})$ und $\mathrm{E}[S]$ durch μ, Σ und n aus.

$$\mathrm{E}[\bar{X}] = \mathrm{E}\left[\frac{1}{n}\sum_j X_j\right] = \frac{1}{n}\sum_j \mathrm{E}[X_j] = \frac{1}{n}\sum_j \mu = \mu$$

$$\mathrm{Var}(\bar{X}) = \mathrm{Var}\left(\frac{1}{n}\sum_j X_j\right) = \frac{1}{n^2}\sum_j \mathrm{Var}(X_j) = \frac{1}{n^2}\cdot n\Sigma = \frac{1}{n}\cdot\Sigma$$

$$\mathrm{E}[S] = \mathrm{E}\left[\frac{1}{n}\sum_j (X_j-\bar{X})(X_j-\bar{X})^T\right] = \frac{1}{n}\cdot\mathrm{E}\left[\left(\sum_j X_j X_j^T\right) - \bar{X}\left(\sum_j X_j^T\right) - \left(\sum_j X_j\right)\bar{X}^T + n\bar{X}\bar{X}^T\right]$$

$$= \frac{1}{n}\cdot\sum_j \mathrm{E}[X_j X_j^T] - \mathrm{E}[\bar{X}\bar{X}^T] = \frac{1}{n}\sum_j(\mathrm{Var}(X_j) + \mathrm{E}[X_j]\cdot\mathrm{E}[X_j^T]) - \mathrm{Var}(\bar{X}) - \mathrm{E}[\bar{X}]\cdot\mathrm{E}^T[\bar{X}]$$

$$] = \Sigma + \mu\mu^T - \frac{1}{n}\Sigma - \mu\mu^T = \frac{n-1}{n}\Sigma$$

Übungsblatt 3

Aufgabe 8

Es seien X_1 und X_2 unabhängige exponentialverteilte Zufallsvariablen mit Parameter $\lambda > 0$. Bestimmen Sie die gemeinsame Dichte von

$$Y_1 = X_1 + X_2, Y_2 = \frac{X_1}{X_2}$$

Sind Y_1 und Y_2 unabhängig?

Es seien X_i und T_i wie in Beispiel 15 der Vorlesung. Berechnen Sie für jedes $t > 0$

$$\mathrm{E}[X_1 | X_2 = t]$$

Aufgabe 9

Der Zufallsvektor $X = (X_1, X_2, X_3)^T$ habe die Dichte

$$f(x_1,x_2,x_3) = \begin{cases} 2x_1(x_2+x_3), & 0 \leqslant x_i \leqslant 1, i = 1,2,3 \\ 0, & \text{sonst} \end{cases}$$

(i)

Berechnen Sie für jedes $x_3 \in [0,1]$

$$\mathrm{E}\left[\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \mid X_3 = x_3\right]$$

(ii)

Berechnen Sie für jedes $x_3 \in [0,1]$

$$\mathrm{Var}\left[\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \mid X_3 = x_3\right]$$

$$X = (X_1, X_2)^T, \ \mathrm{E}[g(X_1)|X_2 = x_2] = \int g(x_1) f_{X_1}(x_1|X_2 = x_2)\,\mathrm{d}x_1$$

Die Funktion $x_2 \to \mathrm{E}(X_1|X_2 = x_2)$ wird als Regressionsfunktion bezeichnet für die Regression von X_1 auf X_2. Ist die Funktion von der Form $\alpha + \beta x_2$, so spricht man von linearer Regression.

Falls X_1 und X_2 unabhängig sind, dann gilt

$$\mathrm{E}[g(X_1)|X_2 = x_2] = \int g(x_1) f_{X_1}(x_1|X_2 = x_2)\,\mathrm{d}x_1 = \int g(x_1) f_{X_1}(x_1)\,\mathrm{d}x_1 = \mathrm{E}[g(X_1)] \quad \forall x_2$$

und die Regressionsfunktion ist konstant.
Beispiel 16 $(X_1, X_2)^T$ *sei wie in Beispiel 5. Berechne Regressionsfunktion* $\mathrm{E}(X_1|X_2 = x_2)$ *und* $\mathrm{Var}(X_1|X_2 = x_2)$, $0 \leqslant x_2 \leqslant 1$.

$$\mathrm{E}(X_1|X_2 = x_2) = \int x_1 f_{X_1}(x_1|X_2 = x_2)\,\mathrm{d}x_1 = \int_0^1 x_1 \frac{x_1 + x_2}{0.5 + x_2}\,\mathrm{d}x_1$$

$$= \frac{1}{0.5 + x_2} \int_0^1 x_1^2 + x_1 x_2 \,\mathrm{d}x_1 = \frac{\frac{1}{3} + \frac{x_2}{2}}{0.5 + x_2}$$

$$\mathrm{E}(X_1^2|X_2 = x_2) = \frac{1}{0.5 + x_2} \int_0^1 x_1^3 + x_1^2 x_2 \,\mathrm{d}x_1 = \frac{0.25 + \frac{x_2}{3}}{0.5 + x_2}$$

$$\Longrightarrow \ \mathrm{Var}(X_1|X_2 = x_2) = \frac{\frac{1}{4} + \frac{x_2}{3}}{\frac{1}{2} + x_2} - \left(\frac{\frac{1}{3} + \frac{x_2}{2}}{0.5 + x_2}\right)^2 = \frac{1 + 6x_2 + 6x_2^2}{18(1 + 2x_2)^2}$$

Sei $X = (X_1, X_2)^T$ p-dimensionaler Zufallsvektor mit Dichte f, X_1 k-dimensional, X_2 $(p-k)$-dimensional.

Setze

$$\psi(x_2) = \mathrm{E}(X_1|X_2 = x_2) \ \forall x_2 \in \mathbb{R}^{p-k} : f_{X_2}(x_2) < 0$$

Der k-dimensionale Zufallsvektor $\psi(X_2)$ hat folgende Eigenschaften:

1. $\mathrm{E}[\psi(X_2)] = \mathrm{E}(X_1)$
2. $\psi(X_2)$ ist die beste Approximation von X_1 durch einen Zufallsvektor der Form $g(X_2)$ im Sinne der mittleren quadratischen Abweichung, d.h. für jede Funktion $g : \mathbb{R}^{p-k} \to \mathbb{R}^k$ gilt

$$\mathrm{E}(\|X_1 - g(X_2)\|^2) \geqslant \mathrm{E}(\|X_1 - \psi(X_2)\|^2) \quad \text{wobei} \quad \|u\| = \sqrt{u^T u} \forall u \in \mathbb{R}^k$$

Begründung unter der Annahme $f_{X_2}(x_2) > 0 \ \forall x_2 \in \mathbb{R}^{p-k}$.

1.

$$\mathrm{E}[\psi(X_2)] = \int \psi(x_2) f_{X_2}(x_2)\,\mathrm{d}x_2 = \iint x_1 f_{X_1}(x_1|X_2 = x_2)\,\mathrm{d}x_1 f_{X_2}(x_2)\,\mathrm{d}x_2$$

$$= \iint \frac{x_1 f(x_1,x_2)}{f_{X_2}(x_2)} f_{X_2}(x_2)\,\mathrm{d}x_1\,\mathrm{d}x_2 = \iint x_1 f(x_1,x_2)\,\mathrm{d}x_1\,\mathrm{d}x_2 = \mathrm{E}(X_1)$$

2.

$$\mathrm{E}(\|X_1 - g(X_2)\|^2) = \mathrm{E}(\|X_1 - \psi(X_2) + \psi(X_2) - g(X_2)\|^2)$$

mit der Definition der euklidischen Norm folgt

$$\mathrm{E}(\|X_1 - \psi(X_2)\|^2) + \mathrm{E}(\|\psi(X_2) - g(X_2)\|^2) + 2\mathrm{E}[(X_1 - \psi(X_2))^T(\psi(X_2) - g(X_2))]$$

$$\mathrm{E}[(X_1 - \psi(X_2))^T(\psi(X_2) - g(X_2))] = \iint (x_1 - \psi(x_2))^T - (\psi(x_2) - g(x_2)) f(x_1,x_2)\,\mathrm{d}x_1\,\mathrm{d}x_2$$

$$\int \left[\int (x_1 - \psi(x_2)) f(x_1,x_2)\,\mathrm{d}x_1\right]^T (\psi(x_2) - g(x_2))\,\mathrm{d}x_2$$

und für jedes $x_2 \in \mathbb{R}^{p-k}$ gilt

$$\int (x_1 - \psi(x_2)) f(x_1,x_2)\,\mathrm{d}x_1 = \int x_1 f(x_1,x_2)\,\mathrm{d}x_1 - \int\int t_1 f_{X_1}(t_1|X_2 = x_2) f(x_1,x_2)\,\mathrm{d}x_1$$

$$\int x_1 f(x_1,x_2)\,\mathrm{d}x_1 - \int t_1 \underbrace{f_{X_1}(t_1|X_2 = x_2) \underbrace{\int f(x_1,x_2)\,\mathrm{d}x_1}_{=f_{X_2}(x_2)}}_{=f(t_1,x_2)}\,\mathrm{d}t_1 = 0$$

$$\Longrightarrow\ \mathrm{E}(\|X_1 - g(X_2)\|^2) = \mathrm{E}(\|X_1 - \psi(X_2)\|^2) + \mathrm{E}(\|\psi(X_2) - g(X_2)\|^2) \geqslant \mathrm{E}(\|X_1 - \psi(X_2)\|^2)\ \blacksquare$$

Der Zufallsvektor $\psi(X_2)$ wird auch mit $\mathrm{E}(X_1|X_2)$ bezeichnet. Sind X_1 und X_2 unabhängig, dann gilt

$$\psi(x_2) = \mathrm{E}(X_1|X_2 = x_2) = \mathrm{E}(X_1)$$

also $\mathrm{E}(X_1|X_2) = E(X_1)$.

2 Multivariate Normalverteilung

Lemma 17 *Seien $X_1,\dots,X_p$ unabhängige $N(0,1)$ verteilte Zufallsvariablen.*

$$A \in \mathbb{R}^{p\times p} \quad \text{regulär} \quad , b \in \mathbb{R}^p,\ X = (X_1,\dots,X_p)^T$$

$$f_Y(y) = \frac{1}{(2\pi)^{\frac{p}{2}}\sqrt{\det B}} \exp\left(-\frac{1}{2}(y-b)^T B^{-1}(y-b)\right),\ y \in \mathbb{R}^p$$

wobei $B = AA^T$ und es gilt $E(Y) = b$, $\mathrm{Var}(Y) = B$.

Beweis

X_i hat Dichte $f_{X_i}(x_i) = \frac{1}{\sqrt{2\pi}} e^{\frac{x_i^2}{2}}$ Daher hat X die Dichte

$$f_X(x) = \prod_{i=1}^{p} f_{X_i}(x_i) = \frac{1}{(2\pi)^{\frac{p}{2}}} e^{-\frac{1}{2}\sum_{i=1}^{p} x_i^2} = \frac{1}{(2\pi)^{\frac{p}{2}}} \exp(-\frac{1}{2} x^T x$$

Wende Satz 1.12 an mit $h(x) = Ax + b$

$$h^{-1}(y) = A^{-1}(y - b),\ J(x) = |\det Dh(x)| = |\det A| \neq 0$$

Daher $Y = h(X)$ hat die Dichte

$$f_Y(y) = \frac{f_X(h^{-1}(y))}{|J(h^{-1}(y))|} = \frac{1}{(2\pi)^{\frac{p}{2}} |\det A|} \exp(-\frac{1}{2}(h^{-1}(y))^T h^{-1}(y))$$

$$= \frac{1}{(2\pi)^{\frac{p}{2}} \sqrt{\det B}} \exp(-\frac{1}{2}(y - b)^T (A^{-1})^T A^{-1}(y - b))$$

$$\underbrace{=}_{B^{-1}=(AA^T)^{-1}=(A^T)^{-1}A^{-1}=(A^{-1})^T A^{-1}} \qquad \mathrm{E}(Y) = A\mathrm{E}(X) + b = b$$

$$\mathrm{Var}(Y) = A\,\mathrm{Var}(X)A^T \underbrace{=}_{\mathrm{Var}(X)=\text{Einheitsmatrix}} AA^T = B$$

2.1 Übungsblatt 3

Aufgabe 8

$X = (X_1, X_2)^T$ hat die Dichte

$$f_X(x_1, x_2) = \begin{cases} \lambda^2 e^{\lambda(x_1+x_2)} & x_1, x_2 > 0 \\ 0 & \text{sonst} \end{cases}$$

Wende Satz 1.12 an mit

$$M = (0, \infty),\ h(x_1, x_2) = (x_1 + x_2, x_1/x_2)^T,\ x_1 + x_2 = y \quad \text{und} \quad x_1/x_2 = y_2$$

$$\iff x_2 y_2 + x_2 = y \quad \text{und} \quad x_1 = x_2 x_2 \iff x_2 = \frac{y_1}{1 + x_2} \quad \text{und} \quad \frac{y_1 y_1}{1 + y_2}$$

$$\text{also} \quad h^{-1}(y) = \frac{y_1}{1 + y_2},\ y = (y_1, y_2)^T \in h(M) = (0, \infty)^2$$

$$J(x) = \det Dh(x) = \det \begin{pmatrix} 1 & 1 \\ 1/x_2 & x_1/x_2 \end{pmatrix} = \frac{x_2}{x_2^2} - \frac{1}{x_2} = -\frac{x_1 + x_2}{x_2^2} \neq 0\ \forall X \in;$$

$\implies y = (y_1, y_2)^T = h(X)$ hat die Dichte

$$f_Y(y) \frac{f_X(h^{-1}(y))}{|J(h^{-1}(y))|} = \frac{\lambda^2 e^{-\lambda y_1}}{|-\frac{y_1}{y_1^2/(1+y_2)^2}|}$$

$$= \begin{cases} \frac{\lambda^2 y_1 e^{-\lambda y_2}}{(1+y_2)^2} & y \in (0, \infty)^2 \\ 0 & \text{sonst} \end{cases}$$

$\underbrace{\implies}_{\text{Aufgabe 3}}$ Y_1 und Y_2 sind unabhängig.

Wähle z.B. $h_1(y_1) = \lambda^2 y_1 e^{-\lambda y_1} \mathbb{1}_{(0,\infty)}(y_1),\ h_2(y_2) = \frac{1}{(1+y_2)^2} \mathbb{1}_{(0,\infty)}(y_2)$.

Aufgabe 9

X_1, X_2 seien unabhängige exponentialverteilte Zufallsvariablen mit Erwartungswert $\mathrm{E}(X_i) = \frac{1}{\lambda}$, $T_1 = X_1$, $T_2 = X_1 + X_2$. Nach Beispiel 1.15 ist die Dichte von $T = (T_1, T_2)^T$ gegeben durch

$$f_T(t_1, t_2) = \begin{cases} \lambda^2 e^{-\lambda t_2} & 0 < t_1 < t_2 \\ 0 & \text{sonst} \end{cases}$$

Betrachte Randverteilung zur Berechnung der bedingten Dichte:

Für $t_2 > 0$ gilt

$$f_{T_2}(t_2) = \int_0^\infty f_T(t_1, t_2)\, \mathrm{d}t_1 = \int_0^{t_2} \lambda^2 e^{-\lambda t_2}\, \mathrm{d}t_1 = t_2 \lambda^2 e^{-\lambda t_2}$$

und

$$\begin{cases} f_{T_1}(t_1|T_2=t_2) = \frac{f_T(t_1,t_2)}{f_T(t_2)} = \frac{\lambda^2 e^{-\lambda t_2}}{t_2 \lambda e^{-\lambda t_2}} = \frac{1}{t_2} & 0 < t_1 < t_2 \\ 0 & \text{sonst} \end{cases}$$

$$\implies \mathrm{E}(X_1|t_2=t) = \mathrm{E}(T_1|T_2=t) = \int_0^t \frac{t_1}{t}\,\mathrm{d}t_1 = \frac{t}{2},\ t > 0.$$

Alternativ: Heuristisches Argument

$$\mathrm{E}(X_1|T_2=t) = \mathrm{E}(X_2|T_2=t) \quad \text{wegen Symmetrie}$$

$$\mathrm{E}(X_1|T_2=t) + \mathrm{E}(X_2|T_2=t) = \mathrm{E}(X_1+X_2|T_2=t) = \mathrm{E}(T_2|T_2=t) = t$$

$$\implies \mathrm{E}(X_1|T_2=t) = \mathrm{E}(X_2|T_2=t) = \frac{t}{2} \ \blacksquare$$

Aufgabe 10

Für $x_3 \in [0,1]$ gilt

$$f_{X_3}(x_3) = \iint f(x_1,x_2,x_3)\,\mathrm{d}x_1\,\mathrm{d}x_2 = \int_0^1\int_0^1 2x_1x_2 + x_1x_3\,\mathrm{d}x_1\,\mathrm{d}x_2 = 2\int_0^1 x_1\,\mathrm{d}x_1 \int_0^1 x_2\,\mathrm{d}x_2 + 2x_3\int_0^1 x_1\,\mathrm{d}x_1$$

$$= 2\frac{1}{2}\frac{1}{2} + 2x_3\frac{1}{2} = \frac{1}{2} + x_3$$

Und

$$f_{(X_1,X_2)^T}(x_1,x_2|X_3=x_3) = \frac{f(x_1,x_2,x_3)}{f_{X_3}(x_3)} = \frac{2x_1x_2 + 2x_1x_3}{0.5 + x_3},\ x_1,x_2 \in [0,1]$$

a)

$$\mathrm{E}[(X_1,X_2)^T|X_3=x_3] = \int_0^1\int_0^1 (X_1,X_2)^T \frac{4x_1x_2 + 4x_1x_3}{1+2x_3}\,\mathrm{d}x_1\,\mathrm{d}x_2$$

$$= \frac{4}{1+2x_3}\begin{pmatrix} \int_0^1 x_1^2\,\mathrm{d}x_1 \int_0^1 x_2\,\mathrm{d}x_2 + x_3\int_0^1 x_1^2\,\mathrm{d}x_1 \\ \int_0^1 x_1\,\mathrm{d}x_1 \int_0^1 x_2^2\,\mathrm{d}x_2 + x_3\int_0^1 x_1\,\mathrm{d}x_1 \int_0^1 x_2\,\mathrm{d}x_2 \end{pmatrix}$$

$$\left(\frac{2}{3}, \frac{2+3x_3}{3+6x_3}\right)^T$$

b) Analog zu a)

$$\mathrm{E}[(X_1^2,X_2^2)^T|X_3=x_3] = \int_0^1\int_0^1 (X_1,X_2)^T \frac{4x_1x_2 + 4x_1x_3}{1+2x_3}\,\mathrm{d}x_1\,\mathrm{d}x_2 = \cdots = \begin{pmatrix} \frac{1}{2} \\ \frac{3+4x_3}{6+12x_3} \end{pmatrix}$$

$$f_{(X_1,X_2)^T}(x_1,x_2|X_3=x_3) = 2x_1 \mathbb{1}_{[0,1]}(x_1)\frac{x_2+x_3}{0.5+x_3}\mathbb{1}_{[0,1]}(x_2)$$

$$\underbrace{\Longrightarrow}_{\text{Aufgabe 3}} \mathrm{E}(X_1\cdot X_2|X_3=x_3) = \mathrm{E}(X_1|X_3=x_3)\mathrm{E}(X_2|X_3=x_3)$$

$$\Longrightarrow \mathrm{Var}[(X_1,X_2)^T|X_3=x_3] = \begin{pmatrix} \frac{1}{2}-(\frac{2}{3})^2 & 0 \\ 0 & \frac{3+4x_3}{6+12x_3}-(\frac{2+3x_3}{3+6x_3})^2 \end{pmatrix}$$

$$= \begin{pmatrix} \frac{1}{18} & 0 \\ 0 & \frac{1+6x_3+6x_3^2}{18(1+2x^3)^2} \end{pmatrix}$$

Ergänzung außerhalb der Vorlesung: Eine quadratische Matrix A heisst *regulär*, falls ihre Determinante von 0 verschieden ist, also $detA \neq 0$. Andernfalls heißt A singulär, nämlich falls $\det A = 0$.

Für $A \in \mathbb{R}^{m\times n}$ gilt: A regulär $\iff$ A hat eine Inverse A^{-1} $\iff$ Für jedes $b \in \mathbb{R}^n$ existiert genau ein $x \in \mathbb{R}^n$ mit $Ax = b$ $\iff$ $\{Ax:\ x \in \mathbb{R}^n\} = \mathbb{R}^n$ $\iff$ $Ax = 0,\ x \in \mathbb{R}^n$, gilt nur für $x = 0$ $\iff$ A^T ist regulär

Für jede reguläre Matrix A gilt $(A^T)^{-1} = (A^{-1})^T$.

Sind $A, B \in \mathbb{R}^{n\times m}$ regulär, dann ist AB auch regulär und $(AB)^{-1} = B^{-1}A^{-1}$.

Definition 2 Eine symmetrische Matrix $A \in \mathbb{R}^{p\times p}$ heisst *positiv definit* (Schreibweise in Zukunft $A > 0$), falls $x^TAx > 0$ für alle $x \in \mathbb{R}^p\backslash\{0\}$. A heisst *nicht negativ definit*, falls $x^TAx \geqslant 0$ für alle $x \in \mathbb{R}^p$ (in Zukunft $A \geqslant 0$).

Bemerkung 3 a) Jede Kovarianzmatrix ist symmetrisch und nicht-negativ definit (Lemma 1.7 f)). b) Ist X ein p-dimensionaler Zufallsvektor und ist $\mathrm{Var}(X)$ nicht positiv definit, dann existiert $a \in \mathbb{R}^p,\ a \neq 0$ mit $a^T\mathrm{Var}(X)a = 0$.

$$\underbrace{\Longrightarrow}_{\text{Lemma 1.7c}} \mathrm{Var}(a^TX) = 0 \Longrightarrow P((a^TX) = a^T\mathrm{E}(X)) = 1$$

Das heisst X liegt mit Wahrscheinlichkeit 1 in einer Hyperebene.

c) Aus $A \geqslant 0$ und $A \neq 0$ folgt nicht $A > 0$ (Lemma 1.7f)). Zum Beispiel gilt für $A = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}$, $A \neq 0$, $(x_1,x_2)A(x_1,x_2)^T = (x_1+x_2)^2 \geqslant 0$ für alle $x_1,x_2 \in \mathbb{R}$. Aber $(1,-1)A(1,-1)^T = 0$, d.h. "$A > 0$" gilt nicht.

d) Für jede symmetrische nicht-negativ definite Matrix $A \in \mathbb{R}^{p\times p}$ gilt $A > 0 \iff A$ ist regulär.

"$\Longrightarrow$" $x^TAx > 0$ für alle $x \neq 0 \Longrightarrow Ax \neq 0$ für alle $x \neq 0 \Longrightarrow A$ regulär.

"$\Longleftarrow$" Sei A regulär und für $x \in \mathbb{R}^p$ gelte $x^TAx = 0$. Zeige $x = 0$. Für $y \in \mathbb{R}^p$ gelte $p_y(t) = (x+ty)‘A(x+ty),\ t \in \mathbb{R}$. $\Longrightarrow$ $p_y(0) = 0,\ p_y(t) \geqslant 0$ für alle $t \in \mathbb{R}$ $\Longrightarrow$ $0 = p_y‘(0) = y^TAx + x^TAy = 2y^TAx$ also $y^TAx = 0$ für jedes $y \in \mathbb{R}^p$. $\Longrightarrow$ $Ax = 0 \underbrace{\Longrightarrow}_{A \text{ regulär}} x = 0$ ■

$\mu \in \mathbb{R}^p,\ \Sigma \in \mathbb{R}^{p\times p}, \Sigma > 0$ Ein p-dimensionaler Zufallsvektor X heisst p-dimensional normalverteilt, falls X folgende Dichte hat:

$$f_X(x) = \frac{1}{(2\pi)^{\frac{p}{2}}\sqrt{\det\Sigma}} \exp\left(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right),\ x \in \mathbb{R}^p$$

$$X \sim N_p(\mu, \Sigma)$$