Multivariate Datenanalyse

MSc Psychologie WiSe 2021/22

Prof. Dr. Dirk Ostwald

(3) Wahrscheinlichkeitstheorie

Realisierungen von Zufallsvariablen

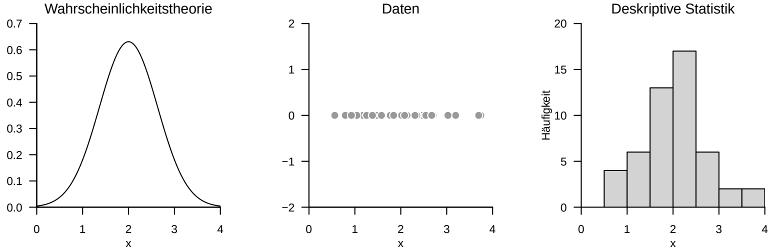
```
# Univariate Normalverteilungsparameter
mu      = 2.0                       # Erwartungswertparameter
sigsqr  = 0.5                       # Varianzparameter
n       = 10                        # Anzahl von u.i.v. Realisierungen

# 10 Realisierungen
X       = rnorm(n,mu,sqrt(sigsqr))  # X_i \sim N(\mu,\sigma^2), i = 1,...,n
print(X)
```

```
>  [1] 1.010 2.181 0.277 1.996 2.440 2.812 0.712 1.825 1.827 1.800
```

```
# 10 Realisierungen
X       = rnorm(n,mu,sqrt(sigsqr))  # X_i \sim N(\mu,\sigma^2), i = 1,...,n
print(X)
```

```
>  [1] 1.608 2.445 3.460 0.847 2.362 0.683 1.631 1.963 2.384 1.354
```

Wahrscheinlichkeitstheorie, Daten, Deskriptive Statistik



Realisierungen von Zufallsvektoren

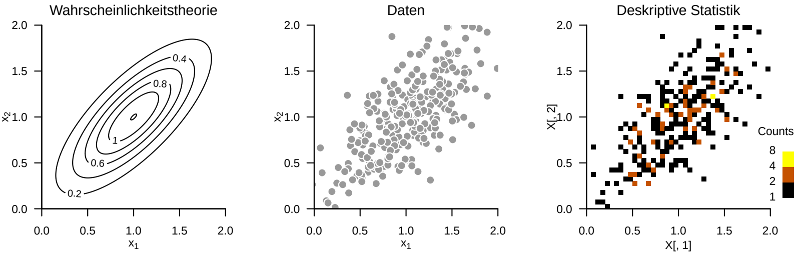
```r
# R Paket für multivariate Normalverteilungsrealisierung
library(MASS)

# Multivariate Normalverteilungsparameter
mu       = c(2.0,5.0)                      # Erwartungswertparameter
Sigma    = matrix(c(0.5,0.1,               # Kovarianzmatrixparamter
                    0.1,0.5),
                  nrow = 2,
                  byrow = TRUE)
n        = 10                              # Anzahl von u.i.v. Realisierungen

# 10 Realisierungen
X        = t(mvrnorm(n,mu,Sigma))          # X_i \sim N(\mu,\Sigma), i = 1,...,n
print(X)
```

```
>      [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10]
> [1,] 1.84 2.12 1.18 1.68 2.98 2.01 2.34 2.12 1.69 0.694
> [2,] 5.67 5.28 4.39 6.13 6.08 4.89 3.63 4.87 4.41 4.649
```

```r
# 10 Realisierungen
X        = t(mvrnorm(n,mu,Sigma))          # X_i \sim N(\mu,\sigma^2), i = 1,...,n
print(X)
```

```
>      [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10]
> [1,] 1.65 1.76 2.06 1.85 2.23 3.26 3.04 1.12 2.23 0.857
> [2,] 5.42 4.54 4.88 4.87 5.26 6.76 4.01 6.52 4.90 4.048
```

Multivariate Wahrscheinlichkeitstheorie, Multivariate Daten, Multivariate Deskriptive Statistik



Zufallsvektoren und multivariate Verteilungen

Marginale und bedingte Verteilungen

Erwartungswerte und Kovarianzmatrizen

Multivariate Normalverteilungen

Selbstkontrollfragen

Zufallsvektoren und multivariate Verteilungen

Marginale und bedingte Verteilungen

Erwartungswerte und Kovarianzmatrizen

Multivariate Normalverteilungen

Selbstkontrollfragen

Definition (Zufallsvektor)

$(\Omega, \mathcal{A}, \mathbb{P})$ sei ein Wahrscheinlichkeitsraum und $(\mathcal{X}, \mathcal{S})$ sei ein m-dimensionaler Messraum. Ein m-dimensionaler *Zufallsvektor* ist definiert als eine Abbildung

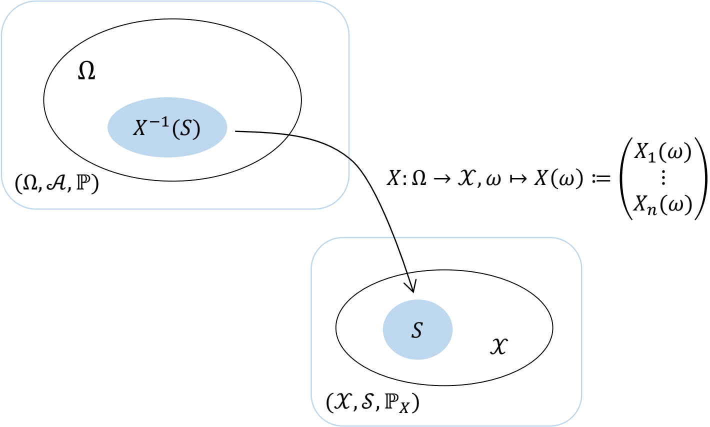
$$X : \Omega \to \mathcal{X}, \omega \mapsto X(\omega) := \begin{pmatrix} X_1(\omega) \\ \vdots \\ X_m(\omega) \end{pmatrix} \tag{1}$$

mit der *Messbarkeitseigenschaft*

$$\{\omega \in \Omega | X(\omega) \in S\} \in \mathcal{A} \text{ für alle } S \in \mathcal{S}. \tag{2}$$

Bemerkungen

- X ist messbar, wenn die Komponentenfunktionen $X_1, ..., X_m$ messbar sind.
- Die Komponentenfunktionen eines Zufallsvektors sind Zufallsvariablen.
- Ein m-dimensionaler Zufallsvektor ist die Konkatenation von m Zufallsvariablen.
- Für einen Zufallsvektor schreiben wir auch häufig $X := (X_1, ..., X_m)$.
- Für $m := 1$ ist ein Zufallsvektor eine Zufallsvariable.



$$\mathbb{P}\left(X^{-1}(S)\right) = \mathbb{P}(\{\omega \in \Omega | X(\omega) \in S\}) =: \mathbb{P}_X(S)$$

Definition (Multivariate Verteilung)

$(\Omega, \mathcal{A}, \mathbb{P})$ sei ein Wahrscheinlichkeitsraum, $(\mathcal{X}, \mathcal{S})$ sei ein m-dimensionaler Messraum und

$$X : \Omega \to \mathcal{X}, \omega \mapsto X(\omega) \tag{3}$$

sei ein Zufallsvektor. Dann heißt das Wahrscheinlichkeitsmaß $\mathbb{P}_X$, definiert durch

$$\mathbb{P}_X : \mathcal{S} \to [0,1], S \mapsto \mathbb{P}_X(S) := \mathbb{P}(X^{-1}(S)) = \mathbb{P}\left(\{\omega \in \Omega | X(\omega) \in S\}\right) \tag{4}$$

die *multivariate Verteilung des Zufallsvektor* X.

Bemerkungen

- Der Einfachheit halber spricht man oft auch nur von "der Verteilung des Zufallsvektors X".
- Die Notationskonventionen für Zufallsvariablen gelten für Zufallsvektoren analog, z.B.

$$\begin{aligned}
\mathbb{P}_X(X \in S) &:= \mathbb{P}\left(\{X \in S\}\right) = \mathbb{P}\left(\{\omega \in \Omega | X(\omega) \in S\}\right) \\
\mathbb{P}_X(X = x) &:= \mathbb{P}\left(\{X = x\}\right) = \mathbb{P}\left(\{\omega \in \Omega | X(\omega) = x\}\right) \\
\mathbb{P}_X(X \le x) &:= \mathbb{P}\left(\{X \le x\}\right) = \mathbb{P}\left(\{\omega \in \Omega | X(\omega) \le x\}\right) \\
\mathbb{P}_X(x_1 \le X \le x_2) &:= \mathbb{P}\left(\{x_1 \le X \le x_2\}\right) = \mathbb{P}\left(\{\omega \in \Omega | x_1 \le X(\omega) \le x_2\}\right)
\end{aligned} \tag{5}$$

- Relationsoperatoren wie $\le$ werden hier *komponentenweise* verstanden.
- Zum Beispiel heißt $x \le y$ für $x, y \in \mathbb{R}^m$, dass $x_i \le y_i$ für alle $i = 1, ..., m$.

Definition (Multivariate kumulative Verteilungsfunktionen)

X sei ein Zufallsvektor mit Ergebnisraum $\mathcal{X}$. Dann heißt eine Funktion der Form

$$P_X : \mathcal{X} \to [0,1],\, x \mapsto P_X(x) := \mathbb{P}_X(X \le x) \tag{6}$$

multivariate kumulative Verteilungsfunktion von X.

Bemerkung

- Multivariate kumulative Verteilungsfunktionen können zur Definition von multivariaten Verteilungen genutzt werden, häufiger ist allerdings die Definition multivariater Verteilungen durch multivariate Wahrscheinlichkeitsmasse- oder Wahrscheinlichkeitsdichtefunktionen.

Definition (Diskreter Zufallsvektor, multivariate WMF)

$(\Omega, \mathcal{A}, \mathbb{P})$ sei ein Wahrscheinlichkeitsraum und $X : \Omega \to \mathcal{X}$ ein Zufallsvektor. X heißt *diskreter Zufallsvektor*, wenn der Ergebnisraum $\mathcal{X}$ endlich viele oder höchsten abzählbar viele Elemente $x_i, i = 1, 2, ...$ enthält. Die *multivariate Wahrscheinlichkeitsmassefunktion (WMF)* eines diskreten Zufallsvektors X wird mit p_X bezeichnet und ist definiert durch

$$p_X : \mathcal{X} \to [0, 1], x \mapsto p_X(x) := \mathbb{P}_X(X = x). \tag{7}$$

Bemerkungen

- Der Begriff der multivariaten WMF ist analog zum Begriff der WMF.
- Man spricht oft einfach von der WMF eines Zufallsvektors.
- Wie univariate WMFen sind multivariate WMFen nicht-negativ und normiert.

Beispiel (Multivariate Wahrscheinlichkeitsmassefunktion)

Wir betrachten einen zweidimensionalen Zufallsvektor $X := (X_1, X_2)$ der Werte in $\mathcal{X} := \mathcal{X}_1 \times \mathcal{X}_2$ annimmt, wobei $\mathcal{X}_1 := \{1, 2, 3\}$ und $\mathcal{X}_2 = \{1, 2, 3, 4\}$ seien.

Eine exemplarische bivariate WMF der Form

$$p_X : \{1, 2, 3\} \times \{1, 2, 3, 4\} \to [0, 1], (x_1, x_2) \mapsto p_X(x_1, x_2) \tag{8}$$

ist dann durch nachfolgende Tabelle definiert.

$p_X(x_1, x_2)$	$x_2 = 1$	$x_2 = 2$	$x_2 = 3$	$x_2 = 4$
$x_1 = 1$	0.1	0.0	0.2	0.1
$x_1 = 2$	0.1	0.2	0.0	0.0
$x_1 = 3$	0.0	0.1	0.1	0.1

Man beachte, dass $\sum_{x_1=1}^{3} \sum_{x_2=1}^{4} p_X(x_1, x_2) = 1$.

Definition (Kontinuierlicher Zufallsvektor, multivariate WDF)

Es sei $(\Omega, \mathcal{A}, \mathbb{P})$ ein Wahrscheinlichkeitsraum. Ein Zufallsvektor der Form $X : \Omega \to \mathbb{R}^m$ heißt *kontinuierlicher* Zufallsvektor. Die *multivariate Wahrscheinlichkeitsdichtefunktion (WDF)* eines kontinuierlichen Zufallsvektors X ist eine Funktion

$$p_X : \mathbb{R}^m \to \mathbb{R}_{\geq 0}, x \mapsto p_X(x), \tag{9}$$

mit den Eigenschaften

(1) $\int_{\mathbb{R}^m} p_X(x)\,dx = 1$

(2) $\mathbb{P}_X(x_1 \leq X \leq x_2) = \int_{x_{1_1}}^{x_{2_1}} \cdots \int_{x_{1_m}}^{x_{2_m}} p_X(s_1, ..., s_m)\,ds_1 \cdots ds_m$

Bemerkungen

- Der Begriff der multivariaten WDF ist analog zum Begriff der WDF.
- Man spricht häufig auch einfach von der WDF eines Zufallsvektors
- Wie univariate WDFen sind multivariate WDFen nicht-negativ und normiert.
- Wie für kontinuierliche Zufallsvariablen gilt für kontinuierliche Zufallsvektoren

$$\mathbb{P}_X(X = x) = \mathbb{P}_X(x \leq X \leq x) = \int_{x_1}^{x_1} \cdots \int_{x_m}^{x_m} p_X(s_1, ..., s_m)\,ds_1 \cdots ds_m = 0 \tag{10}$$

Zufallsvektoren und multivariate Verteilungen

Marginale und bedingte Verteilungen

Erwartungswerte und Kovarianzmatrizen

Multivariate Normalverteilungen

Selbstkontrollfragen

Definition (Univariate Marginalverteilung)

$(\Omega, \mathcal{A}, \mathbb{P})$ sei ein Wahrscheinlichkeitsraum, $(\mathcal{X}, \mathcal{S})$ sei ein m-dimensionaler Messraum, $X : \Omega \to \mathcal{X}$ sei ein Zufallsvektor, $\mathbb{P}_X$ sei die Verteilung von X, $\mathcal{X}_i \subset \mathcal{X}$ sei der Ergebnisraum der $i$ten Komponente X_i von X, und $\mathcal{S}_i$ sei eine σ-Algebra auf X_i. Dann heißt die durch

$$\mathbb{P}_{X_i} : \mathcal{S}_i \to [0,1], S \mapsto \mathbb{P}_X \left(\mathcal{X}_1 \times \cdots \times \mathcal{X}_{i-1} \times S \times \mathcal{X}_{i+1} \times \cdots \times \mathcal{X}_m\right) \text{ für } S \in \mathcal{S}_i \quad (11)$$

definierte Verteilung die *$i$te univariate Marginalverteilung von X*.

Bemerkungen

- Univariate Marginalverteilungen sind die Verteilungen der Komponenten eines Zufallsvektors.
- Univariate Marginalverteilungen sind Verteilungen von Zufallsvariablen.
- Die Festlegung der multivariaten Verteilung von X legt auch die Verteilungen der X_i fest.

Theorem (Marginale Wahrscheinlichkeitsmasse- und dichtefunktionen)

(1) $X = (X_1, ..., X_m)$ sei ein m-dimensionaler diskreter Zufallsvektor mit Wahrscheinlichkeitsmassefunktion p_X und Komponentenergebnisräumen $\mathcal{X}_1, ..., \mathcal{X}_m$. Dann ergibt sich die Wahrscheinlichkeitsmassefunktion der $i$ten Komponente X_i von X als

$$p_{X_i} : \mathcal{X}_i \to [0,1], x_i \mapsto p_{X_i}(x_i) := \sum_{x_1} \cdots \sum_{x_{i-1}} \sum_{x_{i+1}} \cdots \sum_{x_m} p_X(x_1, ..., x_{i-1}, x_i, x_{i+1}, ..., x_m) \tag{12}$$

(2) $X = (X_1, ..., X_m)$ sei ein m-dimensionaler kontinuierlicher Zufallsvektor mit Wahrscheinlichkeitsdichtefunktion p_X und Komponentenergebnisraum $\mathbb{R}$. Dann ergibt sich die Wahrscheinlichkeitsdichtefunktion der $i$ten Komponente X_i von X als

$$p_{X_i} : \mathbb{R} \to \mathbb{R}_{\geq 0}, x_i \mapsto p_{X_i}(x_i) := \int_{x_1} \cdots \int_{x_{i-1}} \int_{x_{i+1}} \cdots \int_{x_m} p_X(x_1, ..., x_{i-1}, x_i, x_{i+1}, ..., x_m)\, dx_1 \ldots dx_{i-1}\, dx_{i+1} \ldots dx_m. \tag{13}$$

Bemerkungen

- Wir verzichten auf einen Beweis.
- Die WMFen der univariaten Marginalverteilungen diskreter Zufallsvektoren ergeben sich durch Summation.
- Die WDFen der univariaten Marginalverteilungen kontinuierlicher Zufallsvektoren ergeben sich durch Integration.

Marginale und bedingte Verteilungen

Beispiel (Marginale Wahrscheinlichkeitsmassefunktionen)

Wir betrachten erneut den zweidimensionalen Zufallsvektor $X := (X_1, X_2)$ der Werte in $\mathcal{X} := \mathcal{X}_1 \times \mathcal{X}_2$ annimmt, wobei $\mathcal{X}_1 := \{1, 2, 3\}$ und $\mathcal{X}_2 = \{1, 2, 3, 4\}$ seien.

Basierend auf der oben definierten WMF ergeben sich folgende marginalen WMFen p_{X_1} und p_{X_2}

$p_X(x_1, x_2)$	$x_2 = 1$	$x_2 = 2$	$x_2 = 3$	$x_2 = 4$	$p_{X_1}(x_1)$
$x_1 = 1$	0.1	0.0	0.2	0.1	0.4
$x_1 = 2$	0.1	0.2	0.0	0.0	0.3
$x_1 = 3$	0.0	0.1	0.1	0.1	0.3
$p_{X_2}(x_2)$	0.2	0.3	0.3	0.2	

Man beachte, dass $\sum_{x_1=1}^{3} p_{X_1}(x_1) = 1$ und $\sum_{x_2=1}^{3} p_{X_2}(x_2) = 1$ gilt.

Marginale und bedingte Verteilungen

Vorbemerkungen

Wir erinnern uns, dass für einen Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mathbb{P})$ und zwei Ereignisse $A, B \in \mathcal{A}$ mit $\mathbb{P}(B) > 0$ die bedingte Wahrscheinlichkeit von A gegeben B definiert ist als

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}. \tag{14}$$

Analog wird für zwei Zufallsvariablen X_1, X_2 mit Ereignisräumen $\mathcal{X}_1, \mathcal{X}_2$ und (messbaren) Mengen $S_1 \in \mathcal{X}_1, S_2 \in \mathcal{X}_2$ die bedingte Verteilung von X_1 gegeben X_2 mithilfe der Ereignisse $A := \{X_1 \in S_1\}$ und $B := \{X_2 \in S_2\}$ definiert.

So ergibt sich zum Beispiel die bedingte Wahrscheinlichkeit, dass $X_1 \in S_1$ gegeben dass $X_2 \in S_2$ unter der Annahme, dass $\mathbb{P}(\{X_2 \in S_2\}) > 0$, zu

$$\mathbb{P}(\{X_1 \in S_1\}|\{X_2 \in S_2\}) = \frac{\mathbb{P}(\{X_1 \in S_1\} \cap \{X_2 \in S_2\})}{\mathbb{P}(\{X_2 \in S_2\})}. \tag{15}$$

In der Folge betrachten wir zunächst durch die WMFen/WDFen zweidimensionaler Zufallsvektoren definierte bedingte Verteilungen.

Marginale und bedingte Verteilungen

Definition (Bedingte WMF, diskrete bedingte Verteilung)

$X := (X_1, X_2)$ sei ein diskreter Zufallsvektor mit Ergebnisraum $\mathcal{X} := \mathcal{X}_1 \times \mathcal{X}_2$, WMF $p_X = p_{X_1,X_2}$ und marginalen WMFen p_{X_1} und p_{X_2}. Die bedingte WMF von X_1 gegeben $X_2 = x_2$ ist dann für $p_{X_2}(x_2) > 0$ definiert als

$$p_{X_1|X_2=x_2} : \mathcal{X}_1 \to [0,1], x_1 \mapsto p_{X_1|X_2=x_2}(x_1|x_2) := \frac{p_{X_1,X_2}(x_1,x_2)}{p_{X_2}(x_2)} \tag{16}$$

Analog ist für $p_{X_1}(x_1) > 0$ die bedingte WMF von X_2 gegeben $X_1 = x_1$ definiert als

$$p_{X_2|X_1=x_1} : \mathcal{X}_2 \to [0,1], x_2 \mapsto p_{X_2|X_1=x_2}(x_1|x_2) := \frac{p_{X_1,X_2}(x_1,x_2)}{p_{X_1}(x_1)} \tag{17}$$

Die bedingten Verteilungen mit WMFen $p_{X_1|X_2=x_2}$ und $p_{X_2|X_1=x_1}$ heißen dann die *diskreten bedingten Verteilungen von X_1 gegeben $X_2 = x_2$ und X_2 gegeben $X_1 = x_1$*, respektive.

Bemerkungen

- In Analogie zur Definition der bedingten Wahrscheinlichkeit von Ereignissen gilt also

$$p_{X_1|X_2}(x_1|x_2) = \frac{p_{X_1,X_2}(x_1,x_2)}{p_{X_2}(x_2)} = \frac{\mathbb{P}(\{X_1 = x_1\} \cap \{X_2 = x_2\})}{\mathbb{P}(\{X_2 = x_2\})}. \tag{18}$$

- Bedingte Verteilungen sind (lediglich) normalisierte gemeinsame Verteilungen.

Marginale und bedingte Verteilungen

Beispiel (Bedingte Wahrscheinlichkeitsmassefunktionen)

Wir betrachten erneut den zweidimensionalen Zufallsvektor $X := (X_1, X_2)$ der Werte in $\mathcal{X} := \mathcal{X}_1 \times \mathcal{X}_2$ annimmt, wobei $\mathcal{X}_1 := \{1, 2, 3\}$ und $\mathcal{X}_2 = \{1, 2, 3, 4\}$ seien.

Basierend auf der oben definierten WMF und den entsprechenden oben evaluierten marginalen WMFen ergeben sich folgende bedingte WMFen für $p_{X_2|X_1=x_1}$

$p_{X_1\|X_2}(x_1, x_2)$	$x_2 = 1$	$x_2 = 2$	$x_2 = 3$	$x_2 = 4$
$p_{X_2\|X_1=1}(x_2\|x_1 = 1)$	$\frac{0.1}{0.4} = 0.25$	$\frac{0.0}{0.4} = 0.00$	$\frac{0.2}{0.4} = 0.50$	$\frac{0.1}{0.4} = 0.25$
$p_{X_2\|X_1=2}(x_2\|x_1 = 2)$	$\frac{0.1}{0.3} = 0.3\bar{3}$	$\frac{0.2}{0.3} = 0.6\bar{6}$	$\frac{0.0}{0.3} = 0.00$	$\frac{0.0}{0.3} = 0.00$
$p_{X_2\|X_1=3}(x_2\|x_1 = 3)$	$\frac{0.0}{0.3} = 0.00$	$\frac{0.1}{0.3} = 0.3\bar{3}$	$\frac{0.1}{0.3} = 0.3\bar{3}$	$\frac{0.1}{0.3} = 0.3\bar{3}$

Bemerkungen

- Man beachte, dass $\sum_{x_2=1}^{4} p_{X_2|X_1=x_1}(x_2|x_1) = 1$ für alle $x_1 \in \mathcal{X}_1$.
- Man beachte die qualitative Ähnlichkeit der WMFen $p_{X_1,X_2}(x_1, x_2)$ und $p_{X_2|X_1}(x_2|x_1)$.
- Bedingte Verteilungen sind (lediglich) normalisierte gemeinsame Verteilungen.

Definition (Bedingte WDF, kontinuierliche bedingte Verteilungen)

$X := (X_1, X_2)$ sei ein kontinuierlicher Zufallsvektor mit Ergebnisraum $\mathbb{R}^2$, WDF $p_X = p_{X_1,X_2}$ und marginalen WDFen p_{X_1} und p_{X_2}. Die bedingte WDF von X_1 gegeben $X_2 = x_2$ ist dann für $p_{X_2}(x_2) > 0$ definiert als

$$p_{X_1|X_2=x_2} : \mathbb{R} \to \mathbb{R}_{\geq 0}, x_1 \mapsto p_{X_1|X_2=x_2}(x_1|x_2) := \frac{p_{X_1,X_2}(x_1, x_2)}{p_{X_2}(x_2)} \tag{19}$$

Analog ist für $p_{X_1}(x_1) > 0$ die bedingte WMF von X_2 gegeben $X_1 = x_1$ definiert als

$$p_{X_2|X_1=x_1} : \mathbb{R} \to \mathbb{R}_{\geq 0}, x_2 \mapsto p_{X_2|X_1=x_2}(x_2|x_1) := \frac{p_{X_1,X_2}(x_1, x_2)}{p_{X_1}(x_1)} \tag{20}$$

Die Verteilungen mit WDFen $p_{X_1|X_2=x_2}$ und $p_{X_2|X_1=x_1}$ heißen dann die *kontinuierlichen bedingten Verteilungen von X_1 gegeben $X_2 = x_2$ und X_2 gegeben $X_1 = x_1$*, respektive.

Bemerkung

- Im kontinuierlichen Fall gilt zwar $\mathbb{P}(X = x) = 0$, aber nicht notwendig auch $p_X(x) = 0$.

Definition (Erwartungswert)

X sei ein m-dimensionaler Zufallvektor. Dann ist der *Erwartungwert* von X definiert als der m-dimensionale Vektor

$$\mathbb{E}(X) := \begin{pmatrix} \mathbb{E}(X_1) \\ \vdots \\ \mathbb{E}(X_m) \end{pmatrix}. \tag{21}$$

Bemerkungen

- Der Erwartungswert von X ist der Vektor der Erwartungswerte $\mathbb{E}(X_1), ..., \mathbb{E}(X_m)$.
- Im allgemeinen linearen Modell $X = D\beta + \varepsilon$ gilt zum Beispiel $\mathbb{E}(\varepsilon) = 0_m$ und $\mathbb{E}(X) = D\beta$.

Definition (Kovarianzmatrix)

X sei ein m-dimensionaler Zufallsvektor. Die *Kovarianzmatrix* von X ist definiert als die $m \times m$ matrix

$$\mathbb{C}(X) := \mathbb{E}\left((X - \mathbb{E}(X))(X - \mathbb{E}(X))^T\right) = \begin{pmatrix} \mathbb{C}(X_1, X_1) & \cdots & \mathbb{C}(X_1, X_m) \\ \vdots & \ddots & \vdots \\ \mathbb{C}(X_m, X_1) & \cdots & \mathbb{C}(X_m, X_m) \end{pmatrix}. \quad (22)$$

Bemerkungen

- Die Kovarianzmatrix $\mathbb{C}(X)$ ist die Matrix der Kovarianzen $\mathbb{C}(X_i, X_j), i, j = 1, ..., m.$
- Die Korrelation von X_i and X_j ist definiert als

$$\rho_{ij} = \frac{\mathbb{C}(X_i, X_j)}{\sqrt{\mathbb{C}(X_i, X_i)}\sqrt{\mathbb{C}(X_j, X_j)}} \in [-1, 1]. \quad (23)$$

- Die Kovarianzmatrix repräsentiert also die Korrelationsstruktur der Zufallsvariablen $X_1, ..., X_m$
- Im ALM mit sphärischer Kovarianmatrix gilt per Definition $\mathbb{C}(\varepsilon) = \mathbb{C}(X) = \sigma^2 I_m.$

Erwartungswerte und Kovarianzmatrizen

Die Äquivalenz der Kovarianzschreibweisen folgt mit

$$
\begin{aligned}
&\mathbb{E}\left((X-\mathbb{E}(X))(X-\mathbb{E}(X))^T\right) \\
&= \mathbb{E}\left(\left(\begin{pmatrix} X_1 \\ \vdots \\ X_m \end{pmatrix} - \begin{pmatrix} \mathbb{E}(X_1) \\ \vdots \\ \mathbb{E}(X_m) \end{pmatrix}\right)\left(\begin{pmatrix} X_1 \\ \vdots \\ X_m \end{pmatrix} - \begin{pmatrix} \mathbb{E}(X_1) \\ \vdots \\ \mathbb{E}(X_m) \end{pmatrix}\right)^T\right) \\
&= \mathbb{E}\left(\begin{pmatrix} X_1-\mathbb{E}(X_1) \\ \vdots \\ X_m-\mathbb{E}(X_m) \end{pmatrix}\begin{pmatrix} X_1-\mathbb{E}(X_1) \\ \vdots \\ X_m-\mathbb{E}(X_m) \end{pmatrix}^T\right) \\
&= \mathbb{E}\left(\begin{pmatrix} X_1-\mathbb{E}(X_1) \\ \vdots \\ X_m-\mathbb{E}(X_m) \end{pmatrix}\begin{pmatrix} X_1-\mathbb{E}(X_1) & \dots & X_m-\mathbb{E}(X_m) \end{pmatrix}\right) \\
&= \mathbb{E}\begin{pmatrix} (X_1-\mathbb{E}(X_1))(X_1-\mathbb{E}(X_1)) & \dots & (X_1-\mathbb{E}(X_1))(X_m-\mathbb{E}(X_m) \\ \vdots & \ddots & \vdots \\ (X_m-\mathbb{E}(X_m))(X_1-\mathbb{E}(X_1)) & \dots & (X_m-\mathbb{E}(X_m))(X_m-\mathbb{E}(X_m)) \end{pmatrix} \\
&= \left(\mathbb{E}\left((X_i-\mathbb{E}(X_i))(X_j-\mathbb{E}(X_j))\right)\right)_{1\le i,j\le m} \\
&=: \left(\mathbb{C}(X_i, X_j)\right)_{1\le i,j\le m} \\
&=: \mathbb{C}(X).
\end{aligned}
$$

Definition (Stichprobenmittel, Stichprobenkovarianmatrix)

$X^{(1)}, ..., X^{(n)}$ seien n m-dimensionale Zufallsvektoren. Dann ist das Stichprobenmittel der $X^{(1)}, ..., X^{(n)}$ definiert als

$$\bar{X} := \frac{1}{n}\sum_{i=1}^{n} X^{(i)} \tag{24}$$

und die Stichprobenkovarianzmatrix der $X^{(1)}, ..., X^{(m)}$ ist definiert als

$$C := \frac{1}{n-1}\sum_{i=1}^{n}(X^{(i)} - \bar{X})(X^{(i)} - \bar{X})^T. \tag{25}$$

Bemerkungen

- $\bar{X}$ ist ein unverzerrter Schätzer von $\mathbb{E}(X)$.
- C ist ein unverzerrter Schätzer von $\mathbb{C}(X)$.

Theorem (Datenmatrix und Stichprobenkovarianzmatrix)

$$Y := \begin{pmatrix} X^{(1)} & X^{(2)} & \cdots & X^{(n)} \end{pmatrix} \in \mathbb{R}^{m \times n} \tag{26}$$

sei eine $m \times n$ *Datenmatrix*, die durch die spaltenweise Konkatenation von n m-dimensionalen Zufallvektoren $X^{(1)}, ..., X^{(n)}$ gegeben sei. Dann kann das Stichprobenmittel berechnet werden als

$$\bar{Y} := \frac{1}{n} Y 1_n = \bar{X} \tag{27}$$

Weiterhin sei

$$Y_c := Y - \bar{Y} 1_n^T \in \mathbb{R}^{m \times n} \tag{28}$$

die $m \times n$ *zentrierte Datenmatrix*. Dann kann die Stichprobenkovarianzmatrix der $X^{(1)}, ..., X^{(n)}$ durch

$$C = \frac{1}{n-1} Y_c Y_c^T \text{ und } C = \frac{1}{n-1}\left(Y\left(I_n - \frac{1}{n} 1_{nn}\right) Y^T\right) \tag{29}$$

berechnet werden.

Bemerkungen

- $C = \frac{1}{n-1} Y_c Y_c^T$ ist in der Theorie der Hauptkomponentenanalyse von Bedeutung.
- $C = \frac{1}{n-1}\left(Y\left(I_n - \frac{1}{n} 1_{nn}\right) Y^T\right)$ erleichtert die numerische Implementation.

Beweis

Die Darstellung des Stichprobenmittels ergibt sich direkt aus

$$
\begin{aligned}
\bar{Y} &:= \frac{1}{n} Y 1_n \\
&= \frac{1}{n}\left(\begin{pmatrix} X_1^{(1)} & \cdots & X_1^{(n)} \\ \vdots & \ddots & \vdots \\ X_m^{(1)} & \cdots & X_m^{(n)} \end{pmatrix}\begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}\right) \\
&= \frac{1}{n}\begin{pmatrix} \sum_{i=1}^n X_1^{(i)} \\ \vdots \\ \sum_{i=1}^n X_m^{(i)} \end{pmatrix} \\
&= \frac{1}{n}\sum_{i=1}^n X^{(i)} =: \bar{X}
\end{aligned}
\tag{30}
$$

Beweis (fortgeführt)

Wir halten weiterhin zunächt fest, dass gilt

$$
\begin{aligned}
Y_c &= Y - \bar{X}1_n^T \\
&= \begin{pmatrix} X_1^{(1)} & \dots & X_1^{(n)} \\ \vdots & \ddots & \vdots \\ X_m^{(1)} & \dots & X_m^{(n)} \end{pmatrix} - \begin{pmatrix} \bar{X}_1 \\ \vdots \\ \bar{X}_m \end{pmatrix} \begin{pmatrix} 1 & \dots & 1 \end{pmatrix} \\
&= \begin{pmatrix} X_1^{(1)} & \dots & X_1^{(n)} \\ \vdots & \ddots & \vdots \\ X_m^{(1)} & \dots & X_m^{(n)} \end{pmatrix} - \begin{pmatrix} \bar{X}_1 & \dots & \bar{X}_1 \\ \vdots & \ddots & \vdots \\ \bar{X}_m & \dots & \bar{X}_m \end{pmatrix} \\
&= \begin{pmatrix} X_1^{(1)} - \bar{X}_1 & \dots & X_1^{(n)} - \bar{X}_1 \\ \vdots & \ddots & \vdots \\ X_m^{(1)} - \bar{X}_m & \dots & X_m^{(n)} - \bar{X}_m \end{pmatrix} \\
&= \begin{pmatrix} X^{(1)} - \bar{X} & \dots & X^{(n)} - \bar{X} \end{pmatrix}
\end{aligned}
$$

Beweis (fortgeführt)

Dann aber gilt

$$
\begin{aligned}
\frac{1}{n-1} Y_c Y_c^T &= \frac{1}{n-1} \left(X^{(1)} - \bar{X} \quad \dots \quad X^{(n)} - \bar{X} \right) \begin{pmatrix} \left(X^{(1)} - \bar{X} \right)^T \\ \vdots \\ \left(X^{(n)} - \bar{X} \right)^T \end{pmatrix} \\
&= \frac{1}{n-1} \sum_{i=1}^{n} \left(X^{(i)} - \bar{X} \right) \left(X^{(i)} - \bar{X} \right)^T \\
&= C.
\end{aligned}
$$

$\square$

Definition (Normalverteilte Zufallsvariable)

X sei eine Zufallsvariable mit Ergebnisraum $\mathbb{R}$ und WDF

$$p : \mathbb{R} \to \mathbb{R}_{>0}, x \mapsto p(x) := \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right). \tag{31}$$

Dann sagen wir, dass X einer *Normalverteilung (oder Gauß-Verteilung) mit Parametern $\mu \in \mathbb{R}$ und $\sigma^2 > 0$* unterliegt und nennen X eine *normalverteilte Zufallsvariable*. Wir kürzen dies mit $X \sim N\left(\mu, \sigma^2\right)$ ab. Die WDF einer normalverteilten Zufallsvariable bezeichnen wir mit

$$N\left(x; \mu, \sigma^2\right) := \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right). \tag{32}$$

Bemerkungen

- Der Parameter μ entspricht dem Wert höchster Wahrscheinlichkeitsdichte.
- Der Parameter σ^2 spezifiziert die Breite der WDF.

Definition (Multivariate Normalverteilung)

X sei ein n-dimensionaler Zufallsvektor mit Ergebnisraum $\mathbb{R}^n$ und WDF

$$p : \mathbb{R}^n \to \mathbb{R}_{>0},\, x \mapsto p(x) := (2\pi)^{-\frac{n}{2}} |\Sigma|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right). \quad (33)$$

Dann sagen wird, dass X einer *multivariaten (oder n-dimensionalen) Normalverteilung* mit *Erwartungswertparameter* $\mu \in \mathbb{R}^n$ und positive-definitem *Kovarianzmatrixparameter* $\Sigma \in \mathbb{R}^{n\times n}$ unterliegt und nennen X einen *(multivariat) normalverteilten Zufallsvektor*. Wir kürzen dies mit $X \sim N(\mu, \Sigma)$ ab. Die WDF eines multivariat normalverteilten Zufallsvektors bezeichnen wir mit

$$N(x;\mu,\Sigma) := (2\pi)^{-\frac{n}{2}} |\Sigma|^{-\frac{1}{2}} \exp\left(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right). \quad (34)$$

Bemerkungen

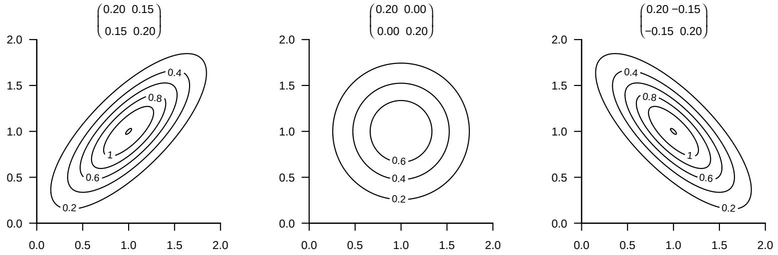
- Der Parameter $\mu \in \mathbb{R}^n$ entspricht dem Wert höchster Wahrscheinlichkeitsdichte
- Die Diagonalelemente von Σ spezifizieren die Breite der WDF bezüglich $X_1, ..., X_n$.
- Das i, jte Element von Σ spezifiziert die Kovarianz on X_i und X_j.
- Der Term $(2\pi)^{-\frac{n}{2}} |\Sigma|^{-\frac{1}{2}}$ ist die Normalsierungskonstante für den Exponentialfunktionsterm.

Visualisierung bivariater Normalverteilungsdichtefunktionen

```
# multivariate Normalverteilungstools
# install.packages("mvtnorm")
library(mvtnorm)


# Ergebnisraumdefintion
x_min  = 0                                          # x_i Minimum
x_max  = 2                                          # x_i Maxim
x_res  = 1e3                                        # x_i Auflösung
x_1    = seq(x_min, x_max, length.out = x_res)      # x_1 Raum
x_2    = seq(x_min, x_max, length.out = x_res)      # x_2 Raum
X      = expand.grid(x_1,x_2)                       # X = (x_1,x_2)^T Raum

# Parameterdefinition
mu     = c(1,1)                                     # \mu \in \mathbb{R}^2
S      = list(matrix(c(0.2,  0.15,  0.15, 0.2), 2), # \Sigma in \mathbb{R}^{2 \times 2}
              matrix(c(0.2,  0.00,  0.00, 0.2), 2), # \Sigma in \mathbb{R}^{2 \times 2}
              matrix(c(0.2, -0.15, -0.15, 0.2), 2)) # \Sigma in \mathbb{R}^{2 \times 2}

# Kovarianzparametervariantenschleife
for (Sigma in S){

  # Wahrscheinlichkeitsdichtefunktionauswertung
  p      = matrix(                                  # Matrixkonversion des von
                dmvnorm(as.matrix(X), mu, Sigma),   # dmvnorm() ausgegebenen Vektors
                nrow = x_res)

  # Visualisierung
  contour(
  x_1,
  x_2,
  p,
  xlim      =  c(x_min,x_max),
  ylim      =  c(x_min,x_max),
  nlevels   = 5)
}
```

Multivariate Normalverteilung

Visualisierung bivariater Normalverteilungsdichtefunktionen



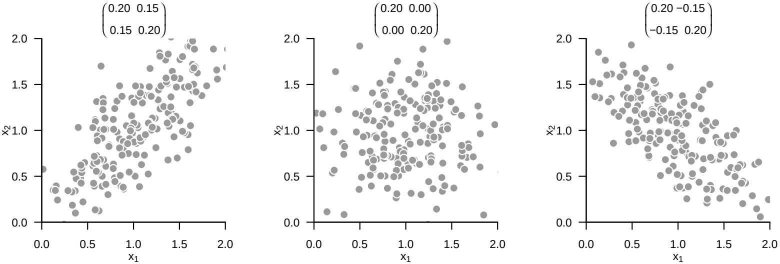
Multivariate Normalverteilung

Realisierung bivariater normalverteilter Zufallsvektoren

```
# R Paket für multivariate Normalverteilungen
library(mvtnorm)

# Parameterdefinition
mu     = c(1,1)                                   # \mu \in \mathbb{R}^2
Sigma  = matrix(c(0.2,  0.15,  0.15, 0.2), 2)  # \Sigma in \mathbb{R}^{2 \times 2}

# Zufallsvektorrealisierungen
rmvnorm(n = 10, mu, Sigma)
```

```
>        [,1]  [,2]
>  [1,] 1.316 0.927
>  [2,] 1.409 1.105
>  [3,] 0.340 0.386
>  [4,] 1.458 1.250
>  [5,] 0.886 0.397
>  [6,] 0.453 0.845
>  [7,] 1.589 1.343
>  [8,] 1.634 1.456
>  [9,] 1.227 1.067
> [10,] 1.054 0.995
```

Realisierung bivariater normalverteilter Zufallsvektoren



Theorem (Marginale Normalverteilungen)

Es sei $m := k + l$ und $X = (X_1, ..., X_m)$ sei ein m-dimensionaler normalverteilter Zufallsvektor mit Erwartungswertparameter

$$\mu = \begin{pmatrix} \mu_y \\ \mu_z \end{pmatrix} \in \mathbb{R}^m, \tag{35}$$

mit $\mu_y \in \mathbb{R}^k$ and $\mu_z \in \mathbb{R}^l$ und Kovarianzmatrixparameter

$$\Sigma = \begin{pmatrix} \Sigma_{yy} & \Sigma_{yz} \\ \Sigma_{zy} & \Sigma_{zz} \end{pmatrix} \in \mathbb{R}^{m \times m}, \tag{36}$$

mit $\Sigma_{yy} \in \mathbb{R}^{k \times k}$, $\Sigma_{yz} \in \mathbb{R}^{k \times l}$, $\Sigma_{zy} \in \mathbb{R}^{l \times k}$, und $\Sigma_{zz} \in \mathbb{R}^{l \times l}$. Dann sind $Y := (X_1, ..., X_k)$ und $Z := (X_{k+1}, ..., X_m)$ k- und l-dimensionale normalverteilte Zufallsvektoren, respektive, und es gilt

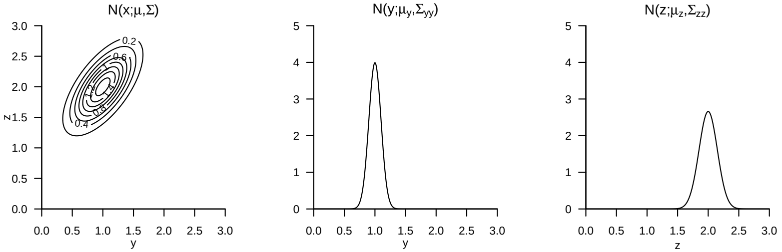
$$Y \sim N(\mu_y, \Sigma_{yy}) \text{ and } Z \sim N(\mu_z, \Sigma_{xx}), \tag{37}$$

Bemerkungen

- Wir verzichten auf einen Beweis
- Die Marginalverteilungen einer multivariaten Normalverteilung sind auch Normalverteilungen.
- Die Parameter der Marginalverteilungen ergeben sich aus den Parametern der gemeinsamen Verteilung.

Marginale Normalverteilungen

$$m := 2, k = 1, l = 1, \mu := \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \Sigma := \begin{pmatrix} 0.10 & 0.08 \\ 0.08 & 0.15 \end{pmatrix}$$



Theorem (Gemeinsame Normalverteilunge)

X sei ein m-dimensionaler normalverteilter Zufallsvektor mit WDF

$$p_X : \mathbb{R}^m \to \mathbb{R}_{>0},\ x \mapsto p_X(x) := N(x; \mu_x, \Sigma_{xx}) \text{ mit } \mu_x \in \mathbb{R}^m, \Sigma_{xx} \in \mathbb{R}^{m \times m}, \tag{38}$$

$A \in \mathbb{R}^{n \times m}$ sei eine Matrix, $b \in \mathbb{R}^n$ sei ein Vektor und Y sei ein n-dimensionaler bedingt normalverteilter Zufallsvektor mit bedingter WDF

$$p_{y|X}(\cdot|x) : \mathbb{R}^n \to \mathbb{R}_{>0},\ y \mapsto p_{Y|X}(y|x) := N(y; AX + b, \Sigma_{yy}) \text{ mit } \Sigma_{yy} \in \mathbb{R}^{n \times n}. \tag{39}$$

Dann ist der $m + n$-dimensionale Zufallsvektor (X, Y) normalverteilt mit (gemeinsamer) WDF

$$p_{X,Y} : \mathbb{R}^{m+n} \to \mathbb{R}_{>0},\ (x, y) \mapsto p_{X,Y}(x, y) = N\left((x, y); \mu_{x,y}, \Sigma_{x,y}\right), \tag{40}$$

mit $\mu_{x,y} \in \mathbb{R}^{n+m}$ and $\Sigma_{x,y} \in \mathbb{R}^{m+n \times m+n}$ und insbesondere

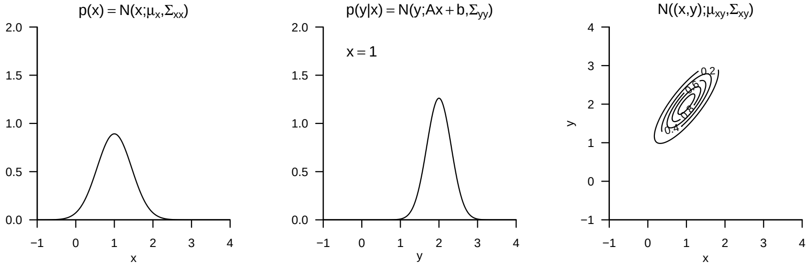
$$\mu_{x,y} = \begin{pmatrix} \mu_x \\ A\mu_x + b \end{pmatrix} \text{ und } \Sigma_{x,y} = \begin{pmatrix} \Sigma_{xx} & \Sigma_{xx}A^T \\ A\Sigma_{xx} & \Sigma_{yy} + A\Sigma_{xx}A^T \end{pmatrix}. \tag{41}$$

Bemerkungen

- Wir verzichten auf einen Beweis.
- Eine marginale und eine bedingte multivariate Normalverteilung induzieren eine gemeinsame Normalverteilung.
- Die Parameter der gemeinsamen Verteilungen ergeben sich als linear-affine Transformation der Parameter der induzierenden Verteilungen.

Gemeinsame Normalverteilungen

$$m := 1, n := 1, \mu_x := 1, \Sigma_{xx} := 0.2, A := 1, b := 1, \Sigma_{yy} := 0.1$$



Theorem (Bedingte Normalverteilungen)

(X, Y) sei ein $m + n$-dimensionaler normalverteilter Zufallsvektor mit WDF

$$p_{X,Y} : \mathbb{R}^{m+n} \to \mathbb{R}_{>0}, (x, y) \mapsto p_{X,Y}(x, y) := N\Big((x, y); \mu_{x,y}, \Sigma_{x,y}\Big), \tag{42}$$

mit

$$\mu_{x,y} = \begin{pmatrix} \mu_x \\ \mu_y \end{pmatrix}, \Sigma_{x,y} = \begin{pmatrix} \Sigma_{xx} & \Sigma_{xy} \\ \Sigma_{yx} & \Sigma_{yy} \end{pmatrix}, \tag{43}$$

mit $x, \mu_x \in \mathbb{R}^m, y, \mu_y \in \mathbb{R}^n$ and $\Sigma_{xx} \in \mathbb{R}^{m\times m}, \Sigma_{xy} \in \mathbb{R}^{m\times n}, \Sigma_{yy} \in \mathbb{R}^{n\times n}$. Dann ist die bedingte Verteilung von X gegeben Y eine m-dimensionale Normalverteilung mit bedingter WDF

$$p_{X|Y}(\cdot|y) : \mathbb{R}^m \to \mathbb{R}_{>0}, x \mapsto p_{X|Y}(x|y) := N(x; \mu_{x|y}, \Sigma_{x|y}) \tag{44}$$

mit

$$\mu_{x|y} = \mu_x + \Sigma_{xy}\Sigma_{yy}^{-1}(Y - \mu_y) \in \mathbb{R}^m \tag{45}$$

und

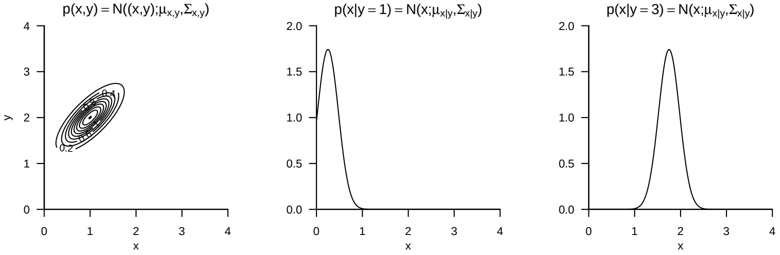
$$\Sigma_{x|y} = \Sigma_{xx} - \Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{yx} \in \mathbb{R}^{m\times m}. \tag{46}$$

Bemerkungen

- Wir verzichten auf einen Beweis.
- Die Parameter einer bedingten (multivariaten) Normalverteilung ergeben sich aus den Parametern einer gemeinsamen multivariaten Normalverteilung. Im Zusammenspiel mit den vorherigen Theoremen können die Parameter bedingter und marginale Normalverteilungen aus den Parametern der komplementären bedingten und marginalen Normalverteilungen bestimmt werden.

Bedingte Normalverteilungen

$$m := 2, n := 1, \mu := \begin{pmatrix} 1 \\ 2 \end{pmatrix}, \Sigma := \begin{pmatrix} 0.12 & 0.09 \\ 0.09 & 0.12 \end{pmatrix}$$



Zufallsvektoren und multivariate Verteilungen

Marginale und bedingte Verteilungen

Erwartungswerte und Kovarianzmatrizen

Multivariate Normalverteilungen

Selbstkontrollfragen

Selbstkontrollfragen

1. Definieren Sie den Begriff des Zufallsvektors.
2. Definieren Sie den Begriff der multivariaten Verteilung eines Zufallsvektors.
3. Definieren Sie den Begriff der multivariaten WMF.
4. Definieren Sie den Begriff der multivariaten WDF.
5. Definieren Sie den Begriff der univariaten Marginalverteilung eines Zufallsvektors.
6. Wie berechnet man die WMF der $i$ten Komponente eines diskreten Zufallsvektors?
7. Wie berechnet man die WDF der $i$ten Komponente eines kontinuierlichen Zufallsvektors?
8. Definieren Sie die Begriffe der bedingten WMF und der diskreten bedingten Verteilung.
9. Definieren Sie die Begriffe der bedingten WDF und der kontinuierlichen bedingten Verteilung.
10. Geben Sie die Definition des Erwartungswerts eines Zufallsvektors wieder.
11. Geben Sie die Definition der Kovarianzmatrix eines Zufallsvektors wieder.
12. Geben Sie die Definition des Stichprobenmittels und der Stichprobenkovarianzmatrix wieder.
13. Erläutern Sie, warum für eine Stichprobenkovarianzmatrix $C = \frac{1}{n-1} Y_c Y_c^T$ gilt.
14. Definieren Sie die WDF einer univariaten normalverteilten Zufallsvariable und erläutern Sie diese.
15. Definieren Sie die WDF eines multivariaten normalverteilten Zufallsvektors wieder und erläutern Sie diese.
16. Geben Sie das Theorem zu Marginalen Normalverteilungen wieder.
17. Geben Sie das Theorem zu Gemeinsamen Normalverteilungen wieder.
18. Geben Sie das Theorem zu Bedingten Normalverteilungen wieder.