

90 JAHRE LINDEBERG-METHODE

PETER EICHELSBACHER¹ AND MATTHIAS LÖWE²

Zusammenfassung: Eines der wohl bekanntesten Resultate der Wahrscheinlichkeitstheorie ist der *zentrale Grenzwertsatz*. Es geht um die Beobachtung, dass in vielen verschiedenen Modellen der Stochastik im Limes die Verteilung von Zufallsgrößen gegen die universelle Gauß-Verteilung mit Dichte $\frac{1}{\sqrt{2\pi}} \exp(-x^2/2)$ konvergiert. LINDEBERG publizierte in den Jahren 1920-1922 in drei Arbeiten eine Beweismethode, die im Vergleich zu anderen Beweismethoden überraschend elementar ist. Sie kann vor allem auf tiefere analytische Werkzeuge sowie auf knifflige kombinatorische Überlegungen verzichten. In jüngster Zeit erlebt LINDEBERGs Ansatz eine beeindruckende Renaissance. Erweiterungen seiner Methode haben zu teilweise bahnbrechenden neuen Resultaten unter anderem in der Theorie der Zufallsmatrizen geführt. Wir stellen die LINDEBERG Methode vor und geben einen Einblick in die jüngste Entwicklung.

1. EINLEITUNG

Die Bezeichnung *zentraler Grenzwertsatz* wird relativ einheitlich seit der Arbeit [25] von GEORG PÓLYA aus dem Jahre 1920 mit dem Titel *Über den zentralen Grenzwertsatz der Wahrscheinlichkeitsrechnung und das Momentenproblem* immer dann verwendet, wenn als Limesverteilung in einem stochastischen Modell die Gaußsche Dichte auftritt. PÓLYA meinte mit seiner Wortwahl durchaus die *zentrale* Bedeutung dieser Gruppe von Grenzwertsätzen. Es gibt in der Wahrscheinlichkeitstheorie auch eine andere Begründung für die Wahl des Wortes *zentral*: die Gaußsche Limesverteilung tritt etwa bei Partialsummen von unabhängigen Zufallsvariablen immer dann auf, wenn nicht Ausreißer das Verhalten der Partialsummen bestimmen, sondern das Verhalten im Zentrum der Verteilung der Partialsumme dominiert. In der moderneren Wahrscheinlichkeitstheorie spricht man daher ganz allgemein von einem zentralen Grenzwertsatz,

¹Ruhr-Universität Bochum, Fakultät für Mathematik, NA 3/67, D-44780 Bochum, Germany, peter.eichelsbacher@ruhr-uni-bochum.de

²Westfälische Wilhelms-Universität Münster, Fachbereich Mathematik, Einsteinstraße 62, D-48149 Münster, Germany, maloeve@math.uni-muenster.de

wenn das Verhalten des Zentrums der Verteilung in einem stochastischen Modell, und nicht das Randverhalten (tail-Verhalten) beschrieben wird. Wir stellen in Kapitel 2 zunächst die LINDEBERG Methode für Partialsummen unabhängiger Zufallsvariablen vor. Die Methode ist sehr elementar. Im Wesentlichen verwendet man eine Taylor-Entwicklung. Es überrascht ein wenig, dass dieser Beweis in nur sehr wenigen Lehrbüchern aufgenommen wurde, unter ihnen das Buch von BILLINGSLEY [4] und das von BREIMAN [6]. Wir werben in diesem Artikel für eine Auseinandersetzung mit der Beweis-Methode von LINDEBERG nicht nur, weil die Methode elegant ist, sondern weil sich Modifikationen der Methode ebenfalls recht elementar darstellen lassen, diese aber wiederum sehr viel mehr Informationen enthalten, als der ursprüngliche Weg von LINDEBERG es vermuten lässt. In Kapitel 3 geben wir eine kurze historische Einordnung der Arbeiten von LINDEBERG im Kontext des zentralen Grenzwertsatzes. In Kapitel 4 zeigen wir exemplarisch für *zufällige Partialsummen* sowie für *Martingale*, wie die Methode in den Jahrzehnten nach LINDEBERGs Grundlegung verwendet und ausgebaut wurde. Im Anschluss wird in Kapitel 5 eine jüngere Arbeit von CHATTERJEE sowie in Kapitel 6 eine Arbeit von TAO und VU vorgestellt, die zu bedeutenden Resultaten in der jungen Erfolgsgeschichte der Theorie der *Zufallsmatrizen* führte. Beide Arbeiten basieren auf LINDEBERGs Methode und beschreiben eindrucksvoll die Renaissance seiner Idee.

2. DIE METHODE IM EINFACHSTEN FALL

2.1 Ein verblüffend einfacher Beweis

Im Folgenden wollen wir den sehr eleganten Beweisweg von LINDEBERG in der denkbar einfachsten Situation der Wahrscheinlichkeitstheorie vorstellen: Gegeben sind reellwertige Zufallsvariablen $X_1, X_2, \dots$, die unabhängig seien. Wir nehmen ohne Einschränkung an, dass die Erwartungswerte $\mathbb{E}(X_i) = \int X_i dP = 0$ sind (P bezeichnet das Wahrscheinlichkeitsmaß des Wahrscheinlichkeitsraumes, auf dem die Zufallsvariablen definiert sind). Weiter bezeichnen wir mit $\sigma_i^2 := \mathbb{V}(X_i) = \mathbb{E}((X_i - \mathbb{E}(X_i))^2)$ die Varianz der Zufallsvariable X_i . Dann ist auf Grund der angenommenen Unabhängigkeit der Zufallsvariablen $s_n^2 := \sum_{i=1}^n \sigma_i^2$ die Varianz der Summe $\sum_{i=1}^n X_i$.

Die Frage nach der Gültigkeit eines zentralen Grenzwertsatzes in dieser Situation ist die Frage nach den Bedingungen an die Zufallsvariablen X_i , so dass für jedes $x \in \mathbb{R}$ und für

$$W_n := \frac{1}{s_n} (X_1 + X_2 + \dots + X_n)$$

gilt:

$$(2.1) \quad \lim_{n \rightarrow \infty} P(W_n \leq x) = \Phi(x)$$

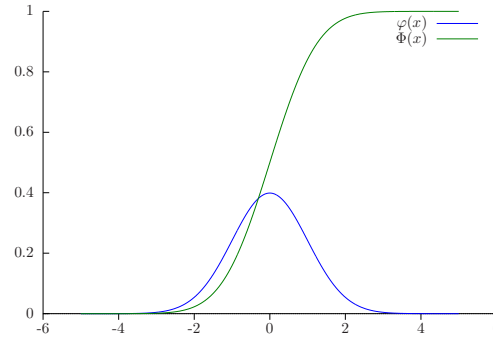


ABBILDUNG 1. Dichte φ und Verteilungsfunktion Φ der Standardnormalverteilung.

mit $\Phi(x) = \int_{-\infty}^x \varphi_{0,1}(t) dt$ und $\varphi_{0,1}(t) = \frac{1}{\sqrt{2\pi}} \exp(-t^2/2)$. Wir bemerken, dass $\mathbb{E}(W_n) = 0$ (da $\mathbb{E}(X_i) = 0$ für jedes i) und $\mathbb{V}(W_n) = \frac{1}{s_n^2} \mathbb{V}(\sum_{i=1}^n X_i) = 1$ gilt, also die Zufallsvariable W_n für jedes $n \geq 1$ in den ersten beiden Momenten $\mathbb{E}(W_n)$ und $\mathbb{E}(W_n^2)$ mit den Momenten der Gauß-Verteilung übereinstimmt. Wir kommen auf dieses „Momente-Matching“, später zurück.

Um den Beweis nach LINDEBERG transparent zu halten, betrachten wir zunächst den Fall, in dem die Zufallsvariablen X_i die gleiche Verteilung besitzen ($P(X_i \leq t) = P(X_1 \leq t)$ für jedes i) und die Varianz $\mathbb{V}(X_1) = \mathbb{E}((X_1 - \mathbb{E}(X_1))^2) = 1$ ist. Es seien $Z_1, Z_2, \dots$ unabhängige, identisch verteilte Zufallsvariablen, die $N(0, 1)$ -verteilt sind. Dies bedeutet $P(Z_i \leq x) = \Phi(x)$ für alle $x \in \mathbb{R}$. Dann gilt, dass

$$Z = \frac{1}{\sqrt{n}}(Z_1 + Z_2 + \dots + Z_n)$$

auch $N(0, 1)$ verteilt ist. Dies ist eine bekannte Tatsache, die wir später noch genauer kommentieren werden. Es wird zentral sein, dass diese Beobachtung eine charakterisierende Eigenschaft der Normalverteilung ist.

Die Idee (eine der Ideen) von LINDEBERG ist, in dem obigen W_n sukzessive (Teleskop-Summe) die Summanden X_i durch die normalverteilten Z_i zu ersetzen und somit W_n durch Z zu approximieren. Dieses Austauschverfahren (im Englischen *replacement trick* oder auch *swapping trick* genannt) führt schnell zum Beweis des zentralen Grenzwertsatzes. Dieser Ansatz führt darüberhinaus zu der Möglichkeit, den zentralen Grenzwertsatz für nicht notwendig identisch verteilte Zufallsvariablen zu beweisen, allerdings unter einer zusätzlichen Bedingung, die seit den Arbeiten von LINDEBERG nach ihm benannt ist.

Eine kleine technische Vorbereitung ist die folgende: wir schreiben $P(W_n \leq x) = \mathbb{E}(1_{W_n \leq x})$ mit der Notation der Indikatorfunktion $1_{W_n \leq x}$, die den Wert 1 auf dem Ereignis $\{W_n \leq x\}$ annimmt, und sonst den Wert 0. Wir wollen zeigen, dass $\lim_{n \rightarrow \infty} \mathbb{E}(1_{W_n \leq x}) = \mathbb{E}(1_{Z \leq x})$ gilt. Dazu approximieren wir die Indikatorfunktion $1_{(-\infty, x]}$ durch eine hinreichend glatte Funktion f . Für

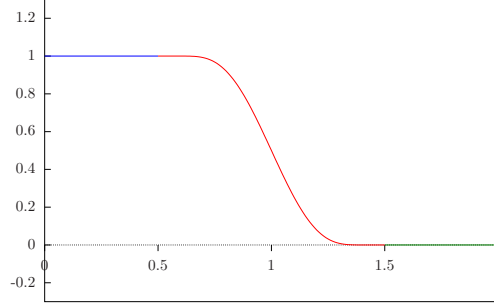


ABBILDUNG 2. gute Approximierende einer Indikatorfunktion

unsere Ziele genügt es ein f zu wählen, welches dreimal differenzierbar ist und dessen Ableitungen stetig und beschränkt sind (diese *Testfunktionen* bilden eine Konvergenz-determinierende Klasse für die Konvergenz in (2.1), siehe auch Abbildung 2.1).

Zu zeigen ist für jede solche Funktion f

$$\lim_{n \rightarrow \infty} \mathbb{E}(f(W_n)) = \mathbb{E}(f(Z)).$$

Mittels TAYLOR-Entwicklung existiert nun ein η und ein $g(h)$ mit

$$(2.2) \quad g(h) \leq \eta \min(h^2, |h|^3)$$

mit

$$(2.3) \quad |f(x + h_1) - f(x + h_2) - f'(x)(h_1 - h_2) - \frac{1}{2}f''(x)(h_1^2 - h_2^2)| \leq g(h_1) + g(h_2).$$

Hierbei ist η eine Konstante, die durch die Werte $M_i := \sup_{x \in \mathbb{R}} |f^{(i)}(x)| < \infty$ für $i = 0, 1, 2, 3$ bestimmt ist. Alternativ ist

$$(2.4) \quad g(h) \leq \eta |h|^3$$

mit $\eta = M_3/6$ eine simple Abschätzung. Nun definieren wir $T_k := X_1 + \dots + X_{k-1} + Z_{k+1} + \dots + Z_n$. Also ist $\frac{1}{\sqrt{n}}(T_n + X_n) = W_n$ und $\frac{1}{\sqrt{n}}(T_1 + Z_1) = Z$ und mittels *Teleskop-Summen* Bildung folgt

$$\begin{aligned} \mathbb{E}(f(W_n) - f(Z)) &= \mathbb{E}\left[f\left(\frac{1}{\sqrt{n}}(T_n + X_n)\right) - f\left(\frac{1}{\sqrt{n}}(T_1 + Z_1)\right)\right] \\ &= \sum_{k=1}^n \mathbb{E}\left[f\left(\frac{1}{\sqrt{n}}(T_k + X_k)\right) - f\left(\frac{1}{\sqrt{n}}(T_k + Z_k)\right)\right]. \end{aligned}$$

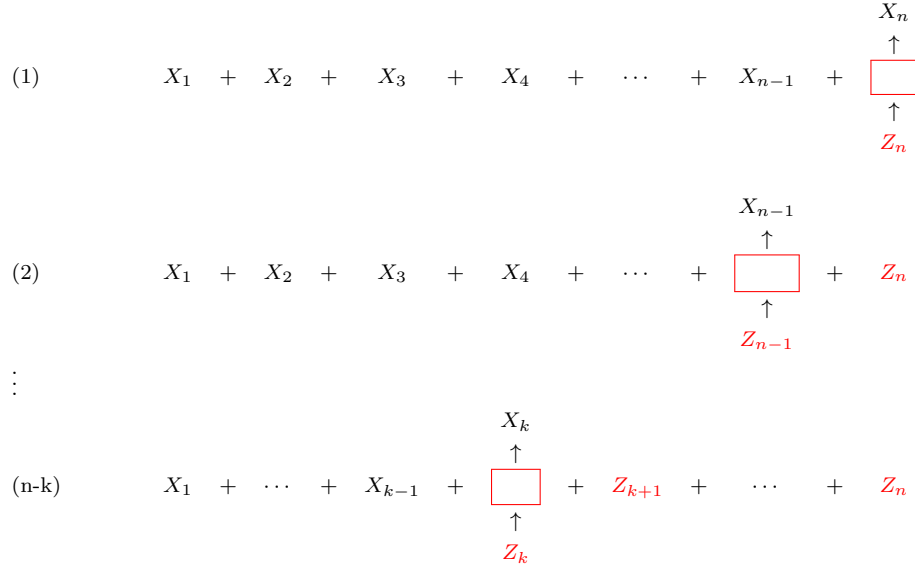


ABBILDUNG 3. Schema der LINDEBERG-Methode

Es folgt mit $\mathbb{E}(X_k) = \mathbb{E}(Z_k) = 0$ und $\mathbb{E}(X_k^2) = \mathbb{E}(Z_k^2) (= 1)$ sowie der stochastischen Unabhängigkeit von T_k zu X_k und Z_k (für jedes k)

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

Die Unabhängigkeit impliziert etwa $\mathbb{E}(f'(T_k)(X_k - Z_k)) = \mathbb{E}(f'(T_k)) \mathbb{E}(X_k - Z_k)$.

Die vorgestellte Abschätzung basiert alleine auf der Annahme an die X_k , in den ersten beiden Momenten $\mathbb{E}(X_k)$ und $\mathbb{E}(X_k^2)$ mit den Momenten einer $N(0, 1)$ -verteilten Zufallsvariable übereinzustimmen. Es folgt mit der obigen TAYLOR-Vorbereitung (2.3) und der Annahme, dass alle X_i bzw. Z_i identisch verteilt sind:

$$(2.5) \quad |\mathbb{E}(f(W_n) - f(Z))| \leq n \mathbb{E}g\left(\frac{X_1}{\sqrt{n}}\right) + n \mathbb{E}g\left(\frac{Z_1}{\sqrt{n}}\right).$$

Nun wenden wir die Abschätzung (2.2) für g an und erhalten durch Zerlegung des Integrals

$$n \mathbb{E}g\left(\frac{X_1}{\sqrt{n}}\right) \leq n \left[\eta \int_{|X_1| \leq \varepsilon \sqrt{n}} \left| \frac{X_1}{\sqrt{n}} \right|^3 dP + \eta \int_{|X_1| > \varepsilon \sqrt{n}} \left| \frac{X_1}{\sqrt{n}} \right|^2 dP \right].$$

Der erste Summand kann durch $n \eta \varepsilon \int_{|X_1| \leq \varepsilon \sqrt{n}} \left| \frac{X_1}{\sqrt{n}} \right|^2 dP \leq \eta \varepsilon$ abgeschätzt werden, da $\int X_1^2 dP = 1$ nach Voraussetzung. Weiter steigt die Folge der Ereignisse $\{|X_1| > \varepsilon \sqrt{n}\}$ für wachsende n gegen die leere Menge ab, so dass der zweite Summand gegen Null für $n \rightarrow \infty$ konvergiert. Da die gleiche Überlegung für den Term $n \mathbb{E}g\left(\frac{Z_1}{\sqrt{n}}\right)$ verwendet werden kann, folgt bereits der Beweis des zentralen Grenzwertsatzes für Partialsummen unabhängig, identisch

verteilter Zufallsvariablen mit existierendem Erwartungswert und endlicher Varianz. Wir schauen auf den Beweis zurück:

2.2 Das Potential der LINDBERG-Methode

1. Eine Inspektion des obigen Beweises zeigt schnell, dass die Annahme *identisch* verteilter Zufallsvariablen nicht notwendig ist. Entlang der obigen Argumente entdeckt man die zweite fundamentale Beobachtung von LINDBERG, die sogenannte LINDBERG-Bedingung. Sind die Zufallsvariablen *nicht identisch verteilt*, so ergibt sich unmittelbar anstelle von (2.5):

$$(2.6) \quad |\mathbb{E}(f(W_n) - f(Z))| \leq \sum_{i=1}^n \left(\mathbb{E}g\left(\frac{X_i}{s_n}\right) + \mathbb{E}g\left(\frac{Z_i}{s_n}\right) \right).$$

Hierbei sind nun die Z_i unabhängige, normalverteilte Zufallsvariablen mit Erwartungswert 0 und Varianz $\sigma_i^2 = \mathbb{V}(X_i)$ und s_n^2 ist erneut die Varianz der Summe $\sum_i X_i$ bzw. $\sum_i Z_i$ (Unabhängigkeit). Die gleiche Zerlegung des Integrals führt zu der Abschätzung

$$(2.7) \quad \eta \sum_{i=1}^n \int_{|X_i| \leq \varepsilon s_n} \left| \frac{X_i}{s_n} \right|^3 dP + \eta \sum_{i=1}^n \int_{|X_i| > \varepsilon s_n} \left| \frac{X_i}{s_n} \right|^2 dP.$$

Der erste Summand kann nun analog durch $\eta \varepsilon$ abgeschätzt werden. Für den zweiten Summanden erhält man

$$(2.8) \quad \eta \frac{1}{s_n^2} \sum_{i=1}^n \int_{|X_i| > \varepsilon s_n} X_i^2 dP.$$

Man muss nun fordern, dass dieser Ausdruck gegen Null konvergiert für $n \rightarrow \infty$, und genau dies ist die LINDBERG-Bedingung. Wenn man dann noch zeigt, dass die Gauß-Variablen Z_i diese Bedingung erfüllen, was wir hier nicht ausführen, so haben wir bewiesen:

Theorem 2.1 (Zentraler Grenzwertsatz von LINDBERG, 1922). *Gegeben seien unabhängige, reellwertige Zufallsvariablen $X_1, X_2, \dots$ mit $\mathbb{E}(X_i) = 0$ und $\sigma_i^2 := \mathbb{V}(X_i) > 0$ für jedes i . Es sei $s_n^2 := \sum_{i=1}^n \sigma_i^2$. Gilt für jedes $\varepsilon > 0$*

$$(2.9) \quad \lim_{n \rightarrow \infty} \frac{1}{s_n^2} \sum_{i=1}^n \int_{|X_i| > \varepsilon s_n} X_i^2 dP = 0,$$

dann folgt die Aussage (2.1).

WILLIAM FELLER hat in [12] bewiesen, dass eine Art Umkehrung des Satzes von LINDBERG gilt. Setzt man voraus, dass in der Situation des obigen Satzes (gegeben seien unabhängige, reellwertige Zufallsvariablen $X_1, X_2, \dots$ mit $\mathbb{E}(X_i) = 0$ und $\sigma_i^2 := \mathbb{V}(X_i) > 0$ für jedes i und es sei $s_n^2 := \sum_{i=1}^n \sigma_i^2$) die Aussage (2.1) gilt und darüber hinaus $\sigma_n/s_n \rightarrow 0$ und $s_n \rightarrow \infty$ für $n \rightarrow \infty$, so folgt die Gültigkeit der LINDBERG-Bedingung (2.9). Wir gehen auf dieses wichtige Resultat (der Satz von LINDBERG und FELLER) nicht weiter ein.

2. Eine weitere Inspektion des obigen Beweises zeigt, dass die LINDEBERG-Methode eine Information über die *Konvergenzgeschwindigkeit* im zentralen Grenzwertsatz enthält. Dies betrachten wir nur im Sonderfall identisch verteilter Zufallsgrößen X_i . Tatsächlich hat LINDEBERG in seiner ersten Arbeit im Jahre 1920 eine stärkere Bedingung an die Zufallsvariablen X_i gestellt. Er forderte die Endlichkeit des dritten absoluten Momentes $\mathbb{E}|X_i|^3$ für jedes X_i . Erst in den Arbeiten aus dem Jahre 1922 hat er diese Bedingung zur Bedingung (2.9) abgeschwächt. Schauen wir auf den Spezialfall identisch verteilter Zufallsvariablen und nehmen wir $\mathbb{E}|X_1|^3 < \infty$ an, so kann jeder Summand in (2.5) mittels (2.4) wie folgt abgeschätzt werden:

$$\mathbb{E}g(X_1/\sqrt{n}) \leq \eta \mathbb{E}|X_1|^3 \frac{1}{\sqrt{n}}.$$

Wir erhalten also im identisch verteilten Fall für unsere Testfunktionen f mit $\|f'''\| \leq \eta$:

$$|\mathbb{E}(f(W_n) - f(Z))| = \eta \mathcal{O}\left(\frac{1}{\sqrt{n}} \mathbb{E}|X_1|^3\right).$$

Dies ist eine schwache Version einer *Konvergenzrate* im Sinne der Theorie von BERRY und ESSÉEN. *Schwach* bedeutet, dass dieses Resultat für die beschriebene Testfunktionen-Klasse nicht impliziert, dass $\sup_{x \in \mathbb{R}} |P(W_n \leq x) - \Phi(x)|$ (der sogenannte KOLMOGOROV-Abstand) ebenfalls in der Größenordnung $n^{-1/2}$ liegt. Es ist aber bekannt, dass die optimale Konvergenzrate für den KOLMOGOROV-Abstand in unserer Situation $\mathcal{O}(n^{-1/2})$ ist. Dies nennt man die BERRY-ESSÉEN-Rate. Wir kommen darauf in Kapitel 4 zurück.

Man kann darüberhinaus leicht erkennen, dass bei Berücksichtigung höherer Ordnungsterme der TAYLOR-Entwicklung von f und der Forderung der Endlichkeit und Übereinstimmung höherer Momente der X_i mit denen der Z_i die Konvergenzrate verbessert werden kann. Dieser Aspekt wird in unserem Kapitel zur Theorie der Zufallsmatrizen eine wichtige Rolle spielen.

3. Schließlich mag man sich auch fragen, wieso im Zentralen Grenzwertsatz die Normalverteilung als Limes auftaucht und keine andere Grenzverteilung. Wesentlich ist dabei offensichtlich die Eigenschaft, dass die Summe unabhängiger, normal-verteilter Zufallsvariablen wieder normal-verteilt ist. Diese Eigenschaft ist in der Literatur als *unbegrenzte Teilbarkeit* der Normalverteilung bekannt. Es gibt allerdings eine ganze Klasse von Verteilungen, die diese Eigenschaft mit der Normalverteilung teilen. Die Summe von n unabhängigen $N(0, 1)$ verteilten Zufallsvariablen ist $N(0, n)$ verteilt. Was ist das ausgezeichnete der Normalverteilung? Wichtig ist, dass mit den obigen Z_i die Zufallsvariable $\frac{1}{\sqrt{n}} \sum_{i=1}^n Z_i$ wieder $N(0, 1)$ -verteilt ist, d.h. auf der Verteilungsebene und der gewählten Skala $1/\sqrt{n}$ ist die $N(0, 1)$ -Verteilung ein Fixpunkt der Abbildung

$$(2.10) \quad g(X_1, \dots, X_n) = \frac{1}{\sqrt{n}} \sum_{i=1}^n X_i.$$

Ihre herausragende Rolle resultiert in gewisser Hinsicht daraus, dass sie der *einzig*e derartige Fixpunkt ist. Die Inspektion des LINDBERG-Beweises zeigt aber gleichzeitig, dass der teleskopische Austauschprozess auch für andere *Vergleichssummen* $\frac{1}{s_n} \sum_{i=1}^n Z_i$ verwendet werden kann, man wird nur eine andere Limesverteilung beobachten (was im Jargon der Wahrscheinlichkeitstheorie ein *nichtzentraler* Grenzwertsatz genannt wird). Dieser Aspekt wird uns in Kapitel 4 und 5 begegnen.

3. HISTORISCHER ABRISS

Zum Thema *Der Zentrale Grenzwertsatz* gibt es umfassende Werke und Übersichten zur Geschichte, etwa das Buch von FISCHER [13] und die Referenzen dort. Zu Beginn steht ABRAHAM DE MOIVRE und PIERRE SIMON DE LAPLACE und Vorarbeiten von den BERNOULLIS. DE MOIVRE bewies 1733, dass im Falle einer binomialverteilten Zufallsvariable (die Anzahl der Erfolge bei einem n -malig unabhängig wiederholten Münzwurf mit Erfolgswahrscheinlichkeit p) diese standardisiert gegen die Normalverteilung konvergiert. Tatsächlich zeigt er nur den symmetrischen Fall $p = 1/2$. LAPLACE zeigt etwa 1810 diese Aussage für jedes $p \in (0, 1)$. In einer einführenden Stochastik-Vorlesung ist es heute noch üblich, zunächst im Detail diese spezielle Situation $P(X_i = 1) = p = 1 - P(X_i = 0)$ zu betrachten. Dies liegt daran, dass man hier lokal das Histogramm im Zentrum der Binomialverteilung durch die Gaußsche Glockenkurve approximieren kann und somit einen anschaulichen Beweis eines zentralen Grenzwertsatzes voran stellt. LAPLACE deutet an, dass er einen allgemeinen Grenzwertsatz beweisen kann (jenseits des Münzwurf-Modells), aber sein Beweis ist unvollständig. In der Folge hat PAFNUTII TSCHEBYCHEV mit Hilfe der Momente-Methode einen Beweis gegeben, der ebenfalls nicht vollständig ist. Der erste bekannte rigorose Beweis eines zentralen Grenzwertsatzes, der *universell* für die Partialsumme W_n (also unabhängig von der konkreten Wahl der Verteilung der X_i) gilt, stammt von ALEKSANDR LYAPUNOV aus dem Jahre 1901. LYAPUNOV stellte an die Momente $\mathbb{E}(X_i^{2+\delta})$ die folgende heute nach ihm benannte Bedingung: sind $n(X_i)_i$ unabhängige Zufallsvariablen mit $\mathbb{E}(X_i) = 0$, $0 < \sigma_i^2 = \mathbb{V}(X_i) < \infty$ für alle $i \in \mathbb{N}$ und ist $s_n^2 = \sum_{i=1}^n \sigma_i^2$, so genügt diese Folge der LYAPUNOV-Bedingung genau dann, wenn ein $\delta > 0$ existiert mit

$$\lim_{n \rightarrow \infty} \frac{1}{s_n^{2+\delta}} \sum_{i=1}^n \mathbb{E} [|X_i|^{2+\delta}] = 0.$$

Aus der LYAPUNOV-Bedingung folgt die LINDBERG-Bedingung und somit die Gültigkeit des zentralen Grenzwertsatzes. LYAPUNOV führte in seinen Arbeiten die Verwendung der *charakteristischen Funktion* als Beweis-Technik in die Wahrscheinlichkeitstheorie ein. Seine Arbeit war lange Zeit ausserhalb Russlands nicht bekannt. Methodisch kann man in der Historie nun zu LINDBERG übergehen. Sein Beweis war, wie wir gesehen haben, wunderbar elementar. JARL WALDEMAR LINDBERG (1876–1932) war ein finnischer Mathematiker, Sohn eines Dozenten am Polytechnischen Institut Helsinki. Er wurde in Helsinki geboren und starb dort auch. Er



ABBILDUNG 4. Der finnische Mathematiker J.W. LINDEBERG

studierte in Paris und an der Universität seiner Heimatstadt. Er beschäftigte sich mit partiellen Differentialgleichungen und promovierte 1902. Bei der Verfassung seiner ersten Arbeit [21] zum zentralen Grenzwertsatz 1920 kannte er die Resultate von LYAPUNOV nicht, wohl aber schwächere Resultate von VON MISES. In 1922 schrieb LINDEBERG die Arbeiten [23, 22], in denen seine Methode und die nach ihm benannte Bedingung vollständig entwickelt wurde. Der schwedische Mathematiker HARALD CRAMÉR lernte LINDEBERG 1922 kennen und berichtete über diese Begegnung (siehe [9], Seite 514): *When he was reproached for not being sufficiently active in his scientific work, he said well, I am a farmer. And if somebody happened to say that his farm was not properly cultivated, his answer was of course my real job is to be a professor. I was very fond of him and saw him often during the following years.* Paul Lévy schrieb 1925 sein berühmtes Buch *Calcul des probabilités*, in dem er eine gewisse Form des LINDEBERG Beweises reproduzierte. Er nutzte aber entscheidend charakteristische Funktionen, wie LYAPUNOV, und konnte so später einen zentralen Grenzwertsatz für Martingale (1934) beweisen. Dies gilt als Ursache dafür, dass die LINDEBERG-Methode in den Folgejahren kaum in Lehrbüchern auftauchte. FELLER, der die erwähnte Rückrichtung bewiesen hat, formulierte die Vermutung, dass durch LÉVYS Beweis die LINDEBERG-Methode durch die Verwendung der Fourier-Theorie ersetzt werde. Eindrücklich ist die Würdigung der LINDEBERG-Methode durch LE CAM in [20, Kapitel 3] sowie der Kommentar dazu von POLLARD am Ende des Artikels [20]. POLLARD schliesst seinen Kommentar mit den Worten *Lindeberg's argument still has something to offer.* Über 25 Jahre später kann man über eine Renaissance der LINDEBERG-Methode berichten, und genau dieses Ziel verfolgen wir in den nun folgenden Kapiteln.

4. DIE LINDEBERG METHODE IN ANDEREN MODELLEN

In diesem Kapitel beschreiben wir punktuell Weiterentwicklungen der LINDEBERG-Methode. Wir nehmen die Situation aus Kapitel 2 erneut in den Blick und betrachten Partialsummen von unabhängigen und nicht identisch verteilten Zufallsvariablen, wobei jetzt die Anzahl der

Eine neue Herleitung des Exponentialgesetzes in der Wahrscheinlichkeitsrechnung.

Von

J. W. Lindeberg in Helsingfors (Finnland).

1. In einer Arbeit „Über das Exponentialgesetz in der Wahrscheinlichkeitsrechnung“¹⁾ habe ich einige Sätze bewiesen, die sich auf die Frage beziehen, unter welchen Bedingungen die Summe einer großen Anzahl von unabhängigen Wahrscheinlichkeitsgrößen dem Gaußschen Gesetze folgt. Zur Zeit der Redaktion dieser Arbeit hielt ich einen von Herrn v. Mises angegebenen Satz²⁾ für das schärfste bisher in dieser Frage gewonnene Resultat. Nunmehr finde ich, daß schon Liapounoff³⁾ allgemeine Resultate dargelegt hat, die nicht nur über diejenigen des Herrn v. Mises hinausgehen, sondern aus denen auch die meisten der von mir in der oben genannten Arbeit bewiesenen Tatsachen abgeleitet werden können.

Das Studium der Arbeiten von Liapounoff hat mich veranlaßt, die von mir angewandte Methode aufs neue zu prüfen. Hierbei ist mir der Umstand, daß meine Entwicklungen nur an endliche Reihen von Wahrscheinlichkeitsgrößen knüpfen, immer deutlicher als formale Überlegenheit gegenüber der früheren Darstellungsweise hervorgetreten. Man bemerke in dieser Hinsicht, daß der Satz II der nachfolgenden Darstellung, der mir für die mathematische Statistik unbedingt notwendig scheint, aus dem allgemeinen Satze, in welchem Liapounoff seine Resultate zusammenfaßt⁴⁾, nicht gefolgert werden kann, obgleich die Hilfsmittel Liapounoffs

¹⁾ Annales Academiæ Scientiarum Fennicæ 16 (1920), S. 1–23.

²⁾ Fundamentalsätze der Wahrscheinlichkeitsrechnung, Mathematische Zeitschrift 4 (1919), (S. 1–97), S. 78.

³⁾ Sur une proposition de la théorie des probabilités, Bulletin de l'Académie impériale des sciences de St. Pétersbourg 13 (1900), S. 359–386. — Nouvelle forme du théorème sur la limite de probabilité, Mémoires de l'Académie impériale des sciences de St. Pétersbourg 13 (1902), S. 1–24.

⁴⁾ Seite 3 der zweiten der oben zitierten Arbeiten.

sehr wohl zur Ableitung desselben geeignet sind. Weiter habe ich gefunden, daß meine Methode noch erheblich vereinfacht werden kann und daß eine kleine Abänderung derselben zu einer nicht unwesentlichen sachlichen Erweiterung aller mir jetzt bekannten früheren Resultate führt.

Im folgenden wird eine Darstellung meiner Methode und der daraus hervorgehenden Resultate gegeben. Hierbei wird zunächst versucht, einen möglichst einfachen Beweis des oben erwähnten Satzes II zu geben, denn dies scheint mir die wichtigste Aufgabe der Theorie zu sein. Sodann wird der Beweisgang so abgeändert, daß ein möglichst umfassendes Resultat erreicht wird.

Hinsichtlich des allgemeinen Charakters der Wahrscheinlichkeitsgrößen, die wir in Betracht ziehen, machen wir von Anfang an keine andere Voraussetzung, als daß sie Verteilungsfunktionen⁵⁾ besitzen. Demnach wird im folgenden von Integralen immer im Sinne von Stieltjes die Rede sein.

Betreffe der im folgenden benutzten Bezeichnungsweise ist zu bemerken, daß wir die obere Grenze eines Integrals nicht hinschreiben, falls dieselbe $+\infty$ ist; desgleichen wird die untere Grenze $-\infty$ weggelassen. Ferner werden wir uns durchgehend der abkürzenden Bezeichnung

$$\varphi(x, \sigma) = \frac{e^{-\frac{x^2}{2\sigma^2}}}{\sigma\sqrt{2\pi}}$$

bedienen, wo σ eine positive Zahl bedeutet.

2. Es seien $U_1(x), U_2(x), \dots, U_n(x)$ die Verteilungsfunktionen von n voneinander unabhängigen Wahrscheinlichkeitsgrößen $u_1, u_2, \dots, u_n$. Wir machen zunächst die folgenden Voraussetzungen.

Die Mittelwerte der Größen u_μ sind sämtlich Null, d. h. es ist für $\mu = 1, 2, \dots, n$

$$\int x dU_\mu(x) = 0.$$

Wenn die Streuungen der Größen u_μ mit σ_μ bezeichnet werden, d. h. wenn

$$\int x^2 dU_\mu(x) = \sigma_\mu^2$$

gesetzt wird, so ist

$$\sum_{\mu=1}^n \sigma_\mu^2 = 1.$$

Es sei $U(x)$ die Verteilungsfunktion der Summe $u_1 + u_2 + \dots + u_n = u$, also die durch die Gleichung

$$(1) \quad U(x) = \int \dots \int U_n(x - t_1 - t_2 - \dots - t_{n-1}) dU_{n-1}(t_{n-1}) \dots dU_1(t_1)$$

⁵⁾ Wegen des Begriffes der Verteilungsfunktion verweise ich auf die schon zitierte Arbeit des Herrn v. Mises und die Fortsetzung derselben in Bd. 5.

ABBILDUNG 5. Die ersten beiden Seiten der Arbeit [22] aus 1922 von LINDBERG

Summanden zufällig ist. Interessant wird sein, dass dabei nicht nur die Gauß-Verteilung als Limesverteilung auftritt. Der Frage nach der Konvergenzrate mittels der LINDBERG-Methode gehen wir im Anschluss nach. Wir stellen einen Zugang von BOLTHAUSEN vor, bei dem BERRY-ESSÉEN-Raten für Martingal-Summen hergeleitet werden.

4.1 Zufällige Partialsummen und nicht-zentrale Grenzwertsätze

Gegeben seien unabhängige Zufallsvariablen $(X_i)_i$ mit $\mathbb{E}(X_i) = 0$ und $\sigma_i^2 = \mathbb{V}(X_i) < \infty$. Weiter sei N eine Zufallsvariable mit Werten in $\mathbb{N} = \{1, 2, \dots\}$ mit $\mathbb{V}(N) < \infty$, und sie sei unabhängig von den $(X_i)_i$ gewählt. Wir betrachten

$$W_N = \frac{1}{\sqrt{\mathbb{E}(s_N^2)}} (X_1 + \dots + X_N)$$

mit $s_N^2 := \sum_{i=1}^N \sigma_i^2$. Diese zufälligen Summen sind ein viel studiertes Objekt in der Wahrscheinlichkeitstheorie. Sie treten in Modellen der mathematischen Biologie (Theorie der Verzweigungsprozesse) sowie in der Ökonomie (Risiko-Theorie) auf. Sei etwa N_n die Anzahl der männlichen

Nachkommen in der n -ten Nachkommengeneration, und hat der j -te dieser Nahkommen X_j^{n+1} Söhne, so ist

$$N_{n+1} = \sum_{j=1}^{N_n} X_j^{n+1}.$$

Man nennt $(N_n)_n$ BIENAYMÉ-GALTON-WATSON Prozess. Zu weiteren Modellen zufälliger Summen siehe etwa [16] oder [19]. Unter welchen Bedingungen an die Momente der X_i und N kann Konvergenz in Verteilung bewiesen werden und wird ein zentraler Grenzwertsatz gelten? Um die Argumente etwas zu vereinfachen, betrachten wir zunächst den Spezialfall gleicher Varianzen $\sigma_i^2 = 1$ für alle i . Dann ist $\mathbb{E}(s_N^2) = \sum_{n \geq 1} P(N = n) \sum_{i=1}^n \sigma_i^2 = \mathbb{E}(N)$. Wir untersuchen dann die Limesverteilung von $W_N := \frac{1}{\sqrt{\mathbb{E}(N)}}(X_1 + \cdots + X_N)$. Sind die $(Z_i)_i$ erneut unabhängige $N(0, 1)$ -verteilte Zufallsvariablen, unabhängig von N , so untersuchen wir $\mathbb{E}(f(W_N) - f(Z^N))$ mittels der Teleskop-Summen Zerlegung von LINDEBERG. Hierbei ist $Z^N := \frac{1}{\sqrt{\mathbb{E}(N)}}(Z_1 + \cdots + Z_N)$. Nun kennen wir aber die Verteilung von Z^N nicht. Für welche N ist Z^N selbst (oder im Limes) normalverteilt? Wir kümmern uns um diese Frage zunächst nicht, sondern wenden mechanisch die LINDEBERG-Methode an. Wir erhalten für jede Testfunktion f , gewählt wie in Kapitel 2,

$$\mathbb{E}(f(W_N) - f(Z^N)) = \sum_{n \geq 1} P(N = n) \mathbb{E}(f(W_n) - f(Z^n)).$$

Dies folgt via bedingter Wahrscheinlichkeiten bzw. bedingter Erwartungswerte. Wir bedingen auf das Ereignis $\{N = n\}$ für jedes $n \geq 1$. Für festes $n \geq 1$ gilt mit $\alpha := \mathbb{E}(N)$ analog zu (2.6)

$$(4.1) \quad |\mathbb{E}(f(W_n) - f(Z^n))| \leq \sum_{i=1}^n \left(\mathbb{E}g\left(\frac{X_i}{\sqrt{\alpha}}\right) + \mathbb{E}g\left(\frac{Z_i}{\sqrt{\alpha}}\right) \right).$$

Die erste Summe kann nun durch (2.7) abgeschätzt werden, wobei s_n durch α zu ersetzen ist. Analog wird die zweite Summe durch (2.7) abgeschätzt, wobei zusätzlich die X_i durch die Z_i ersetzt werden müssen. Es folgt via Dreiecksungleichung $|\mathbb{E}(f(W_N) - f(Z^N))| \leq U_1 + U_2$ mit

$$(4.2) \quad U_1 \leq 2 \sum_{n \geq 1} P(N = n) \eta \varepsilon \frac{1}{\mathbb{E}(N)} \sum_{i=1}^n \sigma_i^2 = 2\eta \varepsilon,$$

wobei wir verwendet haben, dass wir nur den einfachen Fall $\sigma_i^2 = 1$ betrachten und $\mathbb{E}(N) = \sum_{n \geq 1} n P(N = n)$ nach Definition des Erwartungswertes. Weiter ist

$$U_2 \leq \frac{\eta}{\mathbb{E}(N)} \sum_{n \geq 1} P(N = n) \sum_{i=1}^n \left(\int_{|X_i| > \varepsilon \sqrt{\alpha}} X_i^2 dP + \int_{|Z_i| > \varepsilon \sqrt{\alpha}} Z_i^2 dP \right).$$

Benötigte man für die Abschätzung von U_1 nur die Existenz des Erwartungswertes von N , so wird U_2 ohne eine weitere Voraussetzung an N nicht klein. Dies kann man schon im Spezialfall von unabhängig und identisch verteilten $(X_i)_i$ sehen: jetzt ist die endliche (innere) Summe gleich $n(\int_{|X_1| > \varepsilon \sqrt{\alpha}} X_1^2 dP + \int_{|Z_1| > \varepsilon \sqrt{\alpha}} Z_1^2 dP)$, und somit folgt $U_2 \leq \eta(\int_{|X_1| > \varepsilon \sqrt{\alpha}} X_1^2 dP + \int_{|Z_1| > \varepsilon \sqrt{\alpha}} Z_1^2 dP)$. Nur für $\alpha \rightarrow \infty$ werden die beiden Integrale beliebig klein. Also muss $\alpha = \mathbb{E}(N)$ wachsen. Dies sollte nicht verwundern, denn wir untersuchen eine zufällige Summe von N Summanden:

Die erwartete Anzahl der Summanden sollte bei unserer Grenzwertanalyse wachsen. Im nicht-identisch verteilten Fall folgt nun durch Umsummation

$$\frac{1}{\mathbb{E}(N)} \sum_{n \geq 1} P(N = n) \sum_{i=1}^n \int_{|X_i| > \varepsilon \sqrt{\alpha}} X_i^2 dP = \frac{1}{\mathbb{E}(N)} \sum_{i=1}^{\infty} P(N \geq i) \int_{|X_i| > \varepsilon \sqrt{\alpha}} X_i^2 dP$$

und analog für den Ausdruck mit Z_i anstelle von X_i . Die Bedingung ist nun, dass die Reihe auf der rechten Seite klein wird mit $\mathbb{E}(N) \rightarrow \infty$. Die Verteilung von N (und die der X_i sowie der Z_i für jedes i) muss diese *neue* LINDBERG-Bedingung erfüllen. Diskutieren wir noch den Ausdruck U_1 im Fall nicht-identisch verteilter X_i . Wenn wir zusätzlich fordern, dass $\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \sigma_i^2 = \sigma^2 > 0$, so entnehmen wir im Fall unterschiedlicher Varianzen σ_i^2 der Abschätzung (4.2): $U_1 \leq 2\eta \varepsilon M$ für ein $M > 0$. Zu bemerken ist, dass im Fall nicht identisch verteilter X_i und Z_i mit Varianz $\sigma_i^2 \neq 1$ beide Zufallssummen W_N und Z^N eigentlich mit $\frac{1}{\sqrt{\mathbb{E}(s_N^2)}}$ und nicht mit $\frac{1}{\sqrt{\mathbb{E}(N)}}$ skaliert werden. Wir bleiben aber bei der letzt genannten Skalierung und müssen dann in Kauf nehmen, dass die Limesverteilung von den σ_i^2 abhängt.

Zusammenfassend erhalten wir unter den Bedingungen, die wir im Laufe der Rechnungen entdeckt haben, dass $|\mathbb{E}(W_N) - \mathbb{E}(Z^N)|$ klein wird. Zur Erinnerung: ist N deterministisch, also $P(N = n) = 1$, so ist Z^N normalverteilt. Für eine beliebige Zufallsvariable N mit Werten in $\mathbb{N}$ haben wir diese Information nicht. Andere Limesverteilungen können eventuell auftreten. Aber entlang des Teleskop-Ansatzes von LINDBERG haben wir zunächst den folgenden Satz bewiesen, den wir in dieser Form nicht in der Literatur gefunden haben.

Theorem 4.1. *Gegeben seien unabhängige, reellwertige Zufallsvariablen $X_1, X_2, \dots$ mit $\mathbb{E}(X_i) = 0$ und $\sigma_i^2 := \mathbb{V}(X_i) > 0$ für jedes i . Angenommen es gilt $\sigma^2 := \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \sigma_i^2 > 0$. Weiter sei N eine Zufallsvariable mit Werten in $\mathbb{N}$, unabhängig von den $(X_i)_i$ mit $\mathbb{E}(N) \rightarrow \infty$. Es gelte für jedes $\varepsilon > 0$*

$$(4.3) \quad \sum_{i=1}^{\infty} \frac{P(N \geq i)}{\mathbb{E}(N)} \int_{|X_i| > \varepsilon \sqrt{\mathbb{E}(N)}} X_i^2 dP \rightarrow 0$$

und die gleiche Lindeberg-Bedingung, wenn die $(X_i)_i$ durch die $(Z_i)_i$ ersetzt werden (unabhängige, $N(0, \sigma_i^2)$ -normalverteilte Zufallsvariablen), so folgt

$$|\mathbb{E}(f(W_N) - f(Z^N))| \rightarrow 0.$$

Der Fall einer geometrisch verteilten Zufallsvariable N wurde jüngst in [28] betrachtet. N heißt geometrisch verteilt, wenn $P(N = n) = (1 - p)^{n-1} p$ mit $0 < p < 1$ und $n \in \mathbb{N}$ gilt. Dies ist im Modell des unabhängigen Münzwurfes die Wahrscheinlichkeit, dass zum Zeitpunkt n zum ersten Mal Erfolg eingetreten ist. Es gilt $\mathbb{E}(N) = \frac{1}{p}$ und $\mathbb{V}(N) = \frac{1-p}{p^2}$. Wir betrachten also den Fall $p \rightarrow 0$, um einen Grenzwert für die Verteilung von W_N finden zu können. Da

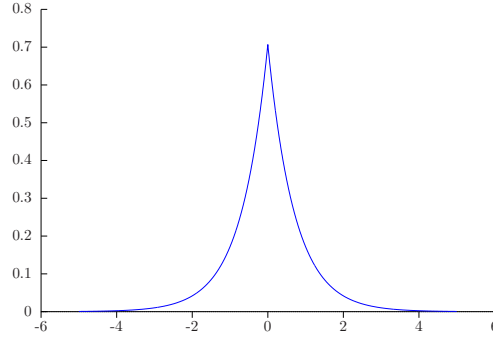


ABBILDUNG 6. Dichte einer LAPLACE-Verteilung

$P(N \geq i) = (1 - p)^{i-1}$, lautet die Lindeberg-Bedingung (4.3) hier (vergleiche [28, (2.1)]):

$$\lim_{p \rightarrow 0} \sum_{i=1}^{\infty} (1 - p)^{i-1} p \int_{|X_i| > \varepsilon p^{-1/2}} X_i^2 dP = 0.$$

In [28] wurde weiter gezeigt, dass (4.3) für die normalverteilten Z_i gilt, wenn man zusätzlich fordert, dass $\lim_{n \rightarrow \infty} n^{-\gamma} \sigma_n^2 = 0$ für ein $0 < \gamma < 1$ gilt. Wenn wir den Fall identisch verteilter Zufallsgrößen $(X_i)_i$ mit $\sigma_i^2 = 1$ betrachten, so folgt ohne weitere Bedingungen $\lim_{p \rightarrow 0} |\mathbb{E}(f(W_N) - f(Z^N))| = 0$. In [28, Proposition 2.4] wurde gezeigt, dass die charakteristische Funktion $\mathbb{E}(\exp(it Z^N))$ im Limes für $p \rightarrow 0$ gegen $\frac{1}{1+t^2/2}$ konvergiert. Die charakteristische Funktion einer Verteilung legt diese eindeutig fest. Die Limesverteilung mit der angegebenen charakteristischen Funktion hat die Dichte

$$f(x) = \frac{1}{\sqrt{2}} \exp(-\sqrt{2}|x|).$$

Dies ist eine spezielle *symmetrische* LAPLACE-Verteilung.

Die mögliche Klasse der Limesverteilungen ist im übrigen mittels der charakteristischen Funktion von Z^N wie folgt beschrieben. Sind die unabhängigen Z_i $N(0, \sigma_i^2)$ -verteilt, so ist $\mathbb{E}(\exp(it Z^N)) = \exp(-\frac{t^2}{2\mathbb{E}(N)} s_n^2)$, und somit gilt

$$\mathbb{E}(\exp(it Z^N)) = \sum_{n \geq 1} P(N = n) \exp\left(-\frac{n}{\mathbb{E}(N)} \frac{s_n^2}{n} \frac{t^2}{2}\right).$$

Mit $s_n^2/n \rightarrow \sigma^2$ folgt (nicht unmittelbar), dass im geometrisch verteilten Fall die rechte Seite für $p \rightarrow 0$ punktweise gegen $\frac{1}{1+\frac{\sigma^2 t^2}{2}}$ konvergiert. Die zugehörige Dichte ist dann $f(x) = \frac{1}{\sqrt{2}\sigma} \exp(-\frac{\sqrt{2}}{\sigma}|x|)$ (LAPLACE-Dichte).

Die LINDEBERG-Methode wurde also erfolgreich für einen nicht-zentralen Grenzwertsatz vorgestellt. Bleibt eine Andeutung, warum bei einer geometrisch verteilten Anzahl von Summanden kein zentraler Grenzwertsatz gilt. Die Varianz der Anzahl der Summanden beträgt $\frac{1-p}{p^2}$ und

wächst somit für $p \rightarrow 0$ schneller als der Erwartungswert. Die Streuung der Anzahl der Summanden ist somit zu groß, um ein zentrales Verhalten im Sinne eines zentralen Grenzwertsatzes zu ermöglichen. Abschliessend sei ergänzt, dass wir im Falle von zufälligen Partialsummen W_N ebenfalls Konvergenzraten aus dem LINDEBERG-Beweis ableiten können. Für die Situation unabhängiger und identisch verteilter $(X_i)_i$ fordern wir $\gamma := \mathbb{E}|X_1|^3 < \infty$ und betrachten erneut nur den Fall $\sigma_i^2 = 1$. Bei der Abschätzung von $\sum_{i=1}^n \mathbb{E}g\left(\frac{X_i}{\sqrt{\alpha}}\right)$ in (4.1) verwenden wir nun (2.4). Wir erhalten

$$\sum_{n \geq 1} P(N = n) \eta \sum_{i=1}^n \frac{\mathbb{E}|X_i|^3}{\mathbb{E}(N)^{3/2}} = \eta \gamma \mathbb{E}(N)^{-1/2},$$

und daher folgt die Konvergenzrate

$$|\mathbb{E}(f(W_N) - f(Z^N))| = \mathcal{O}\left(\frac{\gamma}{\mathbb{E}(N)^{1/2}}\right).$$

Im Fall der geometrischen Summen ist die Konvergenzrate folglich $p^{1/2}$.

4.2 Zentraler Grenzwertsatz für Martingale via LINDEBERG

In der Wahrscheinlichkeitstheorie ist der Begriff des *Martingals* zu einem ganz Wesentlichen geworden. Ursprünglich diente der Begriff zunächst nur der Formalisierung der Idee eines fairen Spiels. Viele Folgen von Zufallsvariablen sind ein Martingal, etwa die eindimensionale symmetrische Irrfahrt auf $\mathbb{Z}$, Partialsummen unabhängiger Zufallsvariablen, so wie sie in Kapitel 2 betrachtet wurden, Produkte von unabhängigen Zufallsvariablen (wenn die Faktoren Erwartungswert 1 haben), die Belegungszahlen des Urnenmodells von PÓLYA, Größen von Populationen in diversen Verzweigungsmodellen oder auch Handelsstrategien in Finanzmärkten. Zentral für die Definition eines Martingals ist der Begriff des *bedingten Erwartungswertes*. Für eine Zufallsvariable X und eine diskrete Zufallsvariable Y (nimmt abzählbar viele Werte an) ist $\mathbb{E}(X|Y = y)$ ganz simpel in Analogie zur bedingten Wahrscheinlichkeit durch

$$\mathbb{E}(X|Y = y) = \frac{1}{P(Y = y)} \int_{Y=y} X dP$$

erklärt. In nicht-diskreten Situation kann aber $P(Y = y) = 0$ für alle Werte y gelten. Allgemeiner sei $(\Omega, \mathcal{A}, P)$ ein Wahrscheinlichkeitsraum und $\mathcal{F}$ ein Unter- σ -Algebra von $\mathcal{A}$. Dann bezeichnet $\mathbb{E}(X|\mathcal{F})$ zu einem integrierbaren X diejenige Zufallsvariable, die $\mathcal{F}$ -messbar ist und für die $\int_A \mathbb{E}(X|\mathcal{F}) dP = \int_A X dP$ für alle $A \in \mathcal{F}$ gilt. Diese Zufallsvariable existiert und ist P -fast sicher eindeutig bestimmt. Weiter sei zu einer Folge $(X_j)_j$ von Zufallsvariablen $\mathcal{F}_j = \sigma(X_1, \dots, X_j)$ die von den ersten j erzeugte σ -Algebra (man sammelt alle Informationen bis zum Zeitpunkt j einschließlich). Dann ist $(X_j)_j$ ein *Martingal*, wenn jedes X_j integrierbar ist und $\mathbb{E}(X_{j+1}|\mathcal{F}_j) = X_j$ P -fast-sicher gilt. Es folgt dann unmittelbar, dass auch $\mathbb{E}(X_{j+k}|\mathcal{F}_j) = X_j$ für jedes $k \in \mathbb{N}$ gilt und dass $\mathbb{E}(X_1) = \mathbb{E}(X_2) = \dots$. Die Definition von Martingalen kann äquivalent durch *Martingaldifferenzen*

$$\Delta_j = X_j - X_{j-1}, \quad \Delta_1 = X_1$$

erklärt werden, denn die Bedingung $\mathbb{E}(X_{j+1}|\mathcal{F}_j) = X_j$ ist äquivalent zu $\mathbb{E}(\Delta_{j+1}|\mathcal{F}_j) = 0$. Eine Folge mit der letztgenannte Eigenschaft nennt man *Martingaldifferenz-Folge*. Bei Vorgabe einer Martingaldifferenz-Folge $(\Delta_j)_j$ liefert dann $X_j = \Delta_1 + \dots + \Delta_j$ ein Martingal. Daher wird ein Martingal häufig in eine Partialsumme einer Differenzen-Folge umgeschrieben. Eine schöne Einführung in die Martingal-Theorie ist etwa [32]. Wir betrachten in der Folge die Situation:

Gegeben sei eine Folge $X = (X_1, \dots, X_n)$ reellwertiger, quadrat-integrierbarer Zufallsvariablen mit

$$\mathbb{E}(X_i|\mathcal{F}_{i-1}) = 0 \quad P\text{-fast sicher für } 1 \leq i \leq n,$$

wobei erneut $\mathcal{F}_j = \sigma(X_1, \dots, X_j)$ sei. Mit M_n bezeichne die Klasse aller solcher Folgen der Länge n . Wir wollen das Grenzverhalten des Martingals $S_n := \sum_{j=1}^n X_j$ untersuchen und können einen Beweis für den folgenden Satz skizzieren, der so von BOLTHAUSEN in [5] gegeben wurde.

Theorem 4.2 (BOLTHAUSEN, 1982, [5]). *Für $0 < \alpha \leq \beta < \infty$, $0 < \gamma < \infty$ gibt es eine Konstante $0 < L := L(\alpha, \beta, \gamma) < \infty$, so dass für alle $X \in M_n$ mit $\sigma_j^2 := \mathbb{E}(X_j^2|\mathcal{F}_{j-1}) = \mathbb{E}(X_j^2) =: \bar{\sigma}_j^2$ P -fast sicher und $\alpha \leq \bar{\sigma}_j^2 \leq \beta$ für $1 \leq j \leq n$ und $\max_{1 \leq j \leq n} \|X_j\|_3 \leq \gamma$ gilt*

$$\sup_{t \in \mathbb{R}} |P(S_n/s_n \leq t) - \Phi(t)| \leq L n^{-1/4},$$

wobei $s_n^2 := \sum_{j=1}^n \bar{\sigma}_j^2$.

Die Beweisskizze wird daraufhin deuten, mit ihm keine bessere Konvergenzrate als $n^{-1/4}$ erhalten zu können. Diese Rate ist im Vergleich zu der Rate bei einer Partialsumme unabhängig und identisch verteilter Zufallsvariablen allerdings schwach, wie in Kapitel 2 bereits angedeutet. Interessant ist nun, dass BOLTHAUSEN zeigen konnte, dass diese Rate im Falle eines Martingals bereits optimal ist! Dies führen wir hier allerdings nicht aus.

Beweisskizze zu Theorem 4.2. Neben $X = (X_1, \dots, X_n)$, gegeben wie im Satz, betrachten wir unabhängige normalverteilte Zufallsvariablen $Z_1, \dots, Z_n, \xi$ mit Erwartungswert 0 und Varianzen $\mathbb{E}(Z_j^2) = \bar{\sigma}_j^2$, $\mathbb{E}(\xi^2) = \sqrt{n}$. Die Bedeutung der Zufallsvariable ξ wird im Laufe der Skizze klar. Vorab schon diese Bemerkung: Wir werden S_n/s_n durch $S_n/s_n + \xi/s_n$ ersetzen, also eine Gauß-Variable ξ/s_n mit kleiner Varianz heranzufügen (im Sinne der Faltung von Maßen): Nach Voraussetzung des Satzes gilt $\frac{1}{\beta\sqrt{n}} \leq \mathbb{E}((\xi/s_n)^2) \leq \frac{1}{\alpha\sqrt{n}}$. Die Zufallsvariable $\sum_{j=1}^n Z_j/s_n$ ist $N(0, 1)$ -verteilt. Nun wenden ein Lemma in [5] an und erhalten:

$$\sup_{t \in \mathbb{R}} |P(S_n/s_n \leq t) - \Phi(t)| \leq 2 \sup_{t \in \mathbb{R}} |P(S_n/s_n + \xi/s_n \leq t) - \Phi(t)| + c n^{-1/4} \alpha^{-1/2}.$$

Das Lemma besagt, dass der Fehler durch $\|\mathbb{E}((\xi/s_n)^2)\|_\infty^{1/2}$ geschätzt werden kann. Nun schieben wir via Dreiecksungleichung $P(\sum_{j=1}^n Z_j/s_n + \xi/s_n \leq t)$ dazwischen und erhalten mittels des gleichen Lemmas

$$\sup_{t \in \mathbb{R}} |P(S_n/s_n \leq t) - \Phi(t)| \leq 2 \sup_{t \in \mathbb{R}} \left| P(S_n/s_n + \xi/s_n \leq t) - P\left(\sum_{j=1}^n Z_j/s_n + \xi/s_n \leq t\right) \right| + c n^{-1/4} \alpha^{-1/2}$$

mit einer anderen Konstante c . Nun erfolgt der bekannte Teleskop-Summen Trick von LINDBERG. Die Zufallsgröße T_m/s_n aus Kapitel 2.1 zerlegen wir hier in den Teil, der von den $(X_j)_j$ gebildet wird, und den Teil, der von den normalverteilten Zufallsvariablen $(Z_j)_j$ gebildet wird, und vergessen dabei das neue ξ nicht: Es sei für jedes $1 \leq k \leq n$

$$U_k := \sum_{j=1}^{k-1} X_j/s_n, \quad W_k := \sum_{j=k+1}^n Z_j/s_n + \xi/s_n.$$

Dann gilt

$$\begin{aligned} P(S_n/s_n + \xi/s_n \leq t) &= P\left(\sum_{j=1}^n Z_j/s_n + \xi/s_n \leq t\right) \\ &= \sum_{k=1}^n \left(P(U_k + W_k + X_k/s_n \leq t) - P(U_k + W_k + Z_k/s_n \leq t)\right). \end{aligned}$$

Nun schaut BOLTHAUSEN sehr raffiniert auf die LINDBERG-Methode: er nutzt aus, dass W_k normal-verteilt ist mit Erwartungswert 0 und Varianz $\lambda_k^2 = (\sum_{j=k+1}^n \bar{\sigma}_j^2 + \sqrt{n})/s_n^2$, also ist W_k/λ_k standardnormal-verteilt. Da W_k unabhängig ist von U_k, X_k und Z_k , kann die obige Summe auf der rechten Seite geschrieben werden als

$$\sum_{k=1}^n \mathbb{E} \left(\Phi \left(\frac{t - U_k}{\lambda_k} - \frac{X_k}{\lambda_k s_n} \right) - \Phi \left(\frac{t - U_k}{\lambda_k} - \frac{Z_k}{\lambda_k s_n} \right) \right).$$

Nun ist Φ quasi eine Testfunktion und wir betrachten wie üblich die TAYLOR-Entwicklung der letzten Summe:

$$\begin{aligned} (4.4) \quad & \sum_{k=1}^n \mathbb{E} \left(\left(-\frac{X_k}{\lambda_k s_n} + \frac{Z_k}{\lambda_k s_n} \right) \varphi_{0,1} \left(\frac{t - U_k}{\lambda_k} \right) + \left(\frac{X_k^2}{2\lambda_k^2 s_n^2} - \frac{Z_k^2}{2\lambda_k^2 s_n^2} \right) \varphi'_{0,1} \left(\frac{t - U_k}{\lambda_k} \right) \right. \\ & \left. - \frac{X_k^3}{6\lambda_k^3 s_n^3} \varphi''_{0,1} \left(\frac{t - U_k}{\lambda_k} - \theta_k \frac{X_k}{\lambda_k s_n} \right) + \frac{Z_k^3}{6\lambda_k^3 s_n^3} \varphi''_{0,1} \left(\frac{t - U_k}{\lambda_k} - \theta'_k \frac{Z_k}{\lambda_k s_n} \right) \right) \end{aligned}$$

mit $0 \leq \theta_k, \theta'_k \leq 1$. Im Fall unabhängiger Zufallsvariablen verschwanden die ersten beiden Summanden bei der Bildung des Erwartungswertes. Hier kann nun mittels bedingter Erwartungswerte wie folgt argumentiert werden.

$$\mathbb{E} \left(\left(-\frac{X_k}{\lambda_k s_n} + \frac{Z_k}{\lambda_k s_n} \right) \varphi_{0,1} \left(\frac{t - U_k}{\lambda_k} \right) \right) = \mathbb{E} \left(\mathbb{E} \left(\left(-\frac{X_k}{\lambda_k s_n} + \frac{Z_k}{\lambda_k s_n} \right) \varphi_{0,1} \left(\frac{t - U_k}{\lambda_k} \right) \middle| \mathcal{F}_{k-1} \right) \right).$$

Nun ist U_k messbar bezüglich $\mathcal{F}_{k-1}$. Also kann $\varphi_{0,1}(\frac{t-U_k}{\lambda_k})$ aus der bedingten Erwartung faktorisiert werden. Da nach Voraussetzung $\mathbb{E}(X_k|\mathcal{F}_{k-1}) = 0$ fast-sicher und $\mathbb{E}(Z_k|\mathcal{F}_{k-1}) = \mathbb{E}(Z_k) = 0$, verschwindet der erste Summand in (4.4). Der zweite Summand in (4.4) verschwindet, da dasselbe Faktorisierungsargument verwendet werden kann und nach Voraussetzung $\mathbb{E}(X_k^2|\mathcal{F}_{k-1}) = \bar{\sigma}_k^2 = \mathbb{E}(Z_k^2)$ gilt. Da $\varphi_{0,1}$ und seine Ableitungen beschränkt sind und $\max_{1 \leq j \leq n} \|X_j\|_3 \leq \gamma$ nach

Voraussetzung, folgt insgesamt

$$\sup_{t \in \mathbb{R}} |P(S_n/s_n \leq t) - \Phi(t)| \leq c \sum_{k=1}^n \lambda_k^{-3} s_n^{-3} + c' n^{-1/4}$$

für Konstante c, c' , die nur von α, β und γ abhängen. Nun erkennt man die Wahl der Varianz $\mathbb{E}(\xi^2) = \sqrt{n}$, denn es folgt die Aussage des Satzes. $\square$

Erneut wurde ein elementarer Beweis eines zentralen Grenzwertsatzes, hier für Martingale, vorgestellt, der gleichzeitig eine Konvergenzrate liefert (die hier sogar optimal ist, ohne dass wir dies begründet haben). Wohl kaum eine andere Beweismethode ist so elegant und wird von uns daher für eine Vorlesung über Wahrscheinlichkeitstheorie empfohlen.

5. DIE LINDEBERGSCHES METHODE ALS INVARIANZPRINZIP

Die schon in Kapitel 4 beobachtete Robustheit der LINDEBERG-Methode unter Änderungen der Limesverteilung ist der Ausgangspunkt zu einer weitreichenden Entwicklung in den vergangenen fünf Jahren, die man mit Fug und Recht als ein Renaissance der LINDEBERG-Methode bezeichnen kann. Diese Wiedergeburt geht unter anderem auf Arbeiten von CHATTERJEE zurück ([7], [8]). Eine wesentliche Neuerung ist hierbei, dass CHATTERJEE die Grenzverteilung von allgemeineren Funktionen g eines Zufallsvektors $X = (X_1, \dots, X_n)$ betrachtet. Diese Funktion g kann die Gestalt wie unter (2.10) haben, muss es aber nicht. Zudem besteht dann natürlich auch keine Notwendigkeit, dass der Limesvektor eine Gaußsche Struktur aufweisen muss, auch wenn dies in unserem wichtigsten Beispiel der Fall sein wird. CHATTERJEE formuliert den folgenden Satz.

Theorem 5.1 (Chatterjee, 2006, [8]). *Es seien $X = (X_1, \dots, X_n)$ und $Z = (Z_1, \dots, Z_n)$ zwei Vektoren von unabhängigen Zufallsvariablen mit endlichem zweiten Moment mit $\mathbb{E}X_i = \mathbb{E}Z_i$ und $\mathbb{E}X_i^2 = \mathbb{E}Z_i^2$ und $g : \mathbb{R}^n \rightarrow \mathbb{R}$ dreimal differenzierbar in jeder Komponente. Wir setzen $U = g(X)$ und $V = g(Z)$. Dann gilt für jede dreimal stetig differenzierbare Abbildung $f : \mathbb{R} \rightarrow \mathbb{R}$ und jedes $K > 0$*

$$(5.1) \quad \begin{aligned} |\mathbb{E}f(U) - \mathbb{E}f(V)| &\leq C_1(f)\kappa_2(g) \sum_{i=1}^n [\mathbb{E}(X_i^2; |X_i| > K) + \mathbb{E}(Z_i^2; |Z_i| > K)] \\ &\quad + C_2(f)\kappa_3(g) \sum_{i=1}^n [\mathbb{E}(X_i^3; |X_i| \leq K) + \mathbb{E}(Z_i^3; |Z_i| \leq K)]. \end{aligned}$$

Hierbei verwenden wir die Notation $\mathbb{E}(X; |X| > K) := \int_{|X| > K} X dP$. Weiter sind mit der bereits verwendeten Notation $M_i := \sup_{x \in \mathbb{R}} |f^{(i)}(x)|$ die Konstanten $C_1(f) := M_1 + M_2$ und $C_2(f) := \frac{1}{6}M_1 + \frac{1}{2}M_2 + \frac{1}{6}M_3$, und für $r = 2, 3$

$$\kappa_r(g) := \sup\{|\partial_i^p g(x)|^{r/p} : 1 \leq i \leq n, 1 \leq p \leq r, x \in \mathbb{R}^n\}.$$

Die Funktionen f in diesem Satz übernehmen die gleiche Rolle wie im Kapitel 2. Wählt man wie dort $g(x) = \frac{1}{\sqrt{n}} \sum_{i=1}^n x_i$, so lässt sich $\kappa_2(g) = \frac{1}{n}$ und $\kappa_3(g) = \frac{1}{n^{3/2}}$ bestimmen und man erhält für $(X_i)_i$ und $(Z_i)_i$, die sogar unabhängig und identisch verteilt mit Erwartungswert 0 und Varianz 1 sind, und für alle dreifach differenzierbaren f :

$$\left| \mathbb{E}f\left(\frac{1}{\sqrt{n}} \sum_{i=1}^n X_i\right) - \mathbb{E}f\left(\frac{1}{\sqrt{n}} \sum_{i=1}^n Z_i\right) \right| \leq C_1(f)[\mathbb{E}(X_1^2; |X_1| > \varepsilon\sqrt{n}) + \mathbb{E}(Z_1^2; |Z_1| > \varepsilon\sqrt{n})] + 2C_2(f)\varepsilon,$$

wenn man $K = \varepsilon\sqrt{n}$ wählt. Dies ist bis auf Bezeichnungsunterschiede das Resultat des Teilkapitels 2.1. Wir werden nun sehen, dass sich Satz 5.1 dem Wesen nach ebenso beweisen lässt wie der zentrale Grenzwertsatz von LINDBERG.

Beweis von Satz 5.1. Wir betrachten die Funktion $h = f \circ g$ und berechnen ihre partiellen Ableitungen als

$$\begin{aligned} \partial_i^2 h(x) &= f'(g(x))\partial_i^2 g(x) + f''(g(x))(\partial_i g(x))^2, \\ \partial_i^3 h(x) &= f'(g(x))\partial_i^3 g(x) + 3f''(g(x))\partial_i g(x)\partial_i^2 g(x) + f'''(g(x))(\partial_i g(x))^3. \end{aligned}$$

Mit den obigen Bezeichnungen erhalten wir daher für alle i und x : $|\partial_i^2 h(x)| \leq C_1(f)\kappa_2(g)$ und $|\partial_i^3 h(x)| \leq 6C_2(f)\kappa_3(g)$. Nun wenden wir den Ersetzungstrick an, der das Herzstück der LINDBERG-Methode bildet. Wir setzen für $0 \leq i \leq n$ $Y_i := (X_1, \dots, X_{i-1}, X_i, Z_{i+1}, \dots, Z_n)$ und $W_i := (X_1, \dots, X_{i-1}, 0, Z_{i+1}, \dots, Z_n)$, und definieren

$$\begin{aligned} R_i &:= h(Y_i) - h(W_i) - X_i \partial_i h(W_i) - \frac{1}{2} X_i^2 \partial_i^2 h(W_i) \quad \text{und} \\ T_i &:= h(Y_{i-1}) - h(W_i) - Z_i \partial_i h(W_i) - \frac{1}{2} Z_i^2 \partial_i^2 h(W_i). \end{aligned}$$

In Kapitel 2.1 hatten wir immer Partialsummen der $(X_i)_i$ untersucht und dort die Summe der Elemente in W_i mit T_i bezeichnet (womit in diesem Fall die Summe der Elemente in Y_i übereinstimmt mit $T_i + X_i$). Ähnlich wie im zweiten Kapitel erhält man mittels TAYLOR-Entwicklung die Schranken

$$\begin{aligned} |R_i| &\leq \min\{C_2(f)\kappa_3(g)|X_i|^3, C_1(f)\kappa_2(g)|X_i|^2\} \quad \text{und} \\ |T_i| &\leq \min\{C_2(f)\kappa_3(g)|Z_i|^3, C_1(f)\kappa_2(g)|Z_i|^2\}. \end{aligned}$$

Aufgrund der Unabhängigkeit von X_i, Z_i und W_i für jedes i bekommt man

$$\mathbb{E}[X_i \partial_i g(W_i)] - \mathbb{E}[Z_i \partial_i g(W_i)] = 0 \quad \text{und} \quad \mathbb{E}[X_i^2 \partial_i^2 g(W_i)] - \mathbb{E}[Z_i^2 \partial_i^2 g(W_i)] = 0.$$

Zusammen erhält man

$$\begin{aligned}
|\mathbb{E}f(U) - \mathbb{E}f(V)| &= \left| \sum_{i=1}^n \mathbb{E}(h(Y_i) - h(Y_{i-1})) \right| \\
&= \left| \sum_{i=1}^n \mathbb{E} \left(X_i \partial_i h(W_i) + \frac{1}{2} X_i^2 \partial_i^2 h(W_i) + R_i \right) \right. \\
&\quad \left. - \sum_{i=1}^n \mathbb{E} \left(Z_i \partial_i h(W_i) + \frac{1}{2} Z_i^2 \partial_i^2 h(W_i) + T_i \right) \right| \\
&\leq C_1(f) \kappa_2(g) \sum_{i=1}^n [\mathbb{E}(X_i^2; |X_i| > K) + \mathbb{E}(Z_i^2; |Z_i| > K)] \\
&\quad + C_2(f) \kappa_3(g) \sum_{i=1}^n [\mathbb{E}(X_i^3; |X_i| \leq K) + \mathbb{E}(Z_i^3; |Z_i| \leq K)].
\end{aligned}$$

□

Man kann und sollte sich natürlich fragen, was man mit einer solchen Verallgemeinerung gewinnt. Wir wollen eine Anwendung von Satz 5.1 kennenlernen, die ein in der modernen Wahrscheinlichkeitstheorie viel beachtetes Gebiet berührt, die Theorie *zufälliger Matrizen*. Unter einer zufälligen Matrix verstehen wir dabei eine Matrix, deren Einträge reell- oder komplexwertige Zufallsvariablen sind. Wir wollen uns in der Folge auf symmetrische $N \times N$ Matrizen M_N beschränken. Deren Eigenwerte sind offensichtlich alle reell und man kann sich nach ihrer Verteilung fragen. Hierzu betrachten wir zunächst Matrizen $M_N := M_N(X) := (X(i, j)/\sqrt{N})_{1 \leq i, j \leq N}$, wobei die $X(i, j)$ für $i \leq j$ unabhängige und identisch verteilte Zufallsvariablen sind. Man definiert die empirische Eigenwertverteilung der Eigenwerte $\lambda_1 \leq \dots \leq \lambda_N$ von M_N als

$$F_N(t) := \frac{1}{N} \#\{i : \lambda_i \leq t\}.$$

WIGNER betrachtete BERNOULLI-verteilte $X(i, j)$ und zeigte in [29], dass F_N schwach in Wahrscheinlichkeit gegen das sogenannte Halbkreisgesetz mit Dichte $\frac{1}{2\pi} \sqrt{4 - x^2}$ auf $[-2, 2]$ konvergiert. Er erkannte in [30], dass dies auch für normal-verteilte $X(i, j)$ und weitere Verteilungen gilt. Die Gültigkeit des Halbkreisgesetzes für unabhängige und identisch verteilte $X(i, j)$ (unter Momentebedingungen) geht auf ARNOLD [2] zurück. In der Sprache der Physiker konnte damit *Universalität* des Halbkreisgesetzes gezeigt werden: zunächst entdeckt man Gesetze für spezielle Verteilungen der $X(i, j)$ (in der Regel normalverteilt), dann untersucht man, ob die Gesetze von der speziellen Verteilung abhängen. Situation mit abhängigen $X(i, j)$ wurden u.a. in [3], [17] und [15], [14] studiert. Wir wollen uns hier fragen, unter welchen Bedingungen an die $X(i, j)$ sich das WIGNERSche Halbkreisgesetz als Limesverteilung von F_N herleiten lässt. Hierbei wird die Anwendung von Satz 5.1 zur Entdeckung einer geeigneten LINDEBERG-Bedingung führen.

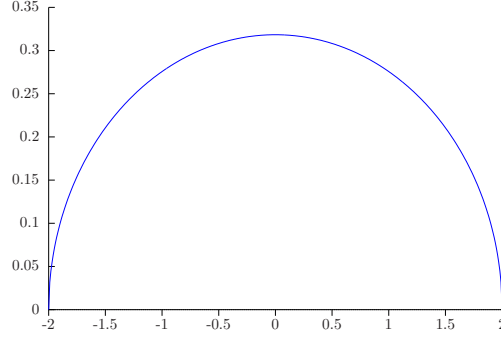


ABBILDUNG 7. Dichte der Halbkreis-Verteilung

Es gibt verschiedene Arten, WIGNERS Resultat zu zeigen. Eine wichtige Methode beruht auf der Resolventen $G(X, z) := (M_N(X) - zI)^{-1}$ der Zufallsmatrix $M_N(X)$, wobei $z = u + iv \in \mathbb{C}$ und I die $N \times N$ -Einheitsmatrix ist. Man betrachtet $g(X, z) = \frac{1}{N} \text{Tr} G(X, z)$, wobei Tr für die Spur einer Matrix steht. Konvergiert $g(X, z)$ für einen zufälligen Vektor X (in $\mathbb{R}^{N(N+1)/2}$, der Symmetrie der Matrix wegen) und jedes feste z in Wahrscheinlichkeit gegen die STIELTJES-Transformierte des Halbkreis-Gesetzes, so impliziert dies die behauptete schwache Konvergenz in Wahrscheinlichkeit der empirischen Eigenwertverteilungsfunktion F_N gegen das Halbkreisgesetz. Genauer gilt: F_N konvergiert genau dann (in Wahrscheinlichkeit) gegen die Verteilungsfunktion einer Wahrscheinlichkeitsverteilung Q , falls für alle $z \in \mathbb{C} \setminus \mathbb{R}$ gilt

$$g(X, z) \rightarrow \int_{-\infty}^{\infty} \frac{1}{x - z} Q(dx) \quad \text{wenn } N \rightarrow \infty.$$

Der Ausdruck auf der rechten Seite ist die STIELTJES-Transformierte von Q in z . Man kann die STIELTJES-Transformierte der Halbkreisverteilung ausrechnen und erhält die Funktion $-\frac{1}{2}(z - \sqrt{z^2 - 4})$. Der Name für die obige Funktion g ist nicht zufällig gewählt, denn auf genau dieses g wollen wir Satz 5.1 anwenden. Dazu sei $z = u + iv \in \mathbb{C}$ mit $v \neq 0$ fest. Da gemäß der klassischen Linearen Algebra die Eigenwerte von $M_N(X)$ sämtlich reell sind und daher $\det(M_N(X) - zI) \neq 0$ gilt und sich die Inverse einer Matrix in diesem Fall als rationale Funktion der Matrixeinträge berechnen lässt (adjungieren, Determinante bilden, etc. sind rationale Funktionen), ist $G : \mathbb{R}^n \rightarrow \mathbb{C}^{N \times N}$ mit $n = N(N + 1)/2$ bezüglich aller n Matrixeinträge differenzierbar. Weiter setzen wir $g : \mathbb{R}^n \rightarrow \mathbb{R}$ als

$$g(x) := \frac{1}{N} \text{Tr}((M_N(x) - zI)^{-1}).$$

Aus der definitionsgemäßen Identität $(M_N(x) - I)G(x) = I$ für jedes x folgt $\frac{\partial}{\partial x(i,j)}(M_N(x) - I)G(x) = 0$, und daher

$$\frac{\partial G}{\partial x(i,j)} = -G \frac{\partial M_N}{\partial x(i,j)} G.$$

Da zudem alle zweifachen Ableitungen von $M_N(x)$ nach einem der Matrixeinträge identisch verschwinden, ergibt sich

$$(5.2) \quad \frac{\partial g}{\partial x(i, j)} = -\frac{1}{N} \text{Tr} \left(\frac{\partial M_N}{\partial x(i, j)} G^2 \right),$$

$$(5.3) \quad \frac{\partial^2 g}{\partial x(i, j)^2} = \frac{2}{N} \text{Tr} \left(\frac{\partial M_N}{\partial x(i, j)} G \frac{\partial M_N}{\partial x(i, j)} G^2 \right),$$

$$(5.4) \quad \frac{\partial^3 g}{\partial x(i, j)^3} = -\frac{6}{N} \text{Tr} \left(\frac{\partial M_N}{\partial x(i, j)} G \frac{\partial M_N}{\partial x(i, j)} G \frac{\partial M_N}{\partial x(i, j)} G^2 \right).$$

Um diese Größen zu beschränken erinnern wir an die HILBERT-SCHMIDT-Norm einer Matrix $A = (a_{i,j})_{i,j=1}^N$ definiert als $\|A\| := (\sum_{i,j} a_{i,j}^2)^{1/2}$. Für diese gilt für jede $N \times N$ Matrix A

$$|\text{Tr}(AB)| \leq \|A\| \|B\|,$$

$$\|UA\| = \|AU\| = \|A\| \quad \text{für jede unitäre } N \times N \text{ Matrix } U,$$

$$\max\{\|BA\|, \|AB\|\} \leq \max_i |\lambda_i| \cdot \|A\| \quad \text{für jede normale Matrix } B \text{ mit Eigenwerten } \lambda_1, \dots, \lambda_N$$

All dieses findet sich beispielsweise in [31]. Man überprüft nun, dass die Matrizen G sowie alle Ableitungen von $M_N(X)$ normal sind. Nach dem Satz von GERSCHGORIN sind die Eigenwerte von G durch $1/|v|$ beschränkt, wobei $v = \text{Im}(z)$. Zieht man zudem die einfache Struktur von $\partial M_N \partial x(i, j)$ in Betracht, so ergibt sich $\|\frac{\partial g}{\partial x(i, j)}\|_\infty \leq \frac{2}{N^{3/2}|v|^2}$. Ähnlich sieht man, dass

$$\left\| \frac{\partial^2 g}{\partial x(i, j)^2} \right\|_\infty \leq \frac{4}{N^2|v|^3} \quad \text{und} \quad \left\| \frac{\partial^3 g}{\partial x(i, j)^3} \right\|_\infty \leq \frac{12}{N^{5/2}|v|^4}.$$

gilt. Damit ergibt sich

$$\kappa_2(g) \leq 4 \max\{|v|^{-3}, |v|^{-4}\} N^{-2} \quad \text{und} \quad \kappa_3(g) \leq 12 \max\{|v|^{-4}, |v|^{-6}\} N^{-5/2}.$$

Die Idee ist nun mithilfe von Satz 5.1 zu zeigen, dass unter einer LINDEBERG-Bedingung die oben definierte Funktion $g(X)$ nahe an der entsprechenden Größe $g(Z)$ ist, wenn man für Z einen Vektor in $\mathbb{R}^n$ mit i.i.d. standard-normal-verteilten Einträgen wählt. Von letzterer wissen wir nach dem Satz von WIGNER, dass sie gegen die STIELTJES-Transformierte des Halbkreisgesetzes konvergiert. Es sei also X eine symmetrische Matrix mit unabhängigen Einträgen $X(i, j)$ für $i \leq j$. Weiter sei Z eine symmetrische Matrix mit standard-normal-verteilten Einträgen $Z(i, j)$, die oberhalb der Diagonalen unabhängig sind. Sei $U = \text{Re } g(X)$ und $V = \text{Re } g(Z)$ und f sei dreimal differenzierbar von $\mathbb{R}$ nach $\mathbb{R}$. $\text{Re } g$ ist eine differenzierbare Funktion und es gilt $\kappa_r(\text{Re } g) \leq \kappa_r(g)$ für alle r . Setzt man (wie oben) $K = \varepsilon\sqrt{N}$, so ergibt Satz 5.1, dass es eine Konstante $C(v, f)$ gibt, sodass $|\mathbb{E}f(U) - \mathbb{E}f(V)|$ beschränkt ist durch

$$C(v, f) \frac{1}{N^2} \sum_{1 \leq i \leq j \leq N} [\mathbb{E}(X^2(i, j); |X(i, j)| > \varepsilon\sqrt{N}) + \mathbb{E}(Z^2(i, j); |Z(i, j)| > \varepsilon\sqrt{N})] + \varepsilon.$$

Das Argument lässt sich für den Imaginärteil von g wiederholen. Damit der letzte Term gegen 0 geht, muss also genau das Verschwinden des ersten der beiden Summanden fordern. Also konvergiert auch die empirische Eigenwertverteilung einer Matrix $M_N(X)$ mit zentrierten Einträgen

$X(i, j)$ mit Varianz 1, die oberhalb der Diagonalen unabhängig sind, gegen das Halbkreisgesetz, wenn die folgende LINDBERG-Bedingung gilt: Für alle $\varepsilon > 0$ ist

$$\lim_{N \rightarrow \infty} \frac{1}{N^2} \sum_{i \leq j=1}^N [\mathbb{E}(X^2(i, j); |X(i, j)| > \varepsilon \sqrt{N})] = 0.$$

In der Literatur ist diese Bedingung als die PASTUR-Bedingung bekannt, siehe [24]. Sie ist etwa im Fall von identisch verteilten $X(i, j)$ erfüllt. Satz 5.1 findet weitere sehr interessante Anwendungen, zum Beispiel in der Theorie der Spin-Gläser sowie bei der Untersuchung von Maxima von Zufallsfeldern, siehe [7]. Wir betrachten dies hier nicht.

6. DAS 4-MOMENTE THEOREM VON TAO UND VU

Wir erinnern an die Situation in Kapitel 5. Gegeben ist eine symmetrische Matrix $M_N := M_N(X) := (X(i, j)/\sqrt{N})_{1 \leq i, j \leq N}$, wobei die $X(i, j)$ für $i \leq j$ unabhängige Zufallsvariablen sind. Das Halbkreisgesetz liefert unter Momentebedingungen für die *globale* Statistik $F_N(t) := \frac{1}{N} \#\{i : \lambda_i \leq t\}$ universell eine Limesverteilung. Seit Beginn der Untersuchung von Zufallsmatrizen wurden viele andere Statistiken, gebildet aus den zufälligen Eigenwerten $(\lambda_i)_i$, untersucht. So interessiert etwa die Verteilung von Lücken aufeinanderfolgender Eigenwerte (wie viele $1 \leq i \leq N$ gibt es mit $\lambda_{i+1} - \lambda_i \leq s$). Kann die Korrelation von k Eigenwerten im Limes beschrieben werden (die sogenannte k -Punkt Korrelationsfunktion) ? Was ist die Verteilung einzelner Eigenwerte λ_i , kann die Limesverteilung der gemeinsamen Verteilung von k Eigenwerten $(\lambda_{i_1}, \dots, \lambda_{i_k})$ gefunden werden? Die genannten Statistiken nennt man *lokale* Eigenwert-Statistiken, ihre Untersuchung ist in der Regel komplizierter als die der globalen Statistiken. Beispiele weiterer globaler Statistiken sind die Determinante der Matrix M_N oder die Anzahl der Eigenwerte $(\lambda_i)_i$, die in einem vorgegebenen Intervall liegen.

Wir betrachten nun eine $N \times N$ WIGNER-hermitesche Matrix $W_N = (X(i, j))_{1 \leq i \leq j \leq N}$: dies ist eine hermitesche Matrix mit unabhängigen $X(i, j)$ und $X(i, j) = \bar{X}(i, j)$. Für $i < j$ seien die $X(i, j)$ identisch verteilt mit Erwartungswert 0 und Varianz 1, für $i = j$ seien die $X(i, i)$ ebenfalls identisch verteilt mit Erwartungswert 0 und Varianz σ^2 . Es seien Realteil und Imaginärteil der Zufallsvariablen unabhängig. Desweiteren existiere eine Konstante C_0 (unabhängig von i, j und N) mit $\mathbb{E}|X(i, j)|^{C_0} \leq C$ für jedes i, j und eine Konstante C , die ebenfalls unabhängig von i, j und N sei. Die Skalierung $M_N := \frac{1}{\sqrt{N}} W_N$ platziert die Eigenwerte in das beschränkte Intervall $[-2, 2]$, während $A_N := \sqrt{N} W_N$ die Abstände (*spacings*) zweier Eigenwerte grob konstant hält. Ein wichtiger Spezialfall einer WIGNER-hermiteschen Matrix ist die Wahl von normalverteilten Einträgen (auf der Diagonalen reelle $N(0, 1)$ verteilte Zufallsvariable, sonst komplexwertige $N(0, 1)$ verteilte). Man nennt dies das *Gaußsche unitäre Ensemble* (GUE), denn die Verteilung von W_N ist dann invariant unter Konjugation durch unitäre Matrizen. Der wichtigste Vorteil des GUE ist, dass nach der Formel von GINIBRE die gemeinsame Verteilung der nicht geordneten

Eigenwerte geschlossen durch den Ausdruck

$$\varrho(\lambda_1, \dots, \lambda_N) = Z_N^{-1} \prod_{1 \leq i < j \leq N} |\lambda_i - \lambda_j|^2 \exp\left(-\frac{1}{2} \sum_{i=1}^n x_i^2\right)$$

mit einer geeigneten Normierung Z_N dargestellt werden kann. Viele grundlegende Gesetze für GUE können in dem Buch [1] gefunden werden. Ein vereinfachter Rückblick auf viele Arbeiten der vergangenen Jahre besagt, dass die Formel von GINIBRE die Untersuchung der Limesverteilung diverser lokaler Statistiken der Eigenwerte eines GUE ermöglichte, wobei hierzu diverse Tools (etwa die Theorie von Determinanten-Punktprozessen oder die Methode der orthogonalen Polynome) verwendet wurden, die hier aber nicht dargestellt werden. Wir beschränken uns auf die Nennung der folgenden Resultate für GUE Matrizen: bezeichnet $\lambda_1(M_N)$ den kleinsten Eigenwert einer GUE Matrix, so konvergiert $(\lambda_1(M_N) + 2)N^{2/3}$ gegen die sogenannte TRACY-WIDOM Verteilung (lokale Statistik). Bezeichnet $N_I(M_N)$ die Anzahl der Eigenwerte der GUE Matrix M_N , die im Intervall I liegen, so konvergiert die standardisierte Zufallsvariable $(N_I(M_N) - \mathbb{E}(N_I(M_N)))/\sqrt{\mathbb{V}(N_I(M_N))}$ gegen eine standard-normalverteilte Zufallsvariable, wann immer $\mathbb{V}(N_I(M_N)) \rightarrow \infty$ für $N \rightarrow \infty$ (globale Statistik). Es gilt sogar lokal ein zentraler Grenzwertsatz: Für $i(N)$ so, dass $i(N)/N \rightarrow c$ für $N \rightarrow \infty$ und $c \in (0, 1)$, konvergiert $(\lambda_{i(N)} - \alpha(i(N)))/(\beta(i(N)))$ gegen eine standard-normalverteilte Zufallsvariable, wobei $\alpha(i(N))$ und $\beta(i(N))$ geeignete Skalierungen sind, die den erwarteten Ort von $\lambda_{i(N)}$ sowie die Standardabweichung dieses Ortes angeben. Für die genaue Wahl dieser Skalierungen siehe [18].

Es wurde seit langem vermutet, dass diese (und viele andere) Verteilungsgesetze *universell* für alle WIGNER-hermiteschen Matrizen oder für noch größere Matrizenklassen gültig sind. Neben vielen mathematisch sehr anspruchsvollen Arbeiten auf dem Weg zu Beweisen dieser Universalität betrachten wir hier die bahnbrechende Arbeit von TAO und VU [26], in der einige der Universalitätsfragen gelöst werden konnten. Die Arbeit basiert auf der Methode von LINDEBERG. Man betrachte zwei unabhängige WIGNER-hermitesche Matrizen M_N und M'_N . Für diverse Statistiken F soll $\mathbb{E}(F(M_N)) - \mathbb{E}(F(M'_N))$ kontrolliert werden. Betrachte die Matrix $\tilde{M}_N$, die aus M_N dadurch gebildet wird, dass entweder einer der Diagonaleinträge $X(i, i)$ von M_N durch den korrespondierenden Eintrag $X'(i, i)$ von M'_N ausgetauscht wird, oder einer der nicht-Diagonaleinträge $X(i, j)$ von M_N durch den korrespondierenden Eintrag $X'(i, j)$ von M'_N ausgetauscht wird (und damit auch $X(j, i)$ durch $X'(j, i)$). Kann man nun zeigen, dass $\mathbb{E}(F(M_N)) - \mathbb{E}(F(\tilde{M}_N)) = o(1/n)$ bei Austausch eines Diagonalelements und $\mathbb{E}(F(M_N)) - \mathbb{E}(F(\tilde{M}_N)) = o(1/n^2)$ bei Austausch eines nicht-Diagonalelements, so würde der Teleskop-Ansatz von LINDEBERG

$$(6.1) \quad \mathbb{E}(F(M_N)) - \mathbb{E}(F(M'_N)) = o(1)$$

implizieren. In Rückblick auf Kapitel 2 wählen wir ein GUE-Element M'_N , tauschen also sukzessive die Matricelemente durch Gauß-verteilte Größen. Kapitel 5 legt alternativ nahe, allgemeiner M'_N aus der gleichen Matrizen-Klasse zuzulassen. Das 4-Momente Theorem von TAO und VU besagt nun grob, dass man (6.1) erreichen kann, sobald die beiden Matrizen M_N und M'_N so gewählt sind, dass die ersten 4 Momente der Matrix-Einträge übereinstimmen. *Wie-so 4?* Man kann dies heuristisch bei einem Rückblick auf Kapitel 2.1 schnell motivieren: dort konnte man bei der Annahme, dass die ersten beiden Momente der Summanden X_i mit der $N(0, 1)$ -Verteilung übereinstimmen, sehen, dass beim Austausch eines Summanden ein Fehler der Größe $O(1/n^{3/2})$ entstand, und bei n -maligem Tauschen so die Rate $O(1/n^{1/2})$ entstand. Jedes weitere Moment der Übereinstimmung mit $N(0, 1)$ lieferte eine Verbesserung der Fehlerrate um $O(1/n^{1/2})$, bei 4 Momenten der Übereinstimmung wäre beim Austausch eines Summanden ein Fehler der Größe $O(1/n^{5/2})$ entstanden. Da wir bei Matrizen nun grob n^2 mal austauschen, wäre dies ausreichend, um (6.1) erzielen zu können. Es wundert somit im Folgenden auch nicht, dass die Einträge von M_N und M'_N auf der Diagonale in den Momenten bis zur 2. Ordnung, und bis zur 4. Ordnung ausserhalb der Diagonalen übereinstimmen sollen. Zunächst geben wir eine genaue Definition der Übereinstimmung von Momenten:

Definition 6.1. *Zwei komplexwertige Zufallsvariablen X und Y stimmen in den Momenten bis zur Ordnung k überein, wenn*

$$\mathbb{E}(\operatorname{Re}(X)^a \operatorname{Im}(X)^b) = \mathbb{E}(\operatorname{Re}(Y)^a \operatorname{Im}(Y)^b)$$

für jedes $a, b \geq 0$ mit $a + b \leq k$. Unter der Annahme, dass Real- und Imaginärteil von X und Y unabhängige Zufallsvariablen sind, vereinfacht sich diese Bedingung zu $\mathbb{E}(\operatorname{Re}(X)^a) = \mathbb{E}(\operatorname{Re}(Y)^a)$ und $\mathbb{E}(\operatorname{Im}(X)^b) = \mathbb{E}(\operatorname{Im}(Y)^b)$ für jedes $0 \leq a, b \leq k$.

Theorem 6.2 (4 Momente Theorem von TAO und VU). *Es sei $c_0 > 0$ hinreichend klein gewählt. Seien W_N und W'_N zwei WIGNER-hermitesche Matrizen. Die Momente der Nicht-Diagonaleinträge beider Matrizen mögen bis zur Ordnung 4 und die der Diagonaleinträge bis zur Ordnung 2 übereinstimmen. Es sei $A_N := \sqrt{N}W_N$ und $A'_N := \sqrt{N}W'_N$, sei $1 \leq k \leq N^{c_0}$ eine natürliche Zahl und $G_N : \mathbb{R}^k \rightarrow \mathbb{R}$ eine glatte Funktion mit*

$$(6.2) \quad |\nabla^j G_N(x)| \leq N^{c_0}$$

für alle $0 \leq j \leq 5$ und $x \in \mathbb{R}^k$. Dann gilt für N hinreichend groß:

$$|\mathbb{E}(G(\lambda_{i_1}(A_N), \dots, \lambda_{i_k}(A_N))) - \mathbb{E}(G(\lambda_{i_1}(A'_N), \dots, \lambda_{i_k}(A'_N)))| \leq N^{-c_0}.$$

Eine aktuelle Übersicht über eine ganze Klasse von diversen 4-Momente-Theoremen findet man in [27]. Wir skizzieren den Beweis-Ansatz. Wir bilden die Matrix $\tilde{M}_N$ aus M_N , indem ein einzelner Eintrag $X(p, q)$ für ein $p < q$ von M_N durch den Eintrag $X'(p, q)$ von M'_N ersetzt wird und entsprechend $X(q, p)$, um $\tilde{M}_N$ hermitesch zu halten. Ein erster technischer Punkt ist, dass

damit $\tilde{M}_N$ keine WIGNER-Matrix mehr ist, da die Einträge nicht mehr notwendig identisch verteilt sind. Wir betrachten $\tilde{A}_N = \sqrt{N} \tilde{M}_N$ und wollen den Beweis von

$$\mathbb{E}(G(\lambda_{i_1}(A_N), \dots, \lambda_{i_k}(A_N))) = \mathbb{E}(G(\lambda_{i_1}(\tilde{A}_N), \dots, \lambda_{i_k}(\tilde{A}_N))) + O(N^{-5/2+O(c_0)})$$

skizzieren (denn der Teleskop-Ansatz verwendet $O(n^2)$ Abschätzungen dieser Bauart). Wir schreiben $A_N = A(X(p, q))$ und $\tilde{A}_N = A(X'(p, q))$ mit $A(t) := A(0) + tA'(t)$, wobei $A(0)$ eine WIGNER-Matrix ist, bei der ein Eintrag (und sein adjungierter Eintrag) Null ist, und $A'(0)$ die hermitesche Matrix $A'(0) = e_p e_q^* + e_p^* e_q$. Wir betrachten

$$F(t) := \mathbb{E}(G(\lambda_{i_1}(A(t)), \dots, \lambda_{i_k}(A(t))))$$

und wollen somit

$$(6.3) \quad \mathbb{E}F(X(p, q)) = \mathbb{E}F(X'(p, q)) + O(N^{-5/2+O(c_0)})$$

zeigen. Wie verändert der Austausch eines Matrix-Elementes den Eigenwert? Angenommen, wir können für jedes $1 \leq l \leq k$ eine TAYLOR-Entwicklung der Bauart

$$(6.4) \quad \lambda_{i_l}(A(t)) = \lambda_{i_l}(A(0)) + \sum_{j=1}^4 c_{l,j} t^j + O(N^{-5/2+O(c_0)})$$

zeigen mit Koeffizienten $c_{l,j} = O(N^{-j/2+O(c_0)})$, dann liefert (6.2)

$$F(t) = F(0) + \sum_{j=1}^4 f_j t^j + O(N^{-5/2+O(c_0)}),$$

falls die Koeffizienten $f_j = O(N^{-j/2+O(c_0)})$. Wenn wir nun aber t durch $X(p, q)$ ersetzen und den Erwartungswert bilden (und dabei darauf hoffen, dass f_j nur von F und $A(0)$ abhängt und damit unabhängig von $X(p, q)$ ist), folgt

$$\mathbb{E}F(X(p, q)) = \mathbb{E}F(0) + \sum_{j=1}^4 \mathbb{E}(f_j) \mathbb{E}(X(p, q)^j) + O(N^{-5/2+O(c_0)}).$$

Analog folgt dies für $\mathbb{E}F(X'(p, q))$, und da die ersten vier Momente nach Voraussetzung übereinstimmen sollen, folgt (6.3). Die technisch große Leistung von TAO und VU ist es, (6.4) hergeleitet zu haben. Dort liegt die eigentliche Schwierigkeit und wird mit Hilfe der Anwendung von Variationsformeln von HADAMARD überwunden. Wenn mit $u_i(A(t))$ der i -te Vektor einer Orthogonal-Basis von Eigenvektoren von $A(t)$ bezeichnet wird, so gilt etwa $\frac{d}{dt} \lambda_i(A(t)) = u_i(A(t))^* A'(0) u_i(A(t))$ und

$$\frac{d^2}{dt^2} \lambda_i(A(t)) = -2 \sum_{j \neq i} \frac{|u_i(A(t))^* A'(0) u_j(A(t))|^2}{\lambda_j(A(t)) - \lambda_i(A(t))}.$$

Mit tiefliegenden Resultaten zur sogenannten *Delokalisierung* kann gezeigt werden, dass im letzten Ausdruck der Zähler zumindest mit sehr großer Wahrscheinlichkeit von der Ordnung

$O(n^{-1+o(1)})$ ist. Der Nenner verlangt zu zeigen, dass mit hoher Wahrscheinlichkeit die Eigenwerte von M_N einfach sind. Dahinter steht eine so genannte *Lückeneigenschaft* der Eigenwerte sowie eine lokale Variante des Halbkreisgesetzes. Formeln für die höhere Ableitungen von $\lambda_i(A(t))$ sind bedeutend komplizierter. Die Details dazu wurden in [26] ausgearbeitet. Wir belassen es bei diesen Andeutungen.

Bei geeigneten Wahlen von Abbildungen G konnten TAO und VU zeigen, dass zum Beispiel die zu Beginn genannten Grenzwertsätze von GUE Matrizen auf WIGNER-Matrizen übertragen werden können. So hat man etwa das Gaußsche Verhalten eines einzelnen Eigenwertes innerhalb des Spektrums oder das Gaußsche Verhalten der Anzahl der Eigenwerte in einem Intervall für alle WIGNER-hermiteschen Matrizen mit $\mathbb{E}(X(p, q)^3) = 0$ und $\mathbb{E}(X(p, q)^4) = 3/4$, $p < q$. Die TRACY-WIDOM Verteilung für den Limes des kleinsten Eigenwertes überträgt sich sogar auf alle WIGNER-hermiteschen Matrizen. Details dazu betrachten wir hier nicht. In jüngster Zeit konnten andere Fluktuationen universell hergeleitet werden, etwa so genannte *Prinzipien moderater Abweichungen* für einzelne Eigenwerte, für die relative Anzahl der Eigenwerte einer Matrix in einem Intervall oder für die Determinante von WIGNER-hermiteschen Matrizen, siehe etwa [10], [11].

Wir stimmen POLLARD zu: LINDBERG's *argument still has something to offer*.

LITERATUR

- [1] G. W. Anderson, A. Guionnet, and O. Zeitouni, *An introduction to random matrices*, Cambridge Studies in Advanced Mathematics, vol. 118, Cambridge University press, 2010.
- [2] L. Arnold, *On the asymptotic distribution of the eigenvalues of random matrices*, J. Math. Anal. Appl. **20** (1967), 262–268.
- [3] Z. D. Bai, *Methodologies in spectral analysis of large-dimensional random matrices, a review*, Statist. Sinica **9** (1999), no. 3, 611–677, With comments by G. J. Rodgers and Jack W. Silverstein; and a rejoinder by the author.
- [4] P. Billingsley, *Probability and measure*, Wiley Series in Probability and Statistics, John Wiley & Sons Inc., Hoboken, NJ, 2012, Anniversary edition [of MR1324786], With a foreword by Steve Lalley and a brief biography of Billingsley by Steve Koppes.
- [5] E. Bolthausen, *Exact convergence rates in some martingale central limit theorems*, Ann. Probab. **10** (1982), no. 3, 672–688.
- [6] L. Breiman, *Probability*, Classics in Applied Mathematics, vol. 7, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1992, Corrected reprint of the 1968 original.
- [7] S. Chatterjee, *A simple invariance theorem*, Available at <http://arxiv.org/math.PR/0508213>, 2004.
- [8] ———, *A generalization of the Lindeberg principle*, Ann. Probab. **34** (2006), no. 6, 2061–2076.
- [9] H. Cramér, *Half a century with probability theory: some personal recollections*, Ann. Probability **4** (1976), no. 4, 509–546.
- [10] H. Döring and P. Eichelsbacher, *Moderate deviations for the determinant of Wigner matrices*, to appear in *Limit Theorems in Probability, Statistics and Number Theory*, Springer Proceedings in Mathematics & Statistics 42, dedicated to Friedrich Götze on the occasion of his sixtieth birthday, 2013.

- [11] ———, *Moderate deviations for the eigenvalue counting function of Wigner matrices*, arXiv:1104.0221, to appear in Lat. Am. J. Probab. Math. Stat., 2013.
- [12] W. Feller, *Über den zentralen Grenzwertsatz der Wahrscheinlichkeitsrechnung*, Math. Z. **40** (1936), no. 1, 521–559.
- [13] H. Fischer, *A history of the central limit theorem*, Sources and Studies in the History of Mathematics and Physical Sciences, Springer, New York, 2011, From classical to modern probability theory.
- [14] O. Friesen and M. Löwe, *A phase transition for the limiting spectral density of random matrices*, Electronic Journal of Prob. **18** (2013), no. 17, 1–17 .
- [15] ———, *The semicircle law for matrices with independent diagonals*, J. Theoret. Probab., 2011. doi:10.1007/s10959-011-0383-2, 2012.
- [16] B. V. Gnedenko and V. Yu. Korolev, *Random summation*, CRC Press, Boca Raton, FL, 1996, Limit theorems and applications.
- [17] F. Götze and A. N. Tikhomirov, *Limit theorems for spectra of random matrices with martingale structure*, Stein’s method and applications, Lect. Notes Ser. Inst. Math. Sci. Natl. Univ. Singap., vol. 5, Singapore Univ. Press, Singapore, 2005, pp. 181–193.
- [18] J. Gustavsson, *Gaussian fluctuations of eigenvalues in the GUE*, Ann. Inst. H. Poincaré Probab. Statist. **41** (2005), no. 2, 151–178.
- [19] V. Kalashnikov, *Geometric sums: bounds for rare events with applications*, Mathematics and its Applications, vol. 413, Kluwer Academic Publishers Group, Dordrecht, 1997, Risk analysis, reliability, queueing.
- [20] L. Le Cam, *The central limit theorem around 1935*, Statist. Sci. **1** (1986), no. 1, 78–96, With comments, and a rejoinder by the author.
- [21] J. W. Lindeberg, *Über das Exponentialgesetz in der Wahrscheinlichkeitsrechnung*, Annales academiae scientiarum Fennicae **16** (1920), 1–23.
- [22] ———, *Eine neue Herleitung des Exponentialgesetzes in der Wahrscheinlichkeitsrechnung*, Math. Z. **15** (1922), no. 1, 211–225.
- [23] ———, *Über das Gauss’sche Fehlergesetz*, Skandinavisk Aktuarietidskrift **5** (1922), 217–234.
- [24] L. A. Pastur, *The spectrum of random matrices*, Teoret. Mat. Fiz. **10** (1972), no. 1, 102–112.
- [25] G. Pólya, *Über den zentralen Grenzwertsatz der Wahrscheinlichkeitsrechnung und das Momentenproblem*, Math. Z. **8** (1920), no. 3-4, 171–181.
- [26] T. Tao and V. Vu, *Random matrices: universality of local eigenvalue statistics*, Acta Math. **206** (2011), 127–204.
- [27] ———, *Random matrices: the universality phenomenon for Wigner ensembles*, preprint, arXiv:1202.0068v1, 2012.
- [28] A. A. Toda, *Weak limit of the geometric sum of independent but not identically distributed random variables*, math/1111.1786v2, 2012.
- [29] E. P. Wigner, *Characteristic vectors of bordered matrices with infinite dimensions*, Ann. of Math. (2) **62** (1955), 548–564.
- [30] E. P. Wigner, *On the distribution of the roots of certain symmetric matrices*, Ann. of Math. (2) **67** (1958), 325–327.
- [31] J. H. Wilkinson, *The algebraic eigenvalue problem*, Clarendon Press, Oxford, 1965.
- [32] D. Williams, *Probability with martingales*, Cambridge Mathematical Textbooks, Cambridge University Press, Cambridge, 1991.