Multivariate Verfahren

Multivariate Verteilungen

Annika Hoyer

Sommersemester 2020

Multivariate Verteilungen - Inhalt

Wiederholung: Eindimensionale Zufallsvariablen
- Verteilungsfunktion
- Wahrscheinlichkeitsfunktion, Dichte
- Kennzahlen von Zufallsvariablen

Multivariate Zufallsvariablen
- Definition, Gemeinsame Verteilung
- Erwartungswertvektor
- Kovarianzmatrix, Korrelationsmatrix

Multivariate Normalverteilung
- Beispiele

Weitere multivariate Verteilungen

Wiederholung: Eindimensionale Zufallsvariablen

Zufallsvariablen

In der Statistik werden Daten mit Hilfe von Zufallsvariablen modelliert:

- **Diskrete Zufallsvariablen** nehmen Werte in endlicher oder höchstens abzählbar unendlicher Menge an (endliche Menge, natürliche Zahlen, ganze Zahlen, ...)
- **Stetige Zufallsvariablen** nehmen Werte auf einem Intervall oder den gesamten reellen Zahlen an.
  Wegen des überabzählbaren Wertebereichs gilt $P(X = x) = 0$ für alle $x \in \mathbb{R}$, d.h. die Wahrscheinlichkeit, dass eine stetige Zufallsvariable X einen Wert x genau annimmt, ist Null.

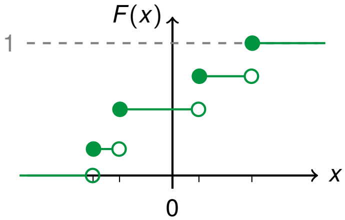
Verteilungsfunktion

Definition: Verteilungsfunktion

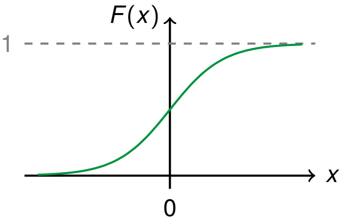
Für eine Zufallsvariable X heißt die Funktion

$$F(x) = P(X \leq x), \quad x \in \mathbb{R}$$

Verteilungsfunktion von X. Durch sie ist X eindeutig bestimmt.



Verteilungsfunktion einer diskreten Zufallsvariable

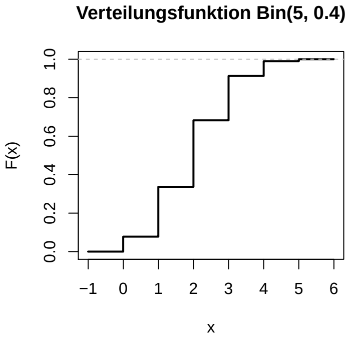


Verteilungsfunktion einer stetigen Zufallsvariable

Beispiel 1 – Diskrete Zufallsvariablen

Binomialverteilte Zufallsvariablen: $X \sim \text{Bin}(n, p)$

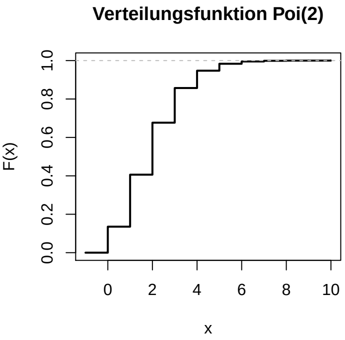
- Wertebereich: $0, 1 \ldots n$ (endliche Menge)
- Parameter: $n \in \mathbb{N}$, $p \in [0, 1]$
- Modellierung: Anzahl von Treffern (Münzwurf, Würfeln,...)



Beispiel 1 – Diskrete Zufallsvariablen

Poissonverteilte Zufallsvariablen: $X \sim \text{Poi}(\lambda)$

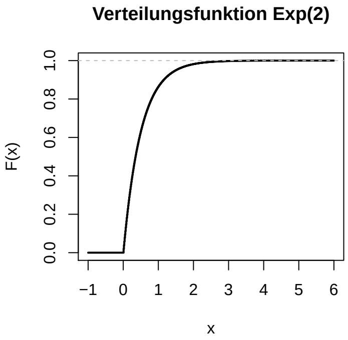
- Wertebereich: $0, 1, 2, \ldots$ (abzählbar unendliche Menge)
- Parameter: $\lambda > 0$
- Modellierung: Zählvorgänge (Krankheitsfälle, Druckfehler,...)



Beispiel 1 – Stetige Zufallsvariablen

Exponentialverteilte Zufallsvariablen: $X \sim \text{Exp}(\lambda)$

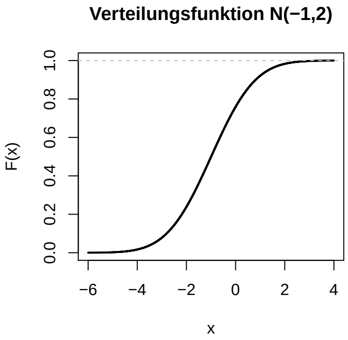
- Wertebereich: $[0, \infty)$ (Intervall)
- Parameter: $\lambda > 0$
- Modellierung: Zeitdauer (Wartezeit, Lebensdauer eines Produkts,...)



Beispiel 1 – Stetige Zufallsvariablen

Normalverteilte Zufallsvariablen: $X \sim \mathrm{N}(\mu, \sigma^2)$

- Wertebereich: $\mathbb{R}$ (reelle Zahlen)
- Parameter: $\mu \in \mathbb{R}$, $\sigma^2 > 0$
- Modellierung: Abweichung von Messwerten von einem Mittelwert (Gewicht, Größe, Alter,...)



Wahrscheinlichkeitsfunktion, Dichte

Statt über die Verteilungsfunktion F werden diskrete Zufallsvariablen häufig über ihre Wahrscheinlichkeitsfunktion, stetige Zufallsvariablen über ihre Dichte beschrieben.

Definition: Wahrscheinlichkeitsfunktion

Für eine diskrete Zufallsvariable X mit Wertebereich $\{x_1, x_2, \dots\}$ ist die **Wahrscheinlichkeitsfunktion** oder **Zähldichte** f an einem Wert $x \in \mathbb{R}$ definiert durch

$$f(x) = \begin{cases} P(X = x_i) & x = x_i \text{ mit } i = 1, 2, \dots \\ 0 & \text{sonst.} \end{cases}$$

Eigenschaften von Wahrscheinlichkeitsfunktionen:

- $0 \leq f(x) \leq 1$ für alle $x \in \mathbb{R}$.
- $\sum_{i=1}^{\infty} f(x_i) = 1$.

Wahrscheinlichkeitsfunktion, Dichte

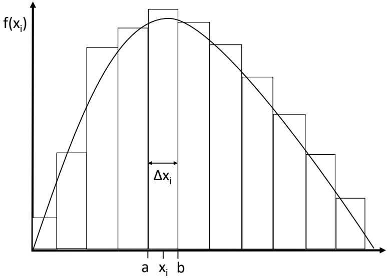
Definition: Dichte

Die **Dichte** f einer stetigen Zufallsvariablen X ist nichtnegativ (d.h. $f(x) \geq 0$ für alle $x \in \mathbb{R}$) und es gilt

$$P(a \leq X \leq b) = \int_a^b f(t)dt$$

für alle Intervalle $[a, b] \subset \mathbb{R}$.

Wahrscheinlichkeitsfunktion, Dichte



- Variable stetig, falls zu zwei Werten $a < b$ auch jeder Zwischenwert im Intervall $[a, b]$ möglich ist
- Wie werden die Wahrscheinlichkeiten bestimmt?
- Diskret: Summe über Wahrscheinlichkeiten für x_i aus $[a, b]$
- Stetig: ausgehend von diskreter Größe

$\rightarrow$ Fläche von Rechteck $(a \cdot b)$: $P(X = x_i) = f(x_i)\Delta x_i$

$\rightarrow$ Δx_i immer kleiner machen

$\rightarrow$ $P(a \leq x \leq b) =$ Fläche zwischen Intervall $[a, b]$ und $f(x)$

Wahrscheinlichkeitsfunktion, Dichte

Definition: Dichte

Die **Dichte** f einer stetigen Zufallsvariablen X ist nichtnegativ (d.h. $f(x) \geq 0$ für alle $x \in \mathbb{R}$) und es gilt

$$P(a \leq X \leq b) = \int_a^b f(t)dt$$

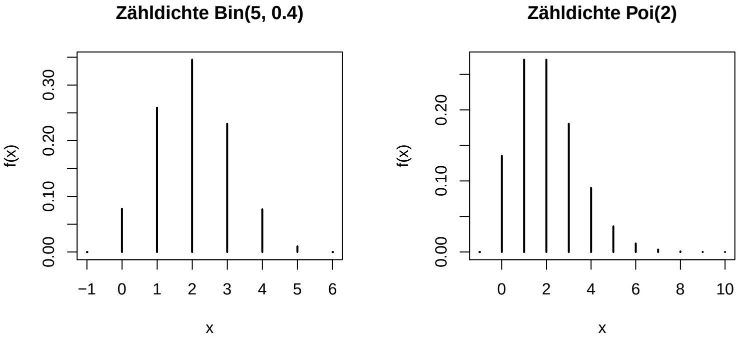
für alle Intervalle $[a, b] \subset \mathbb{R}$.

Eigenschaften von Dichtefunktionen:

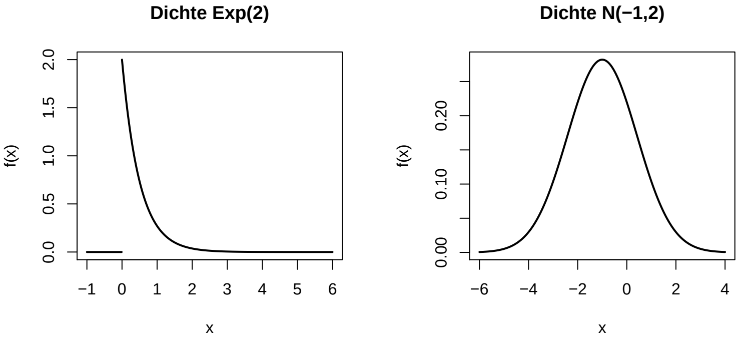
- $\int_{-\infty}^{\infty} f(t)\, dt = 1$
- Zusammenhang mit der Verteilungsfunktion:
  $F(x) = P(-\infty \leq X \leq x) = \int_{-\infty}^{x} f(t)\, dt$

$\rightarrow$ Die Verteilungsfunktion F ist Stammfunktion der Dichte f. Anders formuliert: Die Dichte f ist die Ableitung der Verteilungsfunktion F.

Beispiel 1 – Diskrete Zufallsvariablen



Beispiel 1 – Stetige Zufallsvariablen



Kennzahlen von Zufallsvariablen

Definition: Erwartungswert

Der **Erwartungwert** gibt an, welchen Wert eine Zufallsvariable *im Mittel* annimmt.
Für eine *diskrete* Zufallsvariable mit Wahrscheinlichkeitsfunktion f und Wertebereich $\{x_1, x_2, \dots\}$ gilt

$$\mathbb{E}(X) = \sum_{i=1}^{\infty} x_i \cdot f(x_i).$$

Für eine *stetige* Zufallsvariable mit Dichte f gilt

$$\mathbb{E}(X) = \int_{-\infty}^{\infty} x \cdot f(x) \, dx.$$

$\rightarrow$ $\mathbb{E}(X)$ ist ein fester Wert (**nicht** zufällig!)

Kennzahlen von Zufallsvariablen

Rechenregeln:

Für Zufallsvariablen X, Y und Skalare (Zahlen) $a, b \in \mathbb{R}$ gilt

- $$\begin{aligned} \mathbb{E}(aX + b) &= \sum_{i=1}^{\infty} a x_i f(x_i) + \sum_{i=1}^{\infty} b f(x_i) \\ &= a \sum_{i=1}^{\infty} x_i f(x_i) + b \sum_{i=1}^{\infty} f(x_i) \\ &= a \cdot \mathbb{E}(X) + b \end{aligned}$$
- $\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y)$

Kennzahlen von Zufallsvariablen

Definition: Varianz

Die **Varianz**

$$\mathrm{Var}(X) = \mathbb{E}\left([X - \mathbb{E}(X)]^2\right)$$

beschreibt die erwartete quadratische Abweichung vom Erwartungswert.

Varianzverschiebungssatz

$$\mathrm{Var}(X) = \mathbb{E}(X^2) - [\mathbb{E}(X)]^2$$

Rechenregeln:

Für Zufallsvariablen X, Y und Skalare (Zahlen) $a, b \in \mathbb{R}$ gilt

- $\mathrm{Var}(aX + b) = a^2\,\mathrm{Var}(X)$
- $\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y) + 2\mathrm{Cov}(X, Y)$

Beispiel 1

Erwartungwert und Varianz lassen sich häufig in Abhängigkeit der Verteilungsparameter darstellen.

Beispiele

- $X \sim \text{Bin}(n, p)$: $\mathbb{E}(X) = np, \quad \text{Var}(X) = np(1 - p)$
- $X \sim \text{Poi}(\lambda)$: $\mathbb{E}(X) = \lambda, \quad \text{Var}(X) = \lambda$
- $X \sim \text{Exp}(\lambda)$: $\mathbb{E}(X) = \frac{1}{\lambda}, \quad \text{Var}(X) = \frac{1}{\lambda^2}$
- $X \sim \text{N}(\mu, \sigma^2)$: $\mathbb{E}(X) = \mu, \quad \text{Var}(X) = \sigma^2$

Multivariate Zufallsvariablen

Definition Zufallsvektor

Definition: Zufallsvektor

Betrachte p Zufallsvariablen $X_1 \dots X_p$. Der p-dimensionale Vektor

$$\mathbf{X} = \begin{pmatrix} X_1 \\ \vdots \\ X_p \end{pmatrix} = (X_1, \dots, X_p)^\top \in \mathbb{R}^p$$

heißt **Zufallsvektor** oder p-**dimensionale Zufallsvariable.**

- Der Zufallsvektor $\mathbf{X}$ heißt *diskret*, falls alle Komponenten X_i von $\mathbf{X}$ diskrete Zufallsvariablen sind.
- Der Zufallsvektor $\mathbf{X}$ heißt *stetig*, falls alle Komponenten X_i von $\mathbf{X}$ stetige Zufallsvariablen sind. Dann gilt $P(X_1 = x_1, \dots, X_p = x_p) = 0$ für alle $\mathbf{x} = (x_1 \dots x_p)^\top \in \mathbb{R}^p$.

Gemeinsame Verteilungsfunktion

Definition: Gemeinsame Verteilungsfunktion

Für einen Zufallsvektor $\mathbf{X} \in \mathbb{R}^p$ heißt die Funktion

$$F(\mathbf{x}) = F(x_1, \ldots, x_p) = P(X_1 \leq x_1 \ldots X_p \leq x_p), \quad \mathbf{x} \in \mathbb{R}^p$$

gemeinsame Verteilungsfunktion der Zufallsvariablen $X_1 \ldots X_p$. Durch sie ist $\mathbf{X}$ eindeutig bestimmt.

Analog zum eindimensionalen Fall lassen sich für diskrete bzw. stetige Zufallsvektoren gemeinsame Wahrscheinlichkeitsfunktionen bzw. Dichten definieren.

Gemeinsame Wahrscheinlichkeitsfunktion / Dichte

Definition: Gemeinsame Wahrscheinlichkeitsfunktion

Für einen diskreten Zufallsvektor $\mathbf{X} \in \mathbb{R}^p$ mit Wertebereich $\{\mathbf{x}_1, \mathbf{x}_2, \ldots\}$ ist die (gemeinsame) **Wahrscheinlichkeitsfunktion** oder (gemeinsame) **Zähldichte** f an $\mathbf{x} \in \mathbb{R}^p$ definiert durch

$$f(x) = \begin{cases} P(\mathbf{X} = \mathbf{x}_i) & \mathbf{x} = \mathbf{x}_i \text{ mit } i = 1, 2, \ldots \\ 0 & \text{sonst.} \end{cases}$$

Gemeinsame Wahrscheinlichkeitsfunktion / Dichte

Definition: Gemeinsame Dichte

Die (gemeinsame) **Dichte** f eines stetigen Zufallsvektors $\mathbf{X} \in \mathbb{R}^p$ ist nichtnegativ (d.h. $f(\mathbf{x}) \geq 0$ für alle $\mathbf{x} \in \mathbb{R}^p$) und es gilt

$$\begin{aligned} P(\mathbf{a} \leq \mathbf{X} \leq \mathbf{b}) &= P(a_1 \leq X_1 \leq b_1, \ldots, a_p \leq X_p \leq b_p) \\ &= \int_{a_p}^{b_p} \cdots \int_{a_1}^{b_1} f(t_1, \ldots, t_p)\, dt_1 \ldots dt_p. \end{aligned}$$

für alle $\mathbf{a}, \mathbf{b} \in \mathbb{R}^p$ mit $a_j < b_j,\ j = 1 \ldots p$.

Gemeinsame Wahrscheinlichkeitsfunktion / Dichte

Eigenschaften der gemeinsamen Dichte:

- Wie im eindimensionalen Fall gilt:

$$\int_{-\infty}^{+\infty} \cdots \int_{-\infty}^{+\infty} f(t_1, \ldots, t_p)\, dt_1 \ldots dt_p = 1$$

- Zusammenhang mit der gemeinsamen Verteilungsfunktion:

$$\begin{aligned} F(\mathbf{x}) &= F(x_1, \ldots, x_p) = P(X_1 \leq x_1, \ldots, X_p \leq x_p) \\ &= \int_{-\infty}^{x_p} \cdots \int_{-\infty}^{x_1} f(t_1, \ldots, t_p) dt_1 \ldots dt_p = \int_{-\infty}^{\mathbf{x}} f(\mathbf{t}) \mathrm{d}\mathbf{t} \end{aligned}$$

Erwartungswertvektor

Definition: Erwartungswertvektor

Gegeben sei der Zufallsvektor $\mathbf{X} = (X_1, \ldots, X_p)^\top \in \mathbb{R}^p$. Dann ist der **Erwartungswertvektor** von $\mathbf{X}$ definiert als

$$\mathbb{E}(\mathbf{X}) = \begin{pmatrix} \mathbb{E}(X_1) \\ \vdots \\ \mathbb{E}(X_p) \end{pmatrix} .$$

Rechenregeln:

Für Zufallsvektoren $\mathbf{X}$, $\mathbf{Y} \in \mathbb{R}^p$, eine Matrix $\mathbf{A} \in \mathbb{R}^{q \times p}$ und einen Vektor $\mathbf{b} \in \mathbb{R}^q$ gilt

- $\mathbb{E}(\mathbf{X} + \mathbf{Y}) = \mathbb{E}(\mathbf{X}) + \mathbb{E}(\mathbf{Y})$
- $\mathbb{E}(\mathbf{A} \cdot \mathbf{X} + \mathbf{b}) = \mathbf{A} \cdot \mathbb{E}(\mathbf{X}) + \mathbf{b}$

Kovarianz / Korrelation

Definition: Kovarianz

Die **Kovarianz** von zwei Zufallsvariablen X und Y ist definiert als

$$\mathrm{Cov}(X, Y) = \mathbb{E}[(X - \mathbb{E}(X)) \cdot (Y - \mathbb{E}(Y))].$$

Verschiebungssatz

$$\mathrm{Cov}(X, Y) = \mathbb{E}(X \cdot Y) - \mathbb{E}(X)\mathbb{E}(Y)$$

Sind X und Y stetig, so gilt

$$\mathrm{Cov}(X, Y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} (x - \mathbb{E}(X))(y - \mathbb{E}(Y))\, f(x, y)\, dxdy$$

$$\mathbb{E}(X \cdot Y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} x \cdot y \cdot f(x, y)\, dxdy$$

Kovarianz / Korrelation

Definition: Korrelation

Die **Korrelation** von zwei Zufallsvariablen X und Y ist definiert als

$$\mathrm{Corr}(X, Y) = \frac{\mathrm{Cov}(X, Y)}{\sqrt{\mathrm{Var}(X) \cdot \mathrm{Var}(Y)}}.$$

Eigenschaften:

- $\mathrm{Cov}(X, Y) = \mathrm{Cov}(Y, X)$
  $\mathrm{Corr}(X, Y) = \mathrm{Corr}(Y, X)$
- $\mathrm{Cov}(X, X) = \mathbb{E}(X^2) - [\mathbb{E}(X)]^2 = \mathrm{Var}(X)$
  $\Rightarrow \ \mathrm{Corr}(X, X) = 1$
- $-1 \leq \mathrm{Corr}(X, Y) \leq 1$
- X, Y unabhängig $\Rightarrow \mathrm{Cov}(X, Y) = 0$ und $\mathrm{Corr}(X, Y) = 0$
  Umkehrung ($\Leftarrow$) gilt i.A. nicht!

Kovarianz / Korrelation

Rechenregeln:
Für Zufallsvariablen X, Y und Skalare a, b, c, d gilt

- $\mathrm{Cov}(aX + c, bY + d) = ab\mathrm{Cov}(X, Y)$
- $\mathrm{Corr}(aX + c, bY + d) = \mathrm{Corr}(X, Y)$

Kovarianzmatrix

Definition: Kovarianzmatrix

Gegeben sei der Zufallsvektor $\mathbf{X} = (X_1, \dots, X_p)^\top \in \mathbb{R}^p$. Dann ist die Kovarianzmatrix $\mathrm{Cov}(\mathbf{X})$ von $\mathbf{X}$ definiert durch

$$\mathrm{Cov}(\mathbf{X}) = \mathbf{\Sigma} = \mathbb{E}\left[(\mathbf{X} - \mathbb{E}(\mathbf{X}))(\mathbf{X} - \mathbb{E}(\mathbf{X}))^\top\right]$$

$$= \begin{pmatrix} \mathrm{Cov}(X_1, X_1) & \dots & \mathrm{Cov}(X_1, X_p) \\ \vdots & \ddots & \vdots \\ \mathrm{Cov}(X_p, X_1) & \dots & \mathrm{Cov}(X_p, X_p) \end{pmatrix} \in \mathbb{R}^{p \times p}.$$

Verschiebungssatz

$$\mathrm{Cov}(\mathbf{X}) = \mathbb{E}(\mathbf{X}\mathbf{X}^\top) - \mathbb{E}(\mathbf{X})\mathbb{E}(\mathbf{X})^\top.$$

Kovarianzmatrix

Eigenschaften:

- $\text{Cov}(\mathbf{X})$ ist quadratisch.
- $\text{Cov}(\mathbf{X})$ ist symmetrisch.
- Die Diagonalelemente von $\text{Cov}(\mathbf{X})$ sind gegeben durch

$$\text{Cov}(X_i, X_i) = \text{Var}(X_i) \geq 0 \quad \text{für alle } i = 1 \ldots, p.$$

- $\text{Cov}(\mathbf{X})$ ist positiv semidefinit.

Rechenregeln:
Für einen Zufallsvektor $\mathbf{X} \in \mathbb{R}^p$, eine Matrix $\mathbf{A} \in \mathbb{R}^{q \times p}$ und Vektoren $\mathbf{a} \in \mathbb{R}^q, \mathbf{b} \in \mathbb{R}^q$ gilt

- $\text{Var}(\mathbf{a}^\top \mathbf{X}) = \mathbf{a}^\top \text{Cov}(\mathbf{X})\mathbf{a}$
- $\text{Cov}(\mathbf{A}\mathbf{X} + \mathbf{b}) = \mathbf{A}\text{Cov}(\mathbf{X})\mathbf{A}^\top$

Kovarianzmatrix

Welche Matrizen sind gültige Kovarianzmatrizen?

$$\boldsymbol{\Sigma}_1 = \begin{pmatrix} 0.2 & 0.5 \\ 0.2 & 0.3 \\ 0.5 & 0.3 \end{pmatrix} \qquad \boldsymbol{\Sigma}_2 = \begin{pmatrix} 0.5 & 0.7 & 0.9 \\ 0.3 & 0.9 & 0.3 \\ 0.9 & 0.7 & 0.5 \end{pmatrix}$$

$$\boldsymbol{\Sigma}_3 = \begin{pmatrix} 2 & 0 \\ 0 & 3 \end{pmatrix} \qquad \boldsymbol{\Sigma}_4 = \begin{pmatrix} 0.5 & 0.7 & -0.9 \\ 0.7 & 0.9 & 0.3 \\ -0.9 & 0.3 & -0.5 \end{pmatrix}$$

Kovarianzmatrix

Für zwei Vektoren $\mathbf{X} = (X_1, \ldots, X_p)^\top$ und $\mathbf{Y} = (Y_1, \ldots, Y_q)$ gilt

$$\mathbf{\Sigma}_{XY} = cov(\mathbf{X}, \mathbf{Y}) = \begin{pmatrix} cov(X_1, Y_1) & \ldots & cov(X_1, Y_q) \\ \vdots & \vdots & \vdots \\ cov(X_p, Y_1) & \ldots & cov(X_p, Y_q) \end{pmatrix}$$

$\rightarrow$ $\mathbf{\Sigma}_{XY}$ ist i.A. nicht symmetrisch!

Rechenregeln

- $\mathbb{E}(\mathbf{AX}+\mathbf{b}) = \mathbf{A}\mathbb{E}(\mathbf{X})+\mathbf{b}$
-

$$\begin{aligned}
\mathrm{Cov}(\mathbf{X}) &= \mathbb{E}\left((\mathbf{X}-\boldsymbol{\mu})(\mathbf{X}-\boldsymbol{\mu})^\top\right) \\
&= \mathbb{E}\left(\mathbf{XX}^\top - \mathbf{X}\boldsymbol{\mu}^\top - \boldsymbol{\mu}\mathbf{X}^\top + \boldsymbol{\mu}\boldsymbol{\mu}^\top\right) \\
&= \mathbb{E}\left(\mathbf{XX}^\top - (\mathbf{X}\boldsymbol{\mu}^\top)^\top - \boldsymbol{\mu}\mathbf{X}^\top + \boldsymbol{\mu}\boldsymbol{\mu}^\top\right) \\
&= \mathbb{E}\left(\mathbf{XX}^\top - \boldsymbol{\mu}\mathbf{X}^\top - \boldsymbol{\mu}\mathbf{X}^\top + \boldsymbol{\mu}\boldsymbol{\mu}^\top\right) \\
&= \mathbb{E}(\mathbf{XX}^\top) - 2\boldsymbol{\mu}\mathbb{E}(\mathbf{X}^\top) + \boldsymbol{\mu}\boldsymbol{\mu}^\top \\
&= \mathbb{E}(\mathbf{XX}^\top) - 2\boldsymbol{\mu}\boldsymbol{\mu}^\top + \boldsymbol{\mu}\boldsymbol{\mu}^\top \\
&= \mathbb{E}(\mathbf{XX}^\top) - \boldsymbol{\mu}\boldsymbol{\mu}^\top
\end{aligned}$$

- $\mathrm{Cov}(\mathbf{AX}+\mathbf{b}) = \mathbf{A}\mathrm{Cov}(\mathbf{X})\mathbf{A}^\top$
- $\mathrm{Cov}(\mathbf{X},\mathbf{Y}) = \mathbb{E}\left((\mathbf{X}-\mathbb{E}(\mathbf{X}))(\mathbf{Y}-\mathbb{E}(\mathbf{Y}))^\top\right)$
- $\mathrm{Cov}(\mathbf{AX}+\mathbf{a},\mathbf{BY}+\mathbf{b}) = \mathbf{A}\mathrm{Cov}(\mathbf{X},\mathbf{Y})\mathbf{B}^\top$

Zerlegung der Kovarianzmatrix

Es gilt:

$$\boldsymbol{\Sigma} = \boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{\Sigma}^{\frac{1}{2}} \quad \text{bzw.} \quad \boldsymbol{\Sigma} = \boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{\Sigma}^{\frac{T}{2}}$$

und

$$\boldsymbol{\Sigma}^{-1} = \boldsymbol{\Sigma}^{-\frac{T}{2}}\boldsymbol{\Sigma}^{-\frac{1}{2}}$$

Daraus resultiert:

$$\begin{aligned}
\text{Cov}(\boldsymbol{\Sigma}^{-\frac{1}{2}}\mathbf{X}) &= \boldsymbol{\Sigma}^{-\frac{1}{2}}\text{Cov}(\mathbf{X})\boldsymbol{\Sigma}^{-\frac{T}{2}} \\
&= \boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{\Sigma}\boldsymbol{\Sigma}^{-\frac{T}{2}} \\
&= \boldsymbol{\Sigma}^{-\frac{1}{2}}\boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{\Sigma}^{\frac{T}{2}}\boldsymbol{\Sigma}^{-\frac{T}{2}} \\
&= (\boldsymbol{\Sigma}^{\frac{1}{2}})^{-1}\boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{\Sigma}^{\frac{T}{2}}(\boldsymbol{\Sigma}^{\frac{T}{2}})^{-1} \\
&= \mathbf{I}
\end{aligned}$$

$\Rightarrow \boldsymbol{\Sigma}^{-\frac{1}{2}}(\mathbf{X} - \mathbb{E}(\mathbf{X}))$ hat Erwartungswert $\mathbf{0}$ and Varianz $\mathbf{I}$ (Multivariate Standardisierung).

Spektralzerlegung

Es gilt:

$$\mathbf{\Sigma} = \mathbf{P}\mathbf{\Lambda}\mathbf{P}^\top \quad \text{und} \quad \mathbf{\Sigma}^{-1} = \mathbf{P}\mathbf{\Lambda}^{-1}\mathbf{P}^\top \,,$$

wobei

$\mathbf{P} = (\mathbf{p}_1, \ldots, \mathbf{p}_p)^\top$: Matrix der *orthonormalen* Eigenvektoren und

$\mathbf{\Lambda} = \mathit{diag}(\lambda_1, \ldots, \lambda_p)$: Diagonalmatrix der *geordneten* Eigenwerte.

Korrelationsmatrix

Definition: Korrelationsmatrix

Die **Korrelationsmatrix** eines Zufallsvektors $\mathbf{X} \in \mathbb{R}^p$ ist wie folgt definiert:

$$\mathrm{Corr}(\mathbf{X}) = \begin{pmatrix} \mathrm{Corr}(X_1, X_1) & \dots & \mathrm{Corr}(X_1, X_p) \\ \vdots & \ddots & \vdots \\ \mathrm{Corr}(X_p, X_1) & \dots & \mathrm{Corr}(X_p, X_p) \end{pmatrix} \in \mathbb{R}^{p \times p}.$$

Eigenschaften:

- $\mathrm{Corr}(\mathbf{X})$ ist quadratisch
- $\mathrm{Corr}(\mathbf{X})$ ist symmetrisch
- Diagonale: $\mathrm{Corr}(X_i, X_i) = 1$ für alle $i = 1 \dots, p$
- $\mathrm{Corr}(\mathbf{X})$ ist positiv semidefinit.

Korrelation und Unabhängigkeit

Definition: Unkorrelierte Zufallsvektoren

Seien $\mathbf{Y} \in \mathbb{R}^r$ und $\mathbf{Z} \in \mathbb{R}^q$ zwei Zufallsvektoren mit

$$\mathrm{Cov}\left(\begin{pmatrix} \mathbf{Y} \\ \mathbf{Z} \end{pmatrix}\right) = \begin{pmatrix} \mathrm{Cov}(Y_1, Y_1) & \dots & \mathrm{Cov}(Y_1, Y_r) & \mathrm{Cov}(Y_1, Z_1) & \dots & \mathrm{Cov}(Y_1, Z_q) \\ & \ddots & & & \ddots & \\ \mathrm{Cov}(Y_r, Y_1) & \dots & \mathrm{Cov}(Y_r, Y_r) & \mathrm{Cov}(Y_r, Z_1) & \dots & \mathrm{Cov}(Y_r, Z_q) \\ \mathrm{Cov}(Z_1, Y_1) & \dots & \mathrm{Cov}(Z_1, Y_r) & \mathrm{Cov}(Z_1, Z_1) & \dots & \mathrm{Cov}(Z_1, Z_q) \\ & \ddots & & & \ddots & \\ \mathrm{Cov}(Z_q, Y_1) & \dots & \mathrm{Cov}(Z_q, Y_r) & \mathrm{Cov}(Z_q, Z_1) & \dots & \mathrm{Cov}(Z_q, Z_q) \end{pmatrix}$$

$$= \begin{pmatrix} \mathbf{\Sigma}_Y & \mathbf{\Sigma}_{YZ} \\ \mathbf{\Sigma}_{ZY} & \mathbf{\Sigma}_Z \end{pmatrix} .$$

$\mathbf{Y}$ und $\mathbf{Z}$ heißen **unkorreliert**, falls $\Sigma_{YZ} = \mathbf{0}$.

Bemerkung:

- $\mathbf{Y}$ und $\mathbf{Z}$ sind unabhängig $\Rightarrow$ $\mathbf{Y}$ und $\mathbf{Z}$ sind unkorreliert.
  Die Umkehrung ($\Leftarrow$) gilt i.A. nicht!

Multivariate Normalverteilung

Multivariate Normalverteilung

Definition: Multivariate Normalverteilung

Ein Zufallsvektor $\mathbf{X} \in \mathbb{R}^p$ heißt **multivariat normalverteilt** mit Parametern $\boldsymbol{\mu}$ und $\boldsymbol{\Sigma}$, falls er die folgende Dichte hat

$$f(\mathbf{x}) = \frac{1}{\left(\sqrt{2\pi}\right)^p \sqrt{\det(\boldsymbol{\Sigma})}} \exp\left\{-\frac{1}{2}(\mathbf{x}-\boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1}(\mathbf{x}-\boldsymbol{\mu})\right\}$$

mit $\boldsymbol{\mu} \in \mathbb{R}^p$ und positiv definiter Matrix $\boldsymbol{\Sigma} \in \mathbb{R}^{p \times p}$.

Die Schreibweise ist analog zur eindimensionalen Normalverteilung:

$$\mathbf{X} \sim \mathrm{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$$

Häufig wird der Index p unterdrückt, wenn sich die Dimension aus dem Zusammenhang erschließen lässt.

Zusammenhang zur Spektralzerlegung

Es gilt (quadratische Form):

$$\begin{aligned}(\mathbf{X}-\mu)^\top \mathbf{\Sigma}^{-1}(\mathbf{X}-\mu) =&(\mathbf{X}-\mu)^\top \mathbf{P}\mathbf{\Lambda}^{-1}\mathbf{P}^\top(\mathbf{X}-\mu) = \\ =&\mathbf{Y}\mathbf{\Lambda}^{-1}\mathbf{Y}^\top = \sum_{i=1}^{p}\frac{y_i^2}{\lambda_i} = \sum_{i=1}^{p}\left(\frac{y_i}{\sqrt{\lambda_i}}\right)^2\end{aligned}$$

Die Isodensiten (Kurven mit derselben Dichte) sind damit durch folgende Gleichung bestimmt:

$$(\mathbf{X}-\mu)^\top \mathbf{\Sigma}^{-1}(\mathbf{X}-\mu) = \sum_{i=1}^{p}\left(\frac{y_i}{\sqrt{\lambda_i}}\right)^2 = c$$

$\rightarrow$ Dies stellt in den y-Werten ein Ellipsoid mit den Hauptachsenlängen $\sqrt{\lambda_i}$ dar!

Multivariate Normalverteilung

Eigenschaften:

- $\mathbb{E}(\mathbf{X}) = \boldsymbol{\mu}$, $\text{Cov}(\mathbf{X}) = \boldsymbol{\Sigma}$
- Für $\mathbf{X} \sim N_p(\,\boldsymbol{\mu}\,,\,\boldsymbol{\Sigma}\,)$ und $\mathbf{A} \in \mathbb{R}^{q\times p}, \mathbf{b} \in \mathbb{R}^q$ gilt

$$\mathbf{Y} = \mathbf{A}\,\mathbf{X} + \mathbf{b} \sim N_q(\,\mathbf{A}\,\boldsymbol{\mu} + \mathbf{b}\,,\ \mathbf{A}\,\boldsymbol{\Sigma}\,\mathbf{A}^\top)\,.$$

- $(\mathbf{X} - \boldsymbol{\mu})^\top \boldsymbol{\Sigma}^{-1} (\mathbf{X} - \boldsymbol{\mu}) \sim \chi^2(p)$

Spezialfälle:

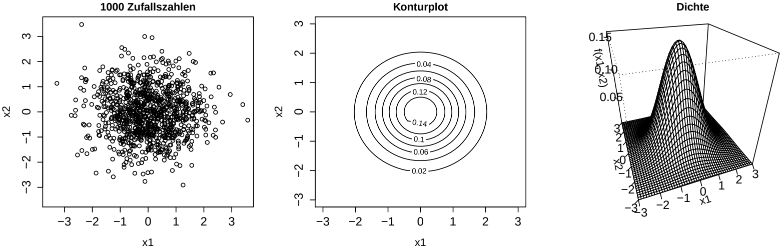
- Für $p = 1$ ergibt sich die univariate Normalverteilung mit Parametern $\boldsymbol{\mu} = \mathbb{E}(X)$ und $\boldsymbol{\Sigma} = \text{Var}(X)$.
- Für $\mathbf{X} \sim N_p(\,\mathbf{0}\,,\,\mathbf{I}\,)$, also mit

$$\boldsymbol{\mu} = \mathbf{0} = \begin{pmatrix} 0 \\ \vdots \\ 0 \end{pmatrix} \qquad \text{und} \qquad \boldsymbol{\Sigma} = \mathbf{I} = \begin{pmatrix} 1 & \cdots & 0 \\ & \ddots & \\ 0 & \cdots & 1 \end{pmatrix},$$

ergibt sich die multivariate **Standardnormalverteilung.**

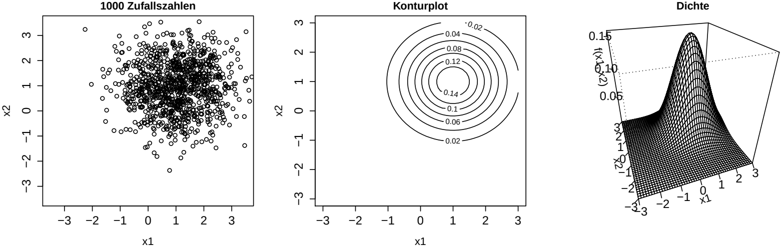
Multivariate Normalverteilung – Beispiele

$$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim \mathrm{N}_2 \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \right)$$



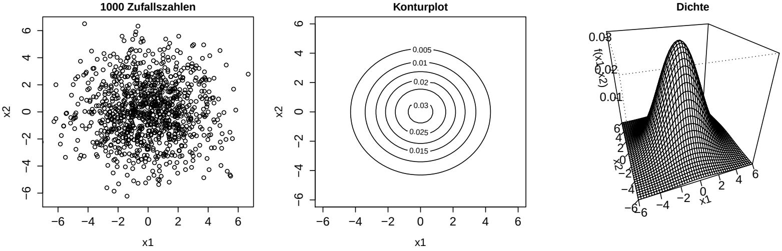
Multivariate Normalverteilung – Beispiele

$$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim \mathrm{N}_2\left(\begin{pmatrix} 1 \\ 1 \end{pmatrix}, \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}\right)$$



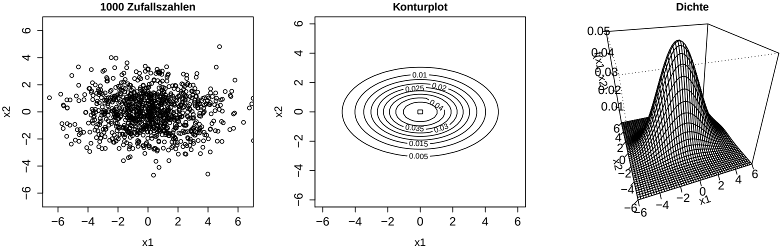
Multivariate Normalverteilung – Beispiele

$$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim \mathrm{N}_2\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 5 & 0 \\ 0 & 5 \end{pmatrix} \right)$$



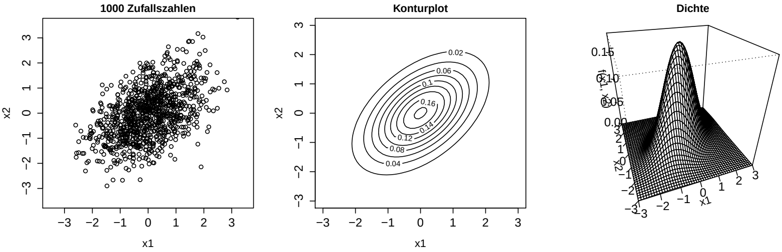
Multivariate Normalverteilung – Beispiele

$$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim \mathrm{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 5 & 0 \\ 0 & 2 \end{pmatrix} \right)$$



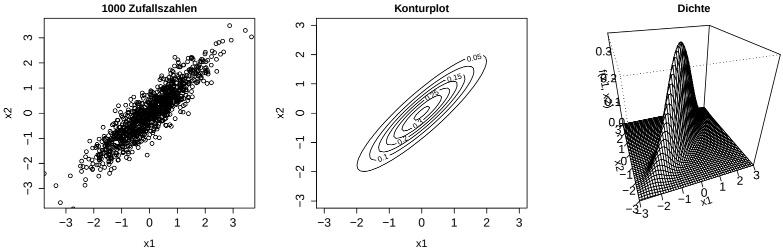
Multivariate Normalverteilung – Beispiele

$$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim \mathrm{N}_2\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0.5 \\ 0.5 & 1 \end{pmatrix}\right)$$



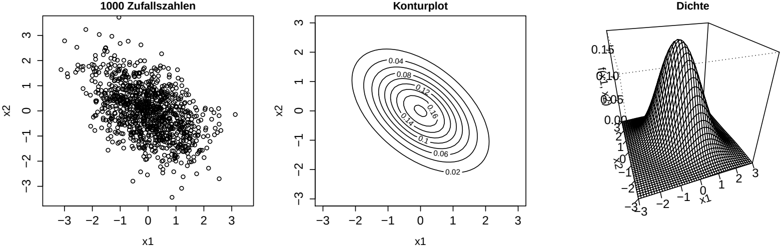
Multivariate Normalverteilung – Beispiele

$$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim \mathrm{N}_2\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 0.9 \\ 0.9 & 1 \end{pmatrix}\right)$$



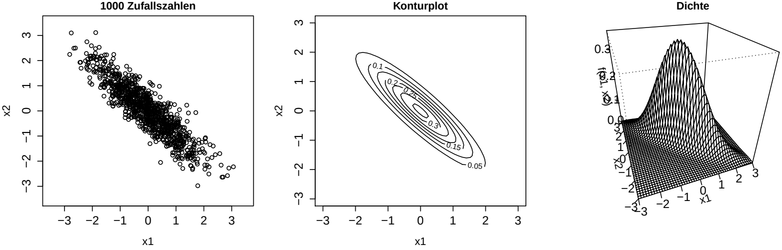
Multivariate Normalverteilung – Beispiele

$$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim \mathrm{N}_2\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & -0.5 \\ -0.5 & 1 \end{pmatrix}\right)$$



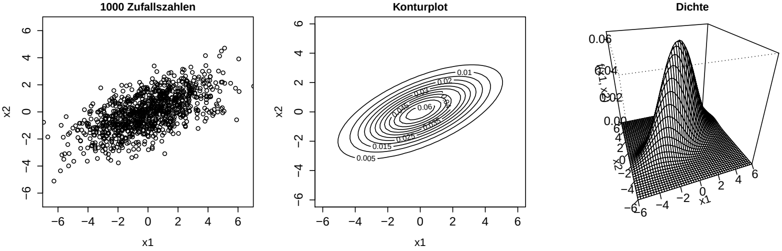
Multivariate Normalverteilung – Beispiele

$$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim \mathrm{N}_2 \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & -0.9 \\ -0.9 & 1 \end{pmatrix} \right)$$



Multivariate Normalverteilung – Beispiele

$$\begin{pmatrix} X_1 \\ X_2 \end{pmatrix} \sim \mathrm{N}_2\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 5 & 2 \\ 2 & 2 \end{pmatrix}\right)$$



Rand- und bedingte Verteilung

Sei $\mathbf{X} \sim \mathrm{N}_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Betrachte die Partition

$$\mathbf{X} = \begin{pmatrix} X_1 \\ \vdots \\ X_r \\ X_{r+1} \\ \vdots \\ X_p \end{pmatrix} = \begin{pmatrix} \mathbf{Y} \\ \mathbf{Z} \end{pmatrix}$$

und damit

$$\boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \vdots \\ \mu_r \\ \mu_{r+1} \\ \vdots \\ \mu_p \end{pmatrix} = \begin{pmatrix} \boldsymbol{\mu}_Y \\ \boldsymbol{\mu}_Z \end{pmatrix} \qquad \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_Y & \boldsymbol{\Sigma}_{YZ} \\ \boldsymbol{\Sigma}_{ZY} & \boldsymbol{\Sigma}_Z \end{pmatrix}.$$

Rand- und bedingte Verteilung

Dann gilt:

- Randverteilungen:

$$\begin{aligned} \mathbf{Y} &\sim \mathrm{N}_r(\boldsymbol{\mu}_Y, \boldsymbol{\Sigma}_Y) \\ \mathbf{Z} &\sim \mathrm{N}_q(\boldsymbol{\mu}_Z, \boldsymbol{\Sigma}_Z) \qquad \text{mit} \quad q = p - r \end{aligned}$$

- Bedingte Verteilung:

$$\mathbf{Y}|\mathbf{Z} = \mathbf{z}_0 \sim \mathrm{N}_r(\boldsymbol{\mu}_{Y|Z}, \boldsymbol{\Sigma}_{Y|Z})$$

$$\text{mit} \quad \boldsymbol{\mu}_{Y|Z} = \boldsymbol{\mu}_Y + \boldsymbol{\Sigma}_{YZ} \cdot \boldsymbol{\Sigma}_Z^{-1}(\mathbf{z}_0 - \boldsymbol{\mu}_Z)$$

$$\boldsymbol{\Sigma}_{Y|Z} = \boldsymbol{\Sigma}_Y - \boldsymbol{\Sigma}_{YZ} \boldsymbol{\Sigma}_Z^{-1} \boldsymbol{\Sigma}_{ZY}$$

- Unabhängigkeit:
  $\boldsymbol{\Sigma}_{YZ} = 0 \Rightarrow$ **Y** und **Z** sind unabhängig
  also: **Y** und **Z** sind unabhängig $\Leftrightarrow \boldsymbol{\Sigma}_{YZ} = 0$

Weitere multivariate Verteilungen

Wishart-Verteilung

Seien $\mathbf{X}_1, \ldots, \mathbf{X}_m \overset{i.i.d.}{\sim} N_p(\mathbf{0}, \mathbf{\Sigma})$, dann ist die Matrix

$$\mathbf{M} = \sum_{i=1}^{m} \mathbf{X}_i \mathbf{X}_i^\top = \mathbf{X}^\top \mathbf{X} \quad \in \mathbb{R}^{pxp}$$

wishart-verteilt mit Parametern $\mathbf{\Sigma}$ und m.

- $\mathbf{M} \sim W_p(\mathbf{\Sigma}, m)$
- Falls $p = 1$, so gilt $\boldsymbol{M} = \sum_{i=1}^{m} X_i^2 \sim \chi^2(m)$, wobei die Komponenten $X_i \sim N(0, \sigma^2)$
- Multivariate Erweiterung der χ^2-Verteilung

Hotellings T^2-Verteilung

Seien $\mathbf{d} \sim N_p(\mathbf{0}, \mathbf{I})$ und $\mathbf{M} \sim W_p(\mathbf{I}, m)$ unabhängig, dann ist

$$u = m\mathbf{d}^\top \mathbf{M}^{-1}\mathbf{d}$$

Hotellings T^2-verteilt mit Parametern p und m.

- $u \sim T^2(p, m)$
- Falls $p = 1$, so ergibt sich die $F(1, m)$-Verteilung
- Seien $\mathbf{d} \sim N_p(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ und $\mathbf{M} \sim W_p(\boldsymbol{\Sigma}, m)$ unabhängig, dann erhält man:
  $m(\mathbf{d} - \boldsymbol{\mu})^\top \mathbf{M}^{-1}(\mathbf{d} - \boldsymbol{\mu}) \sim T^2(p, m)$
- Verwendung bei Testproblemen (multivariate Erweiterung der t-Verteilung bzw. des t-Tests)

Wilks' Λ-Verteilung

Seien $\mathbf{A} \sim W_p(\mathbf{I}, m)$ und $\mathbf{B} \sim W_p(\mathbf{I}, n)$ unabhängig, dann ist

$$\Lambda = \frac{\det(\mathbf{A})}{\det(\mathbf{A} + \mathbf{B})}$$

Wilks' Λ-verteilt mit Parametern p, m und n.

- $\Lambda \sim \Lambda(p, m, n)$
- Falls $p = 1$, so gilt $A \sim \chi^2(m)$ und $B \sim \chi^2(n)$ und damit erhält man: $\Lambda \sim B(m/2, n/2)$
- Verwendung beim Testen im Rahmen der einfaktoriellen Varianzanalyse