Zentraler Grenzwertsatz für i.i.d. Zufallsvariablen mit endlicher Varianz

Es sei $(X_n)_{n\in\mathbb{N}}$ eine Folge von i.i.d. Zufallsvariablen auf $(\Omega, \mathcal{A}, P)$ mit endlicher, positiver Varianz: $0 < \mathrm{Var}_P(X_1) < \infty$. Es sei

$$S_n = \sum_{k=1}^{n} X_k, \tag{1}$$

$$Z_n = \frac{S_n - E_P[S_n]}{\sigma_P(S_n)} = \frac{S_n - E_P[S_n]}{\sqrt{n}\sigma_P(S_1)}. \tag{2}$$

Weiter sei Z eine standardnormalverteilte Zufallsvariable. Dann gilt für alle Intervalle $I \subseteq \mathbb{R}$ (gleichgültig, ob beschränkt oder unbeschränkt, offen, abgeschlossen oder halboffen):

$$P[Z_n \in I] \overset{n\to\infty}{\longrightarrow} P[Z \in I] = \frac{1}{\sqrt{2\pi}} \int_I e^{-z^2/2}\, dz. \tag{3}$$

Bemerkungen.

- Das gibt eine Rechtfertigung für das häufige Auftreten der Normalverteilung für zufällige Schwankungen, wenn diese sich als eine Summe von sehr vielen unabhängigen gleichartigen Beiträgen ergeben.
- Es gelten noch allgemeinere Varianten des Zentralen Grenzwertsatzes, bei denen man auf die identische Verteilung aller X_k verzichten kann, solange die einzelnen Summanden "nur wenig" zur gesamten Summe beitragen. Selbst die Unabhängigkeitsvoraussetzung kann man etwas abschwächen. Solche Varianten werden in höheren Stochastikvorlesungen präzise formuliert und bewiesen.

Wir führen den Zentralen Grenzwertsatz auf die folgende Variante zurück:

Satz 0.1 Es sei $f : \mathbb{R} \to \mathbb{R}$ dreimal stetig differenzierbar und beschränkt mit beschränkten Ableitungen bis zur 3. Stufe (Notation dafür: $f \in C_b^3(\mathbb{R})$.) Dann gilt mit den Bezeichnungen und Voraussetzungen von oben:

$$E_P[f(Z_n)] \overset{n\to\infty}{\longrightarrow} E_P[f(Z)].$$

Beweis: Ohne Beschränkung der Allgemeinheit gelte $E_P[X_n] = 0$ und $\mathrm{Var}_P(X_n) = 1$; sonst ersetzen wir die Zufallsvariablen X_n durch ihre standardisierten Versionen

$$\tilde{X}_n = \frac{X_n - E_P[X_n]}{\sigma_P(X_n)}.$$

Weiter existiere ohne Beschränkung der Allgemeinheit auf dem *gleichen* Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$ eine i.i.d. Folge $(Y_n)_{n\in\mathbb{N}}$ standardnormalverteilter Zufallsvariablen, unabhängig von der Folge $(X_n)_{n\in\mathbb{N}}$. (Wenn nötig ersetzen wir hierzu die Folge $(X_n)_{n\in\mathbb{N}}$ durch eine andere i.i.d. Folge mit der gleichen Verteilung auf einem anderen Wahrscheinlichkeitsraum, z.B. einem Produktraum.) Dann gilt:

$$Z_n = \frac{1}{\sqrt{n}} \sum_{k=1}^{n} X_k.$$

Weiter ist

$$\frac{1}{\sqrt{n}} \sum_{k=1}^{n} Y_k$$

standardnormalverteilt, denn aus der Faltungseigenschaft der Normalverteilung folgt:

$$\begin{aligned}
& Y_1, \dots, Y_n \text{ sind i.i.d. standardnormalverteilt} \\
\Rightarrow \quad & \sum_{k=1}^{n} Y_k \text{ ist normalverteilt mit Erwartung 0 und Varianz } n \\
\Rightarrow \quad & \frac{1}{\sqrt{n}} \sum_{k=1}^{n} Y_k \text{ ist standardnormalverteilt.}
\end{aligned} \tag{4}$$

Wir müssen also zeigen:

$$E_P\left[f\left(\frac{1}{\sqrt{n}}\sum_{k=1}^{n} X_k\right)\right] - E_P\left[f\left(\frac{1}{\sqrt{n}}\sum_{k=1}^{n} Y_k\right)\right] \stackrel{n\to\infty}{\longrightarrow} 0. \tag{5}$$

Hierzu zerlegen wir diese Differenz in eine Teleskopsumme, indem wir sukzessive einen Summanden X_k nach dem anderen durch das entsprechende normalverteilte Y_k austauschen. Für den l-ten Austauschschritt, $l = 1, \dots, n$, bezeichne

$$T_{n,l} := \frac{1}{\sqrt{n}} \sum_{k=1}^{l-1} Y_k + \frac{1}{\sqrt{n}} \sum_{k=l+1}^{n} X_k$$

die Summe derjenigen Summanden, die gerade nicht ausgetauscht werden. Es gilt also

$$\begin{aligned}
T_{n,l} + X_l &= \frac{1}{\sqrt{n}} \sum_{k=1}^{l-1} Y_k + \frac{1}{\sqrt{n}} \sum_{k=l}^{n} X_k, \\
T_{n,l} + Y_l &= \frac{1}{\sqrt{n}} \sum_{k=1}^{l} Y_k + \frac{1}{\sqrt{n}} \sum_{k=l+1}^{n} X_k
\end{aligned} \tag{6}$$

und wir erhalten

$$\begin{aligned}
&E_P\left[f\left(\frac{1}{\sqrt{n}}\sum_{k=1}^n X_k\right)-f\left(\frac{1}{\sqrt{n}}\sum_{k=1}^n Y_k\right)\right]\\
&=\sum_{l=1}^n E_P\left[f(T_{n,l}+\frac{X_l}{\sqrt{n}})-f(T_{n,l}+\frac{Y_l}{\sqrt{n}})\right] \qquad (7)
\end{aligned}$$

Wir entwickeln f nach Taylor mit zwei verschiedenen Darstellungen des Restglieds: Für alle $t, x \in \mathbb{R}$ existieren $\theta_2, \theta_3 \in [0,1]$ mit

$$f(t+x)=f(t)+xf'(t)+\frac{x^2}{2}f''(t)+\frac{x^3}{6}f'''(t+\theta_3 x), \qquad (8)$$

$$f(t+x)=f(t)+xf'(t)+\frac{x^2}{2}f''(t)+\frac{x^2}{2}(f''(t+\theta_2 x)-f''(t)), \qquad (9)$$

also

$$f(t+x)=f(t)+xf'(t)+\frac{x^2}{2}f''(t)+r(t,x) \qquad (10)$$

mit einem Restterm $r(t,x)$, der

$$|r(t,x)|\le \min\left\{x^2\sup_{t\in\mathbb{R}}|f''(t)|,\frac{|x|^3}{6}\sup_{t\in\mathbb{R}}|f'''(t)|\right\}\le c_f\min\{x^2,|x|^3\} \qquad (11)$$

erfüllt, wobei

$$c_f:=\max\left\{\sup_{t\in\mathbb{R}}|f''(t)|,\frac{1}{6}\sup_{t\in\mathbb{R}}|f'''(t)|\right\}. \qquad (12)$$

Es folgt für die Bestandteile des l-ten Summanden in der Zerlegung (7):

$$\begin{aligned}
&E_P\left[f\left(T_{n,l}+\frac{X_l}{\sqrt{n}}\right)\right]\\
&=E_P\left[f(T_{n,l})\right]+E_P\left[\frac{X_l}{\sqrt{n}}f'(T_{n,l})\right]+\frac{1}{2}E_P\left[\frac{X_l^2}{n}f''(T_{n,l})\right]+E_P\left[r\left(T_{n,l},\frac{X_l}{\sqrt{n}}\right)\right] \quad (13)
\end{aligned}$$

und ebenso

$$\begin{aligned}
&E_P\left[f\left(T_{n,l}+\frac{Y_l}{\sqrt{n}}\right)\right]\\
&=E_P\left[f(T_{n,l})\right]+E_P\left[\frac{Y_l}{\sqrt{n}}f'(T_{n,l})\right]+\frac{1}{2}E_P\left[\frac{Y_l^2}{n}f''(T_{n,l})\right]+E_P\left[r\left(T_{n,l},\frac{Y_l}{\sqrt{n}}\right)\right]. \quad (14)
\end{aligned}$$

Bilden wir die Differenz:

$$
\begin{aligned}
&E_P\left[f\left(T_{n,l}+\frac{X_l}{\sqrt{n}}\right)-f\left(T_{n,l}+\frac{Y_l}{\sqrt{n}}\right)\right]\\
=&E_P\left[\frac{X_l}{\sqrt{n}}f'(T_{n,l})\right]-E_P\left[\frac{Y_l}{\sqrt{n}}f'(T_{n,l})\right]+\frac{1}{2}E_P\left[\frac{X_l^2}{n}f''(T_{n,l})\right]-\frac{1}{2}E_P\left[\frac{Y_l^2}{n}f''(T_{n,l})\right]\\
&+E_P\left[r\left(T_{n,l},\frac{X_l}{\sqrt{n}}\right)\right]-E_P\left[r\left(T_{n,l},\frac{Y_l}{\sqrt{n}}\right)\right]. \qquad (15)
\end{aligned}
$$

Nun ist $T_{n,l}$ bezüglich $\sigma(Y_1,\dots,Y_{l-1},X_{l+1},\dots,X_n)$ meßbar und folglich unabhängig von X_l und Y_l. Wir schließen:

$$
E_P\left[\frac{X_l}{\sqrt{n}}f'(T_{n,l})\right]=\frac{1}{\sqrt{n}}E_P[X_l]E_P[f'(T_{n,l})]=0 \qquad (16)
$$

wegen $E_P[X_l]=0$ und

$$
E_P\left[\frac{X_l^2}{n}f''(T_{n,l})\right]=\frac{1}{n}E_P[X_l^2]E_P[f''(T_{n,l})]=\frac{1}{n}E_P[f''(T_{n,l})] \qquad (17)
$$

wegen $E_P[X_l^2]=1$. Analog für Y_l:

$$
E_P\left[\frac{Y_l}{\sqrt{n}}f'(T_{n,l})\right]=\frac{1}{\sqrt{n}}E_P[Y_l]E_P[f'(T_{n,l})]=0 \qquad (18)
$$

wegen $E_P[Y_l]=0$ und

$$
E_P\left[\frac{Y_l^2}{n}f''(T_{n,l})\right]=\frac{1}{n}E_P[Y_l^2]E_P[f''(T_{n,l})]=\frac{1}{n}E_P[f''(T_{n,l})] \qquad (19)
$$

wegen $E_P[Y_l^2]=1$. In (15) eingesetzt bleiben auf der rechten Seite nur die Restterme übrig, und wir erhalten

$$
\begin{aligned}
&\left|E_P\left[f\left(T_{n,l}+\frac{X_l}{\sqrt{n}}\right)-f\left(T_{n,l}+\frac{Y_l}{\sqrt{n}}\right)\right]\right|\\
=&\left|E_P\left[r\left(T_{n,l},\frac{X_l}{\sqrt{n}}\right)\right]-E_P\left[r\left(T_{n,l},\frac{Y_l}{\sqrt{n}}\right)\right]\right|\\
\leq&E_P\left[\left|r\left(T_{n,l},\frac{X_l}{\sqrt{n}}\right)\right|\right]+E_P\left[\left|r\left(T_{n,l},\frac{Y_l}{\sqrt{n}}\right)\right|\right]\\
\leq&c_fE_P\left[\min\left\{\left(\frac{X_l}{\sqrt{n}}\right)^2,\left|\frac{X_l}{\sqrt{n}}\right|^3\right\}\right]+c_fE_P\left[\min\left\{\left(\frac{Y_l}{\sqrt{n}}\right)^2,\left|\frac{Y_l}{\sqrt{n}}\right|^3\right\}\right]\\
\leq&\frac{c_f}{n}E_P\left[\min\left\{X_1^2,\frac{|X_1|^3}{\sqrt{n}}\right\}\right]+\frac{c_f}{n}E_P\left[\min\left\{Y_1^2,\frac{|Y_1|^3}{\sqrt{n}}\right\}\right] \qquad (20)
\end{aligned}
$$

wobei wir die Restgliedabschätzung (11) verwendet haben und im letzten Schritt die Zufallsvariablen X_l bzw. Y_l durch X_1 bzw. Y_1, die jeweils die gleiche Verteilung haben, ersetzt haben. Eingesetzt in die Teleskopsumme (7) haben wir für alle n Summanden die gleiche Abschätzung:

$$\begin{aligned}
&\left| E_P \left[f\left(\frac{1}{\sqrt{n}} \sum_{k=1}^n X_k \right) - f\left(\frac{1}{\sqrt{n}} \sum_{k=1}^n Y_k \right) \right] \right| \\
&\leq \sum_{l=1}^n \left| E_P \left[f(T_{n,l} + \frac{X_l}{\sqrt{n}}) - f(T_{n,l} + \frac{Y_l}{\sqrt{n}}) \right] \right| \\
&\leq \underbrace{n \cdot \frac{c_f}{n}}_{=c_f} \left(E_P \left[\min\left\{ X_1^2, \frac{|X_1|^3}{\sqrt{n}} \right\} \right] + E_P \left[\min\left\{ Y_1^2, \frac{|Y_1|^3}{\sqrt{n}} \right\} \right] \right) \stackrel{n\to\infty}{\longrightarrow} 0,
\end{aligned} \tag{21}$$

denn es gilt für alle $X \in \mathcal{L}^2(\Omega, \mathcal{A}, P)$ (und damit auch für $X = X_1$ und $X = Y_1$):

$$E_P \left[\min\left\{ X^2, \frac{|X|^3}{\sqrt{n}} \right\} \right] \stackrel{n\to\infty}{\longrightarrow} E_P[0] = 0. \tag{22}$$

Die letzte Aussage (22) folgt aus dem Satz von der dominierten Konvergenz. Seine Voraussetzungen sind hier erfüllt:

- Einerseits gilt punktweise Konvergenz:
$$\forall \omega \in \Omega: \quad \min\left\{ X(\omega)^2, \frac{|X(\omega)|^3}{\sqrt{n}} \right\} \stackrel{n\to\infty}{\longrightarrow} 0$$
- Andererseits haben wir die integrierbare Majorante
$$0 \leq \min\left\{ X^2, \frac{|X|^3}{\sqrt{n}} \right\} \leq X^2 \in \mathcal{L}^1(\Omega, \mathcal{A}, P).$$

Die Abschätzung (21) zeigt genau die Behauptung des Satzes. □

Ein Punkt $a \in \mathbb{R}$ heißt *Stetigkeitspunkt* einer Funktion $F : \mathbb{R} \to \mathbb{R}$, wenn F stetig an der Stelle a ist.

Der Zentrale Grenzwertsatz folgt nun aus der Implikation 2)⇒3) im folgenden Satz:

Satz 0.2 Es sei $(Z_n)_{n\in\mathbb{N}}$ eine Folge von reellwertigen Zufallsvariablen und Z eine weitere reellwertige Zufallsvariable. Z besitze die Verteilungsfunktion F. Dann sind äquivalent:

1) Für jede beschränkte stetige Funktion $f : \mathbb{R} \to \mathbb{R}$ gilt
$$E_P[f(Z_n)] \stackrel{n\to\infty}{\longrightarrow} E_P[f(Z)]. \tag{23}$$

2) Für jedes $f \in C_b^3(\mathbb{R})$ gilt
$$E_P[f(Z_n)] \stackrel{n\to\infty}{\longrightarrow} E_P[f(Z)]. \tag{24}$$

3) Für jedes Intervall $I = [a, b],]a, b], [a, b[$ oder $]a, b[$ mit $a, b \in \mathbb{R} \cup \{\pm\infty\}$, so dass a und b Stetigkeitspunkte von F oder $\pm\infty$ sind, gilt:
$$P[Z_n \in I] \stackrel{n\to\infty}{\longrightarrow} P[Z \in I] \tag{25}$$

Bemerkung. Ist Z standardnormalverteilt, so ist F stetig. In diesem Fall liefert die Einschränkung in 3) auf Stetigkeitspunkte von F keine Bedingung.

Sprechweise: Falls die äquivalenten Bedingungen 1)–3) des Satzes gelten, sagt man: "Z_n *konvergiert in Verteilung gegen* Z" oder auch "Z_n *konvergiert schwach gegen* Z".

Wir beweisen hier nur die Implikation 2)⇒3) des Satzes, die wir im Beweis des Zentralen Grenzwertsatzes brauchen:

Beweis zu 2)⇒3): Nach der Stetigkeitsvoraussetzung von F an den Grenzen a, b des Intervalls I gibt es ein offenes Intervall $I_1 \supseteq [a,b] \cap \mathbb{R}$ und ein abgeschlossenes Intervall $I_2 \subseteq]a,b[$ mit

$$P[Z \in I_1] - P[Z \in I] < \epsilon \quad \text{und} \quad P[Z \in I] - P[Z \in I_2] < \epsilon. \tag{26}$$

Wir wählen $f_1, f_2 : \mathbb{R} \to [0,1]$ mit $f_1, f_2 \in C_b^3(\mathbb{R})$ und

$$1_{I_1} \geq f_1 \geq 1_I \geq f_2 \geq 1_{I_2}$$

Solche f_1 und f_2 existieren.[1] Dann gilt aufgrung der Voraussetzung 2):

$$P[Z_n \in I] \leq E_P[f_1(Z_n)] \overset{n\to\infty}{\longrightarrow} E_P[f_1(Z)] \leq P[Z \in I_1] \leq P[Z \in I] + \epsilon, \tag{27}$$
$$P[Z_n \in I] \geq E_P[f_2(Z_n)] \overset{n\to\infty}{\longrightarrow} E_P[f_2(Z)] \geq P[Z \in I_2] \geq P[Z \in I] - \epsilon. \tag{28}$$

Weil $\epsilon > 0$ beliebig war, folgt die Behauptung 3):

$$P[Z_n \in I] \overset{n\to\infty}{\longrightarrow} P[Z \in I]. \tag{29}$$

□

---

[1] Man kann f_1 und f_2 sogar unendlich oft differenzierbar wählen, wenn man will. Hierzu eine Illustration im Beispiel $I = [a, \infty[$ mit $a \in \mathbb{R}$, $I_1 =]a - \delta, \infty[$, $I_2 = [a + \delta, \infty[$: Setzen wir

$$g : \mathbb{R} \to \mathbb{R}, \quad g(x) = \begin{cases} \exp \frac{1}{x^2 - x} & \text{für } 0 < x < 1 \\ 0 & \text{für } x \in \mathbb{R} \setminus]0,1[\end{cases}$$

so ist g beliebig oft differenzierbar, also auch ihre Stammfunktion

$$h : \mathbb{R} \to \mathbb{R}, \quad h(x) = \int_{-\infty}^{x} g(t)\, dt.$$

Normieren wir noch h mittels $h^* := h / \lim_{t\to\infty} h(t)$, so gilt: $h^* \in C^\infty(\mathbb{R})$, $0 \leq h^* \leq 1$, $h^*(x) = 0$ für $x \leq 0$ und $h^*(x) = 1$ für $x \geq 1$. Dann leisten

$$f_1(x) = h((x - a - \delta)/\delta),$$
$$f_2(x) = h((x - a)/\delta)$$

das Gewünschte.