

Vorlesung 7a

Normalverteilung und Zentraler Grenzwertsatz

Teil 3

Die Botschaft des Zentralen Grenzwertsatzes

(Buch S. 77)

Zur Erinnerung:

(Z_1, \dots, Z_n) heißt standard-normalverteilt im \mathbb{R}^n

$:\Leftrightarrow$

Z_1, \dots, Z_n sind unabhängig und $N(0,1)$ -verteilt

Auch $\frac{Z_1 + \dots + Z_n}{\sqrt{n}}$ ist dann $N(0,1)$ -verteilt!

Der Zentrale Grenzwertsatz liefert

eine gewaltige Weiterung der vorigen Aussage

(asymptotisch für große n):

Zentraler Grenzwertsatz (Version (0,1)) salopp formuliert:

“Die durch \sqrt{n} geteilte Summe von
 n unabhängigen, identisch verteilten

nicht notwendig normalverteilten

\mathbb{R} -wertigen Zufallsvariablen

mit Erwartungswert 0 und Varianz 1

ist für große n annähernd standard-normalverteilt”

Satz (ZGWS, “Version (0,1)”)

Seien X_1, X_2, \dots

unabhängige und identisch verteilte Zufallsvariable
mit Erwartungswert 0 und Varianz 1.

Dann gilt für alle $c < d \in \mathbb{R}$

$$\mathbf{P}\left(\frac{X_1 + \dots + X_n}{\sqrt{n}} \in [c, d]\right) \xrightarrow{n \rightarrow \infty} \mathbf{P}(Z \in [c, d]).$$

Dabei ist Z standard-normalverteilt.

Jetzt: (μ, σ^2) statt $(0, 1)$:

Hat X den Erwartungswert μ und die Varianz σ^2 ,
dann hat $\frac{X - \mu}{\sigma}$ Erwartungswert 0 und Varianz 1.

Man nennt $\frac{X - \mu}{\sigma}$ auch die *Standardisierung* von X .

Sind X_1, X_2, \dots unabhängige Kopien von X ,
dann ergibt der ZGWS (Version $(0,1)$)
angewandt auf die Standardisierungen der X_i :

Zentraler Grenzwertsatz (Klassische Version)

Seien X_1, X_2, \dots

unabhängige und identisch verteilte Zufallsvariable mit endlichem Erwartungswert μ und endlicher Varianz $\sigma^2 > 0$.

Dann gilt für alle $c < d \in \mathbb{R}$

$$\mathbf{P}\left(\frac{X_1 + \dots + X_n - n\mu}{\sqrt{n\sigma^2}} \in [c, d]\right) \xrightarrow{n \rightarrow \infty} \mathbf{P}(Z \in [c, d]).$$

Dabei ist Z standard-normalverteilt.

Der Münzwurf passt in den Rahmen des klassischen ZGWS:

(X_i) sei eine Bernoulli-Folge zum Parameter p ,
also insbesondere:

X_1, X_2, \dots unabhängig und identisch verteilt mit

$$\mathbf{E}[X_i] = \mathbf{P}(X_i = 1) = p, \quad \mathbf{Var}[X_i] = pq$$

Dann gilt für alle $c < d \in \mathbb{R}$:

$$\mathbf{P} \left(\frac{X_1 + \dots + X_n - np}{\sqrt{npq}} \in [c, d] \right) \xrightarrow{n \rightarrow \infty} \frac{1}{\sqrt{2\pi}} \int_c^d e^{-a^2/2} da.$$

Das ist der Satz von de Moivre (1733, für $p = 1/2$)
und Laplace (1812, für allgemeines p).

Der klassische Zentrale Grenzwertsatz in anschaulichen Worten:

“Die standardisierte Summe von **VIELEN**
unabhängigen, identisch verteilten
nicht notwendig normalverteilten
 \mathbb{R} -wertigen Zufallsvariablen
mit endlicher Varianz
ist annähernd standard-normalverteilt”

Manchmal noch praktischer so:

“Die standardisierte Summe von VIELEN
unabhängigen, identisch verteilten
nicht notwendig normalverteilten
 \mathbb{R} -wertigen Zufallsvariablen
mit endlicher Varianz
ist annähernd standard-normalverteilt”

Beispiel:

X_1, X_2, \dots seien unabhängig und identisch verteilt
mit Erwartungswert μ und Varianz σ^2

Dann ist für großes n die Zufallsvariable

$$M_n := \frac{X_1 + \dots + X_n}{n}$$

approximativ $N(\mu, \frac{\sigma^2}{n})$ -verteilt.

Spezialfall des vorigen Beispiels:

Ziehen mit Zurücklegen

S sei eine endliche Menge mit $g := \#S$,

$h : S \rightarrow \mathbb{R}$ sei eine Abbildung (ein “individuelles Merkmal”)

$$\mu := \frac{1}{g} \sum_{a \in S} h(a) \quad \dots \text{ Populationsmittelwert}$$

$$\sigma^2 := \frac{1}{g} \sum_{a \in S} (h(a) - \mu)^2 \quad \dots \text{ Populationsvarianz}$$

J_1, J_2, \dots seien unabhängig und uniform auf S verteilt

Die Zufallsvariablen $X_i := h(J_i)$, $i = 1, 2, \dots$

passen dann in den Rahmen des vorigen Beispiels.

“Die standardisierte Summe von VIELEN
unabhängigen, identisch verteilten
nicht notwendig normalverteilten
 \mathbb{R} -wertigen Zufallsvariablen
mit endlicher Varianz
ist annähernd standard-normalverteilt”

Diese Aussage bleibt auch
unter schwächeren Bedingungen bestehen,
sowohl was die Unabhängigkeit,
als auch was die identische Verteiltheit betrifft.

Eine Botschaft zum Mitnehmen ins Leben
(salopp formuliert):

“ Die Summe von vielen
annähernd unabhängigen Zufallsvariablen,
die nicht notwendig identisch verteilt, aber
ungefähr von derselben Größenordnung sind,
ist annähernd normalverteilt.”

Beispiel: **Ziehen ohne Zurücklegen.**

S sei eine endliche Menge mit $g := \#S$,

$h : S \rightarrow \mathbb{R}$ sei eine Abbildung (ein “individuelles Merkmal”)

J_1, J_2, \dots rein zufälliges Ziehen **ohne** Zurücklegen aus S .

Dann ist für großes n und großes $g - n$

$$M_n := \frac{1}{n} \sum_{i=1}^n h(J_i)$$

annähernd normalverteilt.

Beispiel: **Ziehen ohne Zurücklegen.**

S sei eine endliche Menge mit $g := \#S$,

$h : S \rightarrow \mathbb{R}$ sei eine Abbildung (ein “individuelles Merkmal”)

J_1, J_2, \dots rein zufälliges Ziehen **ohne** Zurücklegen aus S .

Dann ist für großes n und großes $g - n$

$$M_n := \frac{1}{n} \sum_{i=1}^n h(J_i)$$

annähernd normalverteilt.

(Zur Berechnung von $\mathbf{Var}[M_n]$ aus der Populationsvarianz siehe ÜA 24 + ein Übungsbeispiel auf dem Weihnachtsblatt.)

Zentraler Grenzwertsatz: Meilensteine in seiner Geschichte

Abraham **de Moivre**:

Der faire Münzwurf (1733)

Pierre-Simon **Laplace**:

Allgemeine binomiale Zufallsgrößen (1812)

Pafnuty Lvovich **Chebyshev**:

Skizze eines Beweises für den allgemeinen Fall (1887)

Aleksandr Mikhailovich **Lyapunov**:

“Klassischer” zentraler Grenzwertsatz (1901)

Noch allgemeiner (1906)

Andrei Andreyevich **Markov**:

weitere Verallgemeinerungen (~ 1910)