Eine Einführung in die Steinsche Methode

Peter Eichelsbacher

Ruhr-Universität Bochum

Sommer-Schule in Oggebbio am Lago Maggiore; September 2008

Inhaltsverzeichnis

Zusammenfassung

Das vorliegende Manuskript basiert auf fünf Vorträgen, die im September 2008 in Oggebbio/Oberitalien in Form eines Minikurses für Doktoranden und Postdoktoranden im Rahmen des SFB/TR 12 gehalten wurden. Ziel war es, einen Eindruck von der STEINschen Methode zu geben. Diese Methode dient der Herleitung von Grenzwertsätzen in der Wahrscheinlichkeitstheorie und verwandten Gebieten. Die Methode erlaubt es, simultan auch Konvergenzgeschwindigkeiten zu untersuchen. Die Methode wurde 1972 von CHARLES STEIN eingeführt. Der vorliegende Kurs betrachtet neben Grundlagen vor allem diejenigen Aspekte, die bei Fragestellungen im SFB-TR Projekt C2 untersucht werden. Wir erinnern in Kapitel 1 zunächst an *klassische Methoden* zur Beweisführung für Normal-Approximation und stellen die Grundzüge der Methode von STEIN vor. In Kapitel 2 stellen wir die STEINSCHE Methode bei Normal-Approximation für den Fall eines *Partialsummenprozesses* von unabhängigen, nicht notwendig identisch verteilten Zufallsvariablen vor. Wir geben in Kapitel 2 desweiteren einen Ausblick auf das Thema der POISSON-Approximation. In Kapitel 3 wird der Zugang mittels *austauschbarer Paare* vorgestellt und an Beispielen vertieft. Die STEINsche Methode wird auch bei Approximation durch andere Verteilungen verwendet. Wir betrachten in Kapitel 4 Beispiele aus der *statistischen Mechanik* sowie der Theorie der *Zufallsmatrizen* und skizzieren den Ansatz von STEIN bei der Untersuchung von Grenzwertsätzen in diesen Gebieten (CURIE-WEISS Modelle, WIGNER-Matrizen). In Kapitel 5 stellen wir einen neueren Ansatz zur Herleitung von *Konzentrations-Ungleichungen* mittels austauschbarer Paare vor.

Dem SFB/TR 12 sei für die Finanzierung der Tagung in Oggebbio (Oberitalien-Lago Maggiore), die in der Zeit vom 22. September - 26. September 2008 stattfand, herzlich gedankt.

1 Normal-Approximation

In der klassischen Ausbildung zur Beweisführung eines zentralen Grenzwertsatzes (z.B. für einen Partialsummenprozess von unabhängigen und identisch verteilten Zufallsvariablen) verwendet man in der Regel einer der drei erst genannten Beweis-Methoden, in jüngster Zeit auch die vierte Methode, um die es in diesem Kurs gehen soll:

- charakteristische Funktion
- Momente-Methode
- LINDEBERG-Methode
- STEINsche Methode, partielle Integration

Wir starten mit einem kleinen Exkurs entlang klassischer Wahrscheinlichkeitstheorie, siehe etwa [2]. Zunächst erinnern wir an *Konvergenz in Verteilung*: Es sei $(\Omega, \mathcal{F}, P)$ ein Wahrscheinlichkeitsraum und

$$X : \ (\Omega, \mathcal{F}, P) \to \mathbb{R}$$

eine Zufallsvariable (bei reellwertigen Zufallsvariablen sprechen wir in der Folge immer von Zufallsgrößen) und

$$F(t) := P(X \leq t) \quad \text{die zugehörige Verteilungsfunktion.}$$

Sei weiter eine Folge von Zufallsvariablen

$$(X_n)_n \text{ mit Verteilungsfunktionen } \ (F_n)_n$$

gegeben. Man sagt, dass X_n *in Verteilung gegen X konvergiert*, wenn

$$\lim_n F_n(t) = F(t) \qquad \text{für alle Stetigkeitspunkte von F gilt,}$$

in Zeichen $X_n \Rightarrow X$. Klassisch ist der folgende Satz (ohne Beweis):

Satz 1 *Es bezeichne* $\mathbb{E}$ *den Erwartungswert bezüglich* P. *Es sind äquivalent:*

(i) $X_n \Rightarrow X$

(ii) $\mathbb{E}f(X_n) \to \mathbb{E}f(X) \quad \forall f \in C_b$ *(schwache Konvergenz)*

(iii) $\mathbb{E}f(X_n) \to \mathbb{E}f(X) \quad \forall f$ *beschränkt und* LIPSCHITZ

Der *zentrale Grenzwertsatz* kann nun so formuliert werden:

Satz 2 *Es seien* $X_1, X_2, \ldots$ *unabhängig, identisch verteilte Zufallsgrößen. Sei*

$$\mathbb{E}(X_1) = 0; \quad \mathbb{E}(X_1^2) = \sigma^2; \quad S_n := \sum_{i=1}^{n} X_i; \quad W_n := \frac{1}{\sigma\sqrt{n}} S_n.$$

Dann gilt

$$P(W_n \leq t) \to \Phi(t) := \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{t} e^{-x^2/2} dx$$

(Zentraler Grenzwertsatz).

Man kann grob sagen, dass eine Aussage wie die des zentralen Grenzwertsatzes in vielen *nicht-unabhängigen* und auch in *nicht-identisch verteilten* Situationen ebenfalls gilt. Viele Zufallsgrößen sind strukturell aufgebaut wie W_n und des gilt (sehr grob): „Summen vieler kleiner Dinge, die einander wenig beeinflussen “, sind in der Limes-Verteilung nicht vom obigen Fall zu unterscheiden.

1.1 charakteristische Funktion

Eine klassische Methode der Beweisführung eines zentralen Grenzwertsatzes ist die Verwendung *charakteristischer Funktionen* (FOURIER-Transformierte): Zu einer Zufallsgröße X ist

$$\varphi_X(t) := \mathbb{E}\left[\exp(itX)\right]$$

die charakteristische Funktion. Es gilt:

$$\varphi_{X_n}(t) \to \varphi_X(t) \quad \forall t \in \mathbb{R} \quad \Leftrightarrow \quad X_n \Rightarrow X$$

(Stetigkeitssatz von Lévy).
Sicher kann man sagen:

- Dies ist eine viel verwendete Methode.
- Die $(X_i)_i$ müssen nicht identisch verteilt sein.
- Es sind auch Konvergenzraten ermittelbar.
- Man betrachtet charakteristische Funktionen nicht nur für Normalapproximation.

Es sei

$$Z \sim N(0,1): \quad \varphi_Z(t) = e^{-t^2/2}.$$

Es gilt

$$X_1, X_2 \text{ unabhängig}: \ \varphi_{X_1+X_2}(t) = \varphi_{X_1}(t)\varphi_{X_2}(t).$$

Skizze des Beweises eines zentralen Grenzwertsatzes:

$$\begin{aligned}
\varphi_{W_n} &= \varphi_{\frac{X_1}{\sigma\sqrt{n}}+\cdots+\frac{X_n}{\sigma\sqrt{n}}}(t) = \left(\varphi_{\frac{X_1}{\sigma\sqrt{n}}}(t)\right)^n \\
&= \left(1 + \frac{it}{\sigma\sqrt{n}}\underbrace{\mathbb{E}(X_1)}_{=0} + \frac{(it)^2}{2(\sigma\sqrt{n})^2}\underbrace{\mathbb{E}(X_1^2)}_{=\sigma^2} + \cdots\right)^n \\
&\cong \left(1 - \frac{t^2}{2n}\right)^n \to e^{-t^2/2} = \varphi_Z(t).
\end{aligned}$$

genauer:

$$= \left(1 - \frac{t^2}{2n} + o\left(\frac{1}{n}\right)\right)^n \to e^{-t^2/2} = \varphi_Z(t).$$

1.2 Momente-Methode

Eine weitere klassische Methode ist die *Momente Methode*:
Es sei $Z \sim N(0,1)$; und $(X_n)_n : X_n$ „habe alle Momente“:

$$\mathbb{E}(X_n^k) < \infty \quad \forall k \in \mathbb{N}.$$

Angenommen wir können

$$\mathbb{E}(X_n^r) \to \mathbb{E}(Z^r) = \begin{cases} (r-1)!! & r \text{ gerade} \\ 0 & \text{sonst} \end{cases} \qquad \forall r \in \mathbb{N}$$

zeigen, dann folgt $X_n \Rightarrow Z$. Ein paar Bemerkungen:

- Erstmals wurde die Momentemethode 1887 von Tschebychev verwendet.
- Die Methode wird gern vermieden: sie ist „schwer“, da sie auf kombinatorisch anspruchsvolle Fragen führt.
- Perci Diaconis sagte einmal: „There are some however who realize that moments generally get the job done without taking five years off to develop special theory“.
- Die Methode liefert keine Konvergenzraten.
- Es müssen *alle* Momente konvergieren.

Wir skizzieren die Methode im Fall unabhängiger, identische verteilter X_i mit $\mathbb{E}(X_1) = 0$ und $\mathbb{E}(X_1^2) = 1$. Wieder sei

$$W_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} X_i.$$

Dann ist

$$\mathbb{E}(W_n) = 0, \quad \mathbb{E}(W_n^2) = 1$$

und

$$\mathbb{E}(W_n^4) = \frac{1}{n^2} \sum_{i_1,i_2,i_3,i_4=1}^{n} \mathbb{E}(X_{i_1} X_{i_2} X_{i_3} X_{i_4}). \tag{1}$$

Nun verschwinden die folgenden Erwartungswerte (verwende Unabhängigkeit und $\mathbb{E}(X_j) = 0$).

$$\mathbb{E}(X_i X_j X_k X_l) = 0, \quad \mathbb{E}(X_i X_j X_k X_k) = 0, \ldots, \mathbb{E}(X_i X_i X_i X_j) = 0.$$

Die Terme $\mathbb{E}(X_i^4)$ sind in der Regel verschieden von Null, wir haben aber nur n dieser Terme. Weiter ist (verwende Unabhängigkeit und $\mathbb{E}(X_i^2) = 1$)

$$\mathbb{E}(X_i X_i X_j X_j) = 1, \quad \mathbb{E}(X_i X_j X_j X_i) = 1, \quad \mathbb{E}(X_i X_j X_i X_j) = 1,$$

wovon je n^2 Terme in obiger Summe vorkommen. Die Summe wird also soviele Beiträge der Größe 1 (im Limes $n \to \infty$) liefern, wieviele Möglichkeiten es gibt, die Menge $\{1, \dots, 4\}$ in 2 Paare zu zerlegen. Also folgt $\lim_{n\to\infty} \mathbb{E}(W_n^4) = 3 = \mathbb{E}(Z^4)$. Allgemeiner konvergiert $\mathbb{E}(W_n^{2k})$ im Limes für $n \to \infty$ gegen die Anzahl der Möglichkeiten, die Menge $\{1, \dots, 2k\}$ in k Paare zu zerlegen. Diese Anzahl ist

$$1 \cdot 3 \cdot 5 \cdots (2k-1) = (2k-1)!!,$$

stimmt also mit dem $2k$-ten Moment der Standard-Normalverteilung überein.

1.3 Lindeberg-Methode

Es seien $Z_1, \dots, Z_n$ unabhängig, identisch $N(0,1)$ verteilt, dann gilt

$$Z = \frac{1}{\sqrt{n}} (Z_1 + \cdots + Z_n) \quad \text{ist } N(0,1) - \text{verteilt.}$$

Die Idee von Lindeberg: approximiere obiges W_n durch Z, indem X_i nacheinander durch Z_i ersetzt wird!

Bemerkung 1 Dies ermöglicht Verallgemeinerungen weg von i.i.d.-Partialsummen.

Bemerkung 2 Sind die X_i unabhängig und identisch verteilte Zufallsgrößen, so gilt:

$$\frac{1}{n} \sum_{i=1}^{n} \int_{\left\{\frac{|X_i|}{\sqrt{n}} > \varepsilon\right\}} x_i^2 d\mu(x_i) \underset{n\to\infty}{\longrightarrow} 0,$$

die Lindeberg-Bedingung, wobei μ die Verteilung der X_i bezeichne.

Wir fordern die Lindeberg-Bedingung anstatt fordern zu müssen, dass die X_i identisch verteilt sind. Wir fordern aber weiterhin unabhängige Summanden.

Wir zeigen (und dies liefert Konvergenz in Verteilung bzw. schwache Konvergenz):

$$\mathbb{E}f(W_n) \to \mathbb{E}f(Z) \quad \forall f \in C_c^\infty.$$

Es gilt mit Hilfe von TAYLOR:
Es existiert ein k und ein $g(h)$ mit $g(h) \leq k \min\{h^2, h^3\}$:

$$\begin{aligned}&\left|f(x+h_1) - f(x+h_2) - f'(x)(h_1 - h_2) - \frac{1}{2}f''(x)(h_1^2 - h_2^2)\right| \\ &\leq g(h_1) + g(h_2).\end{aligned}$$

Wir setzen:

$$T_k := X_1 + \cdots + X_{k-1} + Z_{k+1} + \cdots + Z_n,$$

also

$$T_n + X_n = S_n \quad \text{und} \quad T_1 + Z_1 \sim N(0, n)$$

Es folgt (Teleskop-Summe)

$$\begin{aligned}|\mathbb{E}\left(f(W_n) - f(Z)\right)| &\leq \sum_{k=1}^n \left|\mathbb{E}\left[f\left(\frac{T_k + X_k}{\sqrt{n}}\right) - f\left(\frac{T_k + Z_k}{\sqrt{n}}\right)\right]\right| \\ &= \sum_{k=1}^n \left|\mathbb{E}\left[f\left(\frac{T_k + X_k}{\sqrt{n}}\right) - f\left(\frac{T_k + Z_k}{\sqrt{n}}\right)\right.\right. \\ &\qquad \left.\left. - f'(T_k)\left(\frac{X_k - Z_k}{\sqrt{n}}\right) - \frac{1}{2}f''(T_k)\left(\frac{X_k^2 - Z_k^2}{n}\right)\right]\right|,\end{aligned}$$

denn

$$\mathbb{E}(X_k) = \mathbb{E}(Z_k) \quad \text{und} \quad \mathbb{E}(X_k^2) = \mathbb{E}(Z_k^2).$$

Also folgt für obige Identität die Abschätzung

$$\leq \sum_{k=1}^n \left(\mathbb{E}g\left(\frac{X_k}{\sqrt{n}}\right) + g\left(\frac{Z_k}{\sqrt{n}}\right)\right) = n\mathbb{E}g\left(\frac{X_1}{\sqrt{n}}\right) + n\mathbb{E}g\left(\frac{Z_1}{\sqrt{n}}\right).$$

Nun betrachten wir den ersten Summanden:

$$n\mathbb{E}g\left(\frac{X_1}{\sqrt{n}}\right) \leq n\left[k\int_{\{|x|\leq\varepsilon\sqrt{n}\}}\left|\frac{x}{\sqrt{n}}\right|^3 d\mu(x) + k\int_{\{|x|>\varepsilon\sqrt{n}\}}\left|\frac{x}{\sqrt{n}}\right|^2 d\mu(x)\right],$$

wobei der zweite Summand für $n\to\infty$ und $\forall\varepsilon>0$ gegen Null geht mit Hilfe der LINDEBERG-Bedingung
und

$$nk\int_{\{|x|\leq\varepsilon\sqrt{n}\}}\left|\frac{x}{\sqrt{n}}\right|^3 d\mu(x) \leq nk\varepsilon\int_{\{|x|\leq\varepsilon\sqrt{n}\}}\left|\frac{x}{\sqrt{n}}\right|^2 d\mu(x) \leq k\varepsilon.$$

Analog für

$$n\mathbb{E}g\left(\frac{Z_1}{\sqrt{n}}\right)\dots$$

□

1.4 Steinsche Methode

Nach diesem Exkurs kommen wir nun zur STEINsche Methode (CHARLES STEIN, 1972). Wir nennen hier nur zwei Quellen, in denen man die Grundzüge der Methode nachlesen kann, und zwar [14] und [6].
f habe einen kompakten Träger in (a,b) und sei differenzierbar, dann gilt

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

mittels *partieller Integration*. Es folgt also für $Z\sim N(0,1)$ und $f:\mathbb{R}\to\mathbb{R}$ differenzierbar mit $\mathbb{E}|f'(Z)|<\infty$:

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

Man kann sogar zeigen: gilt $\mathbb{E}\left[f'(Z)-Zf(Z)\right]=0$ für schöne f mit $\mathbb{E}\left[f'(Z)\right]<\infty$ und $\mathbb{E}\left[Zf(Z)\right]<\infty$, so folgt $Z\sim N(0,1)$ (ohne Beweis).
Was ist so nützlich an dieser Beobachtung?

- Wir wollen klären, wie „weit"die Verteilung von W_n von $Z\sim N(0,1)$

entfernt ist!

- Allgemeiner: wir wollen den Abstand zwischen Wahrscheinlichkeitsverteilungen messen!

Intermezzo: *Abstände von Wahrscheinlichkeitsmaßen*:
Es seien μ, ν Wahrscheinlichkeitsmaße (auf $\mathbb{R}$). Wir setzen

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

mit einer Funktionenklasse $\mathcal{D}$.

wichtige Beispiele:

- Totalvariation:

$$\begin{aligned} &\mathcal{D} = \{1_A : A \in \mathcal{B}\}\,; \\ &\mathrm{TV}(\mu, \nu) := \sup_{A \in \mathcal{B}} |\mu(A) - \nu(A)|. \end{aligned}$$

- Wasserstein-Metrik: $\mathcal{D}$ ist die Menge der 1-Lipschitz Funktionen:

$$\mathrm{Wass}(\mu, \nu) := \sup \left\{ \left| \int f d\mu - \int f d\nu \right| : \ f\ 1 - \text{Lipschitz} \right\}.$$

- Kolmogorov-Abstand:

$$\begin{aligned} &\mathcal{D} = \left\{ 1_{(-\infty, x]}, x \in \mathbb{R} \right\} \\ &\mathrm{Kolm}(\mu, \nu) := \sup_{x \in \mathbb{R}} |\mu(-\infty, x] - \nu(-\infty, x]| \leq \mathrm{TV}(\mu, \nu). \end{aligned}$$

Alle drei Metriken sind stärker als die schwache Konvergenz!

Klassisch ist das folgende Theorem von Berry-Esseen:

Satz 3 *Sind die* $(X_i)_i$ *unabhängig und identisch verteilt mit* $\mathbb{E}(X_1) = 0$ *und* $\mathbb{E}(X_1^2) = 1$ *sowie* $\mathbb{E}|X_1|^3 \leq \infty$, *dann gilt*

$$\text{Kolm}(W_n, Z) \leq \frac{const}{\sqrt{n}}.$$

Unser Ziel ist nun: wir wollen die obige Charakterisierung von $N(0,1)$ (nach Stein) nutzen, um einen zentralen Grenzwertsatz und *simultan* das Resultat von Berry-Esseen zur Rate der Konvergenz zu beweisen!

Es sei W eine Zufallsvariable. Wir nehmen an, dass W durch $Z \sim N(0,1)$ gut approximierbar sei. Wir wollen einen Abstand untersuchen:

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

Dies wollen wir nun in Verbindung bringen mit

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

zu einem geeigneten $\mathcal{D}'$. Der letzte Ausdruck sollte „klein“ sein!

Stein's Ansatz kann nun wie folgt beschrieben werden: Sei $\mathcal{D}'$ „die“ Klasse von Funktionen, so dass für jedes $g \in \mathcal{D}$ ein $f := f_g \in \mathcal{D}'$ existiert mit

$$f'(x) - xf(x) = g(x) - \mathbb{E}g(Z), \qquad \text{(Stein-Gleichung mit Input-Funktion } g)$$

so gilt

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

also

$$\sup_{g \in \mathcal{D}} |\mathbb{E}g(W) - \mathbb{E}g(Z)| \leq \sup_{f \in \mathcal{D}'} |\mathbb{E}\left(f'(W) - Wf(W)\right)| \, . \tag{2}$$

Hierbei sollte offensichtlich $\mathcal{D}'$ „möglichst klein“gewählt sein; $f(\cdot), f'(\cdot)$ sind die Lösung und ihre Ableitung der Stein-Gleichung; W beherbergt das stochastisches Modell.

Nun müssen wir uns um die rechte Seite in (2) kümmern.

Die Lösungen der Stein-Gleichung hat man gut im Griff:

Satz 4 *(Stein)*

(a) $g : \mathbb{R} \to \mathbb{R}$ sei beschränkt: es existiert eine fast-sicher differenzierbare Funktion $f : \mathbb{R} \to \mathbb{R}$ mit $f'(x) - xf(x) = g(x) - \mathbb{E}g(Z)\ \forall x$. Es gilt

$$|f|_\infty \leq \sqrt{\frac{\pi}{2}} |g - \mathbb{E}g(Z)|_\infty$$
$$|f'|_\infty \leq 2|g - \mathbb{E}g(Z)|_\infty.$$

(b) Ist $g : \mathbb{R} \to \mathbb{R}$ Lipschitz, nicht notwendig beschränkt, so gilt

$$|f|_\infty \leq |g'|_\infty, \; |f'|_\infty \leq \sqrt{\frac{2}{\pi}} |g'|_\infty, \; |f''|_\infty \leq 2|g'|_\infty.$$

(alle Konstanten sind optimal)

- Um die Lösungen haben wir uns gekümmert (unabhängig von W).
- Die Lösung f_g zu g kann einfach gefunden werden; wir besprechen Teile des Theorems später.
- Der **Erfolg** der Methode von Stein liegt darin, dass sich $\mathbb{E}\left[f'(W) - Wf(W)\right]$ tatsächlich in vielen Fällen „gut“ abschätzen lässt.
- Die Methode ist erfolgreich, da die Unabhängigkeit der Bausteine von W nicht notwendig sein wird (was bei Anwendung der Theorie der charakteristische Funktionen auch so ist, ohne das wir dies ausgeführt haben).
- Die Konvergenzraten „come for free“ für verschiedene Metriken bei Verwendung der Steinschen Methode. Das ist ein feiner(!) Unterschied zur Verwendung charakteristischer Funktionen oder der Momente-Methode.
- Die Methode von Stein kann für andere Verteilungen, die mittels eines geeigneten Differential-/Differenzenoperators charakterisiert werden können, entwickelt und angewendet werden.

Der Zwischenstand unserer bisherigen Betrachtungen kann wie folgt zusammengefaßt werden:

$$\begin{aligned}&\mathrm{Wass}(W,Z) := \mathrm{Wass}(\mathcal{L}(W),\mathcal{L}(Z))\\ &\leq \sup\left\{|\mathbb{E}\left(f'(W)-Wf(W)\right)|:\ |f|_\infty \leq 1, |f'|_\infty \leq \sqrt{2/\pi}, |f''|_\infty \leq 2\right\}.\end{aligned}$$

Nun werden wir

- Ad hoc werden wir $|\mathbb{E}f'(W) - W\mathbb{E}f(W)|$ im Fall einer Partialsumme von unabhängigen Zufallsgrößen äbschätzen.
- Wir werden $|\mathbb{E}f'(W) - W\mathbb{E}f(W)|$ mit einer speziellen Methode abschätzen: die Methode der *austauschbaren Paare* (wir werden Anwendungen dazu skizzieren).
- Steins Theorem zur Abschätzung der Lösung der Stein-Gleichung (und ihrer Ableitungen) diskutieren.

2 Steinsche Methode in Aktion

2.1 Partialsummen mit unabhängigen Summanden

Zunächst seien die Zufallsgrößen $(X_i)_i$ unabhängig mit $\mathbb{E}(X_i) = 0$, $\mathbb{E}(X_i^2) = 1$ und $\mathbb{E}|X_i|^3 < \infty$. Sei erneut

$$W_n := \frac{1}{\sqrt{n}}\sum_{i=1}^n X_i.$$

Weiter sei f so, dass $|f| \leq 1$, $|f'| \leq \sqrt{2/\pi}$, $|f''| \leq 2$. Dann gilt $\mathbb{E}(Wf(W)) = \frac{1}{\sqrt{n}}\sum_{i=1}^n \mathbb{E}(X_i f(W))$. Wir wollen die Unabhängigkeit nutzen und definieren $W_i := W - \frac{X_i}{\sqrt{n}}$, womit X_i und W_i unabhängig sind für jedes i. Also folgt

$$\mathbb{E}(X_i f(W_i)) = \underbrace{\mathbb{E}(X_i)}_{=0}\mathbb{E}(f(W_i)) = 0.$$

Somit ergibt sich

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

Da

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

und

$$W - W_i = \frac{X_i}{\sqrt{n}},$$

folgt

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

und

$$\mathbb{E}[X_i(W - W_i)f'(W_i)] = \frac{1}{\sqrt{n}}\mathbb{E}[X_i^2 f'(W_i)] = \frac{1}{\sqrt{n}}\mathbb{E}[f'(W_i)].$$

Also folgt

$$\left|\mathbb{E}(Wf(W)) - \frac{1}{n}\sum_{i=1}^{n}\mathbb{E}[f'(W)]\right| \leq \frac{1}{\sqrt{n}}\frac{1}{n}\sum_{i=1}^{n}\mathbb{E}|X_i|^3. \tag{1}$$

Weiter ist

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

und mit (1) und (2) folgt daher insgesamt

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

Damit ergibt sich der folgende Satz:

Satz 1 *Sind $X_1, \ldots, X_n$ unabhängig und zentriert mit $\mathit{Var}(X_i) = 1$. Weiter existiere das dritte absolute Moment, dann gilt:*

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

(im Falle identisch verteilter Zufallsgrößen folgt dann : $\leq \frac{3\mathbb{E}|X_1|^3}{\sqrt{n}}$)

Wir halten fest:

- Wir haben insbesondere einen zentralen Grenzwertsatz bewiesen.
- Wir haben zumindest bezüglich der Wasserstein-Metrik eine Konvergenzrate hergeleitet.
- Aber wir haben (derzeit) keine optimale Berry-Esseen-Rate, denn man beachte:

Lemma 1

$$\text{Kolm}(W_n, Z) \leq \frac{2}{(2\pi)^{1/4}} \left(\text{Wass}(W_n, Z)\right)^{1/2},$$

Damit haben wir eine Konvergenzrate der Ordnung $n^{-1/4}$ für den Kolmogorov-Abstand hergeleitet, was nur *suboptimal* ist. Um ein Berry-Esseen-Resultat mittels des Stein-Ansatzes herleiten zu können, muß man zu den speziellen (beschränkten) Testfunktionen der Kolmogorov-Metrik, also zu

$$g(x) = 1_{\{x \leq t\}},\ t \in \mathbb{R},$$

die Lösung der Stein-Gleichung direkt untersuchen. Dabei sind Schranken, die $|g'|_\infty$ verwenden, zu ersetzen, insbesondere muß man auf eine Schranke von $|f''|_\infty$ verzichten. Erfolgreich in diesem Bereich war zuerst Bolthausen in einer Arbeit aus 1984, [3].

Wir wollen nun die Stein-Gleichung

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

ein wenig diskutieren. Es gilt:

Lemma 2 *Es sei $g : \mathbb{R} \to \mathbb{R}$ mit $\mathbb{E}|g(Z)| < \infty$. Dann ist*

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

eine fast sicher differenzierbare Lösung der Stein-Gleichung zu g. Weitere Lösungen sind von der Form $\tilde{f}(x) = f(x) + ce^{x^2/2}$.

Der Beweis erfolgt durch Nachrechnen bzw.:
ist f eine Lösung, so folgt

$$\frac{d}{dx}\left(e^{-x^2/2}f(x)\right) = e^{-x^2/2}(f'(x) - xf(x)) = e^{-x^2/2}(g(x) - \mathbb{E}g(Z)),$$

also ist $(*)$ ein geeigneter Kandidat.
Wir deuten nun sogar an, warum

$$|f|_\infty \le \sqrt{\frac{\pi}{2}}|g - \mathbb{E}g(Z)|_\infty \quad \text{und} \quad |f'|_\infty \le 2|g - \mathbb{E}g(Z)|_\infty$$

gilt. Zunächst gilt

$$\int_{\mathbb{R}} e^{-y^2/2}(g(y) - \mathbb{E}g(Z)) = 0,$$

also

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

Es sei $x > 0$. Dann ist

$$|f(x)| \leq |g - \mathbb{E}g(Z)|_\infty \underbrace{\left(e^{x^2/2}\int_x^\infty e^{-y^2/2}dy\right)}_{(*)}.$$

Man betrachte:

$$\frac{d}{dx}\left(e^{x^2/2}\int_x^\infty e^{-y^2/2}dy\right) = -1 + xe^{x^2/2}\int_x^\infty e^{-y^2/2}dy \leq 0,$$

denn es gilt

$$\int_x^\infty e^{-y^2/2}dy \overset{x>0}{\leq} \int_x^\infty \frac{y}{x}e^{-y^2/2}dy = \frac{1}{x}e^{-x^2/2}$$

(wird Mill's ratio genannt). Also wird $(*)$ maximiert in $x = 0$ auf $[0, \infty)$ und der Wert dort ist $\sqrt{\pi/2}$. Für den Fall $x < 0$ verwende man die erste Darstellung von f.

Kommen wir zu einer Abschätzung der Ableitung. Für $x > 0$ ist

$$\begin{aligned} f'(x) &= xf(x) + g(x) - \mathbb{E}g(Z) \\ &= g(x) - \mathbb{E}g(Z) - xe^{x^2/2}\int_x^\infty e^{-y^2/2}(g(y) - \mathbb{E}g(Z))dy. \end{aligned}$$

Also folgt die Abschätzung

$$|f'(x)| \leq |g - \mathbb{E}g(Z)|_\infty \left(1 + xe^{x^2/2}\int_x^\infty e^{-y^2/2}dy\right) \leq 2|g - \mathbb{E}g(Z)|_\infty,$$

wobei wir $xe^{x^2/2}\int_x^\infty e^{-y^2/2}dy \leq 1$ (Mill's ratio) verwendet haben. Der Fall $x < 0$ läuft analog. Zu guter letzt noch eine Andeutung des Weges, eine gute Abschätzung für f'' zu finden. Es gilt

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

Differenzieren liefert

$$f''(x) - xf'(x) - f(x) = g'(x).$$

Daraus folgt

$$\begin{aligned} f''(x) &= g'(x) + f(x) + xf'(x) \\ &= g'(x) + f(x) + x(g(x) - \mathbb{E}g(Z) + xf(x)) \\ &= g'(x) + x(g(x) - \mathbb{E}g(Z)) + (1 + x^2)f(x), \end{aligned}$$

und nun folgt eine aufwendige Rechnung, siehe etwa [14].

2.2 Poisson-Approximation

Intermezzo: die STEINsche Methode funktioniert nicht nur bei Normal-Approximation. Wir erinnern an den klassischen Grenzwertsatz von POISSON. Es gilt:

$$\underbrace{\binom{n}{k} p_n^k (1 - p_n)^{n-k}}_{\text{Binomial}} \underset{n\to\infty}{\longrightarrow} \underbrace{e^{-\lambda}\frac{\lambda^k}{k!}}_{\text{POISSON}},$$

falls $n\,p_n \to \lambda > 0$.
Die erste Frage betrifft die etwas seltsame Bedingung an die Folge p_n: $n\,p_n \to \lambda$. Gilt diese Bedingung in relevanten Anwendungen?
Eine zweite Frage ist die nach einer Konvergenzrate: kann man für festes n: $\mathrm{TV}(\mathrm{Bin}(n, p), Po(np))$ gut abschätzen?
LE CAM hat 1960 die mittels Kopplung herleitbare Schranke

$$\mathrm{TV}(\mathrm{Bin}(n, p), Po(np)) \leq np^2$$

bewiesen. Wie gut ist diese Rate?
Wir skizzieren kurz die STEINsche Methode für die POISSON-Approximation:
Es gilt für $Z \sim Po(\lambda)$ und jede beschränkte Abbildung $f : N_0 \to \mathbb{R}$

$$\mathbb{E}(Zf(Z)) = \mathbb{E}(\lambda f(Z + 1)).$$

Gilt umgekehrt für alle $f : \mathbb{N}_0 \to \mathbb{R}$ beschränkt

$$\mathbb{E}(Xf(X)) = \mathbb{E}(\lambda f(X+1)),$$

so folgt $X \sim Po(\lambda)$. Wir betrachten in diskreten Modellen den Totalvariations-Abstand. Nun können wir die STEIN-Gleichung aufstellen:

$$\lambda f_{A,\lambda}(j+1) - j f_{A,\lambda}(j) = 1_{\{j\in A\}} - Po(\lambda)\{A\}.$$

$f_{A,\lambda}$ bezeichnet also die Lösung der Diffrenezengleichung zu $1_{\{j\in A\}}$ sowie zu λ. Nun gilt also für jede Zufallsvariable W mit Werten in $\mathbb{N}_0$:

$$\mathbb{E}\left[\lambda f_{A,\lambda}(W+1) - f_{A,\lambda}(W)\right] = P(W \in A) - P_0(\lambda)\{A\}.$$

Die Lösung der STEIN-Gleichung hat die Gestalt

$$f_{A,\lambda}(j+1) = \frac{j!}{\lambda^{j+1}} e^{\lambda}\left[P_0(\lambda)\{A \cap U_j\} - Po(\lambda)\{A\}Po(\lambda)\{U_j\}\right]$$

mit $U_j = \{0, 1, 2, \ldots, j\}$ (einfaches Lösen der Rekursion). LOUIS CHEN hat die Methode von STEIN in seiner Doktorarbeit auf den Fall der POISSON-Approximation ausgeweitet. Man spricht daher im Falle dieser Approximation häufig von den CHEN-STEIN-Methode. Die Lösung $f_{A,\lambda}$ hat man sehr gut im Griff. Es gelten zum Beispiel:

$$||f|| := \sup_{j\geq 0} |f_{A,\lambda}(j)| \leq \min\left(1, \sqrt{\frac{2}{e\lambda}}\right),$$

$$\Delta f := \sup_{j\geq 0} |f_{A,\lambda}(j+1) - f_{A,\lambda}(j)| \leq \frac{1}{\lambda}(1 - e^{\lambda}) \leq \min\left(1, \frac{1}{\lambda}\right).$$

$\frac{1}{\lambda}$ heißt „magic-factor“.

Wir führen hier nun eine beeindruckend kurze Rechnung vor. Es seien $I_1, \ldots, I_n$ unabhängige Zufallsvariablen mit

$$P(I_j = 1) = p_j = 1 - P(I_j = 0).$$

Wir untersuchen die Limesverteilung von

$$W := W_n = \sum_{j=1}^{n} I_j; \quad \lambda := \mathbb{E}(W_n) = \sum_{j=1}^{n} p_j$$

und definieren

$$W_i := \sum_{j=1,\, j\neq i}^{n} I_j.$$

Es gilt

$$\mathbb{E}[\lambda f(W+1) - Wf(W)]$$

mit $f := f_{A,\lambda}$ zu untersuchen. Es gilt

$$\begin{aligned}\mathbb{E}[I_i f(W)] &= \mathbb{E}[I_i f(W_i + 1)] \\ &= \mathbb{E}[I_i]\mathbb{E}[f(W_i + 1)] = p_i \mathbb{E}[f(W_i + 1)].\end{aligned}$$

Also folgt

$$\mathbb{E}[\lambda f(W+1) - Wf(W)] = \sum_{i=1}^{n} p_i \left[\mathbb{E}(f(W+1)) - \mathbb{E}(f(W_i+1))\right].$$

Ist $I_i = 1$ (ein Ereignis, welches mit Wahrscheinlichkeit p_i eintritt), so folgt $W \neq W_i$, sonst gilt Gleichheit. Es folgt somit

$$|P(W \in A) - Po(\lambda)\{A\}| \leq \underbrace{\sup_{j\geq 0} |f(j+1) - f(j)|}_{\leq \frac{1}{\lambda}(1-e^{-\lambda})} \sum_{i=1}^{n} p_i^2.$$

Insbesondere gilt im Fall identische verteilter I_j:

$$\mathrm{TV}(\mathrm{Bin}(n,p), Po(n\,p)) \leq np^2 \min\left(1, \frac{1}{np}\right) = \mathcal{O}(p).$$

Für die Poisson-Approximation ist im abhängigen Fällen sehr viel bewiesen worden (siehe etwa das Buch [1] zur Poisson-Approximation).
Für andere diskrete Verteilungen ist die Steinsche Methode ebenfalls entwi-

ckelt worden. Hierzu noch ein Beispiel aus der elementaren Stochastik-Ausbildung zum Thema Binomial-Verteilung. Man ziehe n Bälle ohne Zurücklegen aus einer Box mit N Bällen, in der R rote seien. Der Vergleich mit einer Hypergeometrischen Verteilung weist die folgende Güte auf:

$$\left|\mathcal{H}(N,R,n)(j) - B\left(n, \frac{R}{N}\right)(j)\right| \leq \frac{n-1}{N-1}$$

via Stein (wobei $\mathcal{H}$ die Hypergeometrische Verteilung und B die Binomialverteilung bezeichnet).
Zu weiteren diskreten Verteilungen siehe etwa [7]. Wir beenden das Intermezzo zu diskreten Verteilungen. Eine Ausführung zu anderen stetigen Dichten erfolgt später.

3 Austauschbare Paare und Beispiele

3.1 Theorie

Bisher haben wir unabhängige Summanden - bei Normalapproximation - betrachtet. Wir betrachten nun eine - von mehreren - allgemeinen Techniken,

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

abzuschätzen.
Dazu sei eine Zufallsgröße W' so gewählt, dass $W' \stackrel{d}{=} W$ (in Verteilung gleich). Weiter gelte

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

fast sicher (auf ein Intermezzo zu bedingten Erwartungen sei hier verzichtet). Weiter nehmen wir $\mathbb{E}W^2 = 1$ an. Es gilt unter diesen Annahmen:

Satz 1 *(Stein [14] und Folgearbeiten)*
Sei $Z \sim N(0,1)$:

$$\mathrm{Wass}(W,Z) \leq \left(\frac{2}{\pi}\mathrm{Var}\left[\mathbb{E}\left(\frac{1}{2\lambda}(W'-W)^2 \middle| W\right)\right]\right)^{1/2} + \frac{1}{3\lambda}\mathbb{E}|W'-W|^3.$$

In der praktischen Anwendung dieses Theorems wird sich zeigen, dass λ klein, zum Beispiel $\sim {}^1\!/\!{}_n$ sein wird. W' wir als eine kleine Störung von W konstruiert, was in der Regel sogar austauschbar gelingt:

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

Der erste Zugang von STEIN verlangte diese Austauschbarkeit, die für uns daher quasi nur aus historischen Gründen die Überschrift des Teilkapitels liefert.

Wir beweisen nun diesen allgemeinen Zugang, also das obige Theorem. Auf Grund der Übereinstimmung der Verteilung von W und W' folgt natürlich

$$\mathbb{E}W = \mathbb{E}W', \quad \mathbb{E}W^2 = \mathbb{E}(W')^2 = 1.$$

Weiter ist

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

und da $\lambda \neq 0$ folgt $\mathbb{E}W = 0$. Es gilt weiter

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

Wieder wählen wir ein $f : \mathbb{R} \to \mathbb{R}$ mit

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

Betrachte

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

Mittels TAYLOR folgt nun

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

mit
$$|R| \leq \frac{1}{6}|W-W'|^3||f''||_\infty \leq \frac{1}{3}|W-W'|^3.$$
Weiter gilt
$$\begin{aligned}
-\lambda\mathbb{E}[Wf(W)] &= \mathbb{E}[(W'-W)f(W)] \\
&= -\mathbb{E}\left[\frac{1}{2}(W'-W)^2 f'(W) + R\right] \\
&= -\mathbb{E}\left[\frac{1}{2}\mathbb{E}\left[(W'-W)^2 \middle| W\right] f'(W)\right] - \mathbb{E}(R).
\end{aligned}$$
Somit folgt zusammenfassend
$$\begin{aligned}
&|\mathbb{E}f'(W) - \mathbb{E}\left[Wf(W)\right]| \\
&\leq \left|\mathbb{E}\left[f'(W)\left(\mathbb{E}\left(\frac{1}{2\lambda}(W-W')^2\middle|W\right)-1\right)\right]\right| + \frac{1}{3\lambda}\mathbb{E}|W-W'|^3 \\
&\leq \sqrt{\frac{2}{\pi}}\mathbb{E}\left|\mathbb{E}\left(\frac{1}{2\lambda}(W-W')^2\middle|W\right)-1\right| + \frac{1}{3\lambda}\mathbb{E}|W-W'|^3 \\
&\leq \left[\frac{2}{\pi}\operatorname{Var}\left(\mathbb{E}\left(\frac{1}{2\lambda}(W'-W)^2\middle|W\right)\right)\right]^{1/2} + \frac{1}{3\lambda}\mathbb{E}|W-W'|^3
\end{aligned}$$
wegen $\mathbb{E}(W-W')^2 = 2\lambda$ und $\mathbb{E}|X| \leq \operatorname{Var}(X)^{1/2}$.
Wir haben den Beweis also schon gebracht. Der Nutzen dieses vermutlich etwas magisch anmutenden Ansatzes wird nun diskutiert.
Wir betrachten - noch einmal - den unabhängigen Fall, um zu sehen, wie man mit obigen Größen umgeht. Die Kernfrage ist natürlich die nach der Konstruktion von W'. Wir betrachten die Schritte:

- $X_1, \ldots, X_n$ seien unabhängig, $\mathbb{E}X_i = 0$, $\operatorname{Var}(X_i) = 1$ und
$$W := W_n := \frac{1}{\sqrt{n}}\sum_{i=1}^n X_i.$$
- $X_1', \ldots, X_n'$ seien unabhängige Kopien von $X_1, \ldots, X_n$.
- Wähle I uniform, zufällig aus $\{1, \ldots, n\}$: $P(I = j) = 1/n$, unabhängig von allen X_i und X_i'.

- Ersetze X_I durch X_I':

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

- Nun gilt (ohne Beweis; hier nur intuitiv)

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

Der Erfolg des Ansatzes kommt jetzt: es gilt

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

Somit ist mit $\lambda = \frac{1}{n}$ unsere zentrale Bedingung erfüllt. Weiter gilt

$$\begin{aligned}
\frac{1}{3\lambda} \mathbb{E}|W' - W|^3 &= \frac{n}{3} \frac{1}{n^{3/2}} \mathbb{E}|X_I' - X_I|^3 \\
&= \frac{1}{3} \frac{1}{n^{3/2}} \sum_{i=1}^{n} \mathbb{E}|X_i' - X_i|^3 \\
&\leq \frac{8}{3} \frac{1}{n^{3/2}} \sum_{i=1}^{n} \mathbb{E}|X_i|^3.
\end{aligned}$$

Im Fall unabhängiger und identisch verteilter X_i ist der letzte Ausdruck

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

Nun bestimmen wir mit $X = (X_1, \dots, X_n)$

$$\mathbb{E}\left[(W' - W)^2 \middle| X\right] = \frac{1}{n} \mathbb{E}\left[(X_I' - X_I)^2 \middle| X\right]$$

$$= \frac{1}{n}\big(\mathbb{E}(X_I^2 \big| X) - 2\mathbb{E}(X_I X_I' | X) + \mathbb{E}((X_I')^2 | X)\big).$$

Da X_I' unabhängig von X nach Konstruktion, folgt

$$\mathbb{E}((X_I')^2 | X) = \frac{1}{n}\sum_{i=1}^n \mathbb{E}((X_i')^2) = \frac{1}{n}\sum_{i=1}^n \mathbb{E}(X_i^2) = 1.$$

Weiter ist

$$\mathbb{E}(X_I^2 | X) = \sum_{i=1}^n \mathbb{E}(1_{\{I=i\}} X_i^2 | X) = \sum_{i=1}^n X_i^2 \mathbb{E}(1_{\{I=i\}} | X) = \frac{1}{n}\sum_{i=1}^n X_i^2,$$

da I und X unabhängig nach Konstruktion. Weiter ist

$$\mathbb{E}(X_I X_I' | X) = \mathbb{E}\big(\mathbb{E}(X_I X_I' | X, I) | X\big) = \mathbb{E}\big(X_I \mathbb{E}(X_I' | X, I) | X\big).$$

Nun ist $\sigma(X_I', I)$ unabhängig von $\sigma(X)$, also ist die innere bedingte Erwartung auf der rechten Seite gleich $\mathbb{E}(X_I' | I)$. Auf Grund der Unabhängigkeit von X_i' und I stimmt dies mit

$$\sum_{i=1}^n 1_{\{I=i\}} \mathbb{E}(X_i^*) = 0$$

überein. Wir erhalten also $\mathbb{E}(X_I X_I' | X) = 0$ und somit

$$\mathbb{E}\left[\frac{1}{2\lambda}(W' - W)^2 \middle| X\right] = \frac{1}{2} + \frac{1}{2n}\sum_{i=1}^n X_i^2. \tag{3}$$

Da W bezüglich $\sigma(X)$ messbar ist, also $\sigma(W) \subset \sigma(X)$, folgt eine fast sichere Übereinstimmung mit $\mathbb{E}\left[\frac{1}{2\lambda}(W' - W)^2 \middle| W\right]$.

Wir müssen also die Varianz der rechten Seite in (3) bestimmen. Es gilt

$$\mathrm{Var}\left(\frac{1}{2} + \frac{1}{2n}\sum_{i=1}^n \mathbb{E}\left[X_i^2 \middle| W\right]\right) = \mathrm{Var}\left(\mathbb{E}\left[\frac{1}{2n}\sum_{i=1}^n X_i^2 \middle| W\right]\right) \le \mathrm{Var}\left(\frac{1}{2n}\sum_{i=1}^n X_i^2\right).$$

Dies folgt mit Hilfe der JENSEN-Ungleichung und $\mathbb{E}\left[\mathbb{E}\left[X|W\right]\right] = \mathbb{E}X$:

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

Nun gilt

$$\mathrm{Var}\left(\frac{1}{2n}\sum_{i=1}^n X_i^2\right) = \frac{1}{4n^2}\sum_{i=1}^n \mathrm{Var}(X_i^2) \leq \frac{1}{4n^2}\sum_{i=1}^n \mathbb{E}X_i^4.$$

Im unabhängig, identisch verteilten Fall gilt ist die rechte Seite gleich $\frac{1}{4n}\mathbb{E}X_1^4$. Wir haben also gezeigt:

$$\mathrm{Wass}(W, Z) \leq \left(\frac{1}{2\pi n^2}\sum_{i=1}^n \mathbb{E}X_i^4\right)^{1/2} + \frac{8}{3n^{3/2}}\sum_{i=1}^n \mathbb{E}|X_i|^3.$$

Im unabhängigen und identisch verteilten Fall folgt die Abschätzung

$$\frac{\mathrm{const}}{\sqrt{n}}\left(\mathbb{E}|X_i|^3 + \mathbb{E}X_i^4\right).$$

Erneut haben wir insbesondere einen zentralen Grenzwertsatz bewiesen sowie eine KOLMOGOROV-Rate der Ordnung $n^{-1/4}$ erreicht.

3.2 Zwei weitere Beispiele

3.2.1 Hoeffdings kombinatorischer zentraler Grenzwertsatz

Es sei π eine zufällige Permutation von $\{1, \ldots, n\}$ (Gleichverteilung). Gegeben sei eine Matrix $(a_{ij})_{i,j=1}^n$ mit gewissen Bedingungen. Wir betrachten die Zufallsgröße

$$W := W_n = \sum_{i=1}^n a_{i\pi(i)}.$$

Unter welchen Bedingungen gilt ein zentraler Grenzwertsatz:

$$\frac{W - \mathbb{E}(W)}{\sqrt{\mathrm{Var}(W)}} \underset{n\to\infty}{\longrightarrow} N(0,1) \qquad ?$$

Beispiele für das obige Modell:

- Abstand zur Identität: wir wählen

$$a_{ij} = |i - j| : \quad W = \sum_{i=1}^{n} |i - \pi(i)|.$$

- Anzahl der Fixpunkte: wir wählen

$$a_{ij} = 1_{\{i=j\}}.$$

- Matchings: das Ménage-Problem:

$$a_{ij} \in \{0, 1\}$$

$$\begin{pmatrix} 1 & 1 & 0 & \dots\dots\dots & & 0 \\ 0 & 1 & 1 & 0 & \dots. & 0 \\ \vdots & & & \ddots & & \vdots \\ 0 & \dots\dots\dots & & 0 & 1 & 1 \end{pmatrix}$$

  $P(W = 0)$: n Mann-Frau-Paare an einem runden Tisch. $P(W = 0)$ ist dann die Wahrscheinlichkeit, dass niemand neben seinem Partner sitzt.

Ohne Rechnung geben wir bekannt, dass wir ohne Beschränkung der Allgemeinheit die folgenden Annahmen treffen können:

$$\sum_{j=1}^{n} a_{ij} = 0 \qquad \sum_{i=1}^{n} a_{ij} = 0$$

und

$$\frac{1}{n-1} \sum_{i,j=1}^{n} a_{ij}^2 = 1.$$

Es folgt unter diesen Annahmen, dass $\mathbb{E}(W) = 0$ und $\mathrm{Var}(W) = 1$.
Wir geben die Konskruktion des austauschbaren Paares bekannt. Es sei

$$(\pi, \pi') : \quad \pi' = \pi \circ (I, J)$$

mit (I,J) die Transposition, gleichmäßig, zufällig aus allen Transpositionen auf $\{1,\ldots,n\}$ ausgewählt. Dann gilt

$$\begin{aligned}\pi'(I) &= \pi(J)\\ \pi'(J) &= \pi(I)\\ \pi'(k) &= \pi(k) \quad \forall k \neq I, J.\end{aligned}$$

Wir wählen nun

$$W' := \sum_{i=1}^{n} a_{i\pi'(i)},$$

womit

$$\begin{aligned}W - W' &= a_{I\pi'(I)} + a_{J\pi'(J)} - a_{I\pi(I)} - a_{J\pi(J)}\\ &= a_{I\pi(J)} + a_{J\pi(I)} - a_{I\pi(I)} - a_{J\pi(J)}.\end{aligned}$$

Damit ist

$$\mathbb{E}\left[W' - W \middle| \pi\right] = \frac{1}{n(n-1)} \sum_{1\leq i\neq j\leq n} \left(a_{i\pi(j)} + a_{j\pi(i)} - a_{i\pi(i)} - a_{j\pi(j)}\right)$$

und

$$\frac{1}{n(n-1)} \sum_{i\neq j} a_{i\pi(i)} = \frac{1}{n}\sum_{i=1}^{n} a_{i\pi(i)} = \frac{W}{n}$$

Die Beobachtung $\sum_{j\neq i} a_{i\pi(j)} = -a_{i\pi(i)}$ liefert

$$\frac{1}{n(n-1)} \sum_{i\neq j} a_{i\pi(j)} = -\frac{1}{n(n-1)} \sum_{i} a_{i\pi(i)} = -\frac{1}{n(n-1)} W,$$

also insgesamt

$$\mathbb{E}\left[W' - W \middle| \pi\right] = \frac{-2(n-2)}{n(n-1)} W := -\lambda W$$

mit $\lambda := \frac{2(n-2)}{n(n-1)}$. Wir geben nun nur noch bekannt, dass in diesem Fall der

Ansatz des austauschbaren Paares liefert:

$$\mathrm{Wass}(W, Z) \leq \frac{\mathrm{const}}{\sqrt{n}}.$$

Bolthausen konnte 1984 in [3] sogar zeigen:

$$\mathrm{Kolm}(W, Z) \leq K \frac{\sum_{i,j} |a_{ij}|^3}{n} = \mathcal{O}\left(\frac{1}{\sqrt{n}}\right).$$

Wir bemerken, dass bei der Wahl von $A = (a_{ij})$ wie oben sowie der Matrix

$$m_{ij} := \begin{cases} 1 & \pi(j) = i \\ 0 & \text{sonst} \end{cases}$$

(Permutationsmatrix) folgt:

$$W = \mathrm{Tr}(AM).$$

Dieses Objekt ist für eine wichtige Klasse anderer Zufallsmatrizen M viel studiert (siehe Kapitel 5).

3.2.2 Curie-Weiss-Modell und ein zentraler Grenzwertsatz

Wir stellen das Curie-Weiss-Modell hier sehr verkürzt vor: Gegeben seien n magnetische Partikel, sogenannte spins mit Werten $+1$ oder -1. Wir bezeichnen die spins mit $\sigma_1, \ldots, \sigma_n$, $\sigma_i \in \{-1, +1\}$. Das Curie-Weiss-Modell ist grob gesprochen so konstruiert, dass es gleiche Spin-Ausrichtung bevorzugt. Wir betrachten die folgende Wahrscheinlichkeitsverteilung:

$$P(\sigma) := \frac{1}{Z_n} \exp\left(\frac{\beta}{n} \sum_{1 \leq i \leq j \leq n} \sigma_i \sigma_j\right),$$

wobei Z_n die Normierung ist, so dass P ein Wahrscheinlichkeitsmaß wird. Z_n ist die sogenannte Zustandssumme. Wir betrachten nun den mittleren spin

$$m(\sigma) := \frac{1}{n} \sum_{i=1}^{n} \sigma_i.$$

Unter dem obigen Maß P kann man herleiten:

$$\begin{aligned} \beta \leq 1: &\quad m(\sigma) \to 0 \quad \text{in Wahrscheinlichkeit} \\ \beta > 1: &\quad m(\sigma) \Rightarrow \frac{1}{2} \left(\delta_{m^*(\beta)} + \delta_{-m^*(\beta)}\right) \end{aligned}$$

mit $m^*(\beta)$, ist die positive Lösungen der Curie-Weiss-Gleichung

$$x = \tanh(\beta x).$$

Weiter ist bekannt:

$$\beta < 1: \quad \sqrt{n} m(\sigma) = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \sigma_i \Rightarrow N\left(0, \frac{1}{1-\beta}\right)$$

und

$$\beta = 1: \quad n^{1/4} m(\sigma) = \frac{1}{n^{3/4}} \sum_{i=1}^{n} \sigma_i \Rightarrow X,$$

wobei X die Dichte (geeignet normiert) $e^{-x^4/12}$ besitzt (siehe etwa [8]).
Nun konstruieren wir ein austauschbare Paar: Zu $\sigma = (\sigma_1, \ldots, \sigma_n)$ konstruiere $\sigma' = (\sigma'_1, \ldots, \sigma'_n)$ wie folgt:

- wähle I aus $\{1, \ldots, n\}$ gleichverteilt, unabhängig von allen vorkommendenn Zufallsgrößen.
- ersetze σ_I durch σ'_I: σ'_I ist realisiert gemäß der bedingten Verteilung von σ_I gegeben $(\sigma_j)_{j \neq I}$, und unabhängig von σ_I.

Wir betrachten

$$s_n^2 := \operatorname{Var}\left(\frac{1}{\sqrt{n}} \sum_i \sigma_i\right).$$

Man kann nachrechnen, dass gilt: $s_n^2 \to \frac{1}{1-\beta}$ für $0 < \beta < 1$. Wir betrachten nun

$$W = W_n = \frac{1}{\sqrt{n} s_n} \sum_{i=1}^{n} \sigma_i.$$

Nun folgt mit $\mathbb{E}\left[\sigma_i' \middle| \sigma\right] = \tanh\left(\frac{\beta}{n} \sum_{j:j\neq i} \sigma_j\right)$ und unter Verwendung der TAYLOR-Entwicklung $\tanh(x) = x - \frac{x^3}{3} + \mathcal{O}(x^5)$:

$$\begin{aligned}
\mathbb{E}\left[W - W' \middle| W\right] &= \frac{1}{\sqrt{n} s_n} \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}\left[\sigma_i - \sigma_i' \middle| \sigma\right] \\
&= \frac{1}{\sqrt{n} s_n} m(\sigma) - \frac{1}{n^{3/2} s_n} \sum_{i=1}^{n} \mathbb{E}\left[\sigma_i' \middle| \sigma\right] \\
&= \lambda W + \overline{R}
\end{aligned}$$

mit

$$\lambda = \frac{1-\beta}{n} + \frac{\beta}{n^2}$$

und

$$\overline{R} = \frac{1}{3} \frac{1}{n^{3/2}} \frac{1}{s_n} \sum_{i=1}^{n} \beta^3 \left(\sum_{j:j\neq i} \frac{\sigma_j}{n} \right)^3.$$

Die Bedingung es austauschbaren Paar-Ansatzes ist also nur bis auf einen Fehlerterm R erfüllt. Daher muss man hier die Methode von STEIN neu aufrollen, konkret eine Vorarbeit [12] von RINOTT und ROTAR aus 1997 verwenden. Tatsächlich erhalten wir eine Konvergenzrate:

$$\operatorname{Kolm}\left(\frac{1}{\sqrt{n} s_n} \sum_i \sigma_i, Z \right) \leq \frac{A}{\sqrt{n}}, \quad 0 < \beta < 1$$

(LÖWE, EICHELSBACHER; 2008)
Bei der kritischen Tempertaur $\beta = -1$ tritt die neue Dichte $\exp\left(-\frac{x^4}{12}\right)$ auf. Nun ist es notwendig, die STEINsche Methode hierfür zu entwickeln. Tatsächlich ist desweiteren ein neuer Ansatz eines austauschbaren Paares zu

entwickeln und man erhält für

$$W = \frac{1}{n^{3/4}} \sum_{i=1}^{n} \sigma_i, \quad W' \text{ analog},$$

$$\mathrm{Wass}(W, \widetilde{Z}) \leq \frac{B}{\sqrt{n}}, \qquad \beta = 1,$$

wobei $\widetilde{Z}$ die Verteilung gemäß dier obigen neuen Dichte habe. Tatsächlich vermuten wir $\mathrm{Kolm}(W, \widetilde{Z}) \leq \frac{\widetilde{B}}{\sqrt{n}}$. Die Untersuchung des Curie-Weiss-Modells eröffnet das neue Arbeitsgebiet, in anderen, komplexeren Modellen der statistischen Mechanik und der Theorie der Spin-Glass-Modelle Konvergenzraten herleiten zu wollen.

4 Andere Dichten, Mean-Field und Zufallsmatrizen

In Kapitel 1 bis 3 haben wir unter den Verteilungen mit absolut stetiger Dichte nur die Standard-Normalverteilung betrachtet und die Methode von Stein vorgestellt. Dies entspricht der historischen Reihenfolge. Lange Zeit wurden primär Normal-Approximation und Poisson-Approximation studiert. Wir deuten in diesem Kapitel an, wie die Methode von Stein bei anderen Dichten funktioniert. Wir haben dabei die folgenden Dichten vor Augen:

- Der Limes des reskalierten mittleren Spins in der kritischen Temperatur $\beta = 1$ des Curie-Weiss Modells:

$$const. \exp\left(-\frac{x^4}{12}\right).$$

- Wigners Halbkreisgesetz:

$$\frac{1}{2\pi}\sqrt{4 - x^2}\, 1_{\{|x| \leq 2\}}(x).$$

- Die Marchenko-Pastur Verteilung (tritt bei Wishart/Laguerre-

Ensemblen auf):

$$\frac{1}{2\pi\lambda x}\sqrt{-x^2+2(1+\lambda)-(1-\lambda)^2}1_{\{[a,b]\}}(x)$$

mit $a := (1-\sqrt{\lambda})^2$ und $b := (1+\sqrt{\lambda})^2$.

Wir betrachten zunächst eine spezielle Klasse von Dichten, worunter das erste Beispiel fallen wird. Eine spezielle Antwort für die Dichten bei Zufalls-Matrix-Ensembles wird sich anschließen.

4.1 reguläre Dichten

Zunächst betrachten wir eine Dichte p auf $[a,b]$, die

- strikt positiv ist;
- p und p' seien regulär;
- es sei $\psi(x) := \frac{p'(x)}{p(x)}$ regulär;

Desweiteren betrachten wir die folgende Funktionenmenge:

$$\mathcal{F} := \left\{ f \text{ regulär} : \int_a^b |f'(x)|p(x)dx < \infty, \int_a^b |f(x)\psi(x)|p(x)dx < \infty \right\}.$$

Dann gilt das folgende Resultat:

Satz 1 *(*Stein, Diaconis, Holmes, Reinert, *[15])*
Z ist verteilt nach p genau dann wenn

$$\mathbb{E}\left[f'(Z)+\psi(Z)f(Z)\right] = f(b-)p(b-) - f(a+)p(a+)$$

für alle $f \in \mathcal{F}$.

Der Beweis ist erneut partielle Integration.

Beispiele 1 (a) $N(0,1)$: hier ist $\psi(x) = -x$. Diese Charakterisierung kennen wir bereits.

(b) $p(x) = \lambda e^{-\lambda x}$, $x \geq 0$ (Exponential-Verteilung). Hier ist $\psi(x) = -\lambda$, $p(\infty) = 0$, $p(0+) = \lambda$. Also folgt

$$\mathbb{E}[f'(Z) - \lambda f(Z)] = -f(0+)\lambda.$$

(c) *const.* $\exp(-x^4/12)$. Hier ist $p(b-) = p(a+) = 0$ und wir erhalten

$$\mathbb{E}\left[f'(Z) - \frac{1}{3}Z^3 f(Z)\right] = 0.$$

Mit Hilfe der jeweiligen (neuen) STEINgleichung könnte man (zum Beispiel) versuchen, den in Kapitel 3 vorgestellten Ansatz austauschbarer Paare zu entwickeln, zu adaptieren. Für die Dichte $\exp(-x^4/12)$ ist dies Bestandteil der Arbeit von EICHELSBACHER und LÖWE.

Wesentlicher Baustein ist, die Lösung und die Ableitungen der Lösung der korrespondierenden STEIN-Gleichung gut abzuschätzen. Wir erinnern an den Fall $N(0,1)$. Hier gilt ist die Gleichung $f_g'(x) - xf_g(x) = g(x) - \mathbb{E}g(Z)$ und es gilt

$$|f_g|_\infty \leq \sqrt{\frac{\pi}{2}}|g - \mathbb{E}g(Z)|_\infty, \quad |f_g'|_\infty \leq 2|g - \mathbb{E}g(Z)|_\infty, \quad |f_g''|_\infty \leq 2|g'|_\infty.$$

Im Fall der Dichte $\exp(-x^4/12)$ ist $f_g'(x) - \frac{x^3}{3}f_g(x) = g(x) - \mathbb{E}g(\tilde{Z})$ die Gleichung und es ist gelungen, $|f_g|_\infty$ und $|f_g'|_\infty$ abzuschätzen. Aber $|f_g''|_\infty$ bereitet uns derzeit noch Probleme. Erste Ansätze liefern

$$|f_g''(t)| \leq \text{const} \cdot |g'(t)| \cdot p(t),$$

wobei $p(t)$ ein Polynom in t ist. Das würde zu neuen Komplikationen führen. Wir geben nun eine Beweisskizze des obigen Satzes. Es gilt

$$\begin{aligned}\mathbb{E}f'(Z) &= \int_a^b f'(z)p(z)dz \\ &= f(b-)p(b-) - f(a+)p(a+) - \int_a^b f(z)p'(z)dz\end{aligned}$$

$$
\begin{aligned}
&= f(b-)p(b-) - f(a+)p(a+) - \int_a^b f(z)\psi(z)p(z)dz \\
&= f(b-)p(b-) - f(a+)p(a+) - \mathbb{E}[f(Z)\psi(Z)].
\end{aligned}
$$

Daraus folgt, dass für ein Z, welches nach p verteilt ist, die behauptete Identität gilt.
Nun gelte die Identität für alle $f \in \mathcal{F}$ und Z sei eine Zufallsvariable. Wir wollen zeigen, dass dann Z nach p verteilt ist. Dazu betrachte

$$
f_g'(z) + \psi(z)f_g(z) = g(z) - \mathbb{E}g(\tilde{Z}),
$$

wobei $\tilde{Z}$ eine p-verteilte Zufallsvariable bezeichne. Für die STEIN-Lösung f_g gilt

$$
f_g(z) = \frac{\int_a^z (g(x) - \mathbb{E}g(Z))p(x)dx}{p(z)}
$$

(vgl. $N(0,1)$-Fall).
Nun muß man zeigen, dass $f_g \in \mathcal{F}$ gilt. Dies zeigen wir hier nicht. Es folgt dann insbesondere für f_g:

$$
\begin{aligned}
0 &= \mathbb{E}\left(f_g'(Z) + \psi(Z)f_g(Z)\right) - \underbrace{f_g(b-)p(b-)}_{=0} + \underbrace{f_g(a+)p(a+)}_{=0} \\
&= \mathbb{E}g(Z) - \mathbb{E}g(\tilde{Z}).
\end{aligned}
$$

Wir wählen nun für g Indikatoren auf BOREL-Mengen. Dann folgt $P(Z \in B) = P(\tilde{Z} \in B)$, womit also Z nach p verteilt ist.

4.2 Wigner-Matizen

Es bezeichne W_n eine komplexe, HERMITEsche Matrix $n \times n$-Matrix mit $W_{lj} = X_{lj} + iY_{lj}$ unabhängig und X_{lj}, Y_{lj} unabhängig. Weiter sei

$$
\begin{aligned}
&\mathbb{E}X_{kj} = \mathbb{E}Y_{kj} = 0 \quad \forall k, j \\
&\mathbb{E}X_{ll}^2 = 1; \quad \mathbb{E}X_{kj}^2 = \mathbb{E}Y_{kj}^2 = \frac{1}{2}, \quad k < j.
\end{aligned}
$$

Man nennt W_n Wigner-Ensemble oder Wigner-Matrix.
Es seien $\lambda_1, \ldots, \lambda_n$ die Eigenwerte von $\frac{1}{\sqrt{n}} W_n$ und

$$F_n(x) := \frac{1}{n} \sum_{j=1}^{n} 1_{\{\lambda_j \leq x\}}(x)$$

die empirische Verteilungsfunktion. Wir interessieren uns für den Kolmogorov-Abstand

$$\sup_{x \in \mathbb{R}} |\mathbb{E} F_n(x) - G(x)|$$

und eine Abschätzung hiervon, wenn G die Verteilungsfunktion des Halbkreis-Gesetzes ist:

$$G(x) = \int_{-\infty}^{x} \frac{1}{2\pi} \sqrt{4 - u^2}\, 1_{\{|u| \leq 2\}}(u)\, du.$$

Das Gausssche Unitäre Ensemble (GUE) ist der Spezialfall, in dem die X_{ij} und Y_{ij} Gauss-verteilt sind. In dieser Situation ist bekannt, dass der obige Kolmogorov-Abstand durch $c_1 n^{-1}$ abgeschätzt werden kann, siehe Arbeiten von Götze und A. Tikomirov, [9], [10]. Im Fall einer beliebigen Verteilung gilt *universell* die Schranke

$$c_2 \max_{i,j} \left(\mathbb{E} |X_{ij}|^4 \right)^{1/2} n^{-1/2}.$$

Man kann versuchen, dies mit Hilfe der Ideen der Steinschen Methode herzuleiten. Wir geben eine Skizze, wie Götze und A. Tikomirov in einer ihrer Arbeiten in etwa vorgeganen sind. Tatsächlich haben die Autoren die Methode von Stein (so wie wir es hier skizzieren) verwendet, um in einem Fall von Wigner-Matrizen mit nicht(!) unabhängigen Einträgen ein Halbkreisgesetz herleiten zu können. Es wird in diesem Fall keine Konvergenzrate hergeleitet. Die oben zitierten Konvergenzraten entstammen Arbeiten, in denen analytische Techniken verwendet wurden (Stichwort: Stieltjes-Transformation anstelle von charakteristischen Funktionen).
Zunächst ist zu bemerken, dass man Vorsicht beim obigen allgemeinen Dichteansatz walten lassen muß, denn nun ist unser p nicht strikt positiv! Die

naive Verwendung des obigen Ansatzes lieferte

$$\psi(x) = \frac{p'(x)}{p(x)} = -\frac{x}{4-x^2}, \tag{$*$}$$

und man sieht unmittelbar, dass es vermultich Probleme bei dem Ansatz

$$f'(x) - \frac{x}{4-x^2} f(x)$$

für die STEIN-Gleichung gibt. Tatsächlich ist eine korrekte STEIN-Gleichung

$$(4-x^2)f'(x) - 3xf(x),$$

die also nicht einfach durch Multiplikation des ersten Ansatzes mit $(4-x^2)$ folgt. Woher kommt die 3? Der Raum der Testfunktionen ist nun

$$\mathcal{F} = \Big\{ f : \mathbb{R} \to \mathbb{R} : \; f \in C^1(\mathbb{R}\backslash\{-2,2\}), \\ \overline{\lim_{|y|\to\infty}} |yf(y)| < \infty, \; \overline{\lim_{y\to\pm 2}} |4-y^2||f'(y)| < c \Big\}.$$

Es sei Z Halbkreis-verteilt. Dann gilt

$$\begin{aligned} \mathbb{E}[(4-Z^2)f'(Z)] &= \int_{-2}^{2} (4-y^2)g(y)f'(y)dy \\ &= \int_{-2}^{2} \frac{1}{2\pi}(4-y^2)^{3/2} f'(y)dy \\ &= \int_{-2}^{2} f(y)[-2yg(y) + (4-y^2)g'(y)]dy \\ &= \int_{-2}^{2} f(y)[-3yg(y)]dy = -3\mathbb{E}(Zf(Z)), \end{aligned}$$

denn $(*)$ liefert $g'(y)(4-y^2) = -yg(y)$.
Es sei bemerkt, dass man das hier skizzierte Vorgehen der Aufstellung einer STEIN-Gleichung auch für den Fall WISHART/LAGUERRE durchführen kann. Wir betrachten nun die obige STEIN-Gleichung und wollen für x die Zufallsmatrix W_n einsetzen. Dazu bemerken wir, dass für ein reell-symmetrisches W

(wir betrachten reell-symmetrische WIGNER-Matrizen alternativ zum obigen HERMITEschen Modell) gilt $W = U^{-1}\Lambda U$ mit U orthogonal und Λ diagonal. Nun setzen wir

$$f(W) := U^{-1}f(\Lambda)U$$

mit

$$f(\Lambda) = \operatorname{diag}(f(\lambda_1), \dots, f(\lambda_n))$$

(und der üblichen Notation, dass diag die Diagonalmatrix mit den angegebenen Diagonaleinträgen bezeichne). Jetzt können wir in der folgenden Art und Weise W_n in die STEIN-Gleichung einsetzen:

$$\frac{1}{n}\mathbb{E}\operatorname{Tr}(4I_n - W_n^2)f'(W_n) - \frac{3}{n}\mathbb{E}\operatorname{Tr} W_n f(W_n).$$

Und nun gilt tatsächlich der folgende Satz, der durch die STEIN-Charakterisierung der Halbkreis-Verteilung motiviert wird:

Satz 2 *Es sei W_n eine Zufallsmatrix für jedes n, so dass für alle $f \in \mathcal{F}$*

$$\frac{1}{n}\mathbb{E}\operatorname{Tr}(4I_n - W_n^2)f'(W_n) - \frac{3}{n}\mathbb{E}\operatorname{Tr} W_n f(W_n) \underset{n\to\infty}{\longrightarrow} 0, \tag{$*$}$$

Dann folgt

$$\sup_{x\in\mathbb{R}} |\mathbb{E}F_n(x) - G(x)| \underset{n\to\infty}{\longrightarrow} 0.$$

Wohl gemerkt, dieser Satz beinhaltet keine Information über die Konvergenzgeschwindigkeit.
Um den obigen Satz zu beweisen, genügt es zu zeigen: wenn in der linken Seite in $(*)$ für f die Lösung f^x der STEINschen Gleichung

$$(4 - y^2)f^x(y)' - 3yf^x(y) = 1_{\{y\le x\}}(y) - G(y) \tag{$**$}$$

eingesetzt wird, folgt $(*) = \mathbb{E}F_n(x) - G(x)$. Dazu verwendet man einen kleinen Trick, der bei der Konstruktion austauschbarer Paare ähnlich vorkam: Es sei

J gleichverteilt aus $\{1, \ldots, n\}$, unabhängig von den X_{ij}, dann ist

$$\begin{aligned} P(\lambda_J \leq x) &= \frac{1}{n} \sum_{j=1}^{n} P(\lambda_j \leq x) \\ &= \frac{1}{n} \sum_{j=1}^{n} \mathbb{E} 1_{\{\lambda_j \leq x\}} = \mathbb{E} F_n(x). \end{aligned}$$

Betrachte in $(**)$ für y nun λ_J :

$$(4 - \lambda_J^2) f^x(\lambda_J)' - 3\lambda_J f^x(\lambda_J) = 1_{\{\lambda_J \leq x\}}(y) - G(x).$$

Nun bilde man den Erwartungswert. Die rechte Seite liefert $\mathbb{E} F_n(x) - G(x)$, und die linke Seite

$$\mathbb{E}[3\lambda_J f^x(\lambda_J)] = \frac{3}{n} \mathbb{E} \left[\sum_{j=1}^{n} \lambda_j f^x(\lambda_j) \right] = \frac{3}{n} \mathbb{E} \left[\operatorname{Tr}(W_n f^x(W_n)) \right]$$

mit

$$\operatorname{Tr}(W_n f^x(W_n)) = \operatorname{Tr}(W_n U^{-1} f^x(\Lambda) U) = \sum_{j=1}^{n} \lambda_j f^x(\lambda_j).$$

Der anderer (zweite) Term wird analog betrachtet.
Was ist der Vorteil des Kriteriums im obigen Satz? Grob gesprochen hat $\operatorname{Tr} W_n f(W_n)$ den Vorteil, linear in den unabhängigen Einträgen von W_n zu sein. Die Entwicklung in X_{ij} und Taylor von $f(W_n)$ in X_{ij} führt zu der Erwartung von Produkten von unabhängigen Zufallsvariablen, die bestimmbar/abschätzbar sind durch erste Momente der X_{ij}. Die ist zwar der Kern, aber an dieser Stelle doch nur eine Andeutung.
Weiter verfolgen Götze und Tikomirov in [10] die Strategie, den obigen Ansatz (siehe Theorem) mit Stieltjes-Funktionen in Verbindung zu bringen. Hierzu wird gezeigt, dass es ausreicht (in einem zu klärenden Sinn), anstelle der obigen Testfunktionen $f \in \mathcal{F}$ die Abbildungen

$$x \mapsto \frac{1}{x - z}, \ z \in \mathbb{C} \backslash \mathbb{R} \quad \text{(wir wählen den Imaginärteil)}$$

zu verwenden. Ist ξ eine Zufallsvariable und F die zugehörige Verteilungsfunktion, so bezeichnet man mit

$$T(z) := \mathbb{E}\frac{1}{\xi - z} = \int_{\mathbb{R}} \frac{1}{x-z} dF(x), \quad z = u + iv, \ v \neq 0$$

die STIELTJES-Transformation von ξ bzw. F. Weiter heißt $f_z(x) := \frac{1}{x-z}$ STILTJES-Funktion und es gilt

$$(f_z(x))'_x = -\frac{1}{(x-z)^2}.$$

Wir berichten kurz über die Verwendung von STIELTJES-Transformierten zur Herleitung des Halbkreis-Gesetzes. Ist ξ Halbkreis-verteilt und sei $S(z) = \mathbb{E}f_z(\xi)$, dann ist $S'(z) = \mathbb{E}\frac{1}{(\xi-z)^2}$ und es gilt

$$(z^2-4)S'(z) - zS(z) - 2 = 0 \quad \text{(nachrechnen)}.$$

Nun gilt für die Transformierte einer empirischen Verteilungsfunktion F_n von Eigenwerten einer Zufallsmatrix W_n

$$\int_{\mathbb{R}} \frac{1}{x-z} dF_n(x) = \frac{1}{n}\sum_{j=1}^{n} \frac{1}{\lambda_j - z} = \frac{1}{n} \operatorname{Tr} R_n(z)$$

mit $R_n(z) := (W_n - zI)^{-1}$, der Resolventenmatrix. Analog

$$\int_{\mathbb{R}} \frac{1}{x-z} d\mathbb{E}F_n(x) = \frac{1}{n}\mathbb{E} \operatorname{Tr} R_n(z).$$

Wenn nun die rechte Seite der letzten Identität gegen $S(z)$ konvergiert, ist das Halbkreisgesetz bewiesen. Man zeigt, dass der Limes der obigen Differentialgleichung genügt. Diese Methode hat PASTUR eingeführt.

Zurück zum STEIN-Ansatz. Wenn wir nun in

$$\frac{1}{n}\mathbb{E}\operatorname{Tr}(4I_n - W_n^2)f'(W_n) - \frac{3}{n}\mathbb{E}\operatorname{Tr} W_n f(W_n)$$

die speziellen Funktionen $f_z(x) = \frac{1}{x-z}$, $f_z(x)'_x = -\frac{1}{(x-z)^2}$ verwenden, moti-

viert die obige Andeutung der Methode der Stieltjes-Transformierten die folgende Aussage:
Man betrachte

$$\frac{1}{n}\mathbb{E}\operatorname{Tr}(4I_n - W_n^2)R_n^2(z) + \frac{3}{n}\mathbb{E}\operatorname{Tr} W_n R_n(z).$$

Wenn man zeigen kann, dass dieser Ausdruck gegen Null konvergiert auf Kompakta von $\mathbb{C}\backslash\mathbb{R}$, und zwar gleichmäßig für $n \to \infty$, dann folgt

$$\sup_{x\in\mathbb{R}} |\mathbb{E}F_n(x) - G(x)| \underset{n\to\infty}{\longrightarrow} 0.$$

Diesen letztgenannten Ansatz verfolgen Götze und Tikomirov erfolgreich in Fällen, in denen die Einträge der Wigner-Matrizen nicht mehr unabhängig (bis auf Symmetrie) gewählt werden. Ein ehrgeiziges Projekt kann es sein, die obigen Techniken für die 10 Ensemble des SFB-TR 12 zu studieren (the ten-fold way).

5 Konzentrationsungleichung via Stein

Wir deuten zunächst an, welcher Typ von Ungleichung in der Wahrscheinlichkeitstheorie Konzentrations-Ungleichung genannt wird. Dazu sei X eine Zufallsvariable und m eine Konstante (etwa der Erwartungswert oder der Median von X unter P). Allgemein ist eine Ungleichung der Bauart

$$P(X - m \geq t) \leq \exp(-f(t))$$

mit $f(0) = 0$ und $f(t) \underset{n\to\infty}{\longrightarrow} \infty$ häufig erreichbar. Man spricht von Konzentration um m.
Wir betrachten den gewohnten Spezialfall unabhängig und identisch verteilter Zufallsgrößen $(X_i)_i$. Es sei $P(X_i = -1) = P(X_i = 1) = 0.5$ (symmetrischer Münzwurf). Dann folgt durch einfache Anwendung der Tscheby-

SCHEV-Ungleichung

$$P\left(\frac{1}{n}\sum_{i=1}^{n} X_i > t\right) \leq \exp\left(-\frac{nt^2}{2}\right).$$

Dies ist einfach und auch ziemlich ungenau, für den Münzwurf. Man sagt, dass $\frac{1}{n}\sum_{i=1}^{n} X_i$ um 0 „konzentriert“ist. Wir halten fest:

$$\begin{aligned} t = n^{-1/2} &: \text{„typische Abweichung“} \\ t << n^{-1/2} &: \ \frac{nt^2}{2} \text{ bei } 0, \text{ keine Info} \\ t >> n^{-1/2} &: \ \frac{nt^2}{n} \text{ groÃ, obere Schranke bei } 0 \end{aligned}$$

Nun stellen wir einen allgemeiner Zugang von S. CHATTERJEE [5] vor, der sich an die Konstruktion austauschbarer Paare (siehe Kapitel 3) anlehnt. Gegeben sei die folgende allgemeine Situation.

$$\begin{aligned} &(X, X') \stackrel{\mathcal{D}}{=} (X', X) \quad \text{ein austauschbares Paar von Zufallsgrößen;} \\ &F(X, X') = -F(X', X) \qquad \text{eine antisymmetrische Funktion } \ F; \\ &\mathbb{E}[F(X, X')|X] = f(X). \end{aligned}$$

Hierbei sei $f(X)$ die zu untersuchende Funktion von X. Weiter sei

$$v(X) := \frac{1}{2}\mathbb{E}\left[|(f(X) - f(X'))F(X, X')|\big|X\right].$$

Dann gilt:

Satz 1 *(CHATTERJEE, 2006) Unter den obigen Annahmen gilt*

(1) $\mathbb{E}[f(X)] = 0$ *und* $\operatorname{Var} f(X) \leq \mathbb{E}[v(X)]$.

(2) Es sei $\mathbb{E}\left[\exp(\theta f(x))|F(X, X')|\right] < \infty$ *für alle* θ *und es sei* $v(X) \leq Bf(X) + C$ *für Konstante* B *und* C*. Dann gilt*

$$P(|f(X)| > t) \leq 2\exp\left(-\frac{t^2}{2Bt + 2C}\right).$$

Wir gehen recht ausführlich durch den Beweis dieses Satzes. Es gilt

$$\mathbb{E}[f(X)] = \mathbb{E}[F(X,X')] = -\mathbb{E}[F(X',X)] = -\mathbb{E}[F(X,X')],$$

also folgt $\mathbb{E}[f(X)] = 0$. Damit gilt

$$\begin{aligned}
\operatorname{Var} f(X) &= \mathbb{E}[f(X)^2] = \mathbb{E}[f(X)F(X,X')] \\
&= \mathbb{E}[f(X')F(X',X)] \qquad \big((X,X') \text{ austauschen}\big) \\
&= -\mathbb{E}[f(X')F(X,X')] = \frac{1}{2}\mathbb{E}[(f(X)-f(X'))F(X,X')] \\
&\le \mathbb{E}[v(X)].
\end{aligned}$$

Nun zum Kern des Satzes, der Konzentrations-Ungleichung. Wir betrachten die Momente-erzeugende Funktion von $f(X)$:

$$m(\theta) := \mathbb{E}[\exp(\theta f(X))],$$

dann folgt

$$m'(\theta) = \mathbb{E}[f(X)\exp(\theta f(X))].$$

In unserer Situation folgt dann

$$\begin{aligned}
m'(\theta) &= \mathbb{E}[F(X,X')\exp(\theta f(X))] \\
&= \frac{1}{2}\mathbb{E}\left[F(X,X')\left(\exp(\theta f(X)) - \exp(\theta f(X'))\right)\right].
\end{aligned}$$

Es gilt die Ungleichung

$$|e^x - e^y| \le \frac{1}{2}|x-y|(e^x + e^y).$$

Daraus folgt

$$\begin{aligned}
|m'(\theta)| &\le \frac{|\theta|}{4}\left|\mathbb{E}\left[F(X,X')(f(X)-f(X'))\left(e^{\theta f(X)} + e^{\theta f(X')}\right)\right]\right| \\
&\le \frac{|\theta|}{2}\mathbb{E}\left[|F(X,X')(f(X)-f(X'))|e^{\theta f(X)}\right] \\
&= |\theta|\mathbb{E}\left[v(X)e^{\theta f(X)}\right] \le B|\theta|m'(\theta) + C|\theta|m(\theta)
\end{aligned}$$

unter Verwendung von $v(X) \leq Bf(X) + C$. Man kann diese gewonnene Ungleichung (klassisch für die Wahrscheinlichkeitstheorie) wie folgt nutzen: Zunächst sei bemerkt, dass m konvex ist mit Minimum in 0 mit $m(0) = 1$. Daher ist

$$m'(\theta)(1 - B\theta) \leq C\theta m(\theta)$$

für $0 < \theta < {}^1\!/_B$. Weiter ist

$$\frac{d}{d\theta} \log m(\theta) = \frac{m'(\theta)}{m(\theta)} \leq \frac{C\theta}{1 - B\theta},$$

und daher

$$\log m(\theta) \leq \int_0^\theta \frac{Ct}{1 - Bt} dt \leq \frac{C\theta^2}{2(1 - B\theta)}.$$

Wir erhalten zusammenfassend

$$\begin{aligned} P(f(X) \geq t) &= P\big(\exp(\theta f(X)) \geq \exp(\theta t)\big) \\ &\leq \exp(-\theta t) m(\theta) \leq \exp\left(-\theta t + \frac{C\theta^2}{2(1 - B\theta)}\right). \end{aligned}$$

Nun optimieren wir in $\theta \in \big(0, \frac{1}{B}\big)$ und erhalten

$$\leq 2 \exp\left(-\frac{t^2}{2Bt + 2C}\right).$$

□

Das wirkt nun noch wie ein recht abstrakter und undurchsichtiger Zugang. Aber wir haben schon mit austauschbaren Paaren gearbeitet. Die beiden folgenden Beispiele zeigen den Wert des Zugangs von Chatterjee.

5.1 Summe unabhängiger Zufallsvariablen

Gegeben seien unabhängige Zufallsgrößen $(Y_i)_i$. Sei

$$X = \sum_i Y_i$$

und $\mu_i := \mathbb{E}Y_i$, $\sigma_i^2 = \mathrm{Var}(Y_i)$. Wir nehmen die folgende Beschränktheitsbedingung an:

$$|Y_i - \mu_i| \leq c_i.$$

Die Konskruktion des austauschbaren Paares geht wie gesehen:

$$X' = \sum_{j \neq I} Y_j + Y_I'.$$

Nun betrachten wir $F(X,Y) := n(x-y)$. F ist antisymmetrisch. Wir rechnen:

$$\begin{aligned}
\mathbb{E}(F(X,X')|Y_1,\ldots,Y_n) &= \frac{1}{n}\sum_{i=1}^n \mathbb{E}(n(Y_i - Y_i')|Y_1,\ldots,Y_n) \\
&= \sum_{i=1}^n (Y_i - \mu_i) = X - \mathbb{E}(X) := f(X).
\end{aligned}$$

Weiter ist

$$\begin{aligned}
v(X) &= \frac{1}{2}\mathbb{E}\left(|(f(X) - f(X'))F(X,X')|\middle|X\right) \\
&= \frac{n}{2}\mathbb{E}\left((X - X')^2\middle|X\right) = \frac{1}{2}\sum_{i=1}^n \mathbb{E}\left((Y_i - Y_i')^2\middle|X\right) \\
&= \frac{1}{2}\sum_{i=1}^n \left(\mathbb{E}((Y_i - \mu_i)^2|X) + \mathbb{E}(Y_i' - \mu_i)^2)\right) \\
&\leq \frac{1}{2}\sum_{i=1}^n (c_i^2 + \sigma_i^2).
\end{aligned}$$

Da $\sigma_i^2 \leq c_i^2$, folgt daraus insgesamt

$$P\left(|X - \mathbb{E}(X)| \geq t\right) = P\left(|f(X)| \geq t\right) \leq 2\exp\left(\frac{-t^2}{\sum_{i=1}^n (c_i^2 + \sigma_i^2)}\right).$$

Dies entspricht der HOEFFDING-Ungleichung.

5.2 Curie-Weiss-Modell

Zu $\sigma = (\sigma_1, \ldots, \sigma_n) \in \{-1, 1\}^n$ war hier

$$P(\sigma) = \frac{1}{Z_\beta} \exp\left(\frac{\beta}{n} \sum_{i<j} \sigma_i \sigma_j\right)$$

und

$$m(\sigma) := \frac{1}{n} \sum_{i=1}^n \sigma_i \quad \in [-1, 1].$$

Nun wählen wir

$$F(\sigma, \sigma') := \sigma_I - \sigma'_I = n(m(\sigma) - m(\sigma')).$$

Dann folgt

$$\mathbb{E}\left[F(\sigma, \sigma')\middle|\sigma\right] = \frac{1}{n} \sum_{i=1}^n \mathbb{E}\left(\sigma_i - \sigma'_i\middle|\sigma\right) = m(\sigma) - \frac{1}{n} \sum_{i=1}^n \mathbb{E}(\sigma_i|\sigma)$$

und es gilt $\mathbb{E}(\sigma_i|\sigma) = \tanh\left(\frac{\beta}{n} \sum_{j,j\neq i} \sigma_j\right)$. Somit ist

$$f(\sigma) = m(\sigma) - \frac{1}{n} \sum_{i=1}^n \tanh(\beta m_i(\sigma)).$$

Weiter gilt

$$\begin{aligned}
|f(\sigma) - f(\sigma')| &\leq |m(\sigma) - m(\sigma')| + \frac{1}{n} \sum_{i=1}^n \big|\tanh(\beta m_i(\sigma')) - \tanh(\beta m_i(\sigma))\big| \\
&\leq \frac{2}{n} + \frac{\beta}{n} \sum_{i=1}^n |m_i(\sigma') - m_i(\sigma)| \\
&\leq \frac{2}{n} + \frac{\beta}{n} \frac{2(n-1)}{n} < \frac{2(1+\beta)}{n},
\end{aligned}$$

da die Funktion $x \mapsto \tanh(x)$ 1-Lipschitz-stetig ist. Somit folgt für v die gewünschte Bedingung:

$$\begin{aligned} v(\sigma) &= \frac{1}{2}\mathbb{E}\left[|(f(\sigma) - f(\sigma'))||F(\sigma,\sigma')|\big|\sigma\right] \\ &\leq \frac{1}{2}\frac{2(1+\beta)}{n}2 = \frac{2(1+\beta)}{n}. \end{aligned}$$

Zusammenfassend erhalten wir

$$P(|f(\sigma)| \geq t) \leq 2\exp\left(-\frac{nt^2}{4(1+\beta)}\right).$$

Da $f(\sigma) = m(\sigma) - \frac{1}{n}\sum_{i=1}^n \tanh(\beta m_i(\sigma))$ nahe bei $m(\sigma) - \tanh(\beta m(\sigma))$ liegt - es gilt

$$|\tanh(\beta m(\sigma)) - \frac{1}{n}\sum_{i=1}^n \tanh(\beta m_i(\sigma))| \leq \frac{n-1}{n^2}\beta \quad -$$

folgt

$$P\left(|m(\sigma) - \tanh(\beta m(\sigma))| \geq \frac{\beta}{n} + t\right) \leq 2\exp\left(-\frac{nt^2}{4(1+\beta)}\right), \quad t \geq 0.$$

Tatsächlich gab es vor der Arbeit von Chatterjee keine Konzentrations-Ungleichung von dieser Güte.

5.3 Zufallsmatrizen

Abschließend kommen wir nochmals auf Zufallsmatrizen zu sprechen. Chatterjee konnte in [4] den folgenden Typ einer Konzentrations-Ungleichung beweisen: $M = M_n$ sei eine $n \times n$ Hermitesche Matrix und

$$F_M(x) := \frac{1}{n}\sum_{i=1}^n 1_{\{\lambda_i \leq x\}}.$$

M und UMU^* für ein unitäres U haben das gleiche Spektrum. Das Haarsches Maß auf der unitären Gruppe der $n \times n$-Matrizen induziert in natürlicher Weise eine Gleichverteilung auf der Menge aller Hermitescher Matrizen mit gleichem Spektrum „wie“ M. Dieses induzierte Maß bezeichnen wir mir ϱ_M.

Satz 2 *(*CHATTERJEE*)*
Es seien M und N HERMITEsch $n \times n$ Matrizen. Es seien $A \sim \varrho_M$ und $B \sim \varrho_N$ unabhängige zufällige HERMITEsche Matrizen und $H := A + B$. Dann gilt

(1)
$$\forall x \in \mathbb{R}: \quad \operatorname{Var}(F_H(x)) \leq \kappa \frac{\log n}{n}.$$

(2)
$$P\left(|F_H(x) - \mathbb{E}(F_H(x))| \geq t\right) \leq 2 \exp\left(-\frac{nt^2}{2\kappa \log n}\right).$$

Die Tatsache, dass $H \mapsto F_H(x)$ „unstetig" ist, ist die Ursache dafür, dass bis dahin entwickelte Konzentrations-Techniken nicht verwendet werden konnten.
Weiter wurden in jüngster Zeit *lineare Funktionen auf klassischen Matrizengruppen* studiert. Wir stellen kurz ein Resultat von E. MECKES [11] vor. Dazu bezeichne O_n die orthogonale Gruppe von orthogonalen $n \times n$ Matrizen. M sei verteilt nach HAAR auf O_n und A sei eine feste $n \times n$-Matrix über $\mathbb{R}$ mit $\operatorname{Tr}(A\,A^t) = n$. Man untersucht die Zufallsgröße

$$W := \operatorname{Tr}(AM)$$

(siehe HOEFFDINGs kombinatorischer Grenzwertsatz: dort war M eine Permutationsmatrix).
D'ARISTOTILE, DIACONIS, NEWMANN konnten

$$\operatorname{Kolm}(W, Z) \underset{n\to\infty}{\longrightarrow} 0$$

mit $Z \sim N(0,1)$ beweisen. MECKES konnte mit Hilfe der Methode von STEIN und der Konstruktion von austauschbaren Paaren zeigen:

$$\operatorname{TV}(W, Z) \leq \frac{2\sqrt{3}}{n-1}, \quad n \geq 2.$$

Hier gibt es für symmetrische Räume etliche spannende offenen Fragen von ähnlichem Typ.

Literatur

[1] Barbour, A. D. ; Holst, L. ; Janson, S. *Poisson approximation.* Oxford Studies in Probability, 2. Oxford Science Publications. The Clarendon Press, Oxford University Press, New York, 1992. x+277 pp. MR MR1163825

[2] Billingsley, P. *Probability and measure.* Third edition. Wiley Series in Probability and Mathematical Statistics. A Wiley-Interscience Publication. John Wiley & Sons, Inc., New York, 1995. xiv+593 pp. MR MR1324786

[3] Bolthausen, E. *An estimate of the remainder in a combinatorial central limit theorem.* Z. Wahrsch. Verw. Gebiete **66** (1984), no. 3, 379–386. MR MR075157

[4] Chatterjee, S. *Concentration of Haar measures, with an application to random matrices.* J. Funct. Anal. **245** (2007), no. 2, 379–389. MR MR2309833

[5] Chatterjee, S. *Stein's method for concentration inequalities.* Probab. Theory Related Fields **138** (2007), no. 1-2, 305–321. MR MR2288072

[6] L. H. Y. Chen and Q.-M. Shao, *Stein's method for normal approximation*, An introduction to Stein's method, Lect. Notes Ser. Inst. Math. Sci. Natl. Univ. Singap., vol. 4, Singapore Univ. Press, Singapore, 2005, pp. 1–59. MR MR2235448

[7] P. Eichelsbacher and G. Reinert, *Stein's method for discrete Gibbs measures*, Ann. Appl. Probab. **18** (2008), no. 4, 1588–1618. MR MR2434182

[8] R. S. Ellis, *Entropy, Large Deviations, and Statistical Mechanics*, Springer-Verlag, New York, 1985.

[9] Götze, F. ; Tikhomirov, A. N. . *The rate of convergence for spectra of GUE and LUE matrix ensembles.* Cent. Eur. J. Math. **3** (2005), no. 4, 666–704 (electronic). MR MR217166

[10] Götze, F. ; Tikhomirov, A. N. *Limit theorems for spectra of random matrices with martingale structure.* Teor. Veroyatn. Primen. **51** (2006), no. 1, 171–192; translation in Theory Probab. Appl. 51 (2007), no. 1, 42–64 MR MR2324173

[11] Meckes, E. *Linear functions on the classical matrix groups.* Trans. Amer. Math. Soc. **360** (2008), no. 10, 5355–5366. MR MR241507

[12] Y. Rinott and V. Rotar, *On coupling constructions and rates in the CLT for dependent summands with applications to the antivoter model and weighted U-statistics*, Ann. Appl. Probab. **7** (1997), no. 4, 1080–1105. MR MR1484798 (99g:60050)

[13] Q.-M. Shao and Z.-G. Su, *The Berry-Esseen bound for character ratios*, Proc. Amer. Math. Soc. **134** (2006), no. 7, 2153–2159 (electronic). MR MR2215787 (2008j:60064)

[14] C. Stein, *Approximate computation of expectations*, Institute of Mathematical Statistics Lecture Notes—Monograph Series, 7, Institute of Mathematical Statistics, Hayward, CA, 1986. MR MR882007 (88j:60055)

[15] C. Stein, P. Diaconis, S. Holmes, and G. Reinert, *Use of exchangeable pairs in the analysis of simulations*, Stein's method: expository lectures and applications, IMS Lecture Notes Monogr. Ser., vol. 46, Inst. Math. Statist., Beachwood, OH, 2004, pp. 1–26. MR MR2118600 (2005j:65005)