Momente höherdimensionaler Zufallsvektoren

Definition 10.8

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

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

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

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

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

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

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

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

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

- Ein n-dimensionaler Zufallsvektor $\mathbf{X} = (X_1, \ldots, X_n)'$ bzw. die n eindimensionalen Zufallsvariablen $X_1, \ldots, X_n$ heißen **unkorreliert**, wenn $\mathrm{V}(\mathbf{X})$ eine Diagonalmatrix ist, also höchstens die Diagonaleinträge von 0 verschieden sind.
- Es gilt in Verallgemeinerung von Folie 272 (bei Existenz der Momente):

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

die Umkehrung gilt *im allgemeinen* nicht!
- Dass stochastische Unabhängigkeit eine stärkere Eigenschaft ist als Unkorreliertheit, zeigt auch der folgende Satz:

Satz 10.2

Seien für $n \in \mathbb{N}$ stochastisch unabhängige Zufallsvariablen $X_1, \ldots, X_n$ über einem gemeinsamen Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, P)$ sowie $\mathcal{B} - \mathcal{B}-$messbare Funktionen $G_i : \mathbb{R} \to \mathbb{R}$ für $i \in \{1, \ldots, n\}$ gegeben.
Dann sind auch die Zufallsvariablen $G_1(X_1), \ldots, G_n(X_n)$ stochastisch unabhängig.

Momente von Summen von Zufallsvariablen

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

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

von n Zufallsvariablen $X_1, \ldots, X_n$ interessiert.
- In Verallgemeinerung von Folie 267 zeigt man für den Erwartungswert leicht

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

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

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

- In Verallgemeinerung von Folie 273 erhält man für die Varianz weiter

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

bzw.

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

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

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

bzw.

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

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

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

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

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

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

Satz 11.1 (Summen spezieller Zufallsvariablen)

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

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

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

1. $X_i \sim B(1, p)$ *für ein* $p \in (0, 1)$, *also insgesamt* $X_i \overset{i.i.d.}{\sim} B(1, p)$, *so gilt* $Y \sim B(n, p)$ *(vgl. Folie 230),*
2. $X_i \sim B(n_i, p)$ *für* $n_i \in \mathbb{N}$ *und ein* $p \in (0, 1)$, *so gilt* $Y \sim B(N, p)$ *mit* $N := n_1 + \cdots + n_n = \sum_{i=1}^{n} n_i$,
3. $X_i \sim \mathrm{Pois}(\lambda_i)$ *für* $\lambda_i \in \mathbb{R}_+$, *so gilt* $Y \sim \mathrm{Pois}(\lambda)$ *mit* $\lambda := \lambda_1 + \cdots + \lambda_n = \sum_{i=1}^{n} \lambda_i$,
4. $X_i \sim \mathrm{N}(\mu_i, \sigma_i^2)$ *für* $\mu_i \in \mathbb{R}$ *und* $\sigma_i^2 > 0$, *so gilt* $Y \sim \mathrm{N}(\mu, \sigma^2)$ *mit* $\mu = \mu_1 + \cdots + \mu_n = \sum_{i=1}^{n} \mu_i$ *und* $\sigma^2 = \sigma_1^2 + \cdots + \sigma_n^2 = \sum_{i=1}^{n} \sigma_i^2$,
5. $X_i \sim \mathrm{N}(\mu, \sigma^2)$ *für ein* $\mu \in \mathbb{R}$ *und ein* $\sigma^2 > 0$, *also insgesamt* $X_i \overset{i.i.d.}{\sim} \mathrm{N}(\mu, \sigma^2)$, *so gilt für* $\overline{X} := \frac{1}{n}(X_1 + \cdots + X_n) = \frac{1}{n}\sum_{i=1}^{n} X_i$ *insbesondere* $\overline{X} \sim \mathrm{N}(\mu, \frac{\sigma^2}{n})$.

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 $\mathrm{E}(X_i) = \mu_X \in \mathbb{R}$ und $\mathrm{Var}(X_i) = \sigma_X^2 > 0$.
- Gilt $n \to \infty$ für die Anzahl n der summierten Zufallsvariablen X_i, gilt für die Summen $Y_n := \sum_{i=1}^n X_i$ offensichtlich $\mathrm{Var}(Y_n) = n \cdot \sigma_X^2 \to \infty$ und für $\mu_X \neq 0$ auch $\mathrm{E}(Y_n) = n \cdot \mu_X \to \pm\infty$, eine Untersuchung der Verteilung von Y_n für $n \to \infty$ ist also schwierig.
- Stattdessen: Betrachtung der **standardisierten** Zufallsvariablen
$$Z_n := \frac{Y_n - \mathrm{E}(Y_n)}{\sqrt{\mathrm{Var}(Y_n)}} = \frac{\left(\sum_{i=1}^n X_i\right) - n\mu_X}{\sigma_X \sqrt{n}} = \frac{\left(\frac{1}{n}\sum_{i=1}^n X_i\right) - \mu_X}{\sigma_X}\sqrt{n}$$
mit $\mathrm{E}(Z_n) = 0$ und $\mathrm{Var}(Z_n) = 1$ für alle $n \in \mathbb{N}$.
- Gemäß folgendem Satz „konvergiert“ die Verteilung der Z_n für $n \to \infty$ gegen eine Standardnormalverteilung:

Satz 11.2 (Zentraler Grenzwertsatz)

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

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

von $X_1, \ldots, X_n$ definiert.
Dann gilt für die Verteilungsfunktionen F_{Z_n} der Zufallsvariablen Z_n

$$\lim_{n\to\infty} F_{Z_n}(z) = F_{\mathrm{N}(0,1)}(z) = \Phi(z) \qquad \textit{für alle } z \in \mathbb{R}\ ,$$

man sagt auch, die Folge Z_n von Zufallsvariablen konvergiere in Verteilung *gegen die Standardnormalverteilung, in Zeichen $Z_n \overset{n\to\infty}{\longrightarrow} \mathrm{N}(0,1)$.*

- 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.

Das schwache Gesetz der großen Zahlen

- Eine direkte Folge von Satz 11.2 ist zum Beispiel das folgende Resultat:

Satz 11.3 (Schwaches Gesetz der großen Zahlen)

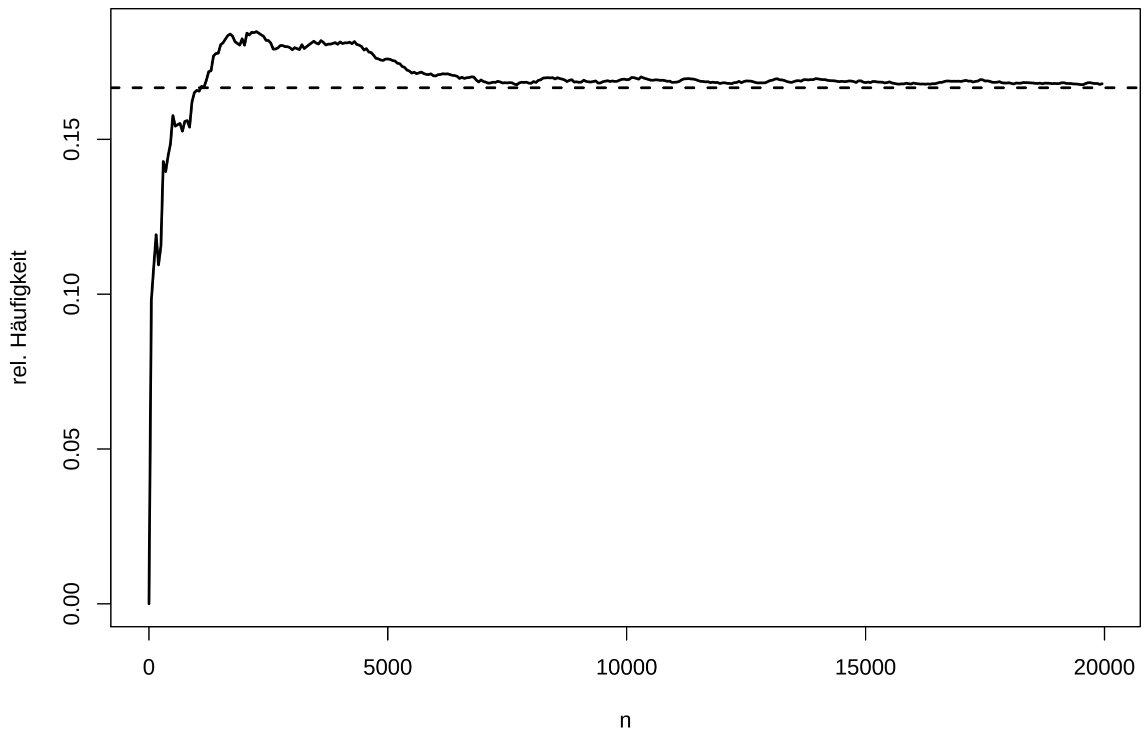
Es seien X_i für $i \in \mathbb{N}$ unabhängig identisch verteilte Zufallsvariablen mit $\mathrm{E}(X_i) = \mu_X \in \mathbb{R}$. Dann gilt:

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

- 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 $n \to \infty$ 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!

Veranschaulichung „Schwaches Gesetz der großen Zahlen"

(relative Häufigkeit des Auftretens der Zahl 6 beim Würfeln)



Anwendung der Grenzwertsätze für Näherungen

- Grenzwertsätze treffen nur Aussagen für $n \to \infty$.
- 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 $\overline{X}_n := \frac{1}{n} Y_n = \frac{1}{n}\sum_{i=1}^{n} X_i$ oder
  - der Standardnormalverteilung für $Z_n := \frac{Y_n - n\mu_X}{\sigma_X\sqrt{n}} = \frac{\overline{X}_n - \mu_X}{\sigma_X}\sqrt{n}$

  mit $\mu_X = \mathrm{E}(X_i)$ und $\sigma_X^2 = \mathrm{Var}(X_i)$ statt der jeweiligen exakten Verteilung.
- Gilt $X_i \overset{i.i.d.}{\sim} N(\mu_X, \sigma_X^2)$, stimmen die „Näherungen" sogar mit den exakten Verteilungen überein (siehe Folie 281)!

Beispiele I

zur Anwendung des zentralen Grenzwertsatzes

Für $Y_n \sim B(n,p)$ erhält man mit $\mathrm{E}(Y_n) = np$ und $\mathrm{Var}(Y_n) = np(1-p)$ 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

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

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 ≤ 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 $X_i \overset{i.i.d}{\sim} \text{Pois}(5)$ für $i \in \{1, \ldots, 100\}$ (es gilt also $\text{E}(X_i) = 5$ 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

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

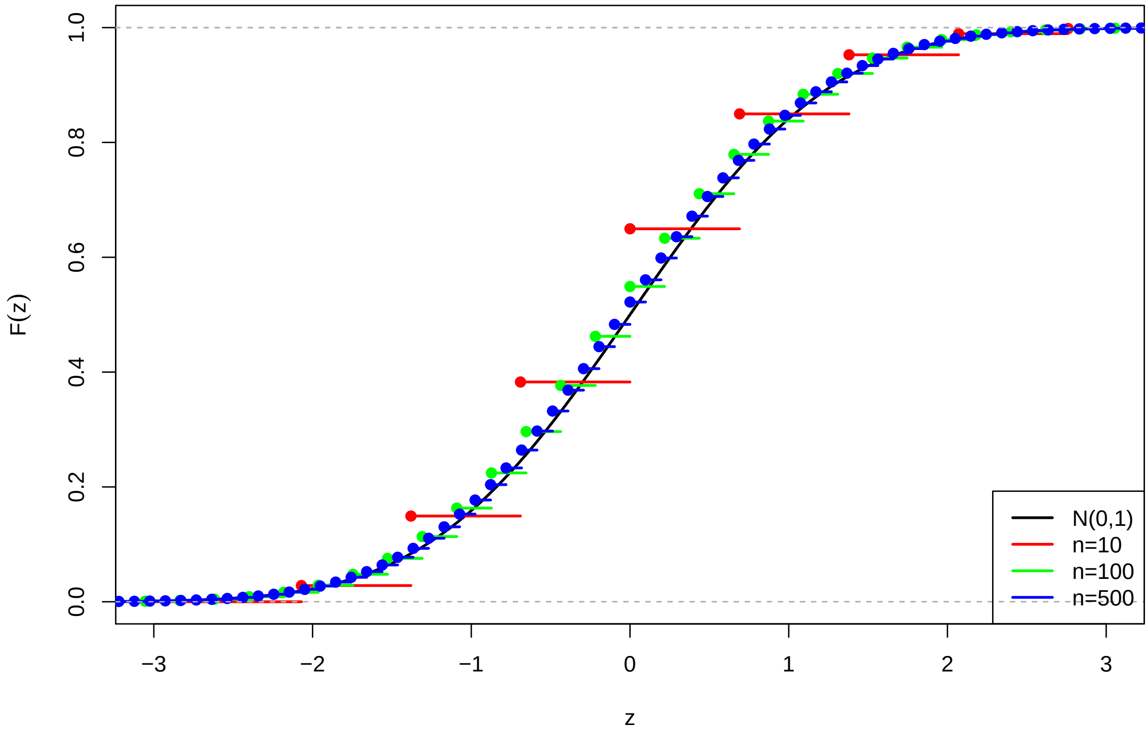
und unter Verwendung von $\Phi(z) = 1 - \Phi(-z)$ bzw. $\Phi(-z) = 1 - \Phi(z)$ weiter

$$\begin{aligned} \Phi\left(\frac{500-y_{0.25}}{\sqrt{500}}\right) &\overset{!}{=} 1 - 0.25 = 0.75 \quad \Rightarrow \quad \frac{500-y_{0.25}}{\sqrt{500}} = \Phi^{-1}(0.75) \approx 0.675 \\ \Rightarrow \quad y_{0.25} &\approx 500 - 0.675 \cdot \sqrt{500} = 484.9065 \end{aligned}$$

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

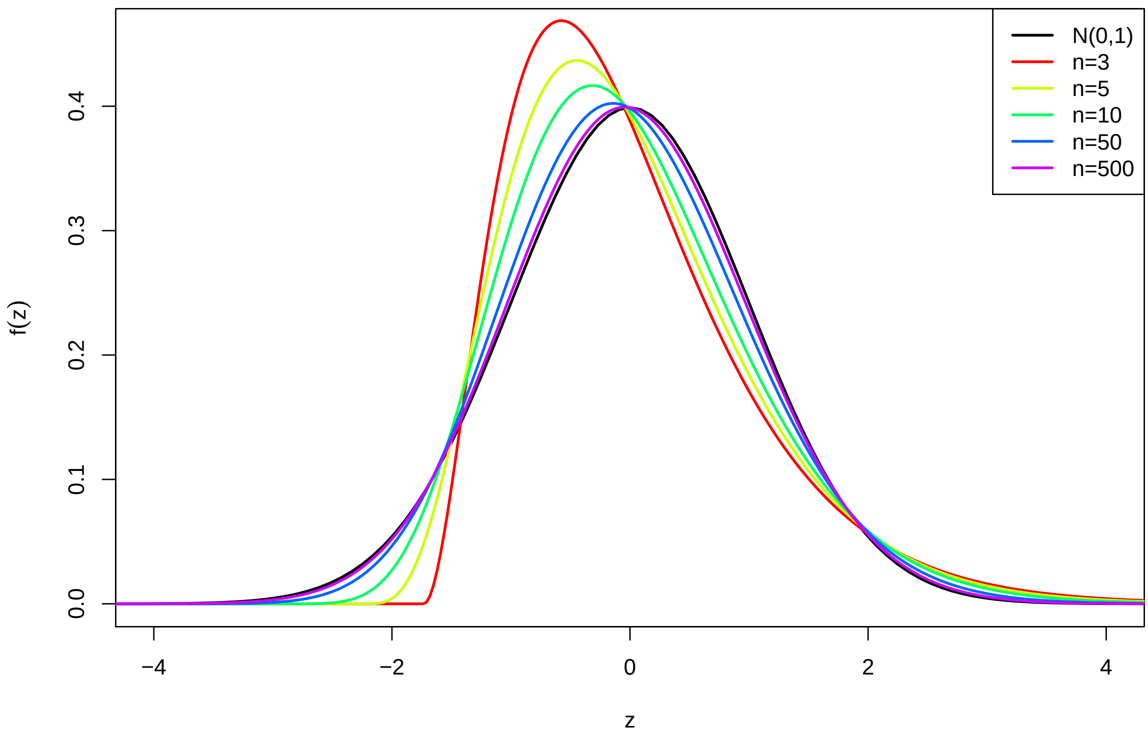
Veranschaulichung des zentralen Grenzwertsatzes

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



Veranschaulichung des zentralen Grenzwertsatzes

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



Veranschaulichung des zentralen Grenzwertsatzes

an der Dichtefunktion standardisierter Summen von $\text{Unif}(20, 50)$-Verteilungen