Kapitel 3

Grenzwertsätze

Contents

1 Ungleichungen

Satz 3.1 *Es sei X eine zufällige Variable. Dann gilt:*

$$\min_{c \in \mathbb{R}} \mathbf{E}(X-c)^2 = \operatorname{Var} X.$$

Beweis: Für alle reellen Zahlen $c \in \mathbb{R}$ gilt:

$$\mathbf{E}(X-c)^2 = \mathbf{E}(X - \mathbf{E}X + \mathbf{E}X - c)^2$$

$$\begin{aligned}
&= \mathbf{E}(X-\mathbf{E}X)^2 + 2\mathbf{E}((X-\mathbf{E}X)(\mathbf{E}X-c)) + (\mathbf{E}X-c)^2 \\
&= \mathbf{E}(X-\mathbf{E}X)^2 + 2(\mathbf{E}X-c)\underbrace{\mathbf{E}(X-\mathbf{E}X)}_{=0} + (\mathbf{E}X-c)^2 \\
&= \operatorname{Var} X + (\mathbf{E}X-c)^2 \\
&\geq \operatorname{Var} X
\end{aligned}$$

Setzen wir $c := \mathbf{E}X$ erhalten wir Gleichheit. $\square$

Satz 3.2 (Ungleichung von JENSEN**)** *Sei X eine zufällige Variable mit $\mathbf{E}X < \infty$ und g eine differenzierbare und konvexe Funktion. Dann gilt:*

$$\mathbf{E}g(X) \geq g(\mathbf{E}X).$$

Beweis: Sei $T(x)$ die Tangente an die Kurve der Funktion g im Punkt x_0,

$$g(x) \geq T(x) = g(x_0) + \underbrace{g'(x_0)}_{\text{Anstieg der Kurve in } x_0} \cdot (x - x_0).$$

Wir setzen nun $x := X$ und $x_0 := \mathbf{E}X$ und erhalten:

$$g(X) \geq g(\mathbf{E}X) + g'(\mathbf{E}X) \cdot (X - \mathbf{E}X).$$

Daraus folgt:

$$\begin{aligned}
\mathbf{E}g(X) &\geq \mathbf{E}(g(\mathbf{E}X) + g'(\mathbf{E}X) \cdot (X - \mathbf{E}X)) \\
&= g(\mathbf{E}X) + g'(\mathbf{E}X) \cdot \underbrace{\mathbf{E}(X - \mathbf{E}X)}_{=0} \\
&= g(\mathbf{E}X)
\end{aligned}$$

□

Folg. 6 *Es sei g differenzierbar und konkav. Weiterhin sei X eine zufällige Variable. Dann gilt:*

$$\mathbf{E}g(X) \leq g(\mathbf{E}X).$$

Beweis: Da die Funktion g nach Voraussetzung konkav ist, ist die Funktion $(-g)$ konvex. Dann gilt nach Satz 3.2:

$$\mathbf{E}((-g)(X)) \geq (-g)(\mathbf{E}X).$$

Daraus folgt:

$$-\mathbf{E}g(X) \geq -g(\mathbf{E}X).$$

$$\mathbf{E}g(X) \leq g(\mathbf{E}X).$$

$\square$

Bsp. 3.1 *1. Es sei $g(x) = x^2$. Dann gilt nach Satz 3.2:*

$$\mathbf{E}X^2 \geq (\mathbf{E}X)^2.$$

Daraus aber folgt (die schon bekannte Aussage):

$$\operatorname{Var} X = \mathbf{E}(X - \mathbf{E}X)^2 = \mathbf{E}X^2 - (\mathbf{E}X)^2 \geq 0.$$

2. Es sei $g(x) = |x|$. Dann gilt nach Satz 3.2:

$$\mathbf{E}|X| \geq |\mathbf{E}X|.$$

3. Es sei $g(x) = \ln x$. Diese Funktion ist konkav. Also gilt nach Folgerung 6:

$$\mathbf{E}(\ln X) \leq \ln(\mathbf{E}X).$$

Satz 3.3 (Ungleichung von MARKOFF) *Sei* $c > 0$. X *sei eine Zufallsgröße. Dann gilt:*

$$P(|X| > c) \leq \frac{\mathbf{E}|X|}{c}.$$

Beweis: Wir definieren eine Zufallsgröße Y:

$$Y(\omega) := \begin{cases} c & \text{, falls } |X(\omega)| > c \\ 0 & \text{, falls } |X(\omega)| \leq c \end{cases}, \quad \forall \omega \in \Omega.$$

Wir können also schreiben:

$$Y : \quad \begin{pmatrix} 0 & c \\ P(|X| \leq c) & P(|X| > c) \end{pmatrix}$$

Offenbar gilt für alle $\omega \in \Omega$:

$$0 \leq Y(\omega) \leq |X(\omega)|,$$

beziehungsweise:

$$0 \leq Y \leq |X|.$$

Daraus folgt: $P(|X| - Y \geq 0) = 1$. Nach Satz 2.4:

$$\begin{aligned} \mathbf{E}(|X| - Y) &\geq 0 \\ \mathbf{E}|X| &\geq EY. \end{aligned}$$

Da die Zufallsgröße Y diskret ist, folgt aus der Definition des Erwartungswertes:

$$\begin{aligned}\mathbf{E}Y &= 0 \cdot P(|X| \leq c) + c \cdot P(|X| > c) \\ &= c \cdot P(|X| > c) \leq \mathbf{E}|X|\end{aligned}$$

Wir stellen um und erhalten:

$$P(|X| > c) \leq \frac{\mathbf{E}|X|}{c}.$$

$\square$

Satz 3.4 (Ungleichung von TSCHEBYCHEW**)** *Es sei $\varepsilon > 0$ und sei Y eine Zufallsgröße. Dann gilt:*

$$P(|Y - \mathbf{E}Y| > \varepsilon) \leq \frac{\operatorname{Var} Y}{\varepsilon^2}.$$

Beweis: Wir verwenden die Ungleichung von MARKOFF:

$$P(|X| > c) \leq \frac{\mathbf{E}|X|}{c}.$$

Wir setzen folgendes in diese Ungleichung ein:

$$X := (Y - \mathbf{E}Y)^{2i} \geq 0, \quad c := \varepsilon^{2i} \qquad (i \in \mathbb{N}).$$

Da $\varepsilon > 0$ gilt, ist die Voraussetzung der Ungleichung von MARKOFF erfüllt. Wir erhalten also:

$$P\left((Y - \mathbf{E}Y)^{2i} > \varepsilon^{2i}\right) = P(|Y - \mathbf{E}Y| > \varepsilon) \leq \frac{\mathbf{E}(Y - \mathbf{E}Y)^{2i}}{\varepsilon^{2i}}.$$

Wenn wir nun $i := 1$ setzen, so ergibt sich:

$$P(|Y - \mathbf{E}Y| > \varepsilon) \leq \frac{\mathbf{E}(Y - \mathbf{E}Y)^2}{\varepsilon^2} = \frac{\operatorname{Var} Y}{\varepsilon^2}.$$

$\square$

Bem.: Aus der Ungleichung von TSCHEBYCHEW folgt:

$$P(|Y - EY| \leq \varepsilon) \geq 1 - \frac{\operatorname{Var} Y}{\varepsilon^2}.$$

Das ist die zweite Variante der Ungleichung von TSCHEBYCHEW.

Bsp. 3.2 *Es sei $X \sim \mathrm{N}(\mu, \sigma^2)$, also*

$$EX = \mu, \qquad \operatorname{Var} X = \sigma^2.$$

Wir setzen $\varepsilon := k \cdot \sigma$ ($k \in \mathbb{N}$) und erhalten dann mit der Ungleichung von TSCHEBYCHEW*:*

$$P(|X - \mu| \leq k \cdot \sigma) \geq 1 - \frac{\sigma^2}{k^2 \cdot \sigma^2} = 1 - \frac{1}{k^2}.$$

k	*exakt* $\Phi(k\sigma) - \Phi(-k\sigma)$	*Tschebychew-Ungleichung* $1 - \frac{1}{k^2}$
1	0.68629	0
2	0.9545	0.75
3	0.9973	0.89
4	≈ 1	0.93
5	≈ 1	0.96

Bem. 16 *Die Tschebyscheff-Ungleichung gilt für beliebig verteilte Zufallsvariablen, die Erwartungswert und Varianz besitzen, insbesondere liegt sie mit Wahrscheinlichkeit* ≥ 0.89 *im* 3σ*-Intervall.*

Bsp. 3.3 *Die Zahl* $med = med(X)$ *heißt Median der Zufallsvariablen* X*, falls*

$$P(X \leq med) \geq \frac{1}{2} \quad \textit{und} \quad P(X \geq med) \geq \frac{1}{2}$$

Aus der Markoff-Ungleichung folgt:

$$\frac{1}{2} \leq P(X \geq med) \leq \frac{\mathbf{E}X}{med}, \textbf{d.h.} med \leq 2 \cdot \mathbf{E}X$$

2 Das Gesetz der großen Zahlen

Der Erwartungswert einer zufälligen Variablen X ist in der Praxis meist nicht bekannt. Um ihn zu bestimmen, sammelt man Beobachtungen $X_1, X_2, \ldots, X_n$ ($n \in \mathbb{N}$) und bildet dann das arithmetische Mittel dieser Beobachtungen:

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

Dabei muß man jedoch beachten, daß die Beobachtungen $X_1, \ldots, X_n$ unabhängig oder wenigstens unkorreliert sind.

Daß diese Vorgehensweise korrekt ist, drückt der folgende Satz formal aus:

Satz 3.5 (Schwaches Gesetz der großen Zahlen) *Es seien $X_1, \ldots, X_n$ unkorrelierte zufällige Variablen mit $\mu := \mathbf{E}X_i$ und $\sigma^2 := \mathrm{Var}\, X_i$ (für alle $i = 1, \ldots, n$). Dann gilt für alle $\varepsilon > 0$:*

$$\lim_{n\to\infty} P(|\overline{X}_n - \mu| > \varepsilon) = 0.$$

Beweis: Zum Beweis des Satzes verwenden wir die Ungleichung von TSCHEBYCHEW (vgl. Satz 3.4).

Da die Zufallsgrößen $X_1, \ldots, X_n$ unkorreliert sind, gilt

$$\begin{aligned}\mathrm{Var}\,\overline{X} &= \mathrm{Var}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\cdot \mathrm{Var}\left(\sum_{i=1}^{n} X_i\right)\\ &= \frac{1}{n^2}\cdot \sum_{i=1}^{n}\mathrm{Var}\,(X_i) = \frac{1}{n^2}\cdot n\cdot \sigma^2 = \frac{\sigma^2}{n}\end{aligned}$$

$$\begin{aligned}\mathbf{E}\overline{X} &= \mathbf{E}\left(\frac{1}{n}\cdot\sum_{i=1}^{n} X_i\right) = \frac{1}{n}\cdot \mathbf{E}\left(\sum_{i=1}^{n} X_i\right)\\ &= \frac{1}{n}\cdot \sum_{i=1}^{n}\mathbf{E}X_i = \frac{1}{n}\cdot n\cdot\mu = \mu\end{aligned}$$

Mittels der TSCHEBYCHEW–Ungleichung erhalten wir:

$$\begin{aligned} P(|\overline{X}_n - \mu| > \varepsilon) &= P(|\overline{X} - \mathbf{E}\overline{X}| > \varepsilon) \\ &\leq \frac{\operatorname{Var}\overline{X}}{\varepsilon^2} \\ &= \frac{\sigma^2}{n \cdot \varepsilon^2} \xrightarrow[n\to\infty]{} 0 \end{aligned}$$

□

Bem. 17 *Aus dem Beweis erkennen wir, daß die Voraussetzungen etwas abgeschwächt werden können, anstelle* $\operatorname{Var} X_i = \sigma^2$ *genügt die Forderung*

$$\lim_{n\to\infty} \frac{1}{n^2} \sum \operatorname{Var} X_i = 0.$$

Def. 3.1 *Wenn*

$$\lim_{n\to\infty} P(|Y_n - Y_0| > \varepsilon) = 0 \quad \forall \varepsilon > 0$$

dann heißt Y_0 *stochastischer Grenzwert der Folge* $\{Y_n\}$ *und man schreibt* $p - \lim Y_n = Y_0$.

Bsp. 3.4 *Es seien* $X_i \sim \mathrm{B}(1, p)$

$$X_i : \quad \begin{pmatrix} 0 & 1 \\ \\ 1-p & p \end{pmatrix}$$

Dann gilt:

$$\mu := \mathbf{E}X = \mathbf{E}X_i = 0 \cdot (1-p) + 1 \cdot p = p$$

und

$$\begin{aligned} \sigma^2 &:= \mathbf{E}(X-p)^2 \\ &= (0-p)^2 \cdot (1-p) + (1-p)^2 \cdot p \\ &= p^2 \cdot (1-p) + (1-p)^2 \cdot p \\ &= p^2 - p^3 + p - 2 \cdot p^2 + p^3 \\ &= p - p^2 \\ &= p \cdot (1-p) \end{aligned}$$

Nach dem schwachen Gesetz der großen Zahlen folgt:

$$P\left(\left| \frac{1}{n} \sum_{i=1}^{n} X_i - p \right| > \varepsilon \right) \xrightarrow[n \to \infty]{} 0.$$

Bsp. 3.5 *Es sei A ein Ereignis, $p = P(A)$ sei unbekannt. Zur Schätzung von p führen wir eine Reihe von unabhängigen Experimenten durch, bei denen A und $\overline{A}$ die einzig möglichen Ausgänge seien.*

n*:* *# der Experimente, die durchgeführt werden.*

$n(A)$*:* *# Auftretens des Ereignisses A.*

$$\hat{p}_n = \frac{n(A)}{n}$$

die relative Häufigkeit des Ereignisses A.

Frage: $\hat{p}_n \xrightarrow[n\to\infty]{} p$*?*

Dazu definieren wir Zufallsgrößen X_i ($i = 1, \ldots, n$),

$$X_i := \begin{cases} 1 & \text{, } A \text{ im } i\text{–ten Experiment eintritt} \\ 0 & \text{, } A \text{ im } i\text{-ten Experiment nicht eintritt} \end{cases}$$

Dann gilt für alle $i = 1, \ldots, n$:

$$X_i \sim \mathrm{B}(1, p)$$

und $P(X_i = 1) = p$ sowie $P(X_i = 0) = 1 - p$.

$$\sigma^2 = \operatorname{Var} X_i = p \cdot (1 - p)$$
$$\mu = \mathbf{E} X_i = p$$

Wir definieren weiterhin:

$$\overline{X} := \tfrac{1}{n} \cdot \sum_{i=1}^{n} X_i = \tfrac{1}{n} \cdot n(A) = \hat{p}_n.$$

Wir wenden nun das schwache Gesetz der großen Zahlen an und erhalten:

$$\begin{aligned} \lim_{n\to\infty} P(|\hat{p}_n - p| > \varepsilon) &= \lim_{n\to\infty} P(|\overline{X}_n - \mu| > \varepsilon) \\ &= 0, \qquad \forall \varepsilon > 0 \end{aligned}$$

Folglich gilt: $\hat{p}_n \xrightarrow[n\to\infty]{} p$.

Bem. 18 *Schätzungen* $\hat{p}_n$*, die gegen den zu schätzenden Parameter konvergieren heißen (schwach) konsistent.*

Satz 3.6 (Gesetz der Großen Zahlen) *Seien die Zufallsvariablen* $X_1, \ldots, X_n$ *identisch verteilt und unabhängig,* $\mathbf{E}|X_i| < \infty$, $\mathbf{E}X_i = \mu$. *Dann gilt*

$$P(\omega : \lim_{n\to\infty} \overline{X}_n = \mu) = 1.$$

Bem. 19 *Das Schwache Gesetz der Großen Zahlen lautet entsprechend:*

$$\{X_i\} : cov(X_i, X_j) = \sigma^2 \delta_{ij}, \quad \mathbf{E}X_i = \mu \Rightarrow p\text{–}\lim \overline{X}_n = \mu.$$

3 Der Satz von GLIVENKO–CANTELLI

Seien $X_1, \ldots, X_n$ unabhängige Beobachtungen mit

$$P(X_i < x) = F(x), \quad \forall i = 1, \ldots, n.$$

Die Verteilungsfunktion $F(x)$ soll durch die Funktion

$$F_n(x) := \frac{\#\{X_i \colon X_i < x, i = 1, \ldots, n\}}{n}$$

approximiert werden.

Def. 3.2 *Seien* $X_1, \ldots, X_n$ *unkorreliert,* $X_i \sim F$*, und* $X_{(1)}, \ldots, X_{(n)}$, $X_{(1)} \leq X_{(2)} \leq \ldots \leq X_{(n)}$ *die geordnete Stichprobe. Die Funktion*

$$\begin{aligned} F_n(x) &= \frac{\#\{X_i \colon X_i < x, i = 1, \ldots, n\}}{n} \\ &= \begin{cases} 0 & \textit{falls} \quad x < X_{(1)} \\ \frac{i}{n} & \textit{falls} \quad X_{(i)} \leq x < X_{(i+1)} \\ 1 & \textit{falls} \quad X_{(n)} < x \end{cases} \end{aligned}$$

heißt empirische Verteilungsfunktion.

`EDF.sas EDF_2.sas`

Satz 3.7 *Seien $X_1, \ldots, X_n$ unkorreliert. Es gilt:*

$$\lim_{n\to\infty} P(|F_n(x) - F(x)| > \varepsilon) = 0 \quad \forall x \in \mathbb{R}.$$

Beweis: Wir definieren Zufallsgrößen Y_{ix} ($i = 1, \ldots, n$, $x \in \mathbb{R}$) durch:

$$Y_{ix} = \begin{cases} 1 & \text{, falls } X_i < x \\ 0 & \text{, sonst} \end{cases}$$

Dann gilt offensichtlich für alle $i = 1, \ldots, n$ und $x \in \mathbb{R}$:

$$Y_{ix}: \quad \begin{pmatrix} 0 & 1 \\ 1 - F(x) & F(x) \end{pmatrix}$$

D.h. $Y_{ix} \sim \mathrm{B}(1, F(x))$. Wir definieren nun für alle $x \in \mathbb{R}$ eine weitere zufällige Variable $\overline{Y}_x$:

$$\overline{Y}_x := \tfrac{1}{n} \sum_{i=1}^{n} Y_{ix}.$$

Vergleichen wir die Zufallsgrößen Y_{ix} und $\overline{Y}_x$, so stellen wir fest, daß gilt:

$$\overline{Y}_x = F_n(x).$$

Aus Beispiel 3.4 folgt, $\mu := \mathbf{E}Y_{ix} = F(x)$. Deshalb folgt aus dem schwachen Gesetz der großen Zahlen:

$$\lim_{n\to\infty} P(|\overline{Y}_x - \mu| > \varepsilon) = 0, \quad \forall \varepsilon > 0.$$

D.h. für alle $\varepsilon > 0$ gilt:

$$\lim_{n\to\infty} P(|F_n(x) - F(x)| > \varepsilon) = 0$$

$\square$

Verschärfung:

Satz 3.8 (Satz von GLIVENKO–CANTELLI) *Es seien $X_1, \ldots, X_n$ unabhängige zufällige Variablen. Dann gilt:*

$$P\left(\lim_{n\to\infty} \sup_{x\in\mathbb{R}} |F_n(x) - F(x)| = 0\right) = 1.$$

Dieser Satz wird auch oft als der Hauptsatz der Statistik bezeichnet.