KAPITEL 13

Der zentrale Grenzwertsatz

Der zentrale Grenzwertsatz besagt, dass eine Summe von sehr vielen unabhängigen identisch verteilten Zufallsvariablen mit endlicher Varianz approximativ normalverteilt ist. Dieser Satz begründet theoretisch die herausragende Rolle, die die Normalverteilung in der Wahrscheinlichkeitstheorie und Statistik spielt. Bevor wir den zentralen Grenzwertsatz beweisen, müssen wir eine weitere Konvergenzart, die *Konvergenz in Verteilung*, definieren.

13.1. Konvergenz in Verteilung

Für eine Zufallsvariable X bezeichnen wir mit

$$S(X) = \{t \in \mathbb{R} : F_X \text{ ist stetig an der Stelle } t\}$$

die Menge der Stetigkeitspunkte der Verteilungsfunktion von X. Die Menge $\mathbb{R}\backslash S$ ist die Menge der Atome von X. Diese Menge ist höchstens abzählbar.

Definition 13.1.1. Eine Folge von Zufallsvariablen $X_1, X_2, \ldots$ konvergiert *in Verteilung* gegen eine Zufallsvariable X, wenn für alle $t \in S(X)$ gilt:

$$\lim_{n\to\infty} F_{X_n}(t) = F_X(t).$$

Bezeichnung. $X_n \xrightarrow[n\to\infty]{d} X$ oder $F_{X_n} \xrightarrow[n\to\infty]{d} F_X$. Dabei steht d für "distribution" (Verteilung).

Bemerkung 13.1.2. In der obigen Definition sind nur die Verteilungsfunktionen von X_n und X relevant. Der zugrundeliegende Wahrscheinlichkeitsraum spielt keine Rolle. Es kann z.B. sogar sein, dass die Zufallsvariablen auf verschiedenen Wahrscheinlichkeitsräumen definiert sind.

Beispiel 13.1.3. In diesem Beispiel zeigen wir, warum in der Definition der Verteilungskonvergenz die Einschränkung auf die Stetigkeitspunkte von F_X sinnvoll ist. Seien $c_1 > c_2 > \ldots > 0$ Konstanten mit $\lim_{n\to\infty} c_n = 0$. Als Beispiel kann man $c_n = \frac{1}{n}$ betrachten. Definiere nun Zufallsvariablen $X_n := c_n$ und $X := 0$. Es gilt:

$$\lim_{n\to\infty} F_{X_n}(t) = \begin{cases} 0, & t < 0 \\ 1, & t > 0 \\ 0, & t = 0 \end{cases} = F_X(t), \text{ für alle } t \neq 0.$$

Für $t = 0$ stimmt die Gleichung $\lim_{n\to\infty} F_{X_n}(0) = F_X(0)$ nicht, denn $F_X(0) = 1$. Da allerdings $0 \notin S(X)$, ist die Bedingung in der Definition der Verteilungskonvergenz erfüllt und wir haben $X_n \xrightarrow[n\to\infty]{d} X$.

Hätten wir aber in der Definition der Verteilungskonvergenz verlangt, dass $\lim_{n\to\infty} F_{X_n}(t) = F_X(t)$ für alle $t \in \mathbb{R}$ gelten soll (und nicht nur für alle $t \in S(X)$), so würde der resultierende Konvergenzbegriff die absurde Eigenschaft haben, dass $\frac{1}{n}$ nicht gegen 0 konvergiert.

Der nächste Satz zeigt, dass die Verteilungskonvergenz die schwächste der von uns eingeführten Konvergenzarten ist.

Satz 13.1.4. *Seien X_n und X Zufallsvariablen mit $X_n \xrightarrow[n\to\infty]{P} X$. Dann gilt: $X_n \xrightarrow[n\to\infty]{d} X$.*

Bemerkung 13.1.5. Man kann nun die Beziehungen zwischen verschiedenen Konvergenzarten folgendermaßen darstellen:

$$\begin{array}{ccccccc} & & & & f.s. & \Rightarrow & P & \Rightarrow & d \\ & & & & & & \Uparrow & & \\ \ldots & \Rightarrow & L^3 & \Rightarrow & L^2 & \Rightarrow & L^1 & & \end{array}$$

Beweis von Satz 13.1.4. Sei $X_n \xrightarrow[n\to\infty]{P} X$. Für alle $\varepsilon > 0$ gilt also:

$$\lim_{n\to\infty} \mathbb{P}[|X_n - X| > \varepsilon] = 0.$$

Schritt 1. Für alle $t \in \mathbb{R}$ gilt:

$$\begin{aligned} F_{X_n}(t) &= \mathbb{P}[X_n \le t] \\ &= \mathbb{P}[X_n \le t, |X_n - X| \le \varepsilon] + \mathbb{P}[X_n \le t, |X_n - X| > \varepsilon] \\ &\le \mathbb{P}[X \le t + \varepsilon] + \mathbb{P}[|X_n - X| > \varepsilon] \\ &= F_X(t+\varepsilon) + \mathbb{P}[|X_n - X| > \varepsilon]. \end{aligned}$$

Mit der Voraussetzung $X_n \xrightarrow[n\to\infty]{P} X$ gilt nun für $n \to \infty$:

$$\limsup_{n\to\infty} F_{X_n}(t) \le F_X(t+\varepsilon).$$

Das gilt für jedes $\varepsilon > 0$. Wir können $\varepsilon \downarrow 0$ gehen lassen und die Rechtsstetigkeit von F_X benutzen:

$$\limsup_{n\to\infty} F_{X_n}(t) \le F_X(t).$$

Schritt 2. Sei $t \in S(X)$ beliebig. Es gilt:

$$\begin{aligned} F_X(t-\varepsilon) &= \mathbb{P}[X \le t - \varepsilon] \\ &= \mathbb{P}[X \le t - \varepsilon, |X_n - X| \le \varepsilon] + \mathbb{P}[X \le t - \varepsilon, |X_n - X| > \varepsilon] \\ &\le F_{X_n}(t) + \mathbb{P}[|X_n - X| > \varepsilon]. \end{aligned}$$

Mit der Voraussetzung $X_n \xrightarrow[n\to\infty]{P} X$ gilt nun für $n \to \infty$:

$$F_X(t-\varepsilon) \le \liminf_{n\to\infty} F_{X_n}(t).$$

Das gilt für jedes $\varepsilon > 0$. Wir können also $\varepsilon \downarrow 0$ gehen lassen und benutzen, dass F_X an der Stelle t stetig ist (denn $t \in S(X)$):

$$F_X(t) \le \liminf_{n\to\infty} F_{X_n}(t).$$

Schritt 3. Fügt man Schritt 1 und Schritt 2 zusammen, so erhält man, dass für alle $t \in S(X)$

$$\limsup_{n\to\infty} F_{X_n}(t) \le F_X(t) \le \liminf_{n\to\infty} F_{X_n}(t).$$

Daraus folgt, dass $F_X(t) = \lim_{n\to\infty} F_{X_n}(t)$ für alle $t \in S(X)$. Somit gilt $X_n \xrightarrow[n\to\infty]{d} X$. $\square$

Beispiel 13.1.6. Wir betrachten eine Folge von normalverteilten Zufallsvariablen $X_n \sim N(0, \sigma_n^2)$ mit $\lim_{n\to\infty} \sigma_n^2 = 0$. Wir zeigen, dass $X_n \xrightarrow[n\to\infty]{d} 0$.

Beweis. Es gilt $X_n \overset{L^2}{\to} 0$, denn

$$\mathbb{E}[(X_n - 0)^2] = \mathbb{E}[X_n^2] = \operatorname{Var} X_n = \sigma_n^2 \underset{n\to\infty}{\longrightarrow} 0.$$

Nun folgt aus der L^2-Konvergenz die stochastische Konvergenz und damit auch die Konvergenz in Verteilung nach dem Schema $L^2 \Rightarrow P \Rightarrow d$. $\square$

Beispiel 13.1.7. Dieses Beispiel soll zeigen, dass die Umkehrung von Satz 13.1.4 im Allgemeinen falsch ist. Wir betrachten eine Zufallsvariable $X \sim N(0,1)$ und die Folge $X_n = -X$ für alle $n \in \mathbb{N}$.

Schritt 1. Zuerst zeigen wir, dass X_n in Verteilung gegen X konvergiert. Es gilt:

$$F_{X_n}(t) = \mathbb{P}[X_n \le t] = \mathbb{P}[-X \le t] = \mathbb{P}[X \le t] = F_X(t).$$

Dabei haben wir benutzt, dass $N(0,1)$ eine symmetrische Verteilung ist, d.h. die Verteilung von X stimmt mit der Verteilung von $-X$ überein. Da $F_{X_n}(t) = F_X(t)$ für alle $t \in \mathbb{R}$, folgt, dass X_n gegen X in Verteilung konvergiert.

Schritt 2. Nun zeigen wir, das X_n jedoch nicht in Wahrscheinlichkeit gegen X konvergiert. Es gilt:

$$\mathbb{P}[|X_n - X| > 1] = \mathbb{P}[2|X| > 1] = \mathbb{P}\left[|X| > \frac{1}{2}\right] = \frac{1}{\sqrt{2\pi}} \int\limits_{|t|>1/2} e^{-\frac{t^2}{2}} dt =: c > 0.$$

Da das Integral unabhängig von n ist, folgt $\lim_{n\to\infty} \mathbb{P}[|X_n - X| > 1] = c > 0$. Somit gilt nicht, dass $X_n \xrightarrow[n\to\infty]{P} X$.

Es gibt allerdings einen Spezialfall, in dem die Umkehrung von Satz 13.1.4 richtig ist. Man muss nämlich voraussetzen, dass die Grenzwertzufallsvariable konstant ist.

Satz 13.1.8. *Seien $X_1, X_2, \ldots$ Zufallsvariablen und $c \in \mathbb{R}$ eine Konstante mit $X_n \xrightarrow[n\to\infty]{d} c$. Dann gilt: $X_n \xrightarrow[n\to\infty]{P} c$.*

Beweis. Sei $X_n \xrightarrow[n\to\infty]{d} c$. Dann gilt für alle $t \ne c$:

$$\lim_{n\to\infty} F_{X_n}(t) = F_X(t) = \begin{cases} 1, & t > c, \\ 0, & t < c. \end{cases}$$

Sei nun $\varepsilon > 0$. Dann gilt:

$$\begin{aligned}\mathbb{P}[|X_n - c| > \varepsilon] &= \mathbb{P}[X_n < c - \varepsilon] + \mathbb{P}[X_n > c + \varepsilon] \\ &\leq \mathbb{P}[X_n \leq c - \varepsilon] + 1 - \mathbb{P}[X_n \leq c + \varepsilon] \\ &= F_{X_n}(c - \varepsilon) + 1 - F_{X_n}(c + \varepsilon) \underset{n\to\infty}{\longrightarrow} 0,\end{aligned}$$

da $\lim_{n\to\infty} F_{X_n}(c-\varepsilon) = 0$ und $\lim_{n\to\infty} F_{X_n}(c+\varepsilon) = 1$. Es gilt also $\lim_{n\to\infty} \mathbb{P}[|X_n - c| > \varepsilon] = 0$ und somit $X_n \xrightarrow[n\to\infty]{P} X$. $\square$

Beispiel 13.1.9. Für jedes $n \in \mathbb{N}$ sei X_n eine Zufallsvariable, die gleichverteilt auf der Menge $\{\frac{1}{n}, \frac{2}{n}, \ldots, \frac{n}{n}\}$ ist. Sei außerdem X gleichverteilt auf $[0,1]$. Somit ist X_n eine diskrete Zufallsvariable, wohingegen X absolut stetig ist. Wir zeigen, dass

$$X_n \xrightarrow[n\to\infty]{d} X.$$

Beweis. Für $t < 0$ hat man $F_{X_n}(t) = F_X(t) = 0$. Für $t > 1$ hat man $F_{X_n}(t) = F_X(t) = 1$. Für $t \in [0,1]$ hat man

$$F_{X_n}(t) = \frac{[tn]}{n} \underset{n\to\infty}{\longrightarrow} t = F_X(t),$$

wobei $[\cdot]$ die Gauß-Klammer ist. Also gilt für jedes $t \in \mathbb{R}$:

$$\lim_{n\to\infty} F_{X_n}(t) = F_X(t).$$

Somit folgt $X_n \xrightarrow[n\to\infty]{d} X$. $\square$

Beispiel 13.1.10 (Poisson-Grenzwertsatz). Wir betrachten eine Folge von Zufallsvariablen mit $X_n \sim \text{Bin}(n, p_n)$, wobei für die Erfolgswahrscheinlichkeit p_n gilt: $\lim_{n\to\infty} np_n = \lambda \in (0,\infty)$. Zum Beispiel kann man $p_n = \frac{\lambda}{n}$ betrachten. Sei außerdem $X \sim \text{Poi}(\lambda)$. Wir zeigen, dass

$$X_n \xrightarrow[n\to\infty]{d} X.$$

Beweis. Der Poisson-Grenzwertsatz besagt, dass für alle $k \in \mathbb{N}_0$ gilt:

$$\lim_{n\to\infty} \mathbb{P}[X_n = k] = e^{-\lambda}\frac{\lambda^k}{k!} = \mathbb{P}[X = k].$$

Die Menge der Stetigkeitspunkte der Verteilungsfunktion von X ist $S(X) = \mathbb{R}\backslash\mathbb{N}_0$. Für $t < 0$ gilt $F_{X_n}(t) = 0 = F_X(t)$. Sei deshalb $t > 0$ mit $t \notin \mathbb{N}_0$. Dann kann man ein $m \in \mathbb{N}_0$ mit $m < t < m+1$ finden. Es gilt

$$F_{X_n}(t) = \mathbb{P}[X_n \leq t] = \sum_{k=0}^{m} \mathbb{P}[X_n = k] \underset{n\to\infty}{\longrightarrow} \sum_{k=0}^{m} e^{-\lambda}\frac{\lambda^k}{k!} = \mathbb{P}[X \leq t] = F_X(t).$$

Daraus folgt: $X_n \xrightarrow[n\to\infty]{d} X$. $\square$

13.2. Eine Charakterisierung der Konvergenz in Verteilung

DEFINITION 13.2.1. Die Menge der stetigen und beschränkten Funktionen bezeichnen wir mit

$$C_b := \left\{ f : \mathbb{R} \to \mathbb{R} \mid f \text{ ist stetig und } \sup_{t \in \mathbb{R}} |f(t)| \leq \infty \right\}.$$

SATZ 13.2.2. *Seien $X, X_1, X_2, \ldots$ Zufallsvariablen. Folgende zwei Aussagen sind äquivalent:*

(1) $X_n \xrightarrow[n\to\infty]{d} X$.
(2) *Für alle $f \in C_b$ gilt:* $\lim_{n\to\infty} \mathbb{E}f(X_n) = \mathbb{E}f(X)$.

BEWEIS VON (1) $\Rightarrow$ (2). Sei $X_n \xrightarrow[n\to\infty]{d} X$. Sei $f \in C_b$ eine stetige und beschränkte Funktion. Wir werden zeigen, dass

$$\lim_{n\to\infty} \mathbb{E}f(X_n) = \mathbb{E}f(X).$$

SCHRITT 1. Für f beschränkt ist, gilt $b := \sup_{t\in\mathbb{R}} |f(t)| < \infty$.

SCHRITT 2. Sei $\varepsilon > 0$. Es gilt $\lim_{c\to\infty} \mathbb{P}[|X| > c] = 0$. Somit existiert ein hinreichend großes c mit

$$\mathbb{P}[|X| > c] < \frac{\varepsilon}{b}.$$

Indem man c vergrößert, kann man außerdem erreichen, dass $c \in S(X)$ und $-c \in S(X)$.

SCHRITT 3. Aus $X_n \xrightarrow[n\to\infty]{d} X$ folgt $\lim_{n\to\infty} \mathbb{P}[|X_n| > c] = \mathbb{P}[|X| > c]$, denn

$$\mathbb{P}[|X_n| > c] = F_{X_n}(-c) + (1 - F_{X_n}(c)) \xrightarrow[n\to\infty]{} F_X(-c) + (1 - F_X(c)) = \mathbb{P}[|X| > c].$$

Für hinreichend großes n gilt somit:

$$\mathbb{P}[|X_n| > c] < \frac{2\varepsilon}{b}.$$

SCHRITT 4. Da die Funktion f stetig ist, kann man eine Funktion g mit folgenden Eigenschaften konstruieren:

(1) $\sup_{t\in\mathbb{R}} |f(t) - g(t)| \leq \varepsilon$.
(2) g ist eine Treppenfunktion:

$$g(t) = \sum_{i=1}^{k} a_i \cdot \mathbb{1}_{t_{i-1} < t \leq t_i},$$

wobei $-c = t_0 < t_1 < \ldots < t_k = c$ und $t_0, \ldots, t_k \in S(X)$.

SCHRITT 5. Für hinreichend großes n gilt:

$$\begin{aligned}
&|\mathbb{E}f(X_n) - \mathbb{E}f(X)| \\
&\leq |\mathbb{E}f(X_n)\mathbb{1}_{|X_n| \leq c} - \mathbb{E}f(X)\mathbb{1}_{|X| \leq c}| + |\mathbb{E}f(X_n)\mathbb{1}_{|X_n| > c}| + |\mathbb{E}f(X)\mathbb{1}_{|X| > c}| \\
&\leq |\mathbb{E}f(X_n)\mathbb{1}_{|X_n| \leq c} - \mathbb{E}f(X)\mathbb{1}_{|X| \leq c}| + 3\varepsilon \\
&\leq |\mathbb{E}f(X_n)\mathbb{1}_{|X_n| \leq c} - \mathbb{E}g(X_n)| + |\mathbb{E}g(X) - \mathbb{E}f(X)\mathbb{1}_{|X| \leq c}| + |\mathbb{E}g(X_n) - \mathbb{E}g(X)| + 3\varepsilon \\
&\leq |\mathbb{E}g(X_n) - \mathbb{E}g(X)| + 5\varepsilon,
\end{aligned}$$

wobei die zweite Ungleichung gilt, da $|f(X_n)| \leq b$, $|f(X)| \leq b$, $\mathbb{P}[|X_n| > c] \leq \frac{2\varepsilon}{b}$ und $\mathbb{P}[|X| > c] \leq \frac{\varepsilon}{b}$, und die letzte Ungleichung gilt, da $|f(x) - g(x)| \leq \varepsilon$.

Schritt 6. Wir betrachten nun den Term $\mathbb{E}g(X_n) - \mathbb{E}g(X)$. Aus $X_n \xrightarrow[n\to\infty]{d} X$ und $t_0, \dots, t_k \in S(X)$ folgt, dass

$$\begin{aligned}
\mathbb{E}g(X_n) &= \sum_{i=1}^{k} a_i \cdot \mathbb{E}\mathbb{1}_{t_{i-1} < X_n \leq t_i} \\
&= \sum_{i=1}^{k} a_i \cdot (F_{X_n}(t_i) - F_{X_n}(t_{i-1})) \\
&\xrightarrow[n\to\infty]{} \sum_{i=1}^{k} a_i \cdot (F_X(t_i) - F_X(t_{i-1})) \\
&= \mathbb{E}g(X).
\end{aligned}$$

Daraus folgt: $\lim_{n\to\infty} \mathbb{E}g(x_n) = \mathbb{E}g(X)$. Für hinreichend großes n gilt somit:

$$|\mathbb{E}g(X_n) - \mathbb{E}g(X)| \leq \varepsilon.$$

Schritt 7. Aus Schritt 5 und Schritt 6 ergibt sich, dass für hinreichend großes n gilt:

$$|\mathbb{E}f(X_n) - \mathbb{E}f(X)| \leq 6\varepsilon.$$

Dabei war $\varepsilon > 0$ beliebig. Daraus folgt, dass $\lim_{n\to\infty} \mathbb{E}f(X_n) = \mathbb{E}f(X)$.

Beweis von (2) $\Rightarrow$ (1). Sei $\lim_{n\to\infty} \mathbb{E}f(X_n) = \mathbb{E}f(X)$ für alle $f \in C_b$. Wir zeigen, dass

$$X_n \xrightarrow[n\to\infty]{d} X.$$

Beweisidee: Dürften wir $f(z) = \mathbb{1}_{z \leq t}$ einsetzen, so würde gelten:

$$\lim_{n\to\infty} F_{X_n}(t) = \lim_{n\to\infty} \mathbb{E}f(X_n) = \mathbb{E}f(X) = F_X(t).$$

Allerdings ist die Indikatorfunktion $\mathbb{1}_{z \leq t}$ nicht stetig. Also werden wir dieese Indikatorfunktion durch stetige Funktionen approximieren.

Schritt 1. Sei $t \in \mathbb{R}$ und $\varepsilon > 0$. Dann existiert eine Funktion $f \in C_b$ mit

$$\mathbb{1}_{y \leq t} \leq f(y) \leq \mathbb{1}_{y \leq t+\varepsilon} \text{ für alle } y \in \mathbb{R}.$$

(1) Es gilt also $\mathbb{1}_{X_n \leq t} \leq f(X_n) \leq \mathbb{1}_{X_n \leq t+\varepsilon}$. Daraus folgt:

$$F_{X_n}(t) = \mathbb{E}\mathbb{1}_{X_n \leq t} \leq \mathbb{E}f(X_n).$$

(2) Außerdem gilt $\mathbb{1}_{X \leq t} \leq f(X) \leq \mathbb{1}_{X \leq t+\varepsilon}$. Es folgt:

$$\mathbb{E}f(X) \leq \mathbb{E}\mathbb{1}_{X \leq t+\varepsilon} = F_X(t+\varepsilon).$$

Fügt man nun (1) und (2) zusammen, erhält man für $n \to \infty$:

$$\limsup_{n\to\infty} F_{X_n}(t) \leq \lim_{n\to\infty} \mathbb{E}f(X_n) = \mathbb{E}f(X) \leq F_X(t+\varepsilon).$$

Das gilt für jedes $\varepsilon > 0$. Da nun F_X rechtsstetig ist, erhalten wir für $\varepsilon \downarrow 0$:

$$\limsup_{n\to\infty} F_{X_n}(t) \leq F_X(t).$$

Schritt 2. Sei $t \in S(X)$ und $\varepsilon > 0$. Dann existiert ein $f \in C_b$ mit $\mathbb{1}_{y\leq t-\varepsilon} \leq f(y) \leq \mathbb{1}_{y\leq t}$ für alle $y \in \mathbb{R}$. Es gilt:

$$F_X(t-\varepsilon) = \mathbb{E}\mathbb{1}_{X\leq t-\varepsilon} \leq \mathbb{E}f(X) = \lim_{n\to\infty} \mathbb{E}f(X_n) \leq \liminf_{n\to\infty} \mathbb{E}\mathbb{1}_{X_n\leq t} = \liminf_{n\to\infty} F_{X_n}(t).$$

Für $\varepsilon \downarrow 0$ erhalten wir dann:

$$F_X(t) \leq \liminf_{n\to\infty} F_{X_n}(t).$$

Schritt 3. Fügt man nun Schritte 1 und 2 zusammen, so folgt für alle $t \in S(X)$:

$$F_X(t) = \lim_{n\to\infty} F_{X_n}(t).$$

Und das bedeutet, dass $X_n \xrightarrow[n\to\infty]{d} X$. □

13.3. Satz von Helly

Für eine Verteilungsfunktion G bezeichnen wir mit $S(G)$ die Menge der Stetigkeitspunkte von G.

Satz 13.3.1 (Helly). *Seien $F_1, F_2, \ldots$ Verteilungsfunktionen. Dann gibt es eine rechtsstetige nichtfallende Funktion G und eine Teilfolge $F_{n_1}, F_{n_2}, \ldots$, so dass für alle $t \in S(G)$ gilt:*

$$\lim_{n\to\infty} F_{n_k}(t) = G(t).$$

Beweis. Die Menge der rationalen Zahlen ist abzählbar, also gibt es eine Abzählung $\mathbb{Q} = \{r_1, r_2, \ldots\}$. Wir werden im Folgenden eine geschachtelte Familie von unendlichen Mengen $\mathbb{N} = K_0 \supset K_1 \supset \ldots$ mit gewissen (unten näher beschriebenen) Eigenschaften konstruieren.

Schritt 1. Sei $K_0 = \mathbb{N}$. Es gilt: $\{F_n(r_1)\}_{n\in\mathbb{N}} \subset [0,1]$. Mit dem Satz von Bolzano–Weierstraß folgt: Es gibt eine Teilmenge $K_1 \subset K_0$ mit $|K_1| = \infty$ und

$$\lim_{n\to\infty,\, n\in K_1} F_n(r_1) = g_1.$$

Schritt 2. Genauso gilt: $\{F_n(r_2)\}_{n\in K_1} \subset [0,1]$. Und damit folgt wiederum mit dem Satz von Bolzano–Weierstraß, dass eine Teilmenge $K_2 \subset K_1$ existiert mit $|K_2| = \infty$ und

$$\lim_{n\to\infty,\, n\in K_2} F_n(r_2) = g_2.$$

Schritt 3. Diese Argumentation kann fortgesetzt werden. Es existieren also Mengen $\mathbb{N} = K_0 \subset K_1 \subset K_2 \subset \ldots$ mit $|K_j| = \infty$ und

$$\lim_{n\to\infty,\, n\in K_j} F_n(r_j) = g_j \text{ für alle } j \in \mathbb{N}.$$

Schritt 4. Nun benutzen wir die sogenannte Cantor–Diagonalisierung. In jeder Menge K_j wählen wir das kleinste Element. Die Vereinigung dieser Elemente nennen wir K, so dass

$K = \{\min K_j\}_{j\in\mathbb{N}}$. Es gilt nach Konstruktion $|K| = \infty$. Außerdem gilt für alle $j \in \mathbb{N}$: K ist bis auf endlich viele Elemente eine Teilmenge von K_j und somit

$$\lim_{n\to\infty,\, n\in K} F_n(r_j) = g_j \text{ für alle } j \in \mathbb{N}.$$

Wir haben gezeigt, dass für alle $j \in \mathbb{N}$ der folgende Grenzwert existiert:

$$H(r_j) := \lim_{n\to\infty,\, n\in K} F_n(r_j).$$

Die Funktion H ist nur auf der Menge der rationalen Zahlen definiert. Wir erweitern die Funktion H auf ganz $\mathbb{R}$:

$$G(t) := \inf\{H(r) : r \in \mathbb{Q}, r > t\}.$$

Für diese Funktion gilt: G ist nichtfallend und rechtsstetig (Übung).

Schritt 5. Nun müssen wir noch zeigen, dass für $t \in S(G)$ gilt:

$$\lim_{n\to\infty,\, n\in K} F_n(t) = G(t).$$

Sei also $t \in S(G)$ und sei $\varepsilon > 0$. Dann existiert $y < t$ mit $G(t)-\varepsilon \leq G(y) \leq G(t)$. Außerdem existieren $r, s \in \mathbb{Q}$ mit $y < r < t < s$ und $G(s) \leq G(t) + \varepsilon$. Nun gilt also:

$$G(t) - \varepsilon \leq G(y) \leq G(r) \leq G(s) \leq G(t) + \varepsilon.$$

Es folgt, dass

$$G(r) \xleftarrow[n\to\infty]{n\in K} F_n(r) \leq F_n(t) \leq F_n(s) \xrightarrow[n\to\infty]{n\in K} G(s).$$

Somit gilt

$$G(t) - \varepsilon \leq \liminf_{n\to\infty, n\in K} F_n(t) \leq \limsup_{n\to\infty, n\in K} F_n(t) \leq G(t) + \varepsilon.$$

Für $\varepsilon \downarrow 0$ erhalten wir dann: $G(t) = \lim_{n\to\infty,\, n\in K} F_n(t)$. □

Beispiel 13.3.2. Die Funktion G aus dem Satz von Helly ist monoton nichtfallend und rechtsstetig, sie muss aber nicht unbedingt eine Verteilungsfunktion sein. Es fehlen nämlich die Eigenschaften $\lim_{t\to-\infty} G(t) = 0$ und $\lim_{t\to+\infty} G(t) = 1$. Um zu sehen, dass die beiden Eigenschaften nicht immer erfüllt sind, betrachten wir die Folge $X_n = n$. Für die Verteilungsfunktion von X_n gilt dann

$$\lim_{n\to\infty} F_{X_n}(t) = \lim_{n\to\infty} \mathbb{1}_{[n,\infty)}(t) = 0.$$

Also ist auch der Grenzwert jeder Teilfolge von F_{X_n} gleich 0. Die Funktion $G = 0$ ist aber keine Verteilungsfunktion.

Definition 13.3.3. Eine Folge $F_1, F_2, \ldots$ von Verteilungsfunktionen (bzw. eine Folge von Zufallsvariablen, die diese Verteilungsfunktionen haben) heißt *straff*, wenn für alle $\varepsilon > 0$ ein $c > 0$ existiert, so dass für alle $n \in \mathbb{N}$ gilt:

$$F_n(-c) + (1 - F_n(c)) < \varepsilon.$$

Bemerkung 13.3.4. Ist F_n die Verteilungsfunktion einer Zufallsvariable X_n, so kann man die obige Bedingung auch so darstellen: für alle $\varepsilon > 0$ existiert ein $c > 0$, so dass

$$\mathbb{P}[|X_n| > c] < \varepsilon \text{ für alle } n \in \mathbb{N}.$$

Beispiel 13.3.5. Die Folge der Verteilungsfunktionen $F_n = \mathbb{1}_{[n,\infty)}$ ist nicht straff.

Beispiel 13.3.6. Seien $X_1, X_2, \ldots$ beliebige Zufallsvariablen, die nur Werte in einem Intervall $[-A, A]$ annehmen. Dann ist diese Folge von Zufallsvariablen straff.

Satz 13.3.7 (Helly, Prochorow). *Sei $F_1, F_2, \ldots$ eine straffe Folge von Verteilungsfunktionen. Dann existieren eine Verteilungsfunktion G und eine Teilfolge $F_{n_1}, F_{n_2}, \ldots$, so dass für alle $t \in S(G)$ gilt:*

$$\lim_{k\to\infty} F_{n_k}(t) = G(t).$$

Bemerkung 13.3.8. Mit anderen Worten: aus einer straffen Folge von Zufallsvariablen kann man eine in Verteilung konvergente Teilfolge extrahieren. Vergleiche das mit dem Satz von Bolzano–Weierstraß: Aus einer beschränkten Folge von Vektoren in $\mathbb{R}^d$ kann man eine konvergente Teilfolge extrahieren.

Beweis. Aus Satz 13.3.1 folgt, dass es eine Teilfolge $n_1, n_2, \ldots$ und eine nichtfallende, rechtsstetige Funktion G mit $\lim_{k\to\infty} F_{n_k}(t) = G(t)$ für alle $t \in S(G)$ gibt. Wir berechnen die Grenzwerte $\lim_{t\to\pm\infty} G(t)$.
Für alle $\varepsilon > 0$ existiert ein $c > 0$ mit $F_{n_k}(-c) < \varepsilon$ und $1 - F_{n_k}(c) < \varepsilon$ für alle $k \in \mathbb{N}$, da die Folge $F_1, F_2, \ldots$ straff ist. Indem wir c vergrößern, können wir zusätzlich annehmen, dass $c, -c \in S(G)$. Nun lassen wir $k \to \infty$ und erhalten, dass

$$G(-c) \le \varepsilon \text{ und } 1 - G(c) \le \varepsilon.$$

Da das für alle $\varepsilon > 0$ gilt, ergibt sich daraus, dass $\lim_{c\to+\infty} G(-c) = 0$ und $\lim_{c\to+\infty} G(c) = 1$. Also ist G eine Verteilungsfunktion. □

13.4. Stetigkeitssatz von Lévy

Der nächste Satz wird im Beweis des zentralen Grenzwertsatzes eine wichtige Rolle spielen. Dieser Satz besagt, dass Konvergenz in Verteilung äquivalent zur punktweisen Konvergenz der entsprechenden charakteristischen Funktionen ist.

Satz 13.4.1 (Stetigkeitssatz von Lévy). *Seien $X, X_1, X_2, \ldots$ Zufallsvariablen mit charakteristischen Funktionen $\varphi_X, \varphi_{X_1}, \varphi_{X_2}, \ldots$. Dann sind folgende zwei Aussagen äquivalent:*

(1) $X_n \xrightarrow[n\to\infty]{d} X$.
(2) $\lim_{n\to\infty} \varphi_{X_n}(t) = \varphi_X(t)$ *für alle* $t \in \mathbb{R}$.

Im Beweis werden wir die vereinfachte Notation $\varphi_{X_n} = \varphi_n$ und $\varphi_X = \varphi$ benutzen. Für die Verteilungsfunktionen von X_n und X benutzen wir die Notation $F_{X_n} = F_n$ und $F_X = F$.

Beweis von (1) ⇒ (2). Sei $X_n \xrightarrow[n\to\infty]{d} X$. Dann gilt mit Satz 13.2.2 für alle $t \in \mathbb{R}$:

$$\varphi_n(t) = \mathbb{E}\cos(tX_n) + i\mathbb{E}\sin(tX_n) \underset{n\to\infty}{\longrightarrow} \mathbb{E}\cos(tX) + i\mathbb{E}\sin(tX) = \varphi(t).$$

Beweis von (2) $\Rightarrow$ (1). Es gelte $\lim_{n\to\infty}\varphi_n(t) = \varphi(t)$ für ale $t\in\mathbb{R}$. Wir wollen zeigen, dass $X_n \xrightarrow[n\to\infty]{d} X$.

Schritt 1. Zuerst zeigen wir, dass die Folge von Verteilungsfunktionen $F_1, F_2, \ldots$ straff ist. Wir definieren eine Funktion $B_n(u)$ folgendermaßen:

$$B_n(u) := \frac{1}{u}\int_{-u}^{u}(1-\varphi_n(t))dt = \frac{1}{u}\int_{-u}^{u}\mathbb{E}[1-e^{itX_n}]dt = \mathbb{E}\left[\frac{1}{u}\int_{-u}^{u}(1-e^{itX_n})dt\right],$$

wobei wir im letzten Schritt die Integrale nach dem Satz von Fubini vertauscht haben. Das war möglich, denn $|1-e^{itX_n}|\le 2$. Da nun $\sin x$ eine gerade Funktion ist und der Imaginärteil somit nach der Integration verschwindet, erhalten wir, dass

$$B_n(u) = 2\mathbb{E}\left(1-\frac{\sin(uX_n)}{uX_n}\right) \ge 2\mathbb{E}\left[1-\frac{1}{|uX_n|}\right] \ge \mathbb{P}\left[|X_n|\ge \frac{2}{u}\right].$$

Es gilt: $\varphi(0)=1$ und φ ist stetig, da φ eine charakteristische Funktion ist. Somit existiert für jedes $\varepsilon>0$ ein hinreichend kleines $u>0$, so dass

$$B(u) = \frac{1}{u}\cdot\int_{-u}^{u}(1-\varphi(t))dt < \varepsilon.$$

Da $\lim_{n\to\infty}\varphi_n(t)=\varphi(t)$ für alle $t\in\mathbb{R}$ gilt, folgt mit der majorisierten Konvergenz:

$$B_n(u) = \frac{1}{u}\int_{-u}^{u}(1-\varphi_n(t))dt \underset{n\to\infty}{\longrightarrow} \frac{1}{u}\int_{-u}^{u}(1-\varphi(t))dt = B(u).$$

Wir durften die majorisierte Konvergenz benutzen, da $|1-\varphi_n(t)|\le 2$. Somit existiert ein n_0, so dass $B_n(u)<2\varepsilon$ für alle $n\ge n_0$. Dann ist auch

$$\mathbb{P}\left[|X_n|\ge\frac{2}{u}\right] \le B_n(u) < 2\varepsilon \text{ für alle } n\ge n_0.$$

Indem wir die Schranke $\frac{2}{u}$ vergrößern, können wir erreichen, dass die obige Ungleichung sogar für alle $n\in\mathbb{N}$ gilt. Daraus folgt, dass die Folge $F_1, F_2, \ldots$ straff ist.

Schritt 2. Nun zeigen wir durch Widerspruch, dass $X_n \xrightarrow[n\to\infty]{d} X$. Wir nehmen also an, dass diese Konvergenz nicht gilt. Es existiert somit ein $t\in\mathbb{R}$ mit

$$\lim_{n\to\infty}F_n(t)\ne F(t) \text{ und } \mathbb{P}[X=t]=0. \tag{13.4.1}$$

Daraus folgt: es gibt ein $\varepsilon>0$ und eine Teilfolge $F_{n_1}, F_{n_2}, \ldots$ mit

$$|F_{n_k}(t)-F(t)|>\varepsilon \text{ für alle } k\in\mathbb{N}.$$

Da die Folge $F_1, F_2, \ldots$ und somit auch die Teilfolge $F_{n_1}, F_{n_2}, \ldots$ straff ist, folgt mit Satz 13.3.7, dass eine Verteilungsfunktion G und eine Teilteilfolge $F_{n_{k_1}}, F_{n_{k_2}}, \ldots$ existieren, für die gilt:

$$F_{n_{k_j}} \xrightarrow[j\to\infty]{d} G. \tag{13.4.2}$$

Wir betrachten zwei Fälle:

Fall 1. G ist nicht stetig an der Stelle t. Dann folgt: $G\ne F$, denn F ist stetig an der Stelle t wegen (13.4.1).

Fall 2. G ist stetig an der Stelle t. Dann folgt aus (13.4.2), dass $\lim_{j\to\infty} F_{n_{k_j}}(t) = G(t) \neq F(t)$, also $G \neq F$.
In beiden Fällen gilt $G \neq F$. Aus (13.4.2) und wegen der bereits bewiesenen Richtung (1) $\Rightarrow$ (2) im Satz folgt, dass

$$\lim_{j\to\infty} \varphi_{n_{k_j}}(t) = \varphi_G(t) \text{ für alle } t \in \mathbb{R}.$$

Dabei bezeichnen wir mit φ_G die charakteristische Funktion von G. Nach Voraussetzung sollte aber auch gelten, dass

$$\lim_{j\to\infty} \varphi_{n_{k_j}}(t) = \varphi(t) \text{ für alle } t \in \mathbb{R}.$$

Somit gilt $\varphi_G = \varphi$. Das heißt, die charakteristischen Funktionen von G und F sind gleich. Dabei haben wir aber gezeigt, dass $G \neq F$. Dies ist ein Widerpruch, da die Verteilungsfunktion eindeutig durch die charakteristische Funktion bestimmt wird. $\square$

13.5. Der zentrale Grenzwertsatz

Seien $X_1, X_2, \ldots$ unabhängige, identisch verteilte Zufallsvariablen mit $\mathbb{E}X_k = \mu$, $\operatorname{Var} X_k = \sigma^2$. Wir betrachten die Summe $S_n = X_1 + \ldots + X_n$. Wir werden nun die Frage beantworten, wie die Summe S_n für $n \to \infty$ verteilt ist. Damit wir eine sinnvole Aussage über die Verteilung von S_n erhalten können, müssen wir zuerst einige Vorbereitungen treffen.

Schritt 1. Durch Abziehen des Erwartungswerts können wir S_n *zentrieren*. Das heißt, wir betrachten die Zufallsvariable

$$S_n - \mathbb{E}S_n = S_n - n\mu.$$

Es gilt dann $\mathbb{E}[S_n - n\mu] = 0$.

Schritt 2. Indem wir durch die Standardabweichung dividieren, können wir die Zufallsvariable auch *normieren*. Das heißt, wir betrachten die Zufallsvariable

$$\frac{S_n - \mathbb{E}S_n}{\sqrt{\operatorname{Var} S_n}} = \frac{S_n - n\mu}{\sigma\sqrt{n}}.$$

Dann gilt $\mathbb{E}[\frac{S_n-n\mu}{\sigma\sqrt{n}}] = 0$ und $\operatorname{Var}[\frac{S_n-n\mu}{\sigma\sqrt{n}}] = 1$.

Satz 13.5.1 (Der zentrale Grenzwertsatz). *Seien $X_1, X_2, \ldots$ unabhängige, identisch verteilte, quadratisch integrierbare Zufallsvariablen mit $\mathbb{E}X_k = \mu$, $\operatorname{Var} X_k = \sigma^2 \in (0, \infty)$. Sei $S_n = X_1 + \ldots + X_n$. Dann gilt:*

$$\frac{S_n - n\mu}{\sigma\sqrt{n}} \xrightarrow[n\to\infty]{d} N,$$

wobei $N \sim N(0,1)$ eine standardnormalverteilte Zufallsvariable ist.

Bemerkung 13.5.2. Mit anderen Worten: Für alle $x \in \mathbb{R}$ gilt

$$\lim_{n\to\infty} \mathbb{P}\left[\frac{S_n - n\mu}{\sigma\sqrt{n}} \leq x\right] = \Phi(x) \text{ mit } \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}}\, dt.$$

Als Spezialfall erhalten wir den folgenden Satz von de Moivre–Laplace.

Satz 13.5.3 (de Moivre (1733), Laplace (1812)). *Sei $S_n \sim Bin(n,p)$ binomialverteilt, wobei $p \in (0,1)$ konstant ist. Dann gilt für alle $x \in \mathbb{R}$:*

$$\lim_{n\to\infty} \mathbb{P}\left[\frac{S_n - np}{\sqrt{np(1-p)}} \le x\right] = \Phi(x).$$

Beweis. Wir definieren unabhägige, identisch verteilte Zufallsvariablen $X_1, X_2, \ldots$ mit

$$\mathbb{P}[X_k = 1] = p \text{ und } \mathbb{P}[X_k = 0] = 1 - p.$$

Dann gilt für die Summe der Zufallsvariablen

$$S_n = X_1 + \ldots + X_n \sim \mathrm{Bin}(n,p).$$

Der Erwartungswert von X_k ist $\mu = p$ und die Varianz ist $\sigma^2 = p \cdot (1-p)$. Mit dem zentralen Grenzwertsatz folgt die Behauptung. □

Bemerkung 13.5.4. Mit dem Satz von de Moivre–Laplace kann man die Binomialverteilung $\mathrm{Bin}(n,p)$ für ein großes n und ein konstantes p durch die Normalverteilung approximieren. Sei nämlich $S_n \sim \mathrm{Bin}(n,p)$. Wir wollen die Wahrscheinlichkeit $\mathbb{P}[a \le S_n \le b]$, dass S_n innerhalb eines gegebenen Intervalls $[a,b]$ liegt, berechnen. Mit dem Satz von de Moivre–Laplace erhalten wir die folgende Approximation:

$$\mathbb{P}[a \le S_n \le b] = \mathbb{P}\left[\frac{a - n\mu}{\sigma\sqrt{n}} \le \frac{S_n - n\mu}{\sigma\sqrt{n}} \le \frac{b - n\mu}{\sigma\sqrt{n}}\right] \approx \mathbb{P}[a^* \le N \le b^*] = \Phi(b^*) - \Phi(a^*),$$

wobei $\frac{a-n\mu}{\sigma\sqrt{n}} = a^*$ und $\frac{b-n\mu}{\sigma\sqrt{n}} = b^*$. Wir haben benutzt, dass $\frac{S_n - n\mu}{\sigma\sqrt{n}} \approx N \sim N(0,1)$ für großes n ist.

Beispiel 13.5.5. Wir betrachten ein $n = 100$-faches Bernoulli-Experiment mit Erfolgswahrscheinlichkeit $p = \frac{1}{2}$. Sei $S = S_{100} \sim \mathrm{Bin}(100, \frac{1}{2})$ die Anzahl der Erfolge. Wir approximieren nun mit dem zentralen Grenzwertsatz die Wahrscheinlichkeit $\mathbb{P}[S \le 55]$.

Lösung. Für den Erwartungswert und die Varianz von S erhalten wir

$$\mathbb{E}S = np = 100 \cdot \frac{1}{2} = 50 \text{ und } \operatorname{Var} S = np(1-p) = 100 \cdot \frac{1}{2} \cdot \frac{1}{2} = 25.$$

Also gilt mit dem Satz von de Moivre–Laplace:

$$\mathbb{P}[S \le 55] = \mathbb{P}\left[\frac{S - 50}{\sqrt{25}} \le \frac{55 - 50}{\sqrt{25}}\right] \approx \mathbb{P}[N \le 1] = \Phi(1) = 0.8413,$$

wobei wir benutzt haben, dass $\frac{S-50}{\sqrt{25}} \approx N$.

Um eine bessere Approximation zu erhalten, kann man den sogenannten $\pm\frac{1}{2}$ Trick anwenden. Da die Zufallsvariable S_n nämlich ganzzahlig ist, sind die Wahrscheinlichkeiten $\mathbb{P}[S \le 55]$ und $\mathbb{P}[S < 56]$ gleich. Sollen wir nun 55 oder 56 in die Approximation einsetzen? Wir werden den Mittelwert 55.5 einsetzen:

$$\mathbb{P}[S \le 55] = \mathbb{P}[S \le 55{,}5] = \mathbb{P}\left[\frac{S - 50}{\sqrt{25}} \le \frac{55{,}5 - 50}{\sqrt{25}}\right] \approx \mathbb{P}[N \le 1{,}1] = \Phi(1{,}1) = 0{,}86433.$$

Der exakte Wert für die Wahrscheinlichkeit ist übrigens $\mathbb{P}[S \le 55] = 0.86437$.

Beispiel 13.5.6. Seien $X_1, X_2, \ldots \sim N(\mu, \sigma^2)$ unabhängige normalverteilte Zufallsvariablen. Dann gilt nicht nur approximativ, sondern sogar exakt, dass

$$\frac{S_n - n\mu}{\sigma\sqrt{n}} \sim N(0,1).$$

13.6. Beweis des zentralen Grenzwertsatzes

Beweis von Satz 13.5.1. Es seien $X_1, X_2, \ldots$ unabhängige identisch verteilte Zufallsvariablen mit $\mathbb{E}X_k = \mu$ und $\operatorname{Var} X_k = \sigma^2 > 0$. Definiere $S_n = X_1 + \ldots + X_n$. Wir zeigen, dass

$$\frac{S_n - n\mu}{\sigma\sqrt{n}} \xrightarrow[n\to\infty]{d} N \sim N(0,1).$$

Anstelle von $X_1, X_2, \ldots$ betrachten wir die normierten Zufallsvariablen

$$Y_k := \frac{X_k - \mu}{\sigma} \text{ mit } \mathbb{E}Y_k = 0 \text{ und } \operatorname{Var} Y_k = 1.$$

Wir definieren die Zufallsvariable

$$V_n := \frac{S_n - n\mu}{\sigma\sqrt{n}} = \frac{Y_1 + \ldots + Y_n}{\sqrt{n}}.$$

Zu zeigen ist, dass

$$V_n \xrightarrow[n\to\infty]{d} N \sim N(0,1).$$

Wir werden zeigen, dass die charakteristische Funktion von V_n gegen die charakteristische Funktion der Standardnormalverteilung punktweise konvergiert. Dann kann man den Stetigkeitssatz 13.4.1 anwenden. Sei $t \in \mathbb{R}$ fest. Für die charakteristische Funktion von V_n gilt:

$$\varphi_{V_n}(t) = \mathbb{E}e^{itV_n} = \mathbb{E}\left[e^{it\frac{Y_1+\ldots Y_n}{\sqrt{n}}}\right] = \mathbb{E}\left[e^{it\frac{Y_1}{\sqrt{n}}}\right] \cdot \ldots \cdot \mathbb{E}\left[e^{it\frac{Y_n}{\sqrt{n}}}\right] = \left(\varphi\left(\frac{t}{\sqrt{n}}\right)\right)^n,$$

wobei $\varphi := \varphi_{Y_1}$ die charakteristische Funktion von Y_1 ist. Nun wollen wir $n \to \infty$ gehen lassen. Da $\varphi(0) = 1$ und φ stetig ist, gilt

$$\lim_{n\to\infty} \varphi\left(\frac{t}{\sqrt{n}}\right) = 1.$$

Also haben wir es mit einer Unbestimmtheit von der Form 1^∞ zu tun. Um die aufzulösen, werden wir die ersten zwei Terme der Taylor-Entwicklung von φ betrachten. Es gilt

$$\varphi(0) = 1, \quad \varphi'(0) = i\mathbb{E}Y_1 = 0, \quad \varphi''(0) = -\mathbb{E}[Y_1^2] = -1.$$

Somit erhalten wir die folgende Taylorentwicklung:

$$\varphi(s) = 1 - \frac{s^2}{2} + o(s^2), \quad s \to 0.$$

Nun setzen wir $s = \frac{t}{\sqrt{n}}$ ein:

$$\varphi\left(\frac{t}{\sqrt{n}}\right) = 1 - \frac{t^2}{2n} + o\left(\frac{t^2}{n}\right), \quad n \to \infty.$$

Und dann gilt für die charakteristische Funktion von V_n:

$$\varphi_{V_n}(t) = \left(\varphi\left(\frac{t}{\sqrt{n}}\right)\right)^n = \left(1 - \frac{t^2}{2n} + o\left(\frac{t^2}{2n}\right)\right)^n \underset{n\to\infty}{\longrightarrow} e^{-\frac{t^2}{2}}.$$

Dabei haben wir die folgende Aussage benutzt: für eine Folge $a_1, a_2, \ldots$ mit $\lim_{n\to\infty} na_n = \lambda$ gilt, dass $\lim_{n\to\infty}(1 + a_n)^n = e^\lambda$. Da $e^{-t^2/2}$ die charakteristische Funktion der Standardnormalverteilung ist, folgt schließlich mit dem Stetigkeitssatz 13.4.1, dass V_n in Verteilung gegen eine Standardnormalverteilte Zufallsvariable konvergiert. Das beweist die Behauptung. $\square$

13.7. Sätze von Lindeberg und Ljapunow

Der folgende Satz ist eine Verallgemeinerung des zentralen Grenzwertsatzes. Er besagt, dass eine Summe von sehr vielen sehr kleinen und unabhängigen zufälligen Fehlern unter ziemlich allgemeinen Bedingungen approximativ normalverteilt ist. In diesem Satz werden wir anstatt einer Folge ein sogenanntes *Dreiecksschema* von Zufallsvariablen betrachten:

$$\begin{array}{lllll} X_{11} & & & & \\ X_{21}, & X_{22} & & & \\ X_{31}, & X_{32}, & X_{33} & & \\ \ldots & & & & \\ X_{n1}, & X_{n2}, & X_{n3}, & \ldots, & X_{nn} \\ \ldots & & & & \end{array}$$

Wir werden annehmen, dass jede Zeile dieses Dreiecks aus unabhängigen Zufallsvariablen besteht. Abhängigkeiten zwischen den Variablen aus verschiedenen Zeilen sind jedoch erlaubt.

Satz 13.7.1 (Lindeberg). *Für alle $n \in \mathbb{N}$ gelte, dass $X_{n1}, X_{n2}, \ldots, X_{nn}$ unabhängige Zufallsvariablen mit $\mathbb{E}X_{nk} = 0$ und $\operatorname{Var} X_{nk} = \sigma_{nk}^2$ sind, wobei $\sum_{k=1}^n \sigma_{nk}^2 = 1$. Außerdem gelte die folgende Lindeberg-Bedingung: für jedes $\varepsilon > 0$*

$$L_n(\varepsilon) := \lim_{n\to\infty} \sum_{k=1}^n \mathbb{E}[X_{nk}^2 \mathbb{1}_{|X_{nk}|>\varepsilon}] = 0. \tag{13.7.1}$$

Dann gilt:

$$X_{n1} + \ldots + X_{nn} \underset{n\to\infty}{\overset{d}{\longrightarrow}} N \sim N(0,1).$$

Bemerkung 13.7.2. Die Lindeberg-Bedingung besagt, dass alle Werte der Zufallsvariablen, die größer als ε sind, einen asymptotisch verschwindenden Beitrag zur Varianz der Summe der Zufallsvariablen leisten.

Bemerkung 13.7.3. Wir beweisen, dass der zentrale Grenzwertsatz ein Spezialfall von Satz 13.7.1 ist. Seien $X_1, X_2, \ldots$ unabhängige, identisch verteilte Zufallsvariablen mit $\mathbb{E}[X_k] = \mu$ und $\operatorname{Var} X_k = \sigma^2 = 0$. Wir definieren die Zufallsvariablen

$$X_{nk} = \frac{X_k - \mu}{\sigma\sqrt{n}}, \quad k = 1, \ldots, n.$$

Dann sind die Zufallsvariablen $X_{n1}, \ldots, X_{nn}$ für jedes $n \in \mathbb{N}$ unabhängig, da $X_1, \ldots, X_n$ nach Voraussetzung unabhängig sind. Außerdem gilt $\mathbb{E}X_{nk} = 0$ und $\sigma_{nk}^2 = \operatorname{Var} X_{nk} = \frac{1}{n}$. Somit gilt tatsächlich $\sum_{k=1}^n \sigma_{nk}^2 = 1$.

Um zu zeigen, dass alle Voraussetzungen aus Satz 13.7.1 erfüllt sind, müssen wir noch die Lindeberg–Bedingung (13.7.1) überprüfen. Sei $\varepsilon > 0$ fest. Es gilt, da die Zufallsvariablen identisch verteilt sind,

$$L_n(\varepsilon) = \sum_{k=1}^{n} \mathbb{E}\left[\left(\frac{X_k - \mu}{\sigma\sqrt{n}}\right)^2 \mathbb{1}_{\left|\frac{X_k-\mu}{\sigma\sqrt{n}}\right|>\varepsilon}\right] = \frac{1}{\sigma^2}\mathbb{E}[(X_1-\mu)^2 \mathbb{1}_{|X_1-\mu|>\varepsilon\sigma\sqrt{n}}].$$

Wir zeigen, dass der Erwartungswert auf der rechten Seite für $n \to \infty$ gegen 0 konvergiert. Es gilt

$$(X_1-\mu)^2 \mathbb{1}_{|X_1-\mu|>\varepsilon\sigma\sqrt{n}} \xrightarrow[n\to\infty]{f.s.} 0 \quad \text{(sogar sicher)},$$

da die Indikatorfunktion für hinreichend großes n gleich 0 wird. Außerdem gilt für alle n die Abschätzung

$$(X_1-\mu)^2 \mathbb{1}_{|X_1-\mu|>\varepsilon\sigma\sqrt{n}} \leq (X_1-\mu)^2 \text{ mit } \mathbb{E}\left[(X_1-\mu)^2\right] < \infty.$$

Damit können wir den Satz von der majorisierten Konvergenz anwenden. Somit gilt die Lindeberg–Bedingung (13.7.1).
Nun können wir Satz 13.7.1 von Lindeberg anwenden:

$$\frac{X_1-\mu}{\sigma\sqrt{n}} + \ldots + \frac{X_n-\mu}{\sigma\sqrt{n}} \xrightarrow[n\to\infty]{d} N \sim N(0,1).$$

Der zentrale Grenzwertsatz ist also ein Spezialfall von Satz 13.7.1.

Bemerkung 13.7.4. Der Satz von Lindeberg ist etwas allgemeiner, als der zentrale Grenzwertsatz. Zum Beispiel wird im Satz von Lindeberg nicht verlangt, dass die Zufallsvariablen $X_{n1}, \ldots, X_{nn}$ identisch verteilt sein sollen.

Beispiel 13.7.5. Definiere $X_{n1} = X$, wobei X eine beliebige (aber nicht normalverteilte) Zufallsvariable mit $\mathbb{E}X = 0$ und $\operatorname{Var} X = 1$ ist. Die restlichen Zufallsvariablen $X_{n2}, \ldots, X_{nn}$ seien gleich 0. Dann gilt

$$X_{n1} + \ldots + X_{nn} = X \nsim N(0,1).$$

In diesem Fall konvergieren die Summen also nicht gegen die Normalverteilung. Die Ursache dafür ist, dass die Lindeberg–Bedingung (13.7.1) nicht erfüllt ist.

Bemerkung 13.7.6. Wenn die Lindeberg–Bedingung (13.7.1) erfüllt ist, so gilt, dass

$$\lim_{n\to\infty} \max_{k=1,\ldots,n} \sigma_{nk}^2 = 0.$$

Beweis. Aus der Gleichung

$$\sigma_{nk}^2 = \mathbb{E}\left[X_{nk}^2\right] = \mathbb{E}\left[X_{nk}^2 \cdot \mathbb{1}_{|X_{nk}|\leq\varepsilon}\right] + \mathbb{E}\left[X_{nk}^2 \cdot \mathbb{1}_{|X_{nk}|>\varepsilon}\right]$$

folgt, dass für die maximale Varianz folgende Abschätzung gilt:

$$M_n := \max_{k=1,\ldots,n} \sigma_{nk}^2 \leq \varepsilon^2 + \sum_{k=1}^{n} \mathbb{E}[X_{nk}^2 \mathbb{1}_{|X_{nk}|>\varepsilon}] = \varepsilon^2 + L_n(\varepsilon).$$

Da die Lindeberg–Summe auf der rechten Seite gegen 0 für $n \to \infty$ konvergiert, erhalten wir, dass $\limsup_{n\to\infty} M_n \leq \varepsilon^2$. Da das für jedes $\varepsilon > 0$ gilt, folgt, dass $\lim_{n\to\infty} M_n = 0$. □

Zum Beweis von Satz 13.7.1 brauchen wir zunächst einige Hilfsaussagen.

Lemma 13.7.7. *Für alle $y_1,\dots,y_n\in\mathbb{C}$ und für alle $z_1,\dots,z_n\in\mathbb{C}$ mit $|y_k|\le 1$ und $|z_k|\le 1$ für alle $k=1,\dots,n$, gilt:*

$$\left|\prod_{k=1}^n y_k-\prod_{k=1}^n z_k\right|\le\sum_{k=1}^n|y_k-z_k|.$$

Beweis. Wir benutzen die Induktion.

Induktionsanfang. Für $n=1$ nimmt unsere Ungleichung die Form $|y_1-z_1|\le|y_1-z_1|$ an. Sie stimmt also.

Induktionsschritt. Wir nehmen an, dass die Ungleichung für ein $n\in\mathbb{N}$ gilt. Nun gilt es, zu zeigen, dass sie auch für $n+1$ gilt. Mit der Dreiecksungleichung erhalten wir, dass

$$\begin{aligned}\left|\prod_{k=1}^{n+1} y_k-\prod_{k=1}^{n+1} z_k\right|&\le\left|\prod_{k=1}^{n+1} y_k-y_{n+1}\prod_{k=1}^{n} z_k\right|+\left|y_{n+1}\prod_{k=1}^{n} z_k-\prod_{k=1}^{n+1} z_k\right|\\&=|y_{n+1}|\left|\prod_{k=1}^{n} y_k-\prod_{k=1}^{n} z_k\right|+\left|\prod_{k=1}^{n} z_k\right|\cdot|y_{n+1}-z_{n+1}|\\&\le\left|\prod_{k=1}^{n} y_k-\prod_{k=1}^{n} z_k\right|+|y_{n+1}-z_{n+1}|,\end{aligned}$$

da $|y_{n+1}|\le 1$ und $|\prod_{k=1}^n z_k|\le 1$. Nun können wir die Induktionsannahme anwenden:

$$\left|\prod_{k=1}^{n+1} y_k-\prod_{k=1}^{n+1} z_k\right|\le\sum_{k=1}^n|y_k-z_k|+|y_{n+1}-z_{n+1}|=\sum_{k=1}^{n+1}|y_k-z_k|.$$

Somit ist die Ungleichung auch für $n+1$ Terme gültig. □

Lemma 13.7.8. *Sei X eine Zufallsvariable mit $\mathbb{E}[|X|^n]<\infty$ für ein $n\in\mathbb{N}$. Dann gilt die Taylor-Entwicklung*

$$\varphi_X(t)=\sum_{k=0}^n\frac{(it)^k}{k!}\,\mathbb{E}[X^k]+R_n(t),$$

wobei für den Restterm $R_n(t)$ die folgende Abschätzung gilt

$$|R_n(t)|\le\mathbb{E}\left[\min\left\{\frac{|tX|^{n+1}}{(n+1)!},2\frac{|tX|^n}{n!}\right\}\right].$$

Bemerkung 13.7.9. Wenn man sich erlaubt, Erwartungswerte und unendliche Summen ohne Begründung zu vertauschen, dann erhält man die Entwicklung

$$\varphi_X(t)=\mathbb{E}[e^{itX}]=\mathbb{E}\left[\sum_{k=0}^\infty\frac{(itX)^k}{k!}\right]=\sum_{k=0}^\infty\frac{(it)^k}{k!}\mathbb{E}[X^k].$$

Das Lemma ist eine exakte Aussage darüber, inwieweit diese Entwicklung möglich ist. Hat die Zufallsvariable n Momente, so können wir in der Entwicklung Terme bis zum Grad n benutzen und den Restterm wie im Lemma abschätzen.

Beweis von Lemma 13.7.8.

Schritt 1. Zuerst beweisen wir eine Abschätzung für den Restterm in der Taylor–Entwicklung der Funktion e^{it}. Wir werden zeigen, dass für alle $t \in \mathbb{R}$

$$\left| e^{it} - \sum_{k=0}^{n} \frac{(it)^k}{k!} \right| \le \min\left\{ \frac{|t|^{n+1}}{(n+1)!}, 2\frac{|t|^n}{n!} \right\}. \tag{13.7.2}$$

Wir führen den Beweis mit Induktion.

Induktionsanfang. Für $n = 0$ stimmt die zu beweisende Ungleichung (13.7.2), denn

$$|e^{it} - 1| \le |t| \text{ und } |e^{it} - 1| \le 2.$$

Induktionsschritt. Wir nehmen an, dass die Ungleichung (13.7.2) für ein $n \in \mathbb{N}$ gilt. Nun ist zu zeigen, dass sie auch für $n+1$ gilt. Sei o.E.d.A. $t > 0$, denn die Funktion auf der rechten Seite von (13.7.2) ist gerade. Es gilt

$$\left| e^{it} - \sum_{k=0}^{n+1} \frac{(is)^k}{k!} \right| = \left| \int_0^t \left(e^{is} - \sum_{k=0}^{n} \frac{(is)^k}{k!} \right) ds \right| \le \int_0^t \left| e^{is} - \sum_{k=0}^{n} \frac{(is)^k}{k!} \right| ds.$$

Nun benutzen wir die Induktionsannahme:

$$\begin{aligned}
\left| e^{it} - \sum_{k=0}^{n+1} \frac{(it)^k}{k!} \right| &= \int_0^t \min\left\{ \frac{s^{n+1}}{(n+1)!}, 2\frac{s^n}{n!} \right\} ds \\
&\le \min\left\{ \int_0^t \frac{s^{n+1}}{(n+1)!}, 2\int_0^t \frac{s^n}{n!} ds \right\} \\
&= \min\left\{ \frac{t^{n+2}}{(n+2)!}, 2\frac{t^{n+1}}{(n+1)!} \right\}.
\end{aligned}$$

Die zu beweisende Ungleichung gilt also auch für $n+1$.

Schritt 2. Nun beweisen wir die Abschätzung für den Restterm $R_n(t)$:

$$|R_n(t)| = \left| \mathbb{E}\left[e^{itX} - \sum_{k=0}^{n} \frac{(itX)^k)}{k!} \right] \right| \le \mathbb{E}\left| e^{itX} - \sum_{k=0}^{n} \frac{(itX)^k}{k!} \right| \le \mathbb{E}\left[\min\left\{ \frac{|tX|^{n+1}}{(n+1)!}, 2\frac{|tX|^n}{n!} \right\} \right].$$

Das Lemma ist somit bewiesen. □

Nun können wir den Satz von Lindeberg beweisen.
Beweis von Satz 13.7.1. Es ist zu zeigen, dass

$$X_{n1} + \ldots + X_{nn} \xrightarrow[n\to\infty]{d} N(0,1).$$

Schritt 1. Nach dem Stetigkeitssatz 13.4.1 reicht es, zu zeigen, dass für alle $t \in \mathbb{R}$ gilt:

$$\lim_{n\to\infty} \prod_{k=1}^{n} \varphi_{nk}(t) = e^{-\frac{t^2}{2}}.$$

Dabei sei $\varphi_{nk}(t) = \mathbb{E}[e^{itX_{nk}}]$ die charakteristische Funktion von X_{nk}. Deshalb betrachten wir die Differenz dieser beiden Terme:

$$\begin{aligned}
\left|\prod_{k=1}^{n} \varphi_{nk}(t) - e^{-\frac{t^2}{2}}\right| &= \left|\prod_{k=1}^{n} \varphi_{nk}(t) - \prod_{k=1}^{n} e^{-\frac{1}{2}\sigma_{nk}^2 t^2}\right| \qquad \left(\text{da } \sum_{k=1}^{n} \sigma_{nk}^2 = 1\right) \\
&\leq \sum_{k=1}^{n} \left|\varphi_{nk}(t) - e^{-\frac{1}{2}\sigma_{nk}^2 t^2}\right| \\
&\leq \sum_{k=1}^{n} \left|\varphi_{nk}(t) - 1 + \frac{1}{2}\sigma_{nk}^2 t^2\right| - \sum_{k=1}^{n} \left|e^{-\frac{1}{2}\sigma_{nk}^2 t^2} - 1 + \frac{1}{2}\sigma_{nk}^2 t^2\right|.
\end{aligned}$$

Nun schätzen wir diese beiden Summen separat ab und definieren hierfür

$$c_{nk} = \left|\varphi_{nk}(t) - 1 + \frac{1}{2}\sigma_{nk}^2 t^2\right|, \quad d_{nk} = \left|e^{-\frac{1}{2}\sigma_{nk}^2 t^2} - 1 + \frac{1}{2}\sigma_{nk}^2 t^2\right|.$$

Schritt 2. Betrachten zunächst c_{nk}. Mit Lemma 13.7.8 erhalten wir

$$c_{nk} = \left|\varphi_{nk}(t) - \left(1 - \frac{1}{2}\sigma_{nk}^2 t^2\right)\right| \leq \mathbb{E}\left[\min\left\{|tX_{nk}|^3, |tX_{nk}|^2\right\}\right].$$

Und nun verwenden wir für $X_{nk} \leq \varepsilon$ die Abschätzung $|tX_{nk}|^3$ und für $X_{nk} > \varepsilon$ die Abschätzung $|tX_{nk}|^2$:

$$c_{nk} \leq \mathbb{E}\left[|tX_{nk}|^3 \mathbb{1}_{|X_{nk}| \leq \varepsilon}\right] + \mathbb{E}\left[|tX_{nk}|^2 \mathbb{1}_{|X_{nk}| > \varepsilon \cdot}\right].$$

Sei $L_n(\varepsilon) = \sum_{k=1}^{n} \mathbb{E}[X_{nk}^2 \mathbb{1}_{|X_{nk}| > \varepsilon}]$ die Summe aus der Lindeberg–Bedingung. Nun schätzen wir die Summe über c_{nk} ab:

$$\begin{aligned}
\sum_{k=1}^{n} c_{nk} &\leq |t|^3 \sum_{k=1}^{n} \mathbb{E}[|X_{nk}|^3 \mathbb{1}_{|X_{nk| \leq \varepsilon}}] + |t|^2 L_n(\varepsilon) \\
&\leq |t|^3 \sum_{k=1}^{n} \varepsilon \cdot \mathbb{E}[|X_{nk}|^2 \mathbb{1}_{|X_{nk}| \leq \varepsilon}] + |t|^2 L_n(\varepsilon) \\
&\leq |t|^3 \varepsilon \sum_{k=1}^{n} \sigma_{nk}^2 + |t|^2 L_n(\varepsilon) \\
&= |t|^3 \varepsilon + |t|^2 L_n(\varepsilon).
\end{aligned}$$

Da die Lindenberg-Summe $L_n(\varepsilon)$ gegen 0 für $n \to \infty$ konvergiert und $\varepsilon > 0$ beliebig klein gewählt werden kann, erhalten wir, dass

$$\lim_{n\to\infty} \sum_{k=1}^{n} c_{nk} = 0.$$

Schritt 3. Wir betrachten nun d_{nk}. Wir verwenden hier die Abschätzung

$$|e^z - (1 - z)| = \left|\sum_{k=2}^{\infty} \frac{z^k}{k!}\right| \leq |z|^2 \sum_{k=2}^{\infty} \frac{|z|^{k-2}}{k!} \leq |z|^2 \sum_{k=2}^{\infty} \frac{|z|^{k-2}}{(k-2)!} \leq |z|^2 e^{|z|}.$$

Und nun substituieren wir $z = -\frac{1}{2}\sigma_{nk}^2 t^2$:

$$d_{nk} = \left| e^{-\frac{1}{2}\sigma_{nk}^2 t^2} - \left(1 - \frac{1}{2}\sigma_{nk}^2 t^2\right) \right| \leq \sigma_{nk}^4 |t|^4 e^{\frac{1}{2}\sigma_{nk}^2 t^2} \leq \sigma_{nk}^4 |t|^4 e^{t^2}.$$

Somit gilt für die Summe

$$\sum_{k=1}^n d_{nk} \leq |t|^4 e^{t^2} \sum_{k=1}^n \sigma_{nk}^4 \leq |t|^4 e^{t^2} \max_{k=1,\ldots,n} \sigma_{nk}^2.$$

Also gilt laut Bemerkung 13.7.6

$$\lim_{n\to\infty} \sum_{k=1}^n d_{nk} = 0.$$

Schritt 4. Wir haben also gezeigt, dass

$$\lim_{n\to\infty} \left| \prod_{k=1}^n \varphi_{nk}(t) - e^{-\frac{t^2}{2}} \right| = 0.$$

Mit dem Stetigkeitssatz 13.4.1 folgt daraus, dass $X_{n1} + \ldots + X_{nn} \xrightarrow[n\to\infty]{d} N(0,1)$. □

Nun beweisen wir einen Satz, der eine vereinfachte Form des Satzes von Lindeberg ist.

Satz 13.7.10 (Ljapunow). *Für alle $n \in \mathbb{N}$ seien $X_{n1}, \ldots, X_{nn}$ unabhängige Zufallsvariablen mit $\mathbb{E}X_{nk} = 0$ und $\operatorname{Var} X_{nk} = \sigma_{nk}^2$, wobei $\sum_{k=1}^n \sigma_{nk}^2 = 1$. Es gelte außerdem die folgende Ljapunow–Bedingung: Es existiert ein $\delta > 0$ mit*

$$\lim_{n\to\infty} \sum_{k=1}^n \mathbb{E}|X_{nk}|^{2+\delta} = 0. \tag{13.7.3}$$

Dann gilt:

$$X_{n1} + \ldots + X_{nn} \xrightarrow[n\to\infty]{d} N \sim N(0,1).$$

Beweis. Wir zeigen, dass die Ljapunow–Bedingung (13.7.3) die Lindeberg–Bedingung (13.7.1) impliziert. Sei also (13.7.3) erfüllt. Sei $\varepsilon > 0$ beliebig. Dann gilt

$$L_n(\varepsilon) = \sum_{k=1}^n \mathbb{E}\left[X_{nk}^2 \mathbb{1}_{|X_{nk}|>\varepsilon}\right] \leq \sum_{k=1}^n \mathbb{E}\left[\frac{|X_{nk}|^{2+\delta}}{\varepsilon^\delta} \mathbb{1}_{|X_{nk}|>\varepsilon}\right] \leq \frac{1}{\varepsilon^\delta} \sum_{k=1}^n \mathbb{E}\left[|X_{nk}|^{2+\delta}\right] \underset{n\to\infty}{\longrightarrow} 0.$$

Somit gilt die Lindeberg–Bedingung (13.7.1). □

Beispiel 13.7.11. Wir betrachten ein n-faches Bernoulli-Experimenten mit Erfolgswahrscheinlichkeit p_n, für die $\lim_{n\to\infty} np_n(1-p_n) = \infty$ gilt. Dann gilt:

$$\frac{S_n - np_n}{\sqrt{np_n(1-p_n)}} \xrightarrow[n\to\infty]{d} N \sim N(0,1). \tag{13.7.4}$$

Bemerkung 13.7.12. Dies ist beispielsweise anwendbar, wenn $p_n = \frac{1}{\sqrt{n}}$, $p_n = 1 - \frac{1}{\sqrt{n}}$ oder wenn $p_n = p \in (0,1)$ konstant ist. Zum Vergleich: Der Poisson–Grenzwertsatz gilt, wenn $\lim_{n\to\infty} np_n = \lambda \in (0,\infty)$.

Beweis von (13.7.4). Den Satz von de Moivre–Laplace können wir nicht anwenden, denn p_n ist im Allgemeinen nicht konstant. Wir werden stattdessen den Satz von Lindeberg anwenden. Wir betrachten unabhängige Zufallsvariablen $Y_{n1}, \ldots, Y_{nn} \sim \operatorname{Bin}(1, p_n)$, d.h.

$$\mathbb{P}[Y_{nk} = 1] = p_n, \quad \mathbb{P}[Y_{nk} = 0] = 1 - p_n.$$

Für die Summe dieser Zufallsvariablen gilt dann

$$S_n := Y_{n1} + \ldots + Y_{nn} \sim \operatorname{Bin}(n, p_n).$$

Wir betrachten auch die normierten Zufallsvariablen

$$X_{nk} := \frac{Y_{nk} - p_n}{\sqrt{np_n(1-p_n)}}.$$

Es gilt $\mathbb{E}X_{nk} = 0$ und $\sum_{k=1}^n \operatorname{Var} X_{nk} = 1$. Für die Summe dieser Zufallsvariablen X_{nk} gilt dann

$$S_n^* := X_{n1} + \ldots + X_{nn} = \frac{S_n - np_n}{\sqrt{np_n(1-p_n)}}.$$

Wir zeigen, dass S_n^* in Verteilung gegen die Standardnormalverteilung konvergiert, indem wir die Lindeberg–Bedingung überprüfen. Sei $\varepsilon > 0$ vorgegeben. Wir können die Zufallsvariable X_{nk} wie folgt abschätzen:

$$|X_{nk}| \leq \frac{1}{\sqrt{np_n(1-p_n)}}.$$

Somit ist $|X_{nk}| < \varepsilon$ für hinreichend großes n, denn $\lim_{n\to\infty} np_n(1-p_n) = \infty$. Es gilt also $L_n(\varepsilon) = \sum_{k=1}^n \mathbb{E}[X_{nk}^2 \mathbb{1}_{|X_{nk}|>\varepsilon}] = 0$ für hinreichend großes n. Somit ist die Lindeberg–Bedingung erfüllt. Nun darf der zentrale Grenzwertsatz von Lindeberg benutzt werden. Er besagt, dass

$$S_n^* \xrightarrow[n\to\infty]{d} N(0,1).$$

□

Beispiel 13.7.13. Seien $X_1, X_2, \ldots$ unabhängige (aber nicht identisch verteilte) Zufallsvariablen mit

$$\mathbb{P}[X_n = n] = \mathbb{P}[X_n = -n] = \frac{1}{2}.$$

Definiere die Partialsummen

$$S_n = X_1 + \ldots + X_n = \pm 1 \pm 2 \pm \ldots \pm n.$$

Wir behaupten nun, dass

$$\sqrt{3} \cdot \frac{S_n}{n^{3/2}} \xrightarrow[n\to\infty]{d} N(0,1).$$

Beweis. Den zentralen Grenzwertsatz können wir nicht anwenden, da die Zufallsvariablen nicht identisch verteilt sind. Stattdessen muss man den Satz von Lindeberg oder den Satz von Ljapunow anwenden. Wir entscheiden uns für den Satz von Ljapunow. Zunächst bemerken wir, dass $\mathbb{E}X_k = 0$ und $\operatorname{Var} X_k = \mathbb{E}[X_k^2] = k^2$. Die Varianz der Summe S_n ist dann

$$D_n^2 := \operatorname{Var} S_n = \sum_{k=1}^n k^2 = \frac{n(n+1)(2n+1)}{6} \sim \frac{n^3}{3}, \quad n \to \infty.$$

Daraus folgt, dass

$$D_n \sim \frac{n^{3/2}}{\sqrt{3}}, \quad n \to \infty.$$

Um nun den zentralen Grenzwertsatz von Ljapunow anwenden zu können, definieren wir uns folgende Zufallsvariablen:

$$X_{nk} = \frac{X_k}{D_n} \text{ für } k = 1, \dots, n.$$

Wir erkennen, dass

$$\mathbb{E}X_{nk} = 0, \quad \sum_{k=1}^{n} \operatorname{Var} X_{nk} = \frac{1}{D_n^2} \sum_{k=1}^{n} \operatorname{Var} X_k = 1.$$

Nun zeigen wir, dass die Ljapunow-Bedingung mit $\delta = 1$ erfüllt ist:

$$\sum_{k=1}^{n} \mathbb{E}[|X_{nk}|^3] = \sum_{k=1}^{n} \left(\frac{k}{D_n}\right)^3 = \frac{1}{D_n^3} \sum_{k=1}^{n} k^3 \xrightarrow[n\to\infty]{} 0,$$

denn $\sum_{k=1}^{n} k^3 \sim \frac{n^4}{4}$ und $D_n^3 \sim \frac{n^{9/2}}{3\sqrt{3}}$ für $n \to \infty$. Der Satz Ljapunow besagt dann, dass

$$\frac{1}{D_n} \sum_{k=1}^{n} X_k = \sum_{k=1}^{n} X_{nk} \xrightarrow[n\to\infty]{d} N(0,1).$$

Also ist die Behauptung bewiesen, da $D_n \sim \frac{n^{3/2}}{\sqrt{3}}$ für $n \to \infty$ ist.