$\langle$ S t a t W t h 1 7 $\rangle$

5. Multivariate Verteilungen

Werner Gurker

Institut für Stochastik und Wirtschaftsmathematik

Technische Universität Wien

Multivariate Verteilungen: Beispiele

Zur Beschreibung von Zufallsexperimenten benötigt man häufig mehrere stochastische Größen. Beispiele:

1. Wir wählen zufällig $n = 10$ Hörer_innen dieser LVA und messen ihre Körpergrößen.

   Das Experiment wird durch eine Folge $X_1, X_2, \ldots, X_n$ von sGn beschrieben.

2. Wir werfen wiederholt eine Münze. Sei $X_i = 1$, wenn der i–te Wurf ein „Kopf" ist, und $X_i = 0$ im anderen Fall.

   Das Experiment lässt sich durch eine Folge $X_1, X_2, \ldots$ von Bernoulli–Größen beschreiben.

3. Wir wählen zufällig eine Person aus einer großen Population und messen ihr Körpergewicht X und ihre Körpergröße Y.

   Das Experiment wird durch den Vektor (X, Y) beschrieben.

Multivariate Verteilungen: Beispiele (Forts.)

Wie lässt sich das Verhalten dieser sGn beschreiben?

Die Spezifikation der einzelnen Verteilungen allein genügt nicht, wir müssen auch den Zusammenhang – oder auch das Fehlen desselben – zwischen den einzelnen Größen beschreiben.

Wenn beispielsweise im dritten Experiment Y groß ist, dann ist sehr wahrscheinlich auch X groß.

Andererseits, in den ersten beiden Experimenten kann man davon ausgehen, dass die einzelnen Größen unabhängig sind.

D. h., wissen wir etwas über eine dieser Größen, so haben wir damit keine zusätzliche Information über die anderen.

M. a. W., wir benötigen die gemeinsame Verteilung der sGn.

Bivariate stochastische Vektoren

Gegeben sei ein Zufallsexperiment mit Merkmalraum Ω.

Man betrachte zwei sGn X_1 und X_2, die jedem Element $\omega \in \Omega$ eine reelle Zahl zuordnen:

$$X_1(\omega) = x_1 \quad \text{und} \quad X_2(\omega) = x_2$$

Dann nennt man (X_1, X_2) einen (2–dimensionalen) stochastischen Vektor (kurz sV) mit dem Merkmalraum:

$$M = \{(x_1, x_2) \,|\, x_1 = X_1(\omega),\ x_2 = X_2(\omega),\ \omega \in \Omega\}$$

Häufig schreibt man:

$$\mathbf{X} = (X_1, X_2)' \quad (= \text{transponierter Zeilenvektor})$$

Verteilungsfunktion

Teilmengen $B \subseteq M$ des Merkmalraums nennt man Ereignisse.

Die Wahrscheinlichkeit $P(\mathbf{X} \in B)$ für den Eintritt von $B \subseteq M$ lässt sich durch die (2-dimensionale) Verteilungsfunktion charakterisieren:

$$F(x_1, x_2) = P(X_1 \leq x_1,\, X_2 \leq x_2)$$

Dabei handelt es sich um eine Funktion von $\mathbb{R}^2$ nach $[0, 1]$:

$$F : (x_1, x_2) \in \mathbb{R}^2 \quad \longmapsto \quad F(x_1, x_2) \in [0, 1]$$

Bem: Der Ausdruck $P(X_1 \leq x_1,\, X_2 \leq x_2)$ ist eine Kurzschreibweise für:

$$P\left(\{X_1 \leq x_1\} \cap \{X_2 \leq x_2\}\right)$$

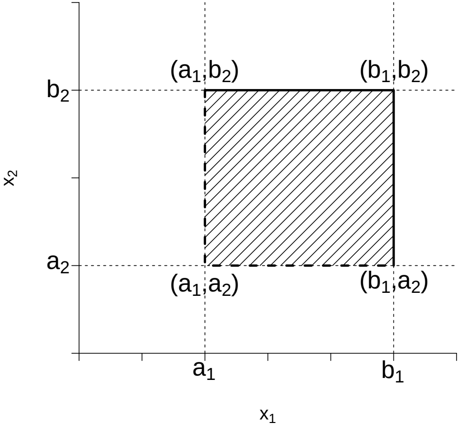
Wahrscheinlichkeit für $(a_1, b_1] \times (a_2, b_2]$

$$P(a_1 < X_1 \leq b_1,\ a_2 < X_2 \leq b_2)$$

$$= F(b_1, b_2)$$

$$- F(a_1, b_2) - F(b_1, a_2)$$

$$+ F(a_1, a_2)$$



Diskrete stochastische Vektoren: W–Funktion

Ist der Merkmalraum M ($\subseteq \mathbb{R}^2$) von $\mathbf{X} = (X_1, X_2)'$ endlich oder abzählbar, handelt es sich um einen diskreten stochastischen Vektor.

Die (gemeinsame) W–Funktion ist gegeben durch:

$$p(x_1, x_2) = P(X_1 = x_1,\ X_2 = x_2) \quad \text{für alle} \quad (x_1, x_2) \in M$$

Die W–Funktion hat die folgenden Eigenschaften:

(1) $0 \leq p(x_1, x_2) \leq 1,\ (x_1, x_2) \in M$

(2) $\displaystyle\sum_{(x_1,x_2)\in M} p(x_1, x_2) = 1$

(3) $\displaystyle P((X_1, X_2) \in B) = \sum_{(x_1,x_2)\in B} p(x_1, x_2) \quad \text{für} \quad B \subseteq M$

Randverteilungen

Die Elemente X_1 und X_2 eines stochastischen Vektors $\mathbf{X} = (X_1, X_2)'$ sind selbst (1–dimensionale) stochastische Größen.

Wie bestimmt man ihre Verteilungen (d. h. W–Funktionen) ?

Die W–Funktionen von X_1 und X_2 sind gegeben durch:

$$X_1 : \; p_1(x_1) = P(X_1 = x_1) = \sum_{x_2} p(x_1, x_2)$$

$$X_2 : \; p_2(x_2) = P(X_2 = x_2) = \sum_{x_1} p(x_1, x_2)$$

Diese Verteilungen nennt man die Randverteilungen von $\mathbf{X} = (X_1, X_2)'$.

Beispiel

In einem Behälter befinden sich drei Würfel.

Würfel 1 ist ein üblicher Würfel.

Würfel 2 hat keine Augenzahl 6, dafür zwei Seiten mit der Augenzahl 5.

Würfel 3 hat keine Augenzahl 5, dafür zwei Seiten mit der Augenzahl 6.

Zufallsexperiment: Wähle zufällig einen Würfel und werfe anschließend den gewählten Würfel.

X_1 = Nummer des gewählten Würfels

X_2 = Geworfene Augenzahl des gewählten Würfels

Gemeinsame Verteilung? Randverteilungen?

Beispiel: Gemeinsame Verteilung

X_2

	$p(x_1, x_2)$	1	2	3	4	5	6
	1	$\frac{1}{18}$	$\frac{1}{18}$	$\frac{1}{18}$	$\frac{1}{18}$	$\frac{1}{18}$	$\frac{1}{18}$
X_1	2	$\frac{1}{18}$	$\frac{1}{18}$	$\frac{1}{18}$	$\frac{1}{18}$	$\frac{1}{9}$	0
	3	$\frac{1}{18}$	$\frac{1}{18}$	$\frac{1}{18}$	$\frac{1}{18}$	0	$\frac{1}{9}$

Beispiel: Gemeinsame Verteilung u. Randverteilungen

X_2

	$p(x_1, x_2)$	1	2	3	4	5	6	$p_1(x_1)$
	1	$\frac{1}{18}$	$\frac{1}{18}$	$\frac{1}{18}$	$\frac{1}{18}$	$\frac{1}{18}$	$\frac{1}{18}$	$\frac{1}{3}$
X_1	2	$\frac{1}{18}$	$\frac{1}{18}$	$\frac{1}{18}$	$\frac{1}{18}$	$\frac{1}{9}$	0	$\frac{1}{3}$
	3	$\frac{1}{18}$	$\frac{1}{18}$	$\frac{1}{18}$	$\frac{1}{18}$	0	$\frac{1}{9}$	$\frac{1}{3}$
	$p_2(x_2)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	1

Bem: Aus den beiden Randverteilungen von X_1 und X_2 allein kann die gemeinsame Verteilung von (X_1, X_2) *nicht* rekonstruiert werden. Grund: X_1 und X_2 sind nicht „unabhängig“.

Stetige stochastische Vektoren

Ist die Verteilungsfunktion $F(x_1, x_2)$ eines stochastischen Vektors (X_1, X_2) eine stetige Funktion, spricht man von einem stetigen sV.

Meist lässt sich die VF eines stetigen sVs wie folgt darstellen:

$$F(x_1, x_2) = \int_{-\infty}^{x_2} \int_{-\infty}^{x_1} f(w_1, w_2)\, dw_1\, dw_2 \quad \text{für} \quad (x_1, x_2) \in \mathbb{R}^2$$

f nennt man die (gemeinsame) Dichtefunktion (kurz Dichte) von (X_1, X_2).

Analog zum 1–dimensionalen Fall gilt an den Stetigkeitspunkten von f:

$$\frac{\partial^2 F(x_1, x_2)}{\partial x_1 \partial x_2} = f(x_1, x_2)$$

Eigenschaften einer 2–dimensionalen Dichte

(1) $f(x_1, x_2) \geq 0 \quad \text{für} \quad (x_1, x_2) \in M$

(2) $$\iint\limits_M f(x_1, x_2)\, dx_1 dx_2 = 1$$

(3) $$P((X_1, X_2) \in B) = \iint\limits_B f(x_1, x_2)\, dx_1 dx_2 \quad \text{für} \quad B \subseteq M$$

D. h., $P((X_1, X_2) \in B)$ entspricht dem Volumen unter der Fläche $z = f(x_1, x_2)$ über der Menge B.

Die Menge alle Punkte $(x_1, x_2) \in M$ mit $f(x_1, x_2) > 0$ bilden den Träger von (X_1, X_2).

Randdichten

Analog zum diskreten Fall bestimmt man die Randdichten von X_1 bzw. X_2 aus der gemeinsamen Dichte $f(x_1, x_2)$ von (X_1, X_2) wie folgt:

$$X_1: \; f_1(x_1) = \int_{-\infty}^{\infty} f(x_1, x_2)\, dx_2 \qquad X_2: \; f_2(x_2) = \int_{-\infty}^{\infty} f(x_1, x_2)\, dx_1$$

Um die Randdichte von X_1 zu bestimmen, ist die gemeinsame Dichte $f(x_1, x_2)$ über x_2 zu integrieren.

Um die Randdichte von X_2 zu bestimmen, ist die gemeinsame Dichte $f(x_1, x_2)$ über x_1 zu integrieren.

Beispiel

$$f(x_1, x_2) = \begin{cases} 6\, x_1^2\, x_2 & 0 < x_1 < 1,\ 0 < x_2 < 1 \\ & \\ 0 & \text{sonst} \end{cases}$$

Wahrscheinlichkeit des Ereignisses $\{0 < X_1 < 3/4\} \cap \{1/3 < X_2 < 2\}$?

$$P\left(0 < X_1 < \frac{3}{4}, \frac{1}{3} < X_2 < 2\right) = \int_{1/3}^{2} \int_{0}^{3/4} f(x_1, x_2)\, dx_1\, dx_2 = \frac{3}{8}$$

[Skriptum S. 189]

Beispiel (Forts.)

Träger von (X_1, X_2): $\quad (0,1) \times (0,1)$

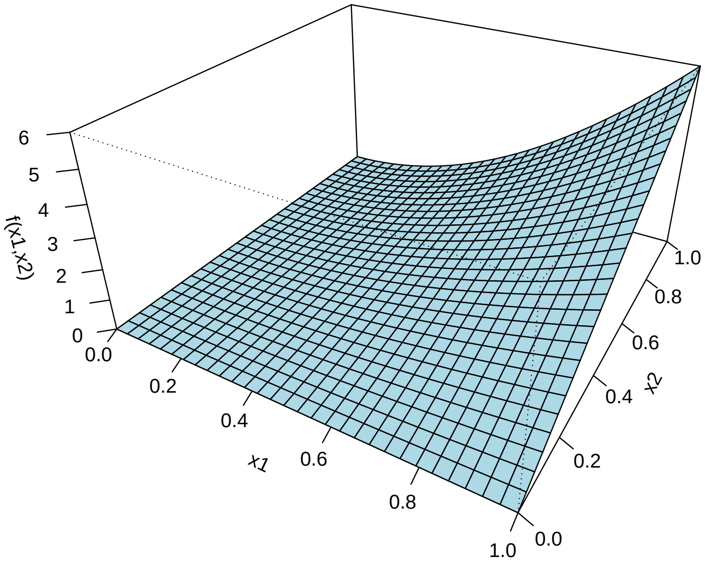
Randdichte von X_1:

$$f_1(x_1) = \int_0^1 6x_1^2 x_2 \, dx_2 = 6x_1^2 \left[\frac{x_2^2}{2} \right]_0^1 = 3x_1^2 \quad \text{für} \quad 0 < x_1 < 1$$

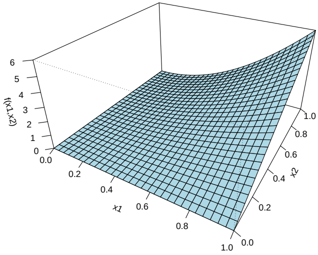
Randdichte von X_2:

$$f_2(x_2) = \int_0^1 6x_1^2 x_2 \, dx_1 = 6x_2 \left[\frac{x_1^3}{3} \right]_0^1 = 2x_2 \quad \text{für} \quad 0 < x_2 < 1$$

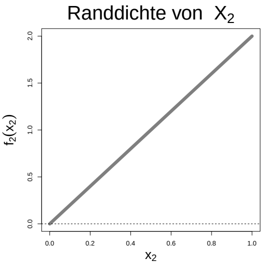
Beispiel (Forts.)



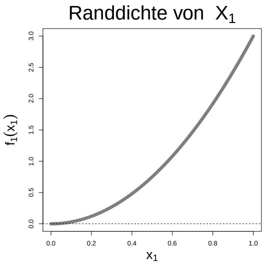
Beispiel (Forts.)



$\longrightarrow$



$\downarrow$



Erwartungswert

Sei (X_1, X_2) ein sV und $g : \mathbb{R}^2 \longrightarrow \mathbb{R}$ eine reellwertige Funktion.

Dann ist $Y = g(X_1, X_2)$ eine (1–dimensionale) sG und existiert ihr Erwartungswert, so ist er gegeben durch:

$$\text{diskret:} \quad \mathbb{E}(Y) = \sum_{x_1} \sum_{x_2} g(x_1, x_2) p(x_1, x_2)$$

$$\text{stetig:} \quad \mathbb{E}(Y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x_1, x_2) f(x_1, x_2) \, dx_1 \, dx_2$$

Dabei wird der (mehrdimensionale) LotUS verwendet.

Beispiel (Forts. von Bsp [15])

Den Erwartungswert von $Y = X_1/X_2$ berechnet man wie folgt:

$$\mathbb{E}(Y) = \mathbb{E}\left(\frac{X_1}{X_2}\right) = \int_0^1\int_0^1 \left(\frac{x_1}{x_2}\right) 6x_1^2 x_2 \, dx_1 dx_2$$

$$= \int_0^1\int_0^1 6x_1^3 \, dx_1 dx_2 = \int_0^1 6\left[\frac{x_1^4}{4}\right]_0^1 dx_2$$

$$= \int_0^1 \frac{3}{2} \, dx_2 = \frac{3}{2}$$

Der Erwartungswert ist ein linearer Operator

(X_1, X_2) sei ein sV und $Y_1 = g_1(X_1, X_2)$ und $Y_2 = g_2(X_1, X_2)$ seien zwei (1–dim.) sGn, deren Erwartungswerte existieren.

Dann gilt für Konstanten k_1 und k_2:

$$\mathbb{E}(k_1 Y_1 + k_2 Y_2) = k_1 \mathbb{E}(Y_1) + k_2 \mathbb{E}(Y_2)$$

Verallgemeinerung: Für m (1–dim.) sGn $Y_j = g_j(X_1, X_2)$, $j = 1, \ldots, m$, und Konstanten k_j gilt:

$$\mathbb{E}\left(\sum_{j=1}^{m} k_j Y_j\right) = \sum_{j=1}^{m} k_j \mathbb{E}(Y_j)$$

Erwartungswert des stoch. Vektors $\mathbf{X} = (X_1, X_2)'$

Der Erwartungswert von $\mathbf{X} = (X_1, X_2)'$ wird definiert durch:

$$\mathbb{E}(\mathbf{X}) = \begin{bmatrix} \mathbb{E}(X_1) \\ \mathbb{E}(X_2) \end{bmatrix} \in \mathbb{R}^2$$

Dabei wird vorausgesetzt, dass die Erwartungswerte von X_1 und X_2 existieren.

Bedingte Verteilungen

Bisher haben wir uns mit der gemeinsamen Verteilung von (X_1, X_2) und den Randverteilungen von X_1 und X_2 beschäftigt.

Häufig kennt man den Wert einer Variablen – d. h. man kennt $X_1 = x_1$ oder $X_2 = x_2$ – und es stellt sich die Frage, welche Auswirkungen sich dadurch für die Verteilung der anderen Variablen ergeben.

Dies führt zum Konzept der bedingten Verteilung.

Bedingte Verteilung: Diskreter Fall

(X_1, X_2) sei ein diskreter sV mit W–Funktion $p(x_1, x_2)$ und $p_1(x_1)$ sei die Randverteilung von X_1.

Sei $x_1 \in S_1$ ein Punkt aus dem Träger S_1 von X_1 (d. h. $p_1(x_1) > 0$).

Dann gilt nach Definition der bedingten Wahrscheinlichkeit:

$$P(X_2 = x_2 \,|\, X_1 = x_1) = \frac{P(X_1 = x_1, X_2 = x_2)}{P(X_1 = x_1)} = \frac{p(x_1, x_2)}{p_1(x_1)}$$

Man nennt: $p(x_2|x_1) := \dfrac{p(x_1, x_2)}{p_1(x_1)}$ für festes $x_1 \in S_1$

die durch $X_1 = x_1$ bedingte W–Funktion von X_2.

Analoge Ausdrücke gelten für $p(x_1|x_2)$, d. h. für die durch $X_2 = x_2$ bedingte W–Funktion von X_1.

Bedingter Erwartungswert (diskreter Fall)

Der durch $X_1 = x_1$ bedingte Erwartungswert von X_2 ist gegeben durch:

$$\mathbb{E}(X_2|x_1) = \sum_{x_2} x_2 \, p(x_2|x_1)$$

Ist $u(X_2)$ eine Funktion von X_2, so ist der durch $X_1 = x_1$ bedingte Erwartungswert von $u(X_2)$ gegeben durch:

$$\mathbb{E}\big[u(X_2)\big|x_1\big] = \sum_{x_2} u(x_2) \, p(x_2|x_1)$$

Bedingte Verteilung: Stetiger Fall

(X_1, X_2) sei ein stetiger sV mit Dichte $f(x_1, x_2)$ und $f_1(x_1)$ sei die Randdichte von X_1.

Sei $x_1 \in S_1$ ein Punkt aus dem Träger S_1 von X_1 (d. h. $f_1(x_1) > 0$).

Dann ist die durch $X_1 = x_1$ bedingte Dichte von X_2 definiert durch:

$$f(x_2|x_1) = \frac{f(x_1, x_2)}{f_1(x_1)} \quad \text{für festes } x_1 \in S_1$$

Analoge Ausdrücke gelten für $f(x_1|x_2)$, d. h. für die durch $X_2 = x_2$ bedingte Dichte von X_1.

Bedingte Dichten haben alle Eigenschaften einer Dichtefunktion.

Bedingter Erwartungswert (stetiger Fall)

Der durch $X_1 = x_1$ bedingte Erwartungswert von X_2 ist gegeben durch:

$$\mathbb{E}(X_2|x_1) = \int_{-\infty}^{\infty} x_2 f(x_2|x_1)\, dx_2$$

Ist $u(X_2)$ eine Funktion von X_2, so ist der durch $X_1 = x_1$ bedingte Erwartungswert von $u(X_2)$ gegeben durch:

$$\mathbb{E}\big[u(X_2)\big|x_1\big] = \int_{-\infty}^{\infty} u(x_2) f(x_2|x_1)\, dx_2$$

Bedingte Varianz (allgemein)

Die durch $X_1 = x_1$ bedingte Varianz von X_2 ist definiert durch:

$$\mathrm{Var}(X_2|x_1) = \mathbb{E}\left\{\left[X_2 - \mathbb{E}(X_2|x_1)\right]^2 \middle| x_1\right\}$$

Sie lässt sich mittels Verschiebungssatz auch wie folgt berechnen:

$$\mathrm{Var}(X_2|x_1) = \mathbb{E}(X_2^2|x_1) - \left[\mathbb{E}(X_2|x_1)\right]^2$$

Beispiel

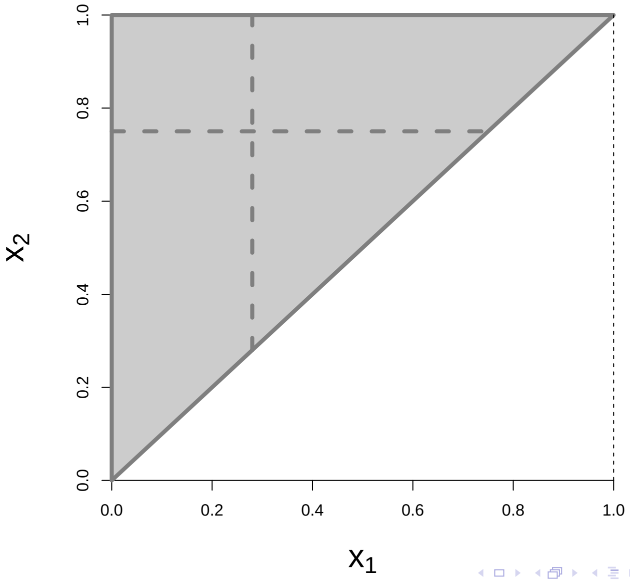
Die gemeinsame Dichte von (X_1, X_2) sei gegeben durch:

$$f(x_1, x_2) = \begin{cases} 2 & 0 < x_1 < x_2 < 1 \\ 0 & \text{sonst} \end{cases}$$

Bei der Bestimmung der Randdichten ist zu beachten, dass der Träger von (X_1, X_2) nicht rechtecksförmig ist.

In der folgenden Abb. des Trägers (grau) von (X_1, X_2) sind auch exemplarisch zwei Integrationswege eingezeichnet (strichliert).

Beispiel (Forts.)



Beispiel (Forts.)

Randdichten:

$$f_1(x_1) = \int_{x_1}^{1} 2\, dx_2 = 2(1 - x_1) \quad \text{für} \quad 0 < x_1 < 1$$

$$f_2(x_2) = \int_{0}^{x_2} 2\, dx_1 = 2x_2 \quad \text{für} \quad 0 < x_2 < 1$$

Beispiel (Forts.)

Bedingte Dichten:

$$f(x_1|x_2) = \frac{f(x_1, x_2)}{f_2(x_2)} = \frac{2}{2x_2} = \frac{1}{x_2} \quad \text{für} \quad 0 < x_1 < x_2 \quad (x_2 \text{ fest})$$

$$\longrightarrow \quad \mathrm{U}(0, x_2) \text{ –Verteilung}$$

$$f(x_2|x_1) = \frac{f(x_1, x_2)}{f_1(x_1)} = \frac{2}{2(1 - x_1)} = \frac{1}{1 - x_1} \quad \text{für} \quad x_1 < x_2 < 1 \quad (x_1 \text{ fest})$$

$$\longrightarrow \quad \mathrm{U}(x_1, 1) \text{ –Verteilung}$$

Beispiel (Forts.)

Bedingte Erwartungswerte:

$$\mathbb{E}(X_1|x_2) = \frac{x_2}{2}, \qquad \mathbb{E}(X_2|x_1) = \frac{1-x_1}{2}$$

Bedingte Varianzen:

$$\mathrm{Var}(X_1|x_2) = \frac{x_2^2}{12}, \qquad \mathrm{Var}(X_2|x_1) = \frac{(1-x_1)^2}{12}$$

Kovarianz

Bem: Wir schreiben im Folgenden X (statt X_1) und Y (statt X_2)

Mittelwerte: $\mu_1 = \mathbb{E}(X), \quad \mu_2 = \mathbb{E}(Y)$

Varianzen: $\sigma_1^2 = \mathrm{Var}(X), \quad \sigma_2^2 = \mathrm{Var}(Y)$

Die Kovarianz von X und Y ist definiert durch:

$$\sigma_{12} = \mathrm{Cov}(X, Y) = \mathbb{E}\big[(X - \mu_1)(Y - \mu_2)\big]$$

Verschiebungssatz: $\quad \mathrm{Cov}(X, Y) = \mathbb{E}(XY) - \mathbb{E}(X)\mathbb{E}(Y)$

Spezialfall: $\quad \mathrm{Cov}(X, X) = \mathrm{Var}(X) = \mathbb{E}(X^2) - \big[\mathbb{E}(X)\big]^2$

Korrelation

Der Korrelationskoeffizient von X und Y ist definiert durch:

$$\rho = \rho_{XY} = \frac{\mathrm{Cov}(X, Y)}{\sqrt{\mathrm{Var}(X)}\,\sqrt{\mathrm{Var}(Y)}}$$

Eigenschaften:

(1) Es gilt: $-1 \leq \rho \leq 1 \quad \Longleftrightarrow \quad |\rho| \leq 1$

(2) Im Grenzfall $\rho = \pm 1$ (oder $|\rho| = 1$) gilt: $P(Y = a + bX) = 1$

D. h., die gesamte Wahrscheinlichkeitsverteilung von (X, Y) konzentriert sich für $|\rho| = 1$ auf einer Geraden.

Für $\rho = 1$ gilt $b = \dfrac{\sigma_2}{\sigma_1} > 0$; für $\rho = -1$ gilt $b = -\dfrac{\sigma_2}{\sigma_1} < 0$.

Beispiel [29]

$$\mathbb{E}(X) = \int_0^1 2x(1-x)\,dx = \frac{1}{3}, \quad \mathbb{E}(Y) = \int_0^1 2y^2\,dy = \frac{2}{3}$$

$$\mathrm{Var}(X) = \mathrm{Var}(Y) = \frac{1}{18}$$

$$\mathbb{E}(XY) = \int_0^1 \int_0^y 2xy\,dxdy = \frac{1}{4}$$

$$\mathrm{Cov}(X,Y) = \frac{1}{4} - \left(\frac{1}{3}\right)\left(\frac{2}{3}\right) = \frac{1}{36}$$

$$\Longrightarrow \quad \rho = \frac{1/36}{\sqrt{1/18}\,\sqrt{1/18}} = \frac{1/36}{1/18} = \frac{1}{2}$$

Interpretation von ρ

Der Korrelationskoeffizient ρ lässt sich als (dimensionsloses) Maß für den linearen Zusammenhang zwischen X und Y interpretieren.

Bsp: Dichte von X: $f(x) = \frac{1}{2} I_{(-1,1)}(x)$

$$\Longrightarrow \quad \mathbb{E}(X) = 0, \quad \mathbb{E}(X^2) = \frac{1}{3}, \quad \mathbb{E}(X^3) = 0$$

Mit $Y := X^2$ gilt:

$$\mathrm{Cov}(X, Y) = \mathrm{Cov}(X, X^2) = \mathbb{E}(X^3) - \mathbb{E}(X)\mathbb{E}(X^2) = 0$$

$$\Longrightarrow \quad \rho_{XY} = 0 \quad (\text{obwohl } Y = X^2 \text{ !})$$

Unabhängigkeit: Motivation

Die gemeinsame Dichte des stochastischen Vektors (X_1, X_2) sei $f(x_1, x_2)$.

Die beiden Randdichten seien $f_1(x_1)$ bzw. $f_2(x_2)$.

Aus der Definition der bedingten Dichte $f(x_2|x_1)$ folgt, dass sich die gemeinsame Dichte wie folgt schreiben lässt:

$$f(x_1, x_2) = f(x_2|x_1)\, f_1(x_1)$$

Hängt die bedingte Dichte $f(x_2|x_1)$ nicht von x_1 ab, d. h. gilt:

$$f(x_2|x_1) = f_2(x_2) \quad \text{für alle } x_1$$

sagt man, dass X_1 und X_2 unabhängig sind.

Unabhängigkeit: Definition

Die (gemeinsame) Dichte (bzw. W–Funktion) von (X_1, X_2) sei $f(x_1, x_2)$ (bzw. $p(x_1, x_2)$).

Die Randdichten (bzw. Randverteilungen) seien $f_1(x_1)$ (bzw. $p_1(x_1)$) und $f_2(x_2)$ (bzw. $p_2(x_2)$).

Die sGn X_1 und X_2 sind (stochastisch) unabhängig (kurz ua.), wenn:

$$\text{stetig:} \quad f(x_1, x_2) \equiv f_1(x_1)\, f_2(x_2)$$

$$\text{diskret:} \quad p(x_1, x_2) \equiv p_1(x_1)\, p_2(x_2)$$

Nicht unabhängige sGn sind (stochastisch) abhängig.

Beispiel 1

Die gemeinsame Dichte von X_1 und X_2 sei gegeben durch:

$$f(x_1, x_2) = \begin{cases} 4x_1x_2 & 0 < x_1 < 1,\ 0 < x_2 < 1 \\ 0 & \text{sonst} \end{cases}$$

Randdichten:

$$f_1(x_1) = \int_0^1 4x_1x_2\, dx_2 = 2x_1 \quad \text{für} \quad 0 < x_1 < 1$$

$$f_2(x_2) = \int_0^1 4x_1x_2\, dx_1 = 2x_2 \quad \text{für} \quad 0 < x_2 < 1$$

Da $f(x_1, x_2) \equiv f_1(x_1)f_2(x_2)$, sind X_1 und X_2 unabhängig.

Beispiel 2

Die gemeinsame Dichte von X_1 und X_2 sei gegeben durch:

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

Randdichten:

$$f_1(x_1) = \int_0^1 (x_1 + x_2)\, dx_2 = x_1 + \frac{1}{2} \quad \text{für} \quad 0 < x_1 < 1$$

$$f_2(x_2) = \int_0^1 (x_1 + x_2)\, dx_1 = x_2 + \frac{1}{2} \quad \text{für} \quad 0 < x_2 < 1$$

Da $f(x_1, x_2) \not\equiv f_1(x_1) f_2(x_2)$, sind X_1 und X_2 abhängig.

Zwei Behauptungen

Behauptung 1: Die gemeinsame Verteilungsfunktion von (X_1, X_2) sei $F(x_1, x_2)$ und die Verteilungsfunktionen von X_1 bzw. X_2 seien $F_1(x_1)$ bzw. $F_2(x_2)$. Dann sind X_1 und X_2 genau dann unabhängig, wenn:

$$F(x_1, x_2) = F_1(x_1)\, F(x_2) \quad \text{für alle} \quad (x_1, x_2) \in \mathbb{R}^2$$

Behauptung 2: Existieren die Erwartungswerte $\mathbb{E}[u(X_1)]$ und $\mathbb{E}[v(X_2)]$ für zwei Funktionen u und v und sind X_1 und X_2 unabhängig, dann gilt:

$$\mathbb{E}[u(X_1)\, v(X_2)] = \mathbb{E}[u(X_1)]\, \mathbb{E}[v(X_2)]$$

Folgerung

X und Y seien zwei sGn mit den Mittelwerten μ_1 und μ_2 und den Varianzen $\sigma_1^2 > 0$ und $\sigma_2^2 > 0$. Dann folgt aus der Unabhängigkeit von X und Y auch die Unkorreliertheit:

$$X,\ Y \text{ ua.} \quad \Longrightarrow \quad \rho_{XY} = 0$$

Achtung: Die Umkehrung gilt nicht, d. h., aus der Unkorreliertheit folgt i. A. nicht die Unabhängigkeit (vgl. Bsp [37]).

Beweis: Es genügt zu zeigen, dass die Kovarianz gleich Null ist:

$$\mathrm{Cov}(X,Y) = \mathbb{E}\big[(X-\mu_1)(Y-\mu_2)\big] = \mathbb{E}(X-\mu_1)\,\mathbb{E}(Y-\mu_2) = 0$$

Dabei wird Behauptung 2 verwendet.

Mehrdimensionale Erweiterungen

Die für zwei sGn entwickelten Konzepte lassen sich unschwer auf mehrere sGn erweitern. (Siehe Skriptum S. 203f.)

Insbesondere definiert man: Die sGn $X_1, X_2, \ldots, X_n$ sind (stochastisch) unabhängig, wenn:

diskret: $$p(x_1, x_2, \ldots, x_n) \equiv p_1(x_1)\, p_2(x_2) \cdots p_n(x_n)$$

stetig: $$f(x_1, x_2, \ldots, x_n) \equiv f_1(x_1)\, f_2(x_2) \cdots f_n(x_n)$$

allgemein: $$F(x_1, x_2, \ldots, x_n) \equiv F_1(x_1)\, F_2(x_2) \cdots F_n(x_n)$$

Sind die sGn $X_1, X_2, \ldots, X_n$ unabhängig mit identischer Verteilung, nennt man sie uiv oder iid (= *independent and identically distributed*).

Varianz–Kovarianzmatrix

$\mathbf{X} = (X_1, X_2, \ldots, X_n)'$ sei ein stochastischer Vektor.

Der Erwartungswert von $\mathbf{X}$ ist der Vektor der Erwartungswerte:

$$\boldsymbol{\mu} = \mathbb{E}(\mathbf{X}) = \left(\mathbb{E}(X_1), \mathbb{E}(X_2), \ldots, \mathbb{E}(X_n)\right)'$$

Varianz–Kovarianzmatrix:

$$\mathrm{Cov}(\mathbf{X}) = \mathbb{E}\left[(\mathbf{X} - \boldsymbol{\mu})(\mathbf{X} - \boldsymbol{\mu})'\right] = [\sigma_{ij}]$$

$$\sigma_{ii} = \sigma_i^2 = \mathrm{Var}(X_i) \ \ldots \ \text{Varianz von } X_i$$

$$\sigma_{ij} = \mathrm{Cov}(X_i, X_j) \ \ldots \ \text{Kovarianz von } X_i \text{ und } X_j$$

Varianz–Kovarianzmatrix (Forts.)

Ausführlich geschrieben lautet $\mathrm{Cov}(\mathbf{X})$ wie folgt:

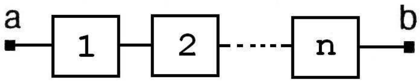
$$\mathrm{Cov}(\mathbf{X}) = \begin{bmatrix} \mathrm{Var}(X_1) & \mathrm{Cov}(X_1, X_2) & \cdots & \mathrm{Cov}(X_1, X_n) \\ \mathrm{Cov}(X_2, X_1) & \mathrm{Var}(X_2) & \cdots & \mathrm{Cov}(X_2, X_n) \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{Cov}(X_n, X_1) & \mathrm{Cov}(X_n, X_2) & \cdots & \mathrm{Var}(X_n) \end{bmatrix}$$

Spezialfall $n = 2$:

$$\mathrm{Cov}(\mathbf{X}) = \begin{bmatrix} \mathrm{Var}(X_1) & \mathrm{Cov}(X_1, X_2) \\ \mathrm{Cov}(X_2, X_1) & \mathrm{Var}(X_2) \end{bmatrix} = \begin{bmatrix} \sigma_1^2 & \rho\sigma_1\sigma_2 \\ \rho\sigma_1\sigma_2 & \sigma_2^2 \end{bmatrix}$$

Minimum

Die sGn $X_1, X_2, \ldots, X_n$ seien Lebensdauern von Komponenten in einem Seriensystem:



Das Seriensystem fällt aus, sobald die erste Komponente ausfällt.

Die Lebensdauer Y_1 des Seriensystems ist also das Minimum der Lebensdauern der Komponenten:

$$Y_1 = \min\{X_1, X_2, \ldots, X_n\}$$

Minimum (Forts.)

Sind die sGn X_i, $i = 1, 2, \ldots, n$, unabhängig, so ist die Verteilungsfunktion von Y_1 gegeben durch:

$$\begin{aligned} F_{\min}(y) = P(Y_1 \leq y) &= 1 - P(Y_1 > y) \\ &= 1 - P(X_1 > y, X_2 > y, \ldots, X_n > y) \\ = 1 - \prod_{i=1}^{n} P(X_i > y) &= 1 - \prod_{i=1}^{n} \left[1 - P(X_i \leq y)\right] \end{aligned}$$

Bezeichnet F_i die Verteilungsfunktion von X_i, so gilt:

$$F_{\min}(y) = 1 - \prod_{i=1}^{n} \left[1 - F_i(y)\right]$$

Beispiel

Für $X_1, X_2, \ldots, X_n$ iid Exp(λ) (oder Exp(τ)) ist $F_{\min}$ gegeben durch:

$$F_{\min}(y) = 1 - \prod_{i=1}^{n} \left[1 - (1 - e^{-\lambda y})\right] = 1 - e^{-n\lambda y} \quad \text{für} \quad y > 0$$

$$\Longrightarrow \quad Y_1 \sim \mathrm{Exp}(n\lambda) \quad \text{oder} \quad Y_1 \sim \mathrm{Exp}(\tau/n)$$

Die Dichte von Y_1 lautet:

$$f_{\min}(y) = F'_{\min}(y) = n\lambda\, e^{-n\lambda y} \quad \text{für} \quad y > 0$$

Erwartungswert: $\mathbb{E}(Y_1) = \dfrac{1}{n\lambda} = \dfrac{\tau}{n}$

Beispiel (Forts.)

Für $X_i \sim \mathrm{Exp}(\lambda_i)$, $i = 1, 2, \ldots, n$, ua., ist $F_{\min}$ gegeben durch:

$$F_{\min}(y) = 1 - e^{-\widetilde{\lambda} y} \quad \text{mit} \quad \widetilde{\lambda} = \sum_{i=1}^{n} \lambda_i$$

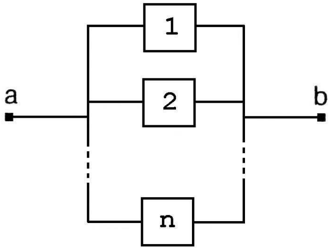
$$\Longrightarrow \quad Y_1 \sim \mathrm{Exp}(\widetilde{\lambda})$$

Dichte von Y_1: $\quad f_{\min}(y) = F'_{\min}(y) = \widetilde{\lambda}\, e^{-\widetilde{\lambda} y} \quad \text{für} \quad y > 0$

Erwartungswert: $\quad \mathbb{E}(Y_1) = \dfrac{1}{\widetilde{\lambda}} = \dfrac{1}{\lambda_1 + \lambda_2 + \cdots + \lambda_n}$

Maximum

Die sGn $X_1, X_2, \ldots, X_n$ seien Lebensdauern von Komponenten in einem Parallelsystem:



Das Parallelsystem fällt aus, sobald alle Komponenten ausgefallen sind.

Die Lebensdauer Y_n des Parallelsystems ist also das Maximum der Lebensdauern der Komponenten:

$$Y_n = \max\{X_1, X_2, \ldots, X_n\}$$

Maximum (Forts.)

Sind die sGn X_i, $i = 1, 2, \ldots, n$, unabhängig, so ist die Verteilungsfunktion von Y_n gegeben durch:

$$F_{\max}(y) = P(Y_n \leq y)$$

$$= P(X_1 \leq y, X_2 \leq y, \ldots, X_n \leq y) = \prod_{i=1}^{n} F_i(y)$$

Dabei bezeichnet F_i die Verteilungsfunktion von X_i.

Beispiel

Für $X_1, X_2, \ldots, X_n$ iid $\text{Exp}(\lambda)$ (oder $\text{Exp}(\tau)$) ist $F_{\max}$ gegeben durch:

$$F_{\max}(y) = \prod_{i=1}^{n} \left(1 - e^{-\lambda y}\right) = \left(1 - e^{-\lambda y}\right)^n \quad \text{für} \quad y > 0$$

Die Dichte von Y_n bekommt man durch Ableiten:

$$f_{\max}(y) = F'_{\max}(y) = n\lambda e^{-\lambda y}\left(1 - e^{-\lambda y}\right)^{n-1} \quad \text{für} \quad y > 0$$

Achtung: Im Gegensatz zum Seriensystem hat die Lebensdauer des Parallelsystems nicht wieder eine Exponentialverteilung.

Erwartungswert: $\mathbb{E}(Y_n) = \dfrac{1}{\lambda}\left(\dfrac{1}{n} + \dfrac{1}{n-1} + \cdots + \dfrac{1}{2} + 1\right)$

k-aus-n–Systeme

Ein k-aus-n–System ist intakt, wenn zumindest k der insgesamt n Komponenten intakt sind.

Die beiden vorhin betrachteten Systeme sind Spezialfälle:

Ein Seriensystem ist ein n-aus-n–System.

Ein Parallelsystem ist ein 1-aus-n–System.

(siehe Skriptum S. 212)

Multinomialverteilung

Ein Experiment bestehe aus n identischen und unabhängigen Versuchen, wobei jeder Versuch mit den (konstanten) Wahrscheinlichkeiten $p_1, \ldots, p_k$, wobei $\sum_{i=1}^{k} p_i = 1$, auf eine von k Arten ausgehen kann.

X_i = Anzahl der Versuche, die auf die i–te Art ausgehen ($i = 1, \ldots, k$)

Der stochastische Vektor $(X_1, X_2, \ldots, X_k)'$ hat eine Multinomialverteilung $\mathrm{M}(n, p_1, p_2, \ldots, p_k)$ mit der W–Funktion:

$$p(x_1, \ldots, x_k) = \binom{n}{x_1, x_2, \ldots, x_k} p_1^{x_1}\, p_2^{x_2} \cdots p_k^{x_k} \quad \text{mit} \quad \sum_{i=1}^{k} x_i = n$$

Dabei ist der Multinomialkoeffizient gegeben durch:

$$\binom{n}{x_1, x_2, \ldots, x_k} = \frac{n!}{x_1!\, x_2! \cdots x_k!} \quad \text{mit} \quad \sum_{i=1}^{k} x_i = n$$

Multinomialverteilung (Forts.)

Achtung: Entgegen der Schreibweise handelt es sich bei der $M(n, p_1, p_2, \ldots, p_k)$–Verteilung wegen $\sum_{i=1}^{k} x_i = n$ nur um eine $(k-1)$–dimensionale Verteilung.

Spezialfall Binomialverteilung $(k = 2)$:

$$p(x) = \binom{n}{x} p^x (1-p)^{n-x} \quad \text{für} \quad x = 0, 1, 2, \ldots, n$$

Spezialfall Trinomialverteilung $(k = 3)$:

$$p(x, y) = \frac{n!}{x!y!(n-x-y)!} \, p_1^x p_2^y (1 - p_1 - p_2)^{n-x-y}$$

$$\text{für} \quad x, y \in \{0, 1, 2, \ldots, n\} \quad \text{mit } x + y \leq n$$

Multinomialverteilung (Forts.)

Es gelte: $(X_1, X_2, \ldots, X_k)' \sim \mathrm{M}(n, p_1, p_2, \ldots, p_k)$

Randverteilungen: Die Randverteilungen sind wieder Multinomialverteilungen. Insbesondere gilt:

(1) $X_i \sim \mathrm{M}(n, p_i, 1 - p_i) \equiv \mathrm{B}(n, p_i)$

(2) $(X_i, X_j) \sim \mathrm{M}(n, p_i, p_j, 1 - p_i - p_j)$ (für $i < j$)

Erwartungswert/Varianz:

$$\mathbb{E}(X_i) = np_i, \quad \mathrm{Var}(X_i) = np_i(1 - p_i)$$

Kovarianz/Korrelation: Für $i \neq j$ gilt:

$$\mathrm{Cov}(X_i, X_j) = -np_ip_j, \quad \rho_{X_iX_j} = -\sqrt{\frac{p_i}{1-p_i}}\sqrt{\frac{p_j}{1-p_j}}$$

Multivariate Normalverteilung

Der stochastische Vektor $\mathbf{X} = (X_1, X_2, \ldots, X_n)'$ hat eine (n–dimensionale) multivariate Normalverteilung, $\mathbf{X} \sim \mathrm{N}_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, wenn seine Dichte gegeben ist durch:

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{n/2} |\boldsymbol{\Sigma}|^{1/2}} \exp\left[-\frac{1}{2} (\mathbf{x} - \boldsymbol{\mu})' \boldsymbol{\Sigma}^{-1} (\mathbf{x} - \boldsymbol{\mu})\right]$$

$$\text{für} \quad \mathbf{x} = (x_1, x_2, \ldots, x_n)' \in \mathbb{R}^n$$

$\boldsymbol{\mu} = (\mu_1, \mu_2, \ldots, \mu_n)'$ ... Vektor aus $\mathbb{R}^n$

$\boldsymbol{\Sigma}$ ... symmetrische u. positiv definite $(n \times n)$–Matrix

$|\boldsymbol{\Sigma}|$ ... Determinante, $\boldsymbol{\Sigma}^{-1}$ ... inverse Matrix

n−dimensionale Standardnormalverteilung

Im Spezialfall $\boldsymbol{\mu} = \mathbf{0}$ und $\boldsymbol{\Sigma} = \mathbf{I}_n$ ($= n \times n-$ Einheitsmatrix) spricht man von einer ($n-$dimensionalen) Standardnormalverteilung und schreibt:

$$\mathbf{Z} = (Z_1, Z_2, \ldots, Z_n)' \sim \mathrm{N}_n(\mathbf{0}, \mathbf{I}_n)$$

Wegen $|\mathbf{I}_n| = 1$ und $\mathbf{I}_n^{-1} = \mathbf{I}_n$ ist die Dichte von $\mathbf{Z}$ gegeben durch:

$$f(\mathbf{z}) = \frac{1}{(2\pi)^{n/2}} \exp\left(-\frac{1}{2}\,\mathbf{z}'\mathbf{z}\right)$$

$$\text{für} \quad \mathbf{z} = (z_1, z_2, \ldots, z_n)' \in \mathbb{R}^n$$

Bivariate Normalverteilung

Im Fall $n = 2$ spricht man von einer bivariate Normalverteilung.

Schreibt man (X, Y) statt (X_1, X_2), so gilt für $(X, Y)' \sim \mathrm{N}_2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$:

$$\boldsymbol{\mu} = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}, \qquad \boldsymbol{\Sigma} = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{bmatrix}$$

μ_1 ... Mittelwert von X, μ_2 ... Mittelwert von Y

σ_1^2 ... Varianz von X, σ_2^2 ... Varianz von Y

$\sigma_{12} = \rho\sigma_1\sigma_2$ ... Kovarianz von X und Y

ρ ... Korrelationskoeffizient von X und Y (Vs.: $\rho^2 < 1$)

Bivariate Normalverteilung (Forts.)

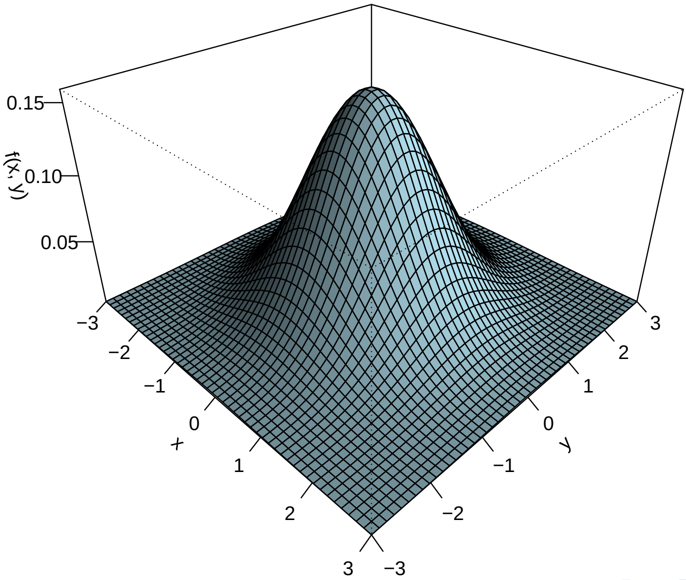
$$|\boldsymbol{\Sigma}| = \sigma_1^2\sigma_2^2(1-\rho^2), \quad \boldsymbol{\Sigma}^{-1} = \frac{1}{\sigma_1^2\sigma_2^2(1-\rho^2)} \begin{bmatrix} \sigma_2^2 & -\rho\sigma_1\sigma_2 \\ -\rho\sigma_1\sigma_2 & \sigma_1^2 \end{bmatrix}$$

Dichte: $f(x,y) = \dfrac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \exp\left(-\dfrac{q}{2}\right)$ für $(x,y) \in \mathbb{R}^2$

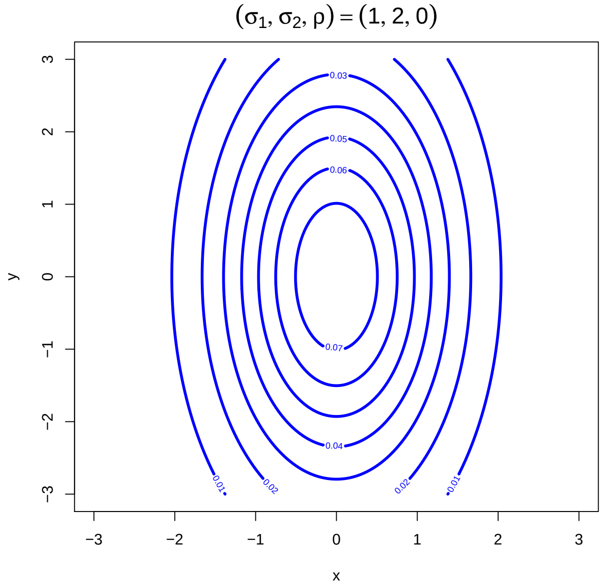
$$q = \frac{1}{1-\rho^2}\left[\left(\frac{x-\mu_1}{\sigma_1}\right)^2 - 2\rho\left(\frac{x-\mu_1}{\sigma_1}\right)\left(\frac{y-\mu_2}{\sigma_2}\right) + \left(\frac{y-\mu_2}{\sigma_2}\right)^2\right]$$

Bem: Man schreibt statt $\mathrm{N}_2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ auch $\mathrm{N}_2(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$.

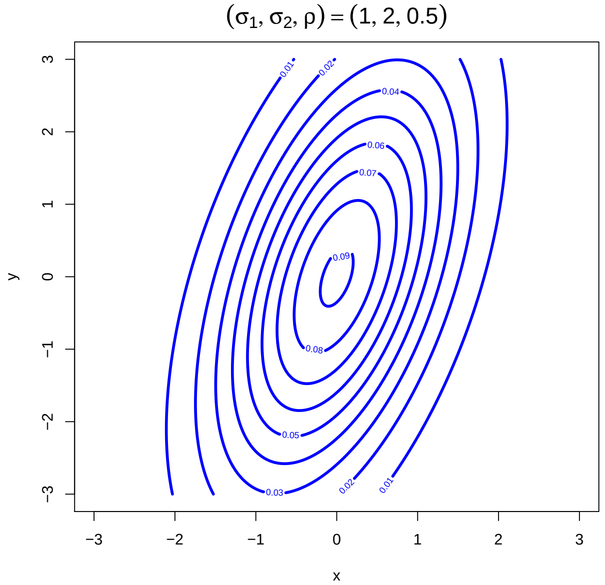
Bivariate Standardnormaldichte



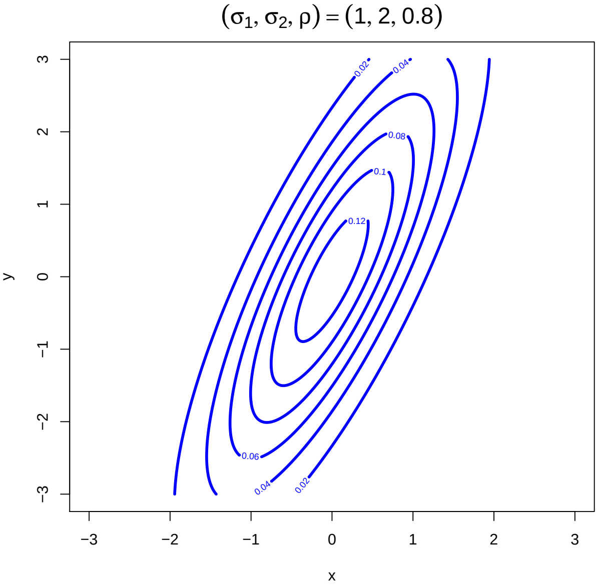
Bivariate Normaldichte: Contourplot



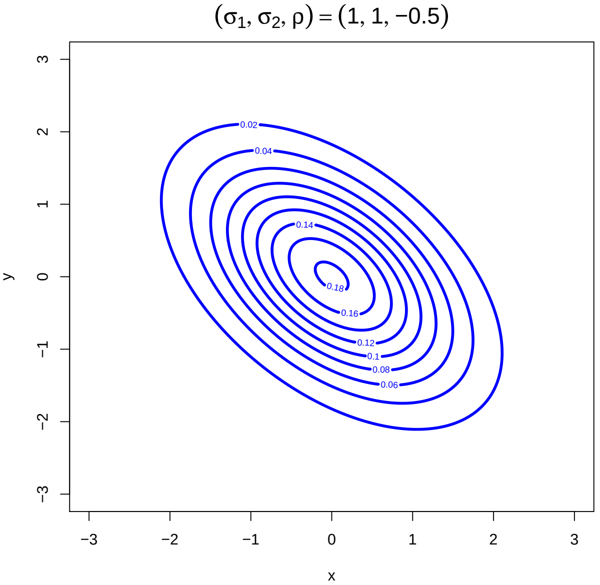
Bivariate Normaldichte: Contourplot



Bivariate Normaldichte: Contourplot



Bivariate Normaldichte: Contourplot



Bivariate Normalverteilung: Randverteilungen

Es gelte: $(X, Y)' \sim \mathrm{N}_2(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$

Randverteilung von X: $X \sim \mathrm{N}(\mu_1, \sigma_1^2)$

Randverteilung von Y: $Y \sim \mathrm{N}(\mu_2, \sigma_2^2)$

Unabhängigkeit/Unkorreliertheit: X und Y sind genau dann unabhängig wenn sie unkorreliert sind:

$$f(x, y) = f_1(x)\, f_2(y), \quad \forall\,(x, y) \in \mathbb{R}^2 \quad \Longleftrightarrow \quad \rho = 0$$

Bem: Bekanntlich folgt aus X, Y ua. stets $\rho_{XY} = 0$. Die Umkehrung gilt i. A. nicht. Die bivariate Normalverteilung bildet diesbezüglich eine wichtige Ausnahme.

Bivariate Normalverteilung: Bedingte Verteilungen

$$Y|X = x \sim \mathrm{N}\left(\mu_2 + \rho\,\frac{\sigma_2}{\sigma_1}\,(x - \mu_1),\, \sigma_2^2(1 - \rho^2)\right)$$

$$\Longrightarrow \quad \mathbb{E}(Y|x) = \mu_2 + \rho\,\frac{\sigma_2}{\sigma_1}\,(x - \mu_1)$$

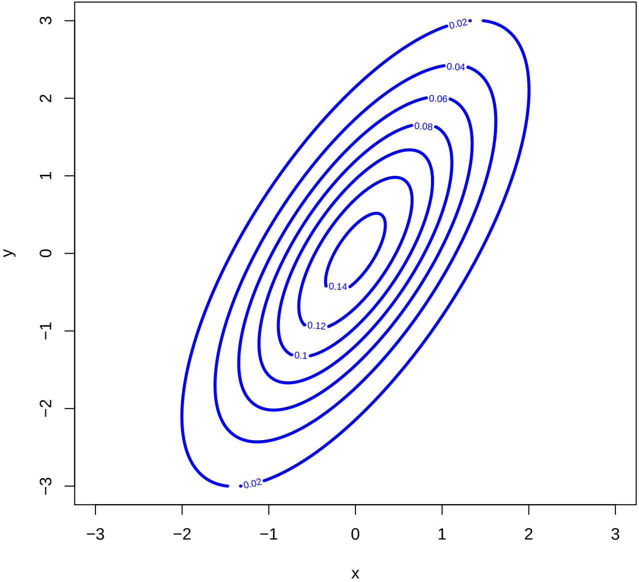
Diese Gerade nennt man auch die Regressionsgerade von Y auf X.

$$X|Y = y \sim \mathrm{N}\left(\mu_1 + \rho\,\frac{\sigma_1}{\sigma_2}\,(y - \mu_2),\, \sigma_1^2(1 - \rho^2)\right)$$

$$\Longrightarrow \quad \mathbb{E}(X|y) = \mu_1 + \rho\,\frac{\sigma_1}{\sigma_2}\,(y - \mu_2)$$

Diese Gerade nennt man auch die Regressionsgerade von X auf Y.

Bivariate Normalverteilung: Regressionsschere



Bivariate Normalverteilung: Regressionsschere