23 Der Zentrale Grenzwertsatz von Lindeberg-Feller

Der Zentrale Grenzwertsatz im i.i.d. Fall geht auf Lindeberg und Lévy zurück und besagt, dass Summenvariablen als näherungsweise normalverteilt angesehen werden können. Die „zentrale“ Bedeutung dieses Satzes liegt darin, dass dann unter entsprechenden Voraussetzungen die Summanden X_k, $k = 1, 2, \ldots$, direkt als normalverteilt angenommen werden und damit statistische Untersuchungen sich erheblich vereinfachen (vgl. etwa die Normalverteilungsannahmen in „linearen Modellen“), denn

$$Z_n \; := \; \frac{\sum\limits_{k=1}^{n} (X_k - a)}{\sqrt{n\sigma^2}} \; \overset{\mathcal{D}}{\approx} \; Z \; = \; \frac{\sum\limits_{k=1}^{n} (Y_k - a)}{\sqrt{n\sigma^2}} \, ,$$

wobei $\{X_k\}$ i.i.d., $EX_1 = a$, $0 < Var(X_1) = \sigma^2 < \infty$, $P_Z = N(0,1)$, $\{Y_k\}$ i.i.d., $P_{Y_1} = N(0,1)$.

In Anwendungen lässt sich zwar häufig die Annahme der Unabhängigkeit, aber nicht die einer identischen Verteilung rechtfertigen, so dass allgemeinere Versionen des Zentralen Grenzwertsatzes wünschenswert sind. Wir betrachten im Folgenden ein (so genanntes) Dreieckschema von zeilenweise unabhängigen, quadratintegrierbaren ZV. auf einem W-Raum $(\Omega, \mathcal{A}, P)$, etwa

$$X_{n1}, X_{n2}, \ldots, X_{nr_n} \qquad (r_n \in \mathbb{N}) \qquad \text{mit}$$

$$EX_{nk} \; =: \; a_{nk} \, , \quad 0 \; \leq \; Var(X_{nk}) \; =: \; \sigma_{nk}^2 \; < \; \infty \, .$$

Wir setzen

$$S_n \; := \; \sum_{k=1}^{r_n} (X_{nk} - a_{nk}) \, ,$$

$$s_n^2 \; := \; Var(S_n) \; = \; \sum_{k=1}^{r_n} \sigma_{nk}^2 \qquad (\overset{!}{>} 0) \, .$$

Hinreichend für die Gültigkeit eines allgemeinen Zentralen Grenzwertsatzes ist die (so genannte) „Lindeberg-Bedingung“

$$(Li) \qquad \lim_{n\to\infty} L_n(\varepsilon) \; := \; \lim_{n\to\infty} \frac{1}{s_n^2} \sum_{k=1}^{r_n} \int\limits_{\{\,|\,X_{nk} - a_{nk}\,| \geq \varepsilon s_n\}} (X_{nk} - a_{nk})^2 \, dP \; = \; 0 \qquad \forall\, \varepsilon > 0 \, .$$

Satz 23.1. *Unter der Voraussetzung* (Li) *gilt für das obige Dreieckschema:*

$$(ZGWS) \qquad \frac{1}{s_n} S_n \xrightarrow{\mathcal{D}} Z \quad (n \to \infty)\,, \quad \textit{wobei} \quad P_Z = N(0,1)\,.$$

Bemerkung 23.1.

a) Satz 23.1 verallgemeinert den Satz von Lindeberg-Lévy (Beispiel 22.3 c). Wählt man nämlich $X_{nk} = X_k$, $k = 1, \ldots, r_n := n$, $\{X_k\}$ i.i.d. Folge, so ist die Lindeberg-Bedingung (Li) erfüllt, denn $s_n^2 = n\sigma^2$ und

$$L_n(\varepsilon) = \frac{1}{n\sigma^2}\, n \int\limits_{\{|X_1 - a| \geq \varepsilon\sqrt{n\sigma^2}\}} (X_1 - a)^2\, dP \longrightarrow 0 \quad (n \to \infty) \qquad \forall\, \varepsilon > 0\,,$$

da $E(X_1 - a)^2 = \sigma^2 < \infty$.

b) Bei Anwendungen kann oft die folgende, stärkere Lyapunov-Bedingung verifiziert werden:

$$(Ly) \qquad \lim_{n\to\infty} \frac{1}{s_n^{2+\delta}} \sum_{k=1}^{r_n} E\,|\,X_{nk} - a_{nk}\,|^{\,2+\delta} = 0 \quad (\exists\, \delta > 0)\,.$$

Die Bedingung (Ly) impliziert (Li), denn für $\varepsilon > 0$ gilt:

$$\begin{aligned}
& \frac{1}{s_n^2} \sum_{k=1}^{r_n} \int\limits_{\{|X_{nk} - a_{nk}| \geq \varepsilon s_n\}} |\,X_{nk} - a_{nk}\,|^2\, dP \\
\leq\ & \frac{1}{s_n^2} \sum_{k=1}^{r_n} \int\limits_{\{|X_{nk} - a_{nk}| \geq \varepsilon s_n\}} |\,X_{nk} - a_{nk}\,|^{2+\delta} \frac{1}{(\varepsilon s_n)^\delta}\, dP \\
\leq\ & \frac{1}{\varepsilon^\delta\, s_n^{2+\delta}} \sum_{k=1}^{r_n} E\,|\,X_{nk} - a_{nk}\,|^{\,2+\delta}\,.
\end{aligned}$$

c) Unter zusätzlichen Voraussetzungen lässt sich auch die Konvergenzgeschwindigkeit im Zentralen Grenzwertsatz abschätzen:

Sei z.B. $\{X_k\}$ eine i.i.d. Folge reeller ZV. mit $EX_1 = a$, $Var(X_1) = \sigma^2 > 0$ und $E\,|\,X_1\,|^3 < \infty$. Bezeichnet F_n die VF. der standardisierten Summe, d.h.

$Z_n = \left(\sum_{k=1}^n X_k - na\right) / \sqrt{n\sigma^2}$, und Φ die VF. der $N(0,1)$-Verteilung, so gilt die folgende (gleichmäßige) Berry-Esseen-Ungleichung:

$$\sup_{z\in\mathbb{R}^1} |\, F_n(z) - \Phi(z)\,| \;\leq\; \frac{c}{\sqrt{n}}\, E\left|\, \frac{X_1 - a}{\sigma}\,\right|^3 ,$$

wobei $\dfrac{1}{\sqrt{2\pi}} \leq c < 0,8$.

Bemerkung 23.2. Der Beweis von Satz 23.1 zeigt, dass aus (Li) die (so genannte) „Feller-Bedingung" folgt:

$$(F) \qquad \max_{k=1,\dots,r_n} \left(\frac{\sigma_{nk}^2}{s_n^2}\right) \longrightarrow 0 \qquad (n\to\infty)\,.$$

Die Feller-Bedingung wiederum impliziert die „asymptotische Vernachlässigbarkeit" der Summanden X_{nk}, d.h.

$$(AV) \qquad \max_{k=1,\dots,r_n} P\left(\frac{|\, X_{nk} - a_{nk}\,|}{s_n} \geq \varepsilon\right) \longrightarrow 0 \qquad (n\to\infty) \qquad \forall\, \varepsilon > 0$$

[über die Tschebychev-Ungleichung].

Der folgende Satz stellt die Beziehungen her zwischen der Gültigkeit des Zentralen Grenzwertsatzes (Satz 23.1) und den obigen Bedingungen:

Satz 23.2. *Für das obige Dreieckschema gelten folgende Äquivalenzen:*

$$(Li) \iff (ZGWS)\wedge(F) \iff (ZGWS)\wedge(AV)\,.$$

Beispiel 23.1. Sei $\{X_k\}_{k=1,2,\dots}$ eine unabhängige Folge (P-) f.s. gleichmäßig beschränkter ZV., etwa $|\, X_k\,| \leq c$ P-f.s. $\forall\, k$. Falls $s_n^2 = Var(S_n) = Var(X_1 + \cdots + X_n) \to \infty$ $(n\to\infty)$, so folgt:

$$\frac{S_n}{s_n} = \frac{\sum_{k=1}^n (X_k - EX_k)}{s_n} \xrightarrow{\mathcal{D}} Z \quad (n\to\infty)\,, \quad \text{wobei} \quad P_Z = N(0,1)\,.$$

Das Gesetz vom iterierten Logarithmus lieferte für eine i.i.d. Folge $\{X_k\}_{k=1,2,...}$ mit $EX_1 = a$, $0 < Var(X_1) = \sigma^2 < \infty$, eine Konvergenzgeschwindigkeitsaussage zum starken Gesetz der großen Zahlen, nämlich

$$|\,\overline{X}_n - a\,| \;=\; \mathcal{O}\Big(\sqrt{\frac{\log\log n}{n}}\Big) \qquad P\text{-f.s.} \qquad (n \to \infty)\,.$$

Der Zentrale Grenzwertsatz liefert entsprechend eine Konvergenzgeschwindigkeitsaussage zum schwachen Gesetz der großen Zahlen, nämlich

$$|\,\overline{X}_n - a\,| \;=\; \mathcal{O}_P\Big(\frac{1}{\sqrt{n}}\Big) \qquad (n \to \infty)\,.$$

Hierbei definiert man wie folgt:

Definition 23.1. (*Symbole* o_P, $\mathcal{O}_P$) *Für zwei Folgen* $\{X_n\}$, $\{Y_n\}$ *reeller ZV. auf* $(\Omega, \mathcal{A}, P)$ *setzt man:*

a) $X_n \;=\; o_P(Y_n) \qquad (n \to \infty) \quad :\Longleftrightarrow \quad \dfrac{X_n}{Y_n} \xrightarrow{P} 0 \quad (n \to \infty)\,;$

b) $X_n \;=\; \mathcal{O}_P(Y_n) \qquad (n \to \infty) \quad :\Longleftrightarrow \quad \forall\, \varepsilon > 0 \;\; \exists\, c = c(\varepsilon),\; n_0 = n_0(\varepsilon) \in \mathbb{N}:$

$$P\Big(\Big|\frac{X_n}{Y_n}\Big| \geq c\Big) \;\leq\; \varepsilon \qquad \forall\, n \geq n_0\,.$$