

Kapitel 1

Mehrdimensionale Zufallsvariablen

Skript Gerhard Tutz, LMU, 18. April 2014

Bei der Beobachtung zufallsgesteuerter Systeme spielen Zufallsvariablen eine bedeutende Rolle. Das Verhalten eindimensionaler Zufallsvariablen läßt sich äquivalent beschreiben durch die Dichte oder die Verteilungsfunktion. Eine stetige Zufallsvariable X besitzt eine *Dichte* $f(x) \geq 0$, für die gilt

$$P(a \leq X \leq b) = \int_a^b f(x) dx.$$

Für diskrete Zufallsvariablen, die nur die Werte x_1, x_2, \dots annehmen können ist die *diskrete Dichte* oder *Massenfunktion* f bestimmt durch

$$f(x_i) = P(X = x_i).$$

Dieselbe Information über den Zufallscharakter der Zufallsvariable ist in der Verteilungsfunktion F enthalten, die die Wahrscheinlichkeit nach links kumuliert,

$$F(x) = P(X \leq x) = \begin{cases} \int_{-\infty}^x f(u) du & X \text{ stetig} \\ \sum_{x_i \leq x} f(x_i) & X \text{ diskret} \end{cases}.$$

1.1 Multivariate Dichten und Verteilungsfunktionen

In der multivariaten Statistik wird der Zusammenhang mehrerer Zufallsvariablen X_1, \dots, X_n untersucht, die zum Zufallsvektor $\mathbf{X} = (X_1, \dots, X_n)^T$ zusammengefasst werden.

Ein Zufallsvektor \mathbf{X} heißt *stetig*, wenn es eine Funktion $f(\mathbf{x}) = f(x_1, \dots, x_p)$ gibt, so daß gilt

$$\begin{aligned} P(a_1 \leq X_1 \leq b_1, \dots, a_p \leq X_p \leq b_p) &= \int_{a_1}^{b_1} \cdots \int_{a_p}^{b_p} f(u_1, \dots, u_p) du_1 \cdots du_p \\ &= \int_{a_1}^{b_1} \cdots \int_{a_p}^{b_p} f(\mathbf{u}) d\mathbf{u}. \end{aligned}$$

Die Funktion f ist die multivariate Dichte des Zufallsvektors \mathbf{X} . Für eine diskrete Zufallsvariable mit den möglichen Werten $\tilde{x}_1, \tilde{x}_2, \dots$ ist die *diskrete* Dichte oder Wahrscheinlichkeitsfunktion bestimmt durch

$$f(\mathbf{x}_i) = P(\mathbf{X} = \tilde{\mathbf{x}}_i).$$

Die zugehörige p -dimensionale Verteilungsfunktion ist gegeben durch

$$\begin{aligned} F(\mathbf{x}) &= F(x_1, \dots, x_p) = P(X_1 \leq x_1, \dots, X_p \leq x_p) \\ &= \begin{cases} \int_{-\infty}^{x_1} \cdots \int_{-\infty}^{x_p} f(u_1, \dots, u_p) du_1 \cdots du_p & \mathbf{x} \text{ stetig} \\ \sum_{\tilde{\mathbf{x}}_i \leq \mathbf{x}} f(\tilde{\mathbf{x}}_i) & \mathbf{x} \text{ diskret} \end{cases} \end{aligned}$$

wobei $\tilde{\mathbf{x}}_i \leq \mathbf{x}$ bedeutet, daß jede Komponente in $\tilde{\mathbf{x}}_i$ kleiner oder gleich der entsprechenden Komponente in \mathbf{x} ist. F besitzt die Eigenschaften:

$$0 \leq F(\mathbf{x}) \leq 1 \text{ für alle } \mathbf{x},$$

$$\lim_{x_i \rightarrow -\infty} F(\mathbf{x}) = 0 \text{ für alle Komponenten } x_i \text{ von } \mathbf{x}^T = (x_1, \dots, x_p),$$

$$\lim_{\mathbf{x} \rightarrow \infty} F(\mathbf{x}) = 1, \text{ wobei } \mathbf{x} \rightarrow \infty \text{ bezeichnet, daß } x_i \rightarrow \infty \text{ für jede Komponente } x_i,$$

$$f(x) = \partial^p F(\mathbf{x}) / \partial x_1 \cdots \partial x_p, \text{ wenn } f(x) \text{ an der Stelle } x \text{ stetig ist.}$$

Um Vektoren von Matrizen unterscheiden zu können werden im folgenden Zufallsvektoren durch $\mathbf{x}^T = (x_1, \dots, x_p)$ bezeichnet. Aus dem Kontext wird klar, ob \mathbf{x} einem Zufallsvektor oder was gelegentlich vorkommt, eine Realisation bezeichnet.

1.2 Erwartungswerte und Kovarianzen

Erwartungswerte

Der Erwartungswert eines Zufallsvektors ist definiert durch die Erwartungswerte der Komponenten. Zum Zufallsvektor $\mathbf{X}^T = (X_1, \dots, X_p)$ heißt

$$E(\mathbf{X}) = (E(X_1), \dots, E(X_p))^T$$

der *Erwartungswert(-vektor)* von \mathbf{X} . Meist wird er durch $\boldsymbol{\mu} = E(\mathbf{X})$ abgekürzt, die Komponenten von $\boldsymbol{\mu}^T = (\mu_1, \dots, \mu_n)$ entsprechen dann den univariaten Erwartungswerten, $\mu_i = E(X_i)$.

Genauso verfährt man mit Zufallsmatrizen $\mathbf{X} = (X_{ij})$, deren Komponenten Zufallsvariablen sind. Als *Erwartungswert von \mathbf{X}* bezeichnet man die Matrix der komponentenweisen Erwartungswerte, d.h.

$$E(\mathbf{X}) = (E(X_{ij})).$$

Kovarianzen

Die Kovarianzen der Komponenten des Vektors $\mathbf{X}^T = (X_1, \dots, X_p)$,

$$\sigma_{ij} = \text{cov}(X_i, X_j) = E(X_i - \mu_i)(X_j - \mu_j)$$

lassen sich zur $(p \times p)$ -Kovarianzmatrix zusammenfassen

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_{11} & \cdots & \sigma_{1p} \\ \vdots & & \\ \sigma_{p1} & \cdots & \sigma_{pp} \end{pmatrix}$$

Diese enthält in der Diagonalen die Varianzen, da

$$\sigma_{ii} = E(X_i - \mu_i)^2 = \text{var}(X_i) = \sigma_i^2.$$

Zu beachten ist die Notation $\sigma_i^2 = \sigma_{ii}$ für die Varianz der i ten Komponente. Naturgemäß ist $\boldsymbol{\Sigma}$ eine symmetrische Matrix, da $\sigma_{ij} = \sigma_{ji}$ gilt.

Die Kovarianzmatrix lässt sich in Vektorschreibweise auch durch

$$\boldsymbol{\Sigma} = \text{cov}(\mathbf{X}) = E\{(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})^T\}$$

bzw. durch

$$\text{cov}(\mathbf{X}) = E(\mathbf{X}\mathbf{X}^T) - \boldsymbol{\mu}\boldsymbol{\mu}^T$$

darstellen.

Die zugehörige Matrix der Korrelationen $\mathbf{P} = (\rho_{ij})$ erhält man aus der Definition

$$\rho_{ij} = \text{cov}(X_i, X_j) / \sqrt{\text{var}(X_i) \text{var}(X_j)}.$$

Kovarianzen zwischen Gruppen von Variablen

Gelegentlich ist es sinnvoll die Kovarianzen zweier Vektoren zu untersuchen. Zu den Vektoren $\mathbf{X}^T = (X_1, \dots, X_p)$ mit $\boldsymbol{\mu} = E(\mathbf{X})$ und $\mathbf{Y}^T = (Y_1, \dots, Y_q)$ mit $\boldsymbol{\nu} = E(\mathbf{Y})$ definiert man als *Kovarianz von X und Y*

$$\boldsymbol{\Sigma}_{\mathbf{XY}} = \text{cov}(\mathbf{X}, \mathbf{Y}) = E\{(\mathbf{X} - \boldsymbol{\mu})(\mathbf{Y} - \boldsymbol{\nu})^T\} = \begin{pmatrix} \text{cov}(X_1, Y_1) & \cdots & \text{cov}(X_1, Y_q) \\ \vdots & & \vdots \\ \text{cov}(X_p, Y_1) & \cdots & \text{cov}(X_p, Y_q) \end{pmatrix}.$$

Während in $\text{cov}(\mathbf{X})$ die Kovarianzen innerhalb des Vektors \mathbf{X} erfasst sind, enthält die $(p \times q)$ -Matrix $\text{cov}(\mathbf{X}, \mathbf{Y})$ nur die Kovarianzen zwischen Komponenten von \mathbf{X} und Komponenten von \mathbf{Y} . Für den Spezialfall $\mathbf{X} = \mathbf{Y}$ erhält man mit $\text{cov}(\mathbf{X}, \mathbf{X}) = \text{cov}(\mathbf{X})$ die Kovarianz von \mathbf{X} wieder.

Faßt man \mathbf{X} und \mathbf{Y} zum Vektor $\mathbf{Z}^T = (\mathbf{X}^T, \mathbf{Y}^T)$ zusammen, erhält man sämtliche Kovarianzen in der strukturierten Matrix

$$\text{cov}(\mathbf{z}) = \begin{pmatrix} \text{cov}(\mathbf{X}) & \text{cov}(\mathbf{X}, \mathbf{Y}) \\ \text{cov}(\mathbf{Y}, \mathbf{X}) & \text{cov}(\mathbf{Y}) \end{pmatrix} = \begin{pmatrix} \boldsymbol{\Sigma}_x & \boldsymbol{\Sigma}_{x,y} \\ \boldsymbol{\Sigma}_{y,x} & \boldsymbol{\Sigma}_y \end{pmatrix}.$$

Eigenschaften

Einige Eigenschaften von Kovarianzen (für geeignet dimensionierte feste Matrizen \mathbf{A} , \mathbf{B} und Vektoren \mathbf{a} , \mathbf{b}) sind

$$E(\mathbf{X} + \mathbf{Y}) = E(\mathbf{X}) + E(\mathbf{Y}), E(\mathbf{A}\mathbf{X} + \mathbf{b}) = \mathbf{A} E(\mathbf{X}) + \mathbf{b},$$

$$\text{var}(\mathbf{a}^T \mathbf{X}) = \mathbf{a}^T \text{cov}(\mathbf{X}) \mathbf{a} = \sum_{i,j=1}^p a_i a_j \sigma_{ij},$$

$$\text{cov}(\mathbf{A}\mathbf{X} + \mathbf{b}) = \mathbf{A} \text{cov}(\mathbf{X}) \mathbf{A}^T,$$

$$\text{cov}(\mathbf{X}, \mathbf{X}) = \text{cov}(\mathbf{X}), \text{cov}(\mathbf{X}, \mathbf{Y}) = \text{cov}(\mathbf{Y}, \mathbf{X})^T,$$

$$\text{cov}(\mathbf{A}\mathbf{X} + \mathbf{a}, \mathbf{B}\mathbf{X} + \mathbf{b}) = \mathbf{A} \text{cov}(\mathbf{X}, \mathbf{Y}) \mathbf{B}^T,$$

$\text{cov}(\mathbf{X})$ und \mathbf{P} sind symmetrisch und positiv semidefinit,

Für symmetrische Matrix \mathbf{A} gilt

$$E(\mathbf{X}^T \mathbf{A} \mathbf{X}) = \text{tr}(\mathbf{A} \mathbf{X}) + \boldsymbol{\mu} \mathbf{A} \boldsymbol{\mu}.$$

Standardisieren von Zufallsvariablen

Eindimensionale Zufallsvariablen X lassen sich einfach standardisieren durch die Transformation

$$Z = \frac{X - E(X)}{\sqrt{\text{var}(X)}}$$

Für Z gilt $E(Z) = 0$ und $\text{var}(Z) = 1$. Das Standardisieren mehrdimensionaler Zufallsvariablen zielt auf den (vektoriellen) Erwartungswert $\boldsymbol{\theta}$ und die Kovarianzmatrix \mathbf{I} (die Einheitsmatrix) ab. Entsprechend genügt es nicht, den Erwartungswert abzuziehen und durch die Varianz zu teilen. Vielmehr muß auch die Kovarianz berücksichtigt werden.

Dazu benutzt man die Wurzel einer Matrix. Für Zahlen $a \in \mathbb{R}$ existiert ein reelle Wurzel \sqrt{a} , wenn $a \geq 0$ ist. Zu einer Matrix \mathbf{A} , die symmetrisch und positiv definit (d.h. $\mathbf{x}^T \mathbf{A} \mathbf{x} > 0$ für alle $\mathbf{x} \neq \boldsymbol{\theta}$) ist, existiert eine Matrix $\mathbf{A} = \mathbf{A}^{1/2}$, so dass die Zerlegung

$$\mathbf{A} = \mathbf{A}^{1/2} \mathbf{A}^{T/2}$$

gilt, wobei $\mathbf{A}^{T/2} = (\mathbf{A}^{1/2})^T$ die Transponierte von $\mathbf{A}^{1/2}$ bezeichnet. Die Matrix $\mathbf{A}^{1/2}$ heißt linke und $\mathbf{A}^{T/2}$ heißt rechte Wurzel von \mathbf{A} . Beide sind nicht singulär, d.h. $\det(\mathbf{A}^{1/2}) \neq 0$.

Häufig betrachtet man die Wurzelzerlegung einer inversen Matrix \mathbf{A}^{-1} . Aus der Zerlegung $\mathbf{A} = \mathbf{A}^{1/2} \mathbf{A}^{T/2}$ folgt $\mathbf{A}^{-1} = (\mathbf{A}^{T/2})^{-1} (\mathbf{A}^{1/2})^{-1} = \mathbf{A}^{-T/2} \mathbf{A}^{-1/2}$, wobei $\mathbf{A}^{-T/2} := (\mathbf{A}^{T/2})^{-1}$, $\mathbf{A}^{-1/2} = (\mathbf{A}^{1/2})^{-1}$.

In dieser Notation ist die Zerlegung von \mathbf{A}^{-1} durch $\mathbf{A}^{-1} = \mathbf{A}^{-T/2} \mathbf{A}^{-1/2}$ gegeben.

Eine wichtige Anwendung ist die Normierung eines Zufallsvektors. Besitze \mathbf{X} die Kovarianz Σ , so ergibt sich für den transformierten Vektor $\Sigma^{-1/2}\mathbf{X}$

$$\text{cov}\left(\Sigma^{-1/2}\mathbf{X}\right) = \mathbf{I}$$

d.h. die Kovarianz von $\Sigma^{-1/2}\mathbf{X}$ ist durch die Einheitsmatrix \mathbf{I} gegeben.

Ersichtlich wird dies durch

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

Zur Bestimmung einer derartigen Wurzel läßt sich die *Cholesky-Zerlegung* anwenden. Sie liefert eine Matrix $\mathbf{A}^{1/2}$, die rechts von der Diagonalen nur Nullen enthält (eine sogenannte untere Dreiecksmatrix).

Generell erhält man für den Zufallsvektor \mathbf{X} mit positiv definiter Kovarianzmatrix Σ durch die Transformation

$$\mathbf{Z} = \Sigma^{-1/2}(\mathbf{X} - \boldsymbol{\mu})$$

einen standardisierten Zufallsvektor mit $E(\mathbf{Z}) = \mathbf{0}$, $\text{cov}(\mathbf{Z}) = \mathbf{I}$.

1.3 Mehrdimensionale Normalverteilung

Die Normalverteilung spielt in der Statistik eine bedeutende Rolle, nicht zuletzt weil sie die Grenzverteilung des zentralen Grenzwertsatzes darstellt. Im eindimensionalen hat die Dichte die vertraute Form

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{1}{2}\left(\frac{x - \mu}{\sigma}\right)^2\right)$$

Für mehrdimensionale Normalverteilungen empfiehlt sich die Darstellung in vektorieller Form. Ein Zufallsvektor $\mathbf{X} = (X_1, \dots, X_p)$ heißt (nicht entartet) *multivariat normalverteilt*, $\mathbf{X} \sim N_p(\boldsymbol{\mu}, \Sigma)$, wenn die Dichte gegeben ist durch

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{p/2} |\Sigma|^{1/2}} \exp\left\{-\frac{1}{2}(\mathbf{x} - \boldsymbol{\mu})^T \Sigma^{-1}(\mathbf{x} - \boldsymbol{\mu})\right\},$$

wobei $\boldsymbol{\mu} = (\mu_1, \dots, \mu_p)^T = E(\mathbf{X})^T$ den Erwartungswert und $\Sigma = \text{cov}(\mathbf{X})$ die Kovarianzmatrix des Zufallsvektors \mathbf{X} bezeichnet. Vorausgesetzt ist, daß Σ positiv definit ist, $|\Sigma|$ bezeichnet die Determinante der Kovarianzmatrix.

Isodensiten

Eine wesentliche Bestimmungsgröße der multivariaten Normalverteilung ist die quadratische Form $(\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu})$. Allgemein spricht man von einer quadratischen Form, wenn man für variables \mathbf{x} und symmetrische Matrix \mathbf{A} die Form $\mathbf{x}^T \mathbf{A} \mathbf{x}$ betrachtet. Eine Matrix heißt positiv definit, wenn $\mathbf{x}^T \mathbf{A} \mathbf{x} > 0$ für alle $\mathbf{x} \neq \mathbf{0}$. Da für $\boldsymbol{\Sigma}$ vorausgesetzt wird, daß sie positiv definit ist (was dann auch für $\boldsymbol{\Sigma}^{-1}$ gilt), gilt für die interessierende quadratische Form

$$(\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu}) > 0.$$

Insbesondere läßt sich nun betrachten, für welche Werte \mathbf{x} die quadratische Form einen festen Wert annimmt. Die Isodensiten oder Höhenlinien sind diejenigen Kurven, die dieselbe Dichte besitzen. Für die Normalverteilung sind die Isodensiten Ellipsoide, die bestimmt sind durch

$$(\mathbf{x} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1}(\mathbf{x} - \boldsymbol{\mu}) = c^2,$$

wobei c eine beliebige Konstante ist.

In Abbildung 1.1, 1.3 und 1.5 sind beispielhaft die Dichten einer zweidimensionalen Normalverteilung dargestellt. In Abbildung ??, 1.4 und 1.6 sieht man die zugehörigen Isodensiten.

BEISPIEL 1.1: Zweidimensionale Normalverteilung

Mit $\mathbf{X} = (X_1, X_2)^T$, $\sigma_i^2 = \text{var}(X_i)$, $\rho = \rho(X_1, X_2)$ ist

$$\boldsymbol{\Sigma} = \begin{pmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{pmatrix} \text{ und } \boldsymbol{\Sigma}^{-1} = \begin{pmatrix} \frac{1}{\sigma_1^2(1-\rho^2)} & -\frac{\rho}{\sigma_1\sigma_2(1-\rho^2)} \\ -\frac{\rho}{\sigma_1\sigma_2(1-\rho^2)} & \frac{1}{\sigma_2^2(1-\rho^2)} \end{pmatrix}.$$

Einsetzen ergibt

$$f(x_1, x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \cdot \exp \left\{ -\frac{1}{2(1-\rho^2)} \left[\left(\frac{x_1 - \mu_1}{\sigma_1} \right)^2 - 2\rho \left(\frac{x_1 - \mu_1}{\sigma_1} \right) \left(\frac{x_2 - \mu_2}{\sigma_2} \right) + \left(\frac{x_2 - \mu_2}{\sigma_2} \right)^2 \right] \right\}$$

□

Lineare Transformation normalverteilter Merkmale

Eine wesentliche Eigenschaft der Normalverteilungen ist, daß der Verteilungstyp bei linearen Transformationen erhalten bleibt. Wird ein normalverteilter p -dimensionaler Zufallsvektor $\mathbf{X} \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ transformiert zu $\mathbf{Y} = \mathbf{A}\mathbf{X} + \mathbf{a}$, wobei \mathbf{A} eine feste $(q \times p)$ -Matrix ist und \mathbf{a} fester $(q \times 1)$ -Vektor, so gilt für den q dimensionalen Vektor $\mathbf{Y} \sim N(\mathbf{A}\boldsymbol{\mu}, \mathbf{A}\boldsymbol{\Sigma}\mathbf{A}^T)$.

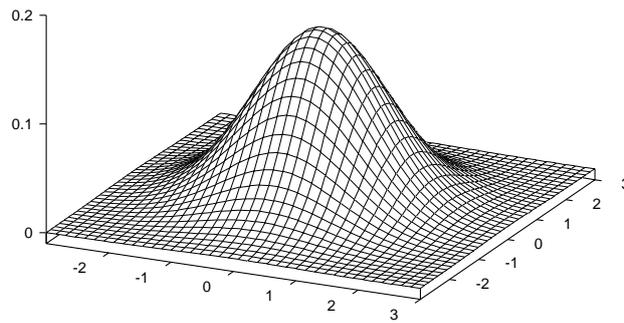


ABBILDUNG 1.1: Zweidimensionale Normalverteilungsdichte für unkorrelierte Merkmale, $\rho = 0$, mit $\mu_1 = \mu_2 = 0, \sigma_1 = \sigma_2 = 1.0$

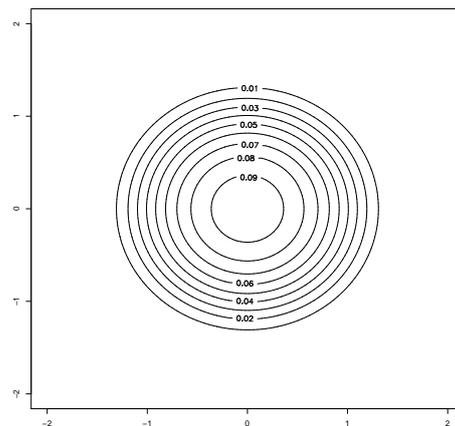


ABBILDUNG 1.2: Isodensiten zu Abbildung 1.1

Randverteilungen und bedingte Dichten

Insbesondere folgt aus der Typkonsistenz, daß alle eindimensionalen Randverteilungen auch Normalverteilungen sind, da sich die Komponenten X_i aus $X_i = (0, \dots, 0, 1, 0, \dots, 0)X$ ergibt, wobei die „1“ an der i ten Stelle steht. Die Aussage gilt auch genereller für beliebige Ränder. Man betrachte den Zufallsvektor $Z^T = (Y^T, X^T)$, $Y = (Y_1, \dots, Y_q)$, $X = (X_1, \dots, X_p)$, wobei

$$z = \begin{pmatrix} Y \\ X \end{pmatrix} \sim N_{q+p}(\mu, \Sigma) \quad \text{mit} \quad \mu = \begin{pmatrix} \mu_y \\ \mu_x \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \Sigma_y & \Sigma_{yx} \\ \Sigma_{xy} & \Sigma_x \end{pmatrix}.$$

Sind Σ_x und Σ_y positiv definit, so folgt speziell

$$X \sim N_p(\mu_x, \Sigma_x), Y \sim N_q(\mu_y, \Sigma_y)$$

d.h. die Randverteilungen sind wieder normalverteilt. Zwar sind die Ränder eines multivariat normalverteilten Zufallsvektors wieder normalverteilt, die Umkehrung

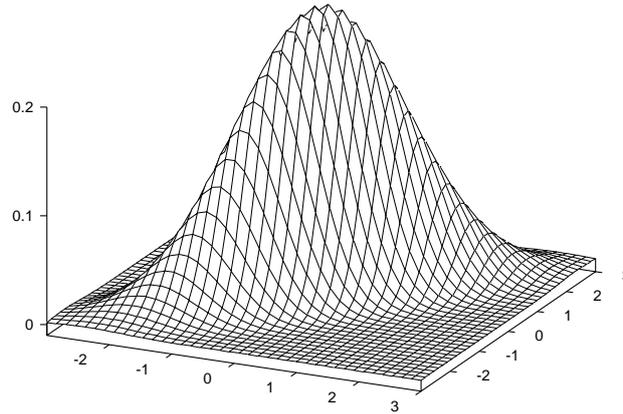


ABBILDUNG 1.3: Zweidimensionale Normalverteilungsdichte, $\rho = 0.8$, $\mu_1 = \mu_2 = 0$, $\sigma_1 = \sigma_2 = 1.0$

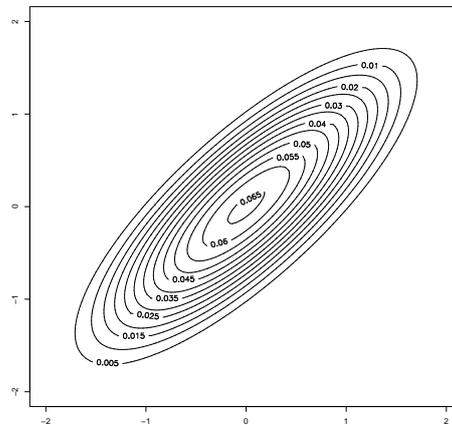


ABBILDUNG 1.4: Isodensiten zu Abbildung 1.3

hingegen gilt nicht. Aus der univariaten Normalverteilung von X_1, \dots, X_p läßt sich nicht schließen, daß $\mathbf{X} = (X_1, \dots, X_p)$ auch normalverteilt ist.

Normalverteilte Vektoren liefern auch normalverteilte bedingte Verteilungen. Gilt $\mathbf{Z}^T = (\mathbf{Y}^T, \mathbf{X}^T) \sim N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, so gilt für die bedingten Verteilungen von $\mathbf{Y}|\mathbf{x}$

$$\mathbf{Y} | \mathbf{x} \sim N_q(\boldsymbol{\mu}_{y|x}, \boldsymbol{\Sigma}_{y|x}),$$

wobei

$$\boldsymbol{\mu}_{y|x} = \boldsymbol{\mu}_y + \boldsymbol{\Sigma}_{yx} \boldsymbol{\Sigma}_x^{-1} (\mathbf{x} - \boldsymbol{\mu}_x),$$

$$\boldsymbol{\Sigma}_{y|x} = \boldsymbol{\Sigma}_y - \boldsymbol{\Sigma}_{yx} \boldsymbol{\Sigma}_x^{-1} \boldsymbol{\Sigma}_{xy}.$$

Daraus ergibt sich insbesondere, dass der bedingte Erwartungswert $\boldsymbol{\mu}_{y|x}$ linear in \mathbf{x} ist, man erhält

$$\boldsymbol{\mu}_{y|x} = \mathbf{b}_0 + \mathbf{B}\mathbf{x},$$

