Data Science 2

Prof. Dr. Mark Trede

Institut für Ökonometrie und Wirtschaftsstatistik

September 2023

Grenzwertsätze

- Gegeben sei eine Folge von unabhängigen, identisch verteilten Zufallsvariablen (i.i.d.)

$$X_1, X_2, X_3, \ldots,$$

- Die Folgenelemente $X_1, X_2, \ldots$ heißen auch unabhängige Wiederholungen von X

- Für gegebenes n ist das arithmetische Mittel

$$\bar{X}_n = \frac{1}{n} \sum_{i=1}^{n} X_i$$

- Achtung: $\bar{X}_n$ ist eine **Zufallsvariable**!

- Was passiert mit der Verteilung von $\bar{X}_n$, wenn $n \to \infty$ geht?
- Sei $E(X) = \mu$ und $Var(X) = \sigma^2$
- Dann gilt

$$E(\bar{X}_n) = \mu$$
$$Var(\bar{X}_n) = \frac{\sigma^2}{n}$$

Grenzwertsätze

Gesetz der großen Zahl

$$E(\bar{X}_n) =$$

$$Var(\bar{X}_n) =$$

Schwaches Gesetz der großen Zahl

Für jedes (noch so kleine) $\varepsilon > 0$ gilt

$$\lim_{n \to \infty} P(|\bar{X}_n - \mu| \geq \varepsilon) = 0$$

Alternative Schreibweise

$$\text{plim}_{n \to \infty} \bar{X}_n = \mu$$

Grenzwertsätze

Gesetz der großen Zahl

- Anschaulich: Die Verteilung von $\bar{X}_n$ zieht sich immer mehr auf μ zusammen
- Spezialfall: X sei Bernoulli-verteilt mit Parameter $P(X=1)=\pi$
- Dann ist $\bar{X}_n$ die relative Häufigkeit der Erfolge
- Wegen $E(X)=\pi$ gilt

$$\lim_{n\to\infty} P\left(\left|\bar{X}_n-\pi\right|\geq\varepsilon\right)=0$$

Beispiel: Ein Würfel wird geworfen

Sei

$$X = \begin{cases} 0 & \text{wenn Augenzahl nicht 5 ist} \\ 1 & \text{wenn Augenzahl 5 ist} \end{cases}$$

Es gilt

$$P(X = 1) = \frac{1}{6}$$

Forts. Beispiel: Ein Würfel wird geworfen

- Der Würfel wird nun sehr oft geworfen (n Mal)
- $X_1, X_2, \ldots$ geben jeweils an, ob eine 5 geworfen wurde
- $\bar{X}_n$ ist der Anteil der Fünfen
- Für großes n geht $\bar{X}_n$ gegen $1/6$
- Simulation in R: [`wlln.R`]

Beispiel: Produktgewicht

- Die Zufallsvariable $X \sim N(201, 4)$ sei das tatsächliche Gewicht einer 200g-Tafel Schokolade
- X_i ist das Gewicht der i-ten Tafel, $i = 1, 2, \ldots$
- $\bar{X}_n$ ist das Durchschnittsgewicht dieser Tafeln
- Für großes n geht $\bar{X}_n$ gegen 201 [`wlln.R`]

- Zentraler Grenzwertsatz (Begründung für die extreme Wichtigkeit der Normalverteilung)
- Definiere das standardisierte arithmetische Mittel

$$U_n = \frac{\bar{X}_n - \mu}{\sqrt{\sigma^2/n}} = \sqrt{n}\frac{\bar{X}_n - \mu}{\sigma}$$

Zentraler Grenzwertsatz

Für alle $u \in \mathbb{R}$ gilt

$$\lim_{n\to\infty} P(U_n \leq u) = \Phi(u),$$

wobei Φ die Verteilungsfunktion der $N(0,1)$ ist

Für großes n gilt approximativ

$$U_n \overset{appr}{\sim} N(0,1)$$

- Folglich gilt auch

$$\sum_{i=1}^{n} X_i \overset{appr}{\sim} N(n\mu, n\sigma^2)$$

und

$$\bar{X}_n \overset{appr}{\sim} N\left(\mu, \frac{\sigma^2}{n}\right)$$

- Die Summe und der Durchschnitt von n **beliebig verteilten** Zufallsvariablen ist approximativ normalverteilt, wenn n groß genug ist!
- Es gibt einige einschränkende Bedingungen, aber in den meisten Situationen gilt der zentrale Grenzwertsatz
- Simulationen in R

- Spezialfall: X Bernoulli-verteilt mit π

$$E(X) = \pi$$
$$V(X) = \pi\,(1-\pi)$$

und daher

$$\sum_{i=1}^{n} X_i \overset{appr}{\sim} N(n\pi, n\pi(1-\pi))$$

- Approximation der Binomialverteilung durch die Normalverteilung (De Moivre, 1733)
- Wegen
$$\sum_{i=1}^{n} X_i \overset{appr}{\sim} N(n\pi, n\pi(1-\pi))$$
gilt
$$P\left(\sum_{i=1}^{n} X_i \leq b\right) \approx \Phi\left(\frac{b - n\pi}{\sqrt{n\pi\,(1-\pi)}}\right)$$

Beispiel: Marketing

Eine Marketing-Abteilung verschickt an $n = 500$ zufällig ausgewählte Kunden Fragebögen.

X_i sind Bernoulli-verteilt mit Parameter $\pi = 0.2$

$$X_i = \begin{cases} 1 & \text{wenn Kunde } i \text{ antwortet} \\ 0 & \text{wenn Kunde } i \text{ nicht antwortet} \end{cases}$$

Sei $Y = \sum_{i=1}^{500} X_i$ die Zahl der Antworten.

Forts. Beispiel: Marketing

Die Wahrscheinlichkeit, dass zwischen 95 und 105 Kunden antworten, ist