Bedingte Verteilungen 2-dim. Zufallsvektoren I

Definition 10.5 (Bedingte Wahrscheinlichkeitsverteilung (diskret))

Seien $\mathbf{X} = (X, Y)$ *ein diskreter zweidimensionaler Zufallsvektor mit Wahrscheinlichkeitsfunktion* $p_{(X,Y)}$, p_Y *die Wahrscheinlichkeitsfunktion zur Randverteilung von* Y, $y \in T(Y)$. *Dann heißt die durch die Wahrscheinlichkeitsfunktion*

$$p_{X|Y=y} : \mathbb{R} \to [0, 1]; p_{X|Y=y}(x) = \frac{p_{(X,Y)}(x, y)}{p_Y(y)}$$

definierte diskrete Wahrscheinlichkeitsverteilung **bedingte Wahrscheinlichkeitsverteilung von** X **unter der Bedingung** $Y = y$.
Analog nennt man für $x \in T(X)$ *die durch die Wahrscheinlichkeitsfunktion*

$$p_{Y|X=x} : \mathbb{R} \to [0, 1]; p_{Y|X=x}(y) = \frac{p_{(X,Y)}(x, y)}{p_X(x)}$$

definierte diskrete Wahrscheinlichkeitsverteilung **bedingte Wahrscheinlichkeitsverteilung von** Y **unter der Bedingung** $X = x$.

Bedingte Verteilungen 2-dim. Zufallsvektoren II

Definition 10.6 (Bedingte Wahrscheinlichkeitsverteilung (stetig))

Seien $\mathbf{X} = (X, Y)$ *ein stetiger zweidimensionaler Zufallsvektor,* $f_{(X,Y)}$ *eine gemeinsame Dichtefunktion von* (X, Y)*,* f_Y *eine Dichtefunktion zur Randverteilung von* Y*,* $y \in \mathbb{R}$ *eine Stetigkeitsstelle von* f_Y *mit* $f_Y(y) > 0$*. Dann heißt die durch die Dichtefunktion*

$$f_{X|Y=y} : \mathbb{R} \to \mathbb{R}; f_{X|Y=y}(x) = \frac{f_{(X,Y)}(x, y)}{f_Y(y)}$$

definierte stetige Wahrscheinlichkeitsverteilung **bedingte Wahrscheinlichkeitsverteilung von** X **unter der Bedingung** $Y = y$.

Analog nennt man für Stetigkeitsstellen $x \in \mathbb{R}$ *von* f_X *mit* $f_X(x) > 0$ *die durch die Dichtefunktion*

$$f_{Y|X=x} : \mathbb{R} \to \mathbb{R}; f_{Y|X=x}(y) = \frac{f_{(X,Y)}(x, y)}{f_X(x)}$$

definierte stetige Wahrscheinlichkeitsverteilung **bedingte Wahrscheinlichkeitsverteilung von** Y **unter der Bedingung** $X = x$.

Bemerkungen I

- Mit den üblichen Bezeichnungen und Abkürzungen für zweidimensionale Zufallsvektoren mit endlichem Träger erhält man mit Definition 10.5 analog zu den bedingten Häufigkeiten bei zweidimensionalen Merkmalen

$$p_{X|Y=y_j}(x_i) = \frac{p_{ij}}{p_{\cdot j}} \qquad \text{und} \qquad p_{Y|X=x_i}(y_j) = \frac{p_{ij}}{p_{i\cdot}}$$

  für $i \in \{1, \ldots, k\}$ und $j \in \{1, \ldots, l\}$.
- Ebenfalls analog zur deskriptiven Statistik sind zwei Zufallsvariablen also genau dann stochastisch unabhängig, wenn die bedingten Verteilungen jeweils mit den zugehörigen Randverteilungen übereinstimmen, also wenn
  - im diskreten Fall $p_{X|Y=y}(x) = p_X(x)$ für alle $x \in \mathbb{R}$ und alle $y \in T(Y)$ sowie $p_{Y|X=x}(y) = p_Y(y)$ für alle $y \in \mathbb{R}$ und alle $x \in T(X)$ gilt.
  - im stetigen Fall (bedingte) Dichtefunktionen f_X, f_Y, $f_{Y|X=x}$, $f_{X|Y=y}$ existieren mit $f_{X|Y=y}(x) = f_X(x)$ für alle $x \in \mathbb{R}$ und alle Stetigkeitsstellen $y \in \mathbb{R}$ von f_Y mit $f_Y(y) > 0$ sowie $f_{Y|X=x}(y) = f_Y(y)$ für alle $y \in \mathbb{R}$ und alle Stetigkeitsstellen $x \in \mathbb{R}$ von f_X mit $f_X(x) > 0$.

Bemerkungen II

- Durch bedingte Dichtefunktionen $f_{X|Y=y}$ bzw. $f_{Y|X=x}$ oder bedingte Wahrscheinlichkeitsfunktionen $p_{X|Y=y}$ bzw. $p_{Y|X=x}$ definierte bedingte Verteilungen können *im weiteren Sinne* als Verteilungen eindimensionaler Zufallsvariablen $X|Y=y$ bzw. $Y|X=x$ aufgefasst werden.
- Damit lassen sich viele für eindimensionale Zufallsvariablen bekannte Methoden/Konzepte in natürlicher Weise übertragen, insbesondere auf
  - bedingte Verteilungsfunktionen $F_{X|Y=y}$ bzw. $F_{Y|X=x}$,
  - bedingte Wahrscheinlichkeitsmaße $P_{X|Y=y}$ bzw. $P_{Y|X=x}$,
  - bedingte Träger $T(X|Y=y)$ bzw. $T(Y|X=x)$,
  - bedingte Erwartungswerte $\mathrm{E}(X|Y=y)$ bzw. $\mathrm{E}(Y|X=x)$,
  - bedingte Varianzen $\mathrm{Var}(X|Y=y)$ bzw. $\mathrm{Var}(Y|X=x)$,
- Den bedingten Erwartungswert $\mathrm{E}(Y|X=x)$ erhält man zum Beispiel
  - im diskreten Fall durch: $\mathrm{E}(Y|X=x) = \sum_{y_i} y_i \cdot p_{Y|X=x}(y_i)$
  - im stetigen Fall durch: $\mathrm{E}(Y|X=x) = \int_{-\infty}^{+\infty} y \cdot f_{Y|X=x}(y)dy$
- Gültig bleibt auch der VZS: $\mathrm{Var}(X|Y=y) = \mathrm{E}(X^2|Y=y) - [\mathrm{E}(X|Y=y)]^2$

Beispiel: Zweidimensionale Normalverteilung I

- Wichtige mehrdimensionale stetige Verteilung: **mehrdimensionale (multivariate) Normalverteilung**
- Spezifikation am Beispiel der zweidimensionalen (bivariaten) Normalverteilung durch Angabe einer Dichtefunktion

$$f_{X,Y}(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} e^{\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_X}{\sigma_X}\right)^2 - 2\rho\left(\frac{x-\mu_X}{\sigma_X}\right)\left(\frac{y-\mu_Y}{\sigma_Y}\right) + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2\right]\right\}}$$

  abhängig von den Parametern $\mu_X, \mu_Y \in \mathbb{R}$, $\sigma_X, \sigma_Y > 0$, $\rho \in (-1, 1)$.
- Man kann zeigen, dass die Randverteilungen von (X, Y) dann wieder (eindimensionale) Normalverteilungen sind, genauer gilt $X \sim N(\mu_X, \sigma_X^2)$ und $Y \sim N(\mu_Y, \sigma_Y^2)$

Beispiel: Zweidimensionale Normalverteilung II

- Sind f_X bzw. f_Y die wie auf Folie 242 definierten Dichtefunktionen zur $N(\mu_X, \sigma_X^2)$- bzw. $N(\mu_Y, \sigma_Y^2)$-Verteilung, so gilt (genau) im Fall $\rho = 0$

  $$f_{X,Y}(x, y) = f_X(x) \cdot f_Y(y) \qquad \text{für alle } x, y \in \mathbb{R} \ ,$$

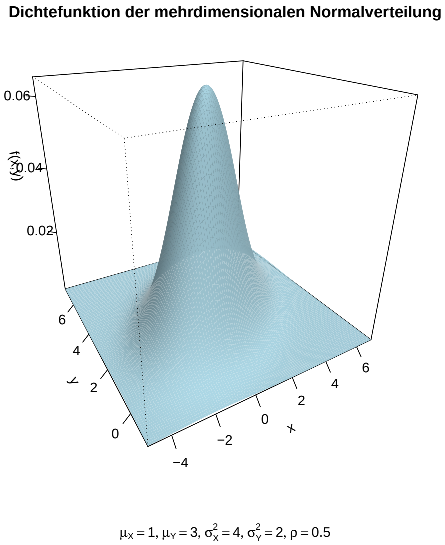
  also sind X und Y (genau) für $\rho = 0$ stochastisch unabhängig.
- Auch für $\rho \neq 0$ sind die bedingten Verteilungen von $X|Y = y$ und $Y|X = x$ wieder Normalverteilungen, es gilt genauer:

  $$X|Y = y \quad \sim \quad N\left(\mu_X + \frac{\rho\sigma_X}{\sigma_Y}(y - \mu_Y), \sigma_X^2(1 - \rho^2)\right)$$

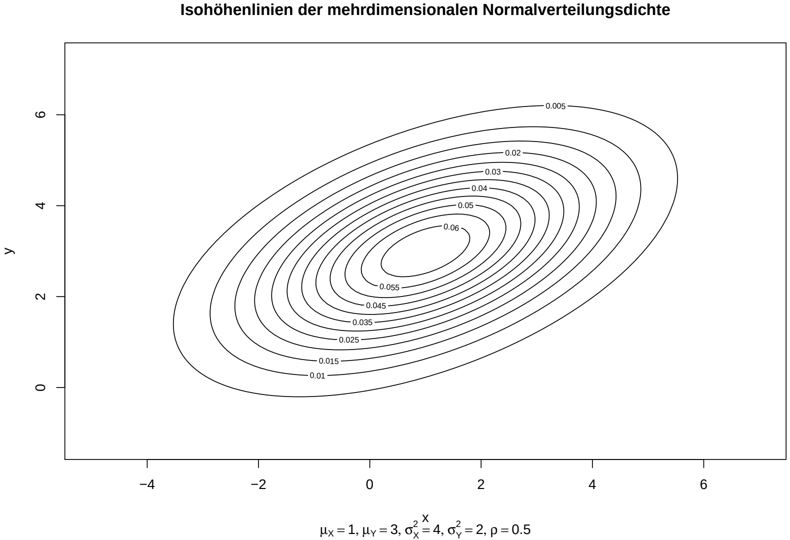
  bzw.

  $$Y|X = x \quad \sim \quad N\left(\mu_Y + \frac{\rho\sigma_Y}{\sigma_X}(x - \mu_X), \sigma_Y^2(1 - \rho^2)\right)$$

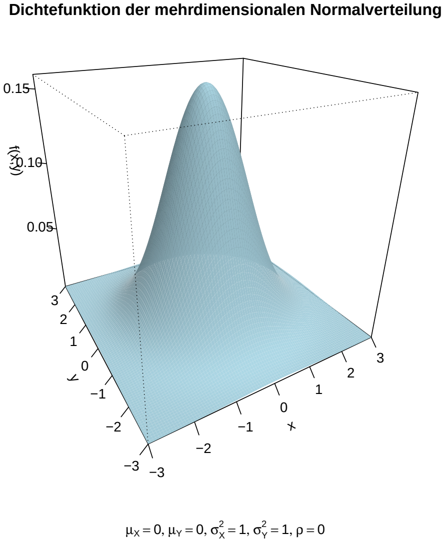
Beispiel: Zweidimensionale Normalverteilung III



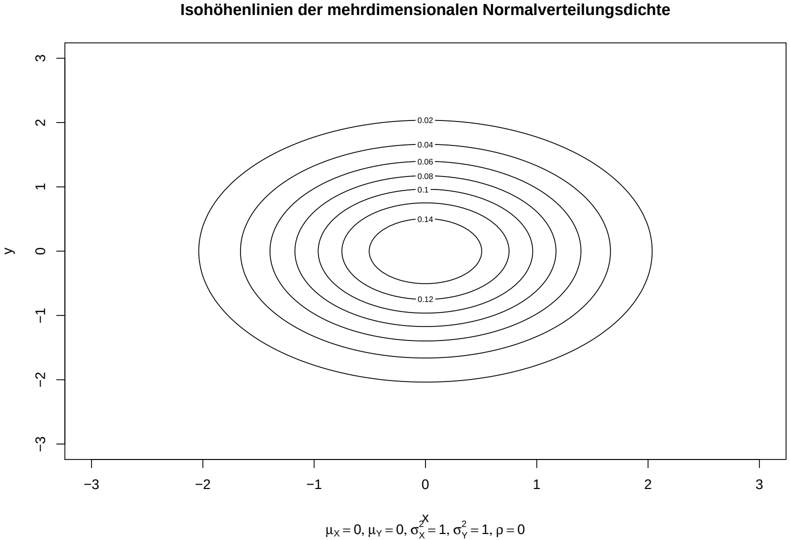
Beispiel: Zweidimensionale Normalverteilung IV



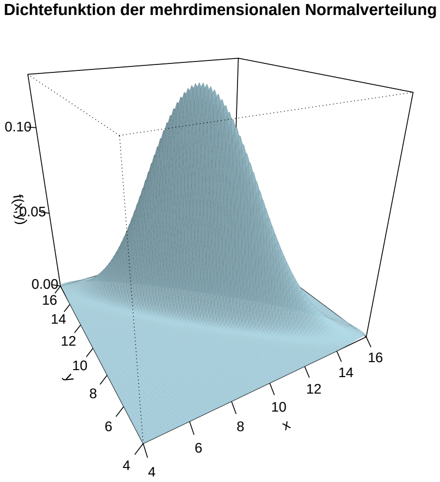
Beispiel: Zweidimensionale Normalverteilung V



Beispiel: Zweidimensionale Normalverteilung VI

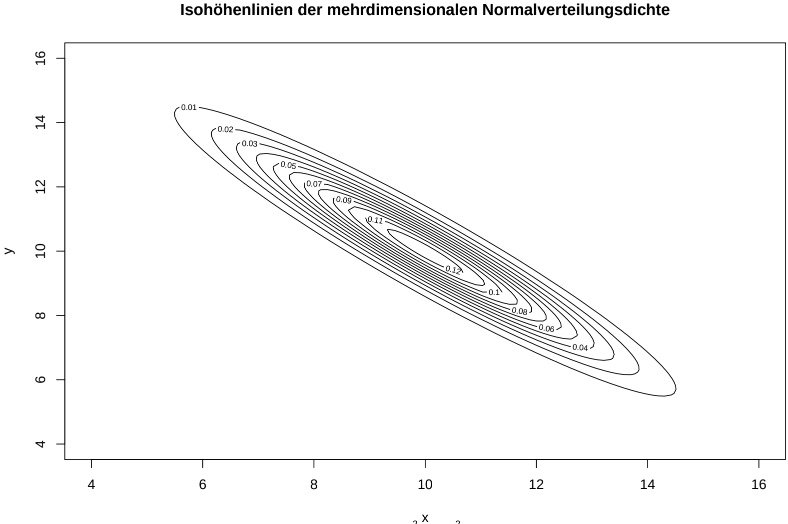


Beispiel: Zweidimensionale Normalverteilung VII



$\mu_X = 10, \mu_Y = 10, \sigma_X^2 = 4, \sigma_Y^2 = 4, \rho = -0.95$

Beispiel: Zweidimensionale Normalverteilung VIII



$\mu_X = 10, \mu_Y = 10, \sigma_X^2 = 4, \sigma_Y^2 = 4, \rho = -0.95$

Erwartungswert von $G(X, Y)$

Definition 10.7

Es seien (X, Y) eine zweidimensionale Zufallsvariable und $G : \mathbb{R}^2 \to \mathbb{R}$ eine $\mathcal{B}^2 - \mathcal{B}-$messbare Abbildung.

- *Ist (X, Y) diskreter Zufallsvektor, sind (x_i, y_j) die Trägerpunkte sowie $p_{(X,Y)}$ die Wahrscheinlichkeitsfunktion von (X, Y) und gilt $\sum_{x_i} \sum_{y_j} |G(x_i, y_j)| \cdot p_{(X,Y)}(x_i, y_j) < \infty$, dann existiert der Erwartungswert* $\mathrm{E}(G(X, Y))$ *und es gilt*

$$\mathrm{E}(G(X, Y)) = \sum_{x_i} \sum_{y_j} G(x_i, y_j) \cdot p_{(X,Y)}(x_i, y_j).$$

- *Ist (X, Y) stetiger Zufallsvektor mit Dichtefunktion $f_{(X,Y)}$ und gilt $\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} |G(x, y)| \cdot f_{(X,Y)}(x, y) dy dx < \infty$, dann existiert der Erwartungswert* $\mathrm{E}(G(X, Y))$ *und es gilt*

$$\mathrm{E}(G(X, Y)) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} G(x, y) \cdot f_{(X,Y)}(x, y) dy dx.$$

- *Beispiel 1:* $G(X, Y) = a \cdot X + b \cdot Y + c$ für $a, b, c \in \mathbb{R}$
  Wegen der Linearität von Summenbildung und Integration gilt stets:
  $$\mathrm{E}(a \cdot X + b \cdot Y + c) = a \cdot \mathrm{E}(X) + b \cdot \mathrm{E}(Y) + c$$
- *Beispiel 2:* $G(X, Y) = X \cdot Y$
  Hier gilt im allgemeinen **nicht** $\mathrm{E}(X \cdot Y) = \mathrm{E}(X) \cdot \mathrm{E}(Y)$!
  Sind X und Y allerdings **stochastisch unabhängig**, so gilt
  - im diskreten Fall wegen $p_{(X,Y)}(x, y) = p_X(x) \cdot p_Y(y)$ insgesamt
    $$\begin{aligned}\mathrm{E}(X \cdot Y) &= \sum_{x_i} \sum_{y_j} x_i \cdot y_j \cdot p_X(x_i) \cdot p_Y(y_j) \\ &= \sum_{x_i} x_i \cdot p_X(x_i) \cdot \sum_{y_j} y_j \cdot p_Y(y_j)\end{aligned}$$
  - und im stetigen Fall wegen $f_{(X,Y)}(x, y) = f_X(x) \cdot f_Y(y)$ insgesamt
    $$\begin{aligned}\mathrm{E}(X \cdot Y) &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} x \cdot y \cdot f_X(x) \cdot f_Y(y) dy dx \\ &= \int_{-\infty}^{+\infty} x \cdot f_X(x) dx \cdot \int_{-\infty}^{+\infty} y \cdot f_Y(y) dy \ ,\end{aligned}$$

  womit man **in diesem Fall** speziell $\mathrm{E}(X \cdot Y) = \mathrm{E}(X) \cdot \mathrm{E}(Y)$ erhält.

Abhängigkeitsmaße

- Analog zur deskriptiven Statistik ist man an Maßzahlen für die Abhängigkeit von zwei Zufallsvariablen X und Y über demselben Wahrscheinlichkeitsraum interessiert.
- Bei stochastischer Unabhängigkeit von X und Y sollten diese Maßzahlen naheliegenderweise den Wert 0 annehmen.
- Wie in deskriptiver Statistik: Maß für **lineare** Abhängigkeit von X und Y vordergründig:

Definition 10.8 (Kovarianz)

Es sei (X, Y) *ein zweidimensionaler Zufallsvektor. Den Erwartungswert*

$$\sigma_{XY} := \mathrm{Cov}(X, Y) := \mathrm{E}\left[(X - \mathrm{E}(X)) \cdot (Y - \mathrm{E}(Y))\right]$$

nennt man (im Falle der Existenz) **Kovarianz** *zwischen* X *und* Y*.*

- Eine dem Varianzzerlegungssatz ähnliche Rechenvorschrift zeigt man auch leicht für die Berechnung der Kovarianz:

Rechenregeln für Kovarianzen

Satz 10.1 (Kovarianzzerlegungssatz)

Ist (X, Y) ein zweidimensionaler Zufallsvektor, so gilt (im Fall der Existenz der beteiligten Erwartungswerte)

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

Sind X, Y und Z Zufallsvariablen (über demselben Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, P)$) und $a, b \in \mathbb{R}$, so gelten außerdem die folgenden Rechenregeln:

1. $\mathrm{Cov}(aX, bY) = ab\,\mathrm{Cov}(X, Y)$
2. $\mathrm{Cov}(X + a, Y + b) = \mathrm{Cov}(X, Y)$
   (Translationsinvarianz)
3. $\mathrm{Cov}(X, Y) = \mathrm{Cov}(Y, X)$
   (Symmetrie)
4. $\mathrm{Cov}(X + Z, Y) = \mathrm{Cov}(X, Y) + \mathrm{Cov}(Z, Y)$
5. $\mathrm{Cov}(X, X) = \mathrm{Var}(X)$
6. X, Y stochastisch unabhängig $\Rightarrow \mathrm{Cov}(X, Y) = 0$

- Die erreichbaren Werte der Größe $\text{Cov}(X, Y)$ hängen nicht nur von der Stärke der linearen Abhängigkeit ab, sondern (wie insbesondere aus Rechenregel 1 von Folie 284 ersichtlich) auch von der Streuung von X bzw. Y.
- Wie in deskriptiver Statistik: Alternatives Abhängigkeitsmaß mit normiertem „Wertebereich", welches invariant gegenüber Skalierung von X bzw. Y ist.
- Hierzu Standardisierung der Kovarianz über Division durch Standardabweichungen von X und Y (falls möglich!):

Definition 10.9

Es sei (X, Y) *ein zweidimensionaler Zufallsvektor mit* $\text{Var}(X) = \sigma_X^2 > 0$ *und* $\text{Var}(Y) = \sigma_Y^2 > 0$*. Man nennt*

$$\rho_{XY} := \text{Korr}(X, Y) := \frac{\sigma_{XY}}{\sigma_X \cdot \sigma_Y} = \frac{\text{Cov}(X, Y)}{+\sqrt{\text{Var}(X) \cdot \text{Var}(Y)}}$$

den **Korrelationskoeffizienten** *(nach Bravais-Pearson) zwischen* X *und* Y*.*

Rechenregeln für Korrelationskoeffizienten

Sind X und Y Zufallsvariablen (über demselben Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, P)$) mit $\text{Var}(X) > 0, \text{Var}(Y) > 0$ und $a, b \in \mathbb{R}$, so gilt:

1. $\text{Korr}(aX, bY) = \begin{cases} \text{Korr}(X, Y) & \text{falls } a \cdot b > 0 \\ -\text{Korr}(X, Y) & \text{falls } a \cdot b < 0 \end{cases}$
2. $\text{Korr}(X + a, Y + b) = \text{Korr}(X, Y)$
   (Translationsinvarianz)
3. $\text{Korr}(X, Y) = \text{Korr}(Y, X)$
   (Symmetrie)
4. $-1 \leq \text{Korr}(X, Y) \leq 1$
5. $\text{Korr}(X, X) = 1$
6. $\left.\begin{array}{rr} \text{Korr}(X, Y) = & 1 \\ \text{Korr}(X, Y) = & -1 \end{array}\right\}$ genau dann, wenn $Y = aX + b$ mit $\begin{cases} a > 0 \\ a < 0 \end{cases}$
7. X, Y stochastisch unabhängig $\Rightarrow \text{Korr}(X, Y) = 0$

Zufallsvariablen X, Y mit $\text{Cov}(X, Y) = 0$ (!) heißen **unkorreliert**.

Zusammenhang Unkorreliertheit und Unabhängigkeit

- Offensichtlich gilt stets (falls die Kovarianz existiert):

$$X, Y \text{ stochastisch unabhängig} \quad \Rightarrow \quad X, Y \text{ unkorreliert}$$

- Die Umkehrung ist allerdings *im allgemeinen* falsch, es gilt **außer in speziellen Ausnahmefällen**:

$$X, Y \text{ unkorreliert} \quad \not\Rightarrow \quad X, Y \text{ stochastisch unabhängig}$$

- Einer dieser Ausnahmefälle ist die bivariate Normalverteilung:
  Ist der Zufallsvektor (X, Y) zweidimensional normalverteilt, so gilt **in dieser Situation** nämlich $\text{Korr}(X, Y) = \rho$ und damit (siehe Folie 274):

$$X, Y \text{ unkorreliert} \quad \Leftrightarrow \quad X, Y \text{ stochastisch unabhängig}$$

Varianzen von Summen zweier Zufallsvariablen

- Durch Verknüpfung verschiedener Rechenregeln aus Folie 284 lässt sich leicht zeigen, dass stets

$$\operatorname{Var}(X+Y) = \operatorname{Var}(X) + 2\operatorname{Cov}(X,Y) + \operatorname{Var}(Y)$$

bzw. für $a, b, c \in \mathbb{R}$ allgemeiner stets

$$\operatorname{Var}(aX+bY+c) = a^2\operatorname{Var}(X) + 2ab\operatorname{Cov}(X,Y) + b^2\operatorname{Var}(Y)$$

gilt.

- **Nur für unkorrelierte** (also insbesondere auch für stochastisch unabhängige) Zufallsvariablen X und Y gilt offensichtlich spezieller

$$\operatorname{Var}(aX+bY+c) = a^2\operatorname{Var}(X) + b^2\operatorname{Var}(Y) \ .$$

Dies kann für mehr als zwei Zufallsvariablen weiter verallgemeinert werden.