Die Stein'sche Methode und Anwendungen

Bachelorarbeit
im Studiengang Wirtschaftsmathematik
am Fachbereich Mathematik und Informatik
der Philipps-Universität Marburg

eingereicht von
Tim Josek
Marburg, September 2012
betreut durch
Prof. Dr. Hajo Holzmann

Inhaltsverzeichnis

Einleitung

In der Wahrscheinlichkeitstheorie und der Stochastik kommen Grenzwertsätzen, wie dem Zentralen Grenzwertsatz oder dem starken Gesetz der großen Zahlen, eine entscheidende Bedeutung zu. Sie werden beispielsweise in der Statistik angewandt, um die Verteilungen von Teststatistiken zu studieren. In diesem Zusammenhang findet meist die Theorie der charakteristischen Funktion und der Fourier-Transformierten Anwendung, welche jedoch für gewöhnlich den Nachteil mitsichbringt, dass sie keine Konvergenzraten liefert. Den Mathematikern A. C. Berry und C.-G. Esseen gelang es 1941 einen Satz zu beweisen, welcher eine Aussage über die Güte der Konvergenz, in Form einer exakten Schranke, im Zentralen Grenzwertsatz der Wahrscheinlichkeitstheorie trifft. Heute ist diese Aussage als Satz von Berry-Esseen bekannt.
Im Jahr 1972 entwickelte Charles Stein, ein amerikanischer Statistiker, ein neues Verfahren zur Bestimmung von Schranken für die Entfernung zweier Wahrscheinlichkeitsverteilungen bezüglich einer entsprechenden Wahrscheinlichkeitsmetrik, die so genannte Stein'sche Methode. Ursprünglich verwand Charles Stein seine Methode zum Beweis des Zentralen Grenzwertsatzes und für die Herleitung von Konvergenzgeschwindigkeiten. Im Jahr 1975 studierte Louis H. Y. Chen, ein Student von Stein, dessen Methode für die Poissonapproximation. In den folgenden Jahren wurde die Stein'sche Methode für einige weitere Bereiche, wie die stochastischen Prozesse oder den multivariaten Zentralen Gerenzwertsatz, weiter entwickelt. In dieser Arbeit, welche auf der Grundlage von zwei Papern von Peter Eichelsbacher ([5],[6]) erarbeitet wurde, stellen wir die Stein'sche Methode für die Normalenapproximation vor.
Im ersten Kapitel dieser Arbeit geben wir eine Wiederholung der grundlegenden Sätze und Definitionen der Wahrscheinlichkeitstheorie, welche im Folgenden als Grundlage für unsere Berechnungen dienen wird.

Das zweite Kapitel stellt die Grundlagen der Stein'schen Methode vor.
Im folgenden dritten Kapitel wenden wir dann die erarbeitete Theorie der Stein'schen Methode für unabhängig verteilte Zufallsvariablen an.
Das vierte Kapitel beschäftigt sich mit der Thematik der sogenannten austauschbaren Paare, eine von mehreren allgemeinen Techniken zur Berechnung der Stein'schen Schranken, welche auch für nicht unabhängige Summanden anwendbar ist.
Im abschließenden fünften Kapitel geben wir mittels Stein'scher Methode einen Beweis, welcher auf E. Bolthausen zurückgeht, für den klassischen Satz von Berry-Esseen aus dem 2. Kapitel. Hierbei wird uns eine Verbesserung der universellen Konstanten von 25 auf 14 (sogar 13 kann erreicht werden) gelingen.

1 Grundlagen der Wahrscheinlichkeitstheorie

In diesem ersten einleitenden Kapitel werden wir zunächst einige wichtige Definitionen nennen und grundlegende Sätze der klassischen Wahrscheinlichkeitstheorie erläutern, welche wir im weiteren Verlauf nutzen werden. Im Anschluss beschäftigen wir uns kurz mit dem Zentralen Grenzwertsatz und gehen auf die klassischen Beweismethoden ein.

Definition 1.0.1. (Wahrscheinlichkeitsraum, Zufallsvariable)

a) Das Tripel $(\Omega, \mathcal{F}, P)$ heißt Wahrscheinlichkeitsraum, genau dann wenn gilt
   1. $\Omega \neq \emptyset$
   2. $\mathcal{F}$ ist σ-Algebra über Ω
   3. P ist Wahrscheinlichkeitsmaß auf $\mathcal{F}$, also
   $P : \mathcal{F} \to [0, 1]$, $P(\emptyset) = 0$, $P(\Omega) = 1$.

b) Sei $(\Omega, \mathcal{F}, P)$ ein Wahrscheinlichkeitsraum und $(\Omega', \mathcal{F}')$ ein Messraum. Dann heißt eine $(\mathcal{F}, \mathcal{F}')$-messbare Funktion $X : \Omega \to \Omega'$ eine Ω'-Zufallsvariable auf Ω (bei reellen Zufallsvariablen spreche ich in Folge von Zufallsgrößen).

c) $F(t) : (P \leq t)$ heißt dann die zu X zugehörige Verteilungsfunktion.

Definition 1.0.2. (Erwartungswert, Varianz)

a) Ist $X \in \mathcal{L}^1(P)$, so heißt X integrierbar und der Erwartungswert von X ist definiert durch

$$\mathbb{E}[X] := \int_{-\infty}^{\infty} X dP.$$

Falls X eine reelle Zufallsvariable mit Wahrscheinlichkeitsdichte f ist, so gilt

$$\mathbb{E}[X] := \int_{-\infty}^{\infty} x f(x) dx,$$

wenn das Integral existiert.
Falls $\mathbb{E}[X] = 0$ gilt, dann heißt X zentriert.

b) Die Varianz einer Zufallsvariablen X ist ein Streumaß von X, also ein Maß für die Abweichung einer Zufallsvariablen X von ihrem Erwartungswert $\mathbb{E}[X]$. Sie ist definiert als

$$Var(X) := \mathbb{E}\left[(X - \mathbb{E}[X])^2\right].$$

Im Folgenden wollen wir einige Rechenregeln für den Erwartungswert und die Varianz erläutern, welche eine wichtige Grundlage für unsere Berechnungen der Stein'schen Methode darstellen.

Satz 1.0.3. (Rechenregeln für den Erwartungswert)
Seien X, Y, X_n, für $n \in \mathbb{N}$ reelle integrierbare Zufallsvariablen auf $(\Omega, \mathcal{A}, P)$. Dann gilt

a) Falls $P_X = P_Y$, so gilt $\mathbb{E}[X] = \mathbb{E}[Y]$.

b) Ist $X \geq 0$ fast sicher, so gilt

$$\mathbb{E}[X] = 0 \Leftrightarrow X = 0 \text{ fast sicher.}$$

c) (Linearität) Sei $c \in \mathbb{R}$. Dann gilt $cX \in \mathcal{L}^1(P)$ und $X + Y \in \mathcal{L}^1(P)$ sowie

$$\mathbb{E}[cX + Y] = c\mathbb{E}[X] + \mathbb{E}[Y].$$

d) (Monotonie) Falls $X \leq Y$ fast sicher gilt, folgt $\mathbb{E}[X] \leq \mathbb{E}[Y]$.

e) (Betragsungleichung) Es gilt $|\mathbb{E}[X]| \leq \mathbb{E}[|X|]$.

f) Ist $X_n \geq 0$ fast sicher für jedes $n \in N$, so gilt $\mathbb{E}\left[\sum\limits_{n=1}^{\infty} X_n\right] = \sum\limits_{n=1}^{\infty} \mathbb{E}[X_n]$.

g) (Multiplikationseigenschaft) Seien $X_1, ..., X_n \in \mathcal{L}^1(P)$ unabhängig, dann ist $(X_1 * ... * X_n) \in \mathcal{L}^1(P)$ und es gilt

$$\mathbb{E}\left[X_1 * ... * X_n\right] = \mathbb{E}[X_1] * ... * \mathbb{E}[X_n].$$

Beweis. a)-f) Auf diese Beweise wollen wir an diesen Stelle nicht explizit eingehen, da sie direkt aus der Definition des Erwartungswertes und den Integraleigenschaften folgen (vgl [7]).

g) Sei $P_{(X_1,...,X_n)}$ die gemeinsame Verteilung der X_i und P_{X_i} die Verteilung von X_i. Nach Voraussetzung gilt $P_{(X_1,...,X_n)} = P_{X_1} \otimes ... \otimes P_{X_n}$. Mit der Transformationsformel und dem Satz von Fubini erhalten wir für
$g(x_1, ..., x_n) = |x_1 * ... * x_n|$ zunächst

$$\begin{aligned}
\mathbb{E}\left[|X_1 * ... * X_n|\right] &= \int_{\mathbb{R}^n} g dP_{(X_1,...,X_n)} \\
&= \int_{\mathbb{R}^n} g d(P_{X_1} \otimes ... \otimes P_{X_n}) \\
&= \int_{\mathbb{R}} ... \int_{\mathbb{R}} |x_1 * ... * x_n| \, dP_{X_1} ... dP_{X_n} \\
&= \int_{\mathbb{R}} |x_1| * ... * \int_{\mathbb{R}} |x_n| \\
&= \mathbb{E}[|X_1|] * ... * \mathbb{E}[|X_n|] < \infty.
\end{aligned}$$

Somit gilt $(X_1 * ... * X_n) \in \mathcal{L}^1(P)$. Analog zeigt sich mit $g(x_1, ..., x_n) = x_1 * ... * x_n$, dass $E\left[X_1 * ... * X_n\right] = \mathbb{E}[X_1] * ... * \mathbb{E}[X_n]$ gilt.

□

Satz 1.0.4. (Rechenregeln für die Varianz)

a) Für die Varianz einer Zufallsvariablen X gilt

$$Var(X) := \mathbb{E}\left[(X - \mathbb{E}[X])^2\right] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2.$$

b) Sei X eine reelle integrierbare Zufallsvariable auf $(\Omega, \mathcal{A}, P)$ und $a, b \in \mathbb{R}$. Dann gilt

$$Var(aX + b) = a^2 Var(X).$$

c) $Var(-X) = Var(X)$

Beweis. a) Folgt nach dem Verschiebungssatz.

b)
$$\begin{aligned} Var(aX + b) &= \mathbb{E}\left[(aX + b - \mathbb{E}(aX + b))^2\right] \\ &= \mathbb{E}\left[(aX + b - b - a\mathbb{E}(X))^2\right] \\ &= \mathbb{E}\left[a^2(X - \mathbb{E}(X))^2\right] \\ &= a^2\mathbb{E}\left[(X - \mathbb{E}(X))^2\right] \\ &= a^2 Var(X) \end{aligned}$$

c) Folgt direkt aus (b) mit $a = -1$.

□

Im Folgenden werden wir einige Konvergenz- und Gleichheitsbegriffe für Zufallsvariablen betrachten.

Definition 1.0.5. (Gleichheitsbegriffe)

a) (Gleichheit in Verteilung)
Seien $X : (\Omega, \mathcal{A}, P) \to (\Omega', \mathcal{A}')$ und $Y : (\Omega_1, \mathcal{A}_1, P_1) \to (\Omega', \mathcal{A}')$ Zufallsvariablen. X und Y heißen in Verteilung gleich oder identisch verteilt, falls sie die gleiche Verteilung besitzen, d.h. wenn $P_X = P_Y$, bzw.

$$P_X(A') = P(X \in A') = P_1(Y \in A') = P_Y(A'), \qquad A' \in \mathcal{A}'.$$

(schreibe X $\stackrel{\text{d}}{=} Y$)

b) (fast sichere Gleichheit)
Seien $X, Y : (\Omega, \mathcal{A}, P) \to (\Omega', \mathcal{A}')$ Zufallsvariablen. Dann nennt man X und Y fast sicher gleich, falls

$$P(X = Y) = P(\omega \in \Omega : X(\omega) = Y(\omega)) = 1.$$

(schreibe X=Y f.s.)

Bemerkung 1.0.6. *Wenn* X=Y *f.s. gilt, so folgt* $X \stackrel{d}{=} Y$. *Die Umkehrung gilt im Allgemeinen nicht.*

Die Existenz verschiedener Gleiheitsbegriffe legt die Vermutung nahe, dass es auch verschiedene Konvergenzbegriffe gibt. Einige dieser Konvergenzbegriffe und deren Eigenschaften wollen wir nun genauer betrachten.

Definition 1.0.7. (Konvergenzbegriffe)
Seien $X_1, X_2, ...$ eine Folge von Zufallsvariablen definiert auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$. Dann sagt man

a) $X_n \to X$ fast sicher, falls

$$P\left\{\omega \in \Omega : X_n(\omega) \to X(\omega)\, n \to \infty\right\} = 1.$$

(schreibe $X_n \stackrel{\text{f.s.}}{\to} X$)

b) $X_n \to X$ in Wahrscheinlichkeit, falls

$$P(|X_n - X| > \epsilon) \to 0 \text{ ,für } n \to \infty.$$

(schreibe $X_n \stackrel{\text{P}}{\to} X$)

c) $X_n \to X$ in Verteilung, falls

$$P(X_n \leq x) \to P(X \leq x), \text{ für } n \to \infty$$

an allen Punkten $x \in \mathbb{R}$, an denen $F_X(x) = P(X \leq x)$ stetig ist.
(schreibe $X_n \stackrel{\text{d}}{\to} X$)

Satz 1.0.8. *Zwischen den Konvergenzarten gelten die folgenden Zusammenhänge*

- $X_n \stackrel{f.s.}{\to} X \Rightarrow X_n \stackrel{P}{\to} X \Rightarrow X_n \stackrel{d}{\to} X.$
- *Falls* $X_n \stackrel{d}{\to} c$, $c \in \mathbb{R}$ *dann gilt auch* $X_n \stackrel{P}{\to} c.$

Satz 1.0.9. *Sei* $X_n \stackrel{d}{\to} X$ *und* $f : \mathbb{R} \to \mathbb{R}$ *stetig. Dann gilt*

$$f(X_n) \stackrel{d}{\to} f(X).$$

Beweis. Nach dem Repräsentationssatz von Skorokhod existiert ein Wahrscheinlichkeitsraum $(\Omega', \mathcal{A}', P)$ und Zufallsvariablen Y_n, Y ,so dass Y_n, Y die selben Verteilungsfunktion besitzen wie X_n, X, und es gilt $Y_n \stackrel{\text{f.s.}}{\to} Y$. Wegen der Stetigkeit von f gilt nun $\omega : f(Y_n(\omega)) \to f(Y(\omega)) \supseteq \omega : Y_n(\omega) \to Y(\omega)$. Damit folgt $f(Y_n) \stackrel{\text{f.s.}}{\to} f(Y)$ und somit nach dem vorangegangenen Satz $f(Y_n) \stackrel{\text{d}}{\to} f(Y)$. Da aber $f(Y_n), f(Y)$ die gleiche Verteilung besitzen wie $f(X_n), f(X)$ folgt die Behauptung. □

An dieser Stelle betrachten wir noch einen wichtigen Satz, der unsere getroffenen Definitionen in ein Verhältnis setzt.

Lemma 1.0.10. (Portmanteau) Für beliebige Zufallsvektoren sind die folgenden Aussagen äquivalent:

1. $P(X_n \leq x) \stackrel{\text{d}}{\to} P(X \leq x)$ für alle Stetigkeitspunkte von $x \mapsto P(X \leq x)$,
2. $\mathbb{E}\left[f(X_n)\right] \stackrel{\text{d}}{\to} \mathbb{E}\left[f(X)\right]$ für alle beschränkten, stetigen Funktionen f,
3. $\mathbb{E}\left[f(X_n)\right] \stackrel{\text{d}}{\to} \mathbb{E}\left[f(X)\right]$ für alle beschränkten, Lipschitz-stetigen Funktionen f,
4. $\underset{x \to \infty}{lim} \, inf \, \mathbb{E}\left[f(X_n)\right] \geq \mathbb{E}\left[f(X)\right]$ für alle nicht negative, stetige Funktionen f,
5. $\underset{x \to \infty}{lim} \, inf \, P(X_n \in G) \geq P(X \in G)$ für alle offenen Mengen G,
6. $\underset{x \to \infty}{lim} \, sup \, P(X_n \in F) \geq P(X \in F)$ für alle abgeschlossenen Mengen F,
7. $P(X_n \in B \stackrel{\text{d}}{\to} P(X \in B)$ für alle Borelleangen B mit $P(X \in \delta B) = 0$.

Beweis. Der Beweis dieses Satzes kann z.B. in [11] nachgelesen werden. □

Nun wollen wir zum Abschluss dieses Kapitels den Zentralen Grenzwertsatz formulieren und kurz auf die meißt verwandten Beweismethoden eingehen.

Satz 1.0.11. (Zentraler Grenzwertsatz) Seien $X_1, X_2, X_3, \ldots$ unabhängig, identisch verteilte Zufallsgrößen mit

$$\mathbb{E}[X_1] = 0, \qquad E[X_1^2] = \sigma^2, \qquad S_n := \sum_{i=1}^{n} X_i, \qquad W_n := \tfrac{1}{\sigma\sqrt{n}} S_n.$$

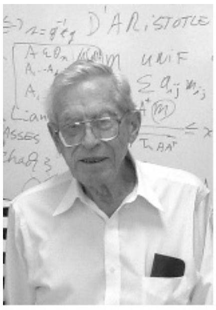
Dann gilt

$$P(W_n \leq t) \xrightarrow{\mathrm{d}} \Phi(t) := \tfrac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-x^2/2} dx.$$

Bemerkung 1.0.12. *Zum Beweis dieses Satzes gibt es drei klassische Methoden. Die vermutlich am häufigsten benutzte Methode verwendet die charakteristische Funktion mit Hilfe der Fourier-Transformierten zum Beweis des Zentralen Grenzwertsatzes (vgl. [8]). Der Vorteil dieser Methode ist, dass sie auch auf andere Verteilungen als die Normalverteilung übertragbar ist, sowie auf den Fall, dass die X_i nicht identisch verteilt sind. Eine andere Methode ist die 1887 erstmals von Tschebychev verwendete Momentenmethode (vgl. [5]). Sie gilt als relativ schwierig und wird daher nicht so häufig angewandt, zudem liefert sie keine Konvergenzraten. Die dritte klassische Methode ist die von Lindeberg entwickelte und nach ihm benannte Lindeberg-Methode (vgl. [1]), auch sie hat wie die Methode der charakteristischen Funktion den Vorteil, dass die identische Verteilung der X_i zum Beweis der Zentralen Grenzwertsatzes nicht notwendig ist - stattdessen wird die so genannte Lindeberg-Bedingung gefordert. Wir wollen den Zentralen Grenzwertsatz an dieser Stelle nicht explizit mit einer der klassischen Methoden beweisen, sondern im nächsten Kapitel mit Hilfe der Stein'schen Methode.*

2 Grundlagen der Stein'schen Methode

Kommen wir nun zur Stein'schen Methode. Diese Methode zum Messen von Abständen zwischen Wahrscheinlichkeitsverteilungen stammt von dem US-amerikanischen Statistiker Charles M. Stein (*22.März 1920) aus dem Jahr 1972. Da Stein sich mit den bis zum Ende der 1960iger Jahre bekannten Beweismethoden zum Zentralen Grenzwertastz nicht „anfreunden“ konnte, entwickelte er die nach ihm benannte Methode, als er diesen Satz für einen Statistik-Vortrag beweisen wollte. Seine bahnbrechende Arbeit [10] wurde bei dem sechsten Berkeley Symposium vorgestellt und später veröffentlicht.
Louis H. Y. Chen, einem Studenten von Stein, entwickelt diese Methode in den folgenden Jahren weiter; deshalb wird die Methode teilweise auch Chen-Stein-Methode genannt.



Charles Stein

Louis H. Y. Chen

Wenn wir eine differenzierbare Funktion f über einem kompaktem Träger in (a, b) gegeben haben, lässt sich mittels partiellen Integration ($f = f(x)$, $g' = x * e^{-x^2/2}$) leicht nachrechnen, dass

$$\int_a^b xf(x)e^{-x^2/2}dx = -f(x)e^{-x^2/2}\Big|_a^b + \int_a^b f'(x)e^{-x^2/2}dx$$

gilt. Hierbei ist zu beachten, dass der linke Teil der rechten Seite für große a, b gegen Null konvergiert.
Da für $Z \sim \mathcal{N}(0,1)$ und $f : \mathbb{R} \to \mathbb{R}$ differenzierbar mit $\mathbb{E}|f'(Z)| < \infty$

$$\mathbb{E}[Zf(Z)] = \int_{-\infty}^{\infty} \frac{e^{-x^2/2}}{\sqrt{2\pi}} xf(x)dx \quad \text{und} \quad \mathbb{E}[Zf'(Z)] = \int_{-\infty}^{\infty} \frac{e^{-x^2/2}}{\sqrt{2\pi}} f'(x)dx$$

gilt, folgt aus der obrigen Integralgleichung

$$\mathbb{E}[Zf(Z)] = \mathbb{E}[f'(Z)].$$

Genauer kann man sogar zeigen:

Satz 2.0.13. (Stein, 1972)
Sei Z eine Zufallsgröße mit Werten in $\mathbb{R}$, dann gilt
Z ist genau dann Standardnormalverteilt, wenn für jede stückweise stetig differenzierbare Funktion $f : \mathbb{R} \to \mathbb{R}$ mit $\int_{\mathbb{R}} |f'(x)| \, e^{-x^2/2} dx < \infty$ gilt, dass $\mathbb{E}\{f'(Z) - Zf(Z)\} = 0$ ist.

Beweis. Um den Beweis dieses Satzes wollen wir uns am Ende dieses Kapitels kümmern. □

Im weiteren Verlauf wollen wir nun klären, wie weit die $W_n := \frac{1}{\sigma\sqrt{n}} S_n$ von der standardnormalverteilten Zufallsgröße Z abweichen. Hierzu geben wir zunächst einen kleinen Exkurs zum Thema Messen von Abständen zwischen Wahrscheinlichkeitsmaßen:
Es seien μ, ν zwei Wahrscheinlichkeitsmaße auf $\mathbb{R}$. Dann definieren wir für eine Funktionenklasse $\mathcal{D}$ und eine Metrik d den Abstand zwischen den Wahrscheinlichkeitsmaßen μ und ν als

$$d_{\mathcal{D}} := sup\{\left|\int f d\mu - \int f d\nu\right| : f \in \mathcal{D}\}.$$

Betrachten wir nun drei wichtige Beispiele solcher Funktionenklassen $\mathcal{D}$ und die daraus entstehenden Abstände.

Beispiel 2.0.14. *1. Totalvariation:*

$$\mathcal{D} = \{1_A : A \in \mathcal{B}\}$$
$$TV(\mu,\nu) := sup_{A\in\mathcal{B}} |\mu(A) - \nu(A)|$$

2. Wasserstein-Metrik:

$$\mathcal{D} = \{f : f \text{ ist 1-Lipschitz}\}$$
$$wass(\mu,\nu) := sup\{\left|\int f d\mu - \int f d\nu\right| : f \text{ ist 1-Lipschitz}\}$$

3. Kolmogorov-Abstand:

$$\mathcal{D} = \left\{1_{(-\infty,x]} : x \in \mathbb{R}\right\}$$
$$kolm(\mu,\nu) := sup_{x\in\mathbb{R}} |\mu(-\infty,x) - \nu(-\infty,x)|$$

Wobei

- $kolm(\mu,\nu) \leq TV(\mu,\nu)$
- *alle drei Metriken sind stärker als die schwache Konvergenz*

gilt.

Eine klassische Aussage über die Güte der Konvergenz im Zentralen Grenzwertsatz trifft der folgende Satz von Berry-Esseen, welchen die beiden Mathematikern A. C. Berry (1941) und C.-G. Essen (1942) unabhängig voneinander gezeigt haben.

Satz 2.0.15. (Berry-Esseen)
Seien $(X_i)_{i\in\mathbb{N}}$ unabhägig, identisch verteilte Zufallsvariablen mit $\mathbb{E}(X_i) = 0, \mathbb{E}(X_i^2) = 1$ und $\mathbb{E}(X_1)^3 \leq \infty$, dann gilt

$$kolm(W_n, Z) \leq \frac{C}{\sqrt{n}}$$

Beweis. Den klassischen Beweis bringen wir an dieser Stellen nicht; er ist z.B. nachzulesen in [9].
In Kapitel 5 werden wir die Aussage von Berry-Esseen mit Hilfe der Stein'schen Methode beweisen. □

Im weiteren Verlauf unserer Betrachtungen wollen wir die Stein'sche Charakterisierung der Standardnormalverteilung nach Satz 2.0.13 nutzen, um einen Zentralen Grenzwertsatz zu beweisen und eine Konvergenzrate herzuleiten.
Hierzu sei W eine Zufallsvariable, welche gut durch $Z \sim \mathcal{N}(0,1)$ approximiert werden kann. Wir wollen nun den Abstand

$$sup_{g \in \mathcal{D}} |\mathbb{E}[g(W)] - \mathbb{E}[g(Z)]|$$

betrachten, und diesen mit

$$sup_{f \in \mathcal{D}'} |\mathbb{E}[f'(W)] - W\mathbb{E}[f(W)]|$$

für eine geeignete Funktionenklasse $\mathcal{D}'$ in Verbindung bringen, wobei der zweite Ausdruck möglichst kein seien sollte.
In seinem Ansatz wählt Stein $\mathcal{D}'$ so, dass $\mathcal{D}'$ die Klasse von Funktionen ist, so dass für jedes $g \in \mathcal{D}$ ein $f := f_g \in \mathcal{D}'$ existiert, welches die sogenannte Stein-Gleichung mit Inputfunktion g

$$f'(x) - xf(x) = g(x) - \mathbb{E}[g(Z)]$$

erfüllt.
Wenn wir nun die so gegebene Stein-Gleichung bezüglich der Verteilung von W integrieren folgt

$$\mathbb{E}\left[f(W) - Wf(W)\right] = \mathbb{E}[g(W)] - \mathbb{E}[g(Z)]$$

und somit

$$sup_{f \in \mathcal{D}'} |\mathbb{E}\left[f'(W) - Wf(W)\right]| \geq sup_{g \in \mathcal{D}} |\mathbb{E}[g(W)] - \mathbb{E}[g(Z)]|$$

wobei $\mathcal{D}'$ möglichst klein seien sollte. f, f' sind dann die Lösung der Stein-Gleichung sowie ihre Ableitung und die Zufallsvariable W enthält das stochastische Modell.
Nun wollen wir einige Eigenschaften der Stein-Gleichung und ihrer Lösung f diskutieren.

Lemma 2.0.16. *Sei $g : \mathbb{R} \to \mathbb{R}$ mit $\mathbb{E}\,|g(Z)| < \infty$ gegeben, dann ist*

$$f(x) = e^{x^2/2} \int_{-\infty}^{x} e^{-y^2/2}(g(y) - \mathbb{E}[g(Z)])dy$$

eine fast sicher differenzierbare Lösung der Stein-Gleichung mit Input-Funktion g. Weiter Lösungen sind von der Form $f^*(x) = f(x) + ce^{x^2/2}$.

Beweis. Sei f eine Lösung, dann zeigt sich durch differenzieren, dass

$$\tfrac{d}{dx}\left(e^{-x^2/2}f(x)\right) \overset{\text{P.I.}}{=} (e^{-x^2/2}(f'(x) - xf(x)) = e^{-x^2/2}\,(g(x) - \mathbb{E}[g(z)])$$

gilt, und somit $f(x) = e^{x^2/2}\int_{-\infty}^{x} e^{-y/2}(g(y) - \mathbb{E}[g(Z)])dy$ eine Lösung ist. Die f.s. Differenzierbarkeit folgt weil f Komposition von f.s. differenzierbarer Funktionen ist, da $g(x) - \mathbb{E}[g(Z)] = f'(x) - xf(x)$ und f nach Satz 2.0.13 stückweise stetig differenzierbar ist. □

Für die Lösung der Stein-Gleichung gilt der folgende Satz.

Satz 2.0.17. (Stein 1986)

1. Sei $g : \mathbb{R} \to \mathbb{R}$ beschränkt, dann existiert eine fast-sicher differenzierbare Funktion $f : \mathbb{R} \to \mathbb{R}$ mit $f'(x) - xf(x) = g(x) - \mathbb{E}g(Z)\ \forall x$ und es gilt
   - $|f|_\infty \leq \sqrt{\frac{2}{\pi}}\,|g - \mathbb{E}[g(Z)]|_\infty$
   - $|f'|_\infty \leq 2\,|g - \mathbb{E}[g(Z)]|_\infty$
   - $|f'|_\infty \leq 2\,|g'|_\infty$
2. Sei $g : \mathbb{R} \to \mathbb{R}$ LIPSCHITZ, nicht notwendigerweise beschränkt, so gilt
   - $|f|_\infty \leq |g'|_\infty$
   - $|f'|_\infty \leq \sqrt{\frac{2}{\pi}}\,|g'|_\infty$
   - $|f''|_\infty \leq 2\,|g'|_\infty$

Bemerkung 2.0.18.
- *Meistens wird die dritte Aussage des ersten Falls weggelassen, weil sie keine aussagekräftige obere Schranke darstellt, da $|g'|$ je nach Ausgangsfunktion g beliebig groß werden kann, weil aus der Beschränktheit*

von g nicht notwendigerweise folgt, dass auch g' beschränkt ist. Deshalb wird häufig L-Stetigkeit gefordert, denn die Ableitung einer L-stetigen Funktion ist per Definition beschränkt.

- *Die ersten beiden Schranken für den Fall der L-stetigen Inputfunktion g gehen nicht direkt auf C. Stein zurück. In der Literatur (z.B [3]) findet man meistens die Schranke $|f|_\infty \leq 2\,|g'|_\infty$.*

Beweis. (zu Satz2.0.17)

1. Die Existenz der fast sicher differenzierbaren Funktion f folgt direkt aus dem vorherigem Lemma.

   - Nach Definition von $\mathbb{E}\left[g(Z)\right]$ gilt
     $$\int_{\mathbb{R}} e^{-y^2/2}\left(g(y) - \mathbb{E}\left[g(Z)\right]\right) = 0,$$
     und somit
     $$\begin{aligned} f(x) &= e^{x^2/2}\int_{-\infty}^{x} e^{-y^2/2}(g(y) - \mathbb{E}\left[g(Z)\right])dy \\ &= -e^{x^2/2}\int_{x}^{\infty} e^{-y^2/2}(g(y) - \mathbb{E}\left[g(Z)\right])dy \end{aligned}$$
     Für $x > 0$ gilt dann die Abschätzung
     $$|f(x)| \leq |g - \mathbb{E}\left[g(Z)\right]|_\infty \underbrace{\left(e^{x^2/2}\int_{x}^{\infty} e^{-y^2/2}dy\right)}_{(*)}.$$
     Um (*) abzuschätzen betrachten wir die Ableitung dieses Ausdrucks. Es gilt
     $$\tfrac{d}{dx}\left(e^{x^2/2}\int_{x}^{\infty} e^{-y^2/2}dy\right) = -1 + xe^{x^2/2}\int_{x}^{\infty} e^{-y^2/2}dy \leq 0, \qquad (**)$$
     wobei wir im letzten Schritt die sogenannte MILL'S ratio
     $$\int_{x}^{\infty} e^{-y^2/2}dy \overset{x>0}{\leq} \int_{x}^{\infty} \frac{y}{x}e^{-y^2/2}dy = \frac{1}{x}e^{-x^2/2}$$
     verwendet haben.
     Somit wird (*) auf $[0,\infty)$ in $x = 0$ maximiert und nimmt dort den Wert $\sqrt{\frac{\pi}{2}}$ an. Für den Fall $x < 0$ wird f analog abgeschätzt.

- Betrachten wir nun die Abschätzung für die Ableitung der Lösung der Stein-Gleichung. Sei zunächst wieder $x > 0$, dann gilt
$$\begin{aligned} f'(x) &= xf(x) + g(x) - \mathbb{E}\left[g(Z)\right] \\ &= g(x) - \mathbb{E}\left[g(Z)\right] - xe^{x^2/2} \int_x^\infty e^{-y^2/2} \left(g(y) - \mathbb{E}\left[g(Z)\right]\right) dy \end{aligned}$$
und somit wieder unter Verwendung der MILL'S ratio die Abschätzung
$$|f'| \leq |g - \mathbb{E}\left[g(Z)\right]|_\infty \left(1 + xe^{x^2/2} \int_x^\infty e^{-y^2/2} dy\right) \overset{(**)}{\leq} 2\left|g - \mathbb{E}\left[g(Z)\right]\right|_\infty.$$
Die selbe Abschätzung zeigt man analog für $x > 0$ und somit folgt die Behauptung.
- Diesen Teil werden wir aus den in Bemerkung 2.0.18 genannten Gründen nicht expliziet ausführen.

2. Zunächst wollen wir vorbereitend zeigen , dass für g L-stetig und $Z \sim N(0,1)$
$$f(x) = -\int_0^1 \frac{1}{2\sqrt{t(1-t)}} \mathbb{E}\left[zg(\sqrt{t}x + \sqrt{1-t}Z)\right] dt \qquad (1)$$
eine Lösung der Stein-Gleichung ist und gleich der Lösung von Lemma 2.0.16 ist.
In der Tat, falls g C-Lipschitz ist, gilt $|g'|_\infty \leq C$. Wenn wir f ableiten und dann die Ableitung in das Integral ziehen erhalten wir
$$f'(x) = -\int_0^1 \frac{1}{2\sqrt{1-t}} \mathbb{E}\left[Zg'(\sqrt{t}x + \sqrt{1-t}Z)\right] dt. \qquad (2)$$
Andererseits gilt nach der Stein'schen Identität
$$\mathbb{E}\left[Zg(\sqrt{t}x + \sqrt{1-t}Z)\right] = \sqrt{1-t}\,\mathbb{E}\left[Zg'(\sqrt{t}x + \sqrt{1-t}Z)\right].$$
Somit gilt nun zusammenfassend

$$
\begin{aligned}
f'(x) - xf(x) &= \int_0^1 \mathbb{E}\left[\left(-\frac{Z}{2\sqrt{1-t}} + \frac{x}{2\sqrt{t}}\right) g'\left(\sqrt{t}x + \sqrt{1-t}Z\right)\right] dt \\
&= \int_0^1 \mathbb{E}\left[\frac{d}{dt} g'\left(\sqrt{t}x + \sqrt{1-t}Z\right)\right] dt \\
&= \mathbb{E}\left[\int_0^1 \frac{d}{dt} g'\left(\sqrt{t}x + \sqrt{1-t}Z\right)\right] dt \\
&= g(x) - \mathbb{E}g(Z),
\end{aligned}
$$

womit wir die oben stehende Aussage gezeigt haben.

- Kommen wir nun zur ersten Abschätzung $|f|_\infty \leq |g'|_\infty$
  Nach der gerade gezeigten Aussage (1) folgt mit
  $$\mathbb{E}\left[Zg(\sqrt{t}x + \sqrt{1-t}Z)\right] = \sqrt{1-t}\mathbb{E}\left[Zg'(\sqrt{t}x + \sqrt{1-t}Z)\right],$$
  dass
  $$f(x) = -\int_0^1 \frac{1}{2\sqrt{t}} \mathbb{E}\left[g'(\sqrt{t}x + \sqrt{1-t}Z)\right] dt$$
  gilt und somit folgt
  $$
  \begin{aligned}
  |f|_\infty &\leq |g'|_\infty \int_0^1 \frac{1}{2\sqrt{t}} \\
  &= |g'|_\infty \left[\sqrt{t}\right]_0^1 \\
  &= |g'|_\infty.
  \end{aligned}
  $$
- Kommen wir nun zu der Abschätzung für die erste Ableitung. Nach (2) folgt, dass
  $$
  \begin{aligned}
  |f'|_\infty &\leq \mathbb{E}\left[|Z|\right] |g'|_\infty \int_0^1 \frac{1}{2\sqrt{1-t}} \\
  &= \mathbb{E}\left[|Z|\right] |g'|_\infty \left[-\sqrt{1-t}\right]_0^1 \\
  &= \sqrt{\frac{2}{\pi}} |g'|_\infty.
  \end{aligned}
  $$
- Für die Abschätzung der zweiten Ableitung müssen wir zunächst einige Berechnungen voranstellen. Durch ableiten der Stein-Gleichung erhalten wir

$$\begin{aligned}
f''(x) &= g'(x) + f(x) + xf'(x) \\
&= g'(x) + f(x) + x\left(g(x) - \mathbb{E}[g(Z)] + xf(x)\right) \\
&= g'(x) + x(g(x) - \mathbb{E}[g(Z)]) + (1 + x^2)f(x). \qquad (3)
\end{aligned}$$

Um diesen Ausdruck weiter abschätzen zu können wollen wir nun die beiden Terme $g(x) - \mathbb{E}[g(Z)]$ und $f(x)$ so umformen, dass sie nur noch von g' abhängen. Für den ersten Term gilt

$$\begin{aligned}
&g(x) - \mathbb{E}[g(Z)] \\
&= \frac{1}{\sqrt{2\pi}} \int_0^1 e^{-y^2/2}(g(x) - g(y))dy \\
&= \frac{1}{\sqrt{2\pi}} \left[\int_{-\infty}^x \int_y^x g'(z)e^{-y^2/2}dzdy - \int_x^\infty \int_x^y g'(z)e^{-y^2/2}dzdy\right] \\
&= \frac{1}{\sqrt{2\pi}} \left[\int_{-\infty}^x g'(z) \int_y^x e^{-y^2/2}dydz - \int_x^\infty g'(z) \int_x^y e^{-y^2/2}dydz\right] \\
&= \int_{-\infty}^x g'(z)\Phi(z)dz - \int_x^\infty g'(z)(1 - \Phi(z))dz,
\end{aligned}$$

wobei $\Phi(z)$ die Verteilungsfunktion der Standardnormalverteilung darstellt. Für $f(x)$ gilt auf analoge Weise

$$\begin{aligned}
f(x) &= e^{x^2/2} \int_{-\infty}^x e^{y^2/2}(g(y) - \mathbb{E}[g(z)])dy \\
&= e^{x^2/2} \int_{-\infty}^x e^{y^2/2} \left(\int_{-\infty}^x g'(z)\Phi(z)dz - \int_y^\infty g'(z)(1 - \Phi(z))\right) dzdy \\
&= e^{x^2/2} \left(\int_{-\infty}^x g'(z)\Phi(z) \int_z^x e^{y^2/2}dydz - \right. \\
&\qquad \left. \int_{-\infty}^\infty g'(z)(1 - \Phi(z)) \int_{-\infty}^{max\{z,x\}} e^{y^2/2}dydz\right) \\
&= \sqrt{2\pi}e^{x^2/2} \left(\int_{-\infty}^x g'(z)\Phi(z)\left((1 - \Phi(z)) - (1 - \Phi(x))\right) dz - \right. \\
&\qquad \left. \int_{-\infty}^x g'(z)(1 - \Phi(z))\Phi(z)dz - \int_x^\infty g'(z)(1 - \Phi(z))\Phi(z)dz\right)
\end{aligned}$$

$$
\begin{aligned}
&= -\sqrt{2\pi}e^{x^2/2}\left[(1-\Phi(x))\int_{-\infty}^{x} g'(z)\Phi(z)dz\right.\\
&\qquad \left.+\Phi(x)\int_{x}^{\infty} g'(z)(1-\Phi(z))dz\right].
\end{aligned}
$$

Wenn wir diese beiden Umformungen nun in (3) einsetzen erhalten wir

$$
\begin{aligned}
f''(x) =& g'(x) + \left(x - \sqrt{2\pi}(1+x^2)e^{x^2/2}(1-\Phi(x))\right)\int_{-\infty}^{x} g'(z)\Phi(z)dz\\
&+ \left(-x - \sqrt{2\pi}(1+x^2)e^{x^2/2}(1-\Phi(x))\right)\int_{x}^{\infty} g'(z)(1-\Phi(z))dz,
\end{aligned}
$$

was uns zu der folgenden Abschätzung führt

$$
\begin{aligned}
&|f''(x)|_\infty\\
&= |g'|_\infty + \left[1 + \left|x - \sqrt{2\pi}(1+x^2)e^{x^2/2}(1-\Phi(x))\right|\int_{-\infty}^{x}\Phi(z)dz\right.\\
&+ \left.\left|-x - \sqrt{2\pi}(1+x^2)e^{x^2/2}(1-\Phi(x))\right|\int_{x}^{\infty}(1-\Phi(z))dz\right]\\
&\leq |g'|_\infty + \left[1 + \left(-x\sqrt{2\pi}(1+x^2)e^{x^2/2}(1-\Phi(x))\right)\left(x\Phi(x) + \frac{e^{-x^2/2}}{\sqrt{2\pi}}\right)\right.\\
&+ \left.\left(x + \sqrt{2\pi}(1+x^2)e^{x^2/2}\Phi(x)\right)\left(-x(1-\Phi(x)) + \frac{e^{-x^2/2}}{\sqrt{2\pi}}\right)\right]\\
&= 2|g'|_\infty.
\end{aligned}
$$

Wobei wir im vorletzten Schritt wieder die Mills Ratio genutzt haben, nach welcher für $x > 0$ gilt

$$
\tfrac{xe^{x^2/2}}{\sqrt{2\pi}(1+x^2)} \leq 1-\Phi(x) \leq \tfrac{e^{x^2/2}}{x\sqrt{2\pi}}.
$$

Eine analoge Abschätzung gibt es auch für $x \geq 0$. Womit

$$
x + \sqrt{2\pi}(1+x^2)e^{x^2/2}\Phi(x) > 0
$$

und

$$
-x + \sqrt{2\pi}(1+x^2)e^{x^2/2}(1-\Phi(x)) > 0
$$

gilt. Zu guter Letzt haben wir bei der Umformung ausgenutzt, dass duch Integration

$$
\int_{-\infty}^{x}\Phi(z)dz = x\Phi(x) + \frac{e^{-x^2/2}}{\sqrt{2\pi}}
$$

und

$$\int_x^\infty (1-\Phi(z))dz = x(1-\Phi(x)) + \frac{e^{-x^2/2}}{\sqrt{2\pi}}$$

gilt.

□

Bemerkung 2.0.19. *Wesentlich einfacher zeigt man für den letzten Fall die Abschätzung $|f''|_\infty \leq 4|g'|_\infty$. Hierzu leitet man die Stein Gleichung ab, was durch Umformung und zu Hilfenahme der Schranke für f zusammen mit der Dreiecksungleichung zu $|g'+f| \leq |g'| + |f| \leq 2|g'|_\infty$ und schließlich zu $|f''|_\infty \leq 2|g'+f|_\infty \leq 4|g'|_\infty$ führt.*

Somit haben wir bis zu dieser Stelle

$$wass(W,Z) \leq sup\left\{|\mathbb{E}[f'(W) - Wf(W)]| : |f|_\infty \leq 1, |f'|_\infty \leq \sqrt{\tfrac{2}{\pi}}, |f''|_\infty \leq 2\right\}$$

gezeigt.
In der Praxis wird sich zeigen, dass es häufig relativ einfach ist die Lösung f_g zu einer gegebenen Input-Funktion g zu finden, und sich $|\mathbb{E}f'(W) - W\mathbb{E}f(W)|$ somit oft genau abschätzen lässt. Ein weiterer großer Vorteil der Stein'schen Methode ist, dass sie für eine Vielzahl von Metriken mit geringem Aufwand die Konvergenzraten liefert.
Kommen wir nun am Ende dieses Kapitels zu dem Beweis der Stein'schen Charakterisierung der Standardnormalverteilung.

Beweis. (zu Satz 2.0.17.)
„$\Rightarrow$" Sei Z standardnormalverteilt und $\int_\mathbb{R} |f'(x)|\, e^{-x^2/2}dx < \infty$. Dann gilt, wie wir bereits vorbereitend zu Satz 2.0.4. gezeigt haben

$$\mathbb{E}[Zf(Z)] = \mathbb{E}[f'(Z)]$$

„$\Leftarrow$" Gelte nun für jede stetige und stückweise stetig differenzierbare Funktion f mit $\int_\mathbb{R} |f'(x)|\, e^{-x^2/2}dx < \infty$, dass $\mathbb{E}\left[f'(Z) - Zf(Z)\right] = 0$. So gilt dies insbesondere für

$$f_\omega(y) := e^{y^2/2} \int_{-\infty}^y \left(g_\omega(x) - \mathbb{E}[g_\omega(Z)]\right) e^{-x^2/2}dx$$

mit

$$g_\omega(x) = \begin{cases} 1, & falls\, x \leq \omega, \\ 0, & falls\, x > \omega \end{cases}$$

und

$$\mathbb{E}\left[g_\omega(Z)\right] := \tfrac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} g_\omega e^{-x^2/2} dx = \Phi(\omega).$$

Nach Ableitungsregeln gilt

$$\begin{aligned} f_\omega'(y) &= e^{y^2/2} \left(g_\omega(y) - \mathbb{E}\left[g_\omega(Z)\right]\right) e^{-y^2/2} \\ &\quad + y e^{y^2/2} \int_{-\infty}^{y} \left(g_\omega(x) - \mathbb{E}\left[g_\omega(Z)\right]\right) e^{-x^2/2} dx \\ &= g_\omega(y) - \mathbb{E}\left[g_\omega(Z)\right] + y f_\omega(y). \end{aligned}$$

Also löst f_ω die Differentialgleichung $f'(y) - yf(y) = g_\omega(y) - \Phi(\omega)$. Somit ergibt sich

$$\begin{aligned} 0 &= \mathbb{E}\left[f_\omega'(Z) - Z f_\omega(Z)\right] \\ &= \mathbb{E}\left[g_\omega(Z) - \mathbb{E}\left[g_\omega(Z)\right]\right] \\ &= P\left(Z \leq \omega\right) - \Phi(\omega). \end{aligned}$$

Abschließend ist noch zu zeigen, dass f_ω stetig und stückweise stetig differenzierbar ist und $\int_{\mathbb{R}} \left|f'(x)\right| e^{-x^2/2} dx < \infty$ gilt. Da f_ω' existiert, ist f_ω stetig. Die stückweise stetige Differenzierbarkeit von f_ω gilt, da f stückweise stetig differenzierbar ist, was aus selbiger Eigenschaft von $g(y) - \mathbb{E}\left[g(Z)\right]$ folgt.
Zur Beendigung des Beweises müssen wir noch zeigen, dass

$$\tfrac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \left|f'(x)\right| e^{-x^2/2} dx < \infty$$

gilt.
Da nach Voraussetzung $\int_{\mathbb{R}} \left|g(x) - \mathbb{E}\left[g(Z)\right]\right| e^{-x^2/2} dx < \infty$ gilt, folgt mit der Definition von $\mathbb{E}[g(Z)]$, dass $\int_{\mathbb{R}} \left(g(x) - \mathbb{E}\left[g(Z)\right]\right) e^{-x^2/2} dx = 0$ gilt. Somit ist

$$\begin{aligned} f(\omega) &= e^{\omega^2/2} \int_{-\infty}^{\omega} \left(g(x) - \mathbb{E}\left[g(Z)\right]\right) e^{-x^2/2} dx \\ &= e^{-\omega^2/2} \int_{\omega}^{-\infty} \left(g(x) - \mathbb{E}\left[g(Z)\right]\right) e^{-x^2/2} dx. \end{aligned}$$

Partielle Integration liefert schließlich

$$\begin{aligned}\int_0^\infty |xf(x)|\, e^{-x^2/2}dx &\leq \int_0^\infty x \left(\int_0^\infty |g(x) - \mathbb{E}\left[g(Z)\right]|\, e^{-y^2/2}dy \right) dx \\ &= \int_0^\infty |g(x) - \mathbb{E}\left[g(Z)\right]|\, e^{-y^2/2}\frac{y^2}{2}dy. \qquad (*)\end{aligned}$$

Wobei (*) für $g = g_\omega$ endlich ist. Analog lässt sich das Integral über ganz $\mathbb{R}$ abschätzen und somit folgt die Behauptung. $\square$

3 Erste Anwendung für Summen unabhängiger Zufallsvariablen

Im kommenden Kapitel wollen wir $|\mathbb{E}[f'(W)] - W\mathbb{E}[f(W)]|$ für den Fall einer Partialsumme von unabhängig verteilten Zufallsvariablen abschätzen.

Satz 3.0.20. *Seien $(X_i)_{i\in\mathbb{N}}$ unabhängige Zufallsgrößen mit $\mathbb{E}(X_i) = 0$, $\mathbb{E}(X_i^2) = 1$ und existiere das absolute dritte Moment, dann gilt:*

$$wass(W_n, Z) \leq \tfrac{3}{n^{3/2}} \sum_{i=1}^{n} \mathbb{E}|X_i|^3,$$

falls die $(X_i)_{i\in\mathbb{N}}$ identisch verteilt sind folgt wegen $\sum\limits_{i=1}^{n} \mathbb{E}|X_i| = n\mathbb{E}|X_1|$, dass $wass(W_n, Z) \leq \frac{3\mathbb{E}|X_1|^3}{\sqrt{n}}$ gilt.

Beweis. Nach Voraussetzung gilt $\mathbb{E}(X_i) = 0$, $\mathbb{E}(X_i^2) = 1$, $\mathbb{E}|X_i| < \infty$. Wähle W_n wieder als

$$W_n := \tfrac{1}{\sqrt{n}} S_n = \tfrac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i$$

Des Weiteren sei f der Form, dass $|f|_\infty \leq 1, |f'|_\infty \leq \sqrt{\frac{2}{\pi}}, |f''|_\infty \leq 2$ erfüllt ist.
Dann gilt

$$\mathbb{E}[Wf(W)] = \tfrac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathbb{E}[X_i f(W)].$$

Wenn wir nun $W_i := W - \frac{X_i}{\sqrt{n}}$ definieren, gilt aufgrund der Unabhängigkeit der $(X_i)_{i\in\mathbb{N}}$, dass X_i und W_i für jedes feste i unabhängig sind und somit

$$\mathbb{E}\left[\left(W_i - \tfrac{X_i}{\sqrt{n}}\right) f(W_i)\right] = \mathbb{E}[X_i f(W_i)] = \mathbb{E}[X_i]\mathbb{E}[f(W_i)] \underset{\text{zent.}}{\overset{X_i}{=}} 0.$$

Mit zweifacher Nullergänzung gilt dann

$$
\begin{aligned}
\mathbb{E}[X_i f(W)] &= \mathbb{E}\left[X_i\left(f(W) - f(W_i)\right)\right] \\
&= \mathbb{E}\left[X_i\left(f(W) - f(W_i) - (W - W_i)f'(W_i)\right] + \mathbb{E}\left[X_i(W - W_i)f'(W_i)\right].
\end{aligned}
$$

Weiter ergibt sich

$$
\begin{aligned}
\left|\mathbb{E}\left[X_i\left(f(W) - f(W_i) - (W - W_i)f'(W_i)\right)\right]\right| &\leq \frac{1}{2}\mathbb{E}\left|X_i(W - W_i)^2\right| |f''|_\infty \\
&= \frac{1}{2}\mathbb{E}\left|\frac{X_i^3}{n}\right| \underbrace{|f''|_\infty}_{\leq 2} \\
&\leq \frac{1}{n}\mathbb{E}\left|X_i\right|^3,
\end{aligned}
$$

wobei wir die Ungleichung

$$
|f(b) - f(a) - (b - a)f'(a)| \leq \tfrac{1}{2}(b - a)^2 |f''|_\infty
$$

mit $b = W$ und $a = W_i$ und die Tatsache, dass

$$
W - W_i = \tfrac{X_i}{\sqrt{n}}
$$

gilt, ausgenutzt haben. Des Weiteren gilt

$$
\begin{aligned}
\mathbb{E}\left[X_i(W - W_i)f'(W_i)\right] &= \mathbb{E}\left[X_i\frac{X_i}{\sqrt{n}}f'(W_i)\right] \\
&= \frac{1}{\sqrt{n}}\mathbb{E}\left[X_i^2 f'(W_i)\right] \\
&= \frac{1}{\sqrt{n}}\mathbb{E}\left[f'(W_i)\right]
\end{aligned}
$$

und somit folgt

$$
\begin{aligned}
\left|\mathbb{E}[Wf(W)] - \frac{1}{n}\sum_{i=1}^n \mathbb{E}\left[f'(W_i)\right]\right| &\leq \frac{1}{\sqrt{n}}\frac{1}{n}\sum_{i=1}^n \mathbb{E}\left|X_i\right|^3 \\
&= \frac{1}{n^{3/2}}\sum_{i=1}^n \mathbb{E}\left|X_i\right|^3. \qquad (1)
\end{aligned}
$$

Andererseits gilt

$$
\begin{aligned}
\left|\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left[f'(W_i)\right]-\mathbb{E}\left[f(W)\right]\right| &\leq \frac{1}{n}\left|f''\right|_\infty \sum_{i=1}^{n}\mathbb{E}\left|W-W_i\right| \\
&\leq \frac{|f''|_\infty}{n}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\mathbb{E}\left|X_i\right| \\
&\leq \frac{2}{n^{3/2}}\sum_{i=1}^{n}\mathbb{E}\left|X_i\right|, \qquad (2)
\end{aligned}
$$

und somit gilt schließlich unter Verwendung von (1) und (2) insgesamt

$$
\begin{aligned}
\left|\mathbb{E}[Wf(W)]-\mathbb{E}[f'(W)]\right| &\leq \frac{2}{n^{3/2}}\sum_{i}\mathbb{E}\left|X_i\right|^3+\frac{2}{n^{3/2}}\sum_{i}\underbrace{\mathbb{E}\left|X_i\right|}_{\leq \mathbb{E}|X_i|^3} \\
&\leq \frac{3}{n^{3/2}}\sum_{i=1}^{n}\mathbb{E}\left|X_i\right|^3.
\end{aligned}
$$

□

Anhand dieses Satzes und des zugehörigen Beweises haben wir insbesondere einen Zentralen Grenzwertsatz bewiesen und für die Wassersteinmetrik eine Konvergenzrate hergeleitet, jedoch zeigt das nachfolgende Lemma, dass diese Konvergenzrate keine optimale Berry-Esseen-Rate darstellt. Denn es gilt

Lemma 3.0.21.

$$kolm(W_n, Z) \leq \tfrac{2}{(2\pi)^{1/4}}(wass(W_n, Z))^{1/2}.$$

Somit haben wir für den Kolmogorov-Abstand keine optimale Konvergenzrate, sondern eine Rate der Ordnung $n^{1/4}$ hergeleitet. Um eine Berry-Esseen-Rate zu erreichen sind einige weitergehende Berechnungen nötig. Dies gelang zuerst E. Bolthausen in seiner Arbeit "An Estimate of the Remainder in a Combinatorial Central Limit Theorem" aus dem Jahr 1984, welche in „Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete“ veröffentlicht wurde [2]. Um das gewünschte Resultat mittels Stein-Ansatz herzuleiten muss man zu $g(X) = 1_{\{x\leq t\}}, t \in \mathbb{R}$, also zu speziellen, beschränkten Testfunktionen der Kolmogorov-Metrik übergehen. Vergleiche hierzu den in Kapitel 5 geführten Beweis des Berry-Esseen Theorems, welcher mittels Stein'scher Methode eine Konvergenzrate der Ordnung $o(n^{1/2})$ liefert.

4 Austauschbare Paare

In diesem Kapitel wollen wir die Methode der austauschbaren Paare erläutern, welche direkt auf C. Stein zurückgeht. Hierzu geben wir zunächst im ersten Abschnitt eine Wiederholung rund um das Thema bedingter Erwartungswert. Im zweiten Abschnitt dieses Kapitels werden wir uns der Theorie der austauschbaren Paare zuwenden, bevor wir diese im letzten Abschnitt anhand unabhängig verteilter Zufallsvariablen genauer erläutern.

4.1 Grundlagen zum bedingten Erwartungswert

Bevor wir uns der Technik der austauschbaren Paare widmen, geben wir in diesem Abschnitt einen kleinen Exkurs zum bedingten Erwartungswert, da dieser für die weiteren Berechnungen in diesem Kapitel von entscheidender Bedeutung sein wird.

Definition 4.1.1. (bedingter Erwartungswert)
Sei X eine Zufallsvariable auf $(\Omega, \mathcal{A}, P)$ und $\mathcal{C} \subset \mathcal{A}$ eine Teil-σ-Algebra. Eine Zufallsvariable $Z := \mathbb{E}(X|\mathcal{C})$ auf $(\Omega, \mathcal{A}, P)$ heißt bedingter Erwartungswert von X unter $\mathcal{C}$, falls

- Z ist $\mathcal{C}$-messbar mit $\mathbb{E}\,|Z| < \infty$ und
- $\displaystyle\int_C Z dP = \int_C X dP \quad \forall C \in \mathcal{C}.$

Bemerkung 4.1.2. *Im Allgemeinen existiert der bedingte Erwartungswert nicht immer. Es gilt aber, falls X in $\mathcal{L}^1(\Omega, \mathcal{A}, P)$ ist, existiert der bedingte Erwartungswert und ist sogar P-fast-sicher eindeutig.*

Satz 4.1.3. (Rechenregeln für den bedingten Erwartungswert)
Seien X, Y $\in \mathcal{L}^1(\Omega, \mathcal{A}, P)$, dann gilt für jedes $\mathcal{C} \subset \mathcal{A}$

1. $\mathbb{E}\left[\mathbb{E}\left[X|\mathcal{C}\right]\right] = \mathbb{E}[X]$
2. $X \, ist \, \mathcal{C} - messbar \Leftrightarrow \mathbb{E}\left[X|\mathcal{C}\right] \overset{\text{f.s.}}{=} X$
3. $X = const \Rightarrow E\left[X|\mathcal{C}\right] = const$
4. $\mathbb{E}\left[aX + Y|\mathcal{C}\right] \overset{\text{f.s.}}{=} a\mathbb{E}[X|\mathcal{C}] + \mathbb{E}[Y|\mathcal{C}] \quad \forall a \in \mathbb{R}$
5. $X \overset{\text{f.s.}}{\leq} Y \Rightarrow \mathbb{E}\left[X|\mathcal{C}\right] \leq \mathbb{E}\left[Y|\mathcal{C}\right]$
6. $|\mathbb{E}(X|\mathcal{C})| \overset{\text{f.s.}}{\leq} \mathbb{E}(|X|\,|\mathcal{C})$
7. falls X $\mathcal{C} - messbar$ und $\mathbb{E}[XY] < \infty$, dann gilt
$$\mathbb{E}\left[XY|\mathcal{C}\right] = X\mathbb{E}\left[Y|\mathcal{C}\right].$$

Beweis.

1. $\mathbb{E}\left[\mathbb{E}\left[X|\mathcal{C}\right]\right] = \int_\Omega \mathbb{E}\left[X|\mathcal{C}\right] dP = \int_\Omega X dP = \mathbb{E}[X]$
2. Folgt direkt aus der Definition des bedingten Erwartungswertes.
3. Die rechte Seite ist $\mathcal{C}$-messbar und für $C \in \mathcal{C}$ gilt
$\mathbb{E}\left[aX + Y|\mathcal{C}\right] = \mathbb{E}\left[(aX + Y)1_C\right] = \mathbb{E}\left[aX1C\right] + \mathbb{E}\left[Y1C\right] = a\mathbb{E}\left[X1C\right] + \mathbb{E}\left[Y1C\right]$
$= \mathbb{E}\left[a\mathbb{E}\left[aX1C\right]\right] + \mathbb{E}\left[\mathbb{E}\left[Y1C\right]\right] = \mathbb{E}\left[a\mathbb{E}\left[aX1C\right]\right] + \mathbb{E}\left[Y1C\right] = a\mathbb{E}[X|\mathcal{C}] + \mathbb{E}[Y|\mathcal{C}]$
4. Folgt direkt aus der Definition des bedingten Erwartungswertes.
5. Für alle C$\in \mathcal{C}$ gilt
$\mathbb{E}\left[\left(\mathbb{E}[Y|\mathcal{C}] - \mathbb{E}[X|\mathcal{C}]\right) 1_C\right] = \mathbb{E}\left[\mathbb{E}[Y - X|\mathcal{C}]1_C\right] = \mathbb{E}[(Y - X)1_C] \geq 0$
also ist $\mathbb{E}[Y|\mathcal{C}] - \mathbb{E}[X|\mathcal{C}] \geq 0$ P-f.s., da die Differenz $\mathcal{C}$-messbar ist.
6. Folgt aus der Linearität und Monotonie mit $X = X^+ - X^-$.
7. Nachzulesen z.B. in [7].

□

Satz 4.1.4. (Jensen-Ungleichung)
Sei $\varphi : \mathbb{R} \to \mathbb{R}$ eine konkave Funktion (d.h. für jeden $x_0 \in \mathbb{R}$ existiert ein $c := c(x_0)$ s.d. $\varphi(x) \geq \varphi(x_0) + c(x - x_0)$). Falls $X, \varphi(X) \in \mathcal{L}^1$, dann gilt

$$\varphi\left(\mathbb{E}\left[X\right]\right) \leq \mathbb{E}\left[\varphi\left(X\right)\right].$$

Beweis. Sei φ so, dass $X, \varphi(X) \in \mathcal{L}^1$. Sei weiter $X = x$ und $\mathbb{E}\left[X\right] = x_0$, dann folgt aus der Konvexität

$$\varphi(X) \geq \varphi(\mathbb{E}\left[X\right]) + c(X - \mathbb{E}\left[X\right]).$$

Wenn wir nun auf beiden Seiten der Ungleichung den Erwartungswert bilden, folgt aufgrund der Linearität des Erwartungswertes

$$\mathbb{E}\left[\varphi(X)\right] \geq \mathbb{E}\left[\varphi\left(\mathbb{E}[X]\right) + c(X - \mathbb{E}\left[X\right])\right] = \varphi(\mathbb{E}\left[X\right]).$$

□

Bemerkung 4.1.5. *Aus der Jensen-Ungleichung folgt mit $\varphi(x) = x^2$ und $\mathbb{E}\left[\mathbb{E}\left[X|W\right]\right] = \mathbb{E}[X]$, dass*

$$\begin{aligned} Var\left(\mathbb{E}\left[X|W\right]\right) &= \mathbb{E}\left[\mathbb{E}\left[X|W\right]^2\right] - \left(\mathbb{E}\left[\mathbb{E}\left[X|W\right]\right]\right)^2 \\ &\overset{Jensen}{\leq} \mathbb{E}\left[\mathbb{E}\left[X^2|W\right]\right] - \mathbb{E}\left[X\right]^2 \\ &= \mathbb{E}\left[X^2\right] - \mathbb{E}\left[X\right]^2 \\ &= Var(X) \end{aligned}$$

gilt.

4.2 Zugang mittels austauschbarer Paare

Definition 4.2.1. (austauschbare Zufallsvariablen)
Seien $(X_i)_{i\in\mathbb{N}}$ Zufallsvariablen, welche auf einem gemeinsamen Wahrscheinlichkeitsraum $(\Omega, \mathbb{A}, P)$ definiert sind. Die $(X_i)_{i\in\mathbb{N}}$ heißen austauschbar, falls für jede Permutation $\pi \in \mathbb{S}_n$

$$(X_1, ..., X_n) \overset{\mathrm{d}}{=} (X_{\pi(1)}, ..., X_{\pi(n)})$$

gilt.
Hierbei stellt $\mathbb{S}_n$ die symmetrische Gruppe dar. Dies ist die Gruppe, welche aus allen Permutationen einer n-elementigen Menge besteht.

Falls n=2 ist - dieser Fall wird im weiteren Verlauf für uns von entscheidender Bedeutung sein - spricht man von austauschbaren Paaren.
Zwei Zufallsvariablen W, W' auf $(\Omega, \mathbb{A}, P)$ bilden also ein austauschbares Paar, falls

$$(W, W') \stackrel{\text{d}}{=} (W', W). \qquad \text{(in Verteilung gleich)}$$

Mit Hilfe dieser Theorie wollen wir nun eine allgemeine Technik zur Abschätzung von

$$\mathbb{E}\left[f'(W) - Wf(W)\right]$$

betrachten, welche direkt auf C. Stein zurückgeht. Vorteil dieser Technik ist, dass wir nicht mehr wie bisher nur unabhängige Summanden betrachten können.
Hierzu sei die Zufallsgröße W' so gewählt, dass $W \stackrel{\text{d}}{=} W'$ gilt mit $\mathbb{E}W^2 = 1$ und

$$\mathbb{E}\left[W' - W|W\right] = -\lambda W, \lambda \in (0, 1)$$

fast sicher. Die letzte Aussage ist dabei äquivalent zu $\mathbb{E}[W'|W] = (1-\lambda)W$, denn es gilt $\mathbb{E}[W' - W|W] = \mathbb{E}[W'|W]) - \mathbb{E}[W|W] = \mathbb{E}[W'|W] - W$ woraus $\mathbb{E}[W'|W] = (1-\lambda)W$ folgt.

Satz 4.2.2. *Sei (W, W') ein austauschbares Paar von Zufallsvariablen auf $(\Omega, \mathbb{A}, P)$ mit endlicher Varianz. Des weitern existiere ein $\lambda \in (0, 1)$ mit $\mathbb{E}\left[W' - W|W\right] = -\lambda W$. Dann gilt*

$$\mathbb{E}[W] = 0 \qquad \textit{und} \qquad \mathbb{E}\left[(W - W')^2\right] = 2\lambda \mathbb{E}[W^2].$$

Beweis. Zunächst gilt wegen $W \stackrel{\text{d}}{=} W'$, dass

$$\mathbb{E}[W] = \mathbb{E}[W'] \text{ und } \mathbb{E}[W^2] = \mathbb{E}[W']^2 = 1.$$

Also gilt

$$\mathbb{E}[W] = \mathbb{E}[W'] = \mathbb{E}\left[\mathbb{E}[W'|W]\right] = \mathbb{E}\left[(1-\lambda)W\right] = (1-\lambda)\mathbb{E}W,$$

und somit folgt, da $\lambda \neq 0$ nach Voraussetzung, dass $\mathbb{E}[W] = 0$. Weiter gilt mit Hilfe der Rechenregeln für den bedingten Erwartungswert

$$\begin{aligned}
\mathbb{E}\left[W' - W\right]^2 &= \mathbb{E}\left[W^2 - 2WW' + W'^2\right] \\
&= \mathbb{E}\left[2W^2 - 2WW'\right] \\
&= \mathbb{E}\left[2W(W - W')\right] \\
&= \mathbb{E}\left[2W\mathbb{E}(W - W'|W)\right] \\
&= \mathbb{E}\left[2W{-}E(W' - W|W)\right] \\
&= \mathbb{E}\left[2\lambda W^2\right] \\
&= 2\lambda.
\end{aligned}$$

□

Unter den Voraussetzungen, welche wir zu Beginn dieses Abschnitts getroffen haben gilt nun der folgende Satz:

Satz 4.2.3. (Stein)
Sei $Z \sim \mathcal{N}(0,1)$, dann gilt

$$wass(W,Z) \leq \sqrt{\tfrac{2}{\pi} Var\left[\mathbb{E}\left(\tfrac{1}{2\lambda}(W' - W)^2|W\right)\right]} + \tfrac{1}{3\lambda}\mathbb{E}\left|W' - W\right|^3.$$

Beweis. Zunächst wählen wir eine Funktion $f : \mathbb{R} \to \mathbb{R}$, welche die Schranken aus Satz 2.0.17 (2) erfüllt, also

$$\|f\|_\infty \leq 1, \qquad \|f'\|_\infty \leq \sqrt{\tfrac{2}{\pi}}, \qquad \|f''\|_\infty \leq 2.$$

Zur weiteren Abschätzung betrachten wir nun

$$F(x) := \int_0^x f(y)dy.$$

Nutzen wir wieder die Übereinstimmung der Verteilungen von W und W', so können wir mittels Satz von Taylor wie folgt abschätzen

$$\begin{aligned}
0 &= \mathbb{E}\left[F(W') - F(W)\right] \\
&= \mathbb{E}\left[(W' - W)f(W) + \frac{1}{2}(W' - W)^2 f'(W) + R\right].
\end{aligned}$$

Dies ist aufgrund der Unabhängigkeit äquivalent zu

$$\mathbb{E}\left[(W'-W)f(W)\right] = -\mathbb{E}\left[\tfrac{1}{2}(W'-W)^2 f'(W) + R\right]. \qquad (*)$$

Wobei wir das Restglied R der Taylorapproximation durch den 3. Term der Taylorreihe abschätzen können. Also gilt für R

$$|R| \leq \tfrac{1}{3!}\left|W'-W\right|^3 \underbrace{\|f''\|_\infty}_{\leq 2} \leq \tfrac{1}{3}\left|W'-W\right|^3.$$

Weiter gilt nach Satz 4.2.2 und den vorrangegangenen Berechnungen

$$\begin{aligned}
-\lambda\mathbb{E}\left[Wf(W)\right] &= \mathbb{E}\left[-\lambda W\right]\mathbb{E}\left[f(W)\right] \\
&= \mathbb{E}\left[W'-W\right]\mathbb{E}\left[f(W)\right] \\
&= \mathbb{E}\left[(W'-W)f(W)\right] \\
&\overset{(*)}{=} -\mathbb{E}\left[\frac{1}{2}(W'-W)^2 f(W) + R\right] \\
&= -\mathbb{E}\left[\frac{1}{2}\mathbb{E}\left[(W'-W)^2|W\right]f'(W)\right] + \mathbb{E}\left[R\right].
\end{aligned}$$

Insgesamt folgt nun mit der abschließenden Abschätzung die Aussage des Satzes

$$\begin{aligned}
&|\mathbb{E}[f'(W)] - \mathbb{E}[Wf(W)]| \\
&\leq \left|\mathbb{E}\left[\sqrt{\frac{2}{\pi}}\right] - \frac{1}{\lambda}\mathbb{E}\left[\frac{1}{2}\mathbb{E}\left[(W'-W)^2|W\right]f'(W)\right] - \frac{1}{\lambda}\mathbb{E}[R]\right| \\
&\leq \left|\mathbb{E}\left[\sqrt{\frac{2}{\pi}}\right] - \frac{1}{\lambda}\mathbb{E}\left[\frac{1}{2}\mathbb{E}\left[(W'-W)^2|W\right]f'(W)\right] - \frac{1}{3\lambda}\mathbb{E}\left|W'-W\right|^3\right| \\
&\leq \left|\mathbb{E}\left[\sqrt{\frac{2}{\pi}}\right] - \frac{1}{\lambda}\mathbb{E}\left[\frac{1}{2}\mathbb{E}\left[(W'-W)^2|W\right]f'(W)\right]\right| + \frac{1}{3\lambda}\mathbb{E}\left|W'-W\right|^3 \\
&\leq \left|\mathbb{E}\left[\sqrt{\frac{2}{\pi}}\right] - \mathbb{E}\left[\frac{1}{2\lambda}\mathbb{E}\left[(W'-W)^2|W\right]\sqrt{\frac{2}{\pi}}\right]\right| + \frac{1}{3\lambda}\mathbb{E}\left|W'-W\right|^3 \\
&= \left|\sqrt{\frac{2}{\pi}}\left[\mathbb{E}\frac{1}{2\lambda}\mathbb{E}\left[(W'-W)^2|W\right] - 1\right]\right| + \frac{1}{3\lambda}\mathbb{E}\left|W'-W\right|^3 \\
&= \sqrt{\frac{2}{\pi}}\mathbb{E}\left|\mathbb{E}\left[\frac{1}{2\lambda}(W'-W)^2|W\right] - 1\right| + \frac{1}{3\lambda}\mathbb{E}\left|W'-W\right|^3 \\
&\overset{(**)}{\leq} \sqrt{\frac{2}{\pi}}\sqrt{Var\left(\mathbb{E}\left[\frac{1}{2\lambda}(W-W')^2|W\right] - 1\right)} + \frac{1}{3\lambda}\mathbb{E}\left|W'-W\right|^3
\end{aligned}$$

$$\leq \sqrt{\frac{2}{\pi}}\sqrt{Var\left(\mathbb{E}\left[\frac{1}{2\lambda}(W-W')^2|W\right]\right)} + \frac{1}{3\lambda}\mathbb{E}\left|W'-W\right|^3$$
$$\leq \sqrt{\frac{2}{\pi}Var\left[\mathbb{E}\left[\frac{1}{2\lambda}(W'-W)^2|W\right]\right]} + \frac{1}{3\lambda}\mathbb{E}\left|W'-W\right|^3.$$

In der Umformung (**) haben wir ausgenutzt, dass $\mathbb{E}\left|Z\right| \leq \sqrt{Var(Z)}$ gilt, wobei wir $Z = \mathbb{E}\left[\frac{1}{2\lambda}(W-W')^2|W\right] - 1$ gewählt haben. $\square$

4.3 Der unabhängige Fall

Zur Verdeutlichung dieser recht theoretisch erscheinenden Technik wollen wir in diesem Abschnitt die Methode der austauschbaren Paare einmal genauer für den Fall unabhängig verteilter Zufallsvariablen betrachten.
Im ersten Schritt betrachten wir die Konstruktion der W'.
Hierzu seien die $(X_i)_{i\in\mathbb{N}}$ unabhängig mit $\mathbb{E}\left[X_i\right] = 0$, $Var\left(X_i\right) = 1$. Sei wieder, wie bei der Einführung der Stein'schen Methode

$$W := W_n := \tfrac{1}{\sqrt{n}}\sum_{i=1}^{n} X_i$$

definiert.
Die Zufallsgrößen $(X_i')_{i\in\mathbb{N}}$ seien unabhängige Kopien der $(X_i)_{i\in\mathbb{N}}$.
Des Weiteren wählen wir I gleichmäßig, zufällig aus $\{1, ..., n\}$ mit $P(I = j) = \frac{1}{n}$ unabhängig von allen X_i', X_i.
Zur Konstruktion von W' ersetzen wir X_I durch X_I' in der Form

$$W' = W_n' = \tfrac{1}{\sqrt{n}}\sum_{j\neq I} X_j + \tfrac{X_I'}{\sqrt{n}} = W + \tfrac{X_I'-X_I}{\sqrt{n}}.$$

Satz 4.3.1. *Seien $X_1, ..., X_n, X_1', ..., X_n', I$ unabhängige Zufallsvariablen, definiert auf dem Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, P)$. Es gelte $X_i \stackrel{d}{=} X_i'$ $\forall i = 1, ..., n$ und I habe beliebige Verteilung auf $\{1, ..., n\}$.*
Seien $W = \frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_i$ und $W' = W + \frac{X_I'-X_I}{\sqrt{n}}$. Dann gilt

$$(W, W') \stackrel{d}{=} (W', W)$$

Beweis. Seien $A, B \in \mathcal{B}$. Dann gilt

$$
\begin{aligned}
P(W \in A, W' \in B) &= P\left(\bigcup_{i=1}^{n} \{I = i\} \cap \{W \in A\} \cap \{W' \in B\}\right) \\
&= \sum_{i=1}^{n} P\left(I = i, \frac{1}{\sqrt{n}} \sum_{j=1}^{n} X_j \in A, \frac{1}{\sqrt{n}} \sum_{\substack{j=1,\\ j\neq i}}^{n} X_j + \frac{X_i'}{\sqrt{n}} \in B\right) \\
&= \sum_{i=1}^{n} P\left(I = i\right) P\left(\sum_{j=1}^{n} \frac{X_j}{\sqrt{n}} \in A, \sum_{\substack{j=1,\\ j\neq i}}^{n} \frac{X_j}{\sqrt{n}} + \frac{X_i'}{\sqrt{n}} \in B\right). \qquad (1)
\end{aligned}
$$

Im nächsten Schritt wählen wir $i \in (1, ..., n)$ fest und definieren Funktionen $S, S' : \mathbb{R}^{n+1} \to \mathbb{R}$ der Form

$$
\begin{aligned}
S(x_1, ..., x_{n+1}) &= \tfrac{1}{\sqrt{n}} \sum_{j=1}^{n} x_j, \\
S'(x_1, ..., x_{n+1}) &= \tfrac{1}{\sqrt{n}} \sum_{\substack{j=1,\\ j\neq i}}^{n} x_j + x_{n+1}.
\end{aligned}
$$

Weiter definieren wir eine Permutation als Funktion $\pi : \mathbb{R}^{n+1} \to \mathbb{R}^{n+1}$ durch

$$\pi\left(x_1, ..., x_{n+1}\right) := (x_1, ..., x_{i-1}, x_{n+1}, x_{i+1}, ..., x_n, x_i).$$

Somit gilt dann

$$S' = S \circ \pi \qquad \textit{und} \qquad \pi^2 = id_{\mathbb{R}^{n+1}} \qquad (*).$$

Nach den Voraussetzungen des Satzes gilt

$$
\begin{aligned}
P^{(X_1, ..., X_i, ..., X_n, X_i')} &= P^{X_1} \otimes ... \otimes P^{X_i} \otimes ... \otimes P^{X_n} \otimes P^{X_i'} \\
&= P^{X_1} \otimes ... \otimes P^{X_i'} \otimes ... \otimes P^{X_n} \otimes P^{X_i} \\
&= P^{(X_1, ..., X_i', ..., X_n, X_i)} \\
&= P^{\pi(X_1, ..., X_i, ..., X_n, X_i')} \qquad (2)
\end{aligned}
$$

Somit folgt insgesamt mit Hilfe der Transformationsformel

$$
\begin{aligned}
&P\left(\sum_{j=1}^{n}\frac{X_j}{\sqrt{n}}\in A,\sum_{\substack{j=1,\\ j\neq i}}^{n}\frac{X_j}{\sqrt{n}}+\frac{X_i'}{\sqrt{n}}\in B\right)\\
&=\int_{\Omega}1_A\left(S\left(X_1,...,X_n,X_i'\right)\right)1_B\left(S'\left(X_1,...,X_n,X_i'\right)\right)dP\\
&=\int_{\mathbb{R}^{n+1}}1_A\left(S\left(x_1,...,x_{n+1}\right)\right)1_B\left(S'\left(x_1,...,x_{n+1}\right)\right)dP^{X_1,...,X_n,X_i'}\left(x_1,...,x_{n+1}\right)\\
&\overset{(*)}{=}\int_{\mathbb{R}^{n+1}}1_A\left(S\circ\pi^2\right)1_B\left(S\circ\pi\right)dP^{X_1,...,X_n,X_i'}\\
&=\int_{\mathbb{R}^{n+1}}1_A\left(S\circ\pi\right)1_B\left(S\right)dP^{\pi\left(X_1,...,X_n,X_i'\right)}\\
&\overset{(2)}{=}\int_{\mathbb{R}^{n+1}}1_A\left(S'\left(x_1,...,x_{n+1}\right)\right)1_B\left(S\left(x_1,...,x_{n+1}\right)\right)dP^{X_1,...,X_n,X_i'}\left(x_1,...,x_{n+1}\right)\\
&=\int_{\Omega}1_A\left(S'\left(X_1,...,X_n,X_i'\right)\right)1_B\left(S\left(X_1,...,X_n,X_i'\right)\right)dP\\
&=P\left(\sum_{j=1}^{n}\frac{X_j}{\sqrt{n}}\in B,\sum_{\substack{j=1,\\ j\neq i}}^{n}\frac{X_j}{\sqrt{n}}+\frac{X_i'}{\sqrt{n}}\in A\right).
\end{aligned}
$$

Einsetzen in (1) ergibt dann schließlich

$$P\left(W\in A,W'\in B\right)=P\left(W\in B,W'\in A\right)$$

also

$$(W,W')\overset{\mathrm{d}}{=}(W',W)$$

Somit haben wir gezeigt, dass W,W' ein austauschbares Paar bilden. □

Als nächstes wollen wir überprüfen, ob unsere Bedingung $\lambda\in(0,1)$ erfüllt ist, ohne welche die weiteren Berechnungen nicht möglich wären. Es gilt

$$
\begin{aligned}
-\lambda W &\underset{\text{Def.}}{\overset{\text{nach}}{=}}\mathbb{E}\left[W'-W|W\right]\\
&=\mathbb{E}\left[W+\frac{X_I'-X_I}{\sqrt{n}}-W|W\right]\\
&=\frac{1}{\sqrt{n}}\mathbb{E}\left[X_I'-X_I|W\right]\\
&=\frac{1}{\sqrt{n}}\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left[X_i'-X_i|W\right]
\end{aligned}
$$

$$
\begin{aligned}
&= \frac{1}{\sqrt{n}}\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left[-X_i|W\right] + \underbrace{\mathbb{E}\left[X_i'|W\right]}_{0} \\
&= \frac{1}{n}\mathbb{E}\left[\frac{1}{\sqrt{n}}\sum_{i=1}^{n} -X_i|W\right] \\
&= -\frac{1}{n}\mathbb{E}\left[\underbrace{\frac{1}{\sqrt{n}}\sum_{i=1}^{n} X_i}_{W}|W\right] \\
&= -\frac{1}{n}\mathbb{E}\left[W|W\right] \\
&= -\frac{1}{n}W.
\end{aligned}
$$

Somit ist mit $\lambda = \frac{1}{n}$ die an λ gestellte Bedingung erfüllt.
Nun wollen wir den letzten Summanden unserer Abschätzung aus Satz 4.2.3 abschätzen.
Es gilt

$$
\begin{aligned}
\frac{1}{3\lambda}\mathbb{E}\left|W'-W\right|^3 &= \frac{1}{3\lambda}\left(\frac{1}{\sqrt{n}}\right)^3\mathbb{E}\left|X_I'-X_I\right|^3 \\
&\overset{\lambda=\frac{1}{n}}{=} \frac{n}{3}\frac{1}{n^{3/2}}\mathbb{E}\left|X_I'-X_I\right|^3 \\
&= \frac{1}{3}\frac{1}{n^{3/2}}\sum_{i=1}^{n}\mathbb{E}\left|X_i'-X_i\right|^3 \\
&= \frac{1}{3}\frac{1}{n^{3/2}}\sum_{i=1}^{n}\mathbb{E}\left|(X_i')^3 - 3(X_i')^2X_i + 3X_i^2X_i' - X_i^3\right| \\
&\le \frac{1}{3}\frac{1}{n^{3/2}}\sum_{i=1}^{n}\mathbb{E}\left|(X_i')^3\right| + 3\mathbb{E}\left|(X_i')^2X_i\right| + 3\mathbb{E}\left|X_i^2X_i'\right| + \mathbb{E}\left|X_i^3\right| \\
&= \frac{1}{3}\frac{1}{n^{3/2}}\sum_{i=1}^{n}\mathbb{E}\left|X_i^3\right| + 3\mathbb{E}\left|X_i^2X_i\right| + 3\mathbb{E}\left|X_i^2X_i\right| + \mathbb{E}\left|X_i^3\right| \\
&= \frac{8}{3}\frac{1}{n^{3/2}}\sum_{i=1}^{n}\mathbb{E}\left|X_i\right|^3.
\end{aligned}
$$

Im Fall unabhängig, identisch verteilter Zufallsvariablen gilt wegen $\mathbb{E}\left|X_i\right| = \mathbb{E}\left|X_1\right|$ und somit $\sum\limits_{i=1}^{n}\mathbb{E}\left|X_i\right|^3 = n\mathbb{E}\left|X_1\right|^3$, dass

$$\tfrac{8}{3}\tfrac{1}{n^{3/2}}\sum_{i=1}^{n}\mathbb{E}\left|X_i\right|^3 = \tfrac{8}{3}\tfrac{1}{\sqrt{n}}\mathbb{E}\left|X_1\right|^3.$$

Kommen wir nun zur Abschätzung des ersten Sumanden aus Satz 4.2.3. Um $\mathbb{E}\left[(W'-W)^2|X\right]$ mit $X=(X_1,...,X_n)$ abzuschätzen zu können, formen wir dies zunächst wie folgt um

$$\begin{aligned}
\mathbb{E}\left[(W'-W)^2|X\right] &= \mathbb{E}\left[\left(W+\frac{X_i'-X_i}{\sqrt{n}}-W\right)^2|W\right]\\
&= \frac{1}{n}\mathbb{E}\left[(X_I'-X_I)^2|X\right]\\
&= \frac{1}{n}\mathbb{E}\left[(X_I')^2-2X_I'X_I+X_I^2|X\right]\\
&= \frac{1}{n}\left(\mathbb{E}\left[(X_I')^2|X\right]-\mathbb{E}\left[2X_I'X_I|X\right]+\mathbb{E}\left[X_I^2|X\right]\right). \qquad (**)
\end{aligned}$$

Für die weitere Berechnung betrachten wir zunächst die drei Terme im Einzelnen.

a) Aufgrund der Unabhängigkeit von X_I' und X gilt durch ausnutzen, dass X_i' unabhängige Kopien der X_i sind

$$\mathbb{E}\left[(X_I')^2|W\right]=\mathbb{E}\left[(X_I')^2\right]=\tfrac{1}{n}\sum_{i=1}^{n}\mathbb{E}\left[(X_i')^2\right]=\tfrac{1}{n}\sum_{i=1}^{n}\underbrace{\mathbb{E}\left[X_i^2\right]}_{1}=1.$$

b) Bei der Betrachtung des nächsten Summanden nutzen wir die Unabhängigkeit von X_I und I aus. Somit gilt

$$\mathbb{E}[X_I^2|X]=\sum_{i=1}^{n}\mathbb{E}\left[1_{\{I=i\}}X_i^2|X\right]=\sum_{i=1}^{n}X_i^2\mathbb{E}\left[1_{\{I=i\}}|X\right]=\sum_{i=1}^{n}X_i^2\mathbb{E}\left[1_{\{I=i\}}\right]=$$
$$\sum_{i=1}^{n}P(I=i)X_i^2=\tfrac{1}{n}\sum_{i=1}^{n}X_i^2.$$

c) Zur Berechnung von $\mathbb{E}\left[X_IX_I'|X\right]$ wollen wir zunächst einmal $\mathbb{E}\left[X_I'|X\right]$ betrachten. Da X_I' und I unabhängig sind gilt

$$\begin{aligned}
\mathbb{E}\left[X_I'|X\right]&=\sum_{i=1}^{n}\mathbb{E}\left[1_{\{I=i\}}X_i'|I\right]=\sum_{i=1}^{n}1_{\{I=i\}}\mathbb{E}\left[X_i'|I\right]\\
&=\sum_{i=1}^{n}1_{\{I=i\}}\underbrace{\mathbb{E}\left[X_i'\right]}_{0}=0. \qquad (**)
\end{aligned}$$

Da $\sigma(X'_I, I)$ unabhängig von $\sigma(X)$ ist, gilt $\mathbb{E}[X'_I|X,I] = \mathbb{E}[X'_I|,I]$. Somit folgt

$$\begin{aligned}\mathbb{E}[X_I X'_I|X] &= \mathbb{E}[\mathbb{E}[X_I X'_I|X,I]\,|X] = \mathbb{E}[X_I \mathbb{E}[X'_I|X,I]\,|X] \\ &= \mathbb{E}[X_I \mathbb{E}[X'_I|I]\,|X] \overset{(**)}{=} \mathbb{E}[X_I * 0|X] = 0.\end{aligned}$$

Insgesamt folgt nun durch Einsetzen von a), b), c) in (*)

$$\begin{aligned}\mathbb{E}\left[\frac{1}{2\lambda}(W'-W)^2|X\right] &= \frac{1}{2\lambda}\frac{1}{n}\left(\mathbb{E}\left[(X'_I)^2|X\right] - \mathbb{E}[2X'_I X_I|X] + \mathbb{E}\left[X_I^2|X\right]\right) \\ &\overset{\text{a,b,c}}{=} \frac{1}{2\lambda}\frac{1}{n}\left(\frac{1}{n}\sum_{i=1}^n X_i^2 - 2*0+1\right) \\ &= \frac{1}{2\lambda}\frac{1}{n}\left(\frac{1}{n}\sum_{i=1}^n X_i^2 + 1\right) \\ &\overset{\lambda=\frac{1}{n}}{=} \frac{1}{2n}\sum_{i=1}^n X_i^2 + \frac{1}{2}.\end{aligned}$$

Dies stimmt fast sicher mit $\mathbb{E}\left[\frac{1}{2\lambda}(W'-W)^2|W\right]$ überein, da aufgrund der Messbarkeit von W bezüglich $\sigma(X)$, $\sigma(W) \subset \sigma(X)$ gilt. Um nun also $wass(W,Z)$ nach Satz 4.2.3 abschließend abschätzen zu können betrachten wir nun die Varianz von $\frac{1}{2n}\sum\limits_{i=1}^n X_i^2 + \frac{1}{2}$. Nach den Rechenregeln für die Varianz gilt

$$\begin{aligned}Var\left(\frac{1}{2n}\sum_{i=1}^n \mathbb{E}\left[X_i^2|W\right] + \frac{1}{2}\right) &= Var\left(\mathbb{E}\left[\frac{1}{2n}\sum_{i=1}^n X_i^2|W\right]\right) \\ &\overset{\text{Bem. 4.1.5}}{\leq} Var\left(\frac{1}{2n}\sum_{i=1}^n X_i^2\right) \\ &= \frac{1}{4n^2}\sum_{i=1}^n Var\left(X_i^2\right) \\ &\leq \frac{1}{4n^2}\sum_{i=1}^n \mathbb{E}\left[X_i^4\right].\end{aligned}$$

Im Fall unabhängig, identisch verteilter Zufallsvariablen ist wegen $\mathbb{E}[X_i] = \mathbb{E}[X_1]$ und somit $\sum\limits_{i=1}^n \mathbb{E}[X_i^4] = n\mathbb{E}[X_1^4]$ dies gleich $\frac{1}{4n}\mathbb{E}[X_1^4]$.

Insgesamt haben wir mit unseren Berechnungen nun also gezeigt, dass wir $wass(W,Z)$ wie folgt abschätzen können

$$
\begin{aligned}
wass(W,Z) &\leq \sqrt{\frac{2}{\pi}Var\left[\mathbb{E}\left(\frac{1}{2\lambda}(W'-W)^2|W\right)\right]} + \frac{1}{3\lambda}\mathbb{E}\left|W'-W\right|^3 \\
&\leq \sqrt{\frac{2}{\pi}\frac{1}{4n^2}\sum_{i=1}^{n}\mathbb{E}\left[X_i^4\right]} + \frac{8}{3n^{3/2}}\sum_{i=1}^{n}\mathbb{E}\left|X_i\right|^3 \\
&= \sqrt{\frac{1}{2\pi n^2}\sum_{i=1}^{n}\mathbb{E}\left[X_i^4\right]} + \frac{8}{3n^{3/2}}\sum_{i=1}^{n}\mathbb{E}\left|X_i\right|^3 .
\end{aligned}
$$

Im Fall unabhängig, identisch verteilter Zufallsvariablen folgt somit

$$
\begin{aligned}
wass(W,Z) &\leq \sqrt{\frac{2}{\pi}\frac{1}{4n}\mathbb{E}\left[X_1^4\right]} + \frac{8}{3}\frac{1}{\sqrt{n}}\mathbb{E}\left|X_1\right|^3 \\
&= \frac{const}{\sqrt{n}}\left(\mathbb{E}\left|X_1\right|^3 + \mathbb{E}\left[X_1^4\right]\right).
\end{aligned}
$$

Somit ist es uns gelungen einen Zentralen Grenzwertsatz zu zeigen, und wie in Kapitel 2 eine Konvergenzrate der Ordnung $n^{1/4}$ für den Kolmogorov-Abstand zu erreichen.

5 Berry-Esseen via Stein'scher Methode

In diesem letzten Kapitel wollen wir einen Beweis für das klassische Berry-Essen Theorem 2.0.15 diskutieren. Der Beweis, welcher auf E. Bolthausen zurückgeht, verwendet eine Variante der Stein'schen Methode. Einen ähnlichen Beweis führten zuvor bereits L. H. Y Chen und S. T. Ho [4], jedoch war ihr Beweis im Gegensatz zu jenem, welchen wir im Folgenden diskutieren wollen, nicht auf nicht-unabhänige Zufallsvariablen anwendbar. Der Ansatz von Bolthausen ist flexibler und auch auf nicht-unabhänige Zufallsvariablen erweiterbar.

Sei $n \in \mathbb{N}$, $\gamma \geq 1$ und $X = (X_1, ..., X_n) \in \mathcal{L}(n, \gamma)$ Zufallsvariablen mit $(X_i)_{i \in \mathbb{N}}$ unabhängig, identisch verteilt mit $\mathbb{E}[X_i] = 0$, $\mathbb{E}[X_i^2] = 1$, $\mathbb{E}\,|X_i^3| = \gamma$.
Falls $\gamma < 1$ gilt $\mathcal{L}(n, \gamma) = \emptyset$, sei also im Folgenden $\gamma \geq 1$.
Sei weiter

$$S_k = \sum_{i=1}^{k} \frac{X_i}{\sqrt{n}},\ i \leq k \leq n \qquad \Rightarrow S_n = \sum_{i=1}^{n} \frac{X_i}{\sqrt{n}} = \sqrt{n} X_n$$

Falls $z, x \in \mathbb{R}, \lambda > 0$ gilt, definiere $h_{z,\lambda} : \mathbb{R} \to \mathbb{R}$ durch

$$h_{z,\lambda}(x) = \begin{cases} 1, & falls\, x \leq z, \\ 1 - \frac{x-z}{\lambda}, & falls\, z < x < z + \lambda, \\ 0 & falls\, x \geq z + \lambda \end{cases}$$

$$h_{z,0}(x) = 1_{(-\infty, z]}(x).$$

Sei weiter

$$\delta(\lambda, \gamma, n) = sup \left\{ \left| \mathbb{E}\left[h_{z,\lambda}(S_n) \right] - \Phi(h_{z,\lambda}) \right| : z \in \mathbb{R}, X \in \mathcal{L}(n, \gamma) \right\}.$$

Wobei $\Phi(x) := \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{1}{2}t^2} dt$ die Verteilungsfunktion der Standardnormalverteilung ist. Für $\delta(0, \gamma, n)$ schreiben wir in Zukunft der Einfachheit halber nur $\delta(\gamma, n)$.

In diesem Kapitel werden wir zeigen, dass eine Konstante $K \in [0, \infty)$ mit

$$\delta(\gamma, n) \leq \tfrac{K\gamma}{\sqrt{n}}$$

existiert.

Zunächst wollen wir vorbereitend den folgenden Satz beweisen.

Satz 5.0.2. *Für alle $n \in \mathbb{N}$, $\lambda > 0$, $\gamma \geq 1$ gilt*

$$\delta(\gamma, n) \leq \delta(\lambda, \gamma, n) + \tfrac{\lambda}{2\sqrt{2\pi}}.$$

Beweis. Da $h_{z-\lambda,0} \leq h_{z,0} \leq h_{z,\lambda}$ gilt, folgt

$$\begin{aligned}\mathbb{E}\left[h_{z-\lambda,\lambda}(S_n)\right] - \Phi(h_{z,0}) &\leq \mathbb{E}\left[h_{z,0}(S_n)\right] - \Phi(h_{z,0}) \\ &\leq \mathbb{E}\left[h_{z,\lambda}(S_n)\right] - \Phi(h_{z,0}).\end{aligned}$$

Somit folgt

$$\begin{aligned}&\left|\mathbb{E}\left[h_{z,0}(S_n)\right] - \Phi(h_{z,0})\right| \\ &\leq max\left\{\left|\mathbb{E}\left[h_{z-\lambda,\lambda}(S_n)\right] - \Phi(h_{z,0})\right|, \left|\mathbb{E}\left[h_{z,\lambda}(S_n)\right] - \Phi(h_{z,0})\right|\right\} \\ &\leq max\left\{\left|\mathbb{E}\left[h_{z-\lambda,\lambda}(S_n)\right] - \Phi(h_{z-\lambda,\lambda})\right| + \left|\Phi(h_{z-\lambda,\lambda}) - \Phi(h_{z,0})\right|\right. \\ &\qquad \left., \left|\mathbb{E}\left[h_{z,\lambda}(S_n)\right] - \Phi(h_{z,\lambda})\right| + \left|\Phi(h_{z,\lambda}) - \Phi(h_{z,0})\right|\right\} \\ &\leq \delta(\lambda, \gamma, n) + max\left\{\left|\Phi(h_{z-\lambda,\lambda}) - \Phi(h_{z,0})\right|, \left|\Phi(h_{z,\lambda}) - \Phi(h_{z,0})\right|\right\}. \qquad (*)\end{aligned}$$

Für die beiden Terme der Maximumsauswahl gilt weiter

$$\begin{aligned}\left|\Phi(h_{z-\lambda,\lambda}) - \Phi(h_{z,0})\right| &= \Phi(h_{z,0}) - \Phi(h_{z-\lambda,\lambda}) \\ &= \int_{z-\lambda}^{z} \left(1 - \left(1 - \frac{y-(z-\lambda)}{\lambda}\right)\right) \varphi(y)dy \\ &\leq \frac{1}{\sqrt{2\pi}} \int_{z-\lambda}^{z} \frac{y-(z-\lambda)}{\lambda} dy \\ &= \frac{1}{\sqrt{2\pi}\lambda} \int_{0}^{\lambda} ydy \\ &= \frac{\lambda}{2\sqrt{2\pi}}\end{aligned}$$

und analog

$$\begin{aligned}
|\Phi(h_{z,\lambda}) - \Phi(h_{z,0})| &= \int_z^{z+\lambda} \left(1 - \frac{y-z}{\lambda}\right) \varphi(y)dy \\
&\leq \frac{1}{\sqrt{2\pi}} \int_z^{z+\lambda} \frac{\lambda - y - z}{\lambda} dy \\
&= \frac{\lambda}{2\sqrt{2\pi}}.
\end{aligned}$$

Wenn wie diese Ergebnisse nun in die Abschätzung (*) einsetzen folgt direkt

$$\delta(\gamma, n) \leq \delta(\lambda, \gamma, n) + \tfrac{\lambda}{2\sqrt{2\pi}}.$$

□

Von nun an werden wir der Einfachheit halber h an Stelle von $h_{z,\lambda}$ schreiben.
Sei $f(x) = e^{x^2/2} \int_{-\infty}^{x} (h(z) - \Phi(h)) e^{-z^2/2} dz$ so gewählt, dass

$$\begin{aligned}
f'(x) - xf(x) &= h(x) - \Phi(h) \\
\Leftrightarrow f'(x) &= h(x) - \Phi(h) + xf(x), \qquad (**)
\end{aligned}$$

also die Stein-Gleichung erfüllt ist.
Im Fall $x \leq 0$ gilt $|f(x)| \leq \Phi(x)/\varphi(x)$, wobei $\varphi(x) := \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$ die Dichte der Standardnormalverteilung ist.
Falls $x > 0$ ist, gilt $|f(x)| \leq (1 - \Phi(x))/\varphi(x)$.
Somit folgen die Abschätzungen

$$|f(x)| \leq 1, \qquad |xf(x)| \leq 1, \qquad |f'(x)| \leq 2 \quad \forall x,$$

wobei wir für die Abschätzung von $|f'(x)|$ die Gleichung (**) genutzt haben.
Damit gilt nun für $x \in \mathcal{L}(n, \gamma)$

$$\begin{aligned}
|f'(x+y) - f'(x)| &= |h(x+y) - \Phi(h) + (x-y)f(x+y) - h(x) + \Phi(x) - xf(x)| \\
&= |yf(x+y) + x(f(x+y) - f(x)) + h(x+y) - h(x)| \\
&\leq |y| \left(|f|_\infty + |x||f'|_\infty + \frac{1}{\lambda} \int_0^1 1_{[z,z+\lambda]}(x+ys)ds \right) \\
&\leq |y| \left(1 + 2|x| + \frac{1}{\lambda} \int_0^1 1_{[z,z+\lambda]}(x+ys)ds \right), \qquad (1)
\end{aligned}$$

wobei wir ausgenutzt haben, dass für alle $y \in \mathbb{R}$ auf Grund der L-Stetigkeit von f, h (somit insbesondere absolut stetig)

$$\begin{aligned} h(x+y) - h(y) &= -\frac{1}{\lambda} \int_x^{x+y} 1_{(z,z+y]} t dt \\ &= -\frac{y}{\lambda} \int_0^1 1_{(z,z+y]}(x+sy) ds \end{aligned}$$

und

$$|f(x+y) - f(x)| \leq |f|_\infty |y|$$

gilt.
Weiter gilt nun mit $S_{n,i} := \frac{1}{\sqrt{n}} \sum_{\substack{j=1, \\ j \neq i}}^n X_j = S_n - \frac{X_i}{\sqrt{n}}$

$$\begin{aligned} &\mathbb{E}\left[h(S_n)\right] - \Phi(h) \\ &= \mathbb{E}\left[f'(S_n) - S_n f(S_n)\right] \\ &= \mathbb{E}\left[f'(S_n) - \sqrt{n} X_n f(S_n)\right] \\ &= \frac{1}{n} \sum_{i=1}^n \mathbb{E}\left[f'(S_n) - \sqrt{n} X_i f(S_n)\right] \\ &= \frac{1}{n} \sum_{i=1}^n \mathbb{E}\left[f'(S_n) - f'(S_{n,i})\right] + \frac{1}{n} \sum_{i=1}^n \mathbb{E}\left[f'(S_{n,i}) - \sqrt{n} X_i f(S_n)\right] \\ &\overset{(***)}{=} \frac{1}{n} \sum_{i=1}^n \mathbb{E}\left[f'(S_n) - f'(S_{n,i})\right] - \frac{1}{n} \sum_{i=1}^n \mathbb{E}\left[\sqrt{n} X_i \left(f(S_n - f(S_{n,i})\right) - X_i^2 f'(S_{n,i})\right] \\ &\overset{(****)}{=} \frac{1}{n} \sum_{i=1}^n \mathbb{E}\left[f'(S_n) - f'(S_{n,i})\right] - \frac{1}{n} \sum_{i=1}^n \mathbb{E}\left[X_i^2 \int_0^1 \left(f'\left(S_{n,i} + t \frac{X_i}{\sqrt{n}}\right) - f'(S_{n,i})\right) dt\right] \end{aligned} \tag{2}$$

.

Wobei wir in Schritt (***)

$$\mathbb{E}\left[f'(S_{n,i})\right] = \mathbb{E}[X_i^2] \mathbb{E}[f'(S_{n,i})] = \mathbb{E}[X_i^2 f'(S_{n,i})]$$

und

$$\mathbb{E}\left[\sqrt{n} X_i f(S_{n,i})\right] = \sqrt{n} \underbrace{\mathbb{E}[X_i]}_{0} \mathbb{E}[f(S_{n,i})] = 0$$

genutzt haben. In Schritt (****) haben wir ausgenutzt, dass

$$f(S_n) - f(S_{n,i}) = \int_{S_{n,i}}^{S_n} f'(y)dy = \frac{X_i}{\sqrt{n}} \int_0^1 f'\left(S_{n,i} + t\frac{X_i}{\sqrt{n}}\right) dt$$

und

$$X_i^2 f'(S_{n,i}) = X_i^2 \int_0^1 f'(S_{n,i})dt$$

gilt.

Nun wollen wir den ersten Teil von Ausdruck (2) weiter abschätzen. Es gilt

$$\begin{aligned}
&|\mathbb{E}\left[f'(S_n) - f'(S_{n,i})\right]| \\
&\leq \mathbb{E}\left|\left[f'(S_n) - f'(S_{n,i})\right|\right] \\
&\stackrel{(1)}{\leq} \mathbb{E}\left[\frac{|X_i|}{\sqrt{n}}\left(1 + 2|S_{n,i}| + \frac{1}{\lambda}\int_0^1 1_{(z,z+\lambda]}\left(S_{n,i} + s\frac{X_i}{\sqrt{n}}\right) ds\right)\right] \\
&= \mathbb{E}\left[\mathbb{E}\left[\frac{|X_i|}{\sqrt{n}}\left(1 + 2|S_{n,i}| + \frac{1}{\lambda}\int_0^1 1_{(z,z+\lambda]}\left(S_{n,i} + s\frac{X_i}{\sqrt{n}}\right) ds\right)\right] |X_i\right] \\
&= \mathbb{E}\left[\frac{|X_i|}{\sqrt{n}}\mathbb{E}\left[1 + 2|S_{n,i}| + \frac{1}{\lambda}\int_0^1 1_{(z,z+\lambda]}\left(S_{n,i} + s\frac{X_i}{\sqrt{n}}\right) ds\right] |X_i\right] \\
&= \mathbb{E}\left[\frac{|X_i|}{\sqrt{n}}\left(1 + 2\mathbb{E}[|S_{n,i}||X_i] + \mathbb{E}\left[\frac{1}{\lambda}\int_0^1 1_{(z,z+\lambda]}\left(S_{n,i} + s\frac{X_i}{\sqrt{n}}\right) ds|X_i\right]\right)\right] \\
&= \mathbb{E}\left[\frac{|X_i|}{\sqrt{n}}\left(1 + 2\mathbb{E}[|S_{n,i}|] + \frac{1}{\lambda}\mathbb{E}\left[\int_0^1 1_{(z,z+\lambda]}\left(S_{n,i} + s\frac{X_i}{\sqrt{n}}\right) ds|X_i\right]\right)\right]. \qquad (3)
\end{aligned}$$

Wobei wir die Unabhängigkeit von $S_{n,i}$ und X_i und somit $\mathbb{E}[|S_{n,i}||X_i] = E[|S_{n,i}|]$ sowie die Linearität des bedingten Erwartungswertes ausgenutzt haben. Betrachten wir nun wieder den letzten Term der so erreichten Ungleichung. Für diesen gilt

$$\begin{aligned}
&\mathbb{E}\left[\int_0^1 1_{(z,z+\lambda]}\left(S_{n,i} + s\frac{X_i}{\sqrt{n}}\right) ds|X_i = x\right] \\
&= \mathbb{E}\left[\int_0^1 1_{(z,z+\lambda]}\left(S_{n,i} + s\frac{x}{\sqrt{n}}\right) ds\right] \\
&\stackrel{\text{Fubini}}{=} \int_0^1 \mathbb{E}\left[1_{(z,z+\lambda]}\left(S_{n,i} + s\frac{x}{\sqrt{n}}\right) ds\right] \\
&\leq \int_0^1 \left(\sqrt{\frac{n}{2\pi(n-1)}}\lambda + 2\delta(\gamma, n-1)\right) ds \\
&= \sqrt{\frac{n}{2\pi(n-1)}}\lambda + 2\delta(\gamma, n-1).
\end{aligned}$$

Bei dieser Umformung haben wir ausgenutzt, dass aufgrund der Jensen-Ungleichung $\mathbb{E}[|S_{n,i}] \leq \sqrt{\mathbb{E}[S_{n,i}^2]} = \sqrt{\frac{n-1}{1}} \leq \sqrt{2} < 1$ und $\mathbb{E}[|X_i|] \leq \sqrt{\mathbb{E}[X_i^2]} = 1$ gilt.
Mit diesen Ergebnisen wollen wir nun Ausdruck (3) noch weiter abschätzen.

$$\begin{aligned}
&|\mathbb{E}\left[f'(S_n) - f'(S_{n,i})\right]| \\
&\leq \mathbb{E}\left[\frac{|X_i|}{\sqrt{n}}\left(1 + 2 + \frac{1}{\lambda}\left(\sqrt{\frac{n}{2\pi(n-1)}}\lambda + 2\delta(\gamma, n-1)\right)\right)\right] \\
&\leq \frac{1}{\sqrt{n}}\left(1 + 2 + \sqrt{\frac{n}{2\pi(n-1)}}\lambda + \frac{2}{\lambda}\delta(\gamma, n-1)\right) \\
&\leq \frac{1}{\sqrt{n}}\left(1 + 2 + \frac{1}{\sqrt{\pi}} + \frac{2}{\lambda}\delta(\gamma, n-1)\right).
\end{aligned}$$

Auf analoge Art und Weise lässt sich der zweite Term von (2) wie folgt abschätzen.

$$\left|\mathbb{E}\left[X_i^2 \int_0^1 \left(f'\left(S_{n,i} + t\frac{X_i}{\sqrt{n}}\right) - f'(S_{n,i})\right) dt\right]\right| \leq \tfrac{\gamma}{\sqrt{n}}\left(\tfrac{1+2}{2} + \tfrac{1}{2\sqrt{\pi}} + \tfrac{1}{\lambda}\delta(\gamma, n-1)\right)$$

Somit erhalten wir insgesamt

$$\begin{aligned}
&\mathbb{E}\left[h(S_n)\right] - \Phi(h) \\
&\leq \frac{1}{\sqrt{n}}\left(1 + 2 + \frac{1}{\sqrt{\pi}} + \frac{2}{\lambda}\delta(\gamma, n-1)\right) + \frac{\gamma}{\sqrt{n}}\left(\frac{1+2}{2} + \frac{1}{2\sqrt{\pi}} + \frac{1}{\lambda}\delta(\gamma, n-1)\right) \\
&\overset{\gamma \geq 1}{\leq} \frac{\gamma}{\sqrt{n}}\left(\frac{3}{2}(1+2) + \frac{3}{2\sqrt{\pi}} + \frac{3}{\lambda}\delta(\gamma, n-1)\right) \\
&= \frac{3\gamma}{\sqrt{n}}\left(\frac{3}{2} + \frac{1}{2\sqrt{\pi}} + \frac{1}{\lambda}\delta(\gamma, n-1)\right).
\end{aligned}$$

Somit gilt insbesondere

$$\delta(\lambda, \gamma, n) \leq \tfrac{3\gamma}{\sqrt{n}}\left(\tfrac{3}{2} + \tfrac{1}{2\sqrt{\pi}} + \tfrac{1}{\lambda}\delta(\gamma, n-1)\right).$$

Damit folgt für $\lambda > 0$, $\forall n \in \mathbb{N}$ nach Satz 5.0.2

$$\delta(\gamma, n) \leq \tfrac{3\gamma}{\sqrt{n}}\left(\tfrac{3}{2} + \tfrac{1}{2\sqrt{\pi}} + \tfrac{1}{\lambda}\delta(\gamma, n-1)\right) + \tfrac{\lambda}{2\sqrt{2\pi}}.$$

Per Induktion (welche wir an dieser Stelle nicht führen wollen) lässt sich zeigen, dass

$$\begin{aligned}
\delta(\gamma, n) &\leq \left(3(1+2) + \frac{9}{\sqrt{\pi}}\right)\frac{\gamma}{\sqrt{n}} \\
&\leq 14,1\frac{\gamma}{\sqrt{n}}
\end{aligned}$$

gilt.
Somit ist es uns gelungen eine Berry-Esseen Rate der Ordnung $o(n^{1/2})$ zu zeigen.

Bemerkung 5.0.3. *Durch kleine Abänderungen in der Beweisführung lässt sich sogar eine universelle Konstante von 13 erreichen.*

Fazit

Als Abschluss dieser Arbeit lässt sich sagen, dass die Stein'sche Methode eine geeignete und aussagekräftige Alternative bildet um Grenzwertsätze und ähnliches zu beweisen. Ein großer Vorteil dieser Methode im Vergleich zu den bisher bekannten und auszugsweise am Ende des zweiten Kapitels vorgestellten ist, dass sie meißt mit relativ geringem Aufwand Konvergenzraten liefert.
Jedoch bleibt andererseits die Stärke der Stein'schen Methode etwas fragwürdig, da einige Probleme offen bleiben, so ist beispielsweise in einigen Fällen die Kontrolle der Stein-Gleichung sehr aufwendig. Auch ist die Stein'sche Methode in einigen anderen Bereichen wie den multivariaten Verteilungen noch recht wenig erforscht.
Nichtdestotrotz findet sie bereits breite Anwendung in Bereichen wie der räumlichen Statistik, beim Bootstrap-Verfahren oder in der Finanz- und Versicherungsmathematik. Abschließend bleibt zu sagen, dass die Stein'sche Methode noch eine Reihe von Weiterentwicklungsmöglichkeiten bietet.

Literaturverzeichnis

[1] Bauer, H. (1990), *Wahrscheinlichkeitstheorie*, de Gruyter Lehrbuch

[2] Bolthausen, E. (1984), *An Estimate of the Remainder in a Combinatorial Central Limit Theorem*, Zeitung für Wahrscheinlichkeitstheorie und verwandte Gebiete - Heft 66

[3] Chen, L. H. Y., Goldstein, L., Shao, Q. (2010), *Normal Approximation by Stein's Method (Probability and Its Applications)*, Springer

[4] Chen, L. H. Y., HO, S. T. (1978), *An L_p bound for the reminder in a combinatorial central limit theorem* , Ann. Probability 6

[5] Eichelsbacher, P. (2008), *Einführung in die Steinsche Methode*, Ruhr-Universität Bochum

[6] Eichelsbacher, P. (2003), *Die Steinsche Methode*, Ruhr-Universität Bochum

[7] Klenke, A. (2008), *Wahrscheinlichkeitstheorie*, Springer

[8] Schmidt, K. D. (2011), *Maß und Wahrscheinlichkeit*, Springer

[9] Shiryaey, A. N. (1995), *Probability*, Springer

[10] Stein, C. (1972). *A bound for the error in the normal approximation to the distribution of a sum of dependent random variables*, Proc. Sixth Berkeley Symp. on Math. Statist. and Prob.

[11] Vaart, A. W. van der (1998), *Asymptotic Statistics*, Cambridge University Press

Erklärung

Hiermit versichere ich, dass ich diese Bachelorarbeit eigenständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe. Diese Arbeit hat in dieser oder einer ähnlichen Form noch nicht im Rahmen einer anderen Prüfung vorgelegen.

Marburg, den 27. September 2012

(Tim Josek, Verfasser)