

## Beispiel: Zweidimensionale Normalverteilung I

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

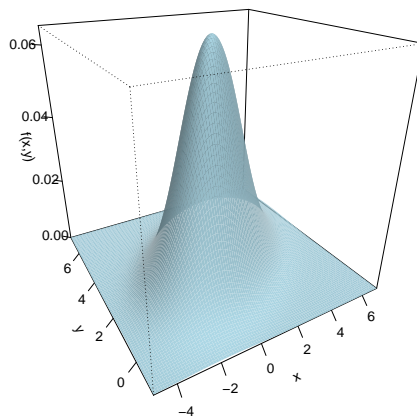
$$f_{X,Y}(x,y) = \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}} e^{\left\{-\frac{1}{2(1-\rho^2)}\left[\left(\frac{x-\mu_X}{\sigma_X}\right)^2 - 2\rho\left(\frac{x-\mu_X}{\sigma_X}\right)\left(\frac{y-\mu_Y}{\sigma_Y}\right) + \left(\frac{y-\mu_Y}{\sigma_Y}\right)^2\right]\right\}}$$

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 III

Dichtefunktion der mehrdimensionalen Normalverteilung



$$\mu_X = 1, \mu_Y = 3, \sigma_X^2 = 4, \sigma_Y^2 = 2, \rho = 0.5$$

## 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  $\rho = 0$

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

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:

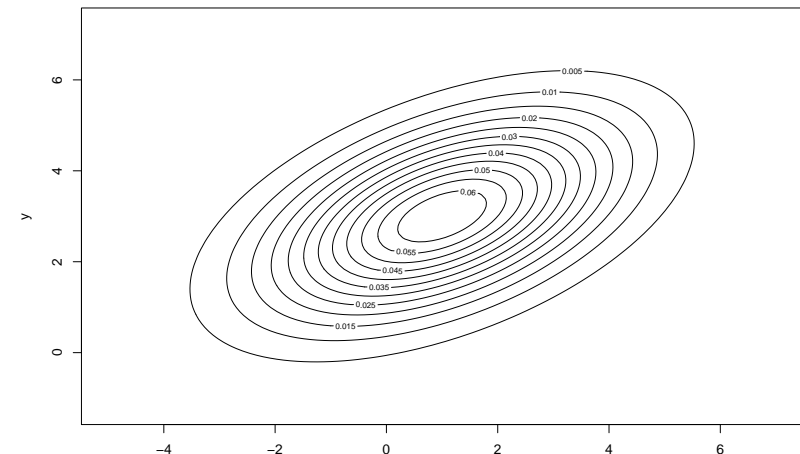
$$X|Y = y \sim N\left(\mu_X + \frac{\rho\sigma_X}{\sigma_Y}(y - \mu_Y), \sigma_X^2(1 - \rho^2)\right)$$

bzw.

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

## Beispiel: Zweidimensionale Normalverteilung IV

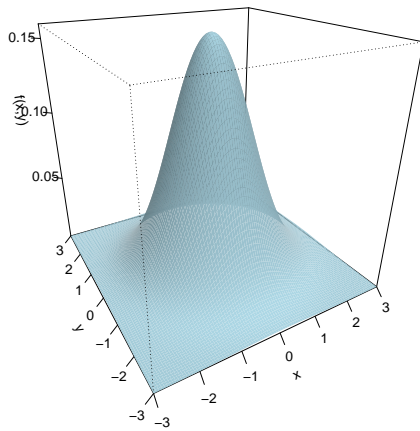
Isöhöhenlinien der mehrdimensionalen Normalverteilungsdichte



$$\mu_X = 1, \mu_Y = 3, \sigma_X^2 = 4, \sigma_Y^2 = 2, \rho = 0.5$$

## Beispiel: Zweidimensionale Normalverteilung V

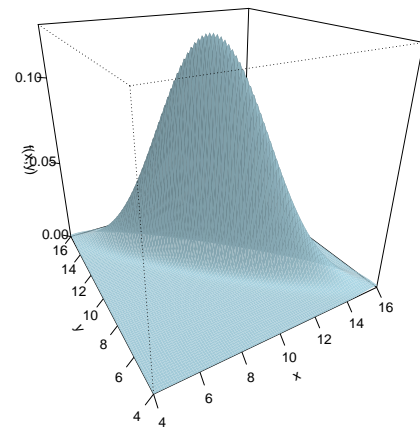
Dichtefunktion der mehrdimensionalen Normalverteilung



$$\mu_x = 0, \mu_y = 0, \sigma_x^2 = 1, \sigma_y^2 = 1, \rho = 0$$

## Beispiel: Zweidimensionale Normalverteilung VII

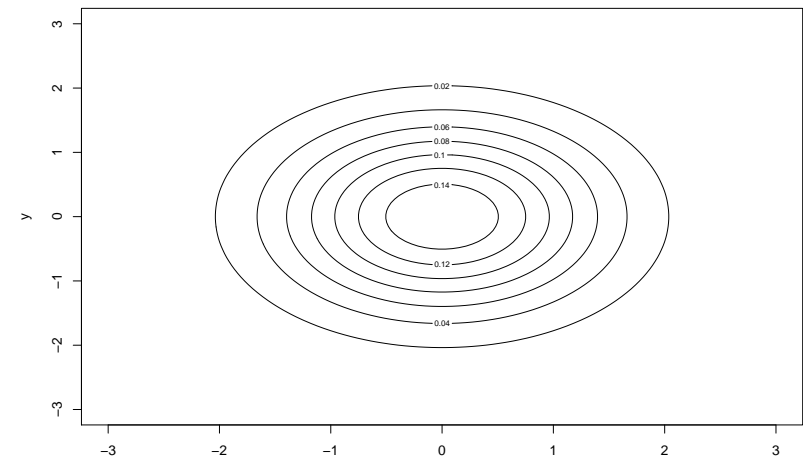
Dichtefunktion der mehrdimensionalen Normalverteilung



$$\mu_x = 10, \mu_y = 10, \sigma_x^2 = 4, \sigma_y^2 = 4, \rho = -0.95$$

## Beispiel: Zweidimensionale Normalverteilung VI

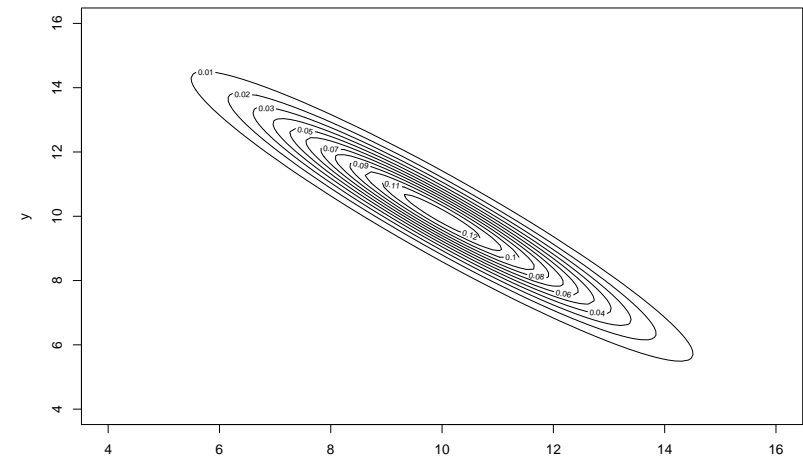
Isohöhenlinien der mehrdimensionalen Normalverteilungsdichte



$$\mu_x = 0, \mu_y = 0, \sigma_x^2 = 1, \sigma_y^2 = 1, \rho = 0$$

## Beispiel: Zweidimensionale Normalverteilung VIII

Isohöhenlinien der mehrdimensionalen Normalverteilungsdichte



$$\mu_x = 10, \mu_y = 10, \sigma_x^2 = 4, \sigma_y^2 = 4, \rho = -0.95$$

## Erwartungswert von $G(X, Y)$

### Definition 10.7

Es seien  $(X, Y)$  eine zweidimensionale Zufallsvariable und  $G: \mathbb{R}^2 \rightarrow \mathbb{R}$  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  $E(G(X, Y))$  und es gilt

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

- 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  $E(G(X, Y))$  und es gilt

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

## 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

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

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:

- 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:

$$E(a \cdot X + b \cdot Y + c) = a \cdot E(X) + b \cdot E(Y) + c$$

- Beispiel 2:**  $G(X, Y) = X \cdot Y$   
Hier gilt im allgemeinen **nicht**  $E(X \cdot Y) = E(X) \cdot E(Y)$ !  
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

$$\begin{aligned} E(X \cdot Y) &= \sum_{x_i} \sum_{y_j} x_i \cdot y_j \cdot p_X(x_i) \cdot p_Y(y_j) \\ &= \sum_{x_i} x_i \cdot p_X(x_i) \cdot \sum_{y_j} y_j \cdot p_Y(y_j) \end{aligned}$$

- und im stetigen Fall wegen  $f_{(X, Y)}(x, y) = f_X(x) \cdot f_Y(y)$  insgesamt

$$\begin{aligned} E(X \cdot Y) &= \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} x \cdot y \cdot f_X(x) \cdot f_Y(y) dy dx \\ &= \int_{-\infty}^{+\infty} x \cdot f_X(x) dx \cdot \int_{-\infty}^{+\infty} y \cdot f_Y(y) dy, \end{aligned}$$

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

## Rechenregeln für Kovarianzen

### Satz 10.9 (Kovarianzzerlegungssatz)

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

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

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:

- $\text{Cov}(aX, bY) = ab \text{Cov}(X, Y)$
- $\text{Cov}(X + a, Y + b) = \text{Cov}(X, Y)$   
(Translationsinvarianz)
- $\text{Cov}(X, Y) = \text{Cov}(Y, X)$   
(Symmetrie)
- $\text{Cov}(X + Z, Y) = \text{Cov}(X, Y) + \text{Cov}(Z, Y)$
- $\text{Cov}(X, X) = \text{Var}(X)$
- $X, Y$  stochastisch unabhängig  $\Rightarrow \text{Cov}(X, Y) = 0$

- 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 10.10

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  $Y$ .

## Zusammenhang Unkorreliertheit und Unabhängigkeit

- Offensichtlich gilt stets (falls die Kovarianz existiert):

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

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

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

- 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):

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

## 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  $\left. \begin{array}{l} \text{Korr}(X, Y) = 1 \\ \text{Korr}(X, Y) = -1 \end{array} \right\}$  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**.

## Varianzen von Summen zweier Zufallsvariablen

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

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

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

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

gilt.

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

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

Dies kann für mehr als zwei Zufallsvariablen weiter verallgemeinert werden.

## Momente höherdimensionaler Zufallsvektoren

### Definition 10.11

Seien  $n \in \mathbb{N}$  und  $\mathbf{X} = (X_1, \dots, X_n)'$  ein  $n$ -dimensionaler Zufallsvektor. Dann heißt der  $n$ -dimensionale Vektor

$$E(\mathbf{X}) := [E(X_1), \dots, E(X_n)]' = \begin{bmatrix} E(X_1) \\ \vdots \\ E(X_n) \end{bmatrix}$$

**Erwartungswertvektor** von  $\mathbf{X}$  und die  $n \times n$ -Matrix

$$V(\mathbf{X}) := E[(\mathbf{X} - E(\mathbf{X})) \cdot (\mathbf{X} - E(\mathbf{X}))'] \\ := \begin{bmatrix} E[(X_1 - E(X_1)) \cdot (X_1 - E(X_1))] & \cdots & E[(X_1 - E(X_1)) \cdot (X_n - E(X_n))] \\ \vdots & \ddots & \vdots \\ E[(X_n - E(X_n)) \cdot (X_1 - E(X_1))] & \cdots & E[(X_n - E(X_n)) \cdot (X_n - E(X_n))] \end{bmatrix}$$

**(Varianz-)Kovarianzmatrix** von  $\mathbf{X}$ .

- Ein  $n$ -dimensionaler Zufallsvektor  $\mathbf{X} = (X_1, \dots, X_n)'$  bzw. die  $n$  eindimensionalen Zufallsvariablen  $X_1, \dots, X_n$  heißen **unkorreliert**, wenn  $V(\mathbf{X})$  eine Diagonalmatrix ist, also höchstens die Diagonaleinträge von 0 verschieden sind.
- Es gilt in Verallgemeinerung von Folie 287 (bei Existenz der Momente):

$$X_1, \dots, X_n \text{ stochastisch unabhängig} \Rightarrow X_1, \dots, X_n \text{ unkorreliert,}$$

die Umkehrung gilt *im allgemeinen* nicht!

- Dass stochastische Unabhängigkeit eine stärkere Eigenschaft ist als Unkorreliertheit, zeigt auch der folgende Satz:

### Satz 10.12

Seien für  $n \in \mathbb{N}$  stochastisch unabhängige Zufallsvariablen  $X_1, \dots, X_n$  über einem gemeinsamen Wahrscheinlichkeitsraum  $(\Omega, \mathcal{F}, P)$  sowie  $\mathcal{B}$ - $\mathcal{B}$ -messbare Funktionen  $G_i: \mathbb{R} \rightarrow \mathbb{R}$  für  $i \in \{1, \dots, n\}$  gegeben.

Dann sind auch die Zufallsvariablen  $G_1(X_1), \dots, G_n(X_n)$  stochastisch unabhängig.

Es gilt  $V(\mathbf{X})$

$$= \begin{bmatrix} \text{Cov}(X_1, X_1) & \text{Cov}(X_1, X_2) & \cdots & \text{Cov}(X_1, X_{n-1}) & \text{Cov}(X_1, X_n) \\ \text{Cov}(X_2, X_1) & \text{Cov}(X_2, X_2) & \cdots & \text{Cov}(X_2, X_{n-1}) & \text{Cov}(X_2, X_n) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \text{Cov}(X_{n-1}, X_1) & \text{Cov}(X_{n-1}, X_2) & \cdots & \text{Cov}(X_{n-1}, X_{n-1}) & \text{Cov}(X_{n-1}, X_n) \\ \text{Cov}(X_n, X_1) & \text{Cov}(X_n, X_2) & \cdots & \text{Cov}(X_n, X_{n-1}) & \text{Cov}(X_n, X_n) \end{bmatrix} \\ = \begin{bmatrix} \text{Var}(X_1) & \text{Cov}(X_1, X_2) & \cdots & \text{Cov}(X_1, X_{n-1}) & \text{Cov}(X_1, X_n) \\ \text{Cov}(X_2, X_1) & \text{Var}(X_2) & \cdots & \text{Cov}(X_2, X_{n-1}) & \text{Cov}(X_2, X_n) \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ \text{Cov}(X_{n-1}, X_1) & \text{Cov}(X_{n-1}, X_2) & \cdots & \text{Var}(X_{n-1}) & \text{Cov}(X_{n-1}, X_n) \\ \text{Cov}(X_n, X_1) & \text{Cov}(X_n, X_2) & \cdots & \text{Cov}(X_n, X_{n-1}) & \text{Var}(X_n) \end{bmatrix}$$

- Der Eintrag  $v_{ij}$  der  $i$ -ten Zeile und  $j$ -ten Spalte von  $V(\mathbf{X})$  ist also gegeben durch  $\text{Cov}(X_i, X_j)$ .
- $V(\mathbf{X})$  ist wegen  $\text{Cov}(X_i, X_j) = \text{Cov}(X_j, X_i)$  offensichtlich symmetrisch.
- Für  $n = 1$  erhält man die „klassische“ Definition der Varianz, das heißt die Matrix  $V(\mathbf{X})$  wird als  $1 \times 1$ -Matrix skalar und stimmt mit  $\text{Var}(X_1)$  überein.

## Momente von Summen von Zufallsvariablen

- Regelmäßig ist man für  $n \in \mathbb{N}$  an der Verteilung bzw. an Maßzahlen der Summe  $\sum_{i=1}^n X_i = X_1 + \dots + X_n$  oder einer gewichteten Summe

$$\sum_{i=1}^n w_i \cdot X_i = w_1 \cdot X_1 + \dots + w_n \cdot X_n \quad (w_1, \dots, w_n \in \mathbb{R})$$

von  $n$  Zufallsvariablen  $X_1, \dots, X_n$  interessiert.

- In Verallgemeinerung von Folie 282 zeigt man für den Erwartungswert leicht

$$E\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n E(X_i) \quad \text{bzw.} \quad E\left(\sum_{i=1}^n w_i \cdot X_i\right) = \sum_{i=1}^n w_i \cdot E(X_i).$$

- Insbesondere gilt für das (arithmetische) Mittel aus  $X_1, \dots, X_n$

$$E\left(\frac{1}{n} \sum_{i=1}^n X_i\right) = \frac{1}{n} \sum_{i=1}^n E(X_i).$$

- In Verallgemeinerung von Folie 288 erhält man für die Varianz weiter

$$\begin{aligned}\text{Var}\left(\sum_{i=1}^n X_i\right) &= \sum_{i=1}^n \sum_{j=1}^n \text{Cov}(X_i, X_j) = \sum_{i=1}^n \text{Var}(X_i) + \sum_{\substack{i,j \in \{1, \dots, n\} \\ i \neq j}} \text{Cov}(X_i, X_j) \\ &= \sum_{i=1}^n \text{Var}(X_i) + 2 \sum_{i=1}^{n-1} \sum_{j=i+1}^n \text{Cov}(X_i, X_j)\end{aligned}$$

bzw.

$$\begin{aligned}\text{Var}\left(\sum_{i=1}^n w_i \cdot X_i\right) &= \sum_{i=1}^n \sum_{j=1}^n w_i \cdot w_j \cdot \text{Cov}(X_i, X_j) \\ &= \sum_{i=1}^n w_i^2 \cdot \text{Var}(X_i) + \sum_{\substack{i,j \in \{1, \dots, n\} \\ i \neq j}} w_i \cdot w_j \cdot \text{Cov}(X_i, X_j) \\ &= \sum_{i=1}^n w_i^2 \cdot \text{Var}(X_i) + 2 \sum_{i=1}^{n-1} \sum_{j=i+1}^n w_i \cdot w_j \cdot \text{Cov}(X_i, X_j).\end{aligned}$$

## Summen von Zufallsvariablen spezieller Verteilungen

- Sind die Zufallsvariablen  $X_1, \dots, X_n$  nicht nur unkorreliert, sondern sogar unabhängig, dann sind einige der erläuterten Verteilungsfamilien „abgeschlossen“ gegenüber Summenbildungen.
- Besitzen darüberhinaus alle  $n$  Zufallsvariablen  $X_1, \dots, X_n$  exakt dieselbe Verteilung, spricht man von unabhängig identisch verteilten Zufallsvariablen gemäß folgender Definition:

### Definition 11.1 (i.i.d. Zufallsvariablen)

Seien für  $n \in \mathbb{N}$  Zufallsvariablen  $X_1, \dots, X_n$  über einem gemeinsamen Wahrscheinlichkeitsraum  $(\Omega, \mathcal{F}, P)$  gegeben mit

- $Q_X := P_{X_1} = P_{X_2} = \dots = P_{X_n}$ ,
- $X_1, \dots, X_n$  stochastisch unabhängig.

Dann heißen die Zufallsvariablen  $X_1, \dots, X_n$  **unabhängig identisch verteilt (independent and identically distributed)** gemäß  $Q_X$ , in Zeichen:  $X_i \stackrel{i.i.d.}{\sim} Q_X$  oder  $X_i \stackrel{u.i.v.}{\sim} Q_X$  für  $i \in \{1, \dots, n\}$ .

- Sind  $X_1, \dots, X_n$  unkorreliert, so vereinfacht sich die Varianz der Summe wegen  $\text{Cov}(X_i, X_j) = 0$  für  $i \neq j$  offensichtlich zu

$$\text{Var}\left(\sum_{i=1}^n X_i\right) = \sum_{i=1}^n \text{Var}(X_i)$$

bzw.

$$\text{Var}\left(\sum_{i=1}^n w_i \cdot X_i\right) = \sum_{i=1}^n w_i^2 \cdot \text{Var}(X_i).$$

- Fasst man die Zufallsvariablen  $X_1, \dots, X_n$  im  $n$ -dimensionalen Zufallsvektor  $\mathbf{X} = (X_1, \dots, X_n)$  und die Gewichte  $w_1, \dots, w_n$  im Vektor  $\mathbf{w} = (w_1, \dots, w_n)' \in \mathbb{R}^n$  zusammen, so lassen sich die Ergebnisse kürzer darstellen als

$$E\left(\sum_{i=1}^n w_i \cdot X_i\right) = \mathbf{w}' E(\mathbf{X}) \quad \text{bzw.} \quad \text{Var}\left(\sum_{i=1}^n w_i \cdot X_i\right) = \mathbf{w}' V(\mathbf{X}) \mathbf{w}.$$

### Satz 11.2 (Summen spezieller Zufallsvariablen)

Seien  $n \in \mathbb{N}$ ,  $X_1, \dots, X_n$  stochastisch unabhängige Zufallsvariablen und

$$Y := X_1 + \dots + X_n = \sum_{i=1}^n X_i.$$

Gilt für  $i \in \{1, \dots, n\}$  weiterhin

- $X_i \sim B(1, p)$  für ein  $p \in (0, 1)$ , also insgesamt  $X_i \stackrel{i.i.d.}{\sim} B(1, p)$ , so gilt  $Y \sim B(n, p)$  (vgl. Folie 232),
- $X_i \sim B(n_i, p)$  für  $n_i \in \mathbb{N}$  und ein  $p \in (0, 1)$ , so gilt  $Y \sim B(N, p)$  mit  $N := n_1 + \dots + n_n = \sum_{i=1}^n n_i$ ,
- $X_i \sim \text{Pois}(\lambda_i)$  für  $\lambda_i \in \mathbb{R}_+$ , so gilt  $Y \sim \text{Pois}(\lambda)$  mit  $\lambda := \lambda_1 + \dots + \lambda_n = \sum_{i=1}^n \lambda_i$ ,
- $X_i \sim N(\mu_i, \sigma_i^2)$  für  $\mu_i \in \mathbb{R}$  und  $\sigma_i^2 > 0$ , so gilt  $Y \sim N(\mu, \sigma^2)$  mit  $\mu = \mu_1 + \dots + \mu_n = \sum_{i=1}^n \mu_i$  und  $\sigma^2 = \sigma_1^2 + \dots + \sigma_n^2 = \sum_{i=1}^n \sigma_i^2$ ,
- $X_i \sim N(\mu, \sigma^2)$  für ein  $\mu \in \mathbb{R}$  und ein  $\sigma^2 > 0$ , also insgesamt  $X_i \stackrel{i.i.d.}{\sim} N(\mu, \sigma^2)$ , so gilt für  $\bar{X} := \frac{1}{n}(X_1 + \dots + X_n) = \frac{1}{n} \sum_{i=1}^n X_i$  insbesondere  $\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$ .

## Der zentrale Grenzwertsatz

- Die Verteilung von Summen (insbesondere) unabhängig identisch verteilter Zufallsvariablen weist ein spezielles Grenzverhalten auf, wenn die Anzahl der summierten Zufallsvariablen wächst.
- Dieses Grenzverhalten ist wesentlicher Grund für großes Anwendungspotenzial der (Standard-)Normalverteilung.
- Betrachte im Folgenden für  $i \in \mathbb{N}$  unabhängig identisch verteilte Zufallsvariablen  $X_i$  mit  $E(X_i) = \mu_X \in \mathbb{R}$  und  $\text{Var}(X_i) = \sigma_X^2 > 0$ .
- Gilt  $n \rightarrow \infty$  für die Anzahl  $n$  der summierten Zufallsvariablen  $X_i$ , gilt für die Summen  $Y_n := \sum_{i=1}^n X_i$  offensichtlich  $\text{Var}(Y_n) = n \cdot \sigma_X^2 \rightarrow \infty$  und für  $\mu_X \neq 0$  auch  $E(Y_n) = n \cdot \mu_X \rightarrow \pm\infty$ , eine Untersuchung der Verteilung von  $Y_n$  für  $n \rightarrow \infty$  ist also schwierig.
- Stattdessen: Betrachtung der **standardisierten** Zufallsvariablen

$$Z_n := \frac{Y_n - E(Y_n)}{\sqrt{\text{Var}(Y_n)}} = \frac{(\sum_{i=1}^n X_i) - n\mu_X}{\sigma_X \sqrt{n}} = \frac{(\frac{1}{n} \sum_{i=1}^n X_i) - \mu_X}{\sigma_X} \sqrt{n}$$

mit  $E(Z_n) = 0$  und  $\text{Var}(Z_n) = 1$  für alle  $n \in \mathbb{N}$ .

- Gemäß folgendem Satz „konvergiert“ die Verteilung der  $Z_n$  für  $n \rightarrow \infty$  gegen eine Standardnormalverteilung:

## Das schwache Gesetz der großen Zahlen

- Eine direkte Folge von Satz 11.3 ist zum Beispiel das folgende Resultat:

### Satz 11.4 (Schwachtes Gesetz der großen Zahlen)

Es seien  $X_i$  für  $i \in \mathbb{N}$  unabhängig identisch verteilte Zufallsvariablen mit  $E(X_i) = \mu_X \in \mathbb{R}$ . Dann gilt:

$$\lim_{n \rightarrow \infty} P \left\{ \left| \left( \frac{1}{n} \sum_{i=1}^n X_i \right) - \mu_X \right| \geq \varepsilon \right\} = 0 \quad \text{für alle } \varepsilon > 0$$

- Das schwache Gesetz der großen Zahlen besagt also, dass die Wahrscheinlichkeit, mit der ein Mittelwert von  $n$  i.i.d. Zufallsvariablen betragsmäßig um mehr als eine vorgegebene (kleine) Konstante  $\varepsilon > 0$  vom Erwartungswert der Zufallsvariablen abweicht, für  $n \rightarrow \infty$  gegen Null geht.
- Insbesondere „konvergiert“ also die Folge der beobachteten relativen Häufigkeiten der Erfolge bei unabhängiger wiederholter Durchführung eines Bernoulli-Experiments gegen die Erfolgswahrscheinlichkeit  $p$ .
- Letztendlich ist dies auch eine Rechtfertigung für den häufigkeitsbasierten Wahrscheinlichkeitsbegriff!

## Satz 11.3 (Zentraler Grenzwertsatz)

Es seien  $X_i$  für  $i \in \mathbb{N}$  unabhängig identisch verteilte Zufallsvariablen mit  $E(X_i) = \mu_X \in \mathbb{R}$  und  $\text{Var}(X_i) = \sigma_X^2 > 0$ . Für  $n \in \mathbb{N}$  seien die Summen  $Y_n := \sum_{i=1}^n X_i$  bzw. die standardisierten Summen bzw. Mittelwerte

$$Z_n := \frac{Y_n - E(Y_n)}{\sqrt{\text{Var}(Y_n)}} = \frac{(\sum_{i=1}^n X_i) - n\mu_X}{\sigma_X \sqrt{n}} = \frac{(\frac{1}{n} \sum_{i=1}^n X_i) - \mu_X}{\sigma_X} \sqrt{n}$$

von  $X_1, \dots, X_n$  definiert.

Dann gilt für die Verteilungsfunktionen  $F_{Z_n}$  der Zufallsvariablen  $Z_n$

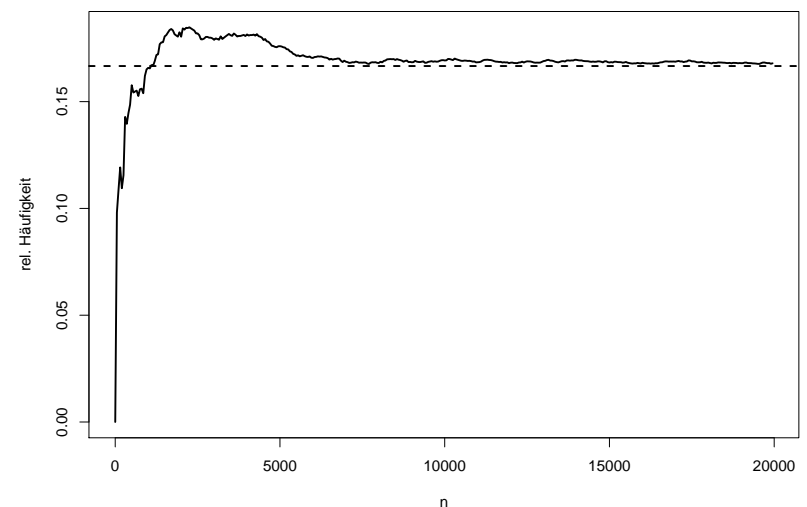
$$\lim_{n \rightarrow \infty} F_{Z_n}(z) = F_{N(0,1)}(z) = \Phi(z) \quad \text{für alle } z \in \mathbb{R},$$

man sagt auch, die Folge  $Z_n$  von Zufallsvariablen konvergiere in Verteilung gegen die Standardnormalverteilung, in Zeichen  $Z_n \xrightarrow{n \rightarrow \infty} N(0, 1)$ .

- Mit Hilfe des zentralen Grenzwertsatzes lassen sich insbesondere
  - weitere (schwächere oder speziellere) Grenzwertsätze zeigen,
  - Wahrscheinlichkeiten von Summen von i.i.d. Zufallsvariablen näherungsweise mit Hilfe der Standardnormalverteilung auswerten.

## Veranschaulichung „Schwachtes Gesetz der großen Zahlen“

(relative Häufigkeit des Auftretens der Zahl 6 beim Würfeln)



## Grenzwertsatz von de Moivre-Laplace

- *Erinnerung:* (siehe Satz 11.2, Folie 296)  
Summen i.i.d. bernoullivertelter Zufallsvariablen sind binomialverteilt.
- Die Anwendung des zentralen Grenzwertsatzes 11.3 auf binomialverteilte Zufallsvariablen als Summen i.i.d. bernoullivertelter Zufallsvariablen führt zur (zumindest historisch wichtigen und beliebten) Näherung von Binomialverteilungen durch Normalverteilungen.
- Resultat ist auch als eigenständiger Grenzwertsatz bekannt:

### Satz 11.5 (Grenzwertsatz von de Moivre-Laplace)

Für  $n \in \mathbb{N}$  und  $p \in (0, 1)$  sei  $Y_n \sim B(n, p)$ . Dann konvergieren die standardisierten Zufallsvariablen

$$Z_n := \frac{Y_n - np}{\sqrt{np(1-p)}}$$

in Verteilung gegen die Standardnormalverteilung, es gilt also

$$\lim_{n \rightarrow \infty} F_{Z_n}(z) = F_{N(0,1)}(z) = \Phi(z) \quad \text{für alle } z \in \mathbb{R}.$$

## Beispiele I

zur Anwendung des zentralen Grenzwertsatzes

Für  $Y_n \sim B(n, p)$  erhält man beispielsweise die Näherung

$$P\{Y_n \leq z\} = F_{Y_n}(z) \approx F_{N(np, np(1-p))}(z) = \Phi\left(\frac{z - np}{\sqrt{np(1-p)}}\right)$$

oder gleichbedeutend

$$\begin{aligned} P\{Y_n \leq z\} &= P\left\{\frac{Y_n - np}{\sqrt{np(1-p)}} \leq \frac{z - np}{\sqrt{np(1-p)}}\right\} = P\left\{Z_n \leq \frac{z - np}{\sqrt{np(1-p)}}\right\} \\ &\approx F_{N(0,1)}\left(\frac{z - np}{\sqrt{np(1-p)}}\right) = \Phi\left(\frac{z - np}{\sqrt{np(1-p)}}\right). \end{aligned}$$

## Anwendung der Grenzwertsätze für Näherungen

- Grenzwertsätze treffen nur Aussagen für  $n \rightarrow \infty$ .
- In praktischen Anwendungen wird verwendet, dass diese Aussagen für endliche, aber hinreichend große  $n$ , schon „näherungsweise“ gelten.
- Eine häufige verwendete „Faustregel“ zur Anwendung des Grenzwertsatzes von de Moivre-Laplace ist zum Beispiel  $np(1-p) \geq 9$ .
- (Äquivalente!) Anwendungsmöglichkeit der Grenzwertsätze für endliches  $n$ : Verwendung

▶ der  $N(n\mu_X, n\sigma_X^2)$ -Verteilung für  $Y_n := \sum_{i=1}^n X_i$  oder

▶ der  $N\left(\mu_X, \frac{\sigma_X^2}{n}\right)$ -Verteilung für  $\bar{X}_n := \frac{1}{n}Y_n = \frac{1}{n}\sum_{i=1}^n X_i$  oder

▶ der Standardnormalverteilung für  $Z_n := \frac{Y_n - n\mu_X}{\sigma_X\sqrt{n}} = \frac{\bar{X}_n - \mu_X}{\sigma_X}\sqrt{n}$

mit  $\mu_X = E(X_i)$  und  $\sigma_X^2 = \text{Var}(X_i)$  statt der jeweiligen exakten Verteilung.

- Gilt  $X_i \stackrel{i.i.d.}{\sim} N(\mu_X, \sigma_X^2)$ , stimmen die „Näherungen“ sogar mit den exakten Verteilungen überein (siehe Folie 296)!

## Beispiele II

zur Anwendung des zentralen Grenzwertsatzes

Gilt  $Y \sim B(5000, 0.2)$ , so erhält man für die Wahrscheinlichkeit, dass  $Y$  Werte  $> 1000$  und  $\leq 1050$  annimmt, näherungsweise

$$\begin{aligned} P\{1000 < Y \leq 1050\} &= F_Y(1050) - F_Y(1000) \\ &\approx \Phi\left(\frac{1050 - 5000 \cdot 0.2}{\sqrt{5000 \cdot 0.2 \cdot (1-0.2)}}\right) - \Phi\left(\frac{1000 - 5000 \cdot 0.2}{\sqrt{5000 \cdot 0.2 \cdot (1-0.2)}}\right) \\ &= \Phi(1.77) - \Phi(0) = 0.9616 - 0.5 = 0.4616 \end{aligned}$$

(Exakte Wahrscheinlichkeit: 0.4538)



## Beispiele III

zur Anwendung des zentralen Grenzwertsatzes

Seien  $X_i \stackrel{i.i.d.}{\sim} \text{Pois}(5)$  für  $i \in \{1, \dots, 100\}$  (es gilt also  $E(X_i) = 5$  und  $\text{Var}(X_i) = 5$ ), sei  $Y := \sum_{i=1}^{100} X_i$ . Dann gilt für das untere Quartil  $y_{0.25}$  von  $Y$

$$F_Y(y_{0.25}) \approx F_{N(100 \cdot 5, 100 \cdot 5)}(y_{0.25}) = \Phi\left(\frac{y_{0.25} - 500}{\sqrt{500}}\right) \stackrel{!}{=} 0.25$$

und unter Verwendung von  $\Phi(z) = 1 - \Phi(-z)$  bzw.  $\Phi(-z) = 1 - \Phi(z)$  weiter

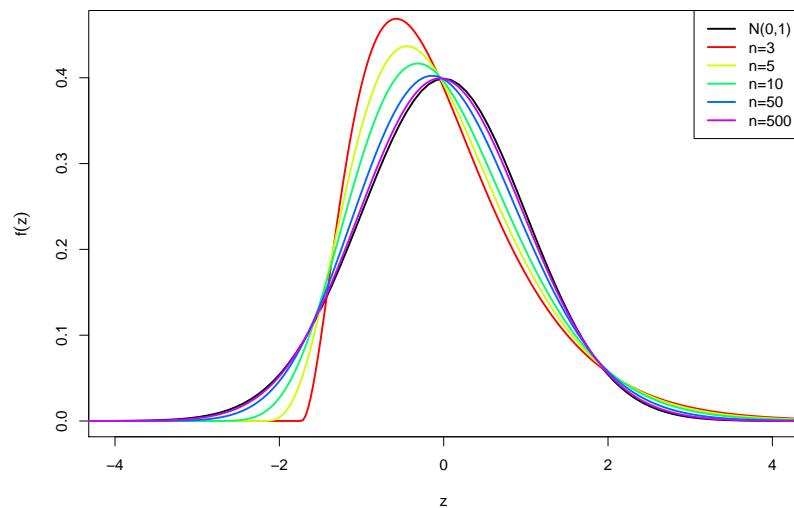
$$\Phi\left(\frac{500 - y_{0.25}}{\sqrt{500}}\right) \stackrel{!}{=} 1 - 0.25 = 0.75 \Rightarrow \frac{500 - y_{0.25}}{\sqrt{500}} = \Phi^{-1}(0.75) \approx 0.675$$

$$\Rightarrow y_{0.25} \approx 500 - 0.675 \cdot \sqrt{500} = 484.9065$$

(exaktes unteres Quartil unter Verwendung von  $Y \sim \text{Pois}(500)$ ):  $y_{0.25} = 485$ )

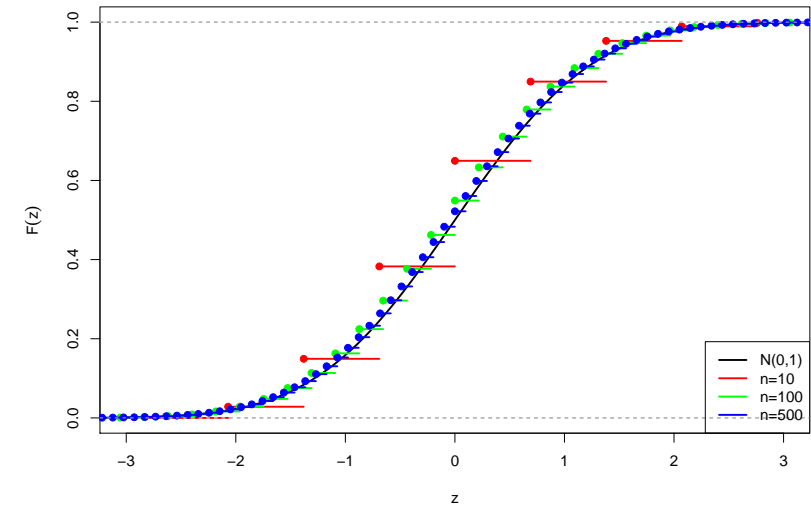
## Veranschaulichung des zentralen Grenzwertsatzes

an der Dichtefunktion standardisierter Summen von Exponentialverteilungen mit  $\lambda = 2$



## Veranschaulichung des zentralen Grenzwertsatzes

an der Verteilungsfunktion standardisierter Binomialverteilungen  $B(n, p)$  mit  $p = 0.3$



## Veranschaulichung des zentralen Grenzwertsatzes

an der Dichtefunktion standardisierter Summen von Unif(20, 50)-Verteilungen

