

Ein Berry-Esséen-Resultat
für
Character Ratios

Diplomarbeit
vorgelegt von
Christian Döbler

Erstgutachter: Prof. Peter Eichelsbacher
Zweitgutachter: Prof. Thomas Kriecherbauer



Fakultät für Mathematik
der Ruhr-Universität Bochum

13. Oktober 2008

Inhaltsverzeichnis

1	Die Steinsche Methode	7
1.1	Normalapproximation mit Hilfe der Steinschen Methode	9
1.1.1	Summen unabhängiger Zufallsvariablen	21
1.1.2	Der Zugang mittels austauschbarer Paare	23
1.2	Der Satz von Berry-Esséen	37
1.2.1	Bolthausens Beweis des Satzes von Berry-Esséen	38
1.3	Die abstrakte Steinsche Methode	52
1.3.1	Der algebraische Zugang von Stein	52
1.3.2	Verallgemeinerung der Stein-Charakterisierung und Stein-Gleichung	55
1.4	Anhang: Beweis von Lemma 1.2.4	58
1.5	Anhang: Beweis von Satz 1.2.5	77
1.6	Anhang: reelle Funktionen	82
2	Darstellungstheorie und der random walk auf endlichen Gruppen	89
2.1	Darstellungstheorie endlicher Gruppen	90
2.1.1	Definitionen, erste Eigenschaften und Beispiele	90
2.1.2	G -Invarianz, Irreduzibilität und Äquivalenz von Darstellungen	98
2.1.3	Charaktere von Darstellungen	109
2.1.4	Zerlegung der Gruppen-Algebra $\mathbb{C}[G]$	118
2.1.5	Induktion und Restriktion von Darstellungen	126
2.2	Darstellungstheorie der symmetrischen Gruppe \mathbb{S}_n	134
2.3	Der random walk auf Gruppen	157
3	Der zentrale Grenzwertsatz für character ratios	171
3.1	Das Plancherel-Maß und character ratios	173
3.2	Das Berry-Esséen-Theorem für character ratios	187

Einleitung

In der ersten Hälfte des zwanzigsten Jahrhunderts bestand die Wahrscheinlichkeitstheorie im Wesentlichen aus der Formulierung und dem Beweis der *klassischen* Grenzwertsätze, wie dem *zentralen Grenzwertsatz*, dem (*starken*) *Gesetz der großen Zahlen* oder auch dem *Gesetz vom iterierten Logarithmus* und zwar stets für *unabhängige* Zufallsvariablen. Die Aussagen dieser Sätze waren dabei stets so formuliert, dass „im Limes“ für $n \rightarrow \infty$ das entsprechende Grenzwertresultat gelten solle. Der zentrale Grenzwertsatz, zum Beispiel, spielt eine wesentliche Rolle in der mathematischen Statistik für die *asymptotische Verteilung* gewisser Teststatistiken. Da dem (statistischen) Anwender aber in der Praxis stets nur endlich viele, sagen wir $n \in \mathbb{N}$, Beobachtungen vorliegen, ist durch *diese* Formulierung des zentralen Grenzwertsatzes nicht klar, ob das Ersetzen der Verteilung der gegebenen Teststatistik durch die Normalverteilung gerechtfertigt ist. Für solche Anwendungen braucht man also ein Maß für die Konvergenzgeschwindigkeit (in $n \in \mathbb{N}$), eine sogenannte *Konvergenzrate*. Ein solches Resultat wurde, weiterhin unter der Annahme der Unabhängigkeit der beteiligten Summanden, erst in den vierziger Jahren des vorigen Jahrhunderts von *Berry* und *Esséen* gefunden. Die Beweistechnik bestand aus einer Kombination der Technik der *charakteristischen Funktionen* oder *Fourier-Transformierten* und einer sogenannten *Glättungs-Ungleichung*. Leider hängt jedoch der Erfolg der Methode der Fourier-Transformierten erheblich von der Unabhängigkeit der Summanden ab, da eine wichtige Faktorisierungsformel nur in dieser Situation gilt. Jedoch lässt sich die Annahme der Unabhängigkeit gerade in vielen statistischen Szenarien nicht aufrechterhalten, bzw. wird sogar durch dem Modell zugrunde liegende Relationen ausgeschlossen. Die in den siebziger Jahren von *Charles Stein* gefundene und nunmehr *Steinsche Methode* genannte Technik zur Herleitung von Verteilungskonvergenz ist grundsätzlich dazu geeignet sowohl Konvergenzraten zu liefern, als auch gewisse Abhängigkeitsstrukturen zu behandeln. Wir werden diese Technik für den Fall der Normalapproximation im ersten Kapitel dieser Arbeit ausführlich vorstellen und einen von *Erwin Bolthausen* gefundenen Beweis des klassischen Resultats von *Berry* und *Esséen* vorstellen und beweisen (vgl. [Bol84]). Hierbei gelingt uns eine Verbesserung der *universellen* Konstante im Grenzwertsatz von 25 auf 13 gegenüber der Ausarbeitung von *Mann* (vgl. [Man94]) von *Bolthausens* Arbeit.

Das zweite Kapitel führt umfassend in die Theorie der Darstellungen endlicher Gruppen und insbesondere in die Darstellungstheorie der symmetrischen Gruppen ein und motiviert dies durch die Ermittlung der Konvergenzgeschwindigkeit eines random walk auf der symmetrischen Gruppe gegen die Gleichverteilung. Dort wird auch der Begriff des *character ratio* eingeführt und seine Bedeutung für diese Konvergenzgeschwindigkeit verdeutlicht.

Im dritten Kapitel schließlich werden wir, der Arbeit *An Inductive Proof of the*

Berry-Esséen Theorem for Character Ratios von *Jason Fulman* folgend, einen zentralen Grenzwertsatz vom Berry-Esséen-Typ für character ratios ausgewertet in der Konjugationsklasse der Transpositionen formulieren und beweisen. Im Gegensatz zu Fulmans Arbeit werden wir auch eine explizite *universelle Berry-Esséen-Konstante* berechnen. Der Zufall kommt in diesem Szenario durch das sogenannte *Plancherel-Maß* zur symmetrischen Gruppe \mathbb{S}_n ins Spiel. Dieses Wahrscheinlichkeitsmaß auf der Menge aller *Partitionen* von n , entspricht der Gleichverteilung auf \mathbb{S}_n gemäß der von Robinson, Schensted und Knuth gefundenen Eins-zu-eins-Korrespondenz (*RSK-Korrespondenz*) zwischen der Gruppe \mathbb{S}_n und der Menge aller Paare von *Standard-Young-Tableaux* gleicher *Form*, welche eine Partition von $n \in \mathbb{N}$ ist.

An dieser Stelle möchte ich noch einigen Personen meinen Dank sagen: Zuallererst bedanke ich mich aufs Herzlichste bei meinen Eltern, Karin und Manfred Döbler, die mich während meiner gesamten Schul- und Studienzeit sowohl finanziell als auch moralisch unterstützt haben, indem sie mir immer wieder Mut zugesprochen haben. Weiterhin danke ich meiner Schwester, Simone Samp, und Nora Kories, dass sie mich immer wieder zur Weiterarbeit ermutigt haben. Ferner gilt mein Dank dem Betreuer dieser Arbeit, Prof. Dr. Peter Eichelsbacher, der in mir durch seine Vorlesungen das Interesse an der Stochastik weckte und mir bereits als Student das Mitwirken in seiner Arbeitsgruppe ermöglichte, sowie den (mir begegneten) Dozenten der Fakultät für Mathematik an der Ruhr-Universität Bochum für die sehr gute Ausbildung, die ich dort erfahren habe. Schließlich danke ich der BDO Deutsche Warentreuhand für die Gewährung eines Studienabschlussstipendiums.

Inhaltsverzeichnis

1 Die Steinsche Methode

Die Steinsche Methode, benannt nach dem Mathematiker Charles Stein, ist ein effizientes Mittel, um den „Abstand“ der Verteilung einer gegebenen Zufallsvariablen W und einer festen sogenannten target-Verteilung μ_0 zu messen. Stein entwickelte diese Methode ursprünglich zur Untersuchung der Normalapproximation einer Summe von abhängigen Zufallsvariablen im Jahre 1972. Seitdem wurde die Methode in verschiedener Hinsicht weiterentwickelt. Zum einen wurden neben der $\mathcal{N}(0, 1)$ -Verteilung auch andere target-Verteilungen, wie zum Beispiel die Poissonverteilung oder die mehrdimensionale Normalverteilung, untersucht und zum anderen wurden verschiedene Zugänge entwickelt, wie man mit Hilfe der Steinschen Methode ein Approximationsproblem angehen kann. Da es in dieser Arbeit ausschließlich um Normalapproximation geht, beschränken wir uns in diesem Kapitel im Wesentlichen auf diesen Sachverhalt, werden aber verschiedene Zugänge innerhalb der Steinschen Methode vorstellen und schließlich einen Induktionsansatz von Bolthausen nutzen, um den Satz von Berry-Esséen zu beweisen. Zum Schluss wollen wir noch auf die allgemeine Steinsche Methode eingehen. Das Kapitel ist genauer wie folgt gegliedert:

In Abschnitt 1.1 erläutern wir ausführlich die Steinsche Methode für die Normalverteilung und stellen Ansätze vor, die sich mit Summen unabhängiger Zufallsvariablen (vgl. Unterabschnitt 1.1.1) und mit der Situation, dass wir ein *austauschbares Paar* zu unserer gegebenen Verteilung finden können (vgl. Unterabschnitt 1.1.2), beschäftigen.

In Abschnitt 1.2 beweisen wir nach Erwin Bolthausen den klassischen Satz von Berry-Esséen mit Hilfe der Steinschen Methode, gepaart mit einem sehr geschickten Induktionsargument. Zudem werden wir die Berry-Esséen-Konstante gegenüber der Arbeit von Bolthausen (vgl. [Bol84]) oder auch der Ausarbeitung von Mann (vgl. [Man94]) von 25 auf 13 verbessern, indem wir die optimale obere Schranke für die Supremumsnorm der Lösungen der in diesem Beweis auftretenden *Stein-Gleichungen* finden.

Schließlich lösen wir uns in Abschnitt 1.3 von der Annahme, dass die Grenzverteilung die Standardnormalverteilung sein muss und diskutieren kurz die abstrakte Steinsche Methode.

In den Anhängen zu Kapitel 1 geben wir zum einen die etwas länglichen Beweise wichtiger Aussagen aus Abschnitt 1.2 und zum anderen stellen wir einige Grundtatsachen über reelle Funktionen bereit, die wir häufig stillschweigend verwenden

1 Die Steinsche Methode

werden und die an einige Fakten erinnern sollen und andere, die nicht unbedingt zum Kanon des Studiums gehören, kurz vorstellen sollen.

1.1 Normalapproximation mit Hilfe der Steinschen Methode

Im Folgenden sei Z stets eine $\mathcal{N}(0, 1)$ -verteilte Zufallsvariable, (Ω, \mathcal{A}, P) sei ein Wahrscheinlichkeitsraum und $W : \Omega \rightarrow \mathbb{R}$ sei eine reellwertige Zufallsvariable. Wie bereits in der Einleitung dieses Kapitels erwähnt wurde, ist die Steinsche Methode eine Technik, um den Abstand der Verteilung P^W zur standardnormalen Verteilung $\mathcal{N}(0; 1)$ zu messen. Um diesen Begriff zu präzisieren, sei \mathcal{H} eine Klasse von Borel-messbaren Funktionen $h : \mathbb{R} \rightarrow \mathbb{R}$ mit $E[|h(W)|] < \infty$ und $E[|h(Z)|] < \infty$. Dann liefert die Größe

$$d_{\mathcal{H}}(P^W, \mathcal{N}(0; 1)) := \sup_{h \in \mathcal{H}} |E[h(Z)] - E[h(W)]|$$

ein Abstandsmaß.

Beispiel 1.1.1. (i) Es sei $\mathcal{H} := \{I_B | B \in \mathcal{B}^1\}$. Dann ist

$$d_{\mathcal{H}}(P^W, \mathcal{N}(0; 1)) = d_{TV}(P^W, \mathcal{N}(0; 1)) = \sup_{B \in \mathcal{B}} |P(W \in B) - \mathcal{N}(0; 1)(B)|$$

der sogenannte Totalvariationsabstand von P^W und $\mathcal{N}(0; 1)$.

(ii) Sei $\mathcal{H} := \{I_{(-\infty; z]} | z \in \mathbb{R}\}$. Dann ist

$$d_{\mathcal{H}}(P^W, \mathcal{N}(0; 1)) = \sup_{z \in \mathbb{R}} |\Phi(z) - P(W \leq z)|$$

der Kolmogorov-Abstand von P^W und $\mathcal{N}(0; 1)$.

(iii) Sei $\mathcal{H} := \{h : \mathbb{R} \rightarrow \mathbb{R} | h \text{ ist Lipschitzstetig mit Konstante } 1\}$. Dann nennt man

$$d_{\mathcal{H}}(P^W, \mathcal{N}(0; 1))$$

den Wasserstein-Abstand von P^W und $\mathcal{N}(0; 1)$.

Leider ist die Größe $\sup_{h \in \mathcal{H}} |E[h(Z)] - E[h(W)]|$ in dieser Form häufig nicht gut abzuschätzen. Dies liegt vor allem daran, dass der Ausdruck $|E[h(Z)] - E[h(W)]|$ zwei Zufallsvariablen, nämlich Z und W , also zwei Verteilungen enthält. Die Steinsche Methode basiert auf folgendem Ausweg aus diesem Problem:

Dazu setzen wir von nun an voraus, dass alle $h \in \mathcal{H}$ beschränkt und stückweise stetig sind (vgl. Definition 1.6.1 (i)). Dann betrachte man zu jedem $h \in \mathcal{H}$ die Funktion f_h mit

$$f_h(x) := \frac{1}{\varphi(x)} \int_{-\infty}^x (h(y) - N(h)) \varphi(y) dy ,$$

1 Die Steinsche Methode

wobei $N(h) := E[h(Z)] = \int_{\mathbb{R}} h(y)\varphi(y)dy = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} h(y)e^{-\frac{y^2}{2}} dy$ sei. Dabei sei φ wie üblich die (einzige stetige) Dichte der Standardnormalverteilung, das heißt für $x \in \mathbb{R}$ gilt $\varphi(x) = \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}}$.

Es ist f_h „Lösung“ der nunmehr *Stein-Gleichung* genannten Differentialgleichung

$$f'(x) = xf(x) + h(x) - N(h), \quad (1.1)$$

wobei „Lösung“ so zu verstehen ist, dass f_h genau in den Stetigkeitspunkten von h stetig differenzierbar ist und dort die Differentialgleichung (1.1) löst. Die Lösungsmenge der zu (1.1) gehörigen homogenen Differentialgleichung ist gegeben durch die skalaren Vielfachen der Funktion

$$\mathbb{R} \ni x \mapsto e^{\frac{x^2}{2}} \in \mathbb{R},$$

so dass man sämtliche Lösungen von (1.1) durch Addition einer solchen Vielfachen zu f_h erhält. Es wird sich herausstellen, dass die Funktion f_h unter den Lösungen von (1.1) als die einzige beschränkte Lösung ausgezeichnet ist, wobei die Eindeutigkeit aufgrund der Unbeschränktheit von $e^{\frac{x^2}{2}}$ klar ist. Ist speziell für $z \in \mathbb{R}$ $h_z := I_{(-\infty; z]}$, so schreiben wir einfach f_z statt f_{h_z} . Wir werden weiter unten sehen, dass mit den Funktionen h auch die Lösungen f_h (und deren Ableitungen) beschränkt sind, so dass Erwartungswerte wie etwa $E[f_h(W)]$ oder $E[Wf_h(W)]$ existieren.

Ersetzt man nun das Argument x durch die Zufallsvariable W und bildet man den Erwartungswert, so folgt aus (1.1) die Identität

$$|E[h(W)] - E[h(Z)]| = |E[f'_h(W)] - E[Wf_h(W)]| \quad (1.2)$$

Der Erfolg der Steinschen Methode im Rahmen der Normalapproximation beruht hauptsächlich darauf, dass die rechte Seite in (1.2) unter der Voraussetzung, dass eine Normalapproximation für W Sinn macht, häufig viel leichter (im Supremum) abgeschätzt werden kann, als die linke. Aber worin ist dies begründet? Zum einen gilt für die exakt standardnormalverteilte Zufallsvariable Z aufgrund der Definition von $N(h)$:

$$E[f'_h(Z)] - E[Zf_h(Z)] = 0,$$

so dass für $W \stackrel{\mathcal{D}}{\approx} \mathcal{N}(0; 1)$ dieser Ausdruck ungefähr gleich 0 sein sollte. Aber dies würde für jede Lösung einer (linearen) Differentialgleichung

$$f^{(n)}(x) - \sum_{j=0}^{n-1} \alpha_j(x)f^{(j)}(x) = h(x) - N(h) \quad (1.3)$$

1.1 Normalapproximation mit Hilfe der Steinschen Methode

gelten!

Eine etwas befriedigendere Antwort lautet wie folgt :

(1) Für $Z \sim \mathcal{N}(0; 1)$ ist die Klasse aller Funktionen f mit

$$E[f'(Z)] - E[Zf(Z)] = 0$$

viel „größer“, als die Menge der Lösungen von (1.1) und

(2) Die Funktionen f_h und ihre Ableitungen können effizient im Supremum abgeschätzt werden.

Zur Präzisierung von (1) stellen wir nun die folgende von Stein gefundene Charakterisierung der Standardnormalverteilung vor (vgl. [Ste86]).

Satz 1.1.2 (Lemma von Stein). *Für eine reellwertige Zufallsvariable W sind die folgenden Bedingungen äquivalent:*

(i) $W \sim \mathcal{N}(0; 1)$

(ii) Für jede stückweise stetig differenzierbare Funktion $f : \mathbb{R} \rightarrow \mathbb{R}$ (vgl. Definition 1.6.1 (ii)) mit

$$E[|f'(Z)|] = \int_{\mathbb{R}} |f'(x)|\varphi(x)dx < +\infty$$

existieren die Größen $E[f'(W)]$ und $E[Wf(W)]$ und stimmen überein.

Damit der Beweis überschaubarer bleibt, stellen wir zunächst ein Lemma voran, welches garantiert, dass die Stein-Lösung $f_z := f_{h_z}$ zur Funktion $h_z = I_{(-\infty; z]}$ die Voraussetzung aus Satz 1.1.2 (ii) erfüllt:

Lemma 1.1.3. *Es sei die Funktion $S : \mathbb{R} \rightarrow \mathbb{R}$ definiert durch*

$$S(x) := \begin{cases} \frac{1}{|x|} & \text{für } x \neq 0 \\ \sqrt{\frac{\pi}{2}} & \text{für } x = 0 \end{cases}$$

Dann gilt für alle $x, z \in \mathbb{R}$:

$$|f_z(x)| \leq S(x) .$$

Inbesondere gilt $\int_{\mathbb{R}} |f'_z(y)|\varphi(y)dy < +\infty$.

1 Die Steinsche Methode

Beweis: Wir zeigen zunächst die erste Behauptung. Sei $z \in \mathbb{R}$. Nach Definition von $N(h_z)$ gilt für alle $x \in \mathbb{R}$:

$$f_z(x) = \frac{1}{\varphi(x)} \int_{-\infty}^x (h_z(y) - N(h_z))\varphi(y)dy = -\frac{1}{\varphi(x)} \int_x^{\infty} (h_z(y) - N(h_z))\varphi(y)dy$$

Wegen $0 \leq h_z(x) \leq 1$ für alle $x \in \mathbb{R}$ gilt $0 \leq N(h_z) \leq 1$ und somit auch $\|h_z - N(h_z)\|_{\infty} \leq 1$.

Für $x > 0$ folgt daraus:

$$\begin{aligned} |f_z(x)| &= \frac{1}{\varphi(x)} \left| \int_x^{\infty} (h_z(y) - N(h_z))\varphi(y)dy \right| \\ &\leq \frac{1}{\varphi(x)} \int_x^{\infty} \|h_z - N(h_z)\|_{\infty} \varphi(y)dy \leq \frac{1}{\varphi(x)} \int_x^{\infty} \varphi(y)dy \\ &\leq \frac{1}{\varphi(x)} \int_x^{\infty} \frac{y}{x} \varphi(y)dy = \frac{1}{\varphi(x)x} \int_x^{\infty} y\varphi(y)dy \\ &= \frac{1}{\varphi(x)x} (-\varphi(y)|_x^{\infty}) = \frac{1}{x} = \frac{1}{|x|}. \end{aligned}$$

Für $x < 0$ folgt analog

$$\begin{aligned} |f_z(x)| &= \frac{1}{\varphi(x)} \left| \int_{-\infty}^x (h_z(y) - N(h_z))\varphi(y)dy \right| \\ &\leq \frac{1}{\varphi(x)} \int_{-\infty}^x \|h_z - N(h_z)\|_{\infty} \varphi(y)dy \leq \frac{1}{\varphi(x)} \int_{-\infty}^x \varphi(y)dy \\ &\leq \frac{1}{\varphi(x)} \int_{-\infty}^x \frac{|y|}{|x|} \varphi(y)dy = \frac{1}{\varphi(x)} \int_{-\infty}^x \frac{-y}{-x} \varphi(y)dy \\ &= \frac{1}{-x\varphi(x)} \int_{-\infty}^x (-y)\varphi(y)dy = \frac{1}{|x|\varphi(x)} (\varphi(y)|_{-\infty}^x) = \frac{1}{|x|}. \end{aligned}$$

Schließlich gilt für $x = 0$

1.1 Normalapproximation mit Hilfe der Steinschen Methode

$$|f_z(x)| = \frac{1}{\varphi(0)} \left| \int_0^\infty (h_z(y) - N(h_z)) \varphi(y) dy \right| \leq \sqrt{2\pi} \int_0^\infty \varphi(y) dy = \frac{\sqrt{2\pi}}{2} = \sqrt{\frac{\pi}{2}}.$$

Damit ist die erste Behauptung gezeigt. Für die zweite beachten wir, dass nach dem gerade Bewiesenen und, da f_z die Stein-Gleichung löst, für $x \neq 0$ gilt

$$|f'_z(x)| = |x f_z(x) + h_z(y) - N(h_z)| \leq \frac{|x|}{|x|} + \|h_z - N(h_z)\|_\infty \leq 2,$$

so dass

$$\int_{\mathbb{R}} |f'_z(y)| \varphi(y) dy = \int_{\mathbb{R} \setminus \{0\}} |f'_z(y)| \varphi(y) dy \leq 2 < +\infty.$$

□

Beweis von Satz 1.1.2

(ii) \Rightarrow (i): Wir zeigen, dass $P(W \leq z) = P(Z \leq z) = \Phi(z)$ für alle $z \in \mathbb{R}$ gilt. Da eine Verteilung auf $(\mathbb{R}, \mathcal{B})$ durch ihre Verteilungsfunktion charakterisiert ist, folgt hieraus die Aussage von (i). Sei also $z \in \mathbb{R}$. Betrachte die Funktion $h_z = I_{(-\infty; z]}$ und die dazugehörige Stein-Lösung f_z . Nach Lemma 1.1.3 gilt

$$E[|f'_z(Z)|] = \int_{\mathbb{R}} |f'_z(y)| \varphi(y) dy < +\infty,$$

so dass wir auf die stückweise stetig differenzierbare Funktion f_z die Aussage von (ii) anwenden können. Es folgt:

$$0 = E[f'_z(W)] - E[W f_z(W)] = E[h_z(W)] - N(h_z) = P(W \leq z) - \Phi(z).$$

(i) \Rightarrow (ii): Es sei $f : \mathbb{R} \rightarrow \mathbb{R}$ stückweise stetig differenzierbar (vgl. Definition 1.6.1 (ii)) mit

$$E[|f'(Z)|] = E[|f'(W)|] = \int_{\mathbb{R}} |f'(x)| \varphi(x) dx < +\infty.$$

1 Die Steinsche Methode

Dann gilt:

$$\begin{aligned}
 E[f'(W)] &= \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} f'(w) e^{-\frac{w^2}{2}} dw \\
 &= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^0 f'(w) \left(\int_{-\infty}^w (-z) e^{-\frac{z^2}{2}} dz \right) dw \\
 &\quad + \frac{1}{\sqrt{2\pi}} \int_0^{\infty} f'(w) \left(\int_w^{\infty} z e^{-\frac{z^2}{2}} dz \right) dw \tag{1.4}
 \end{aligned}$$

Behauptung: Die Abbildungen $F^+, F^- : \mathbb{R}^2 \rightarrow \mathbb{R}$ mit

$$F^+(w, z) := I_{[0;+\infty)}(w) I_{[w;+\infty)}(z) f'(w) z e^{-\frac{z^2}{2}}$$

und

$$F^-(w, z) := I_{(-\infty;0)}(w) I_{(-\infty;w]}(z) f'(w) (-z) e^{-\frac{z^2}{2}}$$

sind λ_2 -integrierbar.

denn: Nach den Sätzen 1.6.3 und 1.6.7 sind die Abbildungen Lebesgue-messbar. Weiter gilt nach dem Satz von Fubini für den nicht-negativen Integranden $|F^+|$:

$$\begin{aligned}
 &\int_{\mathbb{R}^2} |F^+(w, z)| d\lambda_2(w, z) \\
 &= \int_{\mathbb{R}} \left(\int_{\mathbb{R}} I_{[0;+\infty)}(w) I_{[w;+\infty)}(z) |f'(w)| |z| e^{-\frac{z^2}{2}} dz \right) dw \\
 &= \int_0^{\infty} |f'(w)| \left(\int_w^{\infty} z e^{-\frac{z^2}{2}} dz \right) dw = \int_0^{\infty} |f'(w)| \left(-e^{-\frac{z^2}{2}} \Big|_w^{+\infty} \right) dw \\
 &= \int_0^{\infty} |f'(w)| e^{-\frac{w^2}{2}} dw \leq \int_{\mathbb{R}} |f'(w)| e^{-\frac{w^2}{2}} dw \\
 &= \sqrt{2\pi} E[|f'_z(Z)|] < +\infty
 \end{aligned}$$

Analog zeigt man, dass $\int_{\mathbb{R}^2} |F^-(w, z)| d\lambda_2(w, z) < +\infty$ ist.

Somit können wir in (1.4) nach dem Satz von Fubini die Integrationsreihenfolge vertauschen und mit dem Hauptsatz der Differential- und Integralrechnung weiterrechnen:

1.1 Normalapproximation mit Hilfe der Steinschen Methode

$$\begin{aligned}
 & E[f'(W)] \\
 = & \frac{1}{\sqrt{2\pi}} \int_{-\infty}^0 \left(\int_z^0 f'(w) dw \right) (-z) e^{-\frac{z^2}{2}} dz + \frac{1}{\sqrt{2\pi}} \int_0^{\infty} \left(\int_0^z f'(w) dw \right) z e^{-\frac{z^2}{2}} dz \\
 = & \frac{1}{\sqrt{2\pi}} \int_{-\infty}^0 (f(0) - f(z)) (-z) e^{-\frac{z^2}{2}} dz + \frac{1}{\sqrt{2\pi}} \int_0^{\infty} (f(z) - f(0)) z e^{-\frac{z^2}{2}} dz \\
 = & \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} z f(z) e^{-\frac{z^2}{2}} dz - \frac{f(0)}{\sqrt{2\pi}} \int_{\mathbb{R}} z e^{-\frac{z^2}{2}} dz \\
 = & E[W f(W)] - f(0) E[W] = E[W f(W)]
 \end{aligned}$$

□

Die durch Satz 1.1.2 gegebene Charakterisierung der Standardnormalverteilung ist eine erste Rechtfertigung für die Wahl der Stein Gleichung (1.1) aus der großen Klasse von Differentialgleichungen des Typs (1.3).

Eine weitere Rechtfertigung ergibt sich aus dem folgenden Lemma, welches im Wesentlichen schon in [Ste86] bewiesen wird. Wir verzichten an dieser Stelle auf einen Beweis dieser Aussage. Ein vollständiger Beweis lässt sich z.B. in [BC05] finden.

Lemma 1.1.4. *Es sei $h : \mathbb{R} \rightarrow \mathbb{R}$ eine beschränkte, absolut stetige Funktion und es sei f_h die durch*

$$f_h(x) = \frac{1}{\varphi(x)} \int_{-\infty}^x (h(y) - N(h)) \varphi(y) dy$$

gegebene Lösung der Stein-Gleichung (1.1). Dann gilt:

- (a) $\|f_h\|_{\infty} \leq \min\left(\sqrt{\frac{\pi}{2}} \|h - N(h)\|_{\infty}, 2 \|h'\|_{\infty}\right)$
- (b) $\|f'_h\|_{\infty} \leq \min(2 \|h - N(h)\|_{\infty}, 4 \|h'\|_{\infty})$
- (c) $\|f''_h\|_{\infty} \leq 2 \|h'\|_{\infty}$

Schranken dieser Form können natürlich universell für die Normalapproximation verwendet werden, das heißt wann immer eine gegebene Zufallsvariable W auf Normalapproximation untersucht werden soll. Hat man solch eine Zufallsvariable W gegeben, so verwendet man dann ihre spezielle Struktur, um die rechte Seite

1 Die Steinsche Methode

von (1.2) (möglichst gleichmäßig in $h \in \mathcal{H}$) nach oben abzuschätzen. Dies soll im Folgenden für verschiedene Zusammensetzungen der Zufallsvariablen W vorgeführt werden. Die Darstellung folgt dabei in großen Teilen derjenigen aus [BC05]. Dabei werden wir sehen, dass wir für den Fall, dass W eine Summe von unabhängigen und identisch verteilten Zufallsvariablen ist, explizite Konvergenzraten erhalten, wobei unsere Rate für den Kolmogorov-Abstand in diesem Abschnitt stets von der Ordnung $n^{-\frac{1}{4}}$ und somit nicht optimal ist. Jedoch werden wir in Abschnitt 1.2 die optimale Rate mit Hilfe der Steinschen Methode erhalten. Wir beginnen mit einem Resultat, welches eine obere Schranke für den Kolmogorov-Abstand liefert, wenn wir eine obere Schranke für den Wasserstein-Abstand haben.

Satz 1.1.5. *Sei W eine reellwertige Zufallsvariable. Es gebe ein $\delta > 0$, so dass für alle beschränkten und Lipschitz-stetigen Funktionen $h : \mathbb{R} \rightarrow \mathbb{R}$ mit minimaler Lipschitz-Konstante $L_h = \|h'\|_\infty$ (vgl. Bemerkung 1.6.2 (ii)) gilt:*

$$|E[h(W)] - N(h)| \leq \delta \|h'\|_\infty.$$

Dann gilt

$$\sup_{z \in \mathbb{R}} |P(W \leq z) - \Phi(z)| \leq \sqrt{\delta}.$$

Beweis: Für $\lambda > 0$ und $z \in \mathbb{R}$ sei die Funktion $h_{z,\lambda} : \mathbb{R} \rightarrow \mathbb{R}$ gegeben durch

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

Dann ist $h_{z,\lambda}$ Lipschitz-stetig mit minimaler Lipschitz-Konstante $\|h'_{z,\lambda}\|_\infty = \frac{1}{\lambda}$ (vgl. Beispiel 1.6.6). Man hat nun die punktweise Ungleichung

$$h_{z-\lambda,\lambda} \leq h_z \leq h_{z,\lambda} \leq h_{z+\lambda}$$

und somit gilt für jedes $\lambda > 0$

$$\begin{aligned} P(W \leq z) - \Phi(z) &= E[h_z(W)] - E[h_z(Z)] \\ &\leq E[h_{z,\lambda}(W)] - E[h_{z,\lambda}(Z)] + E[h_{z,\lambda}(Z)] - E[h_z(Z)] \\ &\leq \frac{\delta}{\lambda} + \int_z^{z+\lambda} \left(1 - \frac{y-z}{\lambda}\right) \varphi(y) dy \end{aligned}$$

Es ist

1.1 Normalapproximation mit Hilfe der Steinschen Methode

$$\begin{aligned}
 \int_z^{z+\lambda} \left(1 - \frac{y-z}{\lambda}\right) \varphi(y) dy &\leq \frac{1}{\sqrt{2\pi}} \int_z^{z+\lambda} \left(1 - \frac{y-z}{\lambda}\right) dy \\
 &= \frac{1}{\sqrt{2\pi}} \left(\lambda - \frac{1}{\lambda} \int_0^\lambda x dx \right) \\
 &= \frac{1}{\sqrt{2\pi}} \left(\lambda - \frac{\lambda}{2} \right) = \frac{\lambda}{2\sqrt{2\pi}},
 \end{aligned}$$

also

$$P(W \leq z) - \Phi(z) \leq \frac{\delta}{\lambda} + \frac{\lambda}{2\sqrt{2\pi}}.$$

Ebenso gilt

$$\begin{aligned}
 \Phi(z) - P(W \leq z) &\leq \Phi(z) - E[h_{z-\lambda,\lambda}(Z)] + E[h_{z-\lambda,\lambda}(Z)] - E[h_{z-\lambda,\lambda}(W)] \\
 &\leq \Phi(z) - E[h_{z-\lambda,\lambda}(Z)] + \frac{\delta}{\lambda} \\
 &\leq \frac{\delta}{\lambda} + \int_{z-\lambda}^z \left(1 - \left(1 - \frac{y-(z-\lambda)}{\lambda}\right)\right) \varphi(y) dy.
 \end{aligned}$$

Es ist

$$\begin{aligned}
 \int_{z-\lambda}^z \frac{y-(z-\lambda)}{\lambda} \varphi(y) dy &\leq \frac{1}{\sqrt{2\pi}} \int_{z-\lambda}^z \frac{y-(z-\lambda)}{\lambda} dy \\
 &= \frac{1}{\sqrt{2\pi}\lambda} \int_0^\lambda x dx = \frac{\lambda}{2\sqrt{2\pi}},
 \end{aligned}$$

also auch

$$\Phi(z) - P(W \leq z) \leq \frac{\delta}{\lambda} + \frac{\lambda}{2\sqrt{2\pi}}$$

für alle $z \in \mathbb{R}$ und $\lambda > 0$ und somit auch

$$\sup_{z \in \mathbb{R}} |\Phi(z) - P(W \leq z)| \leq \frac{\delta}{\lambda} + \frac{\lambda}{2\sqrt{2\pi}}$$

1 Die Steinsche Methode

für alle $\lambda > 0$. Sei für $\lambda > 0$

$$g(\lambda) := \frac{\delta}{\lambda} + \frac{\lambda}{2\sqrt{2\pi}}.$$

Wir wollen g in $\lambda > 0$ minimieren! Es gilt:

$$\begin{aligned} g'(\lambda) &= -\frac{\delta}{\lambda^2} + \frac{1}{2\sqrt{2\pi}} = 0 \\ \Leftrightarrow \lambda^2 &= \delta 2\sqrt{2\pi} \\ \Leftrightarrow \lambda &= \sqrt{2\delta}(2\pi)^{\frac{1}{4}} \end{aligned}$$

und

$$g''(\lambda) = \frac{2\delta}{\lambda^3} > 0 \text{ für alle } \lambda > 0.$$

Somit hat g in $\lambda = \sqrt{2\delta}(2\pi)^{\frac{1}{4}}$ ein lokales Minimum und wegen

$$\lim_{\lambda \rightarrow 0} g(\lambda) = +\infty = \lim_{\lambda \rightarrow +\infty} g(\lambda)$$

ist $\lambda = \sqrt{2\delta}(2\pi)^{\frac{1}{4}}$ globale Minimalstelle von g . Eingesetzt liefert dies, dass

$$\begin{aligned} \sup_{z \in \mathbb{R}} |\Phi(z) - P(W \leq z)| &\leq \frac{\delta}{\sqrt{2\delta}(2\pi)^{\frac{1}{4}}} + \frac{\sqrt{2\delta}(2\pi)^{\frac{1}{4}}}{2\sqrt{2\pi}} \\ &= \frac{\sqrt{\delta}}{\sqrt{2}(2\pi)^{\frac{1}{4}}} + \frac{\sqrt{\delta}}{\sqrt{2}(2\pi)^{\frac{1}{4}}} \\ &= \frac{\sqrt{2}}{(2\pi)^{\frac{1}{4}}} \sqrt{\delta} \approx 0,893\sqrt{\delta} \\ &\leq \sqrt{\delta} \end{aligned}$$

□

Bevor wir die Normalapproximation in konkreten Fällen betrachten, beweisen wir noch einige Hilfsaussagen, die wir für die Betrachtung der Situation von unabhängigen Summanden verwenden werden. Es soll nicht verschwiegen werden, dass man diese Aussagen, gepaart mit einer sogenannten *Konzentrationsungleichung*, auch für einen anderen Beweis des Satzes von Berry-Esséen, als wir ihn vorstellen werden, verwenden kann. Dazu sei auf das Buch [BC05] verwiesen.

1.1 Normalapproximation mit Hilfe der Steinschen Methode

Lemma 1.1.6. Seien $n \in \mathbb{N}$ und Y_1, \dots, Y_n seien reellwertige Zufallsvariablen, definiert auf einem gemeinsamen Wahrscheinlichkeitsraum (Ω, \mathcal{A}, P) . Es gelte $E[Y_i] = 0$ für $i = 1, \dots, n$, sowie $\sum_{i=1}^n E[Y_i^2] = \sum_{i=1}^n \text{Var}(Y_i) = 1$. Weiter definieren wir $W := \sum_{i=1}^n Y_i$ sowie für $i = 1, \dots, n$ $W_i := W - Y_i = \sum_{j \neq i} Y_j$. Schließlich seien für $i = 1, \dots, n$ die Funktionen $K_i : \mathbb{R} \rightarrow \mathbb{R}$ definiert durch

$$K_i(t) := E[Y_i (I_{\{0 \leq t \leq Y_i\}} - I_{\{Y_i \leq t < 0\}})].$$

Dann gelten die folgenden Aussagen für $i = 1, \dots, n$:

- (i) $K_i(t) \geq 0$ für alle $t \in \mathbb{R}$
- (ii) $\int_{\mathbb{R}} K_i(t) dt = E[Y_i^2]$
- (iii) $\int_{\mathbb{R}} |t| K_i(t) dt = \frac{1}{2} E[|Y_i|^3]$

Beweis: Sei $i \in \{1, \dots, n\}$ beliebig. Für jedes $t \in \mathbb{R}$ ist die Zufallsvariable

$$Z_i(t) := Y_i (I_{\{0 \leq t \leq Y_i\}} - I_{\{Y_i \leq t < 0\}})$$

nicht-negativ und somit auch $K_i(t) = E[Z_i(t)]$ (wohldefiniert und) nicht-negativ. Somit gilt (i).

zu (ii): Da $Z_i(t) \geq 0$ ist, folgt aus dem Satz von Fubini

$$\int_{\mathbb{R}} K_i(t) dt = E[Y_i \int_{\mathbb{R}} (I_{\{0 \leq t \leq Y_i\}} - I_{\{Y_i \leq t < 0\}}) dt] = E[Y_i^2].$$

zu (iii): Die Abbildung $(t, \omega) \mapsto |t| Z_i(t)(\omega)$ ist messbar und nicht-negativ. Somit folgt aus dem Satz von Fubini

$$\begin{aligned} \int_{\mathbb{R}} |t| K_i(t) dt &= \int_{\mathbb{R}} E[|t| Z_i(t)] dt = E\left[\int_{\mathbb{R}} |t| Z_i(t) dt\right] \\ &= E\left[Y_i \int_{\mathbb{R}} |t| (I_{\{0 \leq t \leq Y_i\}} - I_{\{Y_i \leq t < 0\}}) dt\right] \\ &= E\left[Y_i I_{\{Y_i \geq 0\}} \int_0^{Y_i} t dt - Y_i I_{\{Y_i < 0\}} \int_{Y_i}^0 (-t) dt\right] \\ &= E\left[I_{\{Y_i \geq 0\}} \frac{1}{2} Y_i^3 + I_{\{Y_i < 0\}} \left(-\frac{1}{2} Y_i^3\right)\right] \\ &= \frac{1}{2} E[|Y_i|^3]. \end{aligned}$$

□

Sind in der Situation von Lemma 1.1.6 die Zufallsvariablen Y_1, \dots, Y_n stochastisch unabhängig, dann können wir folgende Identitäten herleiten:

Satz 1.1.7. *Seien zusätzlich zu den Voraussetzungen aus Lemma 1.1.6 die Zufallsvariablen Y_1, \dots, Y_n stochastisch unabhängig. Weiter sei $h : \mathbb{R} \rightarrow \mathbb{R}$ eine beschränkte, absolut stetige Funktion und f_h sei die dazugehörige Lösung der Stein-Gleichung. Dann gelten die folgenden Identitäten:*

$$(a) \quad E[W f_h(W)] = \sum_{i=1}^n \int_{\mathbb{R}} E[f'_h(W_i + t)] K_i(t) dt$$

$$(b) \quad E[h(W)] - N(h) = \sum_{i=1}^n \int_{\mathbb{R}} E[f'_h(W) - f'_h(W_i + t)] K_i(t) dt$$

Beweis: Sei $h : \mathbb{R} \rightarrow \mathbb{R}$ beschränkt und absolut stetig. Dann gilt wegen der Unabhängigkeit von Y_i und W_i und wegen $E[Y_i] = 0$:

$$\begin{aligned} E[W f_h(W)] &= \sum_{i=1}^n E[Y_i f_h(W)] = \sum_{i=1}^n E[Y_i (f_h(W) - f_h(W_i))] \\ &= \sum_{i=1}^n E[Y_i \int_{W_i}^W f'_h(t) dt] = \sum_{i=1}^n E[Y_i \int_0^{Y_i} f'_h(W_i + t) dt] \\ &= \sum_{i=1}^n E[\int_{\mathbb{R}} f'_h(W_i + t) Y_i (I_{\{0 \leq t \leq Y_i\}} - I_{\{Y_i \leq t < 0\}}) dt] \\ &= \sum_{i=1}^n E[\int_{\mathbb{R}} f'_h(W_i + t) Z_i(t) dt] \end{aligned}$$

Es ist für $i = 1, \dots, n$ nach Lemma 1.1.6

$$\begin{aligned} &E[\int_{\mathbb{R}} |f'_h(W_i + t) Z_i(t)| dt] \\ &\leq \|f'_h\|_{\infty} E[\int_{\mathbb{R}} Z_i(t) dt] = \|f'_h\|_{\infty} \int_{\mathbb{R}} E[Z_i(t)] dt \\ &= \|f'_h\|_{\infty} \int_{\mathbb{R}} K_i(t) dt = \|f'_h\|_{\infty} E[Y_i^2] < +\infty \end{aligned}$$

1.1 Normalapproximation mit Hilfe der Steinschen Methode

und somit folgt aus dem Satz von Fubini

$$\begin{aligned}
 E[W f_h(W)] &= \sum_{i=1}^n \int_{\mathbb{R}} E[f'_h(W_i + t) Z_i(t)] dt \\
 &\stackrel{W_i, Z_i(t) \text{ unabh.}}{=} \sum_{i=1}^n \int_{\mathbb{R}} E[f'_h(W_i + t)] E[Z_i(t)] dt \\
 &= \sum_{i=1}^n \int_{\mathbb{R}} E[f'_h(W_i + t)] K_i(t) dt.
 \end{aligned}$$

Wegen $\sum_{i=1}^n \int_{\mathbb{R}} K_i(t) dt = \sum_{i=1}^n E[Y_i^2] = 1$, folgt weiter

$$E[f'_h(W)] = \sum_{i=1}^n \int_{\mathbb{R}} E[f'_h(W)] K_i(t) dt.$$

Somit erhalten wir insgesamt, dass

$$E[h(W)] - N(h) = \sum_{i=1}^n \int_{\mathbb{R}} E[f'_h(W) - f'_h(W_i + t)] K_i(t) dt \text{ gilt.}$$

□

Nun wollen wir in einigen Fällen obere Schranken für beschränkte, Lipschitz-stetige Funktionen h ausrechnen. Nach Satz 1.1.5 liefern diese unmittelbar obere Schranken für den Kolmogorov-Abstand.

1.1.1 Summen unabhängiger Zufallsvariablen

Satz 1.1.8. *Seien Y_1, \dots, Y_n unabhängige Zufallsvariablen mit $E[Y_i] = 0$ für $i = 1, \dots, n$ und mit $\sum_{i=1}^n E[Y_i^2] = \sum_{i=1}^n \text{Var}(Y_i) = 1$. Es sei $W := \sum_{i=1}^n Y_i$. Dann kann man Satz 1.1.5 anwenden mit*

$$\delta = 3 \sum_{i=1}^n E[|Y_i|^3].$$

Beweis: Es sei $h : \mathbb{R} \rightarrow \mathbb{R}$ beschränkt und Lipschitz-stetig mit minimaler Lipschitz-Konstanten $\|h'\|_{\infty}$. Dann gilt nach Satz 1.1.7 mit den dort gültigen Bezeichnungen

1 Die Steinsche Methode

$$\begin{aligned} |E[h(W)] - N(h)| &= \left| \sum_{i=1}^n \int_{\mathbb{R}} E[f'_h(W) - f'_h(W_i + t)] K_i(t) dt \right| \\ &\leq \sum_{i=1}^n \int_{\mathbb{R}} E[|f'_h(W) - f'_h(W_i + t)|] K_i(t) dt \end{aligned}$$

Mit Hilfe von Lemma 1.1.4 (c) kann man nun wegen $\|f''_h\|_{\infty} \leq 2\|h'\|_{\infty}$ für jedes i abschätzen:

$$|f'_h(W) - f'_h(W_i + t)| \leq \|f''_h\|_{\infty} |Y_i - t| \leq 2\|h'\|_{\infty} (|Y_i| + |t|)$$

Daraus folgt mit Hilfe von Lemma 1.1.6

$$\begin{aligned} |E[h(W)] - N(h)| &\leq 2\|h'\|_{\infty} \sum_{i=1}^n \int_{\mathbb{R}} E[|Y_i| + |t|] K_i(t) dt \\ &= 2\|h'\|_{\infty} \sum_{i=1}^n \left(E[|Y_i|] \int_{\mathbb{R}} K_i(t) dt + \int_{\mathbb{R}} |t| K_i(t) dt \right) \\ &= 2\|h'\|_{\infty} \sum_{i=1}^n \left(E[|Y_i|] E[Y_i^2] + \frac{1}{2} E[|Y_i|^3] \right) \end{aligned}$$

Nach der Jensen-Ungleichung gilt:

- $E[|Y_i|] \leq (E[|Y_i|^3])^{\frac{1}{3}}$ sowie
- $E[Y_i^2] \leq (E[|Y_i|^3])^{\frac{2}{3}}$,

so dass wir insgesamt

$$|E[h(W)] - N(h)| \leq 3\|h'\|_{\infty} \sum_{i=1}^n E[|Y_i|^3] \text{ erhalten.}$$

□

Bemerkung 1.1.9. Satz 1.1.8 ist natürlich nur dann nützlich, wenn $E[|Y_i|^3] < +\infty$ für alle $i \in \{1, \dots, n\}$ gilt.

Im unabhängig und identisch verteilten Fall erhalten wir sofort das folgende Korollar aus den Sätzen 1.1.8 und 1.1.5:

1.1 Normalapproximation mit Hilfe der Steinschen Methode

Korollar 1.1.10. *Es seien X_1, \dots, X_n unabhängig, identisch verteilte reellwertige Zufallsvariablen mit $E[X_1] = 0$, $\sigma^2 := \text{Var}(X_1) > 0$ und mit $\xi := E[|X_1|^3] < +\infty$. Seien für $i = 1, \dots, n$ $Y_i := \frac{X_i}{\sqrt{n\sigma^2}}$ und $W := \sum_{i=1}^n Y_i$. Dann gelten folgende Aussagen:*

(a) *Ist $h : \mathbb{R} \rightarrow \mathbb{R}$ beschränkt und Lipschitz-stetig mit minimaler Lipschitz-Konstanten $\|h'\|_\infty$, so gilt*

$$|E[h(W)] - N(h)| \leq \frac{3\|h'\|_\infty \xi}{\sigma^3} \frac{1}{\sqrt{n}}$$

(b)

$$\sup_{z \in \mathbb{R}} |P(W \leq z) - \Phi(z)| \leq \frac{\sqrt{3\xi}}{\sigma^{\frac{3}{2}}} \frac{1}{n^{\frac{1}{4}}}$$

□

Offenbar folgt aus Aussage (b) von Korollar 1.1.10 direkt folgender zentraler Grenzwertsatz für eine Folge unabhängig, identisch verteilter Zufallsvariablen mit existierendem dritten Moment:

Satz 1.1.11. *Es sei $(X_i)_{i \in \mathbb{N}}$ eine Folge unabhängiger, identisch verteilter, reellwertiger Zufallsvariablen mit $E[X_1] = 0$, $\sigma^2 := \text{Var}(X_1) > 0$ und $\xi := E[|X_1|^3] < +\infty$. Es sei für $n \in \mathbb{N}$ $S_n^* := \frac{1}{\sqrt{n\sigma^2}} \sum_{i=1}^n X_i$. Dann gilt:*

$$S_n^* \xrightarrow{\mathcal{D}} Z,$$

also

$$P(S_n^* \leq z) \xrightarrow{n \rightarrow \infty} \Phi(z)$$

für alle $z \in \mathbb{R}$.

1.1.2 Der Zugang mittels austauschbarer Paare

Es seien X_1, \dots, X_n Zufallsvariablen, definiert auf einem gemeinsamen Wahrscheinlichkeitsraum (Ω, \mathcal{A}, P) . Diese heißen *austauschbar*, wenn für jede Permutation $\pi \in \mathbb{S}_n$ gilt:

$$(X_1, \dots, X_n) \stackrel{\mathcal{D}}{=} (X_{\pi(1)}, \dots, X_{\pi(n)})$$

Im Folgenden wird für uns vor allem der Fall $n = 2$ von Interesse sein. In diesem Zusammenhang spricht man auch von *austauschbaren Paaren* von Zufallsvariablen. So bilden zwei Zufallsvariablen W, W' auf (Ω, \mathcal{A}, P) ein austauschbares Paar, wenn $(W, W') \stackrel{\mathcal{D}}{=} (W', W)$ gilt. Der in diesem Abschnitt vorgestellte Zugang über

1 Die Steinsche Methode

austauschbare Paare geht direkt auf C. Stein zurück und wird in [Ste86] durchweg verwendet.

Wir nennen weiter eine Funktion $F : \mathbb{R}^2 \rightarrow \mathbb{R}$ *anti-symmetrisch*, wenn für alle $x, y \in \mathbb{R}$ gilt, dass $F(y, x) = -F(x, y)$ ist. Wir beginnen mit einem ersten Lemma:

Lemma 1.1.12. *Es seien (W, W') ein austauschbares Paar auf dem Wahrscheinlichkeitsraum (Ω, \mathcal{A}, P) , sowie $F : \mathbb{R}^2 \rightarrow \mathbb{R}$ eine anti-symmetrische $\mathcal{B}^2 - \mathcal{B}^1$ -messbare Funktion mit $E[|F(W, W')|] < +\infty$. Weiter sei TF eine Version von $w \mapsto E[F(W, W')|W = w]$. Dann gilt:*

$$(i) \quad E[F(W, W')] = 0$$

$$(ii) \quad E[TF(W)] = 0$$

Beweis: zu (i):

Aus $(W, W') \stackrel{\mathcal{D}}{=} (W', W)$ folgt

$$F(W, W') \stackrel{\mathcal{D}}{=} F(W', W), \text{ also auch } E[F(W, W')] = E[F(W', W)].$$

Andererseits folgt aus der Anti-Symmetrie von F :

$$E[F(W', W)] = E[-F(W, W')] = -E[F(W, W')],$$

so dass insgesamt $E[F(W, W')] = 0$ folgt.

zu (ii): Es ist $TF(W)$ eine Version von $E[F(W, W')|W]$, so dass wir mit den Rechenregeln für bedingte Erwartungen nach (i) erhalten:

$$0 = E[F(W, W')] = E[E[F(W, W')|W]] = E[TF(W)]$$

□

Der folgende Satz ist von zentraler Bedeutung für den Zugang mittels austauschbarer Paare und findet sich im Wesentlichen schon in [Ste86].

Satz 1.1.13. *Es sei (W, W') ein austauschbares Paar von Zufallsvariablen auf (Ω, \mathcal{A}, P) . Es gelte $E[W^2] < +\infty$ und es gebe ein $\lambda \in (0; 1)$ mit $E[W'|W] = (1 - \lambda)W$. Dann gilt:*

$$(i) \quad E[W] = 0 \text{ und } E[(W - W')^2] = 2\lambda E[W^2].$$

(ii) *Ist $f : \mathbb{R} \rightarrow \mathbb{R}$ Borel - messbar und gibt es ein $C \in (0; +\infty)$ mit $|f(w)| \leq C(1 + |w|)$ für alle $w \in \mathbb{R}$, so ist*

$$E[Wf(W)] = \frac{1}{2\lambda} E[(W - W')(f(W) - f(W'))]$$

1.1 Normalapproximation mit Hilfe der Steinschen Methode

Beweis: zu(i): Es ist

$$E[W] = E[W'] = E[E[W'|W]] = E[(1 - \lambda)W] = (1 - \lambda)E[W],$$

also $E[W] = 0$. Weiter gilt nach den Rechenregeln für bedingte Erwartungen:

$$\begin{aligned} E[(W - W')^2] &= E[(W')^2 - 2W'W + W^2] \\ &= E[(W')^2] + E[W^2] - 2E[E[W'W|W]] \\ &= 2E[W^2] - 2E[WE[W'|W]] \\ &= 2E[W^2] - 2E[(1 - \lambda)W^2] \\ &= 2E[W^2](1 - (1 - \lambda)) = 2\lambda E[W^2] \end{aligned}$$

zu (ii): Wegen $|wf(w)| \leq C|w| + C|w|^2$ für alle $w \in \mathbb{R}$ und $E[W^2] < +\infty$ gilt $E[|Wf(W)|] < +\infty$. Sei $g : \mathbb{R}^2 \rightarrow \mathbb{R}$ gegeben durch $g(x, y) := (x - y)(f(x) + f(y))$. Dann ist g Borel - messbar und anti-symmetrisch und es gilt

$$\begin{aligned} E[|g(W, W')|] &\leq E[(|W| + |W'|)C(1 + |W| + 1 + |W'|)] \\ &= CE[2|W| + 2|W'| + W^2 + 2|W||W'| + (W')^2] \\ &\stackrel{\text{Cauchy-Schwarz}}{\leq} 4CE[|W|] + 2CE[W^2] + 2C\sqrt{E[W^2]E[(W')^2]} < +\infty \end{aligned}$$

Nach Lemma 1.1.12 (i) folgt nun

$$\begin{aligned} 0 &= E[g(W, W')] = E[(W - W')(f(W) + f(W'))] \\ &= E[(W - W')(f(W') - f(W))] + 2E[(W - W')f(W)] \\ &= E[(W - W')(f(W') - f(W))] + 2E[f(W)E[W - W'|W]] \\ &= E[(W - W')(f(W') - f(W))] + 2\lambda E[Wf(W)] \\ \Rightarrow E[Wf(W)] &= \frac{1}{2\lambda} E[(W - W')(f(W) - f(W'))] \end{aligned}$$

□

Wir wollen nun einen allgemeinen Satz beweisen, der für die Situation der austauschbaren Paare obere Schranken für $|E[h(W)] - N(h)|$ liefert.

Satz 1.1.14. *Es sei (W, W') ein austauschbares Paar, das die Voraussetzungen von Satz 1.1.13 erfüllt. Dann gilt:*

(i) *Man kann Satz 1.1.5 anwenden mit*

$$\delta = 4E\left[1 - \frac{1}{2\lambda} E[(W - W')^2|W]\right] + \frac{1}{2\lambda} E[|W - W'|^3]$$

1 Die Steinsche Methode

(ii) Für jede beschränkte, Lipschitz-stetige Funktion $h : \mathbb{R} \rightarrow \mathbb{R}$ mit minimaler Lipschitz-Konstante $\|h'\|_\infty$ ist

$$|E[h(W)] - N(h)| \leq |E[f'_h(W)(1 - \frac{1}{2\lambda}E[(W - W')^2|W])]| + \frac{1}{2\lambda}\|h'\|_\infty E[|W - W'|^3]$$

Beweis: Sei $h : \mathbb{R} \rightarrow \mathbb{R}$ beschränkt und Lipschitz-stetig und sei f_h die dazugehörige Lösung der Stein-Gleichung. Für $t \in \mathbb{R}$ definieren wir die Zufallsvariablen $Z(t)$ durch

$$Z(t) := (W - W')(I_{\{-(W - W') \leq t \leq 0\}} - I_{\{0 < t \leq -(W - W')\}})$$

Dann gilt $Z(t) \geq 0$ für alle $t \in \mathbb{R}$ und $\omega \in \Omega$. Weiter hat man

$$\int_{\mathbb{R}} Z(t) dt = (W - W')^2, \quad (1.5)$$

was man wie im Beweis zu Lemma 1.1.6 sieht.

Die Stein-Lösung f_h erfüllt nach Lemma 1.1.4 die Voraussetzung von Satz 1.1.13 (ii) und somit gilt:

$$\begin{aligned} E[W f_h(W)] &= \frac{1}{2\lambda} E[(W - W')(f_h(W) - f_h(W'))] \\ &= \frac{1}{2\lambda} E\left[\int_{W'}^W (W - W') f'_h(t) dt\right] \\ &= \frac{1}{2\lambda} E\left[\int_{-(W - W')}^0 f'_h(W + t)(W - W') dt\right] \\ &= \frac{1}{2\lambda} E\left[\int_{\mathbb{R}} f'_h(W + t) Z(t) dt\right] \end{aligned}$$

Nach (1.5) hat man weiterhin

$$E[f'_h(W)] = E[f'_h(W)(1 - \frac{1}{2\lambda}(W - W')^2)] + \frac{1}{2\lambda} E\left[\int_{\mathbb{R}} f'_h(W) Z(t) dt\right]$$

Insgesamt ergibt sich somit

1.1 Normalapproximation mit Hilfe der Steinschen Methode

$$\begin{aligned}
|E[h(W)] - N(h)| &= |E[f'_h(W) - Wf_h(W)]| \\
&= |E[f'_h(W)(1 - \frac{1}{2\lambda}(W - W')^2)] + \frac{1}{2\lambda}E[\int_{\mathbb{R}}(f'_h(W) - f'_h(W+t))Z(t)dt]| \\
&\leq |E[f'_h(W)(1 - \frac{1}{2\lambda}E[(W - W')^2|W])]| + \frac{1}{2\lambda}|E[\int_{\mathbb{R}}(f'_h(W) - f'_h(W+t))Z(t)dt]| \\
&\leq |E[f'_h(W)(1 - \frac{1}{2\lambda}E[(W - W')^2|W])]| + \frac{1}{2\lambda}E[\int_{\mathbb{R}}|f'_h(W) - f'_h(W+t)|Z(t)dt]
\end{aligned} \tag{1.6}$$

Nach Lemma 1.1.4 gilt

$$|f'_h(W) - f'_h(W+t)| \leq \|f''_h\|_{\infty}|t| \leq 2\|h'\|_{\infty}|t|,$$

also

$$\begin{aligned}
&E[\int_{\mathbb{R}}|f'_h(W) - f'_h(W+t)|Z(t)dt] \\
&\leq 2\|h'\|_{\infty}E[\int_{\mathbb{R}}|t|Z(t)dt] \\
&= 2\|h'\|_{\infty}E[\int_{\mathbb{R}}|t|(W - W')(I_{\{- (W - W') \leq t \leq 0\}} - I_{\{0 < t \leq - (W - W')\}})] \\
&= 2\|h'\|_{\infty}\frac{1}{2}E[|W - W'|^3] = \|h'\|_{\infty}E[|W - W'|^3]
\end{aligned} \tag{1.7}$$

Aus (1.6) und (1.7) zusammen ergibt sich sofort (ii). Für (i) schätzen wir nun noch den ersten Summanden in (1.6) wie folgt ab: Nach Lemma 1.1.4 gilt $\|f'_h\|_{\infty} \leq 4\|h'\|_{\infty}$ und es folgt

$$|E[f'_h(W)(1 - \frac{1}{2\lambda}E[(W - W')^2|W])]| \leq 4\|h'\|_{\infty}E[|1 - \frac{1}{2\lambda}E[(W - W')^2|W]|],$$

woraus mit Hilfe von (1.7) sofort

$$|E[h(W)] - N(h)| \leq 4\|h'\|_{\infty}E[|1 - \frac{1}{2\lambda}E[(W - W')^2|W]|] + \frac{\|h'\|_{\infty}}{2\lambda}E[|W - W'|^3]$$

1 Die Steinsche Methode

folgt. □

Wir wollen nun an einem Beispiel zeigen, wie sich der (sehr abstrakte) Satz 1.1.14 im konkreten Fall zur Normalapproximation anwenden lässt. Die folgenden Sätze stellen einige Hilfsresultate zusammen, welche wir dafür benötigen.

Satz 1.1.15. *Es seien $X_1, \dots, X_n, X_1^*, \dots, X_n^*, I$ unabhängige Zufallsvariablen, definiert auf dem Wahrscheinlichkeitsraum (Ω, \mathcal{A}, P) . Es gelte $X_i \stackrel{D}{=} X_i^*$ für $i = 1, \dots, n$ und es habe I eine beliebige Verteilung auf $\{1, \dots, n\}$. Seien $W := \sum_{i=1}^n X_i$ und $W' := W + X_I^* - X_I = \sum_{j \neq I} X_j + X_I^*$. Dann ist (W, W') ein austauschbares Paar von Zufallsvariablen.*

Beweis: Es seien $A, B \in \mathcal{B}$. Dann gilt:

$$\begin{aligned}
 & P(W \in A, W' \in B) \\
 &= P\left(\bigcup_{i=1}^n \{I = i\} \cap \{W \in A\} \cap \{W' \in B\}\right) \\
 &= \sum_{i=1}^n P\left(I = i, \sum_{j=1}^n X_j \in A, \sum_{j=1, j \neq i}^n X_j + X_i^* \in B\right) \\
 &= \sum_{i=1}^n P(I = i) \cdot P\left(\sum_{j=1}^n X_j \in A, \sum_{\substack{j=1 \\ j \neq i}}^n X_j + X_i^* \in B\right) \tag{1.8}
 \end{aligned}$$

Sei jetzt $i \in \{1, \dots, n\}$ fest gewählt. Wir definieren die Funktionen $S, S' : \mathbb{R}^{n+1} \rightarrow \mathbb{R}$ durch

$$S(x_1, \dots, x_{n+1}) := \sum_{j=1}^n x_j$$

und

$$S'(x_1, \dots, x_{n+1}) := \sum_{\substack{j=1 \\ j \neq i}}^n x_j + x_{n+1}$$

und weiter die Funktion $\tau : \mathbb{R}^{n+1} \rightarrow \mathbb{R}^{n+1}$ durch

$$\tau(x_1, \dots, x_{n+1}) := (x_1, \dots, x_{i-1}, x_{n+1}, x_{i+1}, \dots, x_n, x_i).$$

Dann gilt:

$$S' = S \circ \tau \text{ und } \tau^2 = id_{\mathbb{R}^{n+1}} \tag{1.9}$$

1.1 Normalapproximation mit Hilfe der Steinschen Methode

Weiterhin gilt nun nach Voraussetzung

$$\begin{aligned}
 P^{(X_1, \dots, X_n, X_i^*)} &= P^{X_1} \otimes \dots \otimes P^{X_n} \otimes P^{X_i^*} \\
 &= P^{X_1} \otimes \dots \otimes P^{X_n} \otimes P^{X_i} \\
 &= P^{X_1} \otimes \dots \otimes P^{X_{i-1}} \otimes P^{X_i^*} \otimes P^{X_{i+1}} \otimes \dots \otimes P^{X_n} \otimes P^{X_i} \\
 &= P^{(X_1, \dots, X_{i-1}, X_i^*, X_{i+1}, \dots, X_n, X_i)} \\
 &= P^{\tau(X_1, \dots, X_n, X_i^*)}
 \end{aligned} \tag{1.10}$$

Insgesamt folgt somit:

$$\begin{aligned}
 &P\left(\sum_{j=1}^n X_j \in A, \sum_{j=1, j \neq i}^n X_j + X_i^* \in B\right) \\
 &= \int_{\Omega} I_A(S(X_1, \dots, X_n, X_i^*)) I_B(S'(X_1, \dots, X_n, X_i^*)) dP \\
 &= \int_{\mathbb{R}^{n+1}} I_A(S(x_1, \dots, x_{n+1})) I_B(S'(x_1, \dots, x_{n+1})) dP^{(X_1, \dots, X_n, X_i^*)}(x_1, \dots, x_{n+1}) \\
 &\stackrel{(1.9)}{=} \int_{\mathbb{R}^{n+1}} I_A(S \circ \tau^2) I_B(S \circ \tau) dP^{(X_1, \dots, X_n, X_i^*)} \\
 &= \int_{\mathbb{R}^{n+1}} I_A(S \circ \tau) I_B(S) dP^{\tau(X_1, \dots, X_n, X_i^*)} \\
 &\stackrel{(1.10)}{=} \int_{\mathbb{R}^{n+1}} I_A(S'(x_1, \dots, x_{n+1})) 1_B(S(x_1, \dots, x_{n+1})) dP^{(X_1, \dots, X_n, X_i^*)}(x_1, \dots, x_{n+1}) \\
 &\stackrel{\text{s.o.}}{=} P\left(\sum_{j=1}^n X_j \in B, \sum_{j=1, j \neq i}^n X_j + X_i^* \in A\right)
 \end{aligned}$$

Eingesetzt in (1.8) ergibt dies sofort

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

also

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

□

Nun wollen wir zeigen, dass sich für unabhängige Zufallsvariablen die Voraussetzungen von Satz 1.1.14 stets erfüllen lassen.

1 Die Steinsche Methode

Satz 1.1.16. *Es seien $X_1, \dots, X_n, X_1^*, \dots, X_n^*, I$ unabhängige Zufallsvariablen, definiert auf dem Wahrscheinlichkeitsraum (Ω, \mathcal{A}, P) . Es gelte $X_i \stackrel{D}{=} X_i^*$ sowie $E[X_i] = 0$ für $i = 1, \dots, n$. Weiter besitze I eine Verteilung auf $\{1, \dots, n\}$ und es seien $W := \sum_{i=1}^n X_i$ und $W' := W + X_I^* - X_I = \sum_{j \neq I} X_j + X_I^*$. Dann gilt*

$$(i) \quad E[W'|W] = W - \sum_{i=1}^n P(I = i)E[X_i|W]$$

(ii) *Ist I gleichverteilt auf $\{1, \dots, n\}$, so gilt sogar*

$$E[W'|W] = \left(1 - \frac{1}{n}\right)W.$$

Beweis: zu (i): Es ist

$$\begin{aligned} E[W'|W] &= E[W + X_I^* - X_I|W] \\ &= W + E[X_I^*|W] - E[X_I|W] \\ &= W + E[X_I^*] - E[X_I|W], \end{aligned}$$

denn $X_I^* = \sum_{i=1}^n I_{\{I=i\}}X_i^*$ und W sind unabhängig.

Man hat nun wegen der Unabhängigkeit von X_i und I und, da $E[X_i^*] = 0$ gilt

$$E[X_I^*] = E\left[\sum_{i=1}^n I_{\{I=i\}}X_i^*\right] = \sum_{i=1}^n E[I_{\{I=i\}}X_i^*] = \sum_{i=1}^n P(I = i)E[X_i^*] = 0 \quad (1.11)$$

Weiter gilt:

$$\begin{aligned} E[X_I|W] &= E\left[\sum_{i=1}^n I_{\{I=i\}}X_i|W\right] \\ &= \sum_{i=1}^n E[I_{\{I=i\}}X_i|W] \\ &= \sum_{i=1}^n E[E[I_{\{I=i\}}X_i|W, I]|W] \\ &= \sum_{i=1}^n E[I_{\{I=i\}}E[X_i|W, I]|W] \end{aligned}$$

Es ist $\sigma(X_i, W)$ unabhängig von $\sigma(I)$, also gilt nach einem Resultat aus der Theorie bedingter Erwartungen, dass $E[X_i|W, I] = E[X_i|W]$ ist und diese Zufallsvariable ist $\sigma(W)$ -messbar. Somit folgt

1.1 Normalapproximation mit Hilfe der Steinschen Methode

$$\begin{aligned}
 E[X_I|W] &= \sum_{i=1}^n E[I_{\{I=i\}}E[X_i|W]|W] = \sum_{i=1}^n E[X_i|W]E[I_{\{I=i\}}|W] \\
 &\stackrel{I,W \text{ unabh.}}{=} \sum_{i=1}^n E[X_i|W]E[I_{\{I=i\}}] = \sum_{i=1}^n P(I=i)E[X_i|W] \quad (1.12)
 \end{aligned}$$

Insgesamt erhalten wir aus (1.11) und (1.12) die Behauptung von (i):

$$E[W'|W] = W - \sum_{i=1}^n P(I=i)E[X_i|W]$$

zu(ii): Nun gilt zusätzlich $P(I=i) = \frac{1}{n}$ für alle $i \in \{1, \dots, n\}$, so dass nach (i)

$$\begin{aligned}
 E[W'|W] &= W - \frac{1}{n} \sum_{i=1}^n E[X_i|W] = W - \frac{1}{n} E\left[\sum_{i=1}^n X_i|W\right] \\
 &= W - \frac{1}{n} E[W|W] = W - \frac{1}{n} W = \left(1 - \frac{1}{n}\right)W
 \end{aligned}$$

folgt.

□

Satz 1.1.17. *Es seien $X_1, \dots, X_n, X_1^*, \dots, X_n^*, I$ unabhängige Zufallsvariablen, definiert auf dem Wahrscheinlichkeitsraum (Ω, \mathcal{A}, P) . Es gelte $X_i \stackrel{D}{=} X_i^*$, $E[X_i] = 0$ für $i = 1, \dots, n$ sowie $\sum_{i=1}^n E[X_i^2] = 1$. Weiter sei I gleichverteilt auf $\{1, \dots, n\}$ und es seien $W := \sum_{i=1}^n X_i$ und $W' := W + X_I^* - X_I = \sum_{j \neq I} X_j + X_I^*$. Dann gilt*

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

Beweis: Es sei $X := (X_1, \dots, X_n)$. Dann gilt zunächst

$$\begin{aligned}
 E[(W - W')^2|X] &= E[(X_I - X_I^*)^2|X] \\
 &= E[X_I^2 - 2X_I X_I^* + (X_I^*)^2|X] \\
 &= E[X_I^2|X] - 2E[X_I X_I^*|X] + E[(X_I^*)^2|X]
 \end{aligned}$$

Wegen der Unabhängigkeit von X_I^* und X , folgt

1 Die Steinsche Methode

$$\begin{aligned}
 E[(X_I^*)^2|X] &= E[(X_I^*)^2] = E\left[\sum_{i=1}^n I_{\{I=i\}}(X_i^*)^2\right] \\
 &= \sum_{i=1}^n E[I_{\{I=i\}}(X_i^*)^2] \stackrel{I, X_i^* \text{ unabh.}}{=} \sum_{i=1}^n P(I=i)E[(X_i^*)^2] \\
 &= \frac{1}{n} \sum_{i=1}^n E[(X_i^*)^2] = \frac{1}{n} \sum_{i=1}^n E[X_i^2] = \frac{1}{n}
 \end{aligned} \tag{1.13}$$

Weiter ist

$$E[X_I X_I^*|X] = E[E[X_I X_I^*|X, I]|X] = E[X_I E[X_I^*|X, I]|X].$$

Nun ist $\sigma(X_I^*, I)$ unabhängig von $\sigma(X)$, also ist $E[X_I^*|X, I] = E[X_I^*|I]$. Wegen der Unabhängigkeit von I und X_1^*, \dots, X_n^* folgt weiter, dass gilt:

$$\begin{aligned}
 E[X_I^*|I] &= \sum_{i=1}^n E[I_{\{I=i\}} X_i^*|I] = \sum_{i=1}^n I_{\{I=i\}} E[X_i^*|I] \\
 &= \sum_{i=1}^n I_{\{I=i\}} E[X_i^*] = 0.
 \end{aligned}$$

Somit ist auch

$$E[X_I X_I^*|X] = E[X_I E[X_I^*|I]|X] = E[X_I 0|X] = 0. \tag{1.14}$$

Schließlich hat man

$$\begin{aligned}
 E[X_I^2|X] &= E\left[\sum_{i=1}^n I_{\{I=i\}} X_i^2|X\right] = \sum_{i=1}^n E[I_{\{I=i\}} X_i^2|X] \\
 &= \sum_{i=1}^n X_i^2 E[I_{\{I=i\}}|X] \stackrel{X_i, I \text{ unabh.}}{=} \sum_{i=1}^n X_i^2 E[I_{\{I=i\}}] \\
 &= \sum_{i=1}^n P(I=i) X_i^2 = \frac{1}{n} \sum_{i=1}^n X_i^2
 \end{aligned} \tag{1.15}$$

Setzt man (1.13), (1.14) und (1.15) oben ein, so erhält man

1.1 Normalapproximation mit Hilfe der Steinschen Methode

$$E[(W - W')^2|X] = \frac{1}{n} \sum_{i=1}^n X_i^2 - 2 \cdot 0 + \frac{1}{n} = \frac{1}{n} \left(1 + \sum_{i=1}^n E[X_i^2]\right).$$

Da W messbar ist bzgl. $\sigma(X)$ folgt $\sigma(W) \subseteq \sigma(X)$. Somit gilt

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

□

Wir können nun mit Hilfe der Methode des austauschbaren Paares einen mit Satz 1.1.8 vergleichbaren Satz beweisen:

Satz 1.1.18. *Seien X_1, \dots, X_n unabhängige Zufallsvariablen mit $E[X_i] = 0$ für $i = 1, \dots, n$ und $\sum_{i=1}^n E[X_i^2] = 1$. Weiter sei $W := \sum_{i=1}^n X_i$. Dann gilt für jede beschränkte, Lipschitz-stetige Funktion $h : \mathbb{R} \rightarrow \mathbb{R}$:*

$$|E[h(W)] - N(h)| \leq 6 \|h'\|_\infty \sum_{i=1}^n E[|X_i|^3]$$

Beweis: Nach evtl. Übergang zu einem anderen Wahrscheinlichkeitsraum dürfen wir annehmen, dass zusätzliche Zufallsvariablen X_1^*, \dots, X_n^*, I existieren, so dass $X_1, \dots, X_n, X_1^*, \dots, X_n^*, I$ unabhängig sind, dass $X_i^* \stackrel{D}{=} X_i$ für $i = 1, \dots, n$ gilt und dass I gleichverteilt auf $\{1, \dots, n\}$ ist.

(Gegebenenfalls betrachte man auf dem Raum $\mathbb{R}^{2n} \times \{1, \dots, n\}$ mit der σ -Algebra $\mathcal{B}^{2n} \otimes \mathcal{P}(\{1, \dots, n\})$ das Produktmaß

$P^{X_1} \otimes \dots \otimes P^{X_n} \otimes P^{X_1^*} \otimes \dots \otimes P^{X_n^*} \otimes \mathcal{U}(\{1, \dots, n\})$ und wähle als Zufallsvariablen die $2n+1$ Projektionen π_1, \dots, π_{2n+1} .)

Sei $W' := W + X_I^* - X_I$. Nach Satz 1.1.15 bildet (W, W') ein austauschbares Paar, welches nach Satz 1.1.16 (ii) die Bedingung $E[W'|W] = (1 - \frac{1}{n})W$ erfüllt. Daher können wir Satz 1.1.14 (ii) mit $\lambda = \frac{1}{n}$ anwenden und erhalten für beschränktes, Lipschitz-stetiges $h : \mathbb{R} \rightarrow \mathbb{R}$:

$$|E[h(W)] - N(h)| \leq |E[f'_h(W)(1 - \frac{n}{2}E[(W - W')^2|W])]| + \frac{n}{2} \|h'\|_\infty E[|W - W'|^3]$$

Nach Satz 1.1.17 ist $E[(W - W')^2|W] = \frac{1}{n}(1 + \sum_{i=1}^n E[X_i^2|W])$, also folgt für den ersten Summanden

1 Die Steinsche Methode

$$\begin{aligned}
& |E[f'_h(W)(1 - \frac{n}{2}E[(W - W')^2|W])]| \\
= & |E[f'_h(W)(1 - \frac{n}{2} \frac{1}{n}(1 + \sum_{i=1}^n E[X_i^2|W]))]| \\
= & \frac{1}{2}|E[f'_h(W)(1 - \sum_{i=1}^n E[X_i^2|W])]| \\
= & \frac{1}{2}|E[E[f'_h(W)(1 - \sum_{i=1}^n X_i^2)]|W]| \\
\stackrel{\text{totale Erwartung}}{=} & \frac{1}{2}|E[f'_h(W)(1 - \sum_{i=1}^n X_i^2)]| \\
\stackrel{\sum E[X_i^2]=1}{=} & \frac{1}{2}|E[f'_h(W) \sum_{i=1}^n (E[X_i^2] - X_i^2)]|
\end{aligned}$$

Da $f'_h(W - X_i)$ und $E[X_i^2] - X_i^2$ unabhängig sind mit $E[E[X_i^2] - X_i^2] = 0$ ist dies gleich

$$\begin{aligned}
& \frac{1}{2}|E[\sum_{i=1}^n (f'_h(W) - f'_h(W - X_i))(E[X_i^2] - X_i^2)]| \\
\leq & \frac{1}{2}\|f'_h\|_\infty \sum_{i=1}^n E[|X_i| |X_i^2 - E[X_i^2]|] \\
\leq & \frac{1}{2}\|f'_h\|_\infty \sum_{i=1}^n (E[|X_i|]E[X_i^2] + E[|X_i^3|]) \\
\stackrel{\text{Jensen}}{\leq} & \frac{1}{2}\|f'_h\|_\infty \sum_{i=1}^n 2E[|X_i|^3] = \|f'_h\|_\infty \sum_{i=1}^n E[|X_i|^3] \\
\stackrel{\text{Lemma 1.1.4}}{\leq} & 2\|h'\|_\infty \sum_{i=1}^n E[|X_i|^3] \tag{1.16}
\end{aligned}$$

Für den zweiten Summanden berechnen wir zunächst $E[|W - W'|^3]$:

1.1 Normalapproximation mit Hilfe der Steinschen Methode

$$\begin{aligned}
 E[|W - W'|^3] &= E[|X_I - X_I^*|^3] = E\left[\sum_{i=1}^n I_{\{I=i\}} |X_i - X_i^*|^3\right] \\
 &= \sum_{i=1}^n E[I_{\{I=i\}} |X_i - X_i^*|^3] = \sum_{i=1}^n P(I = i) E[|X_i - X_i^*|^3] \\
 &= \sum_{i=1}^n \frac{1}{n} E[|X_i - X_i^*|^3],
 \end{aligned}$$

denn die Zufallsvariablen $I_{\{I=i\}}$ und $|X_i - X_i^*|^3$ sind stochastisch unabhängig für $i = 1, \dots, n$. Für jedes $i \in \{1, \dots, n\}$ gilt nun:

$$\begin{aligned}
 &E[|X_i - X_i^*|^3] \\
 &\leq E[(|X_i| + |X_i^*|)^3] \\
 &= E[|X_i|^3 + 3|X_i|^2|X_i^*| + 3|X_i||X_i^*|^2 + |X_i^*|^3] \\
 &= E[|X_i|^3] + 3E[|X_i|^2|X_i^*|] + 3E[|X_i||X_i^*|^2] + E[|X_i^*|^3] \\
 &= E[|X_i|^3] + 3E[|X_i|^2]E[|X_i^*|] + 3E[|X_i|]E[|X_i^*|^2] + E[|X_i^*|^3],
 \end{aligned}$$

denn X_i und X_i^* sind unabhängig für $i = 1, \dots, n$. Da sie auch identisch verteilt sind, folgt

$$E[|X_i - X_i^*|^3] \leq 2E[|X_i|^3] + 6E[|X_i|]E[|X_i|^2]$$

und nach der Jensen-Ungleichung gilt

$$E[|X_i|] \leq (E[|X_i|^3])^{\frac{1}{3}} \text{ und } [E[|X_i|^2] \leq (E[|X_i|^3])^{\frac{2}{3}}$$

so dass wir

$$E[|X_i - X_i^*|^3] \leq 8E[|X_i|^3]$$

und damit

$$E[|W - W'|^3] \leq \frac{8}{n} \sum_{i=1}^n E[|X_i|^3] \tag{1.17}$$

erhalten. Insgesamt erhalten wir somit aus (1.16) und (1.17), dass

$$\begin{aligned}
 |E[h(W)] - N(h)| &\leq 2\|h'\|_\infty \sum_{i=1}^n E[|X_i|^3] + \|h'\|_\infty \frac{n}{2} \frac{8}{n} E[|X_i|^3] \\
 &= 6\|h'\|_\infty \sum_{i=1}^n E[|X_i|^3]
 \end{aligned}$$

□

Das folgende Korollar liefert, abgesehen von einer etwas schlechteren Konstanten, dasselbe Ergebnis wie Korollar 1.1.10:

Korollar 1.1.19. *Es seien Y_1, \dots, Y_n unabhängig, identisch verteilte reellwertige Zufallsvariablen mit $E[Y_1] = 0$, $\sigma^2 := \text{Var}(Y_1) > 0$ und mit $\xi := E[|Y_1|^3] < +\infty$. Seien für $i = 1, \dots, n$ $X_i := \frac{Y_i}{\sqrt{n\sigma^2}}$ und $W := \sum_{i=1}^n X_i$. Dann gelten folgende Aussagen:*

(a) *Ist $h : \mathbb{R} \rightarrow \mathbb{R}$ beschränkt und Lipschitz-stetig mit minimaler Lipschitz-Konstanten $\|h'\|_\infty$, so gilt*

$$|E[h(W)] - N(h)| \leq \frac{6\|h'\|_\infty \xi}{\sigma^3} \frac{1}{\sqrt{n}}$$

(b)

$$\sup_{z \in \mathbb{R}} |P(W \leq z) - \Phi(z)| \leq \frac{\sqrt{6\xi}}{\sigma^{\frac{3}{2}}} \frac{1}{n^{\frac{1}{4}}}$$

□

Aus diesem Resultat folgt sofort wieder die Aussage von Satz 1.1.11.

1.2 Der Satz von Berry-Esséen

Es seien (Ω, \mathcal{A}, P) ein Wahrscheinlichkeitsraum und $(X_i)_{i \in \mathbb{N}}$ eine Folge stochastisch unabhängiger, reellwertiger Zufallsvariablen aus $L_2(\Omega, \mathcal{A}, P)$ mit $\sigma_i^2 := \text{Var}(X_i) > 0$ für alle $i \in \mathbb{N}$. Sei weiter $\eta_i := E[X_i]$, $i \in \mathbb{N}$, und $s_n := (\sigma_1^2 + \dots + \sigma_n^2)^{\frac{1}{2}}$, $n \in \mathbb{N}$. Der zentrale Grenzwertsatz besagt, dass unter gewissen Zusatzbedingungen für alle $z \in \mathbb{R}$ gilt:

$$P\left(\frac{\sum_{i=1}^n (X_i - \eta_i)}{s_n} \leq z\right) \xrightarrow{n \rightarrow \infty} \Phi(z)$$

Dabei ist Φ wie üblich die Verteilungsfunktion der Standardnormalverteilung. In den Anwendungen liegen aber stets nur endlich viele zufällige Beobachtungen vor, so dass man an Stelle dieses Grenzwertresultats lieber eine konkrete obere Schranke vor allem in Form des Kolmogorov-Abstands hätte:

$$\sup_{z \in \mathbb{R}} \left| P\left(\frac{\sum_{i=1}^n (X_i - \eta_i)}{s_n} \leq z\right) - \Phi(z) \right| \leq K_n$$

Dabei sollte $(K_n)_{n \in \mathbb{N}}$ eine Nullfolge nichtnegativer reeller Zahlen sein, die man mit Hilfe der Verteilungen der X_i berechnen kann.

Berry und Esséen konnten eine derartige Aussage beweisen, sofern die dritten Momente aller X_i existieren:

Satz 1.2.1 (Berry-Esséen, 1945). *Es gibt eine universelle Konstante $K \in [0; +\infty)$ mit folgender Eigenschaft: Sind (Ω, \mathcal{A}, P) ein Wahrscheinlichkeitsraum, $n \in \mathbb{N}$ und $X_1, \dots, X_n \in L_2(\Omega, \mathcal{A}, P)$ unabhängig mit $\sigma_i^2 := \text{Var}(X_i) > 0$ für alle $i \in \{1, \dots, n\}$, so gilt mit $\eta_i := E[X_i]$, sowie $s_n := (\sigma_1^2 + \dots + \sigma_n^2)^{\frac{1}{2}}$, dass*

$$\sup_{z \in \mathbb{R}} \left| P\left(\frac{\sum_{i=1}^n (X_i - \eta_i)}{s_n} \leq z\right) - \Phi(z) \right| \leq \frac{K}{s_n^3} \sum_{i=1}^n E[|X_i - \eta_i|^3]$$

Die Aussage des obigen Satzes ist natürlich trivialerweise erfüllt, falls es ein $i \in \{1, \dots, n\}$ gibt mit $E[|X_i - \eta_i|^3] = +\infty$.

Im Spezialfall, wo die Zufallsvariablen X_1, \dots, X_n unabhängig und identisch verteilt sind mit $\sigma^2 := \text{Var}(X_1)$ und $\xi := E[|X_1 - \eta_1|^3] < +\infty$, so folgt $s_n = \sigma\sqrt{n}$ und es gilt das folgende Korollar:

Korollar 1.2.2 (Berry-Esséen, 1945). *Es gibt eine universelle Konstante $K \in [0; +\infty)$, so dass gilt:*

$$\sup_{z \in \mathbb{R}} \left| P\left(\frac{\sum_{i=1}^n (X_i - \eta_i)}{s_n} \leq z\right) - \Phi(z) \right| \leq \frac{K\xi}{\sigma^3\sqrt{n}}$$

1 Die Steinsche Methode

In diesem Fall kann man also $K_n = \frac{K\xi}{\sigma^3\sqrt{n}}$ wählen. Man kann zeigen, dass $n^{-\frac{1}{2}}$ die optimale Konvergenzrate ist, d.h. zu jedem $\alpha > \frac{1}{2}$ und jedem $K \in [0; +\infty)$ gibt es unabhängige, identisch verteilte Zufallsvariablen X_1, \dots, X_n mit existierendem dritten Moment ξ und

$$\sup_{z \in \mathbb{R}} |P\left(\frac{\sum_{i=1}^n (X_i - \eta_i)}{s_n} \leq z\right) - \Phi(z)| > \frac{K\xi}{\sigma^3 n^\alpha}.$$

Nachdem man eingesehen hat, dass die Konvergenzrate optimal ist, versucht man die Konstante K so klein wie möglich zu machen, da eine kleine Konstante für den Anwender eine Approximation durch die Normalverteilung natürlich eher rechtfertigt, als eine große. Van Beeck hat im Jahre 1972 bewiesen, dass man $K = 0,7975$ nehmen kann. Weiter ist bekannt, dass $K \geq \frac{1}{\sqrt{2\pi}}$ sein muss, siehe [GS77]. Kolmogorov vermutete, dass man $K = \frac{1}{\sqrt{2\pi}}$ nehmen kann, das heißt, dass dies die optimale Konstante ist.

Der „klassische“ Beweis des Satzes von Berry-Esséen benutzt die Theorie der charakteristischen Funktionen und eine sogenannte Glättungs-Ungleichung, siehe [GS77]. Wir wollen im Folgenden eine Aussage vom Berry-Esséen-Typ mit Hilfe der Steinschen Methode beweisen, die etwas schwächer ist als Satz 1.2.1, aber allgemeiner als die Aussage von Korollar 1.2.2. Dieser Beweis geht auf Erwin Bolthausen zurück und verwendet auf eine sehr geschickte Weise ein induktives Argument, vgl. [Bol84].

1.2.1 Bolthausens Beweis des Satzes von Berry-Esséen

In diesem Abschnitt wollen wir einen Beweis des Satzes von Berry und Esséen geben, der im Jahre 1984 von Erwin Bolthausen gefunden wurde und durch Verwendung eines Induktionsarguments die optimale Rate $n^{-\frac{1}{2}}$ liefert (vgl. [Bol84]). Dieser Beweis wurde von Mann in [Man94] ausgearbeitet.

Wir beginnen mit einigen Definitionen:

Für $n \in \mathbb{N}$ und $\gamma \in [1; +\infty)$ sei

$$\begin{aligned} L(n, \gamma) := \{ & X = (X_1, \dots, X_n) : X_1, \dots, X_n \text{ sind unabhängige, reellwertige} \\ & \text{Zufallsvariablen mit } E[X_i] = 0, E[X_i^2] = 1 \text{ und} \\ & E[|X_i|^3] \leq \gamma \text{ für } i = 1, \dots, n\} \end{aligned}$$

Der Grund dafür, dass man nur solche γ mit $\gamma \geq 1$ zulässt, ist, dass aus der Jensen-Ungleichung ansonsten $L(n, \gamma) = \emptyset$ folgen würde: Die Funktion $\psi : [0; +\infty) \rightarrow [0; +\infty)$ mit $\psi(x) := x^{\frac{3}{2}}$ ist konvex, so dass gilt:

$$1 = \psi(E[|X_i|^2]) \leq E[\psi(|X_i|^2)] = E[|X_i|^3] \leq \gamma$$

Wir betrachten weiter für $z \in \mathbb{R}$ und $\lambda \in (0; +\infty)$ die uns bereits bekannte Funktion $h_{z,\lambda} : \mathbb{R} \rightarrow \mathbb{R}$ mit

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

sowie die Funktion $h_z : \mathbb{R} \rightarrow \mathbb{R}$ mit $h_z(x) = I_{(-\infty; z]}(x)$. Wir definieren weiterhin

$$\delta(\lambda, \gamma, n) := \sup\left\{ \left| E\left[h_{z,\lambda}\left(\frac{1}{\sqrt{n}} \sum_{i=1}^n X_i\right) \right] - N(h_{z,\lambda}) \right| : z \in \mathbb{R}, \right. \\ \left. X = (X_1, \dots, X_n) \in L(n, \gamma) \right\}$$

und

$$\delta(\gamma, n) := \sup\left\{ \left| E\left[h_z\left(\frac{1}{\sqrt{n}} \sum_{i=1}^n X_i\right) \right] - N(h_z) \right| : z \in \mathbb{R}, X = (X_1, \dots, X_n) \in L(n, \gamma) \right\}.$$

Nach diesen Vorbereitungen können wir nun die zu beweisende Aussage wie folgt formulieren:

Satz 1.2.3 (Bolthausen, 1984). *Es gibt eine Konstante $K \in [0; +\infty)$ mit*

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

Die weiter oben erwähnte Ausarbeitung von Mann des Originalbeweises von Bolthausen führt zu einer Konstanten $K = 25$, was natürlich viel schlechter ist, als das Ergebnis von Van Beeck. Ein „typischer“ Beweis, wie er etwa in [GS77] zu finden ist, liefert die Konstante $K = 6$. Wir wollen zeigen, dass man, Bolthausens Beweisidee folgend, eine Konstante $K \leq 13$ erhalten kann.

Bevor wir den Satz beweisen, wollen wir jedoch das Ergebnis noch einordnen: Der Satz ist leider nicht so allgemein wie Satz 1.2.1, da wir in der Definition von $L(n, \gamma)$ $E[X_i^2] = \text{Var}(X_i) = 1$ vorausgesetzt haben und dies eine echte Einschränkung bedeutet, wohingegen man natürlich stets ohne Einschränkung annehmen kann, dass die Zufallsvariablen zentriert sind und sich ihre Varianzen zu Eins aufsummieren. Jedoch ist der Satz offenbar allgemeiner als Korollar 1.2.2.

1 Die Steinsche Methode

Nun zum *Beweis* des Satzes:

Zu $z \in \mathbb{R}$ und $\lambda > 0$ sei

$$f_{z,\lambda}(x) := \frac{1}{\varphi(x)} \int_{-\infty}^x (h_{z,\lambda}(y) - N(h_{z,\lambda})) \varphi(y) dy$$

die aus Abschnitt 1.1 bekannte Lösung der Stein-Gleichung

$$f'(x) - xf(x) = h_{z,\lambda}(x) - N(h_{z,\lambda})$$

Da $h_{z,\lambda}$ stetig ist, ist $f_{z,\lambda}$ stetig differenzierbar. Weiter seien

$$S_n := \frac{1}{\sqrt{n}} \sum_{i=1}^n X_i \quad \text{und} \quad S_{n,i} := \frac{1}{\sqrt{n}} \sum_{\substack{j=1 \\ j \neq i}}^n X_j = S_n - \frac{X_i}{\sqrt{n}},$$

so dass

$$\text{Var}(S_n) = 1 \quad \text{und} \quad \text{Var}(S_{n,i}) = \frac{n-1}{n}$$

gilt.

Wir fixieren nun $z \in \mathbb{R}$ und $\lambda > 0$ zunächst und schreiben daher der Übersichtlichkeit halber einfach h statt $h_{z,\lambda}$ sowie f statt $f_{z,\lambda}$. Da f eine Lösung der Steingleichung ist, gilt

$$\begin{aligned} E[h(S_n)] - N(h) &= E[f'(S_n) - S_n f(S_n)] \\ &= \frac{1}{n} \sum_{i=1}^n E[f'(S_n) - \sqrt{n} X_i f(S_n)] \\ &= \frac{1}{n} \sum_{i=1}^n E[f'(S_n) - f'(S_{n,i})] \\ &\quad + \frac{1}{n} \sum_{i=1}^n E[f'(S_{n,i}) - \sqrt{n} X_i f(S_n)] \end{aligned}$$

Man beachte, dass die jeweiligen Erwartungswerte existieren, da man Schranken für die Steinlösungen f und deren Ableitungen f' hat (vgl. Lemma 1.2.4).

Ferner gilt:

- $E[f'(S_{n,i})] = E[X_i^2] E[f'(S_{n,i})] = E[X_i^2 f'(S_{n,i})]$ und
- $E[\sqrt{n} X_i f(S_{n,i})] = \sqrt{n} E[X_i] E[f(S_{n,i})] = 0,$

denn X_i und $S_{n,i}$ sind unabhängig. Somit können wir schreiben

$$\begin{aligned} E[h(S_n)] - N(h) &= \frac{1}{n} \sum_{i=1}^n E[f'(S_n) - f'(S_{n,i})] \\ &- \frac{1}{n} \sum_{i=1}^n E[\sqrt{n}X_i(f(S_n) - f(S_{n,i})) - X_i^2 f'(S_{n,i})] \end{aligned}$$

Nach dem Hauptsatz der Differential- und Integralrechnung gilt nun punktweise in ω

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

und

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

so dass

$$\begin{aligned} E[h(S_n)] - N(h) &= \frac{1}{n} \sum_{i=1}^n E[f'(S_n) - f'(S_{n,i})] \\ &- \frac{1}{n} \sum_{i=1}^n E[X_i^2 \int_0^1 \left(f'(S_{n,i} + t \frac{X_i}{\sqrt{n}}) - f'(S_{n,i}) \right) dt] \quad (1.18) \end{aligned}$$

Der Philosophie der Steinschen Methode folgend verwenden wir nun (in z und λ) gleichmäßige Schranken für die Lösungen $f_{z,\lambda}$, ihre Ableitungen und für die Zuwächse der Ableitungen. Von nun an seien

$$C_1 := \sup\{\|f_{z,\lambda}\|_\infty : z \in \mathbb{R}, \lambda > 0\} \text{ und } C_2 := \sup\{\|f'_{z,\lambda}\|_\infty : z \in \mathbb{R}, \lambda > 0\}.$$

Das folgende Lemma stellt die gewünschten Schranken bereit. Die Verbesserung der Berry-Esséen-Konstante von 25 auf 13 gelingt uns im Folgenden aufgrund der Aussage (a) des Lemmas, dass $C_1 \leq \sqrt{\frac{\pi}{8}}$ gilt. Diese Schranke ist optimal, wie wir in Abschnitt 1.4 zeigen werden. B. Mann verwendet an dieser Stelle die grobe Abschätzung $C_1 \leq 3$.

Lemma 1.2.4. *Für alle $z, x, u \in \mathbb{R}$ und alle $\lambda \in [0; +\infty)$ gilt:*

1 Die Steinsche Methode

(a) $\|f_{z,\lambda}\|_\infty \leq \sqrt{\frac{\pi}{8}}$

(b) $\|id_{\mathbb{R}} \cdot f_{z,\lambda}\|_\infty \leq 1$

(c) $\|f'_{z,\lambda}\|_\infty \leq 2$

(d) $|f'_{z,\lambda}(x+u) - f'_{z,\lambda}(x)| \leq |u| \left(\|f_{z,\lambda}\|_\infty + \|f'_{z,\lambda}\|_\infty |x| + \frac{1}{\lambda} \int_0^1 I_{(z; z+\lambda)}(x+su) ds \right)$

(e) Für alle $a, b \in \mathbb{R}$ mit $a < b$, alle $n \in \mathbb{N}_{\geq 2}$ und jede reelle Konstante c gilt

$$E[I_{(a;b]}(S_{n,i} + c)] \leq \sqrt{\frac{n}{2\pi(n-1)}}(b-a) + 2\delta(\gamma, n-1)$$

Beweis: (a), (b) und (c) werden im Anhang (siehe Abschnitt 1.4) bewiesen.

Wir schreiben weiterhin h statt $h_{z,\lambda}$ sowie f statt $f_{z,\lambda}$:
zu(d): Nach der Steingleichung gilt

$$\begin{aligned} |f'(x+u) - f'(x)| &= |(x+u)f(x+u) + h(x+u) - N(h) - xf(x) - h(x) + N(h)| \\ &= |uf(x+u) + x(f(x+u) - f(x)) + h(x+u) - h(x)| \end{aligned}$$

Für alle $u \in \mathbb{R}$ gilt nun

$$h(x+u) - h(x) = -\frac{1}{\lambda} \int_x^{x+u} I_{(z; z+u]}(t) dt = \frac{-u}{\lambda} \int_0^1 I_{(z; z+u]}(x+su) ds,$$

denn $h = h_{z,\lambda}$ ist als Lipschitz-stetige Funktion absolut stetig (vgl. Anhang 1.6).
Außerdem ist ebenfalls für alle $u \in \mathbb{R}$

$$|f(x+u) - f(x)| \leq \|f'_{z,\lambda}\|_\infty \cdot |u| \leq C_1 |u|$$

also gilt

$$\begin{aligned} |f'(x+u) - f'(x)| &\leq |u| \left(\|f_{z,\lambda}\|_\infty + |x| \|f'_{z,\lambda}\|_\infty + \frac{1}{\lambda} \int_0^1 I_{(z; z+u]}(x+su) ds \right) \\ &\leq |u| \left(C_1 + |x| C_2 + \frac{1}{\lambda} \int_0^1 I_{(z; z+u]}(x+su) ds \right) \end{aligned}$$

Somit ist insbesondere (d) gezeigt.

zu (e): Nach Definition von $\delta(\cdot, \cdot)$ gilt für alle $z \in \mathbb{R}$

$$|E[h_z \left(\sqrt{\frac{n}{n-1}} S_{n,i} \right)] - N(h_z)| \leq \delta(\gamma, n-1),$$

also

$$\begin{aligned} & |E[I_{(a;b)}(S_{n,i} + c)]| = |E[I_{(a-c;b-c)}(S_{n,i})]| \\ = & |E[I_{(-\infty; \sqrt{\frac{n}{n-1}}(b-c))} \left(\sqrt{\frac{n}{n-1}} S_{n,i} \right)] - E[I_{(-\infty; \sqrt{\frac{n}{n-1}}(a-c))} \left(\sqrt{\frac{n}{n-1}} S_{n,i} \right)]| \\ \leq & |E[h_{\sqrt{\frac{n}{n-1}}(b-c)} \left(\sqrt{\frac{n}{n-1}} S_{n,i} \right)] - N(h_{\sqrt{\frac{n}{n-1}}(b-c)})| \\ + & |N(h_{\sqrt{\frac{n}{n-1}}(b-c)}) - N(h_{\sqrt{\frac{n}{n-1}}(a-c)})| \\ + & |E[h_{\sqrt{\frac{n}{n-1}}(a-c)} \left(\sqrt{\frac{n}{n-1}} S_{n,i} \right)] - N(h_{\sqrt{\frac{n}{n-1}}(a-c)})| \\ \leq & 2\delta(\gamma, n-1) + |N(h_{\sqrt{\frac{n}{n-1}}(b-c)}) - N(h_{\sqrt{\frac{n}{n-1}}(a-c)})| \end{aligned}$$

Es gilt

$$\left\{ h_{\sqrt{\frac{n}{n-1}}(b-c)} \neq h_{\sqrt{\frac{n}{n-1}}(a-c)} \right\} \subseteq \left[\sqrt{\frac{n}{n-1}}(a-c); \sqrt{\frac{n}{n-1}}(b-c) \right]$$

sowie $\|\varphi\|_\infty = \varphi(0) = \frac{1}{\sqrt{2\pi}}$, so dass der gesamte Term nach oben durch

$$\begin{aligned} & 2\delta(\gamma, n-1) + \frac{1}{\sqrt{2\pi}} \left(\sqrt{\frac{n}{n-1}}(b-c) - \sqrt{\frac{n}{n-1}}(a-c) \right) \\ & = 2\delta(\gamma, n-1) + \sqrt{\frac{n}{2\pi(n-1)}}(b-a) \end{aligned}$$

abgeschätzt werden kann. □

Wir beginnen nun damit, (1.18) abzuschätzen: Es gilt

1 Die Steinsche Methode

$$\begin{aligned}
& |E[f'(S_n) - f'(S_{n,i})]| \leq E[|f'(S_n) - f'(S_{n,i})|] \\
& \leq E\left[\frac{|X_i|}{\sqrt{n}} \left(C_1 + C_2|S_{n,i}| + \frac{1}{\lambda} \int_0^1 I_{(z; z+\lambda]}(S_{n,i} + s\frac{X_i}{\sqrt{n}}) ds \right)\right] \\
& = E\left[E\left[\frac{|X_i|}{\sqrt{n}} \left(C_1 + C_2|S_{n,i}| + \frac{1}{\lambda} \int_0^1 I_{(z; z+\lambda]}(S_{n,i} + s\frac{X_i}{\sqrt{n}}) ds \right) \middle| X_i\right]\right] \\
& = E\left[\frac{|X_i|}{\sqrt{n}} E\left[C_1 + C_2|S_{n,i}| + \frac{1}{\lambda} \int_0^1 I_{(z; z+\lambda]}(S_{n,i} + s\frac{X_i}{\sqrt{n}}) ds \middle| X_i\right]\right] \\
& = E\left[\frac{|X_i|}{\sqrt{n}} \left(C_1 + C_2 E[|S_{n,i}| | X_i] + E\left[\frac{1}{\lambda} \int_0^1 I_{(z; z+\lambda]}(S_{n,i} + s\frac{X_i}{\sqrt{n}}) ds \middle| X_i\right] \right)\right] \\
& = E\left[\frac{|X_i|}{\sqrt{n}} \left(C_1 + C_2 E[|S_{n,i}|] + \frac{1}{\lambda} E\left[\int_0^1 I_{(z; z+\lambda]}(S_{n,i} + s\frac{X_i}{\sqrt{n}}) ds \middle| X_i\right] \right)\right]
\end{aligned}$$

Dabei haben wir die Aussage von Lemma 1.2.4 (d), die Tatsache, dass $S_{n,i}$ und X_i unabhängig sind (für $E[|S_{n,i}| | X_i] = E[|S_{n,i}|]$), die Linearität der bedingten Erwartung sowie den Satz von der totalen Erwartung ausgenutzt.

Wir benötigen nun das folgende Resultat aus der Theorie bedingter Erwartungen, welches ebenfalls im Anhang (siehe Abschnitt 1.5) bewiesen wird:

Satz 1.2.5. *Seien (Ω, \mathcal{A}, P) ein Wahrscheinlichkeitsraum, $(\mathcal{X}, \mathcal{A}_{\mathcal{X}})$ und $(\mathcal{Y}, \mathcal{A}_{\mathcal{Y}})$ Messräume und seien $X : (\Omega, \mathcal{A}, P) \rightarrow (\mathcal{X}, \mathcal{A}_{\mathcal{X}})$ und $Y : (\Omega, \mathcal{A}, P) \rightarrow (\mathcal{Y}, \mathcal{A}_{\mathcal{Y}})$ zwei stochastisch unabhängige Zufallsvariablen.*

Sei weiter $f : (\mathcal{X} \times \mathcal{Y}, \mathcal{A}_{\mathcal{X}} \otimes \mathcal{A}_{\mathcal{Y}}) \rightarrow (\overline{\mathbb{R}}, \overline{\mathcal{B}})$ messbar mit $E[|f(X, Y)|] < +\infty$. Dann ist die Menge

$$G := \{x \in \mathcal{X} : f(x, Y) \text{ ist } P\text{-integrierbar}\} \in \mathcal{A}_{\mathcal{X}}$$

mit $P(X \in G) = 1$ und die Abbildung $h_f : (\mathcal{X}, \mathcal{A}_{\mathcal{X}}) \rightarrow (\overline{\mathbb{R}}, \overline{\mathcal{B}})$ mit

$$h_f(x) := \begin{cases} E[f(x, Y)] & \text{für } x \in G \\ 0 & \text{sonst} \end{cases}$$

ist eine Version von $x \mapsto E[f(X, Y) | X = x]$.

Wir wollen Satz 1.2.5 nun anwenden auf den Term

$$E\left[\int_0^1 I_{(z; z+\lambda)}\left(S_{n,i} + s\frac{X_i}{\sqrt{n}}\right) ds \mid X_i\right]$$

Nach diesem Satz gilt also wegen der Unabhängigkeit von X_i und $S_{n,i}$

$$\begin{aligned} E\left[\int_0^1 I_{(z; z+\lambda)}\left(S_{n,i} + s\frac{X_i}{\sqrt{n}}\right) ds \mid X_i = x\right] &= E\left[\int_0^1 I_{(z; z+\lambda)}\left(S_{n,i} + s\frac{x}{\sqrt{n}}\right) ds\right] \\ &= \int_0^1 E\left[I_{(z; z+\lambda)}\left(S_{n,i} + s\frac{x}{\sqrt{n}}\right)\right] ds \quad (1.19) \end{aligned}$$

Die letzte Identität folgt dabei aus dem Satz von Fubini, welcher anwendbar ist, da der Integrand $I_{(z; z+\lambda)}\left(S_{n,i} + s\frac{x}{\sqrt{n}}\right)$ nicht-negativ ist. Durch diese „Derandomisierung“ haben wir also die Zufallsvariable X_i durch die Konstante x ersetzt, was es uns ermöglicht, unterm ds -Integral Lemma 1.2.4 (e) anzuwenden. Wir erhalten, dass für jedes $x \in \mathbb{R}$ gilt

$$\begin{aligned} E\left[\int_0^1 I_{(z; z+\lambda)}\left(S_{n,i} + s\frac{x}{\sqrt{n}}\right) ds\right] &= \int_0^1 E\left[I_{(z; z+\lambda)}\left(S_{n,i} + s\frac{x}{\sqrt{n}}\right)\right] ds \\ &\leq \int_0^1 \left(\sqrt{\frac{n}{2\pi(n-1)}} \lambda + 2\delta(\gamma, n-1) \right) ds \\ &= \sqrt{\frac{n}{2\pi(n-1)}} \lambda + 2\delta(\gamma, n-1) \end{aligned}$$

Nach (1.19) ist dann aber auch

$$\left| E\left[\int_0^1 I_{(z; z+\lambda)}\left(S_{n,i} + s\frac{X_i}{\sqrt{n}}\right) ds \mid X_i\right] \right| \leq \sqrt{\frac{n}{2\pi(n-1)}} \lambda + 2\delta(\gamma, n-1)$$

Nach der Jensen-Ungleichung gilt ferner

$$E[|S_{n,i}|] \leq \sqrt{E[S_{n,i}^2]} = \sqrt{\text{Var}(S_{n,i})} = \sqrt{\frac{n-1}{n}} < 1$$

sowie

$$E[|X_i|] \leq \sqrt{E[X_i^2]} = 1.$$

1 Die Steinsche Methode

Somit können wir weiter abschätzen:

$$\begin{aligned}
& |E[f'(S_n) - f'(S_{n,i})]| \\
\leq & E\left[\frac{|X_i|}{\sqrt{n}} \left(C_1 + C_2 + \frac{1}{\lambda} \left(\sqrt{\frac{n}{2\pi(n-1)}} \lambda + 2\delta(\gamma, n-1) \right) \right)\right] \\
\leq & \frac{1}{\sqrt{n}} \left(C_1 + C_2 + \sqrt{\frac{n}{2\pi(n-1)}} + \frac{2}{\lambda} \delta(\gamma, n-1) \right) \\
\stackrel{\sqrt{\frac{n}{n-1}} \leq \sqrt{2}}{\leq} & \frac{1}{\sqrt{n}} \left(C_1 + C_2 + \frac{1}{\sqrt{\pi}} + \frac{2}{\lambda} \delta(\gamma, n-1) \right) \tag{1.20}
\end{aligned}$$

Für die Summanden des zweiten Terms in (1.18) ergibt sich nun auf analoge Weise:

$$\begin{aligned}
& |E[X_i^2 \int_0^1 \left(f'(S_{n,i} + t \frac{X_i}{\sqrt{n}}) - f'(S_{n,i}) \right) dt]| \\
\leq & E[X_i^2 \int_0^1 \left| f'(S_{n,i} + t \frac{X_i}{\sqrt{n}}) - f'(S_{n,i}) \right| dt] \\
\leq & E[X_i^2 \int_0^1 t \frac{|X_i|}{\sqrt{n}} \left(C_1 + C_2 |S_{n,i}| + \frac{1}{\lambda} \int_0^1 I_{(z; z+\lambda]}(S_{n,i} + \frac{stX_i}{\sqrt{n}}) ds \right) dt] \\
= & E[|X_i|^3 \int_0^1 t \frac{1}{\sqrt{n}} \left(C_1 + C_2 |S_{n,i}| + \frac{1}{\lambda} \int_0^1 I_{(z; z+\lambda]}(S_{n,i} + \frac{stX_i}{\sqrt{n}}) ds \right) dt] \\
= & E[E[|X_i|^3 \int_0^1 t \frac{1}{\sqrt{n}} \left(C_1 + C_2 |S_{n,i}| + \frac{1}{\lambda} \int_0^1 I_{(z; z+\lambda]}(S_{n,i} + \frac{stX_i}{\sqrt{n}}) ds \right) dt | X_i]] \\
= & E[|X_i|^3 E[\int_0^1 t \frac{1}{\sqrt{n}} \left(C_1 + C_2 |S_{n,i}| + \frac{1}{\lambda} \int_0^1 I_{(z; z+\lambda]}(S_{n,i} + \frac{stX_i}{\sqrt{n}}) ds \right) dt | X_i]]
\end{aligned}$$

Dies können wir weiter umformen zu

$$\begin{aligned}
 & E\left[\frac{|X_i|^3}{\sqrt{n}} \left(E\left[\int_0^1 t(C_1 + C_2|S_{n,i}|) dt|X_i\right] + E\left[\frac{1}{\lambda} \int_0^1 t \int_0^1 I_{(z; z+\lambda)}\left(S_{n,i} + \frac{stX_i}{\sqrt{n}}\right) ds dt|X_i\right] \right)\right] \\
 &= E\left[\frac{|X_i|^3}{\sqrt{n}} \left(\frac{C_1 + C_2E[|S_{n,i}||X_i]}{2} + E\left[\frac{1}{\lambda} \int_0^1 t \int_0^1 I_{(z; z+\lambda)}\left(S_{n,i} + \frac{stX_i}{\sqrt{n}}\right) ds dt|X_i\right] \right)\right] \\
 &= E\left[\frac{|X_i|^3}{\sqrt{n}} \left(\frac{C_1 + C_2E[|S_{n,i}|]}{2} + E\left[\frac{1}{\lambda} \int_0^1 t \int_0^1 I_{(z; z+\lambda)}\left(S_{n,i} + \frac{stX_i}{\sqrt{n}}\right) ds dt|X_i\right] \right)\right] \\
 &= \frac{E[|X_i|^3]}{\sqrt{n}} \frac{C_1 + C_2E[|S_{n,i}|]}{2} + E\left[\frac{|X_i|^3}{\lambda\sqrt{n}} \left(E\left[\int_0^1 \int_0^1 t I_{(z; z+\lambda)}\left(S_{n,i} + \frac{stX_i}{\sqrt{n}}\right) ds dt|X_i\right] \right)\right] \\
 &\leq \frac{\gamma(C_1 + C_2)}{2\sqrt{n}} + E\left[\frac{|X_i|^3}{\lambda\sqrt{n}} \left(E\left[\int_0^1 \int_0^1 t I_{(z; z+\lambda)}\left(S_{n,i} + \frac{stX_i}{\sqrt{n}}\right) ds dt|X_i\right] \right)\right] \tag{1.21}
 \end{aligned}$$

Für den zweiten Summanden in (1.21) erhalten wir nun wie oben gemäß Satz 1.2.5, dass für alle $x \in \mathbb{R}$ gilt:

$$\begin{aligned}
 & E\left[\int_0^1 \int_0^1 t I_{(z; z+\lambda)}\left(S_{n,i} + \frac{stX_i}{\sqrt{n}}\right) ds dt|X_i = x\right] \\
 &= E\left[\int_0^1 \int_0^1 t I_{(z; z+\lambda)}\left(S_{n,i} + \frac{stx}{\sqrt{n}}\right) ds dt\right] \\
 &= \int_0^1 \int_0^1 t E\left[I_{(z; z+\lambda)}\left(S_{n,i} + \frac{stx}{\sqrt{n}}\right)\right] ds dt \\
 &\leq \int_0^1 \int_0^1 t \left(\sqrt{\frac{n}{2\pi(n-1)}} \lambda + 2\delta(\gamma, n-1) \right) ds dt \\
 &= \frac{1}{2} \left(\sqrt{\frac{n}{2\pi(n-1)}} \lambda + 2\delta(\gamma, n-1) \right)
 \end{aligned}$$

Dabei haben wir wieder die oben erwähnte „Derandomisierung“ sowie den Satz von Fubini für den nicht-negativen Integranden verwendet, um die Aussage von Lemma 1.2.4 (d) anwenden zu können. Es folgt, dass auch

1 Die Steinsche Methode

$$E\left[\int_0^1 \int_0^1 t I_{(z; z+\lambda]}(S_{n,i} + \frac{stX_i}{\sqrt{n}}) ds dt | X_i\right] \leq \frac{1}{2} \left(\sqrt{\frac{n}{2\pi(n-1)}} \lambda + 2\delta(\gamma, n-1) \right)$$

ist. Wir erhalten somit

$$\begin{aligned} & |E[X_i^2 \int_0^1 \left(f'(S_{n,i} + t \frac{X_i}{\sqrt{n}}) - f'(S_{n,i}) \right) dt]| \\ & \leq \frac{\gamma(C_1 + C_2)}{2\sqrt{n}} + \frac{\gamma}{\lambda\sqrt{n}} \frac{1}{2} \left(\sqrt{\frac{n}{2\pi(n-1)}} \lambda + 2\delta(\gamma, n-1) \right) \\ & = \frac{\gamma(C_1 + C_2)}{2\sqrt{n}} + \frac{\gamma}{\sqrt{n}} \left(\frac{1}{2} \sqrt{\frac{n}{2\pi(n-1)}} + \frac{1}{\lambda} \delta(\gamma, n-1) \right) \\ & \stackrel{\frac{n}{n-1} \leq 2}{\leq} \frac{\gamma}{\sqrt{n}} \left(\frac{(C_1 + C_2)}{2} + \frac{1}{2\sqrt{\pi}} + \frac{1}{\lambda} \delta(\gamma, n-1) \right) \end{aligned} \quad (1.22)$$

Insgesamt erhalten wir somit aus (1.18), (1.20) und (1.22), dass für alle $z \in \mathbb{R}$ und $\lambda \in [0; +\infty)$ gilt

$$\begin{aligned} & |E[h_{z,\lambda}(S_n)] - N(h_{z,\lambda})| \\ & \leq \frac{1}{\sqrt{n}} \left(C_1 + C_2 + \frac{1}{\sqrt{\pi}} + \frac{2}{\lambda} \delta(\gamma, n-1) \right) \\ & + \frac{\gamma}{\sqrt{n}} \left(\frac{(C_1 + C_2)}{2} + \frac{1}{2\sqrt{\pi}} + \frac{1}{\lambda} \delta(\gamma, n-1) \right) \\ & \stackrel{\gamma \geq 1}{\leq} \frac{\gamma}{\sqrt{n}} \left(\frac{3}{2}(C_1 + C_2) + \frac{3}{2\sqrt{\pi}} + \frac{3}{\lambda} \delta(\gamma, n-1) \right) \\ & = \frac{3\gamma}{\sqrt{n}} \left(\frac{C_1 + C_2}{2} + \frac{1}{2\sqrt{\pi}} + \frac{1}{\lambda} \delta(\gamma, n-1) \right). \end{aligned}$$

Somit erhalten wir auch

$$\delta(\lambda, \gamma, n) \leq \frac{3\gamma}{\sqrt{n}} \left(\frac{C_1 + C_2}{2} + \frac{1}{2\sqrt{\pi}} + \frac{1}{\lambda} \delta(\gamma, n-1) \right) \quad (1.23)$$

Wenn wir nun $\delta(\gamma, n)$ nach oben durch einen Term „in“ $\delta(\lambda, \gamma, n)$ abschätzen können, so können wir hoffen, durch ein induktives Argument für $\delta(\gamma, n)$ die gewünschte Schranke der Form $\frac{K\gamma}{\sqrt{n}}$ zu erhalten! Dazu das folgende Lemma:

Lemma 1.2.6. Für alle $n \in \mathbb{N}$, $\lambda \in (0; +\infty)$ und $\gamma \in [1; +\infty)$ gilt

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

Beweis: Punktweise gilt $h_{z-\lambda, \lambda} \leq h_z \leq h_{z, \lambda}$, also

$$E[h_{z-\lambda, \lambda}(S_n)] - N(h_z) \leq E[h_z(S_n)] - N(h_z) \leq E[h_{z, \lambda}(S_n)] - N(h_z)$$

Daraus folgt, dass

$$\begin{aligned} & |E[h_z(S_n)] - N(h_z)| \\ & \leq \max(|E[h_{z-\lambda, \lambda}(S_n)] - N(h_z)|, |E[h_{z, \lambda}(S_n)] - N(h_z)|) \\ & \leq \max(|E[h_{z-\lambda, \lambda}(S_n)] - N(h_{z-\lambda, \lambda})| + |N(h_{z-\lambda, \lambda}) - N(h_z)|, \\ & \quad |E[h_{z, \lambda}(S_n)] - N(h_{z, \lambda})| + |N(h_{z, \lambda}) - N(h_z)|) \\ & \leq \delta(\lambda, \gamma, n) + \max(|N(h_{z-\lambda, \lambda}) - N(h_z)|, |N(h_{z, \lambda}) - N(h_z)|) \end{aligned}$$

Nun gilt

$$\begin{aligned} |N(h_{z, \lambda}) - N(h_z)| &= \int_z^{z+\lambda} \left(1 - \frac{y-z}{\lambda}\right) \varphi(y) dy \leq \frac{1}{\sqrt{2\pi}} \int_z^{z+\lambda} \left(1 - \frac{y-z}{\lambda}\right) dy \\ &= \frac{1}{\sqrt{2\pi}} \left(\lambda - \frac{1}{2\lambda} y^2 \Big|_0^\lambda\right) = \frac{1}{\sqrt{2\pi}} \frac{\lambda}{2} = \frac{\lambda}{2\sqrt{2\pi}} \end{aligned}$$

und

$$\begin{aligned} |N(h_{z-\lambda, \lambda}) - N(h_z)| &= N(h_z) - N(h_{z-\lambda, \lambda}) \\ &= \int_{z-\lambda}^z \left(1 - \left(1 - \frac{y - (z-\lambda)}{\lambda}\right)\right) \varphi(y) dy \\ &\leq \frac{1}{\sqrt{2\pi}} \int_{z-\lambda}^z \frac{y - (z-\lambda)}{\lambda} dy \\ &= \frac{1}{\sqrt{2\pi}\lambda} \int_0^\lambda y dy = \frac{\lambda}{2\sqrt{2\pi}}, \end{aligned}$$

also auch

1 Die Steinsche Methode

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

□

Insgesamt folgt somit für jedes $\lambda > 0$ und alle $n \in \mathbb{N}$ aus Lemma 1.2.6 und (1.23)

$$\delta(\gamma, n) \leq \frac{\gamma}{\sqrt{n}} \left(\frac{3}{2}(C_1 + C_2) + \frac{3}{2\sqrt{\pi}} + \frac{3}{\lambda}\delta(\gamma, n-1) \right) + \frac{\lambda}{2\sqrt{2\pi}} \quad (1.24)$$

Um eine Schranke der Form $\frac{K\gamma}{\sqrt{n}}$ für $\delta(\gamma, n)$ zu bekommen, müssen wir somit $\lambda = \frac{L\gamma}{\sqrt{n}}$ mit einer Konstanten $L > 0$ wählen. Wie sollte man L möglichst passend wählen?

Da der Term $\frac{3}{\lambda}\delta(\gamma, n-1)$ auftritt und wir induktiv vorgehen wollen, sollte „die Hälfte“ der Konstanten K von $\delta(\gamma, n-1)$ her kommen, so dass man zunächst an die Wahl $L = 6$ denken könnte. Nun hätten wir aber nach Induktionsvoraussetzung $\delta(\gamma, n-1) \leq \frac{K\gamma}{\sqrt{n-1}}$ und es wäre der Term $\frac{3}{\lambda}\delta(\gamma, n-1)$ nur durch $\frac{1}{2}\frac{K\gamma}{\sqrt{n-1}}$ nach oben abschätzbar! Da aber $\frac{1}{\sqrt{2(n-1)}} \leq \frac{1}{\sqrt{n}}$ für alle $n \geq 2$ gilt, wählen wir $L = 6\sqrt{2}$ und somit $\lambda = \frac{6\sqrt{2}}{\sqrt{n}}\gamma$. Nun gilt folgendes Lemma.

Lemma 1.2.7. *Für alle $\gamma \geq 1$ und alle $n \in \mathbb{N}$ gilt*

$$\delta(\gamma, n) \leq \left(3(C_1 + C_2) + \frac{9}{\sqrt{\pi}} \right) \frac{\gamma}{\sqrt{n}},$$

wobei C_1 und C_2 wie oben definiert sind.

In Lemma 1.2.4 hatten wir $C_1 \leq \sqrt{\frac{\pi}{8}}$ sowie $C_2 \leq 2$ gesehen. Nun gilt

$$3\left(\sqrt{\frac{\pi}{8}} + 2\right) + \frac{9}{\sqrt{\pi}} = 12,957\dots < 13,$$

so dass wir das folgende Korollar erhalten:

Korollar 1.2.8. *Für alle $\gamma \geq 1$ und alle $n \in \mathbb{N}$ gilt*

$$\delta(\gamma, n) \leq 13\frac{\gamma}{\sqrt{n}}.$$

Beweis von Lemma 1.2.7: Wir beweisen die Behauptung für festes $\gamma \geq 1$ durch vollständige Induktion nach $n \in \mathbb{N}$:

Induktionsanfang: $n = 1$. Da die Funktionen $h_{z,\lambda}$ nur Werte zwischen 0 und 1 annehmen und sowohl $E[h_{z,\lambda}(S_1)]$ als auch $N(h_{z,\lambda})$ nicht-negativ sind, folgt

$$\delta(\gamma, 1) \leq 1 \leq \frac{9}{\sqrt{\pi}} \leq \left(3(C_1 + C_2) + \frac{9}{\sqrt{\pi}} \right) \frac{\gamma}{\sqrt{1}}$$

Induktionsschritt: $n - 1 \rightarrow n$ für $n \geq 2$. Mit obiger Wahl $\lambda = \frac{6\sqrt{2}}{\sqrt{n}}\gamma$ gilt nach (1.24) und der Induktionsvoraussetzung:

$$\begin{aligned} & \delta(\gamma, n) \\ \leq & \frac{\gamma}{\sqrt{n}} \left(\frac{3}{2}(C_1 + C_2) + \frac{3}{2\sqrt{\pi}} \right) + \frac{3\gamma\sqrt{n}}{6\sqrt{2}\gamma\sqrt{n}} \delta(\gamma, n-1) + \frac{6\sqrt{2}\gamma}{2\sqrt{2\pi}\sqrt{n}} \\ = & \frac{\gamma}{\sqrt{n}} \left(\frac{3}{2}(C_1 + C_2) + \frac{3}{2\sqrt{\pi}} \right) + \frac{1}{2\sqrt{2}} \delta(\gamma, n-1) + \frac{3\gamma}{\sqrt{\pi}\sqrt{n}} \\ \stackrel{\text{Ind.-Vor.}}{\leq} & \frac{\gamma}{\sqrt{n}} \left(\frac{3}{2}(C_1 + C_2) + \frac{3}{2\sqrt{\pi}} + \frac{3}{\sqrt{\pi}} \right) + \frac{1}{2\sqrt{2}} \left(3(C_1 + C_2) + \frac{9}{\sqrt{\pi}} \right) \frac{\gamma}{\sqrt{n-1}} \\ \stackrel{\sqrt{2(n-1)} \geq \sqrt{n}}{\leq} & \frac{\gamma}{\sqrt{n}} \left(\frac{3}{2}(C_1 + C_2) + \frac{3}{2\sqrt{\pi}} + \frac{6}{2\sqrt{\pi}} + \frac{3}{2}(C_1 + C_2) + \frac{9}{2\sqrt{\pi}} \right) \\ = & \left(3(C_1 + C_2) + \frac{9}{\sqrt{\pi}} \right) \frac{\gamma}{\sqrt{n}} \end{aligned}$$

□

1.3 Die abstrakte Steinsche Methode

Bisher haben wir stets versucht, die Verteilung einer uns gegebenen Zufallsvariablen W durch die Standardnormalverteilung $N(0; 1)$ zu approximieren, indem wir mit einem geeigneten Abstandsmaß gemessen haben, wie weit die beiden Verteilungen voneinander entfernt sind. Dabei haben die *Stein-Gleichung* (1.1) und die Schranken für ihre Lösungen eine entscheidende Rolle gespielt. Was passiert nun, wenn wir als „Grenzverteilung“ gar nicht die Normalverteilung im Sinn haben, etwa wenn wir schon wissen, dass W mit Wahrscheinlichkeit 1 positive Werte annimmt und wir daher lieber den Abstand von W zu einer auf $(0; +\infty)$ konzentrierten Verteilung, etwa einer χ^2 -Verteilung messen wollen? Aufgrund der Stein-Charakterisierung der Standardnormalverteilung (vgl. Satz 1.1.2) ahnen wir, dass wir dann zunächst eine Stein-Gleichung für unsere neue Grenzverteilung benötigen. Wir wollen in diesem Abschnitt zum einen eine in diese Richtung gehende Verallgemeinerung der Steinschen Methode vorstellen, zum anderen aber auch kurz den „Original-Zugang“ von Stein, der in seiner Monographie [Ste86] ausführlich beschrieben ist, vorstellen.

1.3.1 Der algebraische Zugang von Stein

In seiner Monographie „*Approximate Computation of Expectations*“ aus dem Jahre 1986 behandelt Charles Stein die approximative Berechnung von Wahrscheinlichkeiten und allgemeiner von Erwartungswerten von einem abstrakten Standpunkt aus.

Gegeben sei eine reellwertige Zufallsvariable Y , deren Erwartungswert existiere und (zumindest approximativ) berechnet werden soll. Dazu schlägt Stein die folgende Strategie vor:

- (1) Bestimme den Kern $\ker E$ des Erwartungswert-Operators.
- (2) Suche eine Konstante $c \in \mathbb{R}$, so dass $Y - c \approx \tilde{Y}$ und $\tilde{Y} \in \ker E$, d.h. es gilt $E[\tilde{Y}] = 0$.

Dann gilt (unter gewissen Umständen) $E[Y] \approx c$.

Die Situation, die Stein betrachtet, ist dabei stets die folgende: Sei (Ω, \mathcal{A}, P) ein Wahrscheinlichkeitsraum, $(\mathcal{X}, \mathcal{D})$ ein Messraum und $X : \Omega \rightarrow \mathcal{X}$ eine Zufallsvariable. Dann liefert das Bildmaß P^X ein Wahrscheinlichkeitsmaß auf $(\mathcal{X}, \mathcal{D})$. Ist jetzt $h : \mathcal{X} \rightarrow \mathbb{R}$ eine Borel-messbare Abbildung und interessieren wir uns für deren Erwartungswert $E_{P^X}[h] = E[h(X)]$, so sollten wir nach obiger Strategie zunächst den Kern $\ker E_{P^X}$ bestimmen. Dazu schlägt Stein die Konstruktion einer weiteren Zufallsvariablen X' vor, so dass (X, X') ein austauschbares Paar bildet.

(Um dies zu erreichen, muss man eventuell auf einen anderen Grundraum als (Ω, \mathcal{A}, P) und somit auch auf eine andere Zufallsvariable als X zurückgreifen, die dann aber natürlich dieselbe Verteilung hat. Dies macht insofern nichts aus, als Erwartungswerte von Funktionen $h(X)$ nur von der Verteilung von X abhängen.) Ist unter dieser Voraussetzung $F : \mathcal{X}^2 \rightarrow \mathbb{R}$ eine Borel-messbare und anti-symmetrische Funktion (bezüglich der Produkt- σ -Algebra $\mathcal{D} \otimes \mathcal{D}$ auf $\mathcal{X} \times \mathcal{X}$) mit $E[|F(X, X')|] < \infty$, so gilt, wie wir schon gesehen haben, $E[F(X, X')] = 0$. Aus dem Satz von der totalen Erwartung folgt somit, dass die Funktion $f : \mathcal{X} \rightarrow \mathbb{R}$, definiert durch $f(x) := E[F(X, X')|X = x]$ im Kern von E_{PX} liegt. Somit bildet die Menge aller Funktionen f , die von dieser Form sind, einen Untervektorraum U von $\ker E_{PX}$. In [Ste86] wird gezeigt, dass unter gewissen Bedingungen an \mathcal{X} sogar $U = \ker E_{PX}$ gilt, wobei eine dieser Bedingungen die Endlichkeit von \mathcal{X} ist und die Bedingungen somit ziemlich restriktiv sind. Stein behauptet jedoch, dass auch in dem Fall, dass \mathcal{X} unendlich ist hinreichende Bedingungen für $U = \ker E_{PX}$ existieren sollten.

Wir wollen (analog zu [Ste86]) die bisherigen Ergebnisse in einem Diagramm festhalten. Dazu seien

$\mathcal{F} := \{F : \mathcal{X}^2 \rightarrow \mathbb{R} : F \text{ Borel-messbar, antisymmetrisch und } E[|F(X, X')|] < \infty\}$
 sowie $V := L_1(\mathcal{X}, \mathcal{D}, P^X)$. Weiter sei $T : \mathcal{F} \rightarrow V$ definiert durch $TF(x) := E[F(X, X')|X = x]$. Dann ist T linear und man hat das Diagramm

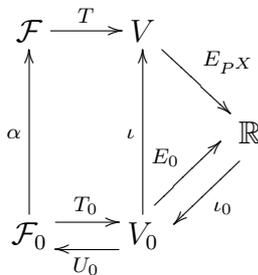
$$\mathcal{F} \xrightarrow{T} V \xrightarrow{E_{PX}} \mathbb{R} \quad (*)$$

mit $\text{Bild}T \subseteq \ker E_{PX}$, oder $E_{PX} \circ T = 0$.

Um das zweite Teilproblem in Steins Strategie, nämlich das Auffinden einer Konstanten $c \in \mathbb{R}$ mit $h - c \approx \tilde{h}$, wobei $\tilde{h} \in \ker E_{PX}$ ist, zu lösen, schlägt Stein vor, das Diagramm (*) um eine „Approximation“

$$\mathcal{F}_0 \xrightarrow{T_0} V_0 \xrightarrow{E_0} \mathbb{R}$$

zu einem Diagramm



1 Die Steinsche Methode

zu erweitern.

Dabei seien \mathcal{F}_0 ein \mathbb{R} -Vektorraum, $V_0 \subseteq V$ ein \mathbb{R} -Untervektorraum, α , T_0 , U_0 und E_0 seien lineare Abbildungen und $\iota : V_0 \rightarrow V$ sowie $\iota_0 : \mathbb{R} \rightarrow V_0$ seien die kanonischen Inklusionen (wobei die reelle Zahl t unter ι_0 mit der konstanten Abbildung $V_0 \rightarrow \mathbb{R}$, die überall den Wert t annimmt, identifiziert wird).

Die Abbildungen sind dabei so zu wählen, dass

- (a) $E_0 : V_0 \rightarrow \mathbb{R}$ eine „Approximation“ an $E_{PX} \circ \iota$ ist und
- (b) $\iota \circ T_0$ eine „Approximation“ an $T \circ \alpha$ ist.

Ferner gelte $T_0 \circ U_0 = id_{V_0} - \iota_0 \circ E_0$. Dann hat man für beliebiges $h \in V_0$:

$$\begin{aligned}
 0 &= (E_{PX} \circ T)((\alpha \circ U_0)(h)) \\
 &= E_{PX}[(\iota \circ T_0 \circ U_0)(h)] + E_{PX}[(T \circ \alpha \circ U_0 - \iota \circ T_0 \circ U_0)(h)] \\
 &= E_{PX}[(\iota \circ (id_{V_0} - \iota_0 \circ E_0))(h)] + E_{PX}[(T \circ \alpha - \iota \circ T_0) \circ U_0)(h)] \\
 &= E_{PX}[\iota(h)] - E_{PX}[(\iota \circ \iota_0 \circ E_0)(h)] + E_{PX}[(T \circ \alpha - \iota \circ T_0) \circ U_0)(h)] \\
 &= E_{PX}[h] - E_0(h) + E_{PX}[(T \circ \alpha - \iota \circ T_0) \circ U_0)(h)] \\
 &= E[h(X)] - E_0(h) + E_{PX}[(T \circ \alpha - \iota \circ T_0) \circ U_0)(h)]
 \end{aligned}$$

Äquivalent dazu ist die Gleichung

$$E[h(X)] - E_0(h) = E_{PX}[(\iota \circ T_0 - T \circ \alpha) \circ U_0)(h)] ,$$

welche die Differenz zwischen dem zu errechnenden Erwartungswert $E[h(X)]$ und dessen „Approximation“ $E_0(h)$ durch den Korrekturterm

$$E_{PX}[(\iota \circ T_0 - T \circ \alpha) \circ U_0)(h)]$$

zum Ausdruck bringt. Dieser sollte bei geschickter Wahl der Abbildungen „klein“ sein!

Möchte man nun als Anwendung dieses abstrakten Konzepts die Approximation einer festen Grenz- oder target-Verteilung μ_0 durch die Verteilung einer gegebenen Zufallsvariablen W betrachten, so muss man zunächst für diese Grenzverteilung die in obigem Diagramm vorkommenden Räume und Abbildungen angeben und insbesondere zeigen, dass die Abbildungen wohldefiniert sind, das heißt in die entsprechenden Räume abbilden. Wie dies für die Standardnormalverteilung geschieht, steht ausführlich in [Ste86] und soll hier nicht weiter verfolgt werden. Es sei nur bemerkt, dass uns einige dieser Abbildungen schon geläufig sind, zum Beispiel gilt

$$E_0(h) = N(h) = \int_{\mathbb{R}} h(y)\varphi(y)dy$$

sowie

$$U_0(h)(x) = f_h(x) = \frac{1}{\varphi(x)} \int_{-\infty}^x (h(y) - N(h))\varphi(y)dy$$

und

$$T_0(f)(x) = f'(x) - xf(x).$$

1.3.2 Verallgemeinerung der Stein-Charakterisierung und Stein-Gleichung

In Verallgemeinerung zur Normalapproximation auf dem Messraum $(\mathbb{R}, \mathcal{B})$ betrachten wir nun die folgende allgemeinere Situation:

Gegeben seien ein Wahrscheinlichkeitsraum $(\mathcal{X}, \mathcal{D}, \mu)$ und ein Unterraum V_0 von $V := L_1(\mu)$. Angenommen, wir wollen für alle $h \in V_0$ die Integrale $\int_{\mathcal{X}} h d\mu$ ausrechnen. Wenn die Verteilung μ eine komplizierte Struktur hat, etwa wenn μ die Verteilung einer Summe abhängiger Zufallsvariablen ist, so ist diese Aufgabe im Allgemeinen schwierig zu lösen. Angenommen, wir haben Anlass zur Vermutung, dass die Verteilung μ „nah“ an einer „einfacheren“ uns besser bekannten Verteilung μ_0 ist, für die wir die entsprechenden Integrale $\int_{\mathcal{X}} h d\mu_0$ leicht ausrechnen können. Dann würden wir uns unter Umständen mit der Approximation $\int_{\mathcal{X}} h d\mu_0$ zufrieden geben, wenn wir denn sicher wüssten, dass der dabei gemachte Fehler nicht zu groß ist.

Die Steinsche Methode ist eine Technik, welche obere Schranken für den bei dieser Approximation begangenen Fehler

$$\left| \int_{\mathcal{X}} h d\mu - \int_{\mathcal{X}} h d\mu_0 \right|$$

liefert. Diese Schranken sind meistens sogar gleichmäßig in $h \in V_0$. Die Grundidee ist dabei stets die folgende:

Sei V_0 ein Untervektorraum von $L_1(\mu) \cap L_1(\mu_0)$. Finde einen Unterraum $\mathcal{F}_0 \subseteq \mathcal{L}_0(\mathcal{X}, \mathcal{D})$ und einen (linearen) Operator $T_0 : \mathcal{F}_0 \rightarrow L_1(\mu)$, so dass für jedes $h \in V_0$ ein $f \in \mathcal{F}_0$ existiert mit

1 Die Steinsche Methode

$$T_0 f = h - \int_{\mathcal{X}} h d\mu_0 \quad (1.25)$$

Dann gilt offenbar

$$\left| \int_{\mathcal{X}} h d\mu - \int_{\mathcal{X}} h d\mu_0 \right| = \left| \int_{\mathcal{X}} T_0 f d\mu \right|, \quad (1.26)$$

sowie

$$\int_{\mathcal{X}} T_0 f d\mu_0 = 0. \quad (1.27)$$

Man nennt T_0 einen *Stein-Operator*, die Gleichung (1.25) eine *Stein-Gleichung* und die Funktion f eine *Steinlösung* von (1.25). Das Ziel ist es nun, einen Stein-Operator T_0 zu finden, für den gute Abschätzungen für $|\int_{\mathcal{X}} T_0 f d\mu|$ gefunden werden können. Wie lässt sich nun zu einer gegebenen „target-Verteilung“ μ_0 ein geeigneter Stein-Operator finden? Dafür gibt es prinzipiell zwei verschiedene Vorgehensweisen:

- (i) Der Stein-Operator kann durch ad-hoc-Überlegungen gefunden werden. Dies haben wir am Beispiel der Normalverteilung gesehen.
- (ii) Es können systematische Verfahren angewendet werden, um einen geeigneten Stein-Operator zu erhalten:
 - (a) Das erste Verfahren ähnelt dem Original-Zugang von Charles Stein, siehe Abschnitt 1.3.1. Man wähle einen Wahrscheinlichkeitsraum (Ω, \mathcal{A}, P) , auf dem ein austauschbares Paar (X, X') von Zufallsvariablen mit Randverteilung μ_0 existiert, d.h. es gilt $(X, X') \stackrel{\mathcal{D}}{=} (X', X)$ und $P^X = \mu_0 (= P^{X'})$. (Im Gegensatz zu Steins Original-Zugang gilt hier nicht $P^X = \mu!$)
 Es seien \mathcal{F} der \mathbb{R} -Vektorraum aller messbaren, anti-symmetrischen Funktionen $F : (\mathcal{X} \times \mathcal{X}, \mathcal{D} \otimes \mathcal{D}) \rightarrow (\mathbb{R}, \mathcal{B})$ mit $E[|F(X, X')|] < +\infty$, $\mathcal{F}_0 \subseteq \mathcal{L}_0(\mathcal{X}, \mathcal{D})$ ein Untervektorraum und $\alpha : \mathcal{F}_0 \rightarrow \mathcal{F}$ eine lineare Abbildung. Weiter sei $T : \mathcal{F} \rightarrow L_1(\mu)$ gegeben durch $TF(x) := E[F(X, X') | X = x]$ und $T_0 := T \circ \alpha$. Dann gilt für alle $f \in \mathcal{F}_0$ mit $F := \alpha(f)$ wegen $F(X, X') \stackrel{\mathcal{D}}{=} F(X', X)$ und wegen der Antisymmetrie von F

$$\begin{aligned}
 \int_{\mathcal{X}} T_0(f) d\mu_0 &= \int_{\mathcal{X}} T(\alpha(f)) d\mu_0 = \int_{\Omega} E[F(X, X') | X] dP \\
 &= \int_{\Omega} F(X, X') dP = \int_{\Omega} F(X', X) dP \\
 &= - \int_{\Omega} F(X, X') dP,
 \end{aligned}$$

also $\int_{\mathcal{X}} T_0(f) d\mu_0 = 0$.

Wie der Raum \mathcal{F}_0 , das austauschbare Paar (X, X') und die lineare Abbildung α gewählt werden sollten, muss dabei im Einzelfall entschieden werden. Außerdem erhält man auf diese Weise im Allgemeinen nicht zu jedem $h \in V_0$ ein $f \in \mathcal{L}_0(\mathcal{X}, \mathcal{D})$ mit $T_0 f = h - \int h d\mu_0$.

- (b) Das zweite Verfahren zur Konstruktion eines Stein-Operators T_0 ist die sogenannte *Generator-Methode*, welche von Barbour (vgl. [Bar88]) und Götze (vgl. [Göt91]) entwickelt wurde. Die Grundidee dieses Verfahrens ist es, den Operator T_0 als den Generator eines zeithomogenen Markov-Prozesses $(X_t)_{t \geq 0}$ mit stationärer Verteilung μ_0 zu wählen. Man definiert dann für alle $t > 0$ und alle „geeigneten“ Funktionen f

$$T_t f(x) := E[f(X_t) | X_0 = x].$$

Dann ist der *Generator* T_0 des Markov-Prozesses $(X_t)_{t \geq 0}$ gegeben durch

$$T_0 f(x) = \lim_{t \searrow 0} \frac{1}{t} (T_t f(x) - f(x)).$$

Ergebnisse aus der Theorie der zeitstetigen Markov-Prozesse besagen dann:

- (1) Eine Zufallsvariable X hat genau dann die Verteilung μ_0 , wenn $E[T_0 f(X)] = 0$ für alle f gilt, für die $T_0 f$ definiert ist.
- (2) Für alle $t > 0$ ist $T_t h - h = T_0 g_t$, wobei $g_t(x) := \int_0^t T_u h(x) du$. Lässt man auf beiden Seiten der Gleichung (formal) $t \rightarrow +\infty$ gehen, so folgt also $\int_{\mathcal{X}} h d\mu_0 - h = T_0 g$, wobei $g(x) := \int_0^\infty T_u h(x) du$.

Somit liefert die Generator-Methode nicht nur einen Stein-Operator, sondern auch einen Kandidaten für die Lösung der zugehörigen Stein-Gleichung, nämlich die Funktion $f_h(x) := - \int_0^\infty T_u h(x) du$.

1.4 Anhang: Beweis von Lemma 1.2.4

zu (a): Wir wollen $\sup_{z \in \mathbb{R}} \sup_{\lambda > 0} \|f_{z,\lambda}\|_\infty \leq \sqrt{\frac{\pi}{8}}$ zeigen. Da wir zunächst $z \in \mathbb{R}$ und $\lambda > 0$ festhalten, schreiben wir vorerst f statt $f_{z,\lambda}$ sowie h statt $h_{z,\lambda}$. Es gilt:

$$\begin{aligned} f(x) &= \frac{1}{\varphi(x)} \int_{-\infty}^x (h(y) - N(h)) \varphi(y) dy \\ &= \frac{1}{\varphi(x)} \int_{-\infty}^x h(y) \varphi(y) dy - \frac{N(h)}{\varphi(x)} \int_{-\infty}^x \varphi(y) dy \\ &= \frac{1}{\varphi(x)} \int_{-\infty}^x h(y) \varphi(y) dy - \frac{\Phi(x)}{\varphi(x)} N(h) \end{aligned}$$

1.Schritt: Berechnung von $N(h)$

Es ist

$$\begin{aligned} N(h) &= \int_{\mathbb{R}} h(y) \varphi(y) dy = \int_{-\infty}^z \varphi(y) dy + \int_z^{z+\lambda} \left(1 - \frac{y-z}{\lambda}\right) \varphi(y) dy \\ &= \Phi(z) + \left(1 + \frac{z}{\lambda}\right) (\Phi(z+\lambda) - \Phi(z)) + \frac{1}{\lambda} \int_z^{z+\lambda} (-y \varphi(y)) dy \\ &= \Phi(z) + \left(1 + \frac{z}{\lambda}\right) (\Phi(z+\lambda) - \Phi(z)) + \frac{1}{\lambda} (\varphi(z+\lambda) - \varphi(z)) \quad (1.28) \end{aligned}$$

$$= \Phi(z+\lambda) + \frac{z}{\lambda} (\Phi(z+\lambda) - \Phi(z)) + \frac{1}{\lambda} (\varphi(z+\lambda) - \varphi(z)) \quad (1.29)$$

$$= \Phi(z+\lambda) + \frac{1}{\lambda} \int_z^{z+\lambda} (z-y) \varphi(y) dy \quad (1.30)$$

2.Schritt: Berechnung von

$$G(x) := G_{z,\lambda}(x) := \int_{-\infty}^x h(y) \varphi(y) dy$$

- Für $x \leq z$ ist $G(x) = \int_{-\infty}^x \varphi(y) dy = \Phi(x)$.

- Für $z \leq x \leq z + \lambda$ ist

$$\begin{aligned}
 G(x) &= \Phi(z) + \int_z^x \left(1 - \frac{y-z}{\lambda}\right) \varphi(y) dy \\
 &= \Phi(z) + \left(1 + \frac{z}{\lambda}\right) (\Phi(x) - \Phi(z)) + \frac{1}{\lambda} (\varphi(x) - \varphi(z)) \\
 &= \Phi(x) + \frac{z}{\lambda} (\Phi(x) - \Phi(z)) + \frac{1}{\lambda} (\varphi(x) - \varphi(z)) \\
 &= \Phi(x) + \frac{1}{\lambda} \int_z^x (z-t) \varphi(t) dt
 \end{aligned}$$

- Für $x \geq z + \lambda$ ist

$$G(x) = \Phi(z) + \int_z^{z+\lambda} \left(1 - \frac{y-z}{\lambda}\right) \varphi(y) dy = N(h).$$

Somit erhalten wir für die Funktion f die beiden Darstellungen

$$f(x) = \begin{cases} \frac{\Phi(x)}{\varphi(x)} (1 - N(h)) & \text{für } x \leq z \\ \frac{\Phi(x)}{\varphi(x)} (1 - N(h)) + \frac{1}{\lambda \varphi(x)} (z(\Phi(x) - \Phi(z)) + \varphi(x) - \varphi(z)) & \text{für } z \leq x \leq z + \lambda \\ \frac{1 - \Phi(x)}{\varphi(x)} N(h) & \text{für } x \geq z + \lambda \end{cases} \quad (1.31)$$

und

$$f(x) = \begin{cases} \frac{\Phi(x)}{\varphi(x)} (1 - N(h)) & \text{für } x \leq z \\ \frac{\Phi(x)}{\varphi(x)} (1 - \Phi(z + \lambda)) + \frac{\int_z^x (z-y) \varphi(y) dy - \Phi(x) \int_z^{z+\lambda} (z-y) \varphi(y) dy}{\lambda \varphi(x)} & \text{für } z \leq x \leq z + \lambda \\ \frac{1 - \Phi(x)}{\varphi(x)} N(h) & \text{für } x \geq z + \lambda \end{cases} \quad (1.32)$$

Wir definieren die beiden Hilfsfunktionen $H_1, H_2 : \mathbb{R} \rightarrow \mathbb{R}$ durch

$$H_1(x) := \frac{\Phi(x)}{\varphi(x)}$$

und

$$H_2(x) := \frac{1 - \Phi(x)}{\varphi(x)} = \frac{\Phi(-x)}{\varphi(x)} = \frac{\Phi(-x)}{\varphi(-x)} = H_1(-x).$$

Diese sind offenbar beliebig oft differenzierbar. Ferner gelten die folgenden Aussagen:

1 Die Steinsche Methode

Lemma 1.4.1. (a) Die Funktion H_1 erfüllt $H_1'(x) > 0$ für alle $x \in \mathbb{R}$, das heißt sie ist insbesondere streng monoton wachsend. Ferner ist sie strikt positiv mit $\lim_{x \rightarrow -\infty} H_1(x) = 0$ und $\lim_{x \rightarrow +\infty} H_1(x) = +\infty$. Weiter gilt $H_1''(x) > 0$ für alle $x \in \mathbb{R}$, so dass H_1 auch strikt konvex ist.

(b) Die Funktion H_2 erfüllt $H_2'(x) < 0$ für alle $x \in \mathbb{R}$, das heißt sie ist insbesondere streng monoton fallend. Ferner ist sie strikt positiv mit $\lim_{x \rightarrow -\infty} H_2(x) = +\infty$ und $\lim_{x \rightarrow +\infty} H_2(x) = 0$. Weiter gilt $H_2''(x) > 0$ für alle $x \in \mathbb{R}$, so dass H_2 auch strikt konvex ist.

Beweis: zu (a): Offensichtlich ist H_1 überall strikt positiv und erfüllt

$$\lim_{x \rightarrow \infty} H_1(x) = +\infty.$$

Wegen

$$\lim_{x \rightarrow -\infty} \Phi(x) = \lim_{x \rightarrow -\infty} \varphi(x) = 0,$$

folgt aus den Regeln von de l'Hospital

$$\lim_{x \rightarrow -\infty} H_1(x) = \lim_{x \rightarrow -\infty} \frac{\Phi'(x)}{\varphi'(x)} = \lim_{x \rightarrow -\infty} \frac{\varphi(x)}{-x\varphi(x)} = 0.$$

Weiterhin ist

$$H_1'(x) = \frac{\varphi(x)^2 + x\varphi(x)\Phi(x)}{\varphi(x)^2} = \frac{\varphi(x) + x\Phi(x)}{\varphi(x)} = 1 + xH_1(x).$$

Sei $Z : \mathbb{R} \rightarrow \mathbb{R}$ gegeben durch $Z(x) := \varphi(x) + x\Phi(x)$. Dann gilt

$$\begin{aligned} \lim_{x \rightarrow -\infty} Z(x) &= 0 + \lim_{x \rightarrow -\infty} x\Phi(x) = \lim_{x \rightarrow -\infty} \frac{\Phi(x)}{\frac{1}{x}} \\ &\stackrel{\text{l'Hospital}}{=} \lim_{x \rightarrow -\infty} \frac{\varphi(x)}{\frac{-1}{x^2}} = \lim_{x \rightarrow -\infty} -x^2\varphi(x) \\ &= 0 \end{aligned}$$

und

$$Z'(x) = -x\varphi(x) + x\varphi(x) + \Phi(x) = \Phi(x) > 0$$

für alle $x \in \mathbb{R}$. Somit ist $Z(x) > 0$ für alle $x \in \mathbb{R}$ und daher auch

$$H_1'(x) = \frac{Z(x)}{\varphi(x)} > 0 \quad \text{für alle } x \in \mathbb{R},$$

so dass H_1 insbesondere streng monoton wächst.
Weiter gilt

$$\begin{aligned} H_1''(x) &= \frac{d}{dx}(1 + xH_1(x)) = \frac{d}{dx}(xH_1(x)) \\ &= \frac{\varphi(x)(\Phi(x) + x\varphi(x)) + x\varphi(x)x\Phi(x)}{\varphi(x)^2} \\ &= \frac{\Phi(x) + x\varphi(x) + x^2\Phi(x)}{\varphi(x)} \end{aligned}$$

Sei $Z_2(x) := \Phi(x) + x\varphi(x) + x^2\Phi(x)$. Es gilt wieder nach l'Hospital

$$\begin{aligned} \lim_{x \rightarrow -\infty} Z_2(x) &= \lim_{x \rightarrow -\infty} x^2\Phi(x) = \lim_{x \rightarrow -\infty} \frac{\Phi(x)}{\frac{1}{x^2}} \\ &= \lim_{x \rightarrow -\infty} \frac{\varphi(x)}{-\frac{2}{x^3}} = \lim_{x \rightarrow -\infty} \frac{-x^3}{2}\varphi(x) \\ &= 0 \end{aligned}$$

und weiter

$$\begin{aligned} Z_2'(x) &= \varphi(x) + \varphi(x) - x^2\varphi(x) + x^2\varphi(x) + 2x\Phi(x) \\ &= 2\varphi(x) + 2x\Phi(x) = 2(\varphi(x) + x\Phi(x)) \\ &= 2Z(x) > 0 \end{aligned}$$

für alle $x \in \mathbb{R}$. Somit ist auch $Z_2(x) > 0$ für alle $x \in \mathbb{R}$ und damit auch

$$H_1''(x) = \frac{Z_2(x)}{\varphi(x)} > 0$$

für alle $x \in \mathbb{R}$ und insbesondere ist H_1 strikt konvex.

zu (b): Wegen

$$H_2(x) = H_1(-x)$$

ist auch H_2 strikt positiv. Ferner gilt:

$$\lim_{x \rightarrow +\infty} H_2(x) = \lim_{x \rightarrow -\infty} H_1(x) = 0$$

und

$$\lim_{x \rightarrow -\infty} H_2(x) = \lim_{x \rightarrow +\infty} H_1(x) = +\infty.$$

1 Die Steinsche Methode

Nach der Kettenregel gilt:

$$H_2'(x) = -H_1'(-x) < 0$$

sowie

$$H_2''(x) = H_1''(-x) > 0$$

für alle $x \in \mathbb{R}$ nach (a). Somit ist H_2 streng monoton fallend und ebenfalls strikt konvex. □

Wegen $0 < N(h) < 1$ und

$$f(x) = \begin{cases} H_1(x)(1 - N(h)) & \text{für } x \leq z \\ H_2(x)N(h) & \text{für } x \geq z + \lambda \end{cases}$$

können wir bisher festhalten:

- $f|_{(-\infty; z]}$ ist strikt positiv, streng monoton wachsend und strikt konvex.
- $f|_{[z+\lambda; +\infty)}$ ist strikt positiv, streng monoton fallend und strikt konvex.
- $\lim_{x \rightarrow -\infty} f(x) = 0$
- $\lim_{x \rightarrow +\infty} f(x) = 0$
- f nimmt an einer Stelle $x^* \in [z; z + \lambda]$ sein globales Maximum an.

Somit müssen wir die Funktion f nur noch auf dem Intervall $[z; z + \lambda]$, wo sie beliebig oft differenzierbar ist, auf ein Maximum hin untersuchen. Bisher haben wir z und λ als feste Parameter angesehen und es war x variabel. Unser Ziel ist es zu zeigen, dass $\sqrt{\frac{\pi}{8}}$ eine obere Schranke ist von

$$\begin{aligned} \sup_{z \in \mathbb{R}} \sup_{\lambda > 0} \sup_{x \in [z; z + \lambda]} f_{z, \lambda}(x) &= \sup_{x \in \mathbb{R}} \sup_{(z, \lambda) \in \overline{M_x} \setminus \{(x, 0)\}} f_{z, \lambda}(x) \\ &\leq \sup_{x \in \mathbb{R}} \sup_{(z, \lambda) \in \overline{M_x}} f_{z, \lambda}(x), \end{aligned}$$

wobei wir für $x \in \mathbb{R}$

$$M_x := \{(z, \lambda) \in \mathbb{R} \times (0; +\infty) : z < x < z + \lambda\}$$

definieren sowie $f_{x, 0} := f_x$ setzen. Die Mengen M_x sind offenbar offen in \mathbb{R}^2 . Wir verändern nun unsere Betrachtungsweise und fassen $x \in \mathbb{R}$ als (zunächst festen) Parameter auf und untersuchen die Funktion

$$g_x : \overline{M_x} \rightarrow \mathbb{R} \quad \text{mit} \quad g_x(z, \lambda) := f_{z,\lambda}(x)$$

für zunächst festes x als Funktion von $(z, \lambda) \in \overline{M_x}$. Es ist g_x auf $\overline{M_x}$ stetig, wobei man die Stetigkeit von g_x im Punkt $(x, 0) \in \partial M_x$ mit Hilfe des Satzes von der majorisierten Konvergenz und der Definition der Funktionen $f_{z,\lambda}$ oder Mithilfe des Mittelwertsatzes (für Funktionen einer Veränderlichen) angewandt auf die Darstellungen (1.29) von $N(h_{z,\lambda})$ und (1.31) von $f_{z,\lambda}$ sehen kann. In den anderen Punkten ist sie aufgrund der Darstellungen für die Funktionen $f_{z,\lambda}$ evident. Anhand der Darstellungen der Funktion $f_{z,\lambda}$ im Intervall $[z; z + \lambda]$ erkennen wir ferner g_x auf M_x als beliebig oft stetig partiell differenzierbar. Ist somit $(z, \lambda) \in M_x$ eine lokale Maximalstelle von g_x , so folgt

$$\nabla g_x(z, \lambda) = 0.$$

Wir berechnen somit die partiellen Ableitungen

$$\frac{\partial}{\partial \lambda} g_x \quad \text{und} \quad \frac{\partial}{\partial z} g_x$$

und setzen diese gleich 0. Es gilt für $(z, \lambda) \in M_x$

$$g_x(z, \lambda) = \frac{\Phi(x)}{\varphi(x)}(1 - N(h)) + \frac{1}{\lambda\varphi(x)}(z(\Phi(x) - \Phi(z)) + \varphi(x) - \varphi(z)),$$

wobei

$$N(h_{z,\lambda}) = \Phi(z + \lambda) + \frac{z}{\lambda}(\Phi(z + \lambda) - \Phi(z)) + \frac{1}{\lambda}(\varphi(z + \lambda) - \varphi(z))$$

ist. Wir berechnen zunächst die partiellen Ableitungen von $N(h_{z,\lambda})$ nach z und λ .

Lemma 1.4.2. *Für alle $z \in \mathbb{R}$ und $\lambda > 0$ gilt:*

(a)

$$\begin{aligned} \frac{\partial N(h_{z,\lambda})}{\partial \lambda} &= \frac{1}{\lambda^2} \int_z^{z+\lambda} (t - z)\varphi(t) dt \\ &= -\frac{1}{\lambda^2}(\varphi(z + \lambda) - \varphi(z) + z(\Phi(z + \lambda) - \Phi(z))) > 0 \end{aligned}$$

1 Die Steinsche Methode

(b)

$$\begin{aligned}\frac{\partial N(h_{z,\lambda})}{\partial z} &= \frac{1}{\lambda} \int_z^{z+\lambda} \varphi(t) dt \\ &= \frac{1}{\lambda} (\Phi(z+\lambda) - \Phi(z)) > 0\end{aligned}$$

Beweis: zu (a):

$$\begin{aligned}\frac{\partial N(h_{z,\lambda})}{\partial \lambda} &= \varphi(z+\lambda) - \frac{z}{\lambda^2} (\Phi(z+\lambda) - \Phi(z)) + \frac{z}{\lambda} \varphi(z+\lambda) \\ &\quad - \frac{1}{\lambda^2} (\varphi(z+\lambda) - \varphi(z)) + \frac{1}{\lambda} (-(z+\lambda)\varphi(z+\lambda)) \\ &= -\frac{1}{\lambda^2} (z(\Phi(z+\lambda) - \Phi(z)) + (\varphi(z+\lambda) - \varphi(z))) \\ &= -\frac{1}{\lambda^2} \int_z^{z+\lambda} (z-t)\varphi(t) dt = \frac{1}{\lambda^2} \int_z^{z+\lambda} (t-z)\varphi(t) dt \\ &> 0\end{aligned}$$

zu (b):

$$\begin{aligned}\frac{\partial N(h_{z,\lambda})}{\partial z} &= \varphi(z+\lambda) + \frac{1}{\lambda} (\Phi(z+\lambda) - \Phi(z)) \\ &\quad + \frac{z}{\lambda} (\varphi(z+\lambda) - \varphi(z)) + \frac{1}{\lambda} (-(z+\lambda)\varphi(z+\lambda) + z\varphi(z)) \\ &= \frac{1}{\lambda} (\Phi(z+\lambda) - \Phi(z)) = \frac{1}{\lambda} \int_z^{z+\lambda} \varphi(t) dt \\ &> 0\end{aligned}$$

Es ist anschaulich klar, dass $N(h_{z,\lambda})$ eine wachsende Funktion sowohl von λ als auch von z ist! □

Nun können wir auch die partiellen Ableitungen von g_x nach λ und z berechnen:

Satz 1.4.3. *Es gilt*

(a)

$$\frac{\partial g_x(z, \lambda)}{\partial \lambda} = \frac{-1}{\varphi(x)\lambda^2} \left(\Phi(x) \int_z^{z+\lambda} (t-z)\varphi(t) dt - \int_z^x (t-z)\varphi(t) dt \right)$$

(b)

$$\frac{\partial g_x(z, \lambda)}{\partial z} = \frac{1}{\lambda \varphi(x)} \left(-\Phi(x) \int_z^{z+\lambda} \varphi(t) dt + \int_z^x \varphi(t) dt \right)$$

(c) Ist insbesondere $(z, \lambda) \in M_x$ ein Punkt mit $\nabla g_x(z, \lambda) = 0$, so erhalten wir die zueinander äquivalenten Gleichungssysteme

$$\begin{aligned} \Phi(x) \int_z^{z+\lambda} \varphi(t) dt &= \int_z^x \varphi(t) dt \\ \wedge \quad \Phi(x) \int_z^{z+\lambda} t \varphi(t) dt &= \int_z^x t \varphi(t) dt \end{aligned} \quad (1.33)$$

und

$$\begin{aligned} \Phi(x) (\Phi(z + \lambda) - \Phi(z)) &= \Phi(x) - \Phi(z) \\ \wedge \quad \Phi(x) (\varphi(z + \lambda) - \varphi(z)) &= \varphi(x) - \varphi(z) \end{aligned} \quad (1.34)$$

Beweis: Es ist

$$g_x(z, \lambda) = \frac{\Phi(x)}{\varphi(x)} (1 - N(h)) + \frac{1}{\lambda \varphi(x)} (z(\Phi(x) - \Phi(z)) + \varphi(x) - \varphi(z)) .$$

Somit folgt

$$\begin{aligned} \frac{\partial g_x(z, \lambda)}{\partial \lambda} &= -\frac{\Phi(x)}{\varphi(x)} \frac{\partial N(h_{z,\lambda})}{\partial \lambda} - \frac{z}{\lambda^2 \varphi(x)} (z(\Phi(x) - \Phi(z))) \\ &\quad - \frac{1}{\lambda^2 \varphi(x)} (\varphi(x) - \varphi(z)) \\ &\stackrel{\text{Lemma 1.4.2 (a)}}{=} -\frac{\Phi(x)}{\varphi(x)} \frac{1}{\lambda^2} \int_z^{z+\lambda} (t - z) \varphi(t) dt - \frac{1}{\varphi(x) \lambda^2} \int_z^x (z - t) \varphi(t) dt \\ &= -\frac{1}{\varphi(x) \lambda^2} \left(\Phi(x) \int_z^{z+\lambda} (t - z) \varphi(t) dt - \int_z^x (t - z) \varphi(t) dt \right) \end{aligned}$$

1 Die Steinsche Methode

und

$$\begin{aligned}
 \frac{\partial g_x(z, \lambda)}{\partial z} &= -\frac{\Phi(x)}{\varphi(x)} \frac{\partial N(h_{z,\lambda})}{\partial z} + \frac{1}{\lambda\varphi(x)} (\Phi(x) - \Phi(z) - z\varphi(z) + z\varphi(x)) \\
 &\stackrel{\text{Lemma 1.4.2 (b)}}{=} -\frac{\Phi(x)}{\varphi(x)} \frac{1}{\lambda} \int_z^{z+\lambda} \varphi(t) dt + \frac{1}{\lambda\varphi(x)} \int_z^x \varphi(t) dt \\
 &= \frac{1}{\lambda\varphi(x)} \left(-\Phi(x) \int_z^{z+\lambda} \varphi(t) dt + \int_z^x \varphi(t) dt \right)
 \end{aligned}$$

Damit sind (a) und (b) gezeigt. Ist jetzt $(z, \lambda) \in M_x$ mit $\nabla g_x(z, \lambda) = 0$, so gilt

$$\frac{\partial g_x(z, \lambda)}{\partial \lambda} = \frac{\partial g_x(z, \lambda)}{\partial z} = 0.$$

Daraus folgt

$$\Phi(x) \int_z^{z+\lambda} \varphi(t) dt = \int_z^x \varphi(t) dt.$$

und

$$0 = \left(\Phi(x) \int_z^{z+\lambda} (t-z)\varphi(t) dt - \int_z^x (t-z)\varphi(t) dt \right)$$

Setzt man die erste in die zweite Gleichung ein, so erhält man das zu diesen beiden Gleichungen äquivalente Gleichungssystem

$$\begin{aligned}
 \Phi(x) \int_z^{z+\lambda} \varphi(t) dt &= \int_z^x \varphi(t) dt \\
 \wedge \quad \Phi(x) \int_z^{z+\lambda} t\varphi(t) dt &= \int_z^x t\varphi(t) dt
 \end{aligned}$$

welches das System (1.33) ist. Das dazu äquivalente Gleichungssystem (1.34) erhält man hieraus durchs Auswerten der Integrale. □

Wir geben nun eine Konstante $S \in (0; +\infty)$ an mit

$$g_x(z, \lambda) \leq S$$

für alle $z, x \in \mathbb{R}$ und $\lambda > 0$.

Satz 1.4.4. Für jedes $x \in \mathbb{R}$ gilt

$$\sup_{(z,\lambda) \in \overline{M_x}} g_x(z, \lambda) \leq S := \sup_{t \in \mathbb{R}} \frac{\Phi(t)}{\varphi(t)} (1 - \Phi(t)).$$

Beweis: Wir machen die folgende Fallunterscheidung:

Fall 1: Die Funktion g_x nimmt ihr globales Maximum in einem Punkt $(z, \lambda) \in M_x$ an.

Dann gilt

$$\nabla g_x(z, \lambda) = 0.$$

Indem wir in die Gleichung für g_x die Gleichungen aus dem System (1.34) einsetzen, die ja in solch einem Punkt nach Satz 1.4.3 erfüllt sind, erhalten wir

$$\begin{aligned} g_x(z, \lambda) &= \frac{\Phi(x)}{\varphi(x)} (1 - N(h_{z,\lambda})) + \frac{1}{\lambda\varphi(x)} (z(\Phi(x) - \Phi(z)) + \varphi(x) - \varphi(z)) \\ &= \frac{\Phi(x)}{\varphi(x)} \left(1 - \Phi(z + \lambda) - \frac{z}{\lambda} (\Phi(z + \lambda) - \Phi(z)) - \frac{1}{\lambda} (\varphi(z + \lambda) - \varphi(z)) \right) \\ &\quad + \frac{z}{\lambda\varphi(x)} (\Phi(x)(\Phi(z + \lambda) - \Phi(z))) + \frac{\Phi(x)}{\lambda\varphi(x)} (\varphi(z + \lambda) - \varphi(z)) \\ &= \frac{\Phi(x)}{\varphi(x)} (1 - \Phi(z + \lambda)) \\ &= H_1(x)(1 - \Phi(z + \lambda)). \end{aligned}$$

Da nach Lemma 1.4.1 (a) die Funktion H_1 streng monoton wachsend ist und $x < z + \lambda$ gilt, folgt hieraus

$$g_x(z, \lambda) < H_1(z + \lambda)(1 - \Phi(z + \lambda)) = \frac{\Phi(z + \lambda)}{\varphi(z + \lambda)} (1 - \Phi(z + \lambda)) \leq S,$$

wie gewünscht.

Fall 2: Die Funktion g_x nimmt ihr Maximum auf dem Rand $\partial\overline{M_x}$ an.

Es gilt:

$$\partial\overline{M_x} = \partial M_x = \{(z, \lambda) : z = x\} \cup \{(z, \lambda) : x = z + \lambda\}$$

(a) $z = x$: Dann gilt nach Gleichung (1.31):

$$\begin{aligned} g_x(z, \lambda) &= \frac{\Phi(x)}{\varphi(x)} (1 - N(h_{z,\lambda})) = \frac{\Phi(z)}{\varphi(z)} (1 - N(h_{z,\lambda})) \\ &\leq \frac{\Phi(z)}{\varphi(z)} (1 - \Phi(z)) \leq S \end{aligned}$$

1 Die Steinsche Methode

(b) $x = z + \lambda$: Dann gilt nach Gleichung (1.31):

$$\begin{aligned} g_x(z, \lambda) &= \frac{1 - \Phi(z + \lambda)}{\varphi(z + \lambda)} N(h_{z, \lambda}) \\ &\leq \frac{1 - \Phi(z + \lambda)}{\varphi(z + \lambda)} \Phi(z + \lambda) \\ &\leq S \end{aligned}$$

Fall 3: Die Funktion g_x besitzt kein Maximum.

Da die stetige Funktion g_x auf jedem Kompaktum $K \subseteq \overline{M_x}$ ihr Maximum annimmt, genügt es, das Verhalten von $g_x(z, \lambda)$ für $\|(z, \lambda)\| \rightarrow +\infty$ zu analysieren. Es sei also $((z_n, \lambda_n))_{n \in \mathbb{N}}$ eine Folge in $\overline{M_x}$ mit

$$\lim_{n \rightarrow \infty} \|(z_n, \lambda_n)\| = +\infty.$$

Wegen der Bedingung

$$z_n \leq x \leq z_n + \lambda_n \quad \text{für alle } n \in \mathbb{N}$$

sind nur zwei Situationen denkbar:

- (a) $(z_n)_{n \in \mathbb{N}}$ beschränkt und $\lambda_n \xrightarrow{n \rightarrow \infty} +\infty$
- (b) $z_n \xrightarrow{n \rightarrow \infty} -\infty$ und $\lambda_n \geq x - z_n \xrightarrow{n \rightarrow \infty} +\infty$

Wir wollen nun zeigen, dass in beiden Situationen (a) und (b) jeweils

$$\lim_{n \rightarrow \infty} g_x(z_n, \lambda_n) = 0$$

und damit insbesondere

$$\limsup_{n \rightarrow \infty} g_x(z_n, \lambda_n) \leq S$$

gilt.

Beweis: zu (a):

Ist $y \in \mathbb{R}$, so gibt es wegen der Eigenschaften der Folgen $(z_n)_{n \in \mathbb{N}}$ und $(\lambda_n)_{n \in \mathbb{N}}$ ein $n_0 \in \mathbb{N}$, so dass für alle $n \geq n_0$ gilt:

$$y \leq z_n \quad \text{oder} \quad z_n \leq y \leq z_n + \lambda_n$$

Ist $y \leq z_n$, so ist $h_{z_n, \lambda_n}(y) = 1$. Für $z_n \leq y \leq z_n + \lambda_n$ gilt

$$h_{z_n, \lambda_n}(y) = 1 - \frac{z_n - y}{\lambda_n} \xrightarrow{n \rightarrow \infty} 1$$

1.4 Anhang: Beweis von Lemma 1.2.4

Somit erhalten wir für jedes $y \in \mathbb{R}$, dass

$$\lim_{n \rightarrow \infty} h_{z_n, \lambda_n}(y) = 1$$

gilt. Es ist φ eine integrierbare Majorante für die Folge von Funktionen $(h_{z_n, \lambda_n} \varphi)_{n \in \mathbb{N}}$, so dass nach dem Satz von der majorisierten Konvergenz folgt:

$$N(h_{z_n, \lambda_n}) = \int_{\mathbb{R}} h_{z_n, \lambda_n}(y) \varphi(y) dy \xrightarrow{n \rightarrow \infty} \int_{\mathbb{R}} \varphi(y) dy = 1$$

Indem wir nun wieder den Satz von der majorisierten Konvergenz anwenden auf die Folge $((h_{z_n, \lambda_n} - N(h_{z_n, \lambda_n}))\varphi)_{n \in \mathbb{N}}$ und wieder mit der gleichen Majorante φ , so ergibt sich

$$\begin{aligned} g_x(z_n, \lambda_n) &= f_{z_n, \lambda_n}(x) = \frac{1}{\varphi(x)} \int_{-\infty}^x (h_{z_n, \lambda_n}(y) - N(h_{z_n, \lambda_n})) \varphi(y) dy \\ &\xrightarrow{n \rightarrow \infty} \frac{1}{\varphi(x)} \int_{-\infty}^x (1 - 1) \varphi(y) dy \\ &= 0 \end{aligned}$$

zu (b): Es sei $y \in \mathbb{R}$. Dann gibt es ein $n_0 \in \mathbb{N}$ mit

$$z_n \leq y \leq z_n + \lambda_n$$

für alle $n \geq n_0$. Für $n \geq n_0$ ergibt sich:

$$\begin{aligned} h_{z_n, \lambda_n}(y) &= 1 - \frac{y - z_n}{\lambda_n} \leq 1 - \frac{y - z_n}{x - z_n} \\ &= 1 - \frac{\frac{y}{z_n} - 1}{\frac{x}{z_n} - 1} \xrightarrow{n \rightarrow \infty} 1 - \frac{0 - 1}{0 - 1} = 0, \end{aligned}$$

so dass

$$\lim_{n \rightarrow \infty} h_{z_n, \lambda_n}(y) = 0$$

gilt. Wie bei (a) erhält man hieraus durch majorisierte Konvergenz zunächst

$$\lim_{n \rightarrow \infty} N(h_{z_n, \lambda_n}) = 0$$

und dann auch

1 Die Steinsche Methode

$$\lim_{n \rightarrow \infty} g_x(z_n, \lambda_n) = \lim_{n \rightarrow \infty} \frac{1}{\varphi(x)} \int_{-\infty}^x (h_{z_n, \lambda_n}(y) - N(h_{z_n, \lambda_n})) \varphi(y) dy = \int_{-\infty}^x (0 - 0) \varphi(y) dy = 0.$$

Dies beendet den Beweis von Satz 1.4.4

□

Es muss somit nur noch gezeigt werden, dass $S \leq \sqrt{\frac{\pi}{8}}$ gilt!

Satz 1.4.5. *Definiere die Funktion $g : \mathbb{R} \rightarrow \mathbb{R}$ durch*

$$g(z) := \frac{\Phi(z)}{\varphi(z)}(1 - \Phi(z)).$$

Dann gilt:

$$S = \|g\|_{\infty} = |g(0)| = \sqrt{\frac{\pi}{8}}.$$

Beweis: Offenbar ist g auf ganz \mathbb{R} strikt positiv. Wegen

$$g(-z) = \frac{\Phi(-z)(1 - \Phi(-z))}{\varphi(-z)} = \frac{(1 - \Phi(z))\Phi(z)}{\varphi(z)} = g(z),$$

genügt es, $z > 0$ zu betrachten. Es gilt weiter

$$\lim_{z \rightarrow +\infty} g(z) = \lim_{z \rightarrow +\infty} H_2(z) \cdot \lim_{z \rightarrow +\infty} \Phi(z) = 0.$$

Für jedes $z \in \mathbb{R}$ gilt weiterhin

$$\begin{aligned} g(z) &\leq \sqrt{\frac{\pi}{8}} = g(0) \\ \Leftrightarrow \Phi(z)(1 - \Phi(z)) &\leq \sqrt{\frac{\pi}{8}} \varphi(z) \\ \Leftrightarrow 0 &\leq \sqrt{\frac{\pi}{8}} \varphi(z) - \Phi(z)(1 - \Phi(z)) =: h_1(z). \end{aligned}$$

Es ist

$$h_1(0) = 0 = \lim_{z \rightarrow +\infty} h_1(z)$$

und weiter

$$\begin{aligned} h_1'(z) &= -z\sqrt{\frac{\pi}{8}}\varphi(z) - \varphi(z) + 2\Phi(z)\varphi(z) \\ &= \varphi(z) \underbrace{\left(-1 - \sqrt{\frac{\pi}{8}}z + 2\Phi(z)\right)}_{=:h_2(z)} = \varphi(z)h_2(z) \end{aligned}$$

Weiter ist

$$\begin{aligned} h_2'(z) &= -\sqrt{\frac{\pi}{8}} + 2\varphi(z) \\ &= -\sqrt{\frac{\pi}{8}} + \frac{2}{\sqrt{2\pi}}e^{-\frac{z^2}{2}} \begin{cases} > 0 & \text{für } z < z_0 \\ = 0 & \text{für } z = z_0 \\ < 0 & \text{für } z > z_0 \end{cases}, \end{aligned}$$

wobei

$$z_0 := \sqrt{2 \log\left(\frac{4}{\pi}\right)},$$

wie man leicht nachrechnet. Somit ist $h_2|_{[0; z_0]}$ streng monoton wachsend und $h_2|_{[z_0; +\infty)}$ streng monoton fallend. Wegen $h_2(0) = 0$ und $\lim_{z \rightarrow +\infty} h_2(z) = -\infty$ muss es ein $z_1 > z_0$ geben mit $h_2(z) > 0$ für alle $0 < z < z_1$ und $h_2(z) < 0$ für alle $z > z_1$. Wegen $h_1' = \varphi \cdot h_2$ ist z_1 die einzige Nullstelle von h_1' und daher eine lokale Maximalstelle von h_1 . Wegen $h_1(0) = \lim_{z \rightarrow +\infty} h_1(z) = 0$ gilt somit $h_1(z) \geq 0$ für alle $z \in [0; +\infty)$, so dass $g(z) \leq \sqrt{\frac{\pi}{8}}$ für alle $z \in \mathbb{R}$ folgt. □

zu (b): Wir wollen

$$\|id_{\mathbb{R}} f_{z,\lambda}\|_{\infty} \leq 1$$

für alle $z \in \mathbb{R}$ und $\lambda > 0$ zeigen. Dies geht genauso wie der Beweis von Lemma 1.1.3. Wir werden zunächst (allgemeiner) zeigen, dass für beliebiges beschränktes und absolut stetiges $h : \mathbb{R} \rightarrow \mathbb{R}$ mit zugehöriger Stein-Lösung f_h gilt:

$$\|id_{\mathbb{R}} \cdot f_h\|_{\infty} \leq \|h - N(h)\|_{\infty}$$

Für f_h haben wir die beiden Darstellungen

$$f_h(x) = \frac{1}{\varphi(x)} \int_{-\infty}^x (h(y) - N(h))\varphi(y)dy = -\frac{1}{\varphi(x)} \int_x^{+\infty} (h(y) - N(h))\varphi(y)dy.$$

1 Die Steinsche Methode

Für $x > 0$ haben wir

$$\begin{aligned}
 |f_h(x)| &= \frac{1}{\varphi(x)} \left| \int_x^\infty (h(y) - N(h))\varphi(y)dy \right| \\
 &\leq \frac{1}{\varphi(x)} \int_x^\infty \|h - N(h)\|_\infty \varphi(y)dy \leq \frac{\|h - N(h)\|_\infty}{\varphi(x)} \int_x^\infty \varphi(y)dy \\
 &\leq \frac{\|h - N(h)\|_\infty}{\varphi(x)} \int_x^\infty \frac{y}{x} \varphi(y)dy = \frac{\|h - N(h)\|_\infty}{\varphi(x)x} \int_x^\infty y\varphi(y)dy \\
 &= \frac{\|h - N(h)\|_\infty}{\varphi(x)x} (-\varphi(y)|_x^\infty) = \frac{\|h - N(h)\|_\infty}{x} \\
 &= \frac{\|h - N(h)\|_\infty}{|x|}.
 \end{aligned}$$

Analog ergibt sich für $x < 0$

$$\begin{aligned}
 |f_h(x)| &= \frac{1}{\varphi(x)} \left| \int_{-\infty}^x (h(y) - N(h))\varphi(y)dy \right| \\
 &\leq \frac{1}{\varphi(x)} \int_{-\infty}^x \|h - N(h)\|_\infty \varphi(y)dy \leq \frac{\|h - N(h)\|_\infty}{\varphi(x)} \int_{-\infty}^x \varphi(y)dy \\
 &\leq \frac{\|h - N(h)\|_\infty}{\varphi(x)} \int_{-\infty}^x \frac{|y|}{|x|} \varphi(y)dy = \frac{\|h - N(h)\|_\infty}{\varphi(x)} \int_{-\infty}^x \frac{-y}{-x} \varphi(y)dy \\
 &= \frac{\|h - N(h)\|_\infty}{-x\varphi(x)} \int_{-\infty}^x (-y)\varphi(y)dy = \frac{\|h - N(h)\|_\infty}{|x|\varphi(x)} (\varphi(y)|_{-\infty}^x) \\
 &= \frac{\|h - N(h)\|_\infty}{|x|}.
 \end{aligned}$$

Wir haben für $x = 0$ die Ungleichung

$$\begin{aligned}
 |f_h(0)| &= \frac{1}{\varphi(0)} \left| \int_{-\infty}^0 (h(y) - N(h)) \varphi(y) dy \right| \leq \frac{1}{\varphi(0)} \int_{-\infty}^0 |h(y) - N(h)| \varphi(y) dy \\
 &\leq \sqrt{2\pi} \|h - N(h)\|_{\infty} \cdot \frac{1}{2} \\
 &= \sqrt{\frac{\pi}{2}} \|h - N(h)\|_{\infty}.
 \end{aligned}$$

Insgesamt haben wir somit gezeigt, dass für alle $x \in \mathbb{R}$ gilt:

$$|f_h(x)| \leq \|h - N(h)\|_{\infty} S(x),$$

wobei

$$S(x) := \begin{cases} \frac{1}{|x|} & \text{für } x \neq 0 \\ \sqrt{\frac{\pi}{2}} & \text{für } x = 0 \end{cases}$$

ist. Für $x \neq 0$ gilt somit

$$|x f_h(x)| \leq \|h - N(h)\|_{\infty}$$

und für $x = 0$ gilt

$$|x f_h(x)| = 0 \leq \|h - N(h)\|_{\infty}.$$

Insgesamt folgt

$$\|id_{\mathbb{R}} \cdot f_h\|_{\infty} \leq \|h - N(h)\|_{\infty}.$$

Ist nun $h = h_{z,\lambda}$, so ist

$$\|h - N(h)\|_{\infty} \leq 1,$$

woraus sofort die Behauptung folgt.

zu (c): Wir möchten $\|f'_{z,\lambda}\|_{\infty} \leq 2$ zeigen. Für jedes $x \in \mathbb{R}$ gilt nach der Stein-Gleichung:

$$f'_{z,\lambda}(x) = x f_{z,\lambda}(x) + h_{z,\lambda}(x) - N(h_{z,\lambda}),$$

so dass aus (b) und $\|h_{z,\lambda} - N(h_{z,\lambda})\|_{\infty} \leq 1$ folgt:

$$|f'_{z,\lambda}(x)| \leq \|id_{\mathbb{R}} f_{z,\lambda}\|_{\infty} + \|h_{z,\lambda} - N(h_{z,\lambda})\|_{\infty} \leq 1 + 1 = 2$$

□

1 Die Steinsche Methode

Bemerkung 1.4.6. (i) Es ist nicht schwierig zu sehen, dass die Konstante $\sqrt{\frac{\pi}{8}}$ in der Aussage (a) optimal ist. Dazu betrachte man für $z \in \mathbb{R}$ die Funktion h_z und die zugehörige Steinlösung f_z , welche gegeben ist durch

$$\begin{aligned} f_z(x) &= \frac{1}{\varphi(x)} \int_{-\infty}^x (h_z(y) - N(h_z)) \varphi(y) dy \\ &= \begin{cases} \frac{\Phi(x)}{\varphi(x)}(1 - \Phi(z)) & \text{für } x \leq z \\ \frac{1 - \Phi(x)}{\varphi(x)}(\Phi(z)) & \text{für } x \geq z \end{cases} = \begin{cases} H_1(x)(1 - \Phi(z)) & \text{für } x \leq z \\ H_2(x)(\Phi(z)) & \text{für } x \geq z, \end{cases} \end{aligned}$$

wobei H_1 und H_2 die Funktionen aus dem Beweis von Lemma 1.2.4 (a) sind. Aus den Eigenschaften dieser Funktionen folgt, dass die Funktion f_z ihr globales Maximum im Punkt z annimmt mit

$$f_z(z) = \frac{\Phi(z)}{\varphi(z)}(1 - \Phi(z)) = g(z).$$

Wir haben weiter oben gesehen, dass

$$\sup_{z \in \mathbb{R}} g(z) = g(0) = \sqrt{\frac{\pi}{8}}$$

ist. Für jedes $z \in \mathbb{R}$ gilt ferner:

$$\begin{aligned} \lim_{\lambda \searrow 0} N(h_{z,\lambda}) &= \lim_{\lambda \searrow 0} \Phi(z + \lambda) + \frac{z}{\lambda}(\Phi(z + \lambda) - \Phi(z)) + \frac{1}{\lambda}(\varphi(z + \lambda) - \varphi(z)) \\ &= \Phi(z) + z\Phi'(z) + \varphi'(z) = \Phi(z) + z\varphi(z) - z\varphi(z) \\ &= \Phi(z) = N(h_z) \end{aligned}$$

Sei $x^* = x^*(z, \lambda)$ die Maximalstelle der Funktion $f_{z,\lambda}$. Wir wissen, dass dann $x^* \in [z; z + \lambda]$ gilt. Es gibt somit ein $t_\lambda \in [0; 1]$ mit

$$x^* = x^*(z, \lambda) = z + t_\lambda \lambda.$$

Außerdem wissen wir, dass

$$f_{z,\lambda}(x^*) = \frac{\Phi(x^*)}{\varphi(x^*)}(1 - N(h_{z,\lambda})) + \frac{1}{\lambda\varphi(x^*)} (z(\Phi(x^*) - \Phi(z)) + \varphi(x^*) - \varphi(z))$$

1.4 Anhang: Beweis von Lemma 1.2.4

gilt. Nach dem Mittelwertsatz der Differentialrechnung existieren ferner $\xi_\lambda, \eta_\lambda \in [0; 1]$ mit

$$\begin{aligned}\Phi(x^*) - \Phi(z) &= \Phi(z + t_\lambda \lambda) - \Phi(z) \\ &= t_\lambda \lambda \Phi'(z + \xi_\lambda t_\lambda \lambda) = t_\lambda \lambda \varphi(z + \xi_\lambda t_\lambda \lambda)\end{aligned}$$

und

$$\begin{aligned}\varphi(x^*) - \varphi(z) &= \varphi(z + t_\lambda \lambda) - \varphi(z) \\ &= t_\lambda \lambda \varphi'(z + \eta_\lambda t_\lambda \lambda) = -t_\lambda \lambda (z + \eta_\lambda t_\lambda \lambda) \varphi(z + \eta_\lambda t_\lambda \lambda)\end{aligned}$$

Für die beiden letzten Summanden in $f_{z,\lambda}$ ergibt sich somit, wenn man noch die Stetigkeit der Funktion φ sowie die Tatsache, dass

$$\lim_{\lambda \searrow 0} x^*(z, \lambda) = z$$

gilt, ausnutzt:

$$\begin{aligned}& \lim_{\lambda \searrow 0} \frac{1}{\lambda \varphi(x^*)} (z(\Phi(x^*) - \Phi(z)) + \varphi(x^*) - \varphi(z)) \\ &= \frac{1}{\varphi(z)} \lim_{\lambda \searrow 0} t_\lambda (z\varphi(z + \xi_\lambda t_\lambda \lambda) - (z + \eta_\lambda t_\lambda \lambda)\varphi(z + \eta_\lambda t_\lambda \lambda)) \\ &= 0,\end{aligned}$$

denn die Zahlen t_λ, ξ_λ und η_λ sind allesamt im Intervall $[0; 1]$ eingesperrt. Für den ersten Summanden erhalten wir nach obigem Resultat für den Grenzwert von $N(h_{z,\lambda})$ wieder aus Stetigkeitsgründen

$$\lim_{\lambda \searrow 0} \frac{\Phi(x^*)}{\varphi(x^*)} (1 - N(h_{z,\lambda})) = \frac{\Phi(z)}{\varphi(z)} (1 - N(h_z)) = \frac{\Phi(z)}{\varphi(z)} (1 - \Phi(z)),$$

so dass

$$\lim_{\lambda \searrow 0} f_{z,\lambda}(x^*) = \frac{\Phi(z)}{\varphi(z)} (1 - \Phi(z)) = f_z(z)$$

gezeigt ist. Für $z = 0$ ergibt sich somit insbesondere

1 Die Steinsche Methode

$$\lim_{\lambda \searrow 0} f_{0,\lambda}(x^*) = f_0(0) = \frac{\Phi(0)}{\varphi(0)}(1 - \Phi(0)) = \sqrt{\frac{\pi}{8}}$$

und daraus die Optimalität dieser Schranke.

(ii) Es lässt sich zeigen (vgl. [BC05]), dass für die Familie $(f_z)_{z \in \mathbb{R}}$ das Supremum der Ableitungen gegeben ist durch

$$\sup_{z \in \mathbb{R}} \|f'_z\|_\infty = 1.$$

Dies lässt vermuten, dass in Analogie zur Aussage (a) von Lemma 1.2.4 auch

$$\sup_{z \in \mathbb{R}} \|f'_{z,\lambda}\|_\infty = 1$$

gilt. Unter Verwendung derselben Methoden wie für die Familie der Funktionen selber, habe ich dies leider (noch) nicht zeigen können.

1.5 Anhang: Beweis von Satz 1.2.5

Beweis: Wir definieren zunächst für nicht-negatives, messbares $f : (\mathcal{X} \times \mathcal{Y}, \mathcal{A}_{\mathcal{X}} \otimes \mathcal{A}_{\mathcal{Y}}) \rightarrow (\overline{\mathbb{R}}, \overline{\mathcal{B}})$ die Abbildung $g_f : (\mathcal{X}, \mathcal{A}_{\mathcal{X}}) \rightarrow (\overline{\mathbb{R}}, \overline{\mathcal{B}})$ durch $g_f(x) := E[f(x, Y)]$. Da die Abbildung $Y \ni y \mapsto f(x, y) \in \overline{\mathbb{R}}$ für jedes feste $x \in \mathcal{X}$ $\mathcal{A}_{\mathcal{Y}} - \overline{\mathcal{B}}$ -messbar ist, ist g_f wohldefiniert für jedes nicht-negative, messbare f . Wir zeigen durch algebraische Induktion, dass g_f $\mathcal{A}_{\mathcal{X}} - \overline{\mathcal{B}}$ -messbar und eine Version von $x \mapsto E[f(X, Y)|X = x]$ ist.

1.Schritt: Sei $f = 1_{A \times B}$ mit $A \in \mathcal{A}_{\mathcal{X}}$ und $B \in \mathcal{A}_{\mathcal{Y}}$. Dann gilt für alle $x \in \mathcal{X}$:

$$g_f(x) = E[1_A(x)1_B(Y)] = 1_A(x) \cdot P(Y \in B)$$

Somit ist g_f $\mathcal{A}_{\mathcal{X}} - \overline{\mathcal{B}}$ -messbar. Ist ferner $\{X \in F\} = X^{-1}(F) \in \sigma(X)$ mit $F \in \mathcal{A}_{\mathcal{X}}$, so folgt

$$\begin{aligned} \int_{\{X \in F\}} g_{1_{A \times B}}(X) dP &= \int_{\{X \in F\}} 1_A(X) P(Y \in B) dP \\ &= P(Y \in B) \cdot P(X \in F \cap A) \stackrel{\text{unabh.}}{=} P(X \in F \cap A, Y \in B) \\ &= P((X, Y) \in ((F \cap A) \times B)) = \int_{\Omega} 1_{((F \cap A) \times B)}(X, Y) dP \\ &= \int_{\Omega} 1_{F \cap A}(X) \cdot 1_B(Y) dP = \int_{\Omega} 1_F(X) 1_A(X) 1_B(Y) dP \\ &= \int_{\{X \in F\}} 1_A(X) 1_B(Y) dP = \int_{\{X \in F\}} f(X, Y) dP \end{aligned}$$

Insgesamt ist g_f daher eine Version von $x \mapsto E[f(X, Y)|X = x]$.

2.Schritt: Sei

$$\mathcal{D} := \{C \in \mathcal{A}_{\mathcal{X}} \otimes \mathcal{A}_{\mathcal{Y}} : g_{1_C} \text{ erfüllt die gewünschte Eigenschaft}\}$$

Wir behaupten, dass \mathcal{D} ein Dynkin-System ist:

- Nach dem ersten Schritt gilt $\mathcal{X} \times \mathcal{Y} \in \mathcal{D}$.
- Es sei $C \in \mathcal{D}$. Dann gilt für alle $x \in \mathcal{X}$

$$g_{1_{C^c}}(x) = E[1_{C^c}(x, Y)] = 1 - E[1_C(x, Y)] = 1 - g_{1_C}(x)$$

1 Die Steinsche Methode

und somit ist auch $g_{1_{C^c}}$ messbar. Ist wieder $\{X \in F\} \in \sigma(X)$, so folgt

$$\begin{aligned} \int_{\{X \in F\}} g_{1_{C^c}}(X) dP &= \int_{\{X \in F\}} (1 - g_{1_C}(X)) dP = P(X \in F) - \int_{\{X \in F\}} g_{1_C}(X) dP \\ &\stackrel{C \in \mathcal{D}}{=} P(X \in F) - \int_{\{X \in F\}} 1_C(X, Y) dP = P(X \in F) - \int_{\{X \in F\}} (1 - 1_{C^c}(X, Y)) dP \\ &= P(X \in F) - P(X \in F) + \int_{\{X \in F\}} 1_{C^c}(X, Y) dP = \int_{\{X \in F\}} 1_{C^c}(X, Y) dP \end{aligned}$$

Somit gilt auch $C^c \in \mathcal{D}$.

- Sei $(C_n)_{n \in \mathbb{N}}$ eine Folge paarweise disjunkter Mengen aus \mathcal{D} und sei $C := \bigcup_{n \in \mathbb{N}} C_n$. Dann gilt für alle $x \in \mathcal{X}$

$$\begin{aligned} g_{1_C}(x) &= E[1_C(x, Y)] = E[1_{\bigcup_{n \in \mathbb{N}} C_n}(x, Y)] \\ &= E\left[\sum_{n=1}^{\infty} 1_{C_n}(x, Y)\right] \\ &\stackrel{\text{mon. Konv.}}{=} \sum_{n=1}^{\infty} E[1_{C_n}(x, Y)] \\ &= \sum_{n=1}^{\infty} g_{1_{C_n}}(x), \end{aligned}$$

so dass g_{1_C} als Reihe messbarer Funktionen wieder messbar ist. Weiter gilt für alle $\{X \in F\} \in \sigma(X)$

$$\begin{aligned}
 \int_{\{X \in F\}} g_{1_C}(X) dP &= \int_{\Omega} \sum_{n=1}^{\infty} 1_F(X) g_{1_{C_n}}(X) dP \stackrel{\text{mon. Konv.}}{=} \sum_{n=1}^{\infty} \int_{\Omega} 1_F(X) g_{1_{C_n}}(X) dP \\
 &= \sum_{n=1}^{\infty} \int_{\{X \in F\}} g_{1_{C_n}}(X) dP \stackrel{\forall n: C_n \in \mathcal{D}}{=} \sum_{n=1}^{\infty} \int_{\{X \in F\}} 1_{C_n}(X, Y) dP \\
 &= \sum_{n=1}^{\infty} \int_{\Omega} 1_F(X) 1_{C_n}(X, Y) dP \stackrel{\text{mon. Konv.}}{=} \int_{\Omega} 1_F(X) \sum_{n=1}^{\infty} 1_{C_n}(X, Y) dP \\
 &= \int_{\{X \in F\}} \sum_{n=1}^{\infty} 1_{C_n}(X, Y) dP = \int_{\{X \in F\}} 1_C(X, Y) dP
 \end{aligned}$$

Somit gilt auch $C = \bigcup_{n \in \mathbb{N}} C_n \in \mathcal{D}$ und \mathcal{D} ist ein Dynkin-System, welches den durchschnittstabilen Erzeuger $\mathcal{R} := \{A \times B : A \in \mathcal{A}_X, B \in \mathcal{A}_Y\}$ von $\mathcal{A}_X \otimes \mathcal{A}_Y$ enthält und somit folgt $D = \mathcal{A}_X \otimes \mathcal{A}_Y$.

3.Schritt: Sei f eine nicht-negative Treppenfunktion, d.h. es gibt ein $s \in \mathbb{N}$, Mengen $C_1, \dots, C_s \in \mathcal{A}_X \otimes \mathcal{A}_Y$ und $\alpha_1, \dots, \alpha_s \geq 0$ mit $f = \sum_{j=1}^s \alpha_j 1_{C_j}$. Dann gilt für alle $x \in \mathcal{X}$

$$\begin{aligned}
 g_f(x) &= E[f(x, Y)] = E\left[\sum_{j=1}^s \alpha_j 1_{C_j}(x, Y)\right] \\
 &= \sum_{j=1}^s \alpha_j E[1_{C_j}(x, Y)] = \sum_{j=1}^s \alpha_j g_{1_{C_j}}(x)
 \end{aligned}$$

D.h. g_f ist als Linearkombination messbarer Funktionen messbar. Ferner gilt wegen der Linearität der bedingten Erwartung

$$\begin{aligned}
 E[f(X, Y)|X] &= E\left[\sum_{j=1}^s \alpha_j 1_{C_j}(X, Y)|X\right] = \sum_{j=1}^s \alpha_j E[1_{C_j}(X, Y)|X] \\
 &\stackrel{\text{2.Schritt}}{=} \sum_{j=1}^s \alpha_j g_{1_{C_j}}(X) = g_f(X),
 \end{aligned}$$

so dass g_f als Version von $x \mapsto E[f(X, Y)|X = x]$ erkannt ist.

1 Die Steinsche Methode

4.Schritt: Sei f messbar und nicht-negativ. Dann gibt es eine monoton wachsende Folge $(f_n)_{n \in \mathbb{N}}$ nicht-negativer Treppenfunktionen, die punktweise gegen f konvergiert. Dann gilt aber auch für alle $x \in \mathcal{X}$, dass $f_n(x, \cdot) \nearrow f(x, \cdot)$, so dass aus dem Satz von der monotonen Konvergenz folgt:

$$g_{f_n}(x) = E[f_n(x, Y)] \nearrow E[f(x, Y)] = g_f(x)$$

Somit ist g_f als punktweiser Limes messbarer Funktionen wieder messbar. Weiter gilt für beliebiges $\{X \in F\} \in \sigma(X)$

$$\begin{aligned} \int_{\{X \in F\}} g_f(X) dP &= \int_{\Omega} 1_F(X) g_f(X) dP \\ \stackrel{\text{mon. Konv.}}{=} \lim_{n \rightarrow \infty} \int_{\Omega} 1_F(X) g_{f_n}(X) dP &= \lim_{n \rightarrow \infty} \int_{\{X \in F\}} g_{f_n}(X) dP \\ \stackrel{\text{3.Schritt}}{=} \lim_{n \rightarrow \infty} \int_{\{X \in F\}} f_n(X, Y) dP &= \lim_{n \rightarrow \infty} \int_{\Omega} 1_F(X) f_n(X, Y) dP \\ &= \int_{\Omega} 1_F(X) f(X, Y) dP = \int_{\{X \in F\}} f(X, Y) dP \end{aligned}$$

Daher ist g_f eine Version von $x \mapsto E[f(X, Y) | X = x]$.

Sei jetzt f so wie im Text des Satzes. Nach Voraussetzung gilt:

$$\begin{aligned} +\infty &> E[|f(X, Y)|] &= \int_{\Omega} |f(X, Y)| dP \\ &= \int_{\mathcal{X} \times \mathcal{Y}} |f(x, y)| dP^{(X, Y)}(x, y) \\ &\stackrel{\text{unabh.}}{=} \int_{\mathcal{X} \times \mathcal{Y}} |f(x, y)| d(P^X \otimes P^Y)(x, y) \\ &\stackrel{\text{Fubini}}{=} \int_{\mathcal{X}} \left(\int_{\mathcal{Y}} |f(x, y)| dP^Y(y) \right) dP^X(x), \end{aligned}$$

d.h. es ist $\int_{\mathcal{Y}} |f(x, y)| dP^Y(y) < +\infty$ für P^X - f.a. $x \in \mathcal{X}$. Wegen

$$\int_{\mathcal{Y}} |f(x, y)| dP^Y(y) = \int_{\Omega} |f(x, Y)| dP = E[|f(x, Y)|] = g_{|f|}(x)$$

ist die Abbildung $x \mapsto \int_{\mathcal{Y}} |f(x, y)| dP^Y(y) = g_{|f|}(x)$ nach dem oben Gezeigten $\mathcal{A}_{\mathcal{X}} - \bar{\mathcal{B}}$ -messbar und somit ist

$$G = \{g_{|f|} < +\infty\} \in \mathcal{A}_{\mathcal{X}}$$

mit $P^X(G) = P(X \in G) = 1$.

Da $G, G^c \in \mathcal{A}_{\mathcal{X}}$ sind und die Nullfunktion messbar ist, muss für die Messbarkeit von h_f nach einem Resultat aus der Maßtheorie nur noch gezeigt werden, dass die Einschränkung $h_f|_G$ messbar bezüglich der Spur- σ -Algebra $\mathcal{A}_{\mathcal{X}}|_G$ ist. Für $x \in G$ gilt aber

$$\begin{aligned} h_f(x) &= E[f(x, Y)] = E[(f(x, Y))^+] - E[(f(x, Y))^-] \\ &= E[f^+(x, Y)] - E[f^-(x, Y)] = g_{f^+}(x) - g_{f^-}(x), \end{aligned}$$

so dass $h_f|_G = g_{f^+}|_G - g_{f^-}|_G$ als Differenz zweier messbarer Funktionen wieder $\mathcal{A}_{\mathcal{X}}|_G - \mathcal{B}$ -messbar ist, denn die Einschränkungen $g_{f^+}|_G$ und $g_{f^-}|_G$ sind messbar bezüglich $\mathcal{A}_{\mathcal{X}}|_G$.

Sei schließlich wieder $\{X \in F\} \in \sigma(X)$ beliebig. Dann gilt wegen $P(X \in G) = 1$:

$$\begin{aligned} \int_{\{X \in F\}} h_f(X) dP &= \int_{\{X \in F \cap G\}} h_f(X) dP - \int_{\{X \in F \cap G^c\}} h_f(X) dP \\ &= \int_{\{X \in F \cap G\}} g_{f^+}(X) dP - \int_{\{X \in F \cap G\}} g_{f^-}(X) dP \\ &= \int_{\{X \in F\}} g_{f^+}(X) dP - \int_{\{X \in F\}} g_{f^-}(X) dP \\ &\stackrel{4.\text{Schritt}}{=} \int_{\{X \in F\}} f^+(X, Y) dP - \int_{\{X \in F\}} f^-(X, Y) dP \\ &= \int_{\{X \in F\}} f(X, Y) dP, \end{aligned}$$

d.h. es ist $h_f(X)$ eine Version von $E[f(X, Y)|X]$ und somit h_f eine Version von $x \mapsto E[f(X, Y)|X = x]$. Damit ist der Satz bewiesen.

□

1.6 Anhang: reelle Funktionen

In diesem Abschnitt wollen wir an die Definitionen bestimmter Klassen reeller Funktionen erinnern, einige Resultate vorstellen und Zusammenhänge zwischen diesen Klassen beleuchten. Es sei $I \subseteq \mathbb{R}$ ein Intervall, das mehr als einen Punkt enthält.

Definition 1.6.1. Eine Funktion $f : I \rightarrow \mathbb{R}$ heißt

- (i) stückweise stetig, wenn es zu jedem kompakten Teilintervall $J = [a; b] \subseteq I$ Punkte $x_0, x_1, \dots, x_m \in J$ gibt mit $a = x_0 < x_1 < \dots < x_m = b$ und so, dass alle eingeschränkten Funktionen $f|_{(x_{i-1}; x_i)}$, $i \in \{1, \dots, m\}$, stetig sind.
- (ii) stückweise stetig differenzierbar, wenn es zu jedem kompakten Teilintervall $J = [a; b] \subseteq I$ Punkte $x_0, x_1, \dots, x_m \in J$ gibt mit $a = x_0 < x_1 < \dots < x_m = b$ und so, dass alle eingeschränkten Funktionen $f|_{[x_{i-1}; x_i]}$, $i \in \{1, \dots, m\}$, stetig differenzierbar sind.
- (iii) absolut stetig, wenn es für jedes kompakte Teilintervall $J = [a; b] \subseteq I$ und zu jedem $\varepsilon > 0$ ein $\delta > 0$ gibt, so dass folgendes gilt:
Sind $n \in \mathbb{N}$ und $x_0, x_1, \dots, x_n \in J$ mit $x_0 < x_1 < \dots < x_n$ und $\sum_{i=1}^n (x_i - x_{i-1}) < \delta$, so ist

$$\sum_{i=1}^n |f(x_i) - f(x_{i-1})| < \varepsilon.$$

- (iv) Lipschitz-stetig, wenn es ein $L \in [0; +\infty)$ gibt mit

$$|f(x) - f(y)| \leq L|x - y| \text{ für alle } x, y \in I.$$

Man nennt L dann eine Lipschitz-Konstante zu f .

Bemerkung 1.6.2. (i) Nach unserer Definition ist eine stückweise stetig differenzierbare Funktion automatisch auch stetig, da wir gefordert haben, dass die auf das **abgeschlossene** Intervall $[x_i; x_{i+1}]$ eingeschränkte Funktion stetig differenzierbar ist.

- (ii) Ist $f : I \rightarrow \mathbb{R}$ Lipschitz-stetig, so heißt

$$L_f := \inf\{L \in [0; +\infty) \mid L \text{ Lipschitz-Konstante für } f\}$$

die minimale Lipschitz-Konstante zu f . Es ist L_f stets auch eine Lipschitz-Konstante zu f . In der Tat, ist $\varepsilon > 0$, so ist $L_\varepsilon := L_f + \varepsilon$ eine Lipschitz-Konstante zu f . Somit gilt für alle $x, y \in I$:

$$\begin{aligned} |f(x) - f(y)| &\leq L_\varepsilon |x - y| = (L_f + \varepsilon) |x - y| \\ &= L_f |x - y| + \varepsilon |x - y| \xrightarrow{\varepsilon \searrow 0} L_f |x - y|, \end{aligned}$$

also

$$|f(x) - f(y)| \leq L_f |x - y|.$$

(iii) Ist $f : I \rightarrow \mathbb{R}$ Lipschitz-stetig mit Konstante L , so ist f auch absolut stetig. denn: Sei $\varepsilon > 0$. Wähle $\delta := \frac{\varepsilon}{L+1}$. Sind dann $x_0 < x_1 < \dots < x_n$ Punkte aus I mit $\sum_{i=1}^n (x_i - x_{i-1}) < \delta$, so folgt

$$\begin{aligned} \sum_{i=1}^n |f(x_i) - f(x_{i-1})| &\leq \sum_{i=1}^n L |x_i - x_{i-1}| = L \sum_{i=1}^n (x_i - x_{i-1}) \\ &< L\delta = L \frac{\varepsilon}{L+1} < \varepsilon. \end{aligned}$$

□

Die Klasse der absolut stetigen Funktionen ist deshalb so bedeutsam, weil sie genau aus den unbestimmten Integralen der Lebesgue-integrierbaren Funktionen besteht. Dies ist eine Teilaussage des Hauptsatzes der Differential- und Integralrechnung für das Lebesgue-Integral (vgl. Satz 1.6.4). Es gilt zunächst:

Satz 1.6.3. Jede absolut stetige Funktion $F : I \rightarrow \mathbb{R}$ ist λ -f.ü. differenzierbar. Setzt man $F'(x) := 0$, wenn F in x nicht differenzierbar ist, so ist die Abbildung $F' : I \rightarrow \mathbb{R}$ Lebesgue-messbar.

Wir formulieren nun den Hauptsatz der Differential- und Integralrechnung für das Lebesgue-Integral.

Satz 1.6.4 (HDI für das Lebesgue-Integral). Seien $a < b$ reelle Zahlen und $I := [a; b]$. Dann gilt

(i) Ist $f : I \rightarrow \mathbb{R}$ Lebesgue-integrierbar und definieren wir $F : I \rightarrow \mathbb{R}$ durch

$$F(x) := \int_a^x f(t) dt,$$

so ist F absolut stetig und es gilt $F'(x) = f(x)$ für λ -fast alle $x \in I$.

1 Die Steinsche Methode

(ii) Ist $F : I \rightarrow \mathbb{R}$ absolut stetig, so ist $F' : I \rightarrow \mathbb{R}$ Lebesgue-integrierbar und es gilt für alle $x \in I$

$$F(x) = F(a) + \int_a^x F'(t) dt .$$

Für einen Beweis dieses wichtigen Satzes sei auf die Literatur verwiesen, vgl. etwa [Els05] oder [HS75].

Wir verwenden die Aussage (ii) von Satz 1.6.4 in dieser Arbeit des Öfteren stillschweigend, indem wir für eine absolut stetige Funktion $h : \mathbb{R} \rightarrow \mathbb{R}$ eine Differenz als Integral über die Ableitung schreiben, also

$$h(x) - h(y) = \int_x^y h'(t) dt .$$

Es sei darauf hingewiesen, dass dies für stetige Funktionen h , deren Ableitung fast überall existiert und Lebesgue-integrierbar ist, nicht gelten muss. Als Gegenbeispiel lässt sich etwa die stetige Cantor-Funktion anführen, die fast überall differenzierbar ist mit Ableitung 0. Daraus erhält man auch Beispiele streng monoton wachsender stetiger Funktionen, deren Ableitungen fast überall existieren und integrierbar sind, aber obige Eigenschaft nicht besitzen. Für Details und Beweise dieser Tatsachen sei wieder auf [Els05] verwiesen.

Wir wollen nun noch zeigen, dass für Lipschitz-stetige Funktionen h die minimale Lipschitzkonstante L_h mit der Supremumsnorm der Ableitung h' übereinstimmt.

Proposition 1.6.5. *Sei $h : I \rightarrow \mathbb{R}$ Lipschitz-stetig mit minimaler Lipschitz-Konstante L_h . Dann ist h λ -f.ü. differenzierbar und es gilt*

$$L_h = \|h'\|_\infty .$$

Beweis: Nach 1.6.2 (iii) ist h absolut stetig, also nach Satz 1.6.3 f.ü. differenzierbar. Wir setzen wieder $h'(x) := 0$ für alle $x \in I$, in denen h nicht differenzierbar ist. Dann gilt nach Satz 1.6.4 (ii) für alle $x, y \in I$:

$$|h(x) - h(y)| = \left| \int_y^x h'(t) dt \right| \leq \int_y^x |h'(t)| dt \leq \|h'\|_\infty |x - y| ,$$

also

$$L_h \leq \|h'\|_\infty .$$

Für die andere Ungleichung wähle man zu beliebigem $\varepsilon > 0$ ein $x_\varepsilon \in I$ mit

$$|h'(x_\varepsilon)| \geq \|h'\|_\infty - \varepsilon.$$

Sei weiter $0 \neq t_\varepsilon$ so gewählt, dass $x_\varepsilon + t_\varepsilon \in I$ ist und, dass

$$\left| h'(x_\varepsilon) - \frac{h(x_\varepsilon + t_\varepsilon) - h(x_\varepsilon)}{t_\varepsilon} \right| < \varepsilon$$

ist. Dann gilt

$$\begin{aligned} |h'(x_\varepsilon)| - \varepsilon &< \left| \frac{h(x_\varepsilon + t_\varepsilon) - h(x_\varepsilon)}{t_\varepsilon} \right| \\ &\leq L_h \frac{|t_\varepsilon|}{|t_\varepsilon|} = L_h, \end{aligned}$$

also insgesamt:

$$\|h'\|_\infty \leq |h'(x_\varepsilon)| + \varepsilon < L_h + 2\varepsilon.$$

Da ε beliebig klein gewählt werden kann, folgt auch $\|h'\|_\infty \leq L_h$. □

Nun behandeln wir noch als Beispiel die Klasse der Funktionen $h_{z,\lambda}$:

Beispiel 1.6.6. Für $\lambda > 0$ und $z \in \mathbb{R}$ sei die Funktion $h_{z,\lambda} : \mathbb{R} \rightarrow \mathbb{R}$ gegeben durch

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

Dann ist $h_{z,\lambda}$ Lipschitz-stetig mit minimaler Lipschitz-Konstante $L_h = \|h'_{z,\lambda}\|_\infty = \frac{1}{\lambda}$.

Abschließend wollen wir noch zeigen, dass eine stückweise stetig differenzierbare Funktion f immer auch absolut stetig ist.

Satz 1.6.7. Es sei $f : I \rightarrow \mathbb{R}$ eine stückweise stetig differenzierbare Funktion. Dann ist f auch absolut stetig.

Beweis: Sei $J = [a; b] \subseteq I$ ein kompaktes Teilintervall von I . Nach Voraussetzung existieren Punkte

$$a = x_0 < x_1 < \dots < x_m = b,$$

1 Die Steinsche Methode

so dass für jedes $i \in \{1, \dots, m\}$ die Einschränkung

$$g_i := f|_{[x_{i-1}; x_i]}$$

stetig differenzierbar ist. Somit ist

$$M := \max_{i=1, \dots, m} \|g'_i\|_\infty < +\infty.$$

Wähle nun zu vorgegebenem $\varepsilon > 0$

$$\delta := \frac{\varepsilon}{M+1}.$$

Seien dann

$$a \leq t_0 < t_1 < \dots < t_n \leq b$$

mit

$$\sum_{j=1}^n (t_j - t_{j-1}) < \delta.$$

Dann gibt es für jedes $j \in \{1, \dots, n\}$ genau ein $i_j \in \{1, \dots, m\}$ mit

$$t_j \in [x_{i_j-1}; x_{i_j}).$$

Ist $i_j = i_{j-1}$, so folgt aus dem Mittelwertsatz die Abschätzung

$$|f(t_j) - f(t_{j-1})| = |g_{i_j}(t_j) - g_{i_j}(t_{j-1})| \leq \|g'_{i_j}\|_\infty (t_j - t_{j-1}) \leq M(t_j - t_{j-1}) \quad (1.35)$$

Ist andererseits $i_j > i_{j-1}$, so folgt:

$$\begin{aligned} |f(t_j) - f(t_{j-1})| &= |f(t_j) - f(x_{i_{j-1}}) + \sum_{l=i_{j-1}+1}^{i_j-1} (f(x_l) - f(x_{l-1})) \\ &\quad + f(x_{i_{j-1}}) - f(t_{j-1})| \\ &\leq |f(t_j) - f(x_{i_{j-1}})| + \sum_{l=i_{j-1}+1}^{i_j-1} |f(x_l) - f(x_{l-1})| \\ &\quad + |f(x_{i_{j-1}}) - f(t_{j-1})| \\ &= |g_{i_j}(t_j) - g_{i_j}(x_{i_{j-1}})| + \sum_{l=i_{j-1}+1}^{i_j-1} |g_l(x_l) - g_l(x_{l-1})| \\ &\quad + |g_{i_{j-1}}(x_{i_{j-1}}) - g_{i_{j-1}}(t_{j-1})| \end{aligned} \quad (1.36)$$

Wieder mit dem Mittelwertsatz kann man nun (1.36) wie folgt weiter nach oben abschätzen:

$$\begin{aligned}
 |f(t_j) - f(t_{j-1})| &\leq \|g'_{i_j}\|_\infty(t_j - x_{i_{j-1}}) + \sum_{l=i_{j-1}+1}^{i_j-1} \|g'_l\|_\infty(x_l - x_{l-1}) \\
 &\quad + \|g'_{i_{j-1}}\|_\infty(x_{i_{j-1}} - t_{j-1}) \\
 &\leq M \left((t_j - x_{i_{j-1}}) + \sum_{l=i_{j-1}+1}^{i_j-1} (x_l - x_{l-1}) + (x_{i_{j-1}} - t_{j-1}) \right) \\
 &= M(t_j - t_{j-1}) \tag{1.37}
 \end{aligned}$$

Aus (1.35) und (1.36) ergibt sich nun, dass

$$\begin{aligned}
 \sum_{j=1}^n |f(t_j) - f(t_{j-1})| &\leq \sum_{j=1}^n M(t_j - t_{j-1}) = M \sum_{j=1}^n (t_j - t_{j-1}) \\
 &< M\delta = \frac{M}{M+1} \varepsilon < \varepsilon
 \end{aligned}$$

ist.

□

1 Die Steinsche Methode

2 Darstellungstheorie und der random walk auf endlichen Gruppen

In diesem Kapitel geben wir eine umfassende Einführung in die Theorie der *Darstellungen* endlicher Gruppen. In diesem klassischen Teilgebiet der Mathematik versucht man eine Gruppe G dadurch besser zu verstehen, dass man sich die *Aktionen* dieser Gruppe auf (endlich-dimensionalen) Vektorräumen ansieht. Wir betrachten hier nur den Fall, dass der dem Vektorraum zugrunde liegende Körper der Körper \mathbb{C} der komplexen Zahlen, also ein algebraisch abgeschlossener Körper der Charakteristik 0, ist. Diese Einschränkung ist jedoch für viele der vorgestellten Resultate nicht notwendig. Auch die geforderte Endlichkeit der Gruppe G wird für manche Sätze nicht benötigt. Wir werden jedoch nicht darauf aufmerksam machen, welche Resultate auch unter schwächeren Bedingungen gelten, sondern verweisen auf die Literatur, wie zum Beispiel [Ser77] für die allgemeine Darstellungstheorie endlicher Gruppen bzw. [Sag01] oder [Jam78] für die Darstellungstheorie der symmetrischen Gruppe. Nachdem wir in Abschnitt 2.1 die allgemeine Darstellungstheorie endlicher Gruppen relativ ausführlich entwickelt haben, betrachten wir in Abschnitt 2.2 den uns am meisten interessierenden Fall der symmetrischen Gruppe S_n . Die Motivation, sich im Rahmen der Stochastik mit Darstellungstheorie (insbesondere der symmetrischen Gruppe) zu beschäftigen, geben wir erst in Abschnitt 2.3, wo wir uns ein konkretes Modell für das Mischen von Kartenspielen ansehen werden, welches uns zu einem sogenannten *random walk*, einer *Markov-Kette* auf der S_n , führen wird. Die Analyse dieses Modells, welche auf Persi Diaconis und Mehrdad Shahshahani (vgl. [DS81] und [Dia88]) zurückgeht, verwendet schließlich Begriffe und Resultate aus der Darstellungstheorie der symmetrischen Gruppe. Desweiteren wird der titelgebende Begriff des *character ratios* vorgestellt und seine Bedeutung für die Konvergenzgeschwindigkeit des random walk auf S_n gegen die Gleichverteilung erläutert.

2.1 Darstellungstheorie endlicher Gruppen

Falls nicht ausdrücklich anders erwähnt, bezeichne im Folgenden stets G eine endliche Gruppe der Ordnung $n := |G| \in \mathbb{N}$ und alle auftretenden Vektorräume seien endlich-dimensional über dem Körper \mathbb{C} der komplexen Zahlen.

2.1.1 Definitionen, erste Eigenschaften und Beispiele

Definition 2.1.1. (a) Es sei V ein endlich-dimensionaler \mathbb{C} -Vektorraum. Eine Darstellung von G in V ist ein Gruppenhomomorphismus

$$\rho : G \rightarrow GL(V)$$

Dabei ist

$$GL(V) := \{f : V \rightarrow V : f \text{ ist } \mathbb{C}\text{-linear und bijektiv}\}$$

die Gruppe aller \mathbb{C} -Vektorraum-Automorphismen auf V .

Die Zahl $d_\rho := \dim_{\mathbb{C}}(V)$ nennt man die Dimension oder auch den Grad der Darstellung ρ . Manchmal nennen wir auch das Paar (V, ρ) eine Darstellung von G .

(b) Es sei $d \in \mathbb{N}$. Unter einer Matrix-Darstellung von G der Dimension d verstehen wir einen Gruppenhomomorphismus

$$X : G \rightarrow GL(d; \mathbb{C}),$$

wobei $GL(d; \mathbb{C})$ die Gruppe aller invertierbaren $d \times d$ -Matrizen mit Einträgen in \mathbb{C} bezeichnet.

Bemerkung 2.1.2. (i) Wir wollen uns davon überzeugen, dass Darstellungen und Matrix-Darstellungen von G im Wesentlichen dasselbe bedeuten:
Sei zunächst

$$\rho : G \rightarrow GL(V)$$

eine Darstellung von G im Sinne von Definition (2.1.1) (a) und es sei

$$\mathcal{B} = (v_1, \dots, v_d)$$

eine Basis von V . Für einen Endomorphismus $\varphi \in \text{End}_{\mathbb{C}}(V)$ bezeichne $M_{\mathcal{B}}(\varphi)$ die darstellende Matrix von φ bezüglich der Basis \mathcal{B} . Ist $\varphi \in GL(V)$, so gilt

2.1 Darstellungstheorie endlicher Gruppen

bekanntlich $M_{\mathcal{B}}(\varphi) \in GL(d; \mathbb{C})$, das heißt man kann $M_{\mathcal{B}}$ (genauer die Einschränkung davon) auffassen als Abbildung

$$M_{\mathcal{B}} : GL(V) \rightarrow GL(d; \mathbb{C})$$

Eine wichtige Rechenregel besagt, dass für $\varphi, \psi \in \text{End}_{\mathbb{C}}(V)$ gilt:

$$M_{\mathcal{B}}(\psi \circ \varphi) = M_{\mathcal{B}}(\psi) \cdot M_{\mathcal{B}}(\varphi)$$

Dies heißt gerade, dass die Abbildung

$$M_{\mathcal{B}} : GL(V) \rightarrow GL(d; \mathbb{C})$$

ein Gruppenhomomorphismus ist und dieser ist bekanntlich sogar ein Isomorphismus. Indem wir nun die Darstellung ϱ mit $M_{\mathcal{B}}$ verknüpfen, erhalten wir eine Abbildung

$$X := M_{\mathcal{B}} \circ \varrho : G \rightarrow GL(d; \mathbb{C}),$$

welche als Komposition zweier Gruppenhomomorphismen wieder ein Gruppenhomomorphismus, also eine Matrix-Darstellung von G ist.

Es sei umgekehrt

$$X : G \rightarrow GL(d; \mathbb{C})$$

eine gegebene Matrix-Darstellung. Dann setzen wir

$$V := \mathbb{C}^d$$

und bemerken, dass für jedes $A \in GL(d; \mathbb{C})$ durch

$$\varphi_A : \mathbb{C}^d \rightarrow \mathbb{C}^d \quad x \mapsto Ax$$

ein Automorphismus $\varphi_A \in GL(V)$ gegeben ist. Weiterhin ist die Abbildung

$$\Phi : GL(d; \mathbb{C}) \rightarrow GL(V) \quad A \mapsto \varphi_A$$

offensichtlich ein Isomorphismus von Gruppen, so dass die Verknüpfung

$$\varrho := \Phi \circ X : G \rightarrow GL(V)$$

einen Gruppenhomomorphismus, das heißt eine Darstellung von G im Sinne von Definition 2.1.1 (a) liefert.

2 Darstellungstheorie und der random walk auf endlichen Gruppen

(ii) Ist $\varrho : G \rightarrow GL(V)$ eine Darstellung von G , so ist für jedes $g \in G$ der Endomorphismus $\varrho(g)$ diagonalisierbar. In der Tat, wegen $n = |G| < \infty$ gilt nach dem kleinen Satz von Fermat

$$g^n = 1 \text{ für jedes } g \in G$$

und somit auch

$$id_V = \varrho(1) = \varrho(g^n) = \varrho(g)^n,$$

das heißt das Polynom

$$f := X^n - 1 \in \mathbb{C}[X]$$

wird von dem Endomorphismus $\varrho(g)$ annulliert. Da \mathbb{C} algebraisch abgeschlossen ist, existiert die Jordansche Normalform von $\varrho(g)$ und das Polynom f wird vom Minimalpolynom $m_{\varrho(g)}$ von $\varrho(g)$ geteilt.

Nun hat f wegen

$$f = \prod_{k=0}^{n-1} (X - e^{2\pi i \frac{k}{n}})$$

nur einfache Nullstellen, nämlich genau die n -ten Einheitswurzeln in \mathbb{C} . Somit sind auch alle Nullstellen in $m_{\varrho(g)}$ einfach und, da die Vielfachheit einer Nullstelle im Minimalpolynom der Größe des größten Jordan-Kästchens zu dieser Nullstelle entspricht, folgt, dass die Jordansche Normalform von $\varrho(g)$ eine Diagonalmatrix ist.

□

Wir schauen uns zwei Beispiele für Darstellungen an:

Beispiel 2.1.3. (a) Sei $G = \mathbb{S}_n$ und $d = n$. Wir betrachten die Matrix-Darstellung $X : \mathbb{S}_n \rightarrow GL(n; \mathbb{C})$ mit

$$X(\pi) := (e_{\pi(1)} | \dots | e_{\pi(n)}).$$

Dabei bezeichnen e_1, \dots, e_n die Standardbasisvektoren des \mathbb{C}^n . Dass X wirklich eine Matrix-Darstellung definiert, lässt sich leicht nachrechnen:

Seien $\sigma, \pi \in \mathbb{S}_n$. Dann gilt:

$$(X(\pi \circ \sigma))_{i,j} = \delta_{i,(\pi \circ \sigma)(j)}$$

und

$$(X(\pi) \cdot X(\sigma))_{i,j} = \sum_{k=1}^n (X(\pi))_{i,k} (X(\sigma))_{k,j} = \sum_{k=1}^n \delta_{i,\pi(k)} \delta_{k,\sigma(j)} = \delta_{i,\pi(\sigma(j))}.$$

Wegen $X(id) = E_n$ folgt hieraus auch, dass X in die $GL(n; \mathbb{C})$ abbildet. Insgesamt erkennen wir X als Matrix-Darstellung von \mathbb{S}_n , welche die definierende Darstellung der \mathbb{S}_n genannt wird.

- (b) Sei wieder $G = \mathbb{S}_n$, aber diesmal $d = 1$. Offenbar können wir $GL(1; \mathbb{C})$ mit der Gruppe \mathbb{C}^\times identifizieren. Dann betrachten wir die Abbildung

$$X : \mathbb{S}_n \rightarrow \mathbb{C}^\times \quad \pi \mapsto \text{sgn}(\pi)$$

Dies ist bekanntlich ein Gruppenhomomorphismus und die Darstellung X wird als sign-Darstellung bezeichnet.

Die Notation $\varrho(g)(v)$ für $g \in G$ und $v \in V$ ist leider etwas sperrig. Auch ist es von Vorteil, einen \mathbb{C} -Vektorraum V , versehen mit einer Darstellung ϱ von G , als Objekt mit einer gewissen algebraischen Struktur aufzufassen, wobei diese algebraische Struktur genau der Darstellung ϱ entsprechen soll. Dies führt uns zum Begriff eines G -Moduls. Bevor wir die Definition geben, erinnern wir an ein analoges Vorgehen in der Theorie der Gruppenoperationen.

Definition 2.1.4. Es seien S eine Menge und G eine (nicht notwendig endliche) Gruppe. Unter einer Operation von G auf S verstehen wir eine Abbildung

$$\cdot : G \times S \rightarrow S, \quad (g, s) \mapsto g \cdot s$$

mit den folgenden beiden Eigenschaften:

(a) $1 \cdot s = s$

(b) $g \cdot (h \cdot s) = (gh) \cdot s$

jeweils für alle $s \in S$ und $g, h \in G$.

Wenn Verwechslungen ausgeschlossen sind, unterdrücken wir manchmal auch den Punkt in $g \cdot s$.

Ist $G \times S \rightarrow S$ eine Operation von G auf S , so ist für festes $g \in G$ die Abbildung

$$\lambda_g : S \rightarrow S, \quad s \mapsto g \cdot s$$

2 Darstellungstheorie und der random walk auf endlichen Gruppen

eine Bijektion auf S , das heißt $\lambda_g \in \mathbb{S}(S)$. Ferner ist die Abbildung

$$\lambda : G \rightarrow \mathbb{S}(S), \quad g \mapsto \lambda_g$$

ein Gruppenhomomorphismus, das heißt eine Operation von G auf S induziert einen Gruppenhomomorphismus $G \rightarrow \mathbb{S}(S)$.

Ist umgekehrt $\lambda : G \rightarrow \mathbb{S}(S)$ ein gegebener Gruppenhomomorphismus, so wird durch die Abbildung

$$G \times S \ni (g, s) \mapsto (\lambda(g))(s) \in S$$

eine Operation von G auf S erklärt, wie man leicht nachrechnet. Diese beiden Zuordnungen sind invers zueinander und insbesondere bijektiv so dass sich also Operationen von G auf S und Gruppenhomomorphismen $G \rightarrow \mathbb{S}(S)$ auf diese Weise 1 : 1 entsprechen.

Ist nun (V, ϱ) eine Darstellung von G , so ist ϱ nach Definition ein Gruppenhomomorphismus von G nach $GL(V)$. Ein G -Modul wird nun das Analogon zu einer Gruppenoperation $G \times S \rightarrow S$ sein, das der Darstellung ϱ entspricht:

Definition 2.1.5. *Ein G -Modul der Dimension $d \in \mathbb{N}$ ist ein Paar (V, \cdot) , wobei V ein d -dimensionaler \mathbb{C} -Vektorraum und $\cdot : G \times V \rightarrow V$ eine Abbildung ist, so dass die folgenden Bedingungen für alle $v, w \in V$, alle $g, h \in G$ und alle $\lambda \in \mathbb{C}$ erfüllt sind:*

$$(i) \quad 1 \cdot v = v$$

$$(ii) \quad g \cdot (h \cdot v) = (gh) \cdot v$$

$$(iii) \quad g \cdot (v + w) = g \cdot v + g \cdot w$$

$$(iv) \quad g \cdot (\lambda v) = \lambda(g \cdot v)$$

Wir wollen uns davon überzeugen, dass ein G -Modul (V, \cdot) und eine Darstellung (V, ϱ) von G im Wesentlichen dasselbe sind:

Sei zunächst $\varrho : G \rightarrow GL(V)$ eine Darstellung von G . Definiere die Abbildung

$$\cdot : G \times V \rightarrow V$$

durch

$$g \cdot v := \varrho(g)(v).$$

Wegen $\varrho(1) = id_V$ gilt (i) und (ii) folgt aus $\varrho(gh) = \varrho(g) \circ \varrho(h)$. Schließlich folgen (iii) und (iv) daraus, dass $\varrho(g)$ für jedes feste $g \in G$ eine \mathbb{C} -lineare Abbildung ist.

2.1 Darstellungstheorie endlicher Gruppen

Somit ist (V, \cdot) ein G -Modul.

Sei jetzt umgekehrt (V, \cdot) ein G -Modul. Wir definieren die Abbildung

$$\varrho : G \rightarrow GL(V)$$

via

$$\varrho(g)(v) := g \cdot v.$$

Um einzusehen, dass ϱ tatsächlich nach $GL(V)$ abbildet, müssen wir zeigen, dass für jedes $g \in G$ die Abbildung $\varrho(g) : V \rightarrow V$ linear und bijektiv ist. Nach (iii) und (iv) in Definition (2.1.5) gilt für alle $v, w \in V$ und $\lambda \in \mathbb{C}$:

$$\varrho(g)(v + \lambda w) = g \cdot (v + \lambda w) = g \cdot v + g \cdot (\lambda w) = g \cdot v + \lambda(g \cdot v) = \varrho(g)(v) + \lambda \varrho(g)(w),$$

d.h. $\varrho(g)$ ist \mathbb{C} -linear. Weiter ist für jedes $v \in V$

$$(\varrho(g^{-1}) \circ \varrho(g))(v) = \varrho(g^{-1})(\varrho(g)(v)) = g^{-1} \cdot (g \cdot v) = (g^{-1}g) \cdot v = 1 \cdot v = v,$$

so dass

$$\varrho(g^{-1}) \circ \varrho(g) = id_V$$

folgt. Analog folgt auch

$$\varrho(g) \circ \varrho(g^{-1}) = id_V,$$

so dass $\varrho(g^{-1})$ die Umkehrabbildung von $\varrho(g)$ ist und insbesondere $\varrho(g) \in GL(V)$ gilt. Es muss somit nur noch gezeigt werden, dass für alle $g, h \in G$ gilt:

$$\varrho(gh) = \varrho(g) \circ \varrho(h)$$

Dies folgt jedoch sofort aus Bedingung (ii) in Definition 2.1.5:
Für jedes $v \in V$ gilt

$$\varrho(gh)(v) = (gh) \cdot v = g \cdot (h \cdot v) = \varrho(g)(h \cdot v) = \varrho(g)(\varrho(h)(v)) = (\varrho(g) \circ \varrho(h))(v),$$

so dass

$$\varrho(gh) = \varrho(g) \circ \varrho(h)$$

folgt. Insgesamt ist das Paar (V, ϱ) eine Darstellung von G . Man sieht leicht, dass die beiden Zuordnungen zueinander inverse Bijektionen sind und sich somit

2 Darstellungstheorie und der random walk auf endlichen Gruppen

die G -Moduln und die Darstellungen von G 1 : 1 entsprechen.

Es sei S eine nichtleere Menge. Dann kann man stets einen \mathbb{C} -Vektorraum V konstruieren, in dem (im Wesentlichen) die Elemente von S eine Basis bilden. Man betrachte nämlich den Untervektorraum

$$V := \mathbb{C}^{(S)} := \{f : S \rightarrow \mathbb{C} : f(s) \neq 0 \text{ für höchstens endlich viele } s \in S\}$$

von \mathbb{C}^S . Die Standardbasis von $\mathbb{C}^{(S)}$ ist gegeben durch die Familie $(I_{\{s\}})_{s \in S}$, die man vermöge $s \mapsto I_{\{s\}}$ mit den Elementen von S identifiziert. Sei von jetzt an zusätzlich eine Operation von G auf S gegeben und sei S als endlich vorausgesetzt. Dann stimmen natürlich \mathbb{C}^S und $\mathbb{C}^{(S)}$ überein. Im Folgenden nehmen wir an, dass s_1, \dots, s_d die paarweise verschiedenen Elemente von S sind und schreiben, wie in der Darstellungstheorie üblich, $\mathbb{C}[S]$ statt $\mathbb{C}^{(S)}$. Dann können wir unter obiger Identifikation jedes Element $v \in V$ eindeutig als Linearkombination

$$v = \lambda_1 s_1 + \dots + \lambda_d s_d \tag{2.1}$$

mit $\lambda_1, \dots, \lambda_d \in \mathbb{C}$ schreiben. Indem wir für $g \in G$ und v wie in (2.1)

$$g \cdot v := \lambda_1 (g \cdot s_1) + \dots + \lambda_d (g \cdot s_d)$$

definieren, erhalten wir eine Fortsetzung der Operation von G auf S auf V . Wir behaupten, dass $V = \mathbb{C}[S]$ dadurch zu einem G -Modul wird. Die Eigenschaften (i)-(iv) aus Definition 2.1.5 lassen sich alle leicht nachrechnen.

Definition 2.1.6. *Es sei $S = \{s_1, \dots, s_d\}$ eine d -elementige Menge und G operiere auf S . Dann heißt der G -Modul $\mathbb{C}[S]$ der Permutations-Modul zur Operation von G auf S . Die dazugehörige Darstellung $\varrho : G \rightarrow GL(\mathbb{C}[S])$ heißt Permutations-Darstellung.*

Bemerkung 2.1.7. (i) *Wir wollen erklären, woher die Bezeichnung Permutations-Modul kommt: Dazu betrachten wir die Standardbasis $\mathcal{B} = (s_1, \dots, s_d)$ von $\mathbb{C}[S]$ und schauen uns für festes $g \in G$ die darstellende Matrix*

$$X(g) := M_{\mathcal{B}}(\varrho(g))$$

an. Da G auf S operiert, bewirkt g eine Permutation der Basisvektoren s_1, \dots, s_d , das heißt es gibt ein von g abhängiges $\pi \in \mathbb{S}_n$ mit

$$g \cdot s_i = s_{\pi(i)}$$

2.1 Darstellungstheorie endlicher Gruppen

für alle $i = 1, \dots, d$. Somit gilt

$$X(g) = (e_{\pi(1)} | \dots | e_{\pi(d)})$$

und dies ist eine sogenannte Permutations-Matrix. Man beachte, dass $i \in \{1, \dots, d\}$ genau dann ein Fixpunkt von π ist, wenn $g \cdot s_i = s_i$ gilt.

(ii) Im Falle $G = \mathbb{S}_n$ mit der gewöhnlichen Operation $\pi \cdot i = \pi(i)$ auf $S = \{1, \dots, n\}$ erhalten wir als Permutations-Darstellung wieder die definierende Darstellung von \mathbb{S}_n .

Wir kommen nun zu einem im Folgenden sehr wichtigen Beispiel eines Permutations-Moduls:

Die Gruppe G operiert auf sich selbst vermöge der Links-Translation

$$G \times G \rightarrow G, \quad (g, h) \mapsto gh$$

Der dazugehörige Permutationsmodul

$$\mathbb{C}[G] := \left\{ \sum_{i=1}^n \lambda_i g_i : \lambda_i \in \mathbb{C} \text{ für alle } i = 1, \dots, n \right\}$$

hat die Standardbasis g_1, \dots, g_n , welche genau aus den Elementen von G besteht und heißt die *Gruppen-Algebra* von G . Die dazu korrespondierende Darstellung nennt man die (*links-*) *reguläre Darstellung* von G und bezeichnet sie mit ρ_{reg} . Wir werden noch sehen, dass die Gruppen-Algebra $\mathbb{C}[G]$ entscheidend ist, um die Darstellungen von G zu analysieren.

Es soll noch erklärt werden, woher der Begriff Gruppen-Algebra rührt:

Indem man für $v = \sum_{i=1}^n \lambda_i g_i$ und $w = \sum_{j=1}^n \mu_j g_j$ aus $\mathbb{C}[G]$ das Produkt

$$v \cdot w := \sum_{i=1}^n \sum_{j=1}^n \lambda_i \mu_j g_i g_j$$

erklärt, wird $\mathbb{C}[G]$, wie man leicht nachprüft, zu einer \mathbb{C} -Algebra. Insbesondere liefert diese Multiplikation die Struktur eines (im Allgemeinen nicht-kommutativen) Ringes auf $\mathbb{C}[G]$. Hat man weiter einen G -Modul (V, \cdot) gegeben, so kann man die Abbildung

$$\cdot : G \times V \rightarrow V$$

fortsetzen zu einer Abbildung

$$\cdot : \mathbb{C}[G] \times V \rightarrow V \quad \left(\sum_{i=1}^n \lambda_i g_i, v \right) \mapsto \sum_{i=1}^n \lambda_i (g_i \cdot v)$$

und mit dieser Skalar-Multiplikation und der vorhandenen Vektoraddition auf V wird V zu einem $\mathbb{C}[G]$ -Links-Modul im Sinne der Algebra. Insofern lässt sich im Nachhinein die Bezeichnung G -Modul mit der bekannten algebraischen Struktur eines R -Moduls über einem Ring R in Verbindung bringen.

2.1.2 G -Invarianz, Irreduzibilität und Äquivalenz von Darstellungen

Definition 2.1.8. Sei $\varrho : G \rightarrow GL(V)$ eine Darstellung und sei $W \subseteq V$ ein Untervektorraum. Dann heißt W G -invariant, wenn für jedes $g \in G$ und jedes $w \in W$ gilt:

$$\varrho(g)(w) \in W,$$

das heißt, wenn W für jedes $g \in G$ ein $\varrho(g)$ -invarianter Unterraum ist. Man nennt gegebenenfalls W auch einen Untermodul des G -Moduls (V, \cdot) , welcher zur Darstellung (V, ϱ) gehört.

Bemerkung 2.1.9. (a) Jeder G -Modul (V, \cdot) hat die sogenannten trivialen Untermodule $\{0\}$ und V .

(b) Sei (V, ϱ) eine Darstellung von G und sei $W \subseteq V$ ein G -invarianter Untervektorraum. Dann induziert für jedes $g \in G$ die Einschränkung $\varrho(g)|_W$ einen Automorphismus von W und somit auch eine Darstellung ϱ' von G in W :

$$\varrho' : G \rightarrow GL(W), \quad g \mapsto \varrho(g)|_W$$

Definition 2.1.10. Ein G -Modul (V, \cdot) (bzw. die zugehörige Darstellung $\varrho : G \rightarrow GL(V)$) heißt irreduzibel, wenn $V \neq \{0\}$ ist und die trivialen Untermodule $\{0\}$ und V die einzigen Untermodule von V sind.

Wir wollen im Folgenden zeigen, dass die irreduziblen Darstellungen von G in gewisser Hinsicht die Bausteine für alle Darstellungen bilden. Dazu müssen wir etwas ausholen und einige Ergebnisse aus der linearen Algebra rekapitulieren. Es seien K ein Körper, V ein K -Vektorraum und U_1, \dots, U_r seien Untervektorräume von V . Dann war die Summe $\sum_{i=1}^r U_i$ der Unterräume U_1, \dots, U_r definiert durch

$$\sum_{i=1}^r U_i := \{u_1 + \dots + u_r : u_i \in U_i \text{ für } i = 1, \dots, r\}$$

und dies war wieder ein Untervektorraum von V .

2.1 Darstellungstheorie endlicher Gruppen

Erinnerung 2.1.11. In obiger Situation sind für $U := \sum_{i=1}^r U_i$ die folgenden Bedingungen äquivalent:

(i) Sind $u_1 \in U_1, \dots, u_r \in U_r$, so gilt:

$$u_1 + \dots + u_r = 0 \quad \Rightarrow \quad u_1 = \dots = u_r = 0$$

(ii) Jedes $u \in U$ besitzt eine eindeutige Darstellung

$$u = u_1 + \dots + u_r$$

mit $u_1 \in U_1, \dots, u_r \in U_r$.

(iii) Für jedes $i \in \{1, \dots, r\}$ gilt

$$U_i \cap \sum_{j \neq i} U_j = \{0\}$$

(iv) Für jedes $i \in \{1, \dots, r\}$ gilt

$$U_i \cap \sum_{j=i+1}^r U_j = \{0\}$$

Gegebenenfalls heißt U die direkte Summe der Unterräume U_1, \dots, U_r und man schreibt dafür

$$U = \bigoplus_{i=1}^r U_i.$$

In der Situation von Erinnerung 2.1.11 spricht man auch genauer davon, dass U die innere direkte Summe der Unterräume U_1, \dots, U_r ist. Eng damit verwandt ist der Begriff der sogenannten konstruierten direkten Summe:

Seien U_1, \dots, U_r K -Vektorräume, die nicht notwendig Untervektorräume eines bestimmten Vektorraumes V sein müssen. Bekanntlich wird dann das kartesische Produkt

$$U := \prod_{i=1}^r U_i = U_1 \times \dots \times U_r$$

durch komponentenweise definierte Addition und Skalar-Multiplikation zu einem K -Vektorraum. Diesen nennt man das direkte Produkt oder auch (da wir es nur mit endlich vielen Vektorräumen zu tun haben) die konstruierte direkte Summe

2 Darstellungstheorie und der random walk auf endlichen Gruppen

der Vektorräume U_1, \dots, U_r . Die Verbindung zwischen innerer und konstruierter direkter Summe liefert der folgende Satz.

Satz 2.1.12. (a) Seien U_1, \dots, U_r Untervektorräume eines K -Vektorraumes V , so dass die Summe

$$U := \sum_{i=1}^r U_i$$

direkt ist. Dann ist die Abbildung

$$\varphi : \prod_{i=1}^r U_i \rightarrow U, \quad (u_1, \dots, u_r) \mapsto \sum_{i=1}^r u_i$$

ein Isomorphismus von K -Vektorräumen.

(b) Seien U_1, \dots, U_r K -Vektorräume und sei $U := \prod_{i=1}^r U_i$ die konstruierte direkte Summe der Räume U_1, \dots, U_r . Sei für $i = 1, \dots, r$

$$U'_i := \{(0, \dots, u_i, 0, \dots, 0) : u_i \in U_i\}.$$

Dann sind U'_1, \dots, U'_r Untervektorräume von U mit $U_i \cong U'_i$ für jedes $i = 1, \dots, r$ und es ist U die innere direkte Summe der Untervektorräume U'_1, \dots, U'_r :

$$U = \bigoplus_{i=1}^r U'_i.$$

Satz (2.1.12) beinhaltet, dass man, wenn es um wesentliche Eigenschaften geht, nicht zwischen inneren und konstruierten direkten Summen zu unterscheiden braucht. Direkte Summen spielen in der Theorie der G -Moduln eine große Rolle:

Definition 2.1.13. (a) Sei $(V; \cdot)$ ein G -Modul und seien U_1, \dots, U_r G -Untermodule von V . Falls im Sinne von \mathbb{C} -Vektorräumen

$$V = \bigoplus_{i=1}^r U_i$$

gilt, so sagt man auch, dass der G -Modul V die (innere) direkte Summe der Untermodule U_1, \dots, U_r ist.

(b) Seien $(U_1, \cdot), \dots, (U_r, \cdot)$ G -Moduln und sei

$$V := \bigoplus_{i=1}^r U_i$$

2.1 Darstellungstheorie endlicher Gruppen

die konstruierte direkte Summe im Sinne von \mathbb{C} -Vektorräumen. Dann wird V durch die Zuordnung

$$g \cdot (u_1, \dots, u_r) := (g \cdot u_1, \dots, g \cdot u_r)$$

ebenfalls zu einem G -Modul, den man die (konstruierte) direkte Summe der G -Moduln U_1, \dots, U_r nennt. Man schreibt hierfür ggfs. ebenfalls

$$V = \bigoplus_{i=1}^r U_i.$$

Die Aussage, die in Teil (b) von Definition 2.1.13 steckt, ist leicht zu beweisen und der Beweis wird daher an dieser Stelle ausgelassen.

Das Konzept der direkten Summe gibt der Aussage, dass die irreduziblen Darstellungen die Bausteine aller Darstellungen von G sind, präzisen Sinn. Als Hilfsmittel für den Beweis dieser Aussage müssen wir zeigen, dass es zu jedem Untermodul U eines G -Moduls V einen weiteren G -Untermodul W von V gibt, so dass

$$V = U \oplus W$$

gilt. Einen solchen Untermodul W nennt man dann auch ein *Komplement* zu U . Um diese Hilfsaussage zu beweisen, führen wir ein geeignetes Skalarprodukt auf V ein:

Sei dazu $\mathcal{B} = (v_1, \dots, v_d)$ eine beliebige Basis von V . Für $v = \sum_{i=1}^d \lambda_i v_i$ und $w = \sum_{j=1}^d \mu_j v_j$ aus V setzen wir

$$\sigma(v, w) := \sum_{i=1}^d \sum_{j=1}^d \lambda_i \mu_j.$$

Dann ist, wie man leicht nachprüft, σ ein Skalarprodukt auf V . Dieses Skalarprodukt erfüllt leider nicht die folgende wünschenswerte Eigenschaft, welche man G -Invarianz nennt: Es gilt nicht

$$\sigma(g \cdot v, g \cdot w) = \sigma(v, w)$$

für $v, w \in V$ und $g \in G$. Indem wir aber nun setzen

$$\langle v, w \rangle := \sum_{h \in G} \sigma(h \cdot v, h \cdot w)$$

erhalten wir ein G -invariantes Skalarprodukt $\langle \cdot, \cdot \rangle$, denn für $g \in G$ gilt:

2 Darstellungstheorie und der random walk auf endlichen Gruppen

$$\langle g \cdot v, g \cdot w \rangle = \sum_{h \in G} \sigma(hg \cdot v, hg \cdot w) = \sum_{s \in G} \sigma(s \cdot v, s \cdot w) = \langle v, w \rangle,$$

denn $G \ni h \mapsto hg \in G$ ist bijektiv für jedes $g \in G$.

Sei jetzt $\varrho : G \rightarrow GL(V)$ die zu dem Modul (V, \cdot) gehörige Darstellung von G . Wegen

$$\langle \varrho(g)(v), \varrho(g)(w) \rangle = \langle g \cdot v, g \cdot w \rangle = \langle v, w \rangle$$

für alle $v, w \in V$ und alle $g \in G$ ist jedes $\varrho(g)$ ein unitärer Endomorphismus (eine Isometrie) des unitären Vektorraumes $(V, \langle \cdot, \cdot \rangle)$. Da unitäre Endomorphismen endlich-dimensionaler \mathbb{C} -Vektorräume stets diagonalisierbar sind, folgt hieraus insbesondere wieder die Aussage von Bemerkung (2.1.2) (ii). Aus der linearen Algebra ist ferner bekannt, dass für jedes $g \in G$ alle Eigenwerte von $\varrho(g)$ auf dem Einheitskreis

$$S^1 := \{z \in \mathbb{C} : |z| = 1\}$$

liegen. Wir kommen nun zurück auf die Zerlegung der G -Moduln:

Lemma 2.1.14. *Sei (V, \cdot) ein G -Modul und sei U ein G -Untermodul von V . Dann existiert ein weiterer G -Untermodul W von V , so dass gilt:*

$$V = U \oplus W$$

Beweis: Es sei

$$W := U^\perp = \{w \in V : \langle w, u \rangle = 0 \text{ für alle } u \in U\},$$

wobei $\langle \cdot, \cdot \rangle$ das oben konstruierte G -invariante Skalarprodukt bezeichne. Aus der linearen Algebra ist bekannt, dass dann (wegen $\dim V < \infty$) im Sinne von \mathbb{C} -Vektorräumen

$$V = U \oplus W$$

gilt. Es verbleibt somit zu zeigen, dass W ein Untermodul von V ist:

Seien dazu $w \in W$ und $g \in G$ beliebig vorgegeben. Dann gilt für jedes $u \in U$

$$\begin{aligned} \langle g \cdot w, u \rangle &\stackrel{G\text{-Invarianz}}{=} \langle g^{-1} \cdot (g \cdot w), g^{-1} \cdot u \rangle = \langle (g^{-1}g) \cdot w, g^{-1} \cdot u \rangle \\ &= \langle w, g^{-1} \cdot u \rangle = 0, \end{aligned}$$

denn $g^{-1} \cdot u \in U$. Somit ist $g \cdot w \in W$ und das Lemma bewiesen. □

Der folgende Satz liefert die versprochene Zerlegung eines G -Moduls in irreduzible G -Moduln:

Satz 2.1.15 (Maschke). *Es sei (V, \cdot) ein G -Modul. Dann gibt es irreduzible G -Untermodule U_1, \dots, U_r von V mit*

$$V = \bigoplus_{i=1}^r U_i.$$

Beweis: Ist V der Nullmodul, so gilt die Behauptung trivialerweise mit der leeren Familie von Untermodulen. Andernfalls führen wir Induktion über

$d := \dim_{\mathbb{C}}(V) \in \mathbb{N}$:

Induktionsanfang: $d = 1$: In diesem Fall ist V selbst irreduzibel und die Behauptung gilt mit $r = 1$ und $U_1 = V$.

Induktionsschritt: $< d \rightarrow d$: Ist V selbst irreduzibel, so folgt wieder die Behauptung mit $r = 1$ und $U_1 = V$. Andernfalls wähle man einen nicht-trivialen G -Untermodule

$$\{0\} \subsetneq U \subsetneq V.$$

Nach Lemma (2.1.14) existiert ein weiterer Untermodul W von V mit

$$V = U \oplus W.$$

Wegen

$$d = \dim_{\mathbb{C}}(V) = \dim_{\mathbb{C}}(U) + \dim_{\mathbb{C}}(W)$$

folgt aus $\dim_{\mathbb{C}}(U) > 0$, dass auch $\dim_{\mathbb{C}}(W) < d$ ist, das heißt wir können sowohl auf U als auch auf W die Induktionsvoraussetzung anwenden und erhalten irreduzible Untermoduln U_1, \dots, U_k von U und U_{k+1}, \dots, U_r von W mit

$$U = \bigoplus_{i=1}^k U_i \quad \text{und} \quad W = \bigoplus_{i=k+1}^r U_i.$$

Dies sind offenbar auch Untermoduln von V und es gilt

$$V = \bigoplus_{i=1}^r U_i,$$

wie behauptet. □

Natürlich stellt sich auch die Frage nach der Eindeutigkeit einer solchen Zerlegung in irreduzible Moduln. Wir werden darauf später noch zurückkommen.

Der Satz von Maschke lehrt uns, dass sich jeder G -Modul in irreduzible Bausteine zerlegen lässt. Es stellt sich nun die Frage, wie viele verschiedene Bausteine es

geben kann. Um diese Frage befriedigend beantworten zu können, müssen wir klären, wann wir zwei G -Moduln als „im Wesentlichen gleich“ ansehen können und daher miteinander identifizieren können. Dieses Vorgehen ist natürlich aus vielen Bereichen der Mathematik bekannt. Mathematisch bedeutet dies, dass wir uns Isomorphismen oder allgemeiner Homomorphismen, das heißt strukturerhaltende Abbildungen, in der *Kategorie* der G -Moduln anschauen müssen.

Definition 2.1.16. Seien (V, \cdot) und (W, \cdot) G -Moduln und es seien $\varrho : G \rightarrow GL(V)$ und $\theta : G \rightarrow GL(W)$ die dazugehörigen Darstellungen.

(a) Eine \mathbb{C} -lineare Abbildung $T : V \rightarrow W$ heißt ein G -Modul-Homomorphismus, wenn

$$T(g \cdot v) = g \cdot T(v)$$

für alle $v \in V$ und alle $g \in G$ gilt. Sie heißt ein Isomorphismus von G -Moduln, wenn T zusätzlich bijektiv ist.

(b) Falls ein G -Modul-Isomorphismus $T : V \rightarrow W$ existiert, so heißen die G -Moduln V und W isomorph, in Zeichen

$$V \cong W.$$

In diesem Fall nennt man auch die Darstellungen ϱ und θ isomorph oder äquivalent.

Zwei Darstellungen (V, ϱ) und (W, θ) sind somit genau dann äquivalent, wenn ein \mathbb{C} -Vektorraum-Isomorphismus $T : V \rightarrow W$ existiert, so dass

$$T \circ \varrho(g) = \theta(g) \circ T$$

für alle $g \in G$ gilt. Wir bezeichnen mit $\text{Hom}_G(V, W)$ den \mathbb{C} -Vektorraum aller G -Modul-Homomorphismen von V nach W . Die nächste Aussage sollte nicht überraschen:

Proposition 2.1.17. Ist $T : (V, \cdot) \rightarrow (W, \cdot)$ ein Isomorphismus von G -Moduln, so auch die Umkehrabbildung $T^{-1} : W \rightarrow V$.

Beweis: Bekanntlich ist T^{-1} \mathbb{C} -linear. Seien $g \in G$ und $w \in W$. Dann gilt

$$\begin{aligned} T^{-1}(g \cdot w) &= T^{-1}(g \cdot T(T^{-1}(w))) = T^{-1}(T(g \cdot T^{-1}(w))) \\ &= g \cdot T^{-1}(w) \end{aligned}$$

□

Bemerkung 2.1.18. Wir haben am Ende des letzten Abschnitts gesehen, dass G -Moduln (V, \cdot) und (W, \cdot) auf natürliche Weise zu $\mathbb{C}[G]$ -Moduln werden. In diesem Zusammenhang wollen wir noch kurz erwähnen, dass eine Abbildung $T : V \rightarrow W$ genau dann ein G -Modul-Homomorphismus ist, wenn sie ein Homomorphismus von $\mathbb{C}[G]$ -Moduln ist.

Begründung: Sei $T \in \text{Hom}_G(V, W)$ und sei

$$x := \sum_{i=1}^d \lambda_i g_i \in \mathbb{C}[G].$$

Dann gilt für $v \in V$:

$$\begin{aligned} T(x \cdot v) &= T\left(\sum_{i=1}^d \lambda_i (g_i \cdot v)\right) = \sum_{i=1}^d \lambda_i T(g_i \cdot v) \\ &= \sum_{i=1}^d \lambda_i (g_i \cdot T(v)) = \left(\sum_{i=1}^d \lambda_i g_i\right) \cdot T(v) \\ &= x \cdot T(v) \end{aligned}$$

□

Wir wollen nun auf ein einfaches aber sehr nützliches Hilfsresultat der Darstellungstheorie, das sogenannte *Lemma von Schur*, hinarbeiten. Dazu benötigen wir das folgende Lemma.

Lemma 2.1.19. Seien (V, \cdot) und (W, \cdot) G -Moduln und sei $T \in \text{Hom}_G(V, W)$. Dann sind $\ker(T)$ und $\text{im}(T)$ G -Untermoduln von V bzw. von W .

Beweis:

(a) $\ker(T)$ G -invariant: Sei $v \in \ker(T)$. Dann gilt für jedes $g \in G$:

$$T(g \cdot v) = g \cdot T(v) = g \cdot 0 = 0.$$

Somit gilt auch $g \cdot v \in \ker(T)$.

(b) $\text{im}(T)$ G -invariant: Sei $w \in \text{im}(T)$ und sei $v \in V$ mit $T(v) = w$. Dann gilt für jedes $g \in G$:

$$T(g \cdot v) = g \cdot T(v) = g \cdot w,$$

also ist auch $g \cdot w \in \text{im}(T)$.

□

Nun können wir das Lemma von Schur formulieren und beweisen:

Satz 2.1.20 (Lemma von Schur). *Seien (V, \cdot) und (W, \cdot) irreduzible G -Moduln und sei $T \in \text{Hom}_G(V, W)$. Dann gilt:*

(a) *Entweder ist $T = 0$ oder T ist ein Isomorphismus von G -Moduln.*

(b) *Gilt zusätzlich $V = W$, so gibt es ein $\lambda \in \mathbb{C}$ mit*

$$T = \lambda \text{id}_V,$$

das heißt T ist eine Homothetie.

Beweis:

(a) : Ist $T = 0$, so kann T wegen $\dim_{\mathbb{C}}(V) \geq 1$ kein Isomorphismus sein. Somit folgt (a) in diesem Fall. Sei jetzt $T \neq 0$ angenommen. Nach Lemma 2.1.19 ist der Unterraum

$$U := \ker(T) \subseteq V$$

G -invariant. Da V irreduzibel ist, folgt

$$U = \{0\} \quad \text{oder} \quad U = V.$$

Da $T \neq 0$ ist, folgt notwendig $U = \{0\}$ und T ist injektiv. Wieder nach Lemma (2.1.19) ist das Bild

$$U' := \text{im}(T) \subseteq W$$

ein G -invarianter Untervektorraum von W . Da auch W irreduzibel ist, folgt

$$U' = \{0\} \quad \text{oder} \quad U' = W.$$

Wegen $T \neq 0$, kann nicht $U' = \{0\}$ gelten, also ist $U' = W$ und T ist auch surjektiv und somit insgesamt bijektiv. Daher ist T ein Isomorphismus von G -Moduln.

2.1 Darstellungstheorie endlicher Gruppen

(b) : Da \mathbb{C} algebraisch abgeschlossen ist, existiert ein Eigenwert $\lambda \in \mathbb{C}$ von T . Sei

$$T' := T - \lambda \text{id}_V.$$

Dann ist auch $T' \in \text{Hom}_G(V, V)$, denn für alle $v \in V$ und $g \in G$ gilt:

$$\begin{aligned} T'(g \cdot v) &= T(g \cdot v) - \lambda(g \cdot v) = g \cdot T(v) - g \cdot (\lambda v) \\ &= g \cdot (T(v) - \lambda v) = g \cdot T'(v). \end{aligned}$$

Wegen $\ker T' \neq \{0\}$ ist T' kein Isomorphismus und somit folgt aus (a), dass

$$T' = 0 \quad \text{also} \quad T = \lambda \text{id}_V$$

gilt.

□

Wegen seiner großen Bedeutung formulieren wir das Lemma von Schur nun noch in einer Version in der Sprache von Darstellungen:

Korollar 2.1.21. *Seien $\varrho : G \rightarrow GL(V)$ und $\theta : G \rightarrow GL(W)$ zwei irreduzible Darstellungen von G und sei $T : V \rightarrow W$ eine \mathbb{C} -lineare Abbildung mit*

$$T \circ \varrho(g) = \theta(g) \circ T$$

für alle $g \in G$. Dann gilt:

(a) Entweder gilt $T = 0$ oder T ist ein \mathbb{C} -Vektorraum-Isomorphismus.

(b) Gilt zusätzlich $V = W$, so gibt es ein $\lambda \in \mathbb{C}$ mit

$$T = \lambda \text{id}_V.$$

Zu Teil (b) des Lemmas von Schur gilt auch die folgende Umkehrung:

Satz 2.1.22 (Umkehrung des Lemmas von Schur). *Es sei $\varrho : G \rightarrow GL(V)$ eine Darstellung von G mit der Eigenschaft, dass die einzigen \mathbb{C} -linearen Abbildungen $T : V \rightarrow V$, die*

$$T \circ \varrho(g) = \varrho(g) \circ T$$

für alle $g \in G$ erfüllen, die Homothetien $T = \lambda \text{id}_V$, $\lambda \in \mathbb{C}$, sind. Dann ist die Darstellung ϱ irreduzibel.

2 Darstellungstheorie und der random walk auf endlichen Gruppen

Beweis: Wir gehen indirekt vor und nehmen an, dass ein G -invarianter Untermodul U von V existiert mit

$$\{0\} \subsetneq U \subsetneq V.$$

Dann wählen wir gemäß Lemma (2.1.14) einen weiteren G -invarianten Unterraum W von V mit

$$V = U \oplus W.$$

jedes $v \in V$ hat dann eine eindeutige Zerlegung

$$v = u + w \quad \text{mit } u \in U \text{ und } w \in W.$$

Definiere nun eine Abbildung $T : V \rightarrow V$ für solch ein v durch

$$T(v) = T(u + w) := iu + w.$$

Wie man leicht einsieht, ist dann T linear und wir behaupten, dass auch

$$T \circ \varrho(g) = \varrho(g) \circ T$$

gilt. Sei dazu wieder $v = u + w$ die Zerlegung eines beliebigen Elements $v \in V$ wie oben. Dann gilt für jedes $g \in G$:

$$u' := \varrho(g)(u) \in U \quad \text{und} \quad w' := \varrho(g)(w) \in W,$$

denn beide Unterräume sind G -invariant. Es folgt somit nach Definition von T :

$$\begin{aligned} (T \circ \varrho(g))(v) &= T(u' + w') = iu' + w' \\ &= i\varrho(g)(u) + \varrho(g)(w) = \varrho(g)(iu + w) \\ &= \varrho(g)(T(v)) = (\varrho(g) \circ T)(v), \end{aligned}$$

also auch

$$T \circ \varrho(g) = \varrho(g) \circ T$$

für jedes $g \in G$. Da aber sowohl U als auch W nicht-trivial sind, kann T keine Homothetie sein, was der Voraussetzung widerspricht. □

2.1.3 Charaktere von Darstellungen

Für die Darstellungstheorie ist es äußerst wichtig, dass man wesentliche Informationen über eine Darstellung ϱ von G in einer Funktion

$$\chi_\varrho : G \rightarrow \mathbb{C}$$

kodieren kann, die man den *Charakter* der Darstellung ϱ nennt. Bevor wir die Definition geben und elementare Eigenschaften herleiten, müssen wir uns an einige Tatsachen aus der linearen Algebra erinnern:

Es seien K ein Körper und $A \in M(d \times d; K)$ eine quadratische Matrix. Dann ist die *Spur* von A definiert durch

$$\mathrm{Tr}(A) := \sum_{i=1}^d a_{ii}.$$

Ist B eine weitere $d \times d$ -Matrix mit Einträgen in K , so gilt

$$\mathrm{Tr}(AB) = \sum_{i=1}^d \sum_{k=1}^d a_{ik} b_{ki} = \sum_{k=1}^d \sum_{i=1}^d b_{ki} a_{ik} = \mathrm{Tr}(BA).$$

Daraus folgt, dass für jedes $S \in GL(d; K)$

$$\mathrm{Tr}(S^{-1}AS) = \mathrm{Tr}((AS)S^{-1}) = \mathrm{Tr}(A(SS^{-1})) = \mathrm{Tr}(A)$$

gilt, das heißt die Spur ist invariant unter *Konjugation* mit der Gruppe $GL(d; K)$. Insbesondere können wir die Spur eines Endomorphismus T eines d -dimensionalen K -Vektorraumes sinnvoll als Spur einer beliebigen darstellenden Matrix A von T definieren, denn je zwei solche Matrizen sind bekanntlich ähnlich. Sind ferner V und W endlich-dimensionale K -Vektorräume, $\varphi \in \mathrm{End}_K(V)$ und ist $\psi : W \rightarrow V$ ein K -Vektorraum-Isomorphismus, so gilt die Formel

$$\mathrm{Tr}(\psi^{-1} \circ \varphi \circ \psi) = \mathrm{Tr}(\varphi),$$

wie man mittels darstellender Matrizen verifiziert.

Nun kommen wir zu der versprochenen Definition des Charakters.

Definition 2.1.23 (Charakter). Sei $\varrho : G \rightarrow GL(V)$ eine Darstellung von G in V . Dann heißt die Funktion $\chi_\varrho : G \rightarrow \mathbb{C}$ mit

$$\chi_\varrho(g) := \mathrm{Tr}(\varrho(g))$$

der Charakter der Darstellung ϱ .

2 Darstellungstheorie und der random walk auf endlichen Gruppen

Sind $g, h \in G$, so heißt g *konjugiert* zu h , wenn es ein $s \in G$ gibt mit $s^{-1}gs = h$. Konjugiertheit ist offenbar eine Äquivalenzrelation auf G . Die Äquivalenzklasse von $g \in G$ nennt man die *Konjugationsklasse* von g und bezeichnet diese mit $C(g)$. Eine Funktion $F : G \rightarrow \mathbb{C}$ nennen wir eine *Klassen-Funktion*, wenn für je zwei konjugierte Elemente $g, h \in G$ $F(g) = F(h)$ gilt, das heißt, wenn F auf den Konjugationsklassen der Gruppe G konstant ist. Den \mathbb{C} -Vektorraum aller Klassenfunktionen auf G bezeichnen wir im Folgenden mit $K(G)$.

Proposition 2.1.24 (elementare Eigenschaften des Charakters). *Sei $\varrho : G \rightarrow GL(V)$ eine Darstellung von G vom Grad $d \in \mathbb{N}$ mit Charakter χ . Dann gilt*

- (a) $\chi(1) = d$
- (b) χ ist eine Klassen-Funktion.
- (c) $\chi(g^{-1}) = \overline{\chi(g)}$ für alle $g \in G$.
- (d) Ist die Darstellung $\theta : G \rightarrow GL(W)$ äquivalent zu ϱ , so gilt

$$\chi_\theta = \chi_\varrho.$$

Beweis: zu (a): Es ist $\varrho(1) = id_V$, also

$$\chi(1) = \text{Tr}(id_V) = \dim_{\mathbb{C}}(V) = d.$$

zu (b): Seien $g, h \in G$ konjugiert und sei $s \in G$ mit $h = sgs^{-1}$. Dann gilt:

$$\begin{aligned} \chi(h) &= \text{Tr}(\varrho(h)) = \text{Tr}(\varrho(sgs^{-1})) = \text{Tr}(\varrho(s) \circ \varrho(g) \circ \varrho(s)^{-1}) \\ &= \text{Tr}(\varrho(s)^{-1} \circ \varrho(s) \circ \varrho(g)) = \text{Tr}(\varrho(g)) = \chi(g). \end{aligned}$$

zu (c): Wir haben schon in der Passage vor Lemma 2.1.14 gesehen, dass wir ein Skalarprodukt $\langle \cdot, \cdot \rangle$ auf V finden können, bzgl. dessen jedes $\varrho(g)$ ein isometrischer Isomorphismus von V ist. Sei \mathcal{B} eine bzgl. $\langle \cdot, \cdot \rangle$ orthonormale Basis von V . Für festes $g \in G$ sei $\varphi := \varrho(g)$ und $A := M_{\mathcal{B}}(\varphi)$. Dann gilt

$$\varrho(g^{-1}) = \varrho(g)^{-1} = \varphi^{-1}$$

und $A \in U(d)$, das heißt

$$M_{\mathcal{B}}(\varphi^{-1}) = (M_{\mathcal{B}}(\varphi))^{-1} = A^{-1} = \bar{A}^T.$$

Daraus folgt

2.1 Darstellungstheorie endlicher Gruppen

$$\begin{aligned}
 \chi(g^{-1}) &= \operatorname{Tr}(\varphi^{-1}) = \operatorname{Tr}(A^{-1}) = \operatorname{Tr}(\bar{A}^T) \\
 &= \sum_{i=1}^d \overline{a_{ii}} = \sum_{i=1}^d a_{ii} = \overline{\operatorname{Tr}(A)} \\
 &= \overline{\operatorname{Tr}(\varphi)} = \overline{\chi(g)}
 \end{aligned}$$

zu (d): Sind ϱ und θ äquivalente Darstellungen, so gibt es einen \mathbb{C} -Vektorraum-Isomorphismus $T : V \rightarrow W$ mit

$$T \circ \varrho(g) = \theta(g) \circ T$$

für jedes $g \in G$. Es folgt für jedes $g \in G$:

$$\begin{aligned}
 \chi_\theta(g) &= \operatorname{Tr}(\theta(g)) = \operatorname{Tr}(T \circ \varrho(g) \circ T^{-1}) \\
 &= \operatorname{Tr}(T^{-1} \circ T \circ \varrho(g)) = \operatorname{Tr}(\varrho(g)) = \chi_\varrho(g).
 \end{aligned}$$

□

Wir wollen im Folgenden beweisen, dass in Proposition 2.1.24 (d) auch die Umkehrung gilt. Dazu betrachten wir auf dem Raum \mathbb{C}^G aller \mathbb{C} -wertigen Funktionen auf G das Skalarprodukt

$$\langle \psi, \varphi \rangle := \frac{1}{|G|} \sum_{g \in G} \psi(g) \overline{\varphi(g)}$$

sowie die symmetrische Bilinearform

$$\beta(\psi, \varphi) := \frac{1}{|G|} \sum_{g \in G} \psi(g) \varphi(g^{-1}).$$

Dass diese beiden Abbildungen tatsächlich die geforderten Eigenschaften besitzen, ist leicht zu sehen. Man beachte, dass nach Proposition 2.1.24 (c) stets

$$\langle \psi, \varphi \rangle = \beta(\psi, \varphi)$$

gilt, falls φ der Charakter einer Darstellung von G ist. Wir beginnen mit der folgenden Folgerung aus dem Lemma von Schur 2.1.20.

Satz 2.1.25. *Seien (V_1, ϱ_1) und (V_2, ϱ_2) zwei irreduzible Darstellungen von G und $T : V_1 \rightarrow V_2$ sei \mathbb{C} -linear. Sei weiter die \mathbb{C} -lineare Abbildung $T_0 : V_1 \rightarrow V_2$ definiert durch*

2 Darstellungstheorie und der random walk auf endlichen Gruppen

$$T_0 := \frac{1}{|G|} \sum_{g \in G} \varrho_2(g)^{-1} \circ T \circ \varrho_1(g).$$

Dann gilt:

- (i) Sind ϱ_1 und ϱ_2 nicht äquivalent, so ist $T_0 = 0$.
- (ii) Gilt $V_1 = V_2 =: V$ sowie $\varrho_1 = \varrho_2 =: \varrho$, so ist

$$T_0 = \lambda id_V \quad \text{mit} \quad \lambda = \frac{1}{d} \text{Tr}(T),$$

wobei $d := \dim_{\mathbb{C}}(V)$ sei.

Beweis: Behauptung: Es gilt für jedes $h \in G$

$$T_0 \circ \varrho_1(h) = \varrho_2(h) \circ T_0.$$

Begründung: Wir haben

$$\begin{aligned} \varrho_2(h)^{-1} \circ T_0 \circ \varrho_1(h) &\stackrel{\varrho_2(h)^{-1} \text{ linear}}{=} \frac{1}{|G|} \sum_{g \in G} \varrho_2(h)^{-1} \circ \varrho_2(g)^{-1} \circ T \circ \varrho_1(g) \circ \varrho_1(h) \\ &= \frac{1}{|G|} \sum_{g \in G} \varrho_2(gh)^{-1} \circ T \circ \varrho_1(gh) \\ &\stackrel{gh=u}{=} \frac{1}{|G|} \sum_{u \in G} \varrho_2(u)^{-1} \circ T \circ \varrho_1(u) \\ &= T_0. \end{aligned}$$

Im Falle von (i) liefert das Lemma von Schur direkt $T_0 = 0$ und im Falle von (ii) $T_0 = \lambda id_V$, wobei gilt:

$$d \cdot \lambda = \text{Tr}(T_0) = \text{Tr}(T),$$

also

$$\lambda = \frac{1}{d} \text{Tr}(T).$$

□

Nun betrachten wir die Aussage von Satz 2.1.25 für den Fall, dass wir es mit irreduziblen Matrix-Darstellungen

$$A : G \rightarrow GL(d_1; \mathbb{C}) \quad \text{und} \quad B : G \rightarrow GL(d_2; \mathbb{C})$$

zu tun haben. Den linearen Abbildungen T und T_0 entsprechen dann Matrizen

$$X = (x_{ij}), Y = (y_{ij}) \in M(d_2 \times d_1; \mathbb{C})$$

mit

$$Y = \frac{1}{|G|} \sum_{g \in G} B(g)^{-1} X A(g).$$

Korollar 2.1.26. *Unter den obigen Bedingungen gilt:*

(i) *Sind A und B nicht äquivalente Matrix-Darstellungen von G , so gilt für beliebige Indizes i, j, k, l*

$$\beta(a_{lj}, b_{ik}) = \frac{1}{|G|} \sum_{g \in G} a_{lj}(g) b_{ik}(g^{-1}) = 0.$$

(ii) *Ist $A = B$, so gilt für alle Indizes $i, j, k, l \in \{1, \dots, d_1\}$*

$$\beta(a_{lj}, b_{ik}) = \frac{1}{|G|} \sum_{g \in G} a_{lj}(g) b_{ik}(g^{-1}) = \frac{1}{d} \delta_{k,l} \delta_{i,j}.$$

Beweis: zu (i): In diesem Fall folgt nach Satz 2.1.25 (i), dass für jede Matrix X die Matrix $Y = 0$ ist. Insbesondere sind alle Einträge y_{ij} von Y gleich 0. Es gilt somit nach den Regeln für die Matrizenmultiplikation

$$0 = y_{ij} = \frac{1}{|G|} \sum_{g \in G} \sum_{k=1}^{d_2} \sum_{l=1}^{d_1} b_{ik}(g^{-1}) x_{kl} a_{lj}(g).$$

Die rechte Seite kann man (für jede Wahl von i, j) auffassen als eine Linearform in der Matrix X , also als ein Element von

$$M(d_2 \times d_1; \mathbb{C})^*$$

und diese Linearform verschwindet. Somit muss der Koeffizient von jedem x_{kl} gleich 0 sein. Wir erhalten somit für alle Indizes i, j, k, l

$$0 = \frac{1}{|G|} \sum_{g \in G} a_{lj}(g) b_{ik}(g^{-1}) = \beta(a_{lj}, b_{ik}).$$

zu (ii): In diesem Fall ist nach Satz 2.1.25 (ii) die Matrix Y für jede Matrix X eine Homothetie

2 Darstellungstheorie und der random walk auf endlichen Gruppen

$$Y = \lambda E_d$$

($d := d_1 = d_2$) mit

$$\lambda = \frac{1}{d} \text{Tr}(X) = \frac{1}{d} \sum_{k=1}^d \sum_{l=1}^d \delta_{k,l} x_{kl}.$$

Wir erhalten somit

$$\frac{1}{d} \sum_{k=1}^d \sum_{l=1}^d \delta_{k,l} \delta_{i,j} x_{kl} = \lambda \delta_{i,j} = y_{ij} = \frac{1}{|G|} \sum_{g \in G} \sum_{k=1}^d \sum_{l=1}^d b_{ik}(g^{-1}) x_{kl} a_{lj}(g).$$

Wie oben sind die linke und die rechte Seite Linearformen in X und deshalb genau dann gleich, wenn die Koeffizienten von jedem x_{kl} übereinstimmen. Daher gilt für alle Indizes i, j, k, l :

$$\frac{1}{d} \delta_{k,l} \delta_{i,j} = \frac{1}{|G|} \sum_{g \in G} a_{lj}(g) b_{ik}(g^{-1}) = \beta(a_{lj}, b_{ik})$$

□

Man beachte, dass Korollar 2.1.26 gerade besagt, dass die Funktionen, die durch die Einträge der paarweise nicht äquivalenten Matrix-Darstellungen von G gegeben sind, orthonormal sind im Raum \mathbb{C}^G bzgl. der symmetrischen Bilinearform β . Für das G -invariante Skalarprodukt auf dem Raum $K(G)$ erhalten wir zunächst das folgende Resultat:

Satz 2.1.27. *Seien $\varrho_1 : G \rightarrow GL(V_1)$ und $\varrho_2 : G \rightarrow GL(V_2)$ zwei nicht äquivalente, irreduzible Darstellungen von G mit Charakteren χ_1 und χ_2 . Dann gilt:*

$$(i) \quad \langle \chi_1, \chi_1 \rangle = \beta(\chi_1, \chi_1) = 1$$

$$(ii) \quad \langle \chi_1, \chi_2 \rangle = \beta(\chi_1, \chi_2) = 0$$

Mit anderen Worten: Die Charaktere der paarweise nicht äquivalenten, irreduziblen Darstellungen von G bilden ein Orthonormalsystem im unitären Raum $(K(G), \langle \cdot, \cdot \rangle)$.

Beweis: Wir können Matrix-Darstellungen $A : G \rightarrow GL(d_1; \mathbb{C})$ und $B : G \rightarrow GL(d_2; \mathbb{C})$ finden, so dass A äquivalent zu ϱ_1 und B äquivalent zu ϱ_2 ist. Die Charaktere ändern sich dadurch nach Proposition 2.1.24 (d) nicht. Die ersten Identitäten in (i) und (ii) wurden auch bereits begründet.

zu (i): Es gilt

2.1 Darstellungstheorie endlicher Gruppen

$$\chi_1 = \sum_{i=1}^{d_1} a_{ii},$$

also nach Korollar 2.1.26 (ii)

$$\beta(\chi_1, \chi_1) = \sum_{i=1}^{d_1} \sum_{j=1}^{d_1} \beta(a_{ii}, a_{jj}) = \sum_{i=1}^{d_1} \frac{1}{d_1} = 1.$$

zu (ii): Wir haben

$$\chi_1 = \sum_{i=1}^{d_1} a_{ii} \quad \text{und} \quad \chi_2 = \sum_{j=1}^{d_2} b_{jj},$$

also folgt wegen der Bilinearität von β aus Korollar 2.1.26 (i)

$$\beta(\chi_1, \chi_2) = \sum_{i=1}^{d_1} \sum_{j=1}^{d_2} \beta(a_{ii}, b_{jj}) = 0.$$

□

Im Folgenden sei stets $R := ((U_i, \varrho_i))_{i \in I}$ ein vollständiges Repräsentantensystem für die irreduziblen Darstellungen von G . Nach Satz 2.1.27 ist dann mit $\chi_i := \chi_{\varrho_i}$ das System $(\chi_i)_{i \in I}$ ein Orthonormalsystem im Raum $(K(G), \langle \cdot, \cdot \rangle)$. Wir werden sehen, dass es sogar eine Orthonormalbasis von $K(G)$ bildet, insbesondere stimmt $|I|$ mit der Dimension $\dim_{\mathbb{C}} K(G)$ überein, welche gleich der Anzahl der verschiedenen Konjugationsklassen von G ist! Zunächst werden wir jedoch andere wichtige Folgerungen ziehen:

Satz 2.1.28. *Sei (V, \cdot) ein G -Modul mit Charakter χ und es sei*

$$V = U_1 \oplus \dots \oplus U_r$$

eine Zerlegung von V in irreduzible Moduln. Sei weiter (U, \cdot) ein gegebener irreduzibler Modul mit Charakter φ . Dann ist die Anzahl der Moduln U_i , die G -isomorph zu U sind, gegeben durch

$$\langle \chi, \varphi \rangle$$

und insbesondere unabhängig von der gewählten Zerlegung. Man nennt diese Zahl auch die Vielfachheit von U in V .

Beweis: Für $i = 1, \dots, r$ sei χ_i der Charakter der zu (U_i, \cdot) gehörigen Darstellung. Indem man Basen der Moduln U_i zu einer Basis von V vereinigt und die darstellenden Matrizen der $\varrho(g)$ bzgl. dieser Basis betrachtet, erhält man, dass

$$\chi = \sum_{i=1}^r \chi_i$$

gilt. Nach Satz 2.1.27 gilt für $i = 1, \dots, r$:

$$\langle \chi_i, \varphi \rangle = \begin{cases} 1 & \text{falls } U_i \cong U \\ 0 & \text{sonst} \end{cases}$$

Somit gilt:

$$\langle \chi, \varphi \rangle = \sum_{i=1}^r \langle \chi_i, \varphi \rangle = \sum_{i=1}^r I_{\{U_i \cong U\}}$$

□

Satz 2.1.28 gibt uns also eine Antwort auf die Frage, inwieweit die Zerlegung eines gegebenen G -Moduls in irreduzible G -Moduln eindeutig ist. Nun können wir die folgende Verallgemeinerung des Lemmas von Schur (vgl. Satz 2.1.20) formulieren und beweisen:

Satz 2.1.29 (Verallgemeinerung des Lemmas von Schur). *Seien (V, \cdot) und (W, \cdot) G -Moduln, wobei V irreduzibel sei. Dann ist*

$$\dim_{\mathbb{C}}(\text{Hom}_G(V, W))$$

gleich der Vielfachheit von V in W .

Beweis: Es sei $k \in \mathbb{N}_0$ die Vielfachheit von V in W und es seien $V_1, \dots, V_k \subseteq W$ zu V isomorphe Untermoduln von W und W' sei ein weiterer Untermodul von W mit

$$W = V_1 \oplus \dots \oplus V_k \oplus W'.$$

Dann gilt:

$$\text{Hom}_G(V, W) \cong \left(\bigoplus_{i=1}^k \text{Hom}_G(V, V_i) \right) \oplus \text{Hom}_G(V, W') \quad (2.2)$$

im Sinne von \mathbb{C} -Vektorräumen. Nach dem Lemma von Schur ist für jedes $i = 1, \dots, k$

$$\text{Hom}_G(V, V_i) \cong \mathbb{C} \quad (2.3)$$

und es ist

$$\mathrm{Hom}_G(V, W') = \{0\}, \quad (2.4)$$

denn ein $T \in \mathrm{Hom}_G(V, W')$ ist entweder injektiv, oder gleich 0, da $\ker T$ ein Untermodul des irreduziblen Moduls V ist. Wäre aber T injektiv, so gälte

$$W' \supseteq T(V) \cong V$$

im Widerspruch zur Wahl von k . Aus (2.2), (2.3) und (2.4) folgt

$$\mathrm{Hom}_G(V, W) \cong \mathbb{C}^k$$

und daraus durch die Betrachtung der Dimension die Behauptung. \square

Wir haben in Proposition 2.1.24 (d) gesehen, dass äquivalente Darstellungen denselben Charakter besitzen. Nun können wir auch die Umkehrung dieser Aussage beweisen und damit insbesondere im Nachhinein die Bezeichnung Charakter rechtfertigen.

Satz 2.1.30. *Zwei Darstellungen (V_1, ϱ_1) und (V_2, ϱ_2) , die denselben Charakter χ besitzen, sind äquivalent.*

Beweis: Nach dem Satz von Maschke, Satz 2.1.15, lassen sich beide G -Moduln V_1 und V_2 als direkte Summen irreduzibler G -Moduln schreiben. Für jeden beliebigen irreduziblen G -Modul U mit Charakter φ ist die Anzahl der irreduziblen Summanden, die isomorph zu U sind, nach Satz 2.1.28 in beiden Zerlegungen gleich $\langle \chi, \varphi \rangle$. Daraus ergibt sich unmittelbar die G -Isomorphie von V_1 und V_2 . \square

Ist $R = ((U_i, \cdot))_{i \in I}$ ein vollständiges Repräsentantensystem für die irreduziblen G -Moduln, so ist es üblich, die Zerlegung eines gegebenen G -Moduls V als

$$V \cong \bigoplus_{i \in I} m_i U_i$$

zu schreiben. Dabei sind die m_i nicht-negative ganze Zahlen, von denen höchstens endlich viele von 0 verschieden sind. Ist wie oben χ_i der Charakter von U_i , so folgt in dieser Situation für den Charakter χ von V die Gleichheit

$$\chi = \sum_{i \in I} m_i \chi_i$$

und weiter gilt nach Satz 2.1.27

$$\begin{aligned}\langle \chi, \chi \rangle &= \sum_{i,j \in I} m_i m_j \langle \chi_i, \chi_j \rangle \\ &= \sum_{i,j \in I} m_i m_j \delta_{i,j} = \sum_{i \in I} m_i^2\end{aligned}$$

Daraus erhalten wir sofort den folgenden Satz:

Satz 2.1.31. *Sei (ϱ, V) eine Darstellung von G mit Charakter χ . Dann gilt:*

- (i) *Es ist $\langle \chi, \chi \rangle$ eine nicht-negative ganze Zahl.*
- (ii) *Die Darstellung ϱ ist genau dann irreduzibel, wenn $\langle \chi, \chi \rangle = 1$ gilt.*

2.1.4 Zerlegung der Gruppen-Algebra $\mathbb{C}[G]$

Das Hauptziel dieses Abschnitts ist es zu zeigen, dass es genauso viele paarweise nicht äquivalente Darstellungen wie Konjugationsklassen von G gibt. Daraus folgt unmittelbar, dass die Charaktere dieser Darstellungen eine Orthonormalbasis für den Raum $K(G)$ der Klassen-Funktionen auf G bilden. Es sei wieder $R = ((U_i, \varrho_i))_{i \in I}$ ein vollständiges Repräsentantensystem für die paarweise nicht äquivalenten, irreduziblen Darstellungen von G . Weiter seien $d_i := \dim_{\mathbb{C}}(U_i)$ und χ_i der Charakter von ϱ_i für $i \in I$. Ferner bezeichnen wir mit ϱ_{reg} die zur Gruppen-Algebra gehörige reguläre Darstellung von G und mit χ_{reg} ihren Charakter. Wir beginnen mit dem folgenden einfachen, aber sehr wichtigen Satz.

Proposition 2.1.32. (a) *Es sei $\emptyset \neq S$ eine endliche Menge, auf der G operiere. Weiter sei $\mathbb{C}[S]$ der dazugehörige Permutationsmodul mit Darstellung ϱ . Dann gilt:*

Für jedes $g \in G$ ist $\chi_{\varrho}(g)$ die Anzahl der Fixpunkte der Abbildung

$$S \ni s \mapsto g \cdot s \in S.$$

(b) *Es ist für $g \in G$*

$$\chi_{reg}(g) = \begin{cases} |G| & \text{für } g = 1 \\ 0 & \text{sonst} \end{cases}$$

2.1 Darstellungstheorie endlicher Gruppen

Beweis: zu (a): Sei $\mathcal{B} := (s_1, \dots, s_d)$ die kanonische Basis von $\mathbb{C}[S]$. Dann gibt es zu jedem $g \in G$ eine Permutation $\pi \in \mathbb{S}_n$ mit $\varrho(g)(s_i) = s_{\pi(i)}$ für $i = 1, \dots, d$. Somit gilt für die Darstellungs-Matrix A von $\varrho(g)$ bzgl. dieser Basis

$$A = (e_{\pi(1)} | \dots | e_{\pi(d)})$$

und weiter ist

$$\chi_\varrho(g) = \text{Tr}(\varrho(g)) = \text{Tr}(A) = \sum_{i=1}^d \delta_{i, \pi(i)}$$

und dies ist die Anzahl der Fixpunkte von π .

zu (b): Die Operation von G auf sich selbst, welche der regulären Darstellung zu Grunde liegt, ist die Linkstranslation

$$(g, s) \mapsto gs.$$

Ist $g \neq 1$, so hat die Abbildung

$$G \ni s \mapsto gs \in G$$

keinen Fixpunkt, denn aus $gs = s$ folgt $g = 1$. Für $g = 1$ hingegen ist diese Abbildung gleich id_G und hat daher $|G|$ Fixpunkte. Die Behauptung folgt nun aus Teil (a). □

Satz 2.1.33. *Jede irreduzible Darstellung (U_i, ϱ_i) aus dem obigen Vertretersystem tritt in der Zerlegung von $\mathbb{C}[G]$ mit Vielfachheit d_i auf, das heißt es gilt*

$$\mathbb{C}[G] \cong \bigoplus_{i \in I} d_i U_i.$$

Insbesondere ist I eine endliche Menge.

Beweis: Nach Satz 2.1.28 ist diese Vielfachheit gegeben durch

$$\begin{aligned} \langle \chi_{reg}, \chi_i \rangle &= \frac{1}{|G|} \sum_{g \in G} \chi_{reg}(g) \chi_i(g^{-1}) \\ &= \frac{1}{|G|} \chi_{reg}(1) \chi_i(1) \\ &= \chi_i(1) = \dim_{\mathbb{C}}(U_i) = d_i. \end{aligned}$$

□

2 Darstellungstheorie und der random walk auf endlichen Gruppen

Daraus erhalten wir unmittelbar die Identität

$$\chi_{reg} = \sum_{i \in I} d_i \chi_i,$$

aus der wir sofort weitere Folgerungen ziehen:

Korollar 2.1.34. *Mit den obigen Bezeichnungen gilt:*

(i) $\sum_{i \in I} d_i^2 = |G|$

(ii) Für jedes $g \in G \setminus \{1\}$ ist $\sum_{i \in I} d_i \chi_i(g) = 0$.

Beweis: zu (i): Es gilt nach Proposition 2.1.32

$$|G| = \chi_{reg}(1) = \sum_{i \in I} d_i \chi_i(1) = \sum_{i \in I} d_i^2.$$

zu (ii): Nach derselben Proposition gilt

$$0 = \chi_{reg}(g) = \sum_{i \in I} d_i \chi_i(g).$$

□

Definition 2.1.35. *Für eine Funktion $f : G \rightarrow \mathbb{C}$ und eine Darstellung (V, ϱ) von G definieren wir die Fouriertransformierte von f in ϱ als die \mathbb{C} -lineare Abbildung $\hat{f}(\varrho) : V \rightarrow V$ mit*

$$\hat{f}(\varrho) := \sum_{g \in G} f(g) \varrho(g).$$

Der folgende Satz wird sich als sehr nützlich erweisen, um die Anzahl $|I|$ der paarweise nicht äquivalenten Darstellungen von G zu bestimmen.

Satz 2.1.36. *Es seien $f \in K(G)$ eine Klassen-Funktion sowie $\varrho : G \rightarrow V$ eine irreduzible Darstellung von G vom Grad d und mit Charakter χ . Dann ist ihre Fouriertransformierte $\hat{f}(\varrho)$ in ϱ eine Homothetie, das heißt es gilt*

$$\hat{f}(\varrho) = \lambda id_V.$$

Ferner gilt

$$\lambda = \frac{1}{d} \sum_{g \in G} f(g) \chi(g) = \frac{|G|}{d} \langle f, \bar{\chi} \rangle.$$

2.1 Darstellungstheorie endlicher Gruppen

Beweis: Nach dem Lemma von Schur, vgl. Satz 2.1.20, genügt es zu zeigen, dass für jedes $h \in G$ gilt:

$$\varrho(h) \circ \hat{f}(\varrho) \circ \varrho(h)^{-1} = \hat{f}(\varrho)$$

Wegen der Linearität von $\varrho(h)$ gilt

$$\begin{aligned} \varrho(h) \circ \hat{f}(\varrho) \circ \varrho(h)^{-1} &= \sum_{g \in G} f(g) (\varrho(h) \circ \varrho(g) \circ \varrho(h^{-1})) \\ &= \sum_{g \in G} f(g) \varrho(hgh^{-1}) \\ &\stackrel{u=hgh^{-1}}{=} \sum_{u \in G} f(h^{-1}uh) \varrho(u) \\ &\stackrel{f \in K(G)}{=} \sum_{u \in G} f(u) \varrho(u) \\ &= \hat{f}(\varrho), \end{aligned}$$

denn

$$G \ni g \mapsto hgh^{-1} \in G$$

ist ein Automorphismus von G . Somit gilt also $\hat{f}(\varrho) = \lambda id_V$ mit

$$\begin{aligned} \lambda d &= \text{Tr}(\lambda id_V) = \text{Tr}(\hat{f}(\varrho)) = \sum_{g \in G} f(g) \text{Tr}(\varrho(g)) \\ &= \sum_{g \in G} f(g) \chi(g) = |G| \langle f, \bar{\chi} \rangle, \end{aligned}$$

woraus unmittelbar die Behauptung folgt. □

Nun können wir den folgenden Satz beweisen!

Satz 2.1.37. Die Familie $(\chi_i)_{i \in I}$ bildet eine Orthonormalbasis des \mathbb{C} -Vektorraums $(K(G), \langle \cdot, \cdot \rangle)$.

Beweis: Nach Satz 2.1.27 ist dies ein Orthonormalsystem. Wir müssen somit nur zeigen, dass

$$U := \text{span}(\chi_i, i \in I) = K(G)$$

gilt. Es ist

2 Darstellungstheorie und der random walk auf endlichen Gruppen

$$I = \dim_{\mathbb{C}}(U) \leq \dim_{\mathbb{C}} K(G) \leq |G| < +\infty$$

und daher

$$K(G) = U \oplus U^{\perp}.$$

Wir müssen zeigen, dass $U^{\perp} = \{0\}$ gilt. Es sei

$$W := \text{span}(\overline{\chi}_i, i \in I).$$

Wegen

$$\dim_{\mathbb{C}}(W) = \dim_{\mathbb{C}}(U) = |I|$$

ist $U^{\perp} = \{0\}$ äquivalent zu $W^{\perp} = \{0\}$, was wir nun zeigen werden.

Sei also $f \in W^{\perp}$. Ist (V, ϱ) eine *irreduzible* Darstellung von G vom Grad d , so ist nach Satz 2.1.37 die Fouriertransformierte $\hat{f}(\varrho)$ von f in ϱ eine Homothetie $\hat{f}(\varrho) = \lambda id_V$ mit

$$\lambda = \frac{1}{d} \sum_{g \in G} f(g) \chi(g) = \frac{|G|}{d} \langle f, \overline{\chi} \rangle = 0,$$

denn $f \in W^{\perp}$ und $\overline{\chi}$ ist Teil einer Basis von W . Es ist somit

$$\hat{f}(\varrho) = 0 \in \text{End}_{\mathbb{C}}(V).$$

Ist ϱ nicht irreduzibel und dann

$$V \cong \bigoplus_{i \in I} m_i U_i$$

die Zerlegung von V in die paarweise nicht äquivalenten irreduziblen Darstellungen (U_i, ϱ_i) , so gilt

$$\begin{aligned} \hat{f}(\varrho) &= \sum_{g \in G} f(g) \underbrace{(\varrho_i(g), \dots, \varrho_i(g))}_{m_i}{}_{i \in I} \\ &= \underbrace{\left(\sum_{g \in G} f(g) \varrho_i(g), \dots, \sum_{g \in G} f(g) \varrho_i(g) \right)}_{m_i}{}_{i \in I} \\ &= 0. \end{aligned}$$

2.1 Darstellungstheorie endlicher Gruppen

Nun wählen wir speziell $\varrho = \varrho_{reg}$. Dann ist also $\hat{f}(\varrho_{reg}) = 0$, also gilt für den Vektor $1 \in G \subseteq \mathbb{C}[G]$

$$0 = \hat{f}(\varrho_{reg})(1) = \left(\sum_{g \in G} f(g) \varrho_{reg}(g) \right) (1) = \sum_{g \in G} f(g)(g1) = \sum_{g \in G} f(g)g.$$

Da die Elemente von G eine Basis der Gruppen-Algebra $\mathbb{C}[G]$ bilden, folgt hieraus $f(g) = 0$ für alle $g \in G$ und somit ist $f = 0$. Somit haben wir

$$W^\perp = \{0\}$$

bewiesen. □

Die folgende Aussage ist wenig überraschend.

Proposition 2.1.38. *Die Dimension des Vektorraumes $K(G)$ aller Klassen-Funktionen auf G ist gleich der Anzahl der verschiedenen Konjugationsklassen von G .*

Beweisskizze: Seien C_1, \dots, C_r die verschiedenen Konjugationsklassen von G . Definiere die Funktionen $f_i : G \rightarrow \mathbb{C}$ durch $f_i := I_{C_i}$ für $i = 1, \dots, r$. Dann sind $f_1, \dots, f_r \in K(G)$ und man rechnet leicht nach, dass sie sogar eine Basis dieses Raumes bilden, woraus die Behauptung unmittelbar folgt. □

Korollar 2.1.39. *Die Anzahl $|I|$ der paarweise nicht äquivalenten irreduziblen Darstellungen von G entspricht der Anzahl der verschiedenen Konjugationsklassen von G .*

Beweis: Dies folgt unmittelbar aus Proposition 2.1.38 und Satz 2.1.37. □

Satz 2.1.40 (Charakter-Relationen 2. Art). *Es seien $g_1, g_2 \in G$ mit Konjugationsklassen $C(g_1)$ und $C(g_2)$. Dann gilt*

$$\sum_{i \in I} \chi_{\varrho_i}(g_1) \overline{\chi_{\varrho_i}(g_2)} = \frac{|G|}{|C(g_1)|} \delta_{C(g_1), C(g_2)}.$$

Beweis: Sei $s := |I|$ die Anzahl der paarweise nicht äquivalenten irreduziblen Darstellungen von G , die nach Korollar 2.1.39 mit der Anzahl der Konjugationsklassen von G übereinstimmt. Seien χ_1, \dots, χ_s sämtliche irreduziblen Charaktere von G und seien C_1, \dots, C_s die Konjugationsklassen von G . Dann definieren wir die Matrix $A \in M(s \times s; \mathbb{C})$ so, dass deren Eintrag in der k -ten Zeile und j -ten

2 Darstellungstheorie und der random walk auf endlichen Gruppen

Spalte gegeben ist durch $a_{k,l} := \sqrt{\frac{|C_j|}{|G|}} \chi_k(C_j)$, wobei $\chi_k(C_j)$ den Wert von χ_k in einem beliebigen Element der Konjugationsklasse C_j bezeichne. Nach Satz 2.1.27 gilt für alle $k, l \in \{1, \dots, s\}$

$$\sum_{j=1}^s a_{k,j} \overline{a_{l,j}} = \sum_{j=1}^s \frac{|C_j|}{|G|} \chi_k(C_j) \overline{\chi_l(C_j)} = \frac{1}{|G|} \sum_{g \in G} \chi_k(g) \overline{\chi_l(g)} = \delta_{k,l}.$$

Dies heißt gerade, dass die Zeilen von A eine Orthonormalbasis des \mathbb{C}^d bilden, so dass $A \in U(d)$ gilt. Dann sind aber auch die Spalten von A orthonormal, so dass für $j, m \in \{1, \dots, s\}$ gilt

$$\begin{aligned} \delta_{j,m} &= \sum_{k=1}^s a_{k,j} \overline{a_{k,m}} = \sum_{k=1}^s \sqrt{\frac{|C_j|}{|G|}} \sqrt{\frac{|C_m|}{|G|}} \chi_k(C_j) \overline{\chi_k(C_m)} \\ &= \frac{\sqrt{|C_j| |C_m|}}{|G|} \sum_{k=1}^s \chi_k(C_j) \overline{\chi_k(C_m)}, \end{aligned}$$

woraus sich durch Umstellen unmittelbar die Behauptung ergibt. □

Im Folgenden werden wir mit

$$Irr(G) := \bigcup_{i \in I} \{(U_i, \varrho_i)\}$$

die Menge der paarweise nicht äquivalenten Darstellungen aus unserem gewählten Vertretersystem bezeichnen. Nach Korollar 2.1.34 (i) gilt

$$1 = \sum_{\varrho \in Irr(G)} \frac{d_\varrho^2}{|G|},$$

wobei d_ϱ den Grad der Darstellung ϱ bezeichne. Somit erhalten wir durch die Festlegung

$$\mathbb{P}(\{\varrho\}) := \frac{d_\varrho^2}{|G|} \tag{2.5}$$

für $\varrho \in Irr(G)$ und eindeutige Ausdehnung für $A \in \mathcal{P}(Irr(G))$ ein Wahrscheinlichkeitsmaß \mathbb{P} auf dem Messraum $(Irr(G), \mathcal{P}(Irr(G)))$.

Definition 2.1.41. *Das durch Gleichung (2.5) definierte Wahrscheinlichkeitsmaß \mathbb{P} auf $(Irr(G), \mathcal{P}(Irr(G)))$ heißt das Plancherel-Maß zur endlichen Gruppe G .*

Der folgende Satz wird sich noch als sehr nützlich erweisen:

Satz 2.1.42. (a) Sei $f : G \rightarrow \mathbb{C}$ eine beliebige Funktion. Dann gilt die Fourier-Umkehrformel: Für jedes $g \in G$ ist

$$f(g) = \frac{1}{|G|} \sum_{\varrho \in \text{Irr}(G)} d_{\varrho} \text{Tr}(\varrho(g^{-1}) \circ \hat{f}(\varrho)).$$

(b) Seien $f, h : G \rightarrow \mathbb{C}$ zwei beliebige Funktionen. Dann gilt die Plancherel-Formel:

$$\sum_{g \in G} f(g^{-1})h(g) = \frac{1}{|G|} \sum_{\varrho \in \text{Irr}(G)} d_{\varrho} \text{Tr}(\hat{f}(\varrho) \circ \hat{h}(\varrho))$$

Beweis: zu (a): Beide Seiten sind linear in $f = \sum_{x \in G} f(x)I_{\{x\}}$. Somit genügt es, die Behauptung für $f = I_{\{x\}}$ zu zeigen. Es gilt zunächst

$$\hat{f}(\varrho) = \sum_{g \in G} f(g)\varrho(g) = \varrho(x)$$

und somit

$$\begin{aligned} \frac{1}{|G|} \sum_{\varrho \in \text{Irr}(G)} d_{\varrho} \text{Tr}(\varrho(g^{-1}) \circ \hat{f}(\varrho)) &= \frac{1}{|G|} \sum_{\varrho \in \text{Irr}(G)} d_{\varrho} \text{Tr}(\varrho(g^{-1}x)) \\ &= \frac{1}{|G|} \sum_{\varrho \in \text{Irr}(G)} d_{\varrho} \chi_{\varrho}(g^{-1}x) = \frac{1}{|G|} \chi_{\text{reg}}(g^{-1}x) \\ &\stackrel{\text{Prop. 2.1.32(b)}}{=} \delta_{g,x} = I_{\{x\}}(g) \\ &= f(g). \end{aligned}$$

zu (b): Wieder sind beide Seiten linear in f . Sei also wieder ohne Einschränkung $f = I_{\{x\}}$. Dann gilt nach (a)

$$\begin{aligned} \sum_{g \in G} f(g^{-1})h(g) &= h(x^{-1}) = \frac{1}{|G|} \sum_{\varrho \in \text{Irr}(G)} d_{\varrho} \text{Tr}(\varrho(x) \circ \hat{h}(\varrho)) \\ &= \frac{1}{|G|} \sum_{\varrho \in \text{Irr}(G)} d_{\varrho} \text{Tr}(\hat{f}(\varrho) \circ \hat{h}(\varrho)). \end{aligned}$$

□

2.1.5 Induktion und Restriktion von Darstellungen

Sei G eine Gruppe der Ordnung $n \in \mathbb{N}$ und es sei H eine Untergruppe von G . Ist $\rho : G \rightarrow GL(V)$ eine Darstellung von G , so erhalten wir eine Darstellung φ von H , indem wir ρ einfach auf H einschränken, das heißt $\varphi = \rho|_H$.

Definition 2.1.43. Die Darstellung $\varphi = \rho|_H$ heißt die Einschränkung von ρ auf die Untergruppe H und wird mit $\text{Res}_H^G(\rho)$ bezeichnet. Ihren Charakter bezeichnen wir mit $\text{Res}_H^G(\chi_\rho)$.

Ist umgekehrt eine Darstellung $\theta : H \rightarrow GL(W)$ von H gegeben, so wollen wir daraus auf möglichst natürliche Weise eine Darstellung ρ von G gewinnen, welche wir dann die *von θ induzierte Darstellung* nennen werden. Wir wollen zunächst erklären, was es heißt, dass eine Darstellung $\rho : G \rightarrow GL(V)$ von G von einer Darstellung $\theta : H \rightarrow GL(W)$ induziert wird. Dazu müssen wir etwas ausholen:

Es bezeichne G/H die Menge der *Linksnebenklassen* von H in G und wie üblich sei $s := [G : H] := |G/H| = \frac{|G|}{|H|}$ der *Index* der Untergruppe H in G . Ferner sei stets x_1, \dots, x_s ein vollständiges Repräsentantensystem für die durch die Linksnebenklassen gegebene Äquivalenzrelation auf G , wobei wir ohne Einschränkung $x_1H = H$ voraussetzen. Wir bezeichnen im Folgenden mit $\sigma_j := x_jH$ die Linksnebenklasse von x_j . Es sei nun $\rho : G \rightarrow GL(V)$ eine Darstellung von G und es sei $W \subseteq V$ ein H -invarianter Untervektorraum von V , das heißt es gilt $\rho(h)(w) \in W$ für alle $h \in H$ und alle $w \in W$. Dann bezeichnen wir mit $\theta : H \rightarrow GL(W)$ die Darstellung von H , die sich durch Einschränkung von ρ auf H und W aus dieser ergibt. Die Untervektorräume $W_{\sigma_j} := \rho(x_j)(W)$ hängen nicht von der Wahl der Repräsentanten $x_j \in \sigma_j$ ab, denn für $h \in H$ und $w \in W$ ist $\rho(x_jh)(w) = \rho(x_j)(\rho(h)(w)) = \rho(x_j)(\tilde{w})$, wobei $\tilde{w} := \rho(h)(w) \in W$ nach Voraussetzung gilt. Außerdem werden die Räume $W_{\sigma_j}, j = 1, \dots, s$ für jedes $g \in G$ von der Abbildung $\rho(g)$ permutiert, so dass insbesondere $\sum_{j=1}^s W_{\sigma_j}$ ein ρ -invarianter Unterraum von V ist.

Definition 2.1.44. Mit den obigen Bezeichnungen heißt die Darstellung ρ von G von der Darstellung θ von H *induziert*, falls V als \mathbb{C} -Vektorraum die direkte Summe der Untervektorräume $W_{\sigma_1}, \dots, W_{\sigma_s}$ ist, das heißt, falls

$$V = \bigoplus_{j=1}^s W_{\sigma_j}$$

im Sinne von \mathbb{C} -Vektorräumen gilt.

Proposition 2.1.45. Wird die Darstellung (V, ρ) von G von der Darstellung (W, θ) von H induziert, so gilt

$$\dim_{\mathbb{C}}(V) = [G : H] \cdot \dim_{\mathbb{C}}(W).$$

Beweis: Wir haben

$$\dim_{\mathbb{C}}(V) = \dim_{\mathbb{C}}\left(\bigoplus_{j=1}^s W_{\sigma_j}\right) = s \cdot \dim_{\mathbb{C}}(W) = [G : H] \cdot \dim_{\mathbb{C}}(W),$$

denn jedes W_{σ_j} ist vermöge der Abbildung $\varrho(x_j)$ isomorph zu W . □

Wir benötigen später einige Eigenschaften induzierter Darstellungen, wollen aber zunächst zeigen, dass diese stets existieren und im Wesentlichen eindeutig bestimmt sind. Dazu benötigen wir das folgende Lemma:

Lemma 2.1.46. *Seien W ein H -Modul und seien V und U G -Moduln, so dass V von W induziert werde. Dann besitzt jeder H -Modul-Homomorphismus $\varphi : W \rightarrow U$ (U ist offenbar auch ein H -Modul!) genau eine Fortsetzung $\tilde{\varphi} : V \rightarrow U$ zu einem G -Modul-Homomorphismus.*

Beweis: Nach Voraussetzung gilt $V = \bigoplus_{j=1}^s W_{\sigma_j}$. Ist $\tilde{\varphi}$ wie im Text des Lemmas, so gilt für $v = x_j \cdot w \in W_{\sigma_j}$:

$$\tilde{\varphi}(x_j \cdot w) = x_j \cdot \tilde{\varphi}(w) = x_j \cdot \varphi(w), \quad (2.6)$$

so dass $\tilde{\varphi}$ eindeutig bestimmt ist. Für die Existenz definieren wir nun $\tilde{\varphi}$ für $v = x_j \cdot w \in W_{\sigma_j}$ durch Gleichung (2.6) und setzen linear auf V fort. Wir zeigen noch, dass diese Definition nicht von der Wahl des Repräsentanten $x_j \in \sigma_j$ abhängt: Sei $h \in H$, dann gilt $v = (x_j h) \cdot (h^{-1} \cdot w)$ mit $h^{-1} \cdot w \in W$ und somit

$$(x_j h) \cdot \varphi(h^{-1} \cdot w) = x_j \cdot \varphi((h h^{-1}) \cdot w) = x_j \cdot \varphi(w),$$

so dass die Definition von $\tilde{\varphi}$ unabhängig von der Wahl der Vertreter ist. Offensichtlich ist $\tilde{\varphi}$ eine Fortsetzung von φ , so dass wir nur noch zeigen müssen, dass es ein G -Modul-Homomorphismus ist: Dazu sei $v = \sum_{j=1}^s x_j \cdot w_j$ das allgemeine Element in V mit $w_j \in W$ für $j = 1, \dots, s$ und es sei $g \in G$ beliebig. Wir setzen für $i = 1, \dots, s$

$$J_i := \{j \in \{1, \dots, s\} : gx_j = x_i h_j \in x_i H\}.$$

2 Darstellungstheorie und der random walk auf endlichen Gruppen

Dann gilt:

$$\begin{aligned}
 \tilde{\varphi}(g \cdot v) &= \tilde{\varphi}\left(g \cdot \sum_{j=1}^s x_j \cdot w_j\right) = \tilde{\varphi}\left(\sum_{j=1}^s (gx_j) \cdot w_j\right) \\
 &= \tilde{\varphi}\left(\sum_{i=1}^s \sum_{j \in J_i} (x_i h_j) \cdot w_j\right) = \tilde{\varphi}\left(\sum_{i=1}^s x_i \cdot \underbrace{\sum_{j \in J_i} h_j \cdot w_j}_{\in W}\right) \\
 &= \sum_{i=1}^s x_i \cdot \varphi\left(\sum_{j \in J_i} h_j \cdot w_j\right) = \sum_{i=1}^s x_i \cdot \sum_{j \in J_i} h_j \cdot \varphi(w_j) \\
 &= \sum_{i=1}^s \sum_{j \in J_i} (x_i h_j) \cdot \varphi(w_j) = \sum_{j=1}^s (gx_j) \cdot \varphi(w_j) \\
 &= g \cdot \sum_{j=1}^s x_j \cdot \varphi(w_j) = g \cdot \tilde{\varphi}\left(\sum_{j=1}^s x_j \cdot w_j\right) \\
 &= g \cdot \tilde{\varphi}(v)
 \end{aligned}$$

□

Satz 2.1.47. Sei $\theta : H \rightarrow GL(W)$ eine Darstellung von H . Dann gibt es eine Darstellung $\rho : G \rightarrow GL(V)$ von G , welche von θ induziert wird. Ferner sind je zwei solche induzierten Darstellungen äquivalent.

Beweis: Es ist bequemer, den Satz in der Sprache der G - bzw. H -Moduln zu beweisen!

zur Existenz: Es sei also W ein H -Modul. Wir betrachten die konstruierte direkte Summe

$$V := \underbrace{W \oplus \dots \oplus W}_{s\text{-mal}} = W \times \dots \times W.$$

Jedes $v = (w_1, \dots, w_s) \in V$ lässt sich dann eindeutig als Summe

$$v = \sum_{j=1}^s (0, \dots, 0, \underbrace{w_j}_{j\text{-te Stelle}}, 0, \dots, 0)$$

schreiben. Indem wir linear fortsetzen, genügt es die Operation von G auf den Elementen der Form $v = (0, \dots, 0, w_j, 0, \dots, 0)$ anzugeben, wobei $w_j \in W$ gelte. Ist $g \in G$ und gilt $gx_j = x_i h_j$ mit $h_j \in H$, so sei

2.1 Darstellungstheorie endlicher Gruppen

$$g \cdot (0, \dots, 0, \underbrace{w_j}_{j\text{-te Stelle}}, 0, \dots, 0) := (0, \dots, 0, \underbrace{h_j \cdot w_j}_{i\text{-te Stelle}}, 0, \dots, 0).$$

Es ist nicht schwierig nachzuprüfen, dass dadurch (und durch lineare Fortsetzung) ein G -Modul für die induzierte Darstellung der durch den H -Modul W gegebenen Darstellung definiert wird.

zur Eindeutigkeit: Seien V und V' zwei G -Moduln, die von dem H -Modul W induziert werden. Sei $\varphi : W \hookrightarrow V'$ die kanonische Inklusion. Diese ist offenbar ein Homomorphismus von H -Moduln und kann daher gemäß Lemma 2.1.46 auf eindeutige Weise zu einem G -Modul-Homomorphismus $\tilde{\varphi} : V \rightarrow V'$ fortgesetzt werden. Da $\tilde{\varphi}$ ein Homomorphismus von G -Moduln ist, der φ fortsetzt, liegen die Untervektorräume $\sigma \cdot W \subseteq V'$, $\sigma \in G/H$, im Bild von $\tilde{\varphi}$ und es folgt die Surjektivität von $\tilde{\varphi}$ aus $V' = \bigoplus_{\sigma \in G/H} \sigma \cdot W$. Wegen $\dim_{\mathbb{C}}(V') = [G : H] \cdot \dim_{\mathbb{C}}(W) = \dim_{\mathbb{C}}(V)$ folgt daraus, dass $\tilde{\varphi}$ auch injektiv und somit ein G -Modul-Isomorphismus ist. \square

Aufgrund der Eindeutigkeitsaussage in Satz 2.1.47 macht es Sinn, für die von $\theta : H \rightarrow GL(W)$ induzierte Darstellung die Bezeichnung $Ind_H^G(\theta)$ und für ihren Charakter die Bezeichnung $Ind_H^G(\chi_\theta)$ einzuführen. Im Kontext eines H -Moduls W schreiben wir ebenso $Ind_H^G(W)$ für den dadurch induzierten G -Modul V . Der folgende Satz zeigt, wie man den Charakter einer induzierten Darstellung mit Hilfe des Charakters der ursprünglichen Darstellung berechnen kann.

Satz 2.1.48. *Seien $\theta : H \rightarrow GL(W)$ eine Darstellung von H mit Charakter χ_θ und es werde die Darstellung $\varrho : G \rightarrow GL(V)$ mit Charakter χ_ϱ von θ induziert. Dann gilt für jedes $g \in G$:*

$$\chi_\varrho(g) = \sum_{\substack{j=1 \\ x_j^{-1}gx_j \in H}}^s \chi_\theta(x_j^{-1}gx_j) = \frac{1}{|H|} \sum_{\substack{x \in G \\ x^{-1}gx \in H}} \chi_\theta(x^{-1}gx)$$

Beweis: Wir beweisen zunächst die erste Identität. Es sei $g \in G$ ein beliebiges Element. Um die Spur $\chi_\varrho(g) = \text{Tr}(\varrho(g))$ zu berechnen, betrachten wir die darstellende Matrix A von $\varrho(g)$ bezüglich einer Basis von V , die durch die geordnete Verkettung von Basen der Untervektorräume $W_{\sigma_j} = \varrho(x_j)(W)$, $j = 1, \dots, s$, entsteht. Da die Spur die Summe der Elemente auf der Hauptdiagonalen ist, leisten nur diejenigen x_j einen Beitrag, für die $gx_j \in x_jH$ gilt, denn für $gx_j = x_ih$ mit $h \in H$ ist $\varrho(g)(\varrho(x_j)(W)) = \varrho(x_i)(W)$. Somit erhalten wir

2 Darstellungstheorie und der random walk auf endlichen Gruppen

$$\chi_{\varrho}(g) = \sum_{j:gx_j \in x_jH} \operatorname{Tr}(\varrho(g)|_{W_{\sigma_j}}) = \sum_{\substack{j=1 \\ x_j^{-1}gx_j \in H}}^s \operatorname{Tr}(\varrho(g)|_{W_{\sigma_j}}).$$

Es bleibt somit nur noch zu zeigen, dass für $x_j^{-1}gx_j =: h \in H$ $\operatorname{Tr}(\varrho(g)|_{W_{\sigma_j}}) = \chi_{\theta}(x_j^{-1}gx_j)$ gilt. Die Abbildung $\varrho(x_j)|_W : W \rightarrow W_{\sigma_j}$ ist ein \mathbb{C} -Vektorraum-Isomorphismus und es gilt

$$\varrho(x_j)|_W \circ \theta(h) = \varrho(g)|_{W_{\sigma_j}} \circ \varrho(x_j)|_W,$$

wie man leicht nachrechnet. Daraus folgt

$$\varrho(g)|_{W_{\sigma_j}} = \varrho(x_j)|_W \circ \theta(h) \circ (\varrho(x_j)|_W)^{-1}$$

und damit aus den Eigenschaften der Spur

$$\operatorname{Tr}(\varrho(g)|_{W_{\sigma_j}}) = \operatorname{Tr}(\theta(h)) = \chi_{\theta}(x_j^{-1}gx_j),$$

wie gewünscht. Nun zur zweiten Identität: Ist $x \in G$ mit $x = x_jh$ und $h \in H$ und gilt $x^{-1}gx \in H$, so folgt

$$\chi_{\theta}(x^{-1}gx) = \chi_{\theta}(h^{-1}(x_j^{-1}gx_j)h) = \chi_{\theta}(x_j^{-1}gx_j),$$

denn χ_{θ} ist eine Klassenfunktion. Wegen $|x_jH| = |H|$ für $j = 1, \dots, s$ folgt die Behauptung. □

Satz 2.1.49 (Frobenius-Reziprozität). *Seien φ bzw. χ Charaktere von Darstellungen von H bzw. von G . Dann gilt:*

$$\langle \operatorname{Ind}_H^G(\varphi), \chi \rangle_G = \langle \varphi, \operatorname{Res}_H^G(\chi) \rangle_H,$$

wobei wir durch den Index zum Ausdruck bringen, bzgl. welcher Gruppe das Skalarprodukt berechnet wird.

Beweis: Wir haben

$$\begin{aligned}
 \langle \text{Ind}_H^G(\varphi), \chi \rangle_G &= \frac{1}{|G|} \sum_{g \in G} \text{Ind}_H^G(\varphi)(g) \chi(g^{-1}) \\
 &\stackrel{\text{Satz 2.1.48}}{=} \frac{1}{|G|} \frac{1}{|H|} \sum_{g \in G} \sum_{\substack{x \in G \\ x^{-1}gx \in H}} \varphi(x^{-1}gx) \chi(g^{-1}) \\
 &= \frac{1}{|G|} \frac{1}{|H|} \sum_{x \in G} \sum_{\substack{g \in G \\ x^{-1}gx \in H}} \varphi(x^{-1}gx) \chi(g^{-1}) \\
 &\stackrel{y:=x^{-1}gx}{=} \frac{1}{|G|} \frac{1}{|H|} \sum_{x \in G} \sum_{y \in H} \varphi(y) \chi(xy^{-1}x^{-1}) \\
 &\stackrel{\chi \in K(G)}{=} \frac{1}{|G|} \frac{1}{|H|} \sum_{x \in G} \sum_{y \in H} \varphi(y) \chi(y^{-1}) \\
 &= \frac{1}{|H|} \sum_{y \in H} \varphi(y) \chi(y^{-1}) \\
 &= \langle \varphi, \text{Res}_H^G(\chi) \rangle_H
 \end{aligned}$$

□

Der nächste Satz sagt aus, dass das Bilden der induzierten Darstellung *transitiv* ist.

Satz 2.1.50 (Transitivität der Induktion von Darstellungen). *Es sei $H \subseteq G \subseteq K$ eine Kette von Untergruppen. Ist $\theta : H \rightarrow GL(W)$ eine Darstellung von H , so gilt*

$$\text{Ind}_H^K(\theta) \cong \text{Ind}_G^K(\text{Ind}_H^G(\theta)).$$

und

$$\text{Ind}_H^K(\chi_\theta)(y) = \frac{1}{|G|} \frac{1}{|H|} \sum_{\substack{z \in K \\ z^{-1}yz \in G}} \left(\sum_{\substack{x \in G \\ x^{-1}z^{-1}yzx \in H}} \chi_\theta(x^{-1}z^{-1}yzx) \right)$$

für jedes $y \in K$.

Beweis: Es seien $s := [G : H]$ und $r := [K : G]$ und es seien $1 = x_1, x_2, \dots, x_s \in G$ ein vollständiges Vertretersystem für die Linksnebenklassen von H in G sowie $1 = y_1, y_2, \dots, y_r \in K$ ein vollständiges Repräsentantensystem für die Linksnebenklassen von G in K . Dann gilt $[K : H] = r \cdot s$ und genauer ist das System $y_i x_j$, $i = 1, \dots, r$, $j = 1, \dots, s$ ein vollständiges Vertretersystem für die Linksnebenklassen von H in K . Gibt es nämlich ein $h \in H$ mit $y_i x_j = y_l x_k h$, so folgt

$$y_i = y_l \underbrace{(x_k h x_j^{-1})}_{\in G} \in y_l G,$$

also $i = l$. Dann folgt $x_j = x_k h$ und daraus dann $j = k$. Somit sind die Linksnebenklassen $y_i x_j H$, $i = 1, \dots, r$, $j = 1, \dots, s$ paarweise verschieden. Ist umgekehrt $y \in K$ beliebig, so wähle man zunächst $i \in \{1, \dots, r\}$ und $g \in G$ mit $y = y_i g$ und dann ein $j \in \{1, \dots, s\}$ sowie ein $h \in H$ mit $g = x_j h$. Dann folgt $y = y_i g = y_i x_j h \in y_i x_j H$. Wir werden zeigen, dass $\varphi := \text{Ind}_G^K(\text{Ind}_H^G(\theta)) : K \rightarrow GL(V)$ die definierenden Eigenschaften der induzierten Darstellung $\text{Ind}_H^K(\theta)$ erfüllt, so dass die erste Behauptung aus der Eindeutigkeitsaussage in Satz 2.1.47 folgt. Es sei $\varrho := \text{Ind}_H^G(\theta) : G \rightarrow GL(U)$. Dann gilt $\varphi = \text{Ind}_G^K(\varrho)$ sowie $U = \bigoplus_{j=1}^s \varrho(x_j)(W)$. Ferner folgt:

$$\begin{aligned} V = \bigoplus_{i=1}^r \varphi(y_i)(U) &= \bigoplus_{i=1}^r \varphi(y_i) \left(\bigoplus_{j=1}^s \varrho(x_j)(W) \right) \\ &= \bigoplus_{i=1}^r \bigoplus_{j=1}^s \varphi(y_i)(\varrho(x_j)(W)) \end{aligned}$$

Man sieht nun leicht ein, dass die Untervektorräume $W_{i,j} := \varphi(y_i)(\varrho(x_j)(W))$ für jedes $y \in K$ unter der Abbildung $\varphi(y)$ permutiert werden und dass $W_{1,1} = W$ ist. Weiter gilt nach Definition von φ und ϱ für alle $h \in H$ und $w \in W \subseteq U \subseteq V$:

$$\varphi(h)(w) \stackrel{h \in G}{=} \varrho(h)(w) \stackrel{h \in H}{=} \theta(h)(w),$$

so dass φ tatsächlich eine von θ induzierte Darstellung von K ist. Für die zweite Behauptung bemerken wir zunächst, dass nach Satz 2.1.48 für alle $g \in G$ gilt:

$$\chi_\varrho(g) = \text{Ind}_H^G(\chi_\theta)(g) = \frac{1}{|G|} \sum_{\substack{x \in G \\ x^{-1}gx \in H}} \chi_\theta(x^{-1}gx) \quad (2.7)$$

Aus Gleichung (2.7) erhalten wir wieder mit Satz 2.1.48 für alle $y \in K$

$$\begin{aligned} \text{Ind}_H^K(\chi_\theta)(y) &= \frac{1}{|G|} \sum_{\substack{z \in K \\ z^{-1}yz \in G}} \chi_\varrho(z^{-1}yz) \\ &= \frac{1}{|G|} \frac{1}{|H|} \sum_{\substack{z \in K \\ z^{-1}yz \in G}} \left(\sum_{\substack{x \in G \\ x^{-1}z^{-1}yzx \in H}} \chi_\theta(x^{-1}z^{-1}yzx) \right), \end{aligned}$$

2.1 Darstellungstheorie endlicher Gruppen

also die Behauptung.

□

2.2 Darstellungstheorie der symmetrischen Gruppe \mathbb{S}_n

In diesem Abschnitt wollen wir die allgemeinen Resultate des letzten Abschnitts am konkreten Beispiel der symmetrischen Gruppe

$$\mathbb{S}_n := \{\pi : \{1, \dots, n\} \rightarrow \{1, \dots, n\} \mid \pi \text{ bijektiv}\}$$

betrachten und insbesondere alle irreduziblen Darstellungen der \mathbb{S}_n bestimmen. Wir wissen aus Korollar 2.1.39, dass es genauso viele irreduzible Darstellungen wie Konjugationsklassen von \mathbb{S}_n gibt. Die Konjugationsklasse einer Permutation ist durch ihren *Zykeltyp* bestimmt:

Bekanntlich lässt sich eine Permutation $\pi \in \mathbb{S}_n$ bis auf die Reihenfolge eindeutig als Produkt paarweise Element-fremder Zyklen darstellen:

$$\pi = (x_1, \dots, x_{n_1}) \circ (x_{n_1+1}, \dots, x_{n_2}) \circ \dots \circ (x_{n_{r-1}+1}, \dots, x_{n_r}) \quad (2.8)$$

Dabei ist ein Zykel $\sigma = (y_1, \dots, y_k)$ eine Permutation von $\{1, \dots, n\}$ mit

$$\begin{aligned} \sigma(y_i) &= y_{i+1} \quad \text{für } 1 \leq i \leq k-1 \\ \sigma(y_k) &= y_1 \\ \sigma(y) &= y \quad \text{für } y \in \{1, \dots, n\} \setminus \{y_1, \dots, y_k\} \end{aligned}$$

Ist π wie in (2.8) und $\varphi \in \mathbb{S}_n$ beliebig, so ist

$$\begin{aligned} &\varphi \circ \pi \circ \varphi^{-1} \quad (2.9) \\ &= (\varphi(x_1), \dots, \varphi(x_{n_1})) \circ (\varphi(x_{n_1+1}), \dots, \varphi(x_{n_2})) \circ \dots \circ (\varphi(x_{n_{r-1}+1}), \dots, \varphi(x_{n_r})) \end{aligned}$$

die kanonische Zykelzerlegung von $\sigma := \varphi \circ \pi \circ \varphi^{-1}$. Aus dieser einfachen Beobachtung folgt, dass zwei Permutationen $\pi, \sigma \in \mathbb{S}_n$ genau dann konjugiert sind, wenn sie denselben Zykeltyp im folgenden Sinne haben:

Definition 2.2.1. Die Permutation $\pi \in \mathbb{S}_n$ hat den Zykeltyp

$$(1^{m_1}, 2^{m_2}, \dots, n^{m_n}),$$

wenn in der kanonischen Zykelzerlegung von π genau m_1 Zyklen der Länge 1, genau m_2 Zyklen der Länge 2, ..., und genau m_n Zyklen der Länge n vorkommen.

2.2 Darstellungstheorie der symmetrischen Gruppe \mathbb{S}_n

Hat π den Zykeltyp $(1^{m_1}, 2^{m_2}, \dots, n^{m_n})$, so gilt offenbar

$$\sum_{j=1}^n j \cdot m_j = \sum_{\substack{j=1 \\ m_j \neq 0}}^n j \cdot m_j = n.$$

Sind $1 \leq j_1 < \dots < j_k \leq n$ die Zahlen mit $m_{j_j} \neq 0$, so ist

$$\lambda := (\underbrace{j_k, \dots, j_k}_{m_{j_k}}, \underbrace{j_{k-1}, \dots, j_{k-1}}_{m_{j_{k-1}}}, \dots, \underbrace{j_1, \dots, j_1}_{m_{j_1}})$$

eine *Partition* von n im folgenden Sinne:

Definition 2.2.2. *Unter einer Partition von $n \in \mathbb{N}$ verstehen wir ein Tupel*

$$\lambda = (\lambda_1, \dots, \lambda_l) \in \mathbb{N}^l$$

positiver ganzer Zahlen mit

$$\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_l \quad \text{und} \quad \sum_{i=1}^l \lambda_i = n.$$

Gegebenenfalls schreiben wir $\lambda \vdash n$ für den Sachverhalt, dass λ eine Partition von n ist.

Proposition 2.2.3. *Es gibt eine Bijektion zwischen der Menge aller Partitionen von n und der Menge $\text{Irr}(\mathbb{S}_n)$ aller paarweise nicht-äquivalenten, irreduziblen Darstellungen der \mathbb{S}_n .*

Beweis: Wir wissen, dass es genau $|\text{Irr}(\mathbb{S}_n)|$ Konjugationsklassen von \mathbb{S}_n gibt. Wir haben gesehen, dass jeder Konjugationsklasse über ihren Zykeltyp genau eine Partition von \mathbb{S}_n entspricht. Daraus folgt unmittelbar die Behauptung. □

Korollar 2.2.4. *Die Charaktere der Darstellungen der symmetrischen Gruppe \mathbb{S}_n nehmen nur reelle Werte an.*

Beweis: Sei $\pi \in \mathbb{S}_n$ gegeben durch (2.8). Dann gilt

$$\pi^{-1} = (x_{n_1}, x_{n_1-1}, \dots, x_1) \circ (x_{n_2}, x_{n_2-1}, \dots, x_{n_1+1}) \circ \dots \circ (x_{n_r}, x_{n_r-1}, \dots, x_{n_{r-1}+1}),$$

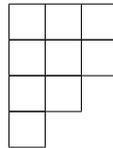
insbesondere haben also π und π^{-1} denselben Zykel-Typ und sind folglich konjugiert. Aus Proposition 2.1.24 (b) und (c) ergibt sich somit für einen beliebigen Charakter χ einer Darstellung der \mathbb{S}_n

$$\overline{\chi(\pi)} = \chi(\pi^{-1}) = \chi(\pi),$$

also $\chi(\pi) \in \mathbb{R}$. □

Es ist alles andere als offensichtlich, wie man einer gegebenen Partition $\lambda \vdash n$ eine irreduzible Darstellung zuordnen soll. Wir wollen dies im Folgenden beschreiben. Unsere Darstellung folgt im Wesentlichen derjenigen in [Sag01], aber teilweise auch derjenigen in [Dia88].

Zunächst wollen wir einer Partition $\lambda = (\lambda_1, \dots, \lambda_l) \vdash n$ ein geometrisches Objekt, nämlich ihr sogenanntes *Young-Diagramm* oder auch *Ferrer-Diagramm* zuordnen. Dies ist einfach eine Aneinanderreihung von n Kästchen mit genau λ_1 Kästchen in der ersten Zeile, genau λ_2 Kästchen in der zweiten Zeile, ... und genau λ_l Kästchen in der l -ten Zeile. Ist beispielsweise $\lambda = (3, 3, 2, 1) \vdash 9$, so ist das zugehörige Young-Diagramm gegeben durch:



Die Zuordnung

$$\{ \text{Partitionen von } n \} \rightarrow \{ \text{Young-Diagramme mit } n \text{ Kästchen} \}$$

ist offenbar bijektiv, das heißt wir können (und werden) eine Partition λ mit ihrem Young-Diagramm identifizieren. Wenn man in ein Young-Diagramm mit n Kästchen die Zahlen $1, \dots, n$ einträgt, so erhält man ein sogenanntes *Young-Tableau*:

Definition 2.2.5. Sei $\lambda = (\lambda_1, \dots, \lambda_l) \vdash n$. Unter einem Young-Tableau oder einfach Tableau t der Form λ verstehen wir ein Young-Diagramm zu λ , in dem die Zahlen $1, \dots, n$ eingetragen sind. Wir schreiben auch $sh(t) = \lambda$ für den Sachverhalt, dass t ein Young-Tableau der Form λ ist.

Offenbar gibt es zu jeder Partition λ von n genau $n!$ verschiedene λ -Tableaux. Ein Young-Tableau t der Form $\lambda = (3, 3, 2, 1)$ ist beispielsweise gegeben durch

$$t = \begin{array}{|c|c|c|} \hline 2 & 7 & 5 \\ \hline 1 & 4 & 3 \\ \hline 9 & 6 & \\ \hline 8 & & \\ \hline \end{array} \tag{2.10}$$

Um die gewünschten \mathbb{S}_n -Moduln konstruieren zu können, müssen wir zunächst zu Äquivalenzklassen von Tableaux, sogenannten *tabloids*, übergehen:

Definition 2.2.6. Sei $\lambda \vdash n$. Zwei Young-Tableaux s, t mit

$$sh(s) = sh(t) = \lambda$$

heißen Zeilen-äquivalent, in Zeichen

$$s \sim t,$$

wenn in jeder Zeile von s dieselben Zahlen eingetragen sind, wie in der entsprechenden Zeile von t . Dies definiert eine Äquivalenzrelation auf der Menge aller Young-Tableaux der Form $\lambda \vdash n$. Die Äquivalenzklassen heißen tabloids und es bezeichne stets $\{t\}$ die Äquivalenzklasse des Tableaus t .

Beispiel 2.2.7. Ist wieder $\lambda = (3, 3, 2, 1)$, so sind die Tableaux

$$t = \begin{array}{|c|c|c|} \hline 2 & 7 & 5 \\ \hline 1 & 4 & 3 \\ \hline 9 & 6 & \\ \hline 8 & & \\ \hline \end{array} \quad \text{und} \quad s = \begin{array}{|c|c|c|} \hline 7 & 2 & 5 \\ \hline 3 & 1 & 4 \\ \hline 6 & 9 & \\ \hline 8 & & \\ \hline \end{array} \quad (2.11)$$

Zeilen-äquivalent.

Man veranschaulicht tabloids dadurch, dass man nur horizontale Linien zwischen die Zeilen einzeichnet, so dass etwa für Beispiel 2.2.7 gilt:

$$\{s\} = \{t\} = \begin{array}{c} \hline 2 \quad 5 \quad 7 \\ \hline 1 \quad 3 \quad 4 \\ \hline 6 \quad 9 \\ \hline 8 \\ \hline \end{array}$$

Für ein Tableau t bezeichne (wie für eine Matrix auch) $t_{i,j}$ den Eintrag in der i -ten Zeile und j -ten Spalte von t , so dass also $t = (t_{i,j})_{i,j}$ gilt. Die Gruppe \mathbb{S}_n operiert auf der Menge aller λ -Tableaus vermöge

$$\pi \cdot t := (\pi(t_{i,j}))_{i,j}.$$

Sind nun s und t Zeilen-äquivalent, so auch $\pi \cdot t$ und $\pi \cdot s$, was man leicht einsieht. Somit überträgt sich die Operation von \mathbb{S}_n auch auf die λ -tabloids via

$$\pi \cdot \{t\} := \{\pi \cdot t\}.$$

Wir haben somit nun eine Gruppen-Operation der Gruppe \mathbb{S}_n auf der Menge aller λ -tabloids und können daher gemäß Definition 2.1.6 den zugehörigen Permutationsmodul bilden. Von nun an sei bis auf Weiteres eine Partition λ von n fixiert und es stets t_1, \dots, t_k ein vollständiges Vertretersystem für die Zeilen-Äquivalenz, das heißt

$$\{t_1\}, \dots, \{t_k\}$$

sind alle verschiedenen tabloids der Form λ .

Definition 2.2.8. *Der zu der Operation von \mathbb{S}_n auf der Menge aller tabloids der Form $\lambda \vdash n$ gehörige Permutationsmodul*

$$M^\lambda := \mathbb{C}[\{t_1\}, \dots, \{t_k\}]$$

heißt der zur Partition $\lambda \vdash n$ gehörige Permutationsmodul.

Die \mathbb{S}_n -Moduln M^λ , $\lambda \vdash n$ sind leider im Allgemeinen nicht irreduzibel, wie wir auch in den folgenden Beispielen sehen werden.

Beispiel 2.2.9. (a) *Sei $\lambda = (n)$. Dann ist*

$$\overline{1 \quad 2 \quad \dots \quad n}$$

das einzige λ -tabloid, so dass der eindimensional Modul $M^{(n)}$ irreduzibel ist und die triviale Darstellung der \mathbb{S}_n trägt.

(b) *Sei $\lambda = (\underbrace{1, \dots, 1}_n)$. Dann bedeutet Zeilen-Äquivalenz nichts anderes als Gleichheit und ein tabloid $\{t\}$ kann mit dem Tableau t identifiziert werden. In diesem sind einfach alle Zahlen von 1 bis n untereinander eingetragen und bilden somit eine (als Spalte geschriebene) Permutation von $\{1, \dots, n\}$. Die Aktion von \mathbb{S}_n auf solch einer Spalte entspricht der Verkettung von Transpositionen, so dass $M^{(1, \dots, 1)}$ ein Modul für die reguläre Darstellung der \mathbb{S}_n und insbesondere nicht irreduzibel ist, falls $n \geq 2$ ist.*

(c) *Ist $\lambda = (n-1, 1)$, so ist jedes λ -tabloid schon eindeutig durch den Eintrag in der zweiten Zeile festgelegt. Dafür gibt es genau n Möglichkeiten. Es ist nicht schwierig zu sehen, dass $M^{(n-1, 1)}$ ein Modul für die definierende Darstellung der \mathbb{S}_n ist (vgl. etwa [Sag01]).*

Ist t ein Young-Tableau mit $sh(t) = \lambda := (\lambda_1, \dots, \lambda_l) \vdash n$, so gibt es offenbar genau

$$\lambda! := \lambda_1! \cdot \lambda_2! \cdot \dots \cdot \lambda_l!$$

zu t Zeilen-äquivalente Tableaus, das heißt es gilt

$$|\{t\}| = \lambda!$$

2.2 Darstellungstheorie der symmetrischen Gruppe \mathbb{S}_n

Als Äquivalenzklassen zu einer Äquivalenzrelation bilden die λ -tabloids eine Partition (im Sinne der Mengenlehre!) von der Menge aller λ -Tableaux, von denen es offensichtlich genau $n!$ Stück gibt. Somit ist die Anzahl k aller λ -tabloids gegeben durch

$$k = \frac{n!}{\lambda!} = \frac{n!}{\lambda_1! \cdot \lambda_2! \cdot \dots \cdot \lambda_l!}.$$

Sind weiter t und s λ -Tableaux, so gibt es (genau) ein $\pi \in \mathbb{S}_n$ mit $\pi \cdot t = s$, also gilt erst recht

$$\pi \cdot \{t\} = \{\pi \cdot t\} = \{s\},$$

das heißt der Modul M^λ ist zyklisch im folgenden Sinne:

Definition 2.2.10. Sei G eine endliche Gruppe. Ein G -Modul (V, \cdot) heißt zyklisch mit Erzeuger $v \in V$, falls gilt:

$$V = \mathbb{C}[G \cdot v] := \left\{ \sum_{i=1}^s \alpha_i (g_i \cdot v) \mid s \in \mathbb{N}_0, \alpha_i \in \mathbb{C}, g_i \in G \text{ für } i = 1, \dots, s \right\}$$

Für die Konstruktion der irreduziblen Untermoduln der M^λ , $\lambda \vdash n$, benötigen wir die folgende Halbordnung „ \triangleright “ auf der Menge aller Partitionen λ von n .

Definition 2.2.11 (Dominanz-Ordnung). Seien $\lambda = (\lambda_1, \dots, \lambda_l)$ und $\mu = (\mu_1, \dots, \mu_m)$ Partitionen von $n \in \mathbb{N}$. Wir sagen, dass λ die Partition μ dominiert, in Zeichen

$$\lambda \triangleright \mu,$$

wenn für alle $1 \leq i \leq \max(l, m)$ die Ungleichung

$$\lambda_1 + \lambda_2 + \dots + \lambda_i \geq \mu_1 + \mu_2 + \dots + \mu_i$$

gilt. Dabei werden die Partitionen λ bzw. μ falls nötig durch Nullen am Ende ergänzt.

Es ist nicht schwierig zu sehen, dass dadurch in der Tat eine Halbordnung auf den Partitionen von n gegeben ist.

Beispiel 2.2.12. Man betrachte die Partitionen $\lambda = (3, 2, 1)$ und $\mu = (2, 2, 1, 1)$ von 6. Dann gilt $\lambda \triangleright \mu$.

$$\lambda \hat{=} \begin{array}{|c|c|c|} \hline \square & \square & \square \\ \hline \square & \square & \\ \hline \square & & \\ \hline \end{array}, \quad \mu \hat{=} \begin{array}{|c|c|} \hline \square & \square \\ \hline \square & \square \\ \hline \square & \\ \hline \square & \\ \hline \end{array}$$

2 Darstellungstheorie und der random walk auf endlichen Gruppen

Anschaulich bedeutet $\lambda \supseteq \mu$, dass das Young-Diagramm zu λ kurz und dick ist, während das zu μ lang und dünn ist.

Das folgende Lemma wird sich als sehr nützlich erweisen:

Lemma 2.2.13 (Dominanz-Lemma für Partitionen). *Seien λ und μ Partitionen von n und seien s, t Young-Tableaux mit*

$$sh(t) = \lambda \quad \text{und} \quad sh(s) = \mu.$$

Falls für jede Zeilenzahl i die Zahlen in der i -ten Zeile von s alle in verschiedenen Spalten von t stehen, dann gilt schon

$$\lambda \supseteq \mu.$$

Beweis: Die Voraussetzung impliziert, dass wir für jedes i die Elemente in jeder Spalte von t so umordnen können, dass alle Elemente aus den ersten i Zeilen von s in den ersten i Zeilen von t stehen. Somit gilt für jedes i :

$$\begin{aligned} \lambda_1 + \lambda_2 + \dots + \lambda_i &= \text{Anzahl der Elemente in den ersten } i \text{ Zeilen von } t \\ &\geq \text{Anzahl der Elemente in den ersten } i \text{ Zeilen von } s \\ &= \mu_1 + \mu_2 + \dots + \mu_i \end{aligned}$$

□

Um für alle $\lambda \vdash n$ den irreduziblen \mathbb{S}_n -Untermodul S^λ von M^λ zu konstruieren, müssen wir uns die folgenden Untergruppen von \mathbb{S}_n ansehen, welche nur innerhalb der Zeilen bzw. Spalten eines gegebenen Tableaus t der Gestalt λ Vertauschungen bewirken. Dazu wollen wir vereinbaren, für eine beliebige Menge A mit $\mathbb{S}(A)$ die *Permutationsgruppe* von A zu bezeichnen, das heißt

$$\mathbb{S}(A) := \{ f : A \rightarrow A \mid f \text{ bijektiv} \}.$$

Definition 2.2.14. *Sei t ein Young-Tableau mit $\lambda := sh(t) \vdash n$. Seien mit R_1, R_2, \dots, R_l die Zeilen und mit C_1, C_2, \dots, C_m die Spalten von t bezeichnet.*

(a) *Die Untergruppe (!) von \mathbb{S}_n aller Permutationen π mit*

$$\pi(R_i) = R_i \quad \text{für } i = 1, \dots, l$$

heißt der Zeilen-Stabilisator von t und wird mit R_t bezeichnet.

2.2 Darstellungstheorie der symmetrischen Gruppe \mathbb{S}_n

(b) Die Untergruppe von \mathbb{S}_n aller Permutationen π mit

$$\pi(C_j) = C_j \quad \text{für } j = 1, \dots, m$$

heißt der Spalten-Stabilisator von t und wird mit C_t bezeichnet.

Bemerkung 2.2.15. (i) Es ist leicht zu sehen, dass

$$R_t \cong \mathbb{S}(R_1) \times \mathbb{S}(R_2) \times \dots \times \mathbb{S}(R_l) \cong \mathbb{S}_{\lambda_1} \times \mathbb{S}_{\lambda_2} \times \dots \times \mathbb{S}_{\lambda_l}$$

gilt, wobei die zweite Isomorphie unmittelbar aus $|R_i| = \lambda_i$ für $i = 1, \dots, l$ folgt. Für die erste Isomorphie ist die Abbildung

$$\varphi : \mathbb{S}(R_1) \times \mathbb{S}(R_2) \times \dots \times \mathbb{S}(R_l) \rightarrow R_t, \quad (\pi_1, \dots, \pi_l) \mapsto \pi_1 \circ \pi_2 \circ \dots \circ \pi_l$$

ein natürlicher Isomorphismus von Gruppen (Dazu beachte man, dass man jedes $\pi_i \in \mathbb{S}(R_i)$ auch als Element von \mathbb{S}_n und somit von R_t auffassen kann, indem man die Elemente aus $\{1, \dots, n\} \setminus R_i$ festlässt).

(ii) Ebenso sieht man, dass

$$C_t \cong \mathbb{S}(C_1) \times \mathbb{S}(C_2) \times \dots \times \mathbb{S}(C_m) \quad \text{ist.}$$

(iii) Mit Hilfe des Zeilen-Stabilisators R_t können wir das tabloid $\{t\}$ zu t nun auch als

$$\{t\} = R_t \cdot t := \{\pi \cdot t \mid \pi \in R_t\}$$

ausdrücken.

Definition 2.2.16. (i) Für eine beliebige Teilmenge $H \subseteq \mathbb{S}_n$ seien

$$H^+ := \sum_{\pi \in H} \pi \in \mathbb{C}[\mathbb{S}_n]$$

und

$$H^- := \sum_{\pi \in H} \text{sgn}(\pi) \pi \in \mathbb{C}[\mathbb{S}_n]$$

Für $H = \{\pi\}$ schreiben wir einfach π^+ bzw. π^- statt $\{\pi\}^+$ bzw. $\{\pi\}^-$.

2 Darstellungstheorie und der random walk auf endlichen Gruppen

(ii) Ist t ein Young-Tableau, so definieren wir

$$\kappa_t := C_t^- = \sum_{\pi \in C_t} \operatorname{sgn}(\pi)\pi \in \mathbb{C}[\mathbb{S}_n].$$

Lemma 2.2.17. Hat das Tableau t die Spalten C_1, C_2, \dots, C_m , so gilt

$$\kappa_t = \kappa_{C_1} \circ \kappa_{C_2} \circ \dots \circ \kappa_{C_m} \quad (\text{Produkt in } \mathbb{C}[\mathbb{S}_n]).$$

Beweis: Es gilt nach Bemerkung 2.2.15 (ii)

$$\begin{aligned} \kappa_t &= \sum_{\pi \in C_t} \operatorname{sgn}(\pi)\pi = \sum_{\pi_1 \in \mathbb{S}(C_1)} \dots \sum_{\pi_m \in \mathbb{S}(C_m)} \operatorname{sgn}(\pi_1 \circ \dots \circ \pi_m)(\pi_1 \circ \dots \circ \pi_m) \\ &= \sum_{\pi_1 \in \mathbb{S}(C_1)} \dots \sum_{\pi_m \in \mathbb{S}(C_m)} \operatorname{sgn}(\pi_1) \cdot \dots \cdot \operatorname{sgn}(\pi_m)(\pi_1 \circ \dots \circ \pi_m) \\ &= \left(\sum_{\pi_1 \in \mathbb{S}(C_1)} \operatorname{sgn}(\pi_1)\pi_1 \right) \circ \left(\sum_{\pi_2 \in \mathbb{S}(C_2)} \operatorname{sgn}(\pi_2)\pi_2 \right) \circ \dots \circ \left(\sum_{\pi_m \in \mathbb{S}(C_m)} \operatorname{sgn}(\pi_m)\pi_m \right) \\ &= \kappa_{C_1} \circ \kappa_{C_2} \circ \dots \circ \kappa_{C_m}. \end{aligned}$$

□

Definition 2.2.18. Für ein Young-Tableau t mit $sh(t) = \lambda \vdash n$ definieren wir das zugehörige polytabloid e_t durch

$$e_t := \kappa_t \cdot \{t\} \in M^\lambda.$$

Man beachte, dass das Element $e_t = \kappa_t \cdot \{t\}$ wohldefiniert ist, da wir in der Passage nach Bemerkung 2.1.7 gesehen haben, dass M^λ auf natürliche Weise ein $\mathbb{C}[\mathbb{S}_n]$ -Modul ist! Weiter beachte man, dass das polytabloid e_t (trotz des Namens) wirklich vom Tableau t und nicht bloß vom zugehörigen tabloid $\{t\}$ abhängt, denn κ_t ist nicht invariant unter Zeilen-Äquivalenz.

Lemma 2.2.19. Sei t ein Young-Tableau mit $\lambda = sh(t) \vdash n$ und sei $\pi \in \mathbb{S}_n$. Dann gilt:

$$(i) \quad R_{\pi \cdot t} = \pi \circ R_t \circ \pi^{-1}$$

$$(ii) \quad C_{\pi \cdot t} = \pi \circ C_t \circ \pi^{-1}$$

$$(iii) \quad \kappa_{\pi \cdot t} = \pi \circ \kappa_t \circ \pi^{-1}$$

2.2 Darstellungstheorie der symmetrischen Gruppe \mathbb{S}_n

(iv) $e_{\pi \cdot t} = \pi \cdot e_t$

Beweis: zu (i): Für $\sigma \in \mathbb{S}_n$ gilt:

$$\begin{aligned} \sigma \in R_{\pi \cdot t} &\Leftrightarrow \sigma\{\pi \cdot t\} = \{\pi \cdot t\} \\ &\Leftrightarrow (\sigma \circ \pi) \cdot \{t\} = \pi \cdot \{t\} \\ &\Leftrightarrow (\pi^{-1} \circ \sigma \circ \pi) \cdot \{t\} = \{t\} \\ &\Leftrightarrow (\pi^{-1} \circ \sigma \circ \pi) \in R_t \\ &\Leftrightarrow \sigma \in \pi \circ R_t \circ \pi^{-1} \end{aligned}$$

zu (ii): Dies zeigt man wie in (i), indem man den Begriff der Spalten-Äquivalenz für Tableaux einführt!

zu (iii): Nach (ii) hat man

$$\begin{aligned} \kappa_{\pi \cdot t} &= \sum_{\sigma \in C_{\pi \cdot t}} \operatorname{sgn}(\sigma) \sigma = \sum_{\sigma \in C_t} \operatorname{sgn}(\pi \circ \sigma \circ \pi^{-1}) \pi \circ \sigma \circ \pi^{-1} \\ &= \sum_{\sigma \in C_t} \operatorname{sgn}(\sigma) \pi \circ \sigma \circ \pi^{-1} = \pi \circ \left(\sum_{\sigma \in C_t} \operatorname{sgn}(\sigma) \right) \circ \pi^{-1} \\ &= \pi \circ \kappa_t \circ \pi^{-1} \end{aligned}$$

zu (iv): Nach (iii) gilt

$$\begin{aligned} e_{\pi \cdot t} &= \kappa_{\pi \cdot t} \circ \{\pi \cdot t\} = (\pi \circ \kappa_t) \cdot \{t\} \\ &= \pi \cdot e_t \end{aligned}$$

□

Nun können wir die sogenannten Specht-Moduln, die sämtliche paarweise nicht-äquivalenten irreduziblen \mathbb{S}_n -Moduln bilden, definieren:

Definition 2.2.20. Für eine Partition λ von n heißt der kleinste \mathbb{S}_n -Untermodul von M^λ , der alle polytabloids e_t mit $sh(t) = \lambda$ enthält, der zugehörige Specht-Modul. Dieser wird mit S^λ bezeichnet.

Wegen $\pi \cdot e_t = e_{\pi \cdot t}$ folgt zunächst die Darstellung

$$S^\lambda = \left\{ \sum_{\substack{t \text{ Tableau} \\ sh(t)=\lambda}} \lambda_t e_t \mid \lambda_t \in \mathbb{C} \right\},$$

2 Darstellungstheorie und der random walk auf endlichen Gruppen

das heißt S^λ ist gleich dem von den polytabloids erzeugten \mathbb{C} -Untervektorraum von M^λ . Ferner ist wegen $e_{\pi \cdot t} = \pi \cdot e_t$ für $\pi \in \mathbb{S}_n$ und, da es zu jedem weiteren λ -Tableau s ein $\pi \in \mathbb{S}_n$ gibt mit $\pi \cdot t = s$, der \mathbb{S}_n -Modul S^λ zyklisch. Wir können also festhalten:

Proposition 2.2.21. *Sei $\lambda \vdash n$. Dann ist der zugehörige Specht-Modul S^λ zyklisch und wird als \mathbb{S}_n -Modul von jedem polytabloid e_t , wobei t ein Tableau mit $sh(t) = \lambda$ ist, erzeugt. Als \mathbb{C} -Untervektorraum von M^λ wird er von sämtlichen polytabloids zu den Tableaux der Form λ erzeugt.*

□

Beispiel 2.2.22. (a) *Sei $\lambda = (n)$. Dann war M^λ nach Beispiel 2.2.9 der triviale Modul und somit ist wegen $S^\lambda \neq \{0\}$ auch $S^\lambda = M^\lambda$ ein Modul für die triviale Darstellung und daher insbesondere irreduzibel.*

(b) *Sei $\lambda = (\underbrace{1, \dots, 1}_n)$ und sei*

$$t := \begin{array}{|c|} \hline 1 \\ \hline 2 \\ \hline \vdots \\ \hline n \\ \hline \end{array}$$

ein fixiertes λ -Tableau. Dann gilt

$$\kappa_t = \sum_{\sigma \in \mathbb{S}_n} \text{sgn}(\sigma) \sigma$$

und somit

$$e_t = \kappa_t \cdot \{t\} = \sum_{\sigma \in \mathbb{S}_n} \text{sgn}(\sigma) \sigma \cdot \{t\}.$$

Nach Lemma 2.2.19 (iv) erhalten wir für jedes $\pi \in \mathbb{S}_n$:

$$\begin{aligned} e_{\pi \cdot t} &= \pi \cdot e_t = \pi \cdot \sum_{\sigma \in \mathbb{S}_n} \text{sgn}(\sigma) \sigma \cdot \{t\} \\ &= \sum_{\sigma \in \mathbb{S}_n} \text{sgn}(\sigma) (\pi \circ \sigma) \cdot \{t\} = \sum_{\sigma \in \mathbb{S}_n} \text{sgn}(\pi^{-1} \circ \pi \circ \sigma) (\pi \circ \sigma) \cdot \{t\} \\ &= \text{sgn}(\pi^{-1}) \sum_{\varphi \in \mathbb{S}_n} \text{sgn}(\varphi) \varphi \cdot \{t\} \\ &= \text{sgn}(\pi) e_t, \end{aligned}$$

2.2 Darstellungstheorie der symmetrischen Gruppe \mathbb{S}_n

das heißt es ist jedes weitere polytabloid ein skalares Vielfaches von e_t und somit erhalten wir, dass

$$S^{(1, \dots, 1)} = \mathbb{C} e_t$$

eindimensional und somit irreduzibel ist. Die Aktion von \mathbb{S}_n auf $S^{(1, \dots, 1)}$ ist gegeben durch

$$\begin{aligned} \pi \cdot (\alpha e_t) &= \alpha(\pi \cdot e_t) = \alpha(\text{sgn}(\pi) e_t) \\ &= \text{sgn}(\pi)(\alpha e_t) \end{aligned}$$

und entspricht daher der sign-Darstellung, vgl. Beispiel 2.1.3.

(c) Ist $\lambda = (n-1, 1)$, so ist das tabloid $\{t\}$ zum λ -Tableau

$$t = \begin{array}{|c|c|c|} \hline i & \dots & j \\ \hline k & & \\ \hline \end{array}$$

gegeben durch

$$\{t\} = \frac{\overline{i \dots j}}{\overline{k}}.$$

Für dieses $\{t\}$ ist $C_t = \{id, (i, k)\}$, also gilt

$$\kappa_t = id - (i, k)$$

und daher

$$e_t = \kappa_t \cdot \{t\} = \{t\} - (i, k) \cdot \{t\}.$$

Wir wählen feste Tableaux t_i^j für $i \neq j$ aus $\{1, \dots, n\}$ der Gestalt

$$t_i^j = \begin{array}{|c|c|c|} \hline i & \dots & k \\ \hline j & & \\ \hline \end{array}$$

und ferner Tableaux t_1, \dots, t_n der Gestalt

$$t_i = \begin{array}{|c|c|c|} \hline * & \dots & * \\ \hline i & & \\ \hline \end{array},$$

so dass also $\{t_1\}, \dots, \{t_n\}$ eine Basis von $M^{(n-1, 1)}$ bilden. Somit gilt

2 Darstellungstheorie und der random walk auf endlichen Gruppen

$$e_{i,j} := e_{t_i^j} = \{t_j\} - \{t_i\}, \quad i \neq j \text{ aus } \{1, \dots, n\}$$

und für Skalare $\alpha_{i,j} \in \mathbb{C}$, $i \neq j$, erhalten wir:

$$\begin{aligned} \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n \alpha_{i,j} e_{i,j} &= \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n \alpha_{i,j} (\{t_j\} - \{t_i\}) \\ &= \sum_{i=1}^n \sum_{\substack{j=1 \\ j \neq i}}^n \alpha_{i,j} \{t_j\} - \sum_{i=1}^n \left(\sum_{\substack{j=1 \\ j \neq i}}^n \alpha_{i,j} \right) \{t_i\} \\ &= \sum_{j=1}^n \left(\sum_{\substack{i=1 \\ i \neq j}}^n \alpha_{i,j} \right) \{t_j\} - \sum_{i=1}^n \left(\sum_{\substack{j=1 \\ j \neq i}}^n \alpha_{i,j} \right) \{t_i\} \\ &= \sum_{l=1}^n \underbrace{\sum_{\substack{m=1 \\ m \neq l}}^n (\alpha_{m,l} - \alpha_{l,m})}_{=: \beta_l} \{t_l\} \end{aligned}$$

Nun gilt:

$$\sum_{l=1}^n \beta_l = \sum_{l=1}^n \sum_{\substack{m=1 \\ m \neq l}}^n \alpha_{m,l} - \sum_{l=1}^n \sum_{\substack{m=1 \\ m \neq l}}^n \alpha_{l,m} = 0,$$

das heißt es folgt

$$S^{(n-1,1)} \subseteq \{ \beta_1 \{t_1\} + \beta_2 \{t_2\} + \dots + \beta_n \{t_n\} : \beta_1 + \beta_2 + \dots + \beta_n = 0 \}.$$

Sind umgekehrt $\beta_1, \dots, \beta_n \in \mathbb{C}$ mit $\sum_{i=1}^n \beta_i = 0$, so sind die $n+1$ Gleichungen

2.2 Darstellungstheorie der symmetrischen Gruppe \mathbb{S}_n

$$\begin{aligned}
 \sum_{i=2}^n \alpha_{i,1} - \sum_{i=2}^n \alpha_{1,i} &= \beta_1 \\
 \sum_{\substack{i=1 \\ i \neq 2}}^n \alpha_{i,2} - \sum_{\substack{i=2 \\ i \neq 2}}^n \alpha_{2,i} &= \beta_2 \\
 &\vdots \\
 \sum_{i=1}^{n-1} \alpha_{i,n} - \sum_{i=1}^{n-1} \alpha_{n,i} &= \beta_n \\
 \sum_{j=1}^n \sum_{\substack{i=1 \\ i \neq j}}^n (\alpha_{i,j} - \alpha_{j,i}) &= 0
 \end{aligned}$$

in den Variablen $\alpha_{i,j}$, $i \neq j$, offenbar lösbar, so dass erst

$$S^{(n-1,1)} \supseteq \{ \beta_1 \{t_1\} + \beta_2 \{t_2\} + \dots + \beta_n \{t_n\} : \beta_1 + \beta_2 + \dots + \beta_n = 0 \}$$

und dann insgesamt Gleichheit folgt.

Sei nun wieder für festes $\lambda \vdash n$ $\{t_1\}, \dots, \{t_k\}$ die Standard-Basis von M^λ . Dann definieren wir für

$$v = \sum_{i=1}^k \alpha_i \{t_i\} \quad \text{und} \quad w = \sum_{i=1}^k \beta_i \{t_i\}$$

durch

$$\langle v, w \rangle := \sum_{i=1}^k \alpha_i \bar{\beta}_i$$

ein \mathbb{S}_n -invariantes Skalarprodukt auf M^λ , wie man leicht nachprüft. Bezüglich dieses Skalarproduktes ist $\{t_1\}, \dots, \{t_k\}$ eine Orthonormalbasis. Das folgende Lemma ist sehr wichtig:

Lemma 2.2.23 (sign lemma). *Sei $H \subseteq \mathbb{S}_n$ eine Untergruppe. Dann gilt:*

(i) *Ist $\pi \in H$, so gilt*

$$\pi \circ H^- = H^- \circ \pi = \text{sgn}(\pi) H^- .$$

2 Darstellungstheorie und der random walk auf endlichen Gruppen

(ii) Für alle $u, v \in M^\lambda$ ist

$$\langle H^- \cdot u, v \rangle = \langle u, H^- \cdot v \rangle.$$

(iii) Ist die Transposition $(i, j) \in H$, dann gibt es ein $z \in \mathbb{C}[\mathbb{S}_n]$ mit

$$H^- = z \circ (\text{id} - (i, j)).$$

(iv) Ist t ein Young-Tableau und sind $i \neq j$ in derselben Zeile von t mit $(i, j) \in H$, so gilt

$$H^- \cdot \{t\} = 0.$$

Beweis: zu (i): Wir haben

$$\begin{aligned} \pi \circ H^- &= \pi \circ \sum_{\sigma \in H} \text{sgn}(\sigma) \sigma = \text{sgn}(\pi^{-1}) \sum_{\sigma \in H} \text{sgn}(\pi \circ \sigma) \pi \circ \sigma \\ &= \text{sgn}(\pi) \sum_{\sigma \in H} \text{sgn}(\pi \circ \sigma) \pi \circ \sigma = \text{sgn}(\pi) \sum_{\varphi \in H} \text{sgn}(\varphi) \varphi \\ &= \text{sgn}(\pi) H^-. \end{aligned}$$

Analog folgt auch $H^- \circ \pi = \text{sgn}(\pi) H^-$.

zu (ii): Da $\langle \cdot, \cdot \rangle_{\mathbb{S}_n}$ -invariant ist, folgt

$$\begin{aligned} \langle H^- \cdot u, v \rangle &= \left\langle \sum_{\sigma \in H} \text{sgn}(\sigma) \sigma \cdot u, v \right\rangle = \sum_{\sigma \in H} \text{sgn}(\sigma) \langle \sigma \cdot u, v \rangle \\ &= \sum_{\sigma \in H} \text{sgn}(\sigma^{-1}) \langle u, \sigma^{-1} \cdot v \rangle = \left\langle u, \sum_{\sigma \in H} \text{sgn}(\sigma^{-1}) \sigma^{-1} \cdot v \right\rangle \\ &= \langle u, H^- \cdot v \rangle. \end{aligned}$$

zu (iii): Es ist $U := \{\text{id}, (i, j)\} \subseteq H$ eine Untergruppe von H . Seien $h_1 \circ U, \dots, h_s \circ U$ die paarweise verschiedenen Linksnebenklassen von U in H , so dass also

$$H = \bigcup_{i=1}^s h_i \circ U$$

ist. Dann gilt:

$$\begin{aligned} H^- &= \sum_{i=1}^s \sum_{u \in U} \text{sgn}(h_i \circ u) h_i \circ u = \underbrace{\left(\sum_{i=1}^s \text{sgn}(h_i) h_i \right)}_{=: z} \circ (\text{id} - (i, j)) \\ &= z \circ (\text{id} - (i, j)) \end{aligned}$$

zu (iv): Wir wählen gemäß (iii) ein $z \in \mathbb{C}[\mathbb{S}_n]$ mit

$$H^- = z \circ (id - (i, j)).$$

Dann gilt:

$$\begin{aligned} H^- \cdot \{t\} &= (z \circ (id - (i, j))) \cdot \{t\} = z \cdot (\{t\} - (i, j) \cdot \{t\}) \\ &= z \cdot (\{t\} - \{t\}) = 0, \end{aligned}$$

denn t und $(i, j) \cdot t$ sind Zeilen-äquivalent. □

Wir ziehen wichtige Folgerungen aus dem sign-lemma:

Korollar 2.2.24. *Seien λ und μ Partitionen von n und seien t und s Young-Tableaux mit $sh(t) = \lambda$ sowie $sh(s) = \mu$.*

(a) *Falls $\kappa_t \cdot \{s\} \neq 0$ ist, so folgt $\lambda \supseteq \mu$.*

(b) *Gilt $\lambda = \mu$, so ist $\kappa_t \{s\} = \pm e_t$.*

Beweis: zu (a): Seien $i \neq j$ aus $\{1, \dots, n\}$, die in derselben Zeile von s stehen. Dann können diese nicht in derselben Spalte von t stehen. Andernfalls gäbe es nämlich nach Lemma 2.2.23 (iii) eine Zerlegung

$$\kappa_t = z \circ (id - (i, j)).$$

Nach Lemma 2.2.23 (iv) wäre dann aber

$$\kappa_t \cdot \{s\} = 0,$$

im Widerspruch zur Voraussetzung. Nach Lemma 2.2.13 folgt somit $\lambda \supseteq \mu$.

zu (b): Ist $\lambda = \mu$, dann gibt es ein $\pi \in C_t$ mit

$$\{s\} = \pi \cdot \{t\},$$

was man mit demselben Argument einsieht, welches der Schlüssel zum Beweis des Dominanz-Lemmas, Lemma 2.2.13, war. Nach Lemma 2.2.23 (i) angewendet auf $H = C_t$ mit $H^- = \kappa_t$ folgt

$$\begin{aligned} \kappa_t \cdot \{s\} &= (\kappa_t \circ \pi) \cdot \{t\} = sgn(\pi)(\kappa_t \cdot \{t\}) = sgn(\pi)e_t \\ &= \pm e_t. \end{aligned}$$

□

2 Darstellungstheorie und der random walk auf endlichen Gruppen

Korollar 2.2.25. Sind $\lambda \vdash n$, t ein λ -Tableau und $u \in M^\lambda$, so gibt es ein $\alpha \in \mathbb{C}$ mit

$$\kappa_t \cdot u = \alpha e_t.$$

Beweis: Es sei

$$u = \sum_{i=1}^k \beta_i \{t_i\},$$

wobei wieder $\{t_1\}, \dots, \{t_k\}$ die verschiedenen λ -tabloids seien. Nach Korollar 2.2.24 (b) gibt es $\delta_1, \dots, \delta_k \in \{\pm 1\}$ mit

$$\begin{aligned} \kappa_t \cdot u &= \sum_{i=1}^k \beta_i (\kappa_t \cdot \{t_i\}) = \sum_{i=1}^k \beta_i \delta_i e_t \\ &= \underbrace{\left(\sum_{i=1}^k \beta_i \delta_i \right)}_{=: \alpha \in \mathbb{C}} e_t = \alpha e_t \end{aligned}$$

□

Wir formulieren und beweisen nun das sogenannte Untermodul-Theorem von James, welches das entscheidende Hilfsmittel für den Beweis der Aussage, dass die S^λ , $\lambda \vdash n$, alle irreduziblen \mathbb{S}_n -Moduln sind, ist.

Satz 2.2.26 (Untermodul-Theorem, James 1976). (a) Sei $\lambda \vdash n$ und U sei ein Untermodul von M^λ . Dann gilt

$$U \supseteq S^\lambda \quad \text{oder} \quad U \subseteq (S^\lambda)^\perp.$$

(b) Die Moduln S^λ , $\lambda \vdash n$, sind sämtlich irreduzibel.

Beweis: zu (a): Sei $u \in U$ und t sei ein λ -Tableau. Nach Korollar 2.2.25 gibt es ein $\alpha \in \mathbb{C}$ mit

$$\kappa_t \cdot u = \alpha e_t.$$

Fall 1: Es gebe ein $u \in U$ und ein λ -Tableau t mit

$$\kappa_t \cdot u = \alpha e_t \neq 0.$$

Dann gilt

2.2 Darstellungstheorie der symmetrischen Gruppe \mathbb{S}_n

$$e_t = \alpha^{-1}(\kappa_t \cdot u) \in U$$

und da S^λ ein zyklischer \mathbb{S}_n -Modul ist, folgt

$$S^\lambda \subseteq U,$$

denn U ist ein Untermodul von M^λ .

Fall 2: Für alle $u \in U$ und alle λ -Tableaux t sei

$$\kappa_t \cdot u = \alpha e_t = 0.$$

Wir behaupten, dass dann $U \subseteq (S^\lambda)^\perp$ gilt.

Seien nämlich $u \in U$ und t ein λ -Tableau. Nach Lemma 2.2.23 (ii) gilt:

$$\begin{aligned} \langle u, e_t \rangle &= \langle u, \kappa_t \cdot \{t\} \rangle = \langle \kappa_t \cdot u, \{t\} \rangle \\ &= \langle 0, \{t\} \rangle = 0 \end{aligned}$$

Da die e_t , $sh(t) = \lambda$, S^λ als \mathbb{C} -Vektorraum erzeugen, folgt $u \in (S^\lambda)^\perp$.

zu (b): Es sei

$$\{0\} \subseteq W \subseteq S^\lambda$$

ein \mathbb{S}_n -Untermodul von S^λ . Nach (a) können zwei Fälle auftreten:

Fall 1: Es ist $W \supseteq S^\lambda$. Dann ist insgesamt $W = S^\lambda$.

Fall 2: Es ist $W \subseteq (S^\lambda)^\perp$. Dann folgt

$$W \subseteq (S^\lambda) \cap (S^\lambda)^\perp = \{0\}.$$

Somit gilt entweder $W = S^\lambda$ oder $W = \{0\}$, so dass S^λ irreduzibel ist. □

Proposition 2.2.27. *Seien λ, μ Partitionen von n und sei $T : S^\lambda \rightarrow M^\mu$ ein von 0 verschiedener \mathbb{S}_n -Modul-Homomorphismus. Dann gilt:*

(a) $\lambda \triangleright \mu$

(b) Im Falle $\lambda = \mu$ ist T die Multiplikation mit einem Skalar $\alpha \in \mathbb{C}$.

2 Darstellungstheorie und der random walk auf endlichen Gruppen

Beweis: zu (a): Da $T \neq 0$ ist und $S^\lambda = \text{span}(e_t, sh(t) = \lambda)$, gibt es ein Young-Tableau t mit $sh(t) = \lambda$, so dass $T(e_t) \neq 0$ ist. Bezüglich des \mathbb{S}_n -invarianten Skalarprodukts $\langle \cdot, \cdot \rangle$ gilt die Zerlegung

$$M^\lambda = S^\lambda \oplus (S^\lambda)^\perp$$

und $(S^\lambda)^\perp$ ist ebenfalls ein \mathbb{S}_n -Untermodul von M^λ . Wir können T fortsetzen zu einem \mathbb{S}_n -Modul-Homomorphismus $\hat{T} : M^\lambda \rightarrow M^\mu$, indem wir

$$\hat{T}|_{(S^\lambda)^\perp} = 0$$

setzen.

Es seien $\{s_1\}, \dots, \{s_m\}$ die paarweise verschiedenen μ -tabloids. Da \hat{T} nach Bemerkung 2.1.18 auch ein $\mathbb{C}[\mathbb{S}_n]$ -Modul-Homomorphismus ist, folgt

$$\begin{aligned} 0 \neq T(e_t) &= \hat{T}(e_t) = \hat{T}(\kappa_t \cdot \{t\}) \\ &\stackrel{!}{=} \kappa_t \cdot \hat{T}(\{t\}). \end{aligned}$$

Dabei haben wir für die letzte Gleichheit gebraucht, dass \hat{T} auf ganz M^λ definiert ist, so dass $\hat{T}(\{t\})$ Sinn ergibt.

Seien $\beta_1, \dots, \beta_m \in \mathbb{C}$ mit

$$\hat{T}(\{t\}) = \sum_{j=1}^m \beta_j \{s_j\}.$$

Dann folgt

$$0 \neq T(e_t) = \kappa_t \cdot \sum_{j=1}^m \beta_j \{s_j\} = \sum_{j=1}^m \beta_j (\kappa_t \cdot \{s_j\}),$$

das heißt es gibt ein $j_0 \in \{1, \dots, m\}$ mit $\kappa_t \cdot \{s_{j_0}\} \neq 0$. Nach Korollar 2.2.24 (a) folgt hieraus $\lambda \geq \mu$.

zu (b): Im Falle $\lambda = \mu$ folgt aus Korollar 2.2.25, dass es ein $\alpha \in \mathbb{C}$ gibt mit

$$T(e_t) = \kappa_t \cdot \hat{T}(\{t\}) = \alpha e_t.$$

Dann gilt aber auch für jedes andere polytabloid $e_{\pi \cdot t}$ zur Partition λ :

2.2 Darstellungstheorie der symmetrischen Gruppe \mathbb{S}_n

$$\begin{aligned} T(e_{\pi \cdot t}) &= T(\pi \cdot e_t) = \pi \cdot T(e_t) \\ &= \pi \cdot (\alpha e_t) = \alpha(\pi \cdot e_t) \\ &= \alpha e_{\pi \cdot t}, \end{aligned}$$

wobei wir zweimal Lemma 2.2.19 (iv) angewandt haben. Da die polytabloids e_t mit $sh(t) = \lambda$ den Modul S^λ als \mathbb{C} -Vektorraum erzeugen, folgt, dass T die Multiplikation mit $\alpha \in \mathbb{C}$ ist. □

Nun können wir all unsere Behauptungen über die Specht-Moduln beweisen:

Satz 2.2.28. *Die Specht-Moduln S^λ , $\lambda \vdash n$, sind ein vollständiges Repräsentantensystem für die irreduziblen \mathbb{S}_n -Moduln.*

Beweis: Nach Satz 2.2.26 (b) sind alle S^λ irreduzibel. Da deren Anzahl stimmt, müssen wir somit nur noch zeigen, dass sie paarweise nicht-äquivalent sind: Seien also $\lambda, \mu \vdash n$ und sei $T : S^\lambda \rightarrow S^\mu$ ein \mathbb{S}_n -Modul-Isomorphismus. Da $S^\mu \subseteq M^\mu$ ein Untermodul ist, kann man T auch auffassen als Element von

$$\text{Hom}_{\mathbb{S}_n}(S^\lambda, M^\mu).$$

Nach Proposition 2.2.27 (a) folgt wegen $T \neq 0$, dass $\lambda \supseteq \mu$ gilt. Indem man dasselbe Argument auf den \mathbb{S}_n -Modul-Homomorphismus $T^{-1} : S^\mu \rightarrow S^\lambda$ anwendet, erhält man ebenso $\mu \supseteq \lambda$, so dass insgesamt $\lambda = \mu$ gezeigt ist. □

Wir behandeln noch die Frage, wie sich die Moduln M^μ für $\mu \vdash n$ aus den Specht-Moduln S^λ , $\lambda \vdash n$, zusammensetzen:

Korollar 2.2.29. *Sei $\mu \vdash n$. Dann gilt*

$$M^\mu = \bigoplus_{\lambda \supseteq \mu} m_{\lambda\mu} S^\lambda,$$

wobei die $m_{\lambda\mu}$ nicht-negative ganze Zahlen sind mit $m_{\mu\mu} = 1$.

Beweis: Nach Satz 2.1.29 gilt

$$m_{\lambda\mu} = \dim_{\mathbb{C}} (\text{Hom}_{\mathbb{S}_n}(S^\lambda, M^\mu))$$

für alle $\lambda \vdash n$. Nach Proposition 2.2.27 (a) folgt somit aus $m_{\lambda\mu} > 0$, dass $\lambda \supseteq \mu$ ist. Ebenso folgt aus Proposition 2.2.27 (b), dass gilt:

$$m_{\mu\mu} = \dim_{\mathbb{C}}(\{\alpha \text{id}_{M^\mu} \mid \alpha \in \mathbb{C}\}) = \dim_{\mathbb{C}}(\mathbb{C}) = 1$$

□

Nachdem wir die Hauptaussage über die Specht-Moduln bewiesen haben, stellen wir (für festes $\lambda \vdash n$) die Frage nach einer (möglichst natürlichen) Basis für S^λ . Wir wissen aus Proposition 2.2.21, dass S^λ als Vektorraum von den polytabloids e_t , $sh(t) = \lambda$, erzeugt wird. Jedoch sind diese im Allgemeinen nicht linear unabhängig. Nach dem *Basis-Auswahlsatz* gibt es auf jeden Fall eine Teilfamilie der polytabloids, welche eine Basis von S^λ bildet. Wir können diese Auswahl auf eine sehr anschauliche Weise treffen. Dazu zunächst:

Definition 2.2.30. *Sei $\lambda \vdash n$. Ein Young-Tableau t mit $sh(t) = \lambda$ heißt ein Standard-Young-Tableau oder einfach Standard-Tableau, wenn in jeder Zeile und in jeder Spalte von t die Einträge (streng) monoton wachsen, sofern man Zeilen von links nach rechts und Spalten von oben nach unten liest.*

Es gilt nun der folgende schöne Satz:

Satz 2.2.31. *Sei $\lambda \vdash n$. Die Familie*

$$(e_t : t \text{ ist ein Standard-Young-Tableau mit } sh(t) = \lambda)$$

bildet eine Basis von S^λ .

Der Beweis ist insgesamt recht lang und wird daher aus Platzgründen ausgelassen. Ein sehr gut verständlicher Beweis ist in [Sag01] zu finden.

□

Ist $\lambda \vdash n$, so bezeichne im Folgenden d_λ die Anzahl aller Standard-Young-Tableaux t mit $sh(t) = \lambda$. Wir erhalten sofort folgende Folgerung aus Satz 2.2.31:

Satz 2.2.32. *Sei $\lambda \vdash n$. Dann gilt:*

$$(i) \dim_{\mathbb{C}}(S^\lambda) = d_\lambda$$

$$(ii) \sum_{\lambda \vdash n} (d_\lambda)^2 = n!$$

Beweis: (i) folgt direkt aus Satz 2.2.31.

zu (ii): Wir wissen aus Korollar 2.1.34 (i), dass für jede endliche Gruppe G

$$|G| = \sum_{(V, \cdot) \in \text{Irr}(G)} (\dim_{\mathbb{C}}(V))^2$$

gilt. Aus Satz 2.2.28 in Verbindung mit (i) folgt daher die Behauptung, denn $|\mathbb{S}_n| = n!$.

□

Bemerkung 2.2.33. Die Aussage aus Satz 2.2.32 (ii) lässt sich auch kombinatorisch beweisen. Und zwar haben Robinson, Schensted und Knuth einen Algorithmus, den sogenannten RSK-Algorithmus, entwickelt, welcher einer Permutation $\pi \in \mathbb{S}_n$ bijektiv ein Paar (t, s) von Standard-Young-Tableaux mit $sh(t) = sh(s) \vdash n$ zuordnet. Eine ausführliche Beschreibung, samt Beweis, findet man in [Sag01].

Für die Zahlen d_λ wurden in der Vergangenheit viele Formeln hergeleitet. Zwei der am häufigsten verwendeten Formeln, welche ebenfalls in [Sag01] bewiesen werden, wollen wir noch in einem Satz, dessen Beweis wir ebenfalls auslassen, zusammenfassen. Dazu brauchen wir den Begriff des *Hakens* (*hook*) und der *Hakenlänge* (*hook length*) einer Box $b \in \lambda$ an der Stelle (i, j) . Wir definieren den Haken $H_{i,j}$ von b durch

$$H_{i,j} := \{(i, k) \in \lambda \mid k \geq j\} \cup \{(m, j) \in \lambda \mid m \geq i\}$$

und die zugehörige Hakenlänge $h_{i,j}$ von B durch

$$h_{i,j} := |H_{i,j}|.$$

Satz 2.2.34. Sei $\lambda = (\lambda_1, \dots, \lambda_l)$ eine Partition von n . Dann gilt

(a) die Determinanten-Formel von Frobenius-Young:

$$d_\lambda = n! \det \left(\frac{1}{(\lambda_i - i + j)!} \right)_{i,j=1,\dots,l}$$

wobei wir $\frac{1}{r!} := 0$ für $r < 0$ setzen.

(b) die Hook-Formel von Frame, Robinson und Thrall:

$$d_\lambda = \frac{n!}{\prod_{(i,j) \in \lambda} h_{i,j}}$$

Wir können (und werden) für jedes $n \in \mathbb{N}$ die Gruppe \mathbb{S}_n mit der Untergruppe

$$U := \{ \pi \in \mathbb{S}_{n+1} : \pi(n+1) = n+1 \}$$

von \mathbb{S}_{n+1} identifizieren. In diesem Sinne stellt sich die Frage, wie man für eine gegebene Partition λ von n gemäß dem Satz von Maschke den von S^λ auf \mathbb{S}_{n+1} induzierten Modul $Ind_{\mathbb{S}_n}^{\mathbb{S}_{n+1}}(S^\lambda)$ als direkte Summe der Moduln S^μ , $\mu \vdash n+1$, schreiben kann. Genauso stellt sich die Frage, wie man für $n \geq 2$ den auf \mathbb{S}_{n-1} eingeschränkten Modul $Res_{\mathbb{S}_{n-1}}^{\mathbb{S}_n}(S^\lambda)$ als direkte Summe der irreduziblen S^μ , $\mu \vdash n-1$, darstellen kann. Der folgende Satz gibt auf diese Frage eine einfache Antwort. Wir vereinbaren dazu für Partitionen $\lambda \vdash n$ und $\mu \vdash n-1$ die Schreibweise

$$\mu \nearrow \lambda,$$

falls λ aus μ durch das Hinzufügen einer einzelnen Box in den zugehörigen Young-Diagrammen erhalten werden kann.

Satz 2.2.35 (Branching-Regel). *Sei $n \in \mathbb{N}$ und $\lambda \vdash n$. Dann gilt:*

$$(i) \text{ Ind}_{\mathbb{S}_n}^{\mathbb{S}_{n+1}}(S^\lambda) \cong \bigoplus_{\lambda \nearrow \mu} S^\mu$$

$$(ii) \text{ Ist } n \geq 2, \text{ so gilt } \text{Res}_{\mathbb{S}_{n-1}}^{\mathbb{S}_n}(S^\lambda) \cong \bigoplus_{\mu \nearrow \lambda} S^\mu$$

Wir verzichten auf einen Beweis dieser Aussage und verweisen dazu auf das Buch von Sagan (vgl. [Sag01]).

2.3 Der random walk auf Gruppen

Sind μ und ν zwei Wahrscheinlichkeitsmaße auf dem Messraum $(\mathbb{R}^d, \mathcal{B}^d)$ und ist $S : \mathbb{R}^d \times \mathbb{R}^d \rightarrow \mathbb{R}^d$ die Summenabbildung, das heißt $S(x, y) := x + y$, so kennen wir die Faltung von μ und ν als das Bildmaß des Produktmaßes $\mu \otimes \nu$ unter der Abbildung S :

$$\mu * \nu := S(\mu \otimes \nu) = (\mu \otimes \nu) \circ S^{-1}.$$

Dies bedeutet nichts anderes, als dass $\mu * \nu$ die Verteilung der Summe $X + Y$ zweier unabhängiger Zufallsvariablen X mit Verteilung μ und Y mit Verteilung ν ist. Wenn wir die Skalarmultiplikation auf dem Vektorraum \mathbb{R}^d vergessen, so bleibt $(\mathbb{R}^d, +)$ eine (abelsche) Gruppe und wir können die Faltung verstehen als das Bildmaß des Produktmaßes $\mu \otimes \nu$ unter der Gruppenverknüpfung $+$. Dies kann man als Ansatz nehmen um das Konzept der Faltung auf beliebige Gruppen (G, \cdot) zu verallgemeinern: Ist G eine Gruppe mit einer σ -Algebra \mathcal{B} , so ist für zwei Wahrscheinlichkeitsmaße μ und ν auf \mathcal{B} die Faltung definiert durch

$$\mu * \nu := p(\mu \otimes \nu) = (\mu \otimes \nu) \circ p^{-1},$$

wobei $p : G \times G \rightarrow G$ die Gruppenverknüpfung $p(x, y) := x \cdot y$ ist.

Wir betrachten jedoch ausschließlich endliche Gruppen und identifizieren Wahrscheinlichkeitsmaße T auf $\mathcal{P}(G)$ mit ihren Zähldichten, so dass wir im Folgenden die Faltung zweier Funktionen $S, T : G \rightarrow \mathbb{C}$ definieren:

Definition 2.3.1. Für zwei \mathbb{C} -wertige Funktionen S, T auf G definieren wir die Faltung $T * S$ von T und S als diejenige Funktion $T * S : G \rightarrow \mathbb{C}$ mit

$$(T * S)(g) := \sum_{h \in G} T(gh^{-1})S(h)$$

für $g \in G$.

Bemerkung 2.3.2. (i) Sind X und Y zwei unabhängige G -wertige Zufallsvariablen auf einem Wahrscheinlichkeitsraum (Ω, \mathcal{A}, P) mit $P^X = T$ und $P^Y = S$, so hat die Zufallsvariable $Z := X \cdot Y$ die Verteilung $P^Z = T * S$.

Beweis: Für $g \in G$ gilt:

2 Darstellungstheorie und der random walk auf endlichen Gruppen

$$\begin{aligned}
 P(Z = g) &= P(X \cdot Y = g) = P\left(\bigcup_{h \in G} \{X = gh^{-1}, Y = h\}\right) \\
 &= \sum_{h \in G} P(X = gh^{-1}, Y = h) \stackrel{\text{unabh.}}{=} \sum_{h \in G} P(X = gh^{-1})P(Y = h) \\
 &= \sum_{h \in G} T(gh^{-1})S(h) \\
 &= (T * S)(g)
 \end{aligned}$$

□

(ii) Wir haben bereits die Gruppenalgebra $\mathbb{C}[G]$ von G kennengelernt. Die Abbildung

$$\varphi : \mathbb{C}^G \rightarrow \mathbb{C}[G], \quad T \mapsto \sum_{g \in G} T(g) g$$

ist dabei ein \mathbb{C} -Vektorraum-Isomorphismus, unter welchem wir $\mathbb{C}[G]$ mit \mathbb{C}^G identifizieren können. Wir wollen noch zeigen, dass der Faltung in \mathbb{C}^G unter dieser Identifikation die Multiplikation in $\mathbb{C}[G]$ entspricht. Daraus folgt, dass wir wichtige Rechenregeln, wie das Assoziativgesetz oder die Distributivgesetze für die Faltung auf \mathbb{C}^G zur Verfügung haben und, dass

$$\varphi : (\mathbb{C}^G, +, *) \rightarrow (\mathbb{C}[G], +, \cdot)$$

ein Isomorphismus von \mathbb{C} -Algebren ist.

Beweis: Für $S, T \in \mathbb{C}^G$ gilt

$$\begin{aligned}
 \varphi(T * S) &= \sum_{g \in G} (T * S)(g) g = \sum_{g \in G} \left(\sum_{h \in G} T(gh^{-1})S(h) \right) g \\
 &= \sum_{h \in G} \left(\sum_{g \in G} T(gh^{-1})S(h)g \right) \stackrel{u=gh^{-1}}{=} \sum_{h \in G} \left(\sum_{u \in G} T(u)S(h)uh \right) \\
 &= \left(\sum_{u \in G} T(u) u \right) \left(\sum_{h \in G} S(h) h \right) = \varphi(T)\varphi(S)
 \end{aligned}$$

□

Wir definieren nun den random walk auf einer endlichen Gruppe (G, \cdot) .

Definition 2.3.3. Es sei T ein Wahrscheinlichkeitsmaß auf $(G, \mathcal{P}(G))$ und es sei $(Z_i)_{i \in \mathbb{N}_0}$ eine Folge unabhängiger und identisch verteilter Zufallsvariablen, definiert auf einem Wahrscheinlichkeitsraum (Ω, \mathcal{A}, P) mit $P^{Z_0} = T$. Für $k \in \mathbb{N}$ sei

$$X_k := \prod_{i=0}^k Z_{k-i} = Z_k \cdot Z_{k-1} \cdot \dots \cdot Z_1 \cdot Z_0.$$

Dann heißt der stochastische Prozess $(X_k)_{k \in \mathbb{N}_0}$ ein random walk auf G zur Verteilung T .

Proposition 2.3.4. Mit den Bezeichnungen aus Definition 2.3.3 gilt für $k \in \mathbb{N}_0$:

$$P^{X_k} = T^{*k+1} := \underbrace{T * \dots * T}_{k+1}$$

Beweis: Dies folgt per Induktion aus Bemerkung 2.3.2 (i). □

Die folgende Aussage ist für den random walk auf \mathbb{Z} wohlbekannt:

Satz 2.3.5. Mit den Bezeichnungen aus Definition 2.3.3 ist der random walk $(X_k)_{k \in \mathbb{N}_0}$ eine zeithomogene Markov-Kette für deren Übergangsmatrix K gilt:

$$K(g, h) = T(hg^{-1})$$

Beweis: Seien $k \in \mathbb{N}_0$, $g_0, \dots, g_{k-1}, g, h \in G$ beliebig mit

$$P(X_0 = g_0, X_1 = g_1, \dots, X_{k-1} = g_{k-1}, X_k = g) > 0.$$

Dann gilt:

$$\begin{aligned} & P(X_{k+1} = h \mid X_0 = g_0, X_1 = g_1, \dots, X_{k-1} = g_{k-1}, X_k = g) \\ = & \frac{P(Z_{k+1} \cdot X_k = h, X_0 = g_0, X_1 = g_1, \dots, X_{k-1} = g_{k-1}, X_k = g)}{P(X_0 = g_0, X_1 = g_1, \dots, X_{k-1} = g_{k-1}, X_k = g)} \\ = & \frac{P(Z_{k+1} \cdot g = h, Z_0 = g_0, Z_1 \cdot Z_0 = g_1, \dots, \prod_{i=0}^{k-1} Z_{k-1-i} = g_{k-1}, \prod_{i=0}^k Z_{k-i} = g)}{P(Z_0 = g_0, Z_1 \cdot Z_0 = g_1, \dots, \prod_{i=0}^{k-1} Z_{k-1-i} = g_{k-1}, \prod_{i=0}^k Z_{k-i} = g)} \\ \stackrel{\text{unabh.}}{=} & P(Z_{k+1} = hg^{-1}) \frac{P(Z_0 = g_0, Z_1 \cdot Z_0 = g_1, \dots, \prod_{i=0}^{k-1} Z_{k-1-i} = g_{k-1}, \prod_{i=0}^k Z_{k-i} = g)}{P(Z_0 = g_0, Z_1 \cdot Z_0 = g_1, \dots, \prod_{i=0}^{k-1} Z_{k-1-i} = g_{k-1}, \prod_{i=0}^k Z_{k-i} = g)} \\ = & P(Z_{k+1} = hg^{-1}) = T(hg^{-1}) \end{aligned}$$

und ebenso

$$\begin{aligned}
 P(X_{k+1} = h \mid X_k = g) &= \frac{P(Z_{k+1} \cdot X_k = h, X_k = g)}{P(X_k = g)} \\
 &= \frac{P(Z_{k+1} = hg^{-1}, \prod_{i=0}^k Z_{k-i} = g)}{P(\prod_{i=0}^k Z_{k-i} = g)} \\
 &= \frac{P(Z_{k+1} = hg^{-1})P(\prod_{i=0}^k Z_{k-i} = g)}{P(\prod_{i=0}^k Z_{k-i} = g)} \\
 &= P(Z_{k+1} = hg^{-1}) = T(hg^{-1})
 \end{aligned}$$

□

Der random walk auf Gruppen findet Anwendungen in vielen Problemen der „realen“ Welt, wie zum Beispiel bei der Erzeugung von Zufallszahlen oder bei verschiedenen Mischverfahren für Kartenspiele. Eine ausführliche Darstellung dieser Thematik findet sich in [Dia88]. Wir wollen im Folgenden nur ein einfaches Mischverfahren für Karten vorstellen, welches eine zufällige Permutation aus zufälligen Transpositionen erzeugt. Dieses Verfahren wurde erstmals in [DS81] untersucht und findet sich ebenfalls in [Dia88].

P. Diaconis und M. Shahshahani diskutieren in [DS81] folgende Strategie zum Mischen von n Spielkarten:

Die Karten liegen verdeckt nebeneinander auf dem Tisch und sind so durchnummeriert, dass die Karte ganz links mit einer eins, die Karte rechts daneben mit einer zwei,..., und die Karte ganz rechts mit der Zahl n versehen ist. Nun werden zufällig und unabhängig voneinander zwei Zahlen $L, R \in \{1, \dots, n\}$ gewählt und es werden die Karten mit den Nummern L und R vertauscht, wobei keine Vertauschung stattfindet, wenn $L = R$ gilt. Indem man die Menge der Karten mit $\{1, \dots, n\}$ identifiziert, erhält man durch dieses Verfahren das folgende Wahrscheinlichkeitsmaß T auf \mathbb{S}_n :

- $T(id) = P(L = R) = n \cdot \frac{1}{n^2} = \frac{1}{n}$.
- $T(\tau_{i,j}) = P(\{L = i, R = j\} \cup \{L = j, R = i\}) = 2P(L = i)P(R = j) = \frac{2}{n^2}$ für eine Transposition $\tau_{i,j}$ mit $i \neq j$ aus $\{1, \dots, n\}$.
- $T(\pi) = 0$ für jede andere Permutation $\pi \in \mathbb{S}_n$.

Der k -fachen Iteration dieses Verfahrens entspricht dann das Produkt

$$X_k = Z_k \circ Z_{k-1} \circ \dots \circ Z_1,$$

wobei die Z_i unabhängig und identisch verteilte, gemäß der Verteilung T „ausgewürfelte“ Permutationen sind. Die Folge $(X_k)_{k \in \mathbb{N}}$ bildet einen random walk auf \mathbb{S}_n zur Verteilung T , so dass nach Proposition 2.3.4 die Verteilung von X_k gleich T^{*k} ist. Eine sinnvolle Minimalforderung an ein solches Mischverfahren sollte es sein, dass die Verteilung von X_k im Limes für $k \rightarrow \infty$ gegen die Gleichverteilung U auf \mathbb{S}_n konvergieren sollte, das heißt für jede Permutation $\pi \in \mathbb{S}_n$ sollte gelten:

$$\lim_{k \rightarrow \infty} P(X_k = \pi) = \lim_{k \rightarrow \infty} T^{*k}(\pi) = U(\pi) = \frac{1}{n!}$$

Dieses Ergebnis können wir im hier vorliegenden Fall wie folgt aus der elementaren Theorie der Markov-Ketten auf höchstens abzählbarem Zustandsraum erhalten:

Die Übergangsmatrix K ist nach Satz 2.3.5 gegeben durch $K(g, h) = T(hg^{-1})$. Wegen

$$\begin{aligned} \sum_{g \in G} K(g, h) &= \sum_{g \in G} T(hg^{-1}) = T\left(\bigcup_{g \in G} \{hg^{-1}\}\right) \\ &= T\left(\bigcup_{u \in G} \{u\}\right) = T(G) = 1 \end{aligned}$$

ist die Matrix K für den random walk auf einer endlichen Gruppe stets doppeltstochastisch und besitzt daher die Gleichverteilung U als stationäre Verteilung. Für unser spezielles Maß T auf \mathbb{S}_n ist die Matrix zusätzlich irreduzibel, da sich jede Permutation als Produkt von Transpositionen schreiben lässt und diese jeweils positive Wahrscheinlichkeit haben. Wegen $T(id) = \frac{1}{n} > 0$ ist die Markov-Kette darüberhinaus aperiodisch, so dass wir nach Ergebnissen aus der Theorie der Markov-Ketten die gewünschte Konvergenz erhalten (vgl. etwa [Bil95]). Ferner ist bekannt, dass diese Konvergenz sogar exponentiell schnell ist.

Wir wollen jedoch erklären wie in [DS81] oder [Dia88] die exakte Konvergenzrate mittels der Darstellungstheorie der symmetrischen Gruppe gefunden wird. Zunächst brauchen wir ein Abstandsmaß für zwei Verteilungen S, T auf der endlichen Gruppe G .

Definition 2.3.6. *Seien S, T zwei Wahrscheinlichkeitsmaße auf der endlichen Gruppe G . Dann heißt*

$$\|S - T\|_{TV} := \sup_{A \subseteq G} |S(A) - T(A)|$$

der Totalvariations-Abstand von S und T (vgl. auch Beispiel 1.1.1 (i)).

Bemerkung 2.3.7. (i) Für den Totalvariations-Abstand gilt die folgende Formel:

$$\|S - T\|_{TV} = \frac{1}{2} \sum_{g \in G} |S(g) - T(g)|,$$

denn das Supremum in der Definition wird angenommen in den beiden folgenden Mengen: $A = \{S \geq T\}$ und $B = \{S < T\}$.

(ii) Man kann den Totalvariations-Abstand für zwei (Wahrscheinlichkeits-) Maße P und Q auf einem beliebigen Messraum $(\mathcal{X}, \mathcal{B})$ genauso definieren. Ist dann μ ein dominierendes Maß für P und Q , zum Beispiel $\mu = \frac{P+Q}{2}$, und sind f und g die Radon-Nikodym-Dichten von P und Q bzgl. μ , so gilt die Formel

$$\|P - Q\|_{TV} = \int_{\{f \geq g\}} (f - g) d\mu = \int_{\{f < g\}} (g - f) d\mu = \frac{1}{2} \int_{\mathcal{X}} |f - g| d\mu$$

Unser Ziel ist es nun, das Verhalten der Größe

$$\|T^{*k} - U\|_{TV} \quad \text{für } k \rightarrow \infty$$

zu untersuchen. Dazu benötigen wir zunächst noch zwei Lemmata. Um das zweite beweisen zu können, benötigen wir eine etwas andere Version der Plancherel-Formel, vgl. Satz 2.1.42.

Satz 2.3.8 (Plancherel-Formel, Version 2). Seien G eine endliche Gruppe und $f, h : G \rightarrow \mathbb{C}$ zwei Funktionen. Für die Darstellungen $\varrho \in \text{Irr}(G)$ sei auf dem zugrundeliegenden Vektorraum jeweils ein ϱ -invariantes Skalarprodukt gewählt und es bezeichne $\hat{h}(\varrho)^*$ den adjungierten Endomorphismus zur Fouriertransformierten $\hat{h}(\varrho)$ von h in ϱ (vgl. Definition 2.1.35) bzgl. dieses Skalarproduktes. Dann gilt

$$\sum_{g \in G} f(g) \overline{h(g)} = \frac{1}{|G|} \sum_{\varrho \in \text{Irr}(G)} d_{\varrho} \text{Tr}(\hat{f}(\varrho) \circ \hat{h}(\varrho)^*).$$

Beweis: Wieder sind beide Seiten linear in f , so dass wir ohne Einschränkung $f = I_{\{x\}}$ annehmen können. Für Endomorphismen φ, ψ eines unitären Vektorraumes gilt die Gleichung

$$\text{Tr}(\varphi \circ \psi^*) = \overline{\text{Tr}(\varphi^* \circ \psi)},$$

wie man leicht mittels darstellender Matrizen verifiziert. Somit gilt:

$$\begin{aligned}
 \frac{1}{|G|} \sum_{\varrho \in \text{Irr}(G)} d_{\varrho} \text{Tr}(\hat{f}(\varrho) \circ \hat{h}(\varrho)^*) &= \overline{\frac{1}{|G|} \sum_{\varrho \in \text{Irr}(G)} d_{\varrho} \text{Tr}(\hat{f}(\varrho)^* \circ \hat{h}(\varrho))} \\
 &= \overline{\frac{1}{|G|} \sum_{\varrho \in \text{Irr}(G)} d_{\varrho} \text{Tr}(\varrho(x)^* \circ \hat{h}(\varrho))} \\
 &= \overline{\frac{1}{|G|} \sum_{\varrho \in \text{Irr}(G)} d_{\varrho} \text{Tr}(\varrho(x^{-1}) \circ \hat{h}(\varrho))} \\
 &\stackrel{\text{Satz 2.1.42}}{=} \overline{\overline{h(x)}} \\
 &= \sum_{g \in G} f(g) \overline{h(g)}
 \end{aligned}$$

Dabei haben wir für die dritte Gleichung verwendet, dass die $\varrho(g)$ unitäre Endomorphismen sind, das heißt, dass gilt:

$$\varrho(g^{-1}) = \varrho(g)^{-1} = \varrho(g)^*$$

□

Lemma 2.3.9. *Seien G eine endliche Gruppe und U die Gleichverteilung auf G . Dann gilt für jede irreduzible Darstellung ϱ von G :*

$$\hat{U}(\varrho) = \begin{cases} 1 & \text{falls } \varrho \text{ trivial} \\ 0 & \text{sonst} \end{cases}$$

Beweis: Wir haben

$$\hat{U}(\varrho) = \sum_{g \in G} U(g) \varrho(g) = \frac{1}{|G|} \sum_{g \in G} \varrho(g).$$

Daraus folgt leicht

$$\hat{U}(\varrho) \circ \varrho(g) = \varrho(g) \circ \hat{U}(\varrho)$$

für jedes $g \in G$. Da ϱ irreduzibel ist, folgt aus dem Lemma von Schur, dass $\hat{U}(\varrho)$ eine Homothetie ist zum Skalar

$$\begin{aligned}
 \lambda &= \frac{1}{d_{\varrho}} \text{Tr}(\hat{U}(\varrho)) = \frac{1}{d_{\varrho}} \frac{1}{|G|} \sum_{g \in G} \chi_{\varrho}(g) \\
 &= \frac{1}{d_{\varrho}} \langle \chi_{\varrho}, \chi_{tr} \rangle,
 \end{aligned}$$

2 Darstellungstheorie und der random walk auf endlichen Gruppen

wobei χ_{tr} der Charakter der trivialen Darstellung von G ist. Dies ist wegen der Orthonormalität der Charaktere

$$= \begin{cases} 1 & \text{falls } \varrho \text{ trivial} \\ 0 & \text{sonst} \end{cases}$$

Daraus ergibt sich sofort die Behauptung. □

Lemma 2.3.10 (Upper bound lemma, Diaconis, Shahshahani). *Sei Q ein Wahrscheinlichkeitsmaß auf der endlichen Gruppe G . Dann gilt*

$$\|Q - U\|_{TV}^2 \leq \frac{1}{4} \sum_{\varrho \in Irr(G) \setminus \{\varrho_{tr}\}} d_{\varrho} \text{Tr}(\hat{Q}(\varrho) \circ \hat{Q}(\varrho)^*),$$

wobei ϱ_{tr} die triviale Darstellung bezeichne.

Beweis:

Nach Bemerkung 2.3.7 (i) und der Cauchy-Schwarz-Ungleichung für den $\mathbb{R}^{|G|}$ folgt:

$$\begin{aligned} 4\|Q - U\|_{TV}^2 &= \left(\sum_{g \in G} |Q(g) - U(g)| \right)^2 = \left(\sum_{g \in G} |Q(g) - U(g)| \cdot 1 \right)^2 \\ &\leq \left(\sqrt{\sum_{g \in G} |Q(g) - U(g)|^2} \sqrt{\sum_{g \in G} 1^2} \right)^2 = \sum_{g \in G} |Q(g) - U(g)|^2 \cdot |G| \\ &= |G| \sum_{g \in G} (Q(g) - U(g))^2 \end{aligned}$$

Nach Satz 2.3.8 erhalten wir daraus weiter

$$4\|Q - U\|_{TV}^2 \leq \sum_{\varrho \in Irr(G)} d_{\varrho} \text{Tr}(\widehat{(Q - U)}(\varrho) \circ \widehat{(Q - U)}(\varrho)^*)$$

Für die triviale Darstellung ϱ_{tr} ergibt sich nach Lemma 2.3.9 und, da Q ein Wahrscheinlichkeitsmaß ist:

$$\hat{Q}(\varrho_{tr}) - \hat{U}(\varrho_{tr}) = 1 - 1 = 0$$

Und ebenfalls nach Lemma 2.3.9 folgt für jedes nicht-triviale $\varrho \in Irr(G)$:

$$\hat{Q}(\varrho) - \hat{U}(\varrho) = \hat{Q}(\varrho),$$

woraus sich unmittelbar die Behauptung ergibt. □

Man kann analog auch folgende untere Schranken für den Totalvariationsabstand angeben:

Satz 2.3.11. *Unter denselben Voraussetzungen wie in Lemma 2.3.10 gilt:*

$$(a) \|Q - U\|_{TV} \geq \frac{1}{2|G|} \sum_{\varrho \in \text{Irr}(G) \setminus \{\varrho_{tr}\}} d_{\varrho} \text{Tr}(\hat{Q}(\varrho) \circ \hat{Q}(\varrho)^*)$$

$$(b) \|Q - U\|_{TV}^2 \geq \frac{1}{4|G|} \sum_{\varrho \in \text{Irr}(G) \setminus \{\varrho_{tr}\}} d_{\varrho} \text{Tr}(\hat{Q}(\varrho) \circ \hat{Q}(\varrho)^*)$$

Beweis: zu (a): Wegen

$$0 \leq |Q(g) - U(g)| \leq 1$$

für jedes $g \in G$ gilt die Ungleichung

$$\begin{aligned} 2|G| \|Q - U\|_{TV} &= |G| \sum_{g \in G} |Q(g) - U(g)| \\ &\geq |G| \sum_{g \in G} |Q(g) - U(g)|^2 \stackrel{\text{wie zuvor}}{=} \sum_{\varrho \in \text{Irr}(G) \setminus \{\varrho_{tr}\}} d_{\varrho} \text{Tr}(\hat{Q}(\varrho) \circ \hat{Q}(\varrho)^*) \end{aligned}$$

zu (b): Indem wir die Ungleichung

$$(a_1 + \dots + a_n)^2 \geq a_1^2 + \dots + a_n^2$$

für nicht-negative reelle Zahlen a_1, \dots, a_n verwenden, erhalten wir:

$$\begin{aligned} 4|G| \|Q - U\|_{TV}^2 &= |G| \left(\sum_{g \in G} |Q(g) - U(g)| \right)^2 \\ &\geq |G| \sum_{g \in G} |Q(g) - U(g)|^2 \\ &= \sum_{\varrho \in \text{Irr}(G) \setminus \{\varrho_{tr}\}} d_{\varrho} \text{Tr}(\hat{Q}(\varrho) \circ \hat{Q}(\varrho)^*) \end{aligned}$$

□

2 Darstellungstheorie und der random walk auf endlichen Gruppen

Eine wichtige Eigenschaft der Fourier-Transformation ist, dass sie Faltung von Funktionen in die Multiplikation (genauer Komposition) der Fourier-Transformierten verwandelt:

Satz 2.3.12. *Seien G eine endliche Gruppe und $S, T : G \rightarrow \mathbb{C}$. Dann gilt für jede Darstellung (V, ϱ) von G :*

$$\widehat{T * S}(\varrho) = \hat{T}(\varrho) \circ \hat{S}(\varrho).$$

Beweis: Wir haben

$$\begin{aligned} \widehat{T * S}(\varrho) &= \sum_{g \in G} (T * S(g)) \varrho(g) = \sum_{g \in G} \left(\sum_{h \in G} T(gh^{-1}) S(h) \right) \varrho(g) \\ &= \sum_{g \in G} \left(\sum_{h \in G} T(gh^{-1}) S(h) \varrho(gh^{-1}h) \right) \\ &= \sum_{h \in G} \left(\sum_{g \in G} T(gh^{-1}) \varrho(gh^{-1}) \right) \circ (S(h) \varrho(h)) \\ &= \sum_{h \in G} \hat{T}(\varrho) \circ (S(h) \varrho(h)) = \hat{T}(\varrho) \circ \sum_{h \in G} S(h) \varrho(h) \\ &= \hat{T}(\varrho) \circ \hat{S}(\varrho) \end{aligned}$$

□

Wir kommen nun zurück auf unsere konkrete Situation und betrachten das Wahrscheinlichkeitsmaß T auf der symmetrischen Gruppe \mathbb{S}_n . Diaconis und Shashahani konnten zeigen (vgl. [DS81] oder [Dia88]):

Satz 2.3.13. *Sei $k \in \mathbb{N}$ von der Form $k = \frac{1}{2}n \log n + cn$. Dann gilt:*

(a) *Es gibt eine Konstante a , die nicht von n oder c abhängt, so dass für $c > 0$ gilt:*

$$\|T^{*k} - U\|_{TV} \leq ae^{-2c}$$

(b) *Ist $c < 0$, so gilt für $n \rightarrow \infty$:*

$$\|T^{*k} - U\|_{TV} \geq \left(\frac{1}{e} - e^{-e^{-2c}} \right) + o(1)$$

Wir wollen Satz 2.3.13 nicht vollständig beweisen, aber erklären, wie in den Beweis von Aussage (a) die Darstellungstheorie der symmetrischen Gruppe einspielt: Nach dem upper bound lemma, Lemma 2.3.10, gilt in dieser Situation

$$\|T^{*k} - U\|_{TV} \leq \frac{1}{4} \sum_{\varrho \in Irr(\mathbb{S}_n) \setminus \{\varrho_{tr}\}} d_\varrho \operatorname{Tr}(\widehat{T^{*k}}(\varrho) \circ \widehat{T^{*k}}(\varrho)^*) \quad (2.12)$$

Nach Satz 2.3.12 gilt für jedes ϱ :

$$\widehat{T^{*k}}(\varrho) = \underbrace{\hat{T}(\varrho) \circ \dots \circ \hat{T}(\varrho)}_k \quad (2.13)$$

Nach Konstruktion ist das Wahrscheinlichkeitsmaß T auch eine Klassen-Funktion der \mathbb{S}_n , so dass nach Satz 2.1.36 die Fourier-Transformierte $\hat{T}(\varrho)$ in ϱ eine Homothetie ist zum Skalar λ mit

$$\begin{aligned} \lambda &= \frac{1}{d_\varrho} \sum_{\pi \in \mathbb{S}_n} T(\pi) \chi_\varrho(\pi) = \frac{1}{d_\varrho} \left(T(id) d_\varrho + \sum_{\tau \text{ Transp.}} T(\tau) \chi_\varrho(\tau) \right) \\ &= \frac{1}{n} + \frac{1}{d_\varrho} \frac{n(n-1)}{2} \frac{2}{n^2} \chi_\varrho(\tau) = \frac{1}{n} + \frac{n-1}{n} \frac{\chi_\varrho(\tau)}{d_\varrho}, \end{aligned}$$

denn es gibt genau $\frac{n(n-1)}{2}$ verschiedene Transpositionen in der \mathbb{S}_n und χ_ϱ nimmt auf jeder Transposition denselben Wert an (τ sei einfach irgendeine Transposition). Nach Gleichung (2.13) ist dann auch $\widehat{T^{*k}}(\varrho)$ eine Homothetie zum Skalar

$$\lambda^k = \left(\frac{1}{n} + \frac{n-1}{n} \frac{\chi_\varrho(\tau)}{d_\varrho} \right)^k \quad (2.14)$$

Ebenso ist der adjungierte Endomorphismus $\widehat{T^{*k}}(\varrho)^*$ eine Homothetie zum Skalar

$$\overline{\lambda^k} = \bar{\lambda}^k = \left(\frac{1}{n} + \frac{n-1}{n} \frac{\overline{\chi_\varrho(\tau)}}{d_\varrho} \right)^k \quad (2.15)$$

$$= \left(\frac{1}{n} + \frac{n-1}{n} \frac{\chi_\varrho(\tau^{-1})}{d_\varrho} \right)^k = \left(\frac{1}{n} + \frac{n-1}{n} \frac{\chi_\varrho(\tau)}{d_\varrho} \right)^k = \lambda^k \quad (2.16)$$

Somit ist nach (2.14) und (2.15) auch $\widehat{T^{*k}}(\varrho) \circ \widehat{T^{*k}}(\varrho)^*$ eine Homothetie zum Skalar

2 Darstellungstheorie und der random walk auf endlichen Gruppen

$$\lambda^{2k} = \left(\frac{1}{n} + \frac{n-1}{n} \frac{\chi_\varrho(\tau)}{d_\varrho} \right)^{2k} \quad (2.17)$$

Indem man (2.17) in (2.12) einsetzt und die Spur auswertet, erhält man

$$\begin{aligned} \|T^{*k} - U\|_{TV} &\leq \frac{1}{4} \sum_{\varrho \in \text{Irr}(\mathbb{S}_n) \setminus \{\varrho_{tr}\}} d_\varrho^2 \left(\frac{1}{n} + \frac{n-1}{n} \frac{\chi_\varrho(\tau)}{d_\varrho} \right)^{2k} \\ &= \frac{1}{4} \sum_{\varrho \in \text{Irr}(\mathbb{S}_n) \setminus \{\varrho_{tr}\}} d_\varrho^2 \left(\frac{1}{n} + \frac{n-1}{n} r_\varrho(\tau) \right)^{2k}, \end{aligned} \quad (2.18)$$

mit

$$r_\varrho(\tau) = \frac{\chi_\varrho(\tau)}{d_\varrho}.$$

Diese Grösse hat einen Namen.

Definition 2.3.14. Die Grösse $r_\varrho(\tau) = \frac{\chi_\varrho(\tau)}{d_\varrho}$ heisst der zur Darstellung ϱ von \mathbb{S}_n gehörige character ratio, ausgewertet in der Transposition τ .

Anhand der Gleichung (2.18) kann man sehen, dass die character ratios der irreduziblen Darstellungen der Gruppe \mathbb{S}_n eine entscheidende Rolle für die Konvergenzgeschwindigkeit des von T erzeugten random walks gegen die Gleichverteilung spielen. Wir wollen noch kurz andeuten, wie Diaconis und Shashahani vorgehen, um den Term in (2.18) weiter nach oben abzuschätzen:

Zunächst machen sie von der Bijektion zwischen den Partitionen $\lambda = (\lambda_1, \dots, \lambda_l)$ von n und den irreduziblen Darstellungen der Gruppe \mathbb{S}_n Gebrauch und schreiben stets $\chi_\lambda, r_\lambda, \dots$ statt $\chi_\varrho, r_\varrho, \dots$. Dann verwenden sie, dass für $\lambda \vdash n$ die folgende Formel von Frobenius für den character ratio, ausgewertet in τ gilt:

$$r_\lambda(\tau) = \frac{1}{n(n-1)} \sum_{j=1}^l (\lambda_j^2 - (2j-1)\lambda_j) \quad (2.19)$$

Weiter zeigen sie, dass für $\lambda, \mu \vdash n$ die folgende wichtige Monotonieregel gilt:

$$\lambda \trianglerighteq \mu \quad \Rightarrow \quad r_\lambda(\tau) \geq r_\mu(\tau) \quad (2.20)$$

Dabei bezeichne \trianglerighteq die Dominanz-Ordnung, vgl. Definition 2.2.11.

Schließlich verwenden sie die Tatsache, dass die Dimension d_λ der zur Partition $\lambda \vdash n$ gehörigen irreduziblen Darstellung von \mathbb{S}_n gleich der Anzahl aller Standard-Young-Tableaus t mit $sh(t) = \lambda$ ist, um zu zeigen, dass für $\lambda = (\lambda_1, \dots, \lambda_l)$ stets

$$d_\lambda \leq \binom{n}{\lambda_1} d_{\lambda^*} \quad (2.21)$$

ist, wobei $\lambda^* = (\lambda_2, \dots, \lambda_l) \vdash n - \lambda_1$.

Unter Verwendung der Fakten (2.19), (2.20) und (2.21) erhalten sie durch Rechnung die Behauptung von Satz 2.3.13 (a).

2 Darstellungstheorie und der random walk auf endlichen Gruppen

3 Der zentrale Grenzwertsatz für character ratios

In Kapitel 2 haben wir das Plancherel-Maß \mathbb{P} zu einer endlichen Gruppe G als Wahrscheinlichkeitsmaß auf der Menge $Irr(G)$ aller paarweise nicht-äquivalenten irreduziblen Darstellungen von G kennengelernt. Der Darstellung $\rho \in Irr(G)$ wird dabei das Gewicht $\mathbb{P}(\{\rho\}) = \frac{d_\rho^2}{|G|}$ zugeordnet, vgl. Definition 2.1.41. In dem Fall, dass $G = \mathbb{S}_n$ eine symmetrische Gruppe ist, werden die irreduziblen Darstellungen von allen Partitionen λ von n über die Specht-Moduln S^λ , $\lambda \vdash n$, parametrisiert, wie wir in Abschnitt 2.2 gesehen haben. Gemäß dieser Parametrisierung können wir nun das Plancherel-Maß \mathbb{P}_n auf $Irr(\mathbb{S}_n)$ auf die Menge aller Partitionen von n übertragen.

Wir haben weiter in Abschnitt 2.3 den random walk auf \mathbb{S}_n betrachtet, der aus zufälligen Transpositionen eine zufällige Permutation erzeugt und gesehen, dass für die Konvergenzgeschwindigkeit dieses random walk gegen die Gleichverteilung in Totalvariation die character ratios $r_\lambda(\tau) = \frac{\chi_\lambda(\tau)}{d_\lambda}$, $\lambda \vdash n$, eine entscheidende Rolle spielen. Unter diesem Gesichtspunkt erscheint es als natürlich, das asymptotische Verhalten der Verteilung von $r_\lambda(\tau)$ für $n \rightarrow \infty$ zu betrachten, wobei die Partition $\lambda \vdash n$ jeweils gemäß dem Plancherel-Maß \mathbb{P}_n auf $Irr(\mathbb{S}_n)$ gewählt wird, welches gemäß der RSK-Korrespondenz der Gleichverteilung auf \mathbb{S}_n entspricht. Es stellt sich heraus, dass die Folge der character ratios, geeignet standardisiert, einem zentralen Grenzwertsatz genügt. Dieses Resultat, der *zentrale Grenzwertsatz für character ratios*, wurde zuerst von Kerov in [Ker93b] bewiesen. Kerov zeigte sogar allgemeiner, dass die gemeinsame asymptotische Verteilung von character ratios, die jeweils in Permutationen mit genau einem nicht-trivialen Zykel ausgewertet werden, auch eine mehrdimensionale Normalverteilung ist, und dass sogar asymptotische Unabhängigkeit dieser character ratios vorliegt. Die Beweismethode, die Kerov verwendete, ist die sogenannte Momente-Methode und diese liefert im Allgemeinen keine Konvergenzrate. Ferner ist Kerovs ursprünglicher Beweis von Grund auf kombinatorisch. Der erste zentrale Grenzwertsatz mit Rate $n^{-\frac{1}{4}}$ und Berry-Esséen-Konstante 40,1 für $r_\lambda(\tau)$ wurde von J. Fulman mit Hilfe des Ansatzes der austauschbaren Paare (siehe Unterabschnitt 1.1.2) innerhalb der Steinschen Methode in [Ful05] bewiesen. In der Arbeit [Ful06b] bewies Fulman mit Martingal-Theorie, dass für jedes $s \in (0; \frac{1}{2})$ eine Konstante $C_s \in [0; +\infty)$ existiert,

3 Der zentrale Grenzwertsatz für character ratios

so dass die Aussage des zentralen Grenzwertsatzes für $r_\lambda(\tau)$ mit der Rate n^{-s} und der Konstanten C_s gilt, wobei keine expliziten Konstanten ausgerechnet wurden. Shao und Su griffen in der Arbeit [SS06] den Beweis aus [Ful05] wieder auf und erhielten, unter Verwendung eines feineren Resultats für den Kolmogorov-Abstand im Rahmen des Ansatzes der austauschbaren Paare, die Konvergenzrate $n^{-\frac{1}{2}}$ mit der Konstanten $C = 150$. Fulman hingegen adaptierte in der Arbeit [Ful06a] den induktiven Ansatz von Bolthausen (siehe Abschnitt 1.2) und erhielt auf diesem Wege ebenfalls die Rate $n^{-\frac{1}{2}}$, berechnete jedoch keine explizite Konstante. Das Ziel dieses Kapitels ist es nun, den Beweis des zentralen Grenzwertsatzes für character ratios aus der Arbeit [Ful06a] vorzustellen und eine Berry-Esséen-Konstante auszurechnen, welche jedoch mit 808 größer als die in [SS06] sein wird. Wir werden zunächst in Abschnitt 3.1 etwas Notation einführen und einige Hilfsaussagen beweisen, um in Abschnitt 3.2 den Hauptsatz dieser Arbeit, das *Berry-Esséen-Theorem für character ratios*, beweisen zu können.

3.1 Das Plancherel-Maß und character ratios

Wie in der Einleitung dieses Kapitels bereits angedeutet, wollen wir, der Arbeit [Ful06a] von Fulman folgend, einen zentralen Grenzwertsatz vom Berry-Esséen-Typ für character ratios ausgewertet in Transpositionen beweisen. Dazu müssen wir den character ratio noch geeignet standardisieren. Zunächst übertragen wir das Plancherel-Maß von $Irr(\mathbb{S}_n)$ auf die Menge

$$\Omega_n := \{ \lambda : \lambda \vdash n \}$$

aller Partitionen von n .

Definition 3.1.1. Das Wahrscheinlichkeitsmaß \mathbb{P}_n auf dem Messraum $(\Omega_n, \mathcal{P}(\Omega_n))$, das durch die Gewichte

$$\mathbb{P}_n(\{\lambda\}) := \frac{d_\lambda^2}{n!}$$

auf den Partitionen λ von n eindeutig festgelegt ist, heißt das Plancherel-Maß auf den Partitionen von $n \in \mathbb{N}$. Den Erwartungswert bezüglich \mathbb{P}_n bezeichnen wir mit \mathbb{E}_n .

Der RSK-Algorithmus aus Bemerkung 2.2.33 definiert eine Bijektion $\pi \mapsto (t(\pi), s(\pi))$ zwischen der \mathbb{S}_n und der Menge aller Paare (t, s) von Standard-Young-Tableaux mit $sh(t) = sh(s) \vdash n$. Ist U_n die Gleichverteilung auf \mathbb{S}_n , so gilt

$$\begin{aligned} U_n(\{\pi \in \mathbb{S}_n : sh(t(\pi)) = sh(s(\pi)) = \lambda\}) &= \frac{|\{\pi \in \mathbb{S}_n : sh(t(\pi)) = sh(s(\pi)) = \lambda\}|}{n!} \\ &= \frac{d_\lambda^2}{n!} = \mathbb{P}_n(\{\lambda\}). \end{aligned}$$

Somit entspricht das Plancherel-Maß auf den Partitionen von n unter der RSK-Korrespondenz der Gleichverteilung auf der \mathbb{S}_n .

Da der Charakter χ_λ der zur Partition $\lambda \vdash n$ gehörigen Darstellung eine Klassenfunktion ist, ist es unerheblich, in welcher Transposition wir den character ratio auswerten. Da die Transposition $(1, 2)$ in jeder Gruppe \mathbb{S}_n , $n \geq 2$, enthalten ist, werden wir ihn stets in dieser Transposition auswerten. Wir definieren nun die Zufallsvariable $T_n : \Omega_n \rightarrow \mathbb{R}$ durch

$$T_n(\lambda) := \sqrt{\binom{n}{2}} r_\lambda(1, 2) = \frac{\sqrt{\binom{n}{2}} \chi_\lambda(1, 2)}{d_\lambda}.$$

Dass die Zufallsvariable T_n die gesuchte Standardisierung ist, ist Gegenstand der folgenden Proposition.

3 Der zentrale Grenzwertsatz für character ratios

Proposition 3.1.2. *Mit den obigen Bezeichnungen gilt:*

(a) $\mathbb{E}_n[T_n] = 0.$

(b) $\text{Var}(T_n) = \mathbb{E}_n[T_n^2] = 1.$

Beweis: zu (a): Nach der Definition des Erwartungswertes gilt:

$$\begin{aligned} \mathbb{E}_n[T_n] &= \sqrt{\binom{n}{2}} \sum_{\lambda \vdash n} \frac{\chi_\lambda(1, 2)}{d_\lambda} \cdot \mathbb{P}_n(\{\lambda\}) = \sqrt{\binom{n}{2}} \sum_{\lambda \vdash n} \frac{\chi_\lambda(1, 2)}{d_\lambda} \frac{d_\lambda^2}{n!} \\ &= \frac{\sqrt{\binom{n}{2}}}{n!} \sum_{\lambda \vdash n} \chi_\lambda(1, 2) d_\lambda = \frac{\sqrt{\binom{n}{2}}}{n!} \chi_{\text{reg}}(1, 2) \\ &= 0, \end{aligned}$$

nach Proposition 2.1.32.

zu (b): Wieder erhalten wir nach der Definition des Erwartungswertes

$$\begin{aligned} \mathbb{E}_n[T_n^2] &= \binom{n}{2} \sum_{\lambda \vdash n} \frac{\chi_\lambda(1, 2)^2}{d_\lambda^2} \cdot \mathbb{P}_n(\{\lambda\}) = \binom{n}{2} \sum_{\lambda \vdash n} \frac{\chi_\lambda(1, 2)^2}{d_\lambda^2} \frac{d_\lambda^2}{n!} \\ &= \frac{\binom{n}{2}}{n!} \sum_{\lambda \vdash n} \chi_\lambda(1, 2)^2 \stackrel{\text{Kor. 2.2.4}}{=} \frac{\binom{n}{2}}{n!} \sum_{\lambda \vdash n} \chi_\lambda(1, 2) \overline{\chi_\lambda(1, 2)} \\ &= 1, \end{aligned}$$

aufgrund der character-Relationen zweiter Art, Satz 2.1.40. □

Wir wollen nun die Zufallsvariable T_n als Summe anderer Zufallsvariablen schreiben, welche die Differenzenfolge eines Martingals bilden. Dazu müssen wir zunächst zu einem größeren Wahrscheinlichkeitsraum als $(\Omega_n, \mathcal{P}(\Omega_n), \mathbb{P}_n)$ übergehen, der das zufällige Wachsen von Partitionen gemäß dem Plancherel-Maß beschreibt. Wir definieren für $j \in \mathbb{N}_{\geq 2}$ Abbildungen $K_j : \Omega_{j-1} \times \mathcal{P}(\Omega_j) \rightarrow \mathbb{R}$ durch

$$K_j(\lambda(j-1), A_j) := \sum_{\substack{\lambda(j) \in A_j: \\ \lambda(j-1) \nearrow \lambda(j)}} \frac{d_{\lambda(j)}}{j d_{\lambda(j-1)}}. \quad (3.1)$$

Wir erinnern daran, dass $\lambda(j-1) \nearrow \lambda(j)$ bedeutet, dass das Young-Diagramm von $\lambda(j)$ aus dem Young-Diagramm von $\lambda(j-1)$ durch das Hinzufügen einer einzelnen Box erhalten werden kann. Die entscheidende Beobachtung von Kerov war es, dass durch Gleichung (3.1) ein Markov-Kern von $(\Omega_{j-1}, \mathcal{P}(\Omega_{j-1}))$ nach $(\Omega_j, \mathcal{P}(\Omega_j))$ definiert wird:

3.1 Das Plancherel-Maß und character ratios

Lemma 3.1.3. Für jedes $j \in \mathbb{N}_{\geq 2}$ wird durch Gleichung (3.1) ein Markov-Kern von $(\Omega_{j-1}, \mathcal{P}(\Omega_{j-1}))$ nach $(\Omega_j, \mathcal{P}(\Omega_j))$ definiert.

Beweis: Sei $j \in \mathbb{N}$ beliebig. Offenbar ist K_j nicht-negativ. Da nur Potenzmengen als σ -Algebren involviert sind, gibt es auch keine Messbarkeitsprobleme. Sei jetzt $\lambda(j-1) \in \Omega_{j-1}$ fest gewählt. Dann müssen wir zeigen, dass die Gleichung

$$\sum_{\substack{\lambda(j) \in \Omega_j: \\ \lambda(j-1) \nearrow \lambda(j)}} \frac{d_{\lambda(j)}}{j d_{\lambda(j-1)}} = 1 \quad (3.2)$$

erfüllt ist. Dies zeigen wir mit Hilfe der Branching-Regel, Satz 2.2.35 (i). Es gilt nach diesem Satz:

$$\text{Ind}_{\mathbb{S}_{j-1}}^{\mathbb{S}_j}(S^{\lambda(j-1)}) \cong \bigoplus_{\substack{\lambda(j) \in \Omega_j \\ \lambda(j-1) \nearrow \lambda(j)}} S^{\lambda(j)} \quad (3.3)$$

Indem wir auf beiden Seiten von (3.3) die Dimensionen betrachten, erhalten wir

$$\begin{aligned} j \cdot d_{\lambda(j-1)} &= \frac{|\mathbb{S}_j|}{|\mathbb{S}_{j-1}|} d_{\lambda(j-1)} = [\mathbb{S}_j : \mathbb{S}_{j-1}] d_{\lambda(j-1)} \\ &\stackrel{\text{Prop. 2.1.45}}{=} \dim_{\mathbb{C}}(\text{Ind}_{\mathbb{S}_{j-1}}^{\mathbb{S}_j}(S^{\lambda(j-1)})) = \sum_{\substack{\lambda(j) \in \Omega_j: \\ \lambda(j-1) \nearrow \lambda(j)}} \dim_{\mathbb{C}}(S^{\lambda(j)}) \\ &= \sum_{\substack{\lambda(j) \in \Omega_j: \\ \lambda(j-1) \nearrow \lambda(j)}} d_{\lambda(j)}. \end{aligned} \quad (3.4)$$

Aus (3.4) ergibt sich nach Division durch $j \cdot d_{\lambda(j-1)}$ unmittelbar die Gleichung (3.2). □

Es sei μ das einzige Wahrscheinlichkeitsmaß auf dem einelementigen Messraum $(\Omega_1, \mathcal{P}(\Omega_1))$. Weiter sei

$$\Omega^{(n)} := \prod_{j=1}^n \Omega_j = \Omega_1 \times \Omega_2 \times \dots \times \Omega_n$$

der Produktraum, ausgestattet mit der Produkt- σ -Algebra

$$\mathcal{P}(\Omega^{(n)}) = \mathcal{P}(\Omega_1) \otimes \dots \otimes \mathcal{P}(\Omega_n).$$

Nun definieren wir das Wahrscheinlichkeitsmaß P_n auf $(\Omega^{(n)}, \mathcal{P}(\Omega^{(n)}))$ durch

$$P_n := \mu \otimes \bigotimes_{j=2}^n K_j, \quad (3.5)$$

3 Der zentrale Grenzwertsatz für character ratios

vgl. etwa [Kle06], Kapitel 14, oder [GS77] für die Definition dieses Maßes, das die Modellierung eines Gesamtexperimentes beschreibt, wenn eine beliebige Start-Verteilung (hier μ) und Markov-Kerne für die bedingten Experimente (hier die K_j) gegeben sind. (Wir könnten natürlich auch mit Hilfe des Satzes von Ionescu-Tulcea ein eindeutig bestimmtes Wahrscheinlichkeitsmaß auf dem unendlichen Produktraum der Ω_j finden, welches dieselben Übergangswahrscheinlichkeiten besitzt!) Wir bezeichnen mit E_n den Erwartungswert bzgl. des Wahrscheinlichkeitsmaßes P_n . Da der Wahrscheinlichkeitsraum $(\Omega^{(n)}, \mathcal{P}(\Omega^{(n)}), P_n)$ diskret ist, ist das Maß P_n bereits durch seine Zähldichte festgelegt. Für einzelne Punkte $(\lambda(1), \lambda(2), \dots, \lambda(n)) \in \Omega^{(n)}$ gilt die bemerkenswerte Formel

$$P_n(\{(\lambda(1), \lambda(2), \dots, \lambda(n))\}) = \frac{d_{\lambda(2)}}{2} \frac{d_{\lambda(3)}}{3d_{\lambda(2)}} \cdot \dots \cdot \frac{d_{\lambda(n)}}{nd_{\lambda(n-1)}} = \frac{d_{\lambda(n)}}{n!}, \quad (3.6)$$

welche sofort aus (3.1) in Verbindung mit einer bekannten Multiplikationsregel für bedingte Wahrscheinlichkeiten folgt. Nach der Formel (3.6) hängt die Wahrscheinlichkeit für das Tupel $(\lambda(1), \lambda(2), \dots, \lambda(n))$ in Wahrheit nur von $\lambda(n)$ ab! Wir bezeichnen die kanonischen Projektionen auf dem Raum $\Omega^{(n)}$ im Folgenden mit $\Lambda_1^{(n)}, \dots, \Lambda_n^{(n)}$, das heißt, es ist für $j = 1, \dots, n$

$$\Lambda_j^{(n)}(\lambda(1), \lambda(2), \dots, \lambda(n)) = \lambda(j).$$

Da die Markov-Kerne K_j nur von $\lambda(j-1)$ und nicht von der ganzen Vorgeschichte abhängen, erkennen wir den Prozess $(\Lambda_j^{(n)})_{j=1, \dots, n}$ als (zeitlich nicht homogene) Markov-Kette (mit endlichem Zeithorizont). Dieser Prozess wird in der Literatur auch *Kerov's growth process* (siehe [Ful06a]) oder *Plancherel growth process* (siehe etwa [Ker96]) genannt. Die zuletzt genannte Bezeichnung wird durch den folgenden wichtigen Satz von Kerov gerechtfertigt:

Satz 3.1.4. *Mit den obigen Bezeichnungen gilt für jedes $n \in \mathbb{N}$ und für alle $1 \leq j \leq n$:*

$$P_n^{\Lambda_j^{(n)}} = \mathbb{P}_j,$$

das heißt, die Verteilung von $\Lambda_j^{(n)}$ ist gleich dem Plancherel-Maß auf den Partitionen von j .

Beweis: Wir beweisen die Behauptung für festes n per Induktion über $1 \leq j \leq n$:

Induktionsanfang: $n = 1$: Es sei $\lambda(1)$ die einzige Partition von 1. Dann gilt:

$$P_n(\Lambda_1^{(n)} = \lambda(1)) = 1 = \frac{d_{\lambda(1)}^2}{1!} = \mathbb{P}_1(\{\lambda(1)\}),$$

3.1 Das Plancherel-Maß und character ratios

denn $S^{\lambda(1)}$ ist der triviale Modul, da \mathbb{S}_1 trivial ist.

Induktionsschritt: $j-1 \rightarrow j$. Es sei $\lambda(j) \in \Omega_j$ beliebig. Dann gilt nach Definition des Prozesses $(\Lambda_j^{(n)})_{j=1, \dots, n}$ und mit der Formel von der totalen Wahrscheinlichkeit:

$$\begin{aligned}
 & P_n(\Lambda_j^{(n)} = \lambda(j)) \\
 = & \sum_{\lambda(j-1) \vdash j-1} P_n(\Lambda_j^{(n)} = \lambda(j) \mid \Lambda_{j-1}^{(n)} = \lambda(j-1)) \cdot P_n(\Lambda_{j-1}^{(n)} = \lambda(j-1)) \\
 = & \sum_{\substack{\lambda(j-1) \in \Omega_{j-1}: \\ \lambda(j-1) \nearrow \lambda(j)}} P_n(\Lambda_j^{(n)} = \lambda(j) \mid \Lambda_{j-1}^{(n)} = \lambda(j-1)) \cdot P_n(\Lambda_{j-1}^{(n)} = \lambda(j-1)) \\
 = & \sum_{\substack{\lambda(j-1) \in \Omega_{j-1}: \\ \lambda(j-1) \nearrow \lambda(j)}} K_j(\lambda(j-1), \{\lambda(j)\}) \cdot P_n(\Lambda_{j-1}^{(n)} = \lambda(j-1)) \\
 \stackrel{\text{Ind.-Vor.}}{=} & \sum_{\substack{\lambda(j-1) \in \Omega_{j-1}: \\ \lambda(j-1) \nearrow \lambda(j)}} \frac{d_{\lambda(j)}}{j d_{\lambda(j-1)}} \cdot \mathbb{P}_{j-1}(\{\lambda(j-1)\}) \\
 = & \sum_{\substack{\lambda(j-1) \in \Omega_{j-1}: \\ \lambda(j-1) \nearrow \lambda(j)}} \frac{d_{\lambda(j)}}{j d_{\lambda(j-1)}} \cdot \frac{d_{\lambda(j-1)}^2}{(j-1)!} \\
 = & \frac{d_{\lambda(j)}}{j!} \sum_{\substack{\lambda(j-1) \in \Omega_{j-1}: \\ \lambda(j-1) \nearrow \lambda(j)}} d_{\lambda(j-1)} \tag{3.7}
 \end{aligned}$$

Nun verwenden wir wieder die Branching-Regel, und zwar Satz 2.2.35 (ii). Dieser besagt, dass

$$\text{Res} \left(\begin{smallmatrix} \mathbb{S}_j \\ \mathbb{S}_{j-1} \end{smallmatrix} S^{\lambda(j)} \right) \cong \bigoplus_{\substack{\lambda(j-1) \in \Omega_{j-1}: \\ \lambda(j-1) \nearrow \lambda(j)}} S^{\lambda(j-1)}$$

gilt. Indem wir auf beiden Seiten die Dimensionen betrachten, folgt

$$d_{\lambda(j)} = \dim_{\mathbb{C}}(\text{Res} \begin{smallmatrix} \mathbb{S}_j \\ \mathbb{S}_{j-1} \end{smallmatrix} S^{\lambda(j)}) = \sum_{\substack{\lambda(j-1) \in \Omega_{j-1}: \\ \lambda(j-1) \nearrow \lambda(j)}} d_{\lambda(j-1)} \tag{3.8}$$

und, indem wir (3.8) in (3.7) einsetzen, erhalten wir

$$P_n(\Lambda_j^{(n)} = \lambda(j)) = \frac{d_{\lambda(j)}^2}{j!} = \mathbb{P}_j(\{\lambda(j)\}),$$

3 Der zentrale Grenzwertsatz für character ratios

also die Behauptung. □

Eine erste Konsequenz aus Satz 3.1.4 ist, dass die Zufallsvariable $T_j^{(n)} := T_j(\Lambda_j^{(n)}) = T_j \circ \Lambda_j^{(n)}$ auf $\Omega^{(n)}$ dieselbe Verteilung hat, wie die Zufallsvariable T_j auf Ω_j , $j = 1, \dots, n$. Anschaulich besagt der Satz, dass in jedem Schritt j , wobei $1 \leq j \leq n$, des zufälligen Wachsens von Young-Diagrammen (bzw. von deren Partitionen) das Diagramm (bzw. die Partition) $\Lambda_j^{(n)}$ nach dem Plancherel-Maß verteilt ist. Wir können somit alle Rechnungen, die wir bzgl. des Plancherel-Maßes anstellen möchten, ohne Einschränkung in diesem Wahrscheinlichkeitsraum, der mehr Informationen enthält, durchführen.

Wir benötigen noch einige weitere Notationen: Es seien $Y_0^{(n)} := Y_1^{(n)} := 0$ und für $2 \leq j \leq n$ sei

$$Y_j^{(n)} := \frac{\binom{j}{2} \chi_{\Lambda_j^{(n)}}(1, 2)}{d_{\Lambda_j^{(n)}}} = \sqrt{\binom{j}{2}} T_j^{(n)}.$$

Weiter sei für $1 \leq j \leq n$ $X_j^{(n)} := Y_j^{(n)} - Y_{j-1}^{(n)}$. Weiter sei für $2 \leq j \leq n$ $B_j^{(n)}$ die Box, die beim Übergang von $\Lambda_{j-1}^{(n)}$ nach $\Lambda_j^{(n)}$ hinzugefügt wird und es sei $C_j^{(n)} := c(B_j^{(n)})$ ihr *Inhalt* im Sinne der folgenden Definition.

Definition 3.1.5. Sei λ eine Partition und es sei $b \in \lambda$ eine Box im zugehörigen Young-Diagramm an der Stelle (i, j) . Dann heißt $c(b) := j - i$ der Inhalt der Box b .

Der Inhalt gibt somit mit Vorzeichen den Abstand einer Box $b \in \lambda$ von der Diagonalen an. Aufgrund unserer Definitionen können wir nun die Zufallsvariable $T_n^{(n)}$, deren Verteilung uns interessiert, wie folgt zerlegen: Es gilt

$$T_n^{(n)} = \frac{1}{\sqrt{\binom{n}{2}}} Y_n^{(n)} = \frac{1}{\sqrt{\binom{n}{2}}} (X_1^{(n)} + X_2^{(n)} + \dots + X_n^{(n)}). \quad (3.9)$$

Der nächste Satz verdeutlicht, wieso der Inhalt einer Box für uns von Interesse ist:

Satz 3.1.6. Für $2 \leq j \leq n$ ist $X_j^{(n)} = C_j^{(n)}$.

Beweis: Als (zufällige) Partition von $j - 1$ können wir $\Lambda_{j-1}^{(n)}$ wie folgt darstellen: Es gibt ein $s \in \mathbb{N}$ und natürliche Zahlen $k_1 \geq k_2 \geq \dots \geq k_s \geq 1$ mit $\Lambda_{j-1}^{(n)} = (k_1, k_2, \dots, k_s)$. Wir setzen $k_{s+1} := 0$. Dann gibt es genau ein $m \in \{1, \dots, s+1\}$ mit

$$\Lambda_j^{(n)} = (k_1, k_2, \dots, k_{m-1}, k_m + 1, k_{m+1}, \dots, k_{s+1}).$$

3.1 Das Plancherel-Maß und character ratios

Indem wir die aus Gleichung (2.19) bekannte Formel von Frobenius für den character ratio ausgewertet in der Transposition $(1, 2)$ verwenden, erhalten wir

$$\begin{aligned}
Y_j^{(n)} &= \binom{j}{2} r_{\Lambda_j^{(n)}}(1, 2) = \frac{1}{2} \left(\sum_{\substack{i=1 \\ i \neq m}}^{s+1} (k_i^2 - (2i-1)k_i) + (k_m+1)^2 - (2m-1)(k_m+1) \right) \\
&= \frac{1}{2} \left(\sum_{\substack{i=1 \\ i \neq m}}^{s+1} (k_i^2 - (2i-1)k_i) + k_m^2 + 2k_m + 1 - (2m-1)k_m - (2m-1) \right) \\
&= \frac{1}{2} \left(\sum_{i=1}^{s+1} (k_i^2 - (2i-1)k_i) + 2k_m + 2 - 2m \right) \\
&= Y_{j-1}^{(n)} + k_m + 1 - m \tag{3.10}
\end{aligned}$$

Wegen $k_m + 1 - m = c(B_j^{(n)})$ folgt aus (3.10) unmittelbar die Behauptung:

$$X_j^{(n)} = Y_j^{(n)} - Y_{j-1}^{(n)} = c(B_j^{(n)}) = C_j^{(n)}$$

□

Somit haben wir für die Größe $X_j^{(n)}$ eine geometrische Interpretation gefunden. Man beachte, dass die entscheidende Formel von Frobenius in dieser Form nur für den character ratio ausgewertet in den Transpositionen richtig ist und wir daher für andere Konjugationsklassen der symmetrischen Gruppen diese Darstellung nicht zur Verfügung haben.

Lemma 3.1.7. *Mit den obigen Bezeichnungen gilt für alle $2 \leq j \leq n$:*

- (i) $E_n[Y_j^{(n)} | \Lambda_{j-1}^{(n)}] = Y_{j-1}^{(n)}$.
- (ii) $E_n[Y_j^{(n)} | Y_n^{(n)}] = \frac{j(j-1)}{n(n-1)} Y_n^{(n)}$.
- (iii) $E_n[X_j^{(n)} | T_n^{(n)}] = E_n[X_j^{(n)} | Y_n^{(n)}] = \frac{j-1}{\sqrt{\binom{n}{2}}} T_n^{(n)}$.
- (iv) $E_n[X_j^{(n)} | \Lambda_{j-1}^{(n)}] = 0$.
- (v) $E_n[(X_j^{(n)})^2 | \Lambda_{j-1}^{(n)}] = j - 1$.
- (vi) $E_n[(X_j^{(n)})^4 | \Lambda_{j-1}^{(n)}] = \binom{j}{2} + 3 \sum_{B \in \Lambda_{j-1}^{(n)}} c(B)^2$.
- (vii) $E_n[(X_j^{(n)})^2] = j - 1$.

3 Der zentrale Grenzwertsatz für character ratios

Beweis: zu (i): Es sei $\lambda(j-1) \in \Omega_{j-1}$ fest gewählt. Dann gilt nach der Transformationsregel angewendet auf den bedingten Erwartungswert:

$$\begin{aligned}
& E_n[Y_j^{(n)} | \Lambda_{j-1}^{(n)} = \lambda(j-1)] \\
&= \sum_{\substack{\lambda(j) \in \Omega_j: \\ \lambda(j-1) \nearrow \lambda(j)}} \frac{\binom{j}{2} \chi_{\lambda(j)}(1, 2)}{d_{\lambda(j)}} \cdot P_n(\Lambda_j^{(n)} = \lambda(j) | \Lambda_{j-1}^{(n)} = \lambda(j-1)) \\
&= \sum_{\substack{\lambda(j) \in \Omega_j: \\ \lambda(j-1) \nearrow \lambda(j)}} \frac{\binom{j}{2} \chi_{\lambda(j)}(1, 2)}{d_{\lambda(j)}} \cdot \frac{d_{\lambda(j)}}{j d_{\lambda(j-1)}} \\
&= \frac{j-1}{2 d_{\lambda(j-1)}} \sum_{\substack{\lambda(j) \in \Omega_j: \\ \lambda(j-1) \nearrow \lambda(j)}} \chi_{\lambda(j)}(1, 2) \\
&= \frac{j-1}{2 d_{\lambda(j-1)}} \cdot \text{Ind}_{\mathbb{S}_{j-1}}^{\mathbb{S}_j}(\chi_{\lambda(j-1)})(1, 2), \tag{3.11}
\end{aligned}$$

nach der Branching-Regel, Satz 2.2.35 (i).

Nach Satz 2.1.48 folgt aus (3.11), dass gilt:

$$E_n[Y_j^{(n)} | \Lambda_{j-1}^{(n)} = \lambda(j-1)] = \frac{j-1}{2 d_{\lambda(j-1)}} \cdot \frac{1}{(j-1)!} \sum_{\substack{\pi \in \mathbb{S}_j: \\ \pi^{-1} \circ (1, 2) \circ \pi \in \mathbb{S}_{j-1}}} \chi_{\lambda(j-1)}(\pi^{-1} \circ (1, 2) \circ \pi) \tag{3.12}$$

Nun gilt $\pi^{-1} \circ (1, 2) \circ \pi \in \mathbb{S}_{j-1}$ genau dann, wenn $\pi(j) \notin \{1, 2\}$ ist. In diesem Fall ist $\tau := \pi^{-1} \circ (1, 2) \circ \pi$ eine Transposition mit $\tau(j) = j$, also $\tau \in \mathbb{S}_{j-1}$, so dass $\chi_{\lambda(j-1)}(\tau) = \chi_{\lambda(j-1)}(1, 2)$ ist. Da es genau $(j-2) \cdot (j-1)!$ solche $\pi \in \mathbb{S}_j$ gibt, folgt aus (3.12)

$$\begin{aligned}
E_n[Y_j^{(n)} | \Lambda_{j-1}^{(n)} = \lambda(j-1)] &= \frac{j-1}{2 d_{\lambda(j-1)}} \frac{(j-2)(j-1)!}{(j-1)!} \chi_{\lambda(j-1)}(1, 2) \\
&= \frac{(j-1)(j-2)}{2 d_{\lambda(j-1)}} \chi_{\lambda(j-1)}(1, 2) = \frac{\binom{j-1}{2} \chi_{\lambda(j-1)}(1, 2)}{d_{\lambda(j-1)}}
\end{aligned}$$

und dies ist gleich $Y_{j-1}^{(n)}$, falls $\Lambda_{j-1}^{(n)} = \lambda(j-1)$ gilt.

zu (ii): Wir werden zeigen, dass für $1 \leq j \leq n$ die Gleichung

$$E_n[Y_j^{(n)} | \Lambda_n^{(n)}] = \frac{j(j-1)}{n(n-1)} Y_n^{(n)} \tag{3.13}$$

3.1 Das Plancherel-Maß und character ratios

gilt. Wegen $\sigma(Y_n^{(n)}) \subseteq \sigma(\Lambda_n^{(n)})$ folgt daraus

$$E_n[Y_j^{(n)}|Y_n^{(n)}] = E_n[E_n[Y_j^{(n)}|\Lambda_n^{(n)}]|Y_n^{(n)}] = E_n\left[\frac{j(j-1)}{n(n-1)}Y_n^{(n)}\middle|Y_n^{(n)}\right] = \frac{j(j-1)}{n(n-1)}Y_n^{(n)},$$

also die Behauptung. Für $j = 1$ ist (3.13) aufgrund der Definition von $Y_1^{(n)}$ klar. Sei nun $j \geq 2$. Wir beweisen zunächst die folgende

Behauptung: Für alle $1 \leq j < i \leq n$ und alle $\lambda(j) \in \Omega_j, \lambda(j+1) \in \Omega_{j+1}, \dots, \lambda(i) \in \Omega_i$ gilt

$$P_n(\Lambda_j^{(n)} = \lambda(j), \Lambda_{j+1}^{(n)} = \lambda(j+1), \dots, \Lambda_i^{(n)} = \lambda(i)) = \frac{d_{\lambda(j)}d_{\lambda(i)}}{i!} \quad (3.14)$$

Beweis: Es gilt nach einer bekannten Formel für bedingte Wahrscheinlichkeiten und der Markov-Eigenschaft:

$$\begin{aligned} & P_n(\Lambda_j^{(n)} = \lambda(j), \Lambda_{j+1}^{(n)} = \lambda(j+1), \dots, \Lambda_i^{(n)} = \lambda(i)) \\ = & P_n(\Lambda_j^{(n)} = \lambda(j)) \cdot \prod_{k=j+1}^i P_n(\Lambda_k^{(n)} = \lambda(k) | \Lambda_j^{(n)} = \lambda(j), \dots, \Lambda_{k-1}^{(n)} = \lambda(k-1)) \\ \stackrel{\text{Satz 3.1.4}}{=} & \mathbb{P}_j(\{\lambda(j)\}) \cdot \prod_{k=j+1}^i P_n(\Lambda_k^{(n)} = \lambda(k) | \Lambda_{k-1}^{(n)} = \lambda(k-1)) \\ = & \frac{d_{\lambda(j)}^2}{j!} \cdot \prod_{k=j+1}^i \frac{d_{\lambda(k)}}{k d_{\lambda(k-1)}} \\ = & \frac{d_{\lambda(j)}^2}{j!} \frac{d_{\lambda(i)}}{d_{\lambda(j)}(j+1)(j+2) \cdots i} \\ = & \frac{d_{\lambda(j)}d_{\lambda(i)}}{i!} \end{aligned}$$

Daraus folgt für beliebige $\lambda(j) \in \Omega_j, \lambda(j+1) \in \Omega_{j+1}, \dots, \lambda(n) \in \Omega_n$ die Formel

$$\begin{aligned} & P_n(\Lambda_j^{(n)} = \lambda(j), \Lambda_{j+1}^{(n)} = \lambda(j+1), \dots, \Lambda_{n-1}^{(n)} = \lambda(n-1) | \Lambda_n^{(n)} = \lambda(n)) \\ = & \frac{P_n(\Lambda_j^{(n)} = \lambda(j), \Lambda_{j+1}^{(n)} = \lambda(j+1), \dots, \Lambda_{n-1}^{(n)} = \lambda(n-1), \Lambda_n^{(n)} = \lambda(n))}{P_n(\Lambda_n^{(n)} = \lambda(n))} \\ \stackrel{\text{Satz 3.1.4}}{=} & \frac{d_{\lambda(j)}d_{\lambda(n)}n!}{n!d_{\lambda(n)}^2} \\ = & \frac{d_{\lambda(j)}}{d_{\lambda(n)}}. \end{aligned} \quad (3.15)$$

3 Der zentrale Grenzwertsatz für character ratios

Aus (3.15) erhalten wir mit Hilfe der Transformationsformel für Erwartungswerte, angewendet auf den bedingten Erwartungswert für beliebiges $\lambda(n) \in \Omega_n$:

$$\begin{aligned}
& E_n[Y_j^{(n)} | \Lambda_n^{(n)} = \lambda(n)] \\
= & \sum_{\substack{\lambda(n-1) \in \Omega_{n-1}: \\ \lambda(n-1) \nearrow \lambda(n)}} \sum_{\substack{\lambda(n-2) \in \Omega_{n-2}: \\ \lambda(n-2) \nearrow \lambda(n-1)}} \cdots \sum_{\substack{\lambda(j) \in \Omega_j: \\ \lambda(j) \nearrow \lambda(j+1)}} \frac{\binom{j}{2} \chi_{\lambda(j)}(1, 2)}{d_{\lambda(j)}} \\
& \cdot P_n(\Lambda_j^{(n)} = \lambda(j), \Lambda_{j+1}^{(n)} = \lambda(j+1), \dots, \Lambda_{n-1}^{(n)} = \lambda(n-1) | \Lambda_n^{(n)} = \lambda(n)) \\
= & \sum_{\substack{\lambda(n-1) \in \Omega_{n-1}: \\ \lambda(n-1) \nearrow \lambda(n)}} \sum_{\substack{\lambda(n-2) \in \Omega_{n-2}: \\ \lambda(n-2) \nearrow \lambda(n-1)}} \cdots \sum_{\substack{\lambda(j) \in \Omega_j: \\ \lambda(j) \nearrow \lambda(j+1)}} \frac{\binom{j}{2} \chi_{\lambda(j)}(1, 2)}{d_{\lambda(j)}} \cdot \frac{d_{\lambda(j)}}{d_{\lambda(n)}} \\
= & \sum_{\substack{\lambda(n-1) \in \Omega_{n-1}: \\ \lambda(n-1) \nearrow \lambda(n)}} \sum_{\substack{\lambda(n-2) \in \Omega_{n-2}: \\ \lambda(n-2) \nearrow \lambda(n-1)}} \cdots \sum_{\substack{\lambda(j) \in \Omega_j: \\ \lambda(j) \nearrow \lambda(j+1)}} \frac{\binom{j}{2} \chi_{\lambda(j)}(1, 2)}{d_{\lambda(n)}} \\
= & \frac{\binom{j}{2}}{d_{\lambda(n)}} \sum_{\substack{\lambda(n-1) \in \Omega_{n-1}: \\ \lambda(n-1) \nearrow \lambda(n)}} \sum_{\substack{\lambda(n-2) \in \Omega_{n-2}: \\ \lambda(n-2) \nearrow \lambda(n-1)}} \cdots \sum_{\substack{\lambda(j) \in \Omega_j: \\ \lambda(j) \nearrow \lambda(j+1)}} \chi_{\lambda(j)}(1, 2) \tag{3.16}
\end{aligned}$$

Da $j \geq 2$ gilt, ist $(1, 2) \in \mathbb{S}_j$. Da die Einschränkung von Darstellungen (wie die von Abbildungen im Allgemeinen) transitiv ist, erhalten wir iterativ mit der Branching-Regel, Satz 2.2.35 (ii), insgesamt aus (3.16), dass gilt:

$$\begin{aligned}
E_n[Y_j^{(n)} | \Lambda_n^{(n)} = \lambda(n)] &= \frac{\binom{j}{2}}{d_{\lambda(n)}} \text{Res}_{(\mathbb{S}_j^n)} (\chi_{\lambda(n)})(1, 2) \\
&= \frac{\binom{j}{2} \chi_{\lambda(n)}(1, 2)}{d_{\lambda(n)}} \\
&= \frac{j(j-1)}{n(n-1)} Y_n^{(n)}
\end{aligned}$$

zu (iii): Die erste Gleichheit ist klar, da sich $Y_n^{(n)}$ und $T_n^{(n)}$ nur durch einen (von Null verschiedenen) konstanten Faktor voneinander unterscheiden. Aus (ii) folgt unmittelbar

$$\begin{aligned}
 E_n[X_j^{(n)}|Y_n^{(n)}] &= E_n[Y_j^{(n)}|Y_n^{(n)}] - E_n[Y_{j-1}^{(n)}|Y_n^{(n)}] \\
 &= \frac{j(j-1)}{n(n-1)}Y_n^{(n)} - \frac{(j-1)(j-2)}{n(n-1)}Y_n^{(n)} \\
 &= \frac{2(j-1)Y_n^{(n)}}{n(n-1)} = \frac{j-1}{\binom{n}{2}}Y_n^{(n)} \\
 &= \frac{j-1}{\sqrt{\binom{n}{2}}}T_n^{(n)}.
 \end{aligned}$$

zu (iv): Es ist $Y_{j-1}^{(n)}$ messbar bezüglich $\sigma(\Lambda_{j-1}^{(n)})$. Somit gilt nach (i)

$$\begin{aligned}
 E_n[X_j^{(n)}|\Lambda_{j-1}^{(n)}] &= E_n[Y_j^{(n)} - Y_{j-1}^{(n)}|\Lambda_{j-1}^{(n)}] = E_n[Y_j^{(n)}|\Lambda_{j-1}^{(n)}] - Y_{j-1}^{(n)} \\
 &= Y_{j-1}^{(n)} - Y_{j-1}^{(n)} = 0.
 \end{aligned}$$

Die einzigen bisher gefundenen Beweise der Aussagen (v) und (vi) sind rein kombinatorisch und werden daher ausgelassen. Für Aussage (v) verweisen wir auf [Ker93a] und für Aussage (vi) auf [Las04].

zu (vii): Dies folgt aus (v) mittels iterierter Erwartung:

$$E_n[(X_j^{(n)})^2] = E_n[E_n[(X_j^{(n)})^2|\Lambda_{j-1}^{(n)}]] = E_n[j-1] = j-1$$

□

Bemerkung 3.1.8. Für $1 \leq j \leq n$ sei $\mathcal{A}_j^{(n)} := \sigma(\Lambda_1^{(n)}, \dots, \Lambda_j^{(n)})$. Aus Lemma 3.1.7 (i) folgt aufgrund der Markov-Eigenschaft der Folge $(\Lambda_j^{(n)})_{1 \leq j \leq n}$ die Gleichung

$$E[Y_{j+1}^{(n)}|\mathcal{A}_j^{(n)}] = E[Y_{j+1}^{(n)}|\Lambda_1^{(n)}, \dots, \Lambda_j^{(n)}] = E[Y_{j+1}^{(n)}|\Lambda_j^{(n)}] = Y_j^{(n)}$$

für $1 \leq j \leq n-1$. Somit ist die Folge $(Y_j^{(n)})_{1 \leq j \leq n}$ ein (endliches) $(\mathcal{A}_j^{(n)})_{1 \leq j \leq n}$ -Martingal.

Seien Z_1, \dots, Z_n Unbestimmte. Dann wird für $1 \leq r \leq n$ das r -te elementarsymmetrische Polynom e_r definiert durch

$$e_r = e_r(Z_1, \dots, Z_n) = \sum_{1 \leq i_1 < \dots < i_r \leq n} Z_{i_1} \cdot \dots \cdot Z_{i_r}$$

Sei $\lambda \vdash n$ eine Partition und sei c_1, \dots, c_n irgendeine Anordnung der Inhalte der Boxen in λ . Dann definieren wir $e_r(\lambda) := e_r(c_1, \dots, c_n)$. Diese Definition ist wegen der Symmetrie des Polynoms e_r unabhängig von der Wahl der Auflistung der Inhalte c_1, \dots, c_n . Wir können somit e_r auch als Zufallsvariable $e_r : (\Omega_n, \mathcal{P}(\Omega_n), \mathbb{P}_n) \rightarrow (\mathbb{R}, \mathcal{B}^1)$ auffassen.

3 Der zentrale Grenzwertsatz für character ratios

Lemma 3.1.9. *Mit den obigen Definitionen und Bezeichnungen gilt $\mathbb{E}_n[e_r] = 0$ für alle $1 \leq r \leq n$.*

Beweisskizze: Im Fall $r = n$ besteht $e_n(\lambda)$ für jedes λ aus nur einem Summanden, in dem der Faktor 0 als Inhalt der Box an der Stelle $(1, 1)$ auftritt. Somit ist e_n konstant gleich 0 und mithin auch $\mathbb{E}_n[e_n] = 0$.

Sei jetzt $1 \leq r < n$ vorausgesetzt. In diesem Fall verwenden wir die Theorie der sogenannten *Murphy-Elemente* R_2, R_3, \dots, R_n . Für $2 \leq i \leq n$ ist das i -te Murphy-Element R_i definiert durch $R_i := \sum_{1 \leq j < i} (i, j) \in \mathbb{C}[\mathbb{S}_n]$. Es sei $c \in \mathbb{C}[\mathbb{S}_n]$ das Element aus der Gruppen-Algebra von \mathbb{S}_n , das die Summe aller Permutationen mit genau $n - r$ nicht-trivialen Zykeln ist. Diaconis und Greene zeigen in [DG89], dass die Identität $c = e_r(R_2, \dots, R_n)$ gilt. Durch Multiplikation von links kann man ein Element der Gruppen-Algebra stets als Endomorphismus auf dem entsprechenden Modul auffassen. Murphy beweist in [Mur81], dass in diesem Sinne die Murphy-Elemente R_2, \dots, R_n als Endomorphismen jedes Specht Moduls S^λ , $\lambda \vdash n$, bzgl. einer Orthogonalbasis simultan diagonalisierbar sind und berechnet auch die entsprechenden Eigenwerte. Eine Konsequenz daraus ist, dass das Element c aufgefasst als Endomorphismus von S^λ eine Homothetie ist zum Skalar $e_r(\lambda)$ und somit die Spur $\text{Tr}_{S^\lambda}(c) = e_r(\lambda)d_\lambda$ hat. Wir wissen aus Satz 2.1.33, dass jeder Modul S^λ in der regulären Darstellung $\mathbb{C}[\mathbb{S}_n]$ mit Vielfachheit d_λ vorkommt, so dass die Spur von c aufgefasst als Endomorphismus von $\mathbb{C}[\mathbb{S}_n]$ gegeben ist durch

$$\text{Tr}_{\mathbb{C}[\mathbb{S}_n]}(c) = \sum_{\lambda \vdash n} d_\lambda \text{Tr}_{S^\lambda}(c) = \sum_{\lambda \vdash n} d_\lambda^2 e_r(\lambda) = n! \mathbb{E}_n[e_r].$$

Sind π_1, \dots, π_m alle Permutationen mit genau $n - r$ nicht-trivialen Zykeln, so gilt andererseits

$$\text{Tr}_{\mathbb{C}[\mathbb{S}_n]}(c) = \sum_{i=1}^m \text{Tr}_{\mathbb{C}[\mathbb{S}_n]}(\pi_i) = \sum_{i=1}^m \chi_{reg}(\pi_i) = 0,$$

nach Proposition 2.1.32 (b). Somit folgt $\mathbb{E}_n[e_r] = 0$ wie behauptet. □

Das nächste Lemma stellt obere Schranken für $E_n[|X_n^{(n)}|^3]$ und für $E_n[|T_{n-1}^{(n)}(X_n^{(n)})^3|]$ bereit.

Lemma 3.1.10. *Es sei $n \geq 3$ eine natürliche Zahl. Dann gelten die folgenden Abschätzungen:*

- (i) $E_n[|X_n^{(n)}|^3] \leq (n - 1)\sqrt{2n - 3}$
- (ii) $E_n[|T_{n-1}^{(n)}(X_n^{(n)})^3|] \leq (n - 1)\sqrt{2n - 3}$

3.1 Das Plancherel-Maß und character ratios

Beweis: Nach der Cauchy-Schwarz-Ungleichung gilt

$$\begin{aligned}
 E_n[|X_n^{(n)}|^3] &= E_n[|X_n^{(n)}| \cdot |X_n^{(n)}|^2] \\
 &\leq \sqrt{E_n[|X_n^{(n)}|^2]} \cdot \sqrt{E_n[|X_n^{(n)}|^4]} \\
 &\stackrel{\text{Lemma 3.1.7(vii)}}{=} \sqrt{n-1} \sqrt{E_n[|X_n^{(n)}|^4]}
 \end{aligned} \tag{3.17}$$

Nach Lemma 3.1.7 (vi) erhalten wir

$$\begin{aligned}
 E_n[|X_n^{(n)}|^4] &= E_n[E_n[|X_n^{(n)}|^4 | \Lambda_{n-1}^{(n)}]] \\
 &= E_n \left[\binom{n}{2} + 3 \sum_{B \in \Lambda_{n-1}^{(n)}} c(B)^2 \right] \\
 &= \binom{n}{2} + 3E_n \left[\sum_{B \in \Lambda_{n-1}^{(n)}} c(B)^2 \right]
 \end{aligned} \tag{3.18}$$

Es ist

$$\begin{aligned}
 \left(\sum_{B \in \Lambda_{n-1}^{(n)}} c(B) \right)^2 &= \sum_{B \in \Lambda_{n-1}^{(n)}} c(B)^2 + \sum_{\substack{B, B' \in \Lambda_{n-1}^{(n)} \\ B \neq B'}} c(B)c(B') \\
 &= \sum_{B \in \Lambda_{n-1}^{(n)}} c(B)^2 + 2 \cdot e_2(\Lambda_{n-1}^{(n)})
 \end{aligned} \tag{3.19}$$

nach Definition von $e_2(\lambda)$ für $\lambda \vdash n-1$. Nach Definition von Kerov's growth process und wegen Satz 3.1.6 gilt

$$\left(\sum_{B \in \Lambda_{n-1}^{(n)}} c(B) \right)^2 = \left(\sum_{j=1}^{n-1} X_j^{(n)} \right)^2 = \binom{n-1}{2} (T_{n-1}^{(n)})^2 \tag{3.20}$$

Aus (3.18), (3.19) und (3.20) erhalten wir

3 Der zentrale Grenzwertsatz für character ratios

$$\begin{aligned}
E_n[|X_n^{(n)}|^4] &= \binom{n}{2} + 3E_n \left[\binom{n-1}{2} (T_{n-1}^{(n)})^2 - 2e_2(\Lambda_{n-1}^{(n)}) \right] \\
&= \binom{n}{2} + 3 \binom{n-1}{2} E_n[(T_{n-1}^{(n)})^2] - 6E_n[e_2(\Lambda_{n-1}^{(n)})] \\
&= \binom{n}{2} + 3 \binom{n-1}{2} E_n[(T_n(\Lambda_{n-1}^{(n)}))^2] - 6E_n[e_2(\Lambda_{n-1}^{(n)})] \\
&\stackrel{\text{Satz 3.1.4}}{=} \binom{n}{2} + 3 \binom{n-1}{2} \mathbb{E}_{n-1}[T_{n-1}^2] - 6\mathbb{E}_{n-1}[e_2] \\
&\stackrel{\substack{\text{Prop. 3.1.2} \\ \text{Lemma 3.1.9}}}{=} \binom{n}{2} + 3 \binom{n-1}{2} - 6 \cdot 0 \\
&= \frac{n(n-1)}{2} + \frac{3(n-1)(n-2)}{2} = \frac{n-1}{2}(n+3n-6) \\
&= (n-1)(2n-3) \tag{3.21}
\end{aligned}$$

Wir weisen darauf hin, dass wir für das vierte Gleichheitszeichen in (3.21) die Aussage von Satz 3.1.4 für die Transformationsformel für Erwartungswerte angewendet haben. Setzen wir (3.21) in (3.17) ein, so erhalten wir sofort die Behauptung von (i).

zu (ii): Da $T_{n-1}^{(n)} = T_{n-1}(\Lambda_{n-1}^{(n)})$ messbar bzgl. $\sigma(\Lambda_{n-1}^{(n)})$ ist, erhalten wir wieder unter Verwendung der Cauchy-Schwarz-Ungleichung:

$$\begin{aligned}
E_n[|T_{n-1}^{(n)}(X_n^{(n)})^3|] &= E_n[E_n[|T_{n-1}^{(n)}(X_n^{(n)})^3| \mid \Lambda_{n-1}^{(n)}]] \\
&= E_n[|T_{n-1}^{(n)}| E_n[|X_n^{(n)}|^3 \mid \Lambda_{n-1}^{(n)}]] \\
&\leq \sqrt{E_n[(T_{n-1}^{(n)})^2] \cdot E_n[(E_n[|X_n^{(n)}|^3 \mid \Lambda_{n-1}^{(n)}])^2]} \\
&\stackrel{\text{Prop. 3.1.2}}{=} \sqrt{E_n[(E_n[|X_n^{(n)}|^3 \mid \Lambda_{n-1}^{(n)}])^2]} \tag{3.22}
\end{aligned}$$

Nach der bedingten Version der Cauchy-Schwarz-Ungleichung (siehe etwa [Kle06]) in Verbindung mit Lemma 3.1.7 (v) gilt

$$\begin{aligned}
(E_n[|X_n^{(n)}|^3 \mid \Lambda_{n-1}^{(n)}])^2 &= (E_n[|X_n^{(n)}| |X_n^{(n)}|^2 \mid \Lambda_{n-1}^{(n)}])^2 \\
&\leq E_n[|X_n^{(n)}|^2 \mid \Lambda_{n-1}^{(n)}] E_n[|X_n^{(n)}|^4 \mid \Lambda_{n-1}^{(n)}] \\
&= (n-1) E_n[|X_n^{(n)}|^4 \mid \Lambda_{n-1}^{(n)}] \tag{3.23}
\end{aligned}$$

Indem wir (3.23) in (3.22) einsetzen, erhalten wir aus dem Satz von der totalen Erwartung:

$$E_n[|T_{n-1}^{(n)}(X_n^{(n)})^3|] \leq \sqrt{(n-1)E_n[E_n[|X_n^{(n)}|^4|\Lambda_{n-1}^{(n)}]]} = \sqrt{(n-1)E_n[|X_n^{(n)}|^4]} \quad (3.24)$$

Aus (3.21) und (3.24) ergibt sich nun unmittelbar die Behauptung von (ii). \square

3.2 Das Berry-Esséen-Theorem für character ratios

In diesem Abschnitt wollen wir den zentralen Grenzwertsatz vom Berry-Esséen-Typ für character ratios, ausgewertet in der Konjugationsklasse der Transpositionen, beweisen und eine explizite Berry-Esséen-Konstante berechnen. Der Beweis folgt der Darstellung im Artikel [Ful06a] und verwendet die Steinsche Methode in Form des aus Abschnitt 1.2 bekannten Induktionsansatzes von Bolthausen. Hier ist jedoch die uns interessierende Zufallsvariable T_n bzw. $T_n^{(n)}$ keine Summe unabhängiger Zufallsvariablen, so dass die im Beweis von Satz 1.2.3 verwendete „Derandomisierung“ mittels Satz 1.2.5 nicht ohne Weiteres übernommen werden kann. So behaupten etwa Shao und Chen in [BC05] sinngemäß, dass das induktive Argument (zum Beispiel das von Bolthausen) im Allgemeinen nur dann gut funktioniert, wenn die bedingte Verteilung der betrachteten Summe von Zufallsvariablen gegeben einen Summanden der unbedingten Verteilung der um diesen Summanden reduzierten Summe ähnelt. Diese Bedingung ist natürlich in der Situation von unabhängigen Summanden, wie in Abschnitt 1.2, stets erfüllt. Als Ausweg muss an einer Stelle im Beweis eine sogenannte Konzentrationsungleichung verwendet werden, welche im Artikel [Ful05] implizit bewiesen wird, um die gewünschte Abschätzung nach oben zu erhalten. Diese Ungleichung formulieren wir als ein vorangehendes Lemma. Dass dieser Weg überhaupt erfolgreich ist, liegt daran, dass die Zufallsvariable $X_n^{(n)}$, auf die wir die Konzentrationsungleichung anwenden wollen, beschränkt ist, wie wir noch sehen werden. Leider führt diese Beweismethode dazu, dass wir eine recht große Konstante ($C=808$) erhalten werden.

In diesem Abschnitt übernehmen wir im Wesentlichen die Bezeichnungen aus dem ersten Kapitel, schreiben aber statt $h_{z,\lambda}$ und $f_{z,\lambda}$ lieber $h_{z,b}$ und $f_{z,b}$, da in diesem Kapitel λ stets eine Partition bezeichnen soll.

Lemma 3.2.1. *Für alle $n \in \mathbb{N}$ gilt die Ungleichung $P_n(|X_n^{(n)}| > 2e\sqrt{n}) \leq 2e^{-2e\sqrt{n}}$.*

Beweis: Als Betrag des Inhalts einer Box in dem Young-Diagramm $\Lambda_n^{(n)}$ mit genau n Boxen, ist $|X_n^{(n)}| \leq \max(Z_1, S_1)$, wobei Z_1 die Länge der ersten Zeile und S_1 die Länge der ersten Spalte von $\Lambda_n^{(n)}$ bezeichnen. Somit gilt

3 Der zentrale Grenzwertsatz für character ratios

$$\{|X_n^{(n)}| > 2e\sqrt{n}\} \subseteq \{Z_1 > 2e\sqrt{n}\} \cup \{S_1 > 2e\sqrt{n}\},$$

so dass

$$P_n(|X_n^{(n)}| > 2e\sqrt{n}) \leq P_n(Z_1 > 2e\sqrt{n}) + P_n(S_1 > 2e\sqrt{n}) \quad (3.25)$$

folgt. Da $\Lambda_n^{(n)}$ nach Satz 3.1.4 gemäß dem Plancherel-Maß \mathbb{P}_n verteilt ist, folgt aus dem RSK-Algorithmus, dass sowohl Z_1 als auch S_1 unter P_n dieselbe Verteilung haben, wie die längste aufsteigende Teilfolge einer zufälligen, gleichverteilten Permutation $\pi \in \mathbb{S}_n$. Es sei $\pi \in \mathbb{S}_n$ eine gleichverteilte Permutation und es bezeichne I_n die Länge ihrer längsten aufsteigenden Teilfolge. Wir wollen nach [Ste97] zeigen, dass $P(I_n \geq 2e\sqrt{n}) < e^{-2e\sqrt{n}}$ gilt. Sei $1 \leq k \leq n$ und seien $1 \leq i_1 < i_2 < \dots < i_k \leq n$ zunächst fest gewählte Indizes. Dann gilt

$$P(\{\pi : \pi(i_1) < \pi(i_2) < \dots < \pi(i_k)\}) = \frac{\binom{n}{k}}{n!} (n-k)! = \frac{1}{k!},$$

denn es gibt $\binom{n}{k}$ Möglichkeiten, die Bildmenge $\{\pi(i_1), \pi(i_2), \dots, \pi(i_k)\}$ zu wählen und für jede solche Wahl gibt es genau $(n-k)!$ Möglichkeiten für die Bilder der Indizes $i \notin \{i_1, i_2, \dots, i_k\}$. Somit folgt

$$\begin{aligned} P(I_n \geq k) &\leq \sum_{1 \leq i_1 < i_2 < \dots < i_k \leq n} P(\{\pi : \pi(i_1) < \pi(i_2) < \dots < \pi(i_k)\}) \\ &= \binom{n}{k} \frac{1}{k!} = \frac{n(n-1) \cdot \dots \cdot (n-k+1)}{(k!)^2}. \end{aligned} \quad (3.26)$$

Nun sei speziell $k := \min\{m \in \mathbb{N} : m \geq 2e\sqrt{n}\}$. Aufgrund der Reihenentwicklung von e^k ist $e^k \geq \frac{k^k}{k!}$. Somit folgt wegen $I_n \in \mathbb{N}$ und nach (3.26)

$$\begin{aligned} P(I_n \geq 2e\sqrt{n}) &= P(I_n \geq k) \leq \frac{n(n-1) \cdot \dots \cdot (n-k+1)}{(k!)^2} \\ &\leq \frac{e^k}{k^k} \frac{e^k}{k^k} n(n-1) \cdot \dots \cdot (n-k+1) \\ &\leq \frac{e^k}{k^k} e^k \left(\frac{n}{k}\right)^k \leq \frac{e^k}{k^k} e^k \left(\frac{\sqrt{n}}{2e}\right)^k \\ &= \frac{e^k (\sqrt{n})^k}{k^k 2^k} = e^{k(1+\log \sqrt{n} - \log k - \log 2)} \\ &\leq e^{k(1+\log \sqrt{n} - 2 \log 2 - 1 - \log \sqrt{n})} = e^{-k 2 \log 2} \\ &\leq e^{-2 \log 2 \cdot 2e\sqrt{n}} < e^{-2e\sqrt{n}}, \end{aligned} \quad (3.27)$$

3.2 Das Berry-Esséen-Theorem für character ratios

denn es ist $2 \log 2 = \log 4 > \log e = 1$. Aus (3.25) und (3.27) folgt dann aber unmittelbar die Behauptung. □

Wir wollen nun das Hauptresultat dieser Arbeit formulieren:

Satz 3.2.2 (Berry-Esséen-Theorem für character ratios). *Es gibt eine Konstante $C > 0$, so dass für alle $n \in \mathbb{N}$ gilt:*

$$\sup_{z \in \mathbb{R}} |\mathbb{P}_n(T_n \leq z) - \Phi(z)| \leq \frac{C}{\sqrt{n}}$$

Beweis: Wir dürfen ohne Einschränkung $n \geq 3$ voraussetzen, da wir andernfalls einfach die Konstante C vergrößern könnten. Wir führen zunächst etwas Notation ein, die der im Beweis von Satz 1.2.3 sehr ähnelt: Für $b > 0$ und $n \in \mathbb{N}$ sei

$$\delta(b, n) := \sup_{z \in \mathbb{R}} |\mathbb{E}_n[h_{z,b}(T_n)] - N(h_{z,b})| \stackrel{\text{Satz 3.1.4}}{=} \sup_{z \in \mathbb{R}} |E_n[h_{z,b}(T_n^{(n)})] - N(h_{z,b})|$$

und für $n \in \mathbb{N}$ sei

$$\begin{aligned} \delta(n) &:= \sup_{z \in \mathbb{R}} |\mathbb{E}_n[h_z(T_n)] - N(h_z)| = \sup_{z \in \mathbb{R}} |\mathbb{P}_n(T_n \leq z) - \Phi(z)| \\ &\stackrel{\text{Satz 3.1.4}}{=} \sup_{z \in \mathbb{R}} |P_n(T_n^{(n)} \leq z) - \Phi(z)|. \end{aligned}$$

Wir fixieren nun zunächst $z \in \mathbb{R}$ und $b > 0$ und schreiben daher auch einfach f für die Stein-Lösung $f_{z,b}$. Dann gilt wie gehabt

$$E_n[h_{z,b}(T_n^{(n)})] - N(h_{z,b}) = E_n[f'(T_n^{(n)}) - T_n^{(n)} f(T_n^{(n)})]. \quad (3.28)$$

Wir wollen nun zunächst die rechte Seite von (3.28) etwas umschreiben: Nach Lemma 3.1.7 (iii) mit $j = n$ gilt

$$\begin{aligned} E_n[X_n^{(n)} f(T_n^{(n)})] &= E_n[E_n[X_n^{(n)} f(T_n^{(n)}) | T_n^{(n)}]] = E_n[f(T_n^{(n)}) E_n[X_n^{(n)} | T_n^{(n)}]] \\ &= \frac{n-1}{\sqrt{\binom{n}{2}}} E_n[T_n^{(n)} f(T_n^{(n)})], \end{aligned}$$

also ist $E_n[T_n^{(n)} f(T_n^{(n)})] = \frac{\sqrt{\binom{n}{2}}}{n-1} E_n[X_n^{(n)} f(T_n^{(n)})]$. Somit gilt

$$E_n[f'(T_n^{(n)}) - T_n^{(n)} f(T_n^{(n)})] = E_n \left[f'(T_n^{(n)}) - \frac{\sqrt{\binom{n}{2}}}{n-1} X_n^{(n)} f(T_n^{(n)}) \right]. \quad (3.29)$$

3 Der zentrale Grenzwertsatz für character ratios

Nach Lemma 3.1.7 (v) (für die letzte Identität) gilt wegen der $\sigma(\Lambda_{n-1}^{(n)})$ -Messbarkeit von $T_{n-1}^{(n)}$:

$$\begin{aligned}
E_n \left[\frac{(X_n^{(n)})^2}{n-1} f' \left(\sqrt{\frac{n-2}{n}} T_{n-1}^{(n)} \right) \right] &= E_n \left[E_n \left[\frac{(X_n^{(n)})^2}{n-1} f' \left(\sqrt{\frac{n-2}{n}} T_{n-1}^{(n)} \right) \middle| \Lambda_{n-1}^{(n)} \right] \right] \\
&= E_n \left[f' \left(\sqrt{\frac{n-2}{n}} T_{n-1}^{(n)} \right) E_n \left[\frac{(X_n^{(n)})^2}{n-1} \middle| \Lambda_{n-1}^{(n)} \right] \right] \\
&= E_n [f' \left(\sqrt{\frac{n-2}{n}} T_{n-1}^{(n)} \right)] \tag{3.30}
\end{aligned}$$

Ebenso gilt nach Aussage (iv) in Lemma 3.1.7 die Identität:

$$\begin{aligned}
E_n \left[X_n^{(n)} f \left(\sqrt{\frac{n-2}{n}} T_{n-1}^{(n)} \right) \right] &= E_n \left[E_n \left[X_n^{(n)} f \left(\sqrt{\frac{n-2}{n}} T_{n-1}^{(n)} \right) \middle| \Lambda_{n-1}^{(n)} \right] \right] \\
&= E_n \left[f \left(\sqrt{\frac{n-2}{n}} T_{n-1}^{(n)} \right) \underbrace{E_n [X_n^{(n)} | \Lambda_{n-1}^{(n)}]}_{=0} \right] \\
&= 0 \tag{3.31}
\end{aligned}$$

Aus (3.29), (3.30) und (3.31) erhalten wir nun die Gleichung

$$\begin{aligned}
&E_n [f'(T_n^{(n)}) - T_n^{(n)} f(T_n^{(n)})] \\
&= E_n [f'(T_n^{(n)})] + E_n \left[\frac{(X_n^{(n)})^2}{n-1} f' \left(\sqrt{\frac{n-2}{n}} T_{n-1}^{(n)} \right) - f' \left(\sqrt{\frac{n-2}{n}} T_{n-1}^{(n)} \right) \right] \\
&\quad - E_n \left[\frac{\sqrt{\binom{n}{2}}}{n-1} X_n^{(n)} f(T_n^{(n)}) - \frac{\sqrt{\binom{n}{2}}}{n-1} X_n^{(n)} f \left(\sqrt{\frac{n-2}{n}} T_{n-1}^{(n)} \right) \right] \\
&= E_n \left[f'(T_n^{(n)}) - f' \left(\sqrt{\frac{n-2}{n}} T_{n-1}^{(n)} \right) \right] \\
&\quad - E_n \left[\frac{\sqrt{\binom{n}{2}} X_n^{(n)}}{n-1} \left(f(T_n^{(n)}) - f \left(\sqrt{\frac{n-2}{n}} T_{n-1}^{(n)} \right) \right) \right] \\
&\quad + E_n \left[\frac{(X_n^{(n)})^2}{n-1} f' \left(\sqrt{\frac{n-2}{n}} T_{n-1}^{(n)} \right) \right]. \tag{3.32}
\end{aligned}$$

3.2 Das Berry-Esséen-Theorem für character ratios

Nach dem Hauptsatz der Differential- und Integralrechnung gilt punktweise

$$\begin{aligned}
 f(T_n^{(n)}) - f\left(\sqrt{\frac{n-2}{n}}T_{n-1}^{(n)}\right) &= f\left(\sqrt{\frac{n-2}{n}}T_{n-1}^{(n)} + \frac{X_n^{(n)}}{\sqrt{\binom{n}{2}}}\right) - f\left(\sqrt{\frac{n-2}{n}}T_{n-1}^{(n)}\right) \\
 &= \int_0^{\frac{X_n^{(n)}}{\sqrt{\binom{n}{2}}}} f'\left(\sqrt{\frac{n-2}{n}}T_{n-1}^{(n)} + u\right) du \\
 &\stackrel{u:=t\frac{X_n^{(n)}}{\sqrt{\binom{n}{2}}}}{=} \frac{X_n^{(n)}}{\sqrt{\binom{n}{2}}} \int_0^1 f'\left(\sqrt{\frac{n-2}{n}}T_{n-1}^{(n)} + t\frac{X_n^{(n)}}{\sqrt{\binom{n}{2}}}\right) dt \quad (3.33)
 \end{aligned}$$

Wegen

$$E_n \left[\frac{(X_n^{(n)})^2}{n-1} f'\left(\sqrt{\frac{n-2}{n}}T_{n-1}^{(n)}\right) \right] = E_n \left[\frac{(X_n^{(n)})^2}{n-1} \int_0^1 f'\left(\sqrt{\frac{n-2}{n}}T_{n-1}^{(n)}\right) dt \right]$$

folgt aus (3.33) und (3.32) die Gleichung:

$$\begin{aligned}
 E_n[f'(T_n^{(n)}) - T_n^{(n)}f(T_n^{(n)})] &= E_n \left[f'(T_n^{(n)}) - f'\left(\sqrt{\frac{n-2}{n}}T_{n-1}^{(n)}\right) \right] \\
 -E_n \left[\frac{(X_n^{(n)})^2}{n-1} \int_0^1 \left(f'\left(\sqrt{\frac{n-2}{n}}T_{n-1}^{(n)} + t\frac{X_n^{(n)}}{\sqrt{\binom{n}{2}}}\right) - f'\left(\sqrt{\frac{n-2}{n}}T_{n-1}^{(n)}\right) \right) dt \right] & \quad (3.34)
 \end{aligned}$$

Wir beginnen nun damit, (3.34) nach oben abzuschätzen. Zunächst betrachten wir den Term $\left| E_n \left[f'(T_n^{(n)}) - f'\left(\sqrt{\frac{n-2}{n}}T_{n-1}^{(n)}\right) \right] \right|$. Nach Lemma 1.2.4 (a), (b) und (c) gilt

$$\begin{aligned}
& \left| E_n \left[f'(T_n^{(n)}) - f' \left(\sqrt{\frac{n-2}{n}} T_{n-1}^{(n)} \right) \right] \right| \\
&= \left| E_n \left[f' \left(\sqrt{\frac{n-2}{n}} T_{n-1}^{(n)} + \frac{X_n^{(n)}}{\sqrt{\binom{n}{2}}} \right) - f' \left(\sqrt{\frac{n-2}{n}} T_{n-1}^{(n)} \right) \right] \right| \\
&\leq E_n \left[\left| f' \left(\sqrt{\frac{n-2}{n}} T_{n-1}^{(n)} + \frac{X_n^{(n)}}{\sqrt{\binom{n}{2}}} \right) - f' \left(\sqrt{\frac{n-2}{n}} T_{n-1}^{(n)} \right) \right| \right] \\
&\leq E_n \left[\frac{|X_n^{(n)}|}{\sqrt{\binom{n}{2}}} \left(\sqrt{\frac{\pi}{8}} + 2\sqrt{\frac{n-2}{n}} |T_{n-1}^{(n)}| + \frac{1}{b} \int_0^1 I_{(z; z+b]} \left(\sqrt{\frac{n-2}{n}} T_{n-1}^{(n)} + s \frac{|X_n^{(n)}|}{\sqrt{\binom{n}{2}}} \right) ds \right) \right] \\
&= \frac{\sqrt{\frac{\pi}{8}} E_n[|X_n^{(n)}|]}{\sqrt{\binom{n}{2}}} + 2 \frac{\sqrt{\frac{n-2}{n}}}{\sqrt{\binom{n}{2}}} E_n[|X_n^{(n)} T_{n-1}^{(n)}|] \\
&+ E_n \left[\frac{|X_n^{(n)}|}{b \sqrt{\binom{n}{2}}} \int_0^1 I_{(z; z+b]} \left(\sqrt{\frac{n-2}{n}} T_{n-1}^{(n)} + s \frac{|X_n^{(n)}|}{\sqrt{\binom{n}{2}}} \right) ds \right] \tag{3.35}
\end{aligned}$$

Nach der Jensen-Ungleichung und Lemma 3.1.7 (vii) gilt $E_n[|X_n^{(n)}|] \leq \sqrt{E_n[|X_n^{(n)}|^2]} = \sqrt{n-1}$. Somit ergibt sich für den ersten Summanden in (3.35) die Abschätzung $\frac{\sqrt{\frac{\pi}{8}} E_n[|X_n^{(n)}|]}{\sqrt{\binom{n}{2}}} \leq \frac{\sqrt{\pi}}{2\sqrt{n}}$. Aus Lemma 3.1.7 (vii) und Proposition 3.1.2 (b) in Verbindung mit der Cauchy-Schwarz-Ungleichung erhalten wir $E_n[|X_n^{(n)} T_{n-1}^{(n)}|] \leq \sqrt{E_n[(X_n^{(n)})^2] E_n[(T_{n-1}^{(n)})^2]} = \sqrt{n-1}$. Somit erhalten wir für den zweiten Summanden in (3.35) die Abschätzung

$$2 \frac{\sqrt{\frac{n-2}{n}}}{\sqrt{\binom{n}{2}}} E_n[|X_n^{(n)} T_{n-1}^{(n)}|] \leq 2 \frac{\sqrt{\frac{n-2}{n}}}{\sqrt{\binom{n}{2}}} \sqrt{n-1} = \frac{2\sqrt{2}\sqrt{\frac{n-2}{n}}}{\sqrt{n}} \leq \frac{2\sqrt{2}}{\sqrt{n}}.$$

Auf den dritten Summanden in (3.35) können wir nun nicht einfach eine zu Gleichung (1.19) analoge „Derandomisierung“ anwenden, da $T_{n-1}^{(n)}$ und $X_n^{(n)}$ nicht stochastisch unabhängig sind, aber die Aussage von Satz 1.2.5 im Allgemeinen nur für unabhängige Zufallsvariablen gilt. Stattdessen schreiben wir ihn als Summe $S_1 + S_2$, wobei

3.2 Das Berry-Esséen-Theorem für character ratios

$$S_1 := E_n \left[I_{\{|X_n^{(n)}| \leq 2e\sqrt{n}\}} \frac{|X_n^{(n)}|}{b\sqrt{\binom{n}{2}}} \int_0^1 I_{(z; z+b]} \left(\sqrt{\frac{n-2}{n}} T_{n-1}^{(n)} + s \frac{|X_n^{(n)}|}{\sqrt{\binom{n}{2}}} \right) ds \right]$$

und

$$S_2 := E_n \left[I_{\{|X_n^{(n)}| > 2e\sqrt{n}\}} \frac{|X_n^{(n)}|}{b\sqrt{\binom{n}{2}}} \int_0^1 I_{(z; z+b]} \left(\sqrt{\frac{n-2}{n}} T_{n-1}^{(n)} + s \frac{|X_n^{(n)}|}{\sqrt{\binom{n}{2}}} \right) ds \right]$$

seien. Offensichtlich ist

$$S_1 \leq \frac{2e\sqrt{n}}{b\sqrt{\binom{n}{2}}} E_n \left[I_{(z - \frac{2e\sqrt{n}}{\sqrt{\binom{n}{2}}}; z + b + \frac{2e\sqrt{n}}{\sqrt{\binom{n}{2}}}] \left(\sqrt{\frac{n-2}{n}} T_{n-1}^{(n)} \right) \right].$$

Nun verwenden wir die folgende Aussage, die der Aussage von Lemma 1.2.4 (e) sehr ähnelt und genauso bewiesen werden kann, so dass wir den Beweis auslassen können.

Für alle reellen Zahlen $a < d$ und alle Konstanten $c \in \mathbb{R}$ gilt

$$E_n[I_{(a;d]}(T_{n-1}^{(n)} + c)] \leq \frac{d-a}{\sqrt{2\pi}} + 2\delta(n-1) \quad (3.36)$$

Daraus erhalten wir zunächst die Abschätzung

$$\begin{aligned} & E_n \left[I_{(z - \frac{2e\sqrt{n}}{\sqrt{\binom{n}{2}}}; z + b + \frac{2e\sqrt{n}}{\sqrt{\binom{n}{2}}}] \left(\sqrt{\frac{n-2}{n}} T_{n-1}^{(n)} \right) \right] \\ &= E_n \left[I_{(\sqrt{\frac{n}{n-2}}(z - \frac{2e\sqrt{n}}{\sqrt{\binom{n}{2}}}); \sqrt{\frac{n}{n-2}}(z + b + \frac{2e\sqrt{n}}{\sqrt{\binom{n}{2}}})] (T_{n-1}^{(n)}) \right] \\ &\leq \sqrt{\frac{n}{(n-2)2\pi}} \left(z + b + \frac{2e\sqrt{n}}{\sqrt{\frac{n(n-1)}{2}}} - (z - \frac{2e\sqrt{n}}{\sqrt{\frac{n(n-1)}{2}}}) \right) + 2\delta(n-1) \\ &= \sqrt{\frac{n}{(n-2)2\pi}} \left(b + \frac{4e\sqrt{2}}{\sqrt{n-1}} \right) + 2\delta(n-1) \\ &= \sqrt{\frac{n}{(n-2)2\pi}} b + \frac{4e\sqrt{n}}{\sqrt{\pi}\sqrt{n-1}\sqrt{n-2}} + 2\delta(n-1) \end{aligned}$$

3 Der zentrale Grenzwertsatz für character ratios

und weiter

$$\begin{aligned}
S_1 &\leq \frac{2e\sqrt{n}}{b\sqrt{\frac{n(n-1)}{2}}} E_n \left[I_{\left(z - \frac{2e\sqrt{n}}{\sqrt{\binom{n}{2}}}; z + b + \frac{2e\sqrt{n}}{\sqrt{\binom{n}{2}}}\right]} \left(\sqrt{\frac{n-2}{n}} T_{n-1}^{(n)} \right) \right] \\
&= \frac{2e\sqrt{2}}{b\sqrt{n-1}} E_n \left[I_{\left(z - \frac{2e\sqrt{n}}{\sqrt{\binom{n}{2}}}; z + b + \frac{2e\sqrt{n}}{\sqrt{\binom{n}{2}}}\right]} \left(\sqrt{\frac{n-2}{n}} T_{n-1}^{(n)} \right) \right] \\
&\leq \frac{2e\sqrt{n}}{\sqrt{\pi}\sqrt{n-2}\sqrt{n-1}} + \frac{8e^2\sqrt{2}\sqrt{n}}{b\sqrt{\pi}\sqrt{n-2}(n-1)} + \frac{4e\sqrt{2}}{b\sqrt{n-1}}\delta(n-1) \\
&\leq \frac{3e\sqrt{2}}{\sqrt{\pi}\sqrt{n}} + \frac{12e^2\sqrt{6}}{b\sqrt{\pi}n} + \frac{4e\sqrt{3}}{b\sqrt{n}}\delta(n-1), \tag{3.37}
\end{aligned}$$

wobei wir für die letzte Ungleichung die Abschätzungen $\frac{n}{n-1} \leq \frac{3}{2}$ und $\frac{n}{n-2} \leq 3$, gültig jeweils für alle $n \geq 3$, verwendet haben. Um S_2 abzuschätzen, verwenden wir zunächst die Tatsache, dass $X_n^{(n)}$ als Inhalt einer Box eines Young-Diagramms mit n Boxen betragsmäßig nach oben durch $n-1$ abgeschätzt werden kann, da der betragsmäßig maximale Inhalt an der Box in der ersten Zeile ganz rechts oder in der ersten Spalte ganz unten erreicht wird. Da solch ein Diagramm aber höchstens n Zeilen oder Spalten haben kann, folgt diese Abschätzung. Indem wir den Indikator unter dem ds -Integral nach oben durch 1 abschätzen, erhalten wir mit Hilfe von Lemma 3.2.1 die Ungleichung

$$\begin{aligned}
S_2 &\leq \frac{n-1}{b\sqrt{\frac{n(n-1)}{2}}} P_n(|X_n^{(n)}| > 2e\sqrt{n}) = \frac{\sqrt{n-1}\sqrt{2}}{b\sqrt{n}} P_n(|X_n^{(n)}| > 2e\sqrt{n}) \\
&\leq \frac{2\sqrt{2}\sqrt{n-1}}{b\sqrt{n}} e^{-2e\sqrt{n}} \leq \frac{2\sqrt{2}}{b} e^{-2e\sqrt{n}}. \tag{3.38}
\end{aligned}$$

Indem wir nun die oberen Schranken aller drei Summanden in (3.35) addieren, erhalten wir die Abschätzung

$$\begin{aligned}
&\left| E_n \left[f'(T_n^{(n)}) - f' \left(\sqrt{\frac{n-2}{n}} T_{n-1}^{(n)} \right) \right] \right| \\
&\leq \frac{\sqrt{\pi}}{2\sqrt{n}} + \frac{2\sqrt{2}}{\sqrt{n}} + \frac{3e\sqrt{2}}{\sqrt{\pi}\sqrt{n}} + \frac{12e^2\sqrt{6}}{b\sqrt{\pi}n} \\
&+ \frac{4e\sqrt{3}}{b\sqrt{n}}\delta(n-1) + \frac{2\sqrt{2}}{b} e^{-2e\sqrt{n}} \tag{3.39}
\end{aligned}$$

3.2 Das Berry-Esséen-Theorem für character ratios

Man kann bereits an dieser Stelle sehen, dass wir eine recht große Berry-Esséen-Konstante erhalten werden: Das Intervall $(a; d]$, auf das wir (3.36) anwenden, hat hier die Länge $\sqrt{\frac{n}{(n-2)}} \left(b + \frac{4e\sqrt{2}}{\sqrt{n-1}}\right)$ (im Vergleich zu b in Abschnitt 1.2) und diese Zahl wird (unter Anderem) noch mit dem Faktor $\frac{2e\sqrt{n}}{b\sqrt{\binom{n}{2}}}$ multipliziert, der aus der Supremums-Abschätzung $|X_n^{(n)}| \leq 2e\sqrt{n}$ herrührt. Im Beweis von Satz 1.2.3 wurde eine so grobe Abschätzung nicht vorgenommen, stattdessen wurde $|X_i|$ nach dem Derandomisierungs-Trick ausintegriert.

Nun wollen wir analog den zweiten Term in (3.34) nach oben abschätzen. Es gilt zunächst:

$$\begin{aligned} & \left| E_n \left[\frac{(X_n^{(n)})^2}{n-1} \int_0^1 \left(f' \left(\sqrt{\frac{n-2}{n}} T_{n-1}^{(n)} + t \frac{X_n^{(n)}}{\sqrt{\binom{n}{2}}} \right) - f' \left(\sqrt{\frac{n-2}{n}} T_{n-1}^{(n)} \right) \right) dt \right] \right| \\ & \leq E_n \left[\frac{(X_n^{(n)})^2}{n-1} \int_0^1 \left| f' \left(\sqrt{\frac{n-2}{n}} T_{n-1}^{(n)} + t \frac{X_n^{(n)}}{\sqrt{\binom{n}{2}}} \right) - f' \left(\sqrt{\frac{n-2}{n}} T_{n-1}^{(n)} \right) \right| dt \right] \end{aligned} \quad (3.40)$$

Nach Lemma 1.2.4 (a), (b) und (c) gilt

$$\begin{aligned} & \left| f' \left(\sqrt{\frac{n-2}{n}} T_{n-1}^{(n)} + t \frac{X_n^{(n)}}{\sqrt{\binom{n}{2}}} \right) - f' \left(\sqrt{\frac{n-2}{n}} T_{n-1}^{(n)} \right) \right| \\ & \leq \frac{|tX_n^{(n)}|}{\sqrt{\binom{n}{2}}} \left(\sqrt{\frac{\pi}{8}} + 2 \left| \sqrt{\frac{n-2}{n}} T_{n-1}^{(n)} \right| + \frac{1}{b} \int_0^1 I_{(z; z+b]} \left(\sqrt{\frac{n-2}{n}} T_{n-1}^{(n)} + st \frac{X_n^{(n)}}{\sqrt{\binom{n}{2}}} \right) ds \right) \end{aligned} \quad (3.41)$$

Aufgrund von (3.41) können wir nun (3.40) nach oben abschätzen durch

$$\begin{aligned} & \frac{1}{2} \sqrt{\frac{\pi}{8}} \frac{1}{(n-1)\sqrt{\binom{n}{2}}} E_n[|X_n^{(n)}|^3] + \frac{\sqrt{\frac{n-2}{n}}}{\sqrt{\binom{n}{2}}(n-1)} E_n[|X_n^{(n)}|^3 |T_{n-1}^{(n)}|] \\ & + \frac{1}{b\sqrt{\binom{n}{2}}(n-1)} E_n \left[|X_n^{(n)}|^3 \int_0^1 t \int_0^1 I_{(z; z+b]} \left(\sqrt{\frac{n-2}{n}} T_{n-1}^{(n)} + st \frac{X_n^{(n)}}{\sqrt{\binom{n}{2}}} \right) ds dt \right] \end{aligned} \quad (3.42)$$

Den ersten Summanden in (3.42) können wir nach Lemma 3.1.10 (i) wie folgt abschätzen:

3 Der zentrale Grenzwertsatz für character ratios

$$\begin{aligned}
\frac{1}{2}\sqrt{\frac{\pi}{8}}\frac{\sqrt{2}}{(n-1)\sqrt{n(n-1)}}E_n[|X_n^{(n)}|^3] &\leq \frac{1}{2}\sqrt{\frac{\pi}{8}}\frac{\sqrt{2}}{(n-1)\sqrt{n(n-1)}}(n-1)\sqrt{2n-3} \\
&= \sqrt{\frac{\pi}{8}}\frac{\sqrt{n-\frac{3}{2}}}{\sqrt{n-1}}\frac{1}{\sqrt{n}} \\
&\leq \sqrt{\frac{\pi}{8}}\frac{1}{\sqrt{n}}
\end{aligned} \tag{3.43}$$

Für den zweiten Summanden in (3.42) gilt nach Lemma 3.1.10 (ii)

$$\begin{aligned}
\frac{\sqrt{\frac{n-2}{n}}}{\sqrt{\frac{n(n-1)}{2}}(n-1)}E_n[|T_n^{(n)}(X_n^{(n)})^3|] &\leq \sqrt{\frac{n-2}{n-1}}\frac{\sqrt{2}}{n(n-1)}(n-1)\sqrt{2n-3} \\
&= \sqrt{\frac{n-2}{n-1}}\frac{2\sqrt{n-\frac{3}{2}}}{n} \leq \frac{2\sqrt{n-2}}{n} \\
&\leq \frac{2}{\sqrt{n}}.
\end{aligned} \tag{3.44}$$

Um auch den dritten Summanden abschätzen zu können, gehen wir analog zu obiger Herangehensweise vor und schreiben ihn zunächst als Summe $S'_1 + S'_2$ mit

$$S'_1 := \frac{\sqrt{2}}{b\sqrt{n(n-1)}(n-1)}A,$$

wobei

$$A := E_n \left[I_{\{|X_n^{(n)}| \leq 2e\sqrt{n}\}} |X_n^{(n)}|^3 \int_0^1 t \int_0^1 I_{(z; z+b]} \left(\sqrt{\frac{n-2}{n}} T_{n-1}^{(n)} + st \frac{X_n^{(n)}}{\sqrt{\binom{n}{2}}} \right) ds dt \right]$$

und

$$S'_2 := \frac{\sqrt{2}}{b\sqrt{n(n-1)}(n-1)}B,$$

wobei

$$B := E_n \left[I_{\{|X_n^{(n)}| > 2e\sqrt{n}\}} |X_n^{(n)}|^3 \int_0^1 t \int_0^1 I_{(z; z+b]} \left(\sqrt{\frac{n-2}{n}} T_{n-1}^{(n)} + st \frac{X_n^{(n)}}{\sqrt{\binom{n}{2}}} \right) ds dt \right].$$

3.2 Das Berry-Esséen-Theorem für character ratios

Unter Verwendung von (3.36) erhalten wir für A :

$$\begin{aligned}
A &\leq 8e^3 n^{\frac{3}{2}} E_n \left[\int_0^1 t \int_0^1 I_{(\sqrt{\frac{n}{n-2}}(z - \frac{2e\sqrt{n}}{\sqrt{n(n-1)}}); \sqrt{\frac{n}{n-2}}(z+b + \frac{2e\sqrt{n}}{\sqrt{n(n-1)}}))} (T_{n-1}^{(n)}) ds dt \right] \\
&= 8e^3 n^{\frac{3}{2}} E_n \left[\int_0^1 t I_{(\sqrt{\frac{n}{n-2}}(z - \frac{2e\sqrt{n}}{\sqrt{n(n-1)}}); \sqrt{\frac{n}{n-2}}(z+b + \frac{2e\sqrt{n}}{\sqrt{n(n-1)}}))} (T_{n-1}^{(n)}) dt \right] \\
&\stackrel{\text{Fubini}}{=} 8e^3 n^{\frac{3}{2}} \int_0^1 t E_n \left[I_{(\sqrt{\frac{n}{n-2}}(z - \frac{2e\sqrt{n}}{\sqrt{n(n-1)}}); \sqrt{\frac{n}{n-2}}(z+b + \frac{2e\sqrt{n}}{\sqrt{n(n-1)}}))} (T_{n-1}^{(n)}) \right] dt \\
&\leq 8e^3 n^{\frac{3}{2}} \int_0^1 t \left(\sqrt{\frac{n}{(n-2)2\pi}} (b + \frac{4e\sqrt{2}}{\sqrt{n-1}}) + 2\delta(n-1) \right) dt \\
&= 4e^3 n^{\frac{3}{2}} \left(\sqrt{\frac{n}{(n-2)2\pi}} (b + \frac{4e\sqrt{2}}{\sqrt{n-1}}) + 2\delta(n-1) \right) \tag{3.45}
\end{aligned}$$

Daraus erhalten wir unmittelbar

$$\begin{aligned}
S'_1 &\leq \frac{4\sqrt{2}e^3 n^{\frac{3}{2}} \sqrt{nb}}{b\sqrt{n(n-1)}(n-1)\sqrt{2\pi}\sqrt{n-2}} + \frac{32e^4 n^{\frac{3}{2}} \sqrt{n}}{b\sqrt{n(n-1)}(n-1)\sqrt{n-2}\sqrt{2\pi}\sqrt{n-1}} \\
&+ \frac{8\sqrt{2}e^3 n^{\frac{3}{2}}}{b\sqrt{n(n-1)}(n-1)} \delta(n-1) \\
&\leq \frac{18e^3}{\sqrt{2\pi}\sqrt{n}} + \frac{72\sqrt{3}e^4}{b\sqrt{2\pi}n} + \frac{12e^3\sqrt{3}}{b\sqrt{n}} \delta(n-1), \tag{3.46}
\end{aligned}$$

wobei wir immer wieder die Abschätzungen $\frac{n}{n-1} \leq \frac{3}{2}$ und $\frac{n}{n-2} \leq 3$ für $n \geq 3$ verwendet haben. Wegen $|X_n^{(n)}| \leq n-1$ erhalten wir unter Verwendung von Lemma 3.2.1 für B die folgende Abschätzung nach oben:

$$\begin{aligned}
B &\leq E_n \left[(n-1)^3 I_{\{|X_n^{(n)}| > 2e\sqrt{n}\}} \int_0^1 t dt \right] \\
&= \frac{(n-1)^3}{2} P_n(|X_n^{(n)}| > 2e\sqrt{n}) \leq (n-1)^3 e^{-2e\sqrt{n}} \tag{3.47}
\end{aligned}$$

Daraus erhalten wir als obere Schranke für S'_2 :

3 Der zentrale Grenzwertsatz für character ratios

$$\begin{aligned}
 S'_2 &\leq \frac{\sqrt{2}(n-1)^3}{b\sqrt{n(n-1)}(n-1)} e^{-2e\sqrt{n}} \\
 &\leq \frac{\sqrt{2}(n-1)}{b} e^{-2e\sqrt{n}} \leq \frac{n\sqrt{2}}{b} e^{-2e\sqrt{n}} \quad (3.48)
 \end{aligned}$$

Aus (3.42), (3.43), (3.44), (3.46) und (3.48) schließen wir, dass wir (3.40) nach oben abschätzen können durch

$$\sqrt{\frac{\pi}{8}} \frac{1}{\sqrt{n}} + \frac{2}{\sqrt{n}} + \frac{18e^3}{\sqrt{2\pi}\sqrt{n}} + \frac{72\sqrt{3}e^4}{b\sqrt{2\pi}n} + \frac{12e^3\sqrt{3}}{b\sqrt{n}} \delta(n-1) + \frac{n\sqrt{2}}{b} e^{-2e\sqrt{n}} \quad (3.49)$$

Nach (3.28), (3.34), (3.39) und (3.49) gilt somit für alle $z \in \mathbb{R}$ und alle $b > 0$

$$\begin{aligned}
 &|E_n[h_{z,b}(T_n^{(n)})] - N(h_{z,b})| \\
 &\leq \frac{\sqrt{\pi}}{2\sqrt{n}} + \frac{2\sqrt{2}}{\sqrt{n}} + \frac{3e\sqrt{2}}{\sqrt{\pi}\sqrt{n}} + \frac{12e^2\sqrt{6}}{b\sqrt{\pi}n} \\
 &\quad + \frac{4e\sqrt{3}}{b\sqrt{n}} \delta(n-1) + \frac{2\sqrt{2}}{b} e^{-2e\sqrt{n}} + \sqrt{\frac{\pi}{8}} \frac{1}{\sqrt{n}} + \frac{2}{\sqrt{n}} \\
 &\quad + \frac{18e^3}{\sqrt{2\pi}\sqrt{n}} + \frac{72\sqrt{3}e^4}{b\sqrt{2\pi}n} + \frac{12e^3\sqrt{3}}{b\sqrt{n}} \delta(n-1) + \frac{n\sqrt{2}}{b} e^{-2e\sqrt{n}} \\
 &= \left(\frac{\sqrt{\pi}}{2} + 2\sqrt{2} + \frac{3e\sqrt{2}}{\sqrt{\pi}} + \sqrt{\frac{\pi}{8}} + 2 + \frac{18e^3}{\sqrt{2\pi}} \right) \frac{1}{\sqrt{n}} \\
 &\quad + \left(\frac{12e^2\sqrt{6}}{\sqrt{\pi}} + \frac{72\sqrt{3}e^4}{\sqrt{2\pi}} \right) \frac{1}{bn} + \left(4e^3 + 12e^3\sqrt{3} \right) \frac{\delta(n-1)}{b\sqrt{n}} \\
 &\quad \quad \quad + \frac{2\sqrt{2}}{b} e^{-2e\sqrt{n}} + \frac{n\sqrt{2}}{b} e^{-2e\sqrt{n}} \\
 &\leq \frac{158}{\sqrt{n}} + \frac{2839}{bn} + 498 \frac{\delta(n-1)}{b\sqrt{n}} + \frac{2\sqrt{2}}{b} e^{-2e\sqrt{n}} + \frac{n\sqrt{2}}{b} e^{-2e\sqrt{n}} \quad (3.50)
 \end{aligned}$$

Nun behaupten wir, dass für alle $n \geq 3$ die Ungleichungen $n^{\frac{3}{2}} e^{-2e\sqrt{n}} \leq \frac{8}{10000} \frac{1}{\sqrt{n}}$ sowie $\sqrt{n} e^{-2e\sqrt{n}} \leq \frac{3}{10000} \frac{1}{\sqrt{n}}$ gelten. Es ist nämlich $3^2 e^{-2e\sqrt{3}} = 0,000732... \leq \frac{8}{10000}$ und die Funktion $g : [1; +\infty) \rightarrow \mathbb{R}$ mit $g(x) := x^2 e^{-2e\sqrt{x}}$ ist wegen $g'(x) = x e^{-2e\sqrt{x}} (2 - ex^{\frac{3}{2}}) < 0$ für $x \geq 1$ streng monoton fallend. Ebenso ist $3 \cdot e^{-2e\sqrt{3}} = 0,000244... \leq \frac{3}{10000}$ und die Funktion $h : [1; +\infty) \rightarrow \mathbb{R}$ mit $h(x) := x e^{-2e\sqrt{x}}$ ist

3.2 Das Berry-Esséen-Theorem für character ratios

wegen $h'(x) = e^{-2e\sqrt{x}}(1 - e\sqrt{x}) < 0$ für alle $x \geq 1$ ebenfalls streng monoton fallend. Wir erhalten aus diesen Abschätzungen in Verbindung mit (3.50)

$$\begin{aligned} & |E_n[h_{z,b}(T_n^{(n)})] - N(h_{z,b})| \\ \leq & \left(158 + \frac{2\sqrt{2}}{b\sqrt{n}} \frac{3}{10000} + \frac{\sqrt{2}}{b\sqrt{n}} \frac{8}{10000} \right) \frac{1}{\sqrt{n}} + \frac{2839}{bn} + 498 \frac{\delta(n-1)}{b\sqrt{n}} \end{aligned} \quad (3.51)$$

und, da diese Abschätzung unabhängig von $z \in \mathbb{R}$ ist, auch

$$\begin{aligned} \delta(b, n) &= \sup_{z \in \mathbb{R}} |E_n[h_{z,b}(T_n^{(n)})] - N(h_{z,b})| \\ &\leq \left(158 + \frac{2\sqrt{2}}{b\sqrt{n}} \frac{3}{10000} + \frac{\sqrt{2}}{b\sqrt{n}} \frac{8}{10000} \right) \frac{1}{\sqrt{n}} + \frac{2839}{bn} + 498 \frac{\delta(n-1)}{b\sqrt{n}}. \end{aligned} \quad (3.52)$$

Nun benötigen wir noch folgende, mit der von Lemma 1.2.6 vergleichbare Aussage, welche auch genauso bewiesen werden kann:

Für alle $n \in \mathbb{N}$ und alle $b > 0$ gilt $\delta(n) \leq \delta(b, n) + \frac{b}{2\sqrt{2\pi}}$.

Aus dieser Aussage mit (3.52) folgt dann die Ungleichung

$$\begin{aligned} \delta(n) &\leq \left(158 + \frac{2\sqrt{2}}{b\sqrt{n}} \frac{3}{10000} + \frac{\sqrt{2}}{b\sqrt{n}} \frac{8}{10000} \right) \frac{1}{\sqrt{n}} \\ &+ \frac{2839}{bn} + 498 \frac{\delta(n-1)}{b\sqrt{n}} + \frac{b}{2\sqrt{2\pi}}. \end{aligned} \quad (3.53)$$

Wir wollen nun $b > 0$ wieder möglichst sinnvoll wählen, um durch ein Induktionsargument eine möglichst gute Konstante $C > 0$ zu bekommen. Dabei wollen wir wieder erreichen, dass die „Hälfte“ $\frac{1}{2} \frac{C}{\sqrt{n}}$ aus der Induktionsvoraussetzung kommt, genauer: Es ist klar, dass b von der Form $b = \frac{L}{\sqrt{n}}$ sein sollte, wobei $L > 0$ nicht von n abhängt. Wir haben dann mit $D := 498$ die Rechnung

$$\begin{aligned} D \frac{\delta(n-1)}{b\sqrt{n}} &= \frac{D}{L} \delta(n-1) \leq \frac{D}{L} \frac{C}{\sqrt{n-1}} \\ &\leq \frac{D}{L} \sqrt{\frac{3}{2}} \frac{C}{\sqrt{n}} \stackrel{!}{=} \frac{1}{2} \frac{C}{\sqrt{n}} \\ &\Leftrightarrow L = \sqrt{6} D \end{aligned}$$

3 Der zentrale Grenzwertsatz für character ratios

Wir errechnen nun mit $b := \frac{498\sqrt{6}}{\sqrt{n}}$ die angestrebte Konstante C : Es ist

$$158 + \frac{2\sqrt{2}}{498\sqrt{6}} \frac{3}{10000} + \frac{\sqrt{2}}{498\sqrt{6}} \frac{8}{10000} + \frac{2839}{498\sqrt{6}} + \frac{498\sqrt{6}}{2\sqrt{2\pi}} = 403,651\dots \leq 404, \quad (3.54)$$

so dass wir $C := 808$ erhalten. Diese Konstante ist, wie bereits angekündigt, recht groß. Wir beweisen nun jedoch noch durch vollständige Induktion über $n \geq 3$, dass die Behauptung des Satzes mit dieser Konstanten gilt:

Der **Induktionsanfang** $n = 3$ ist natürlich trivial.

Induktionsschritt: $n - 1 \rightarrow n$: Wir wählen $b := \frac{498\sqrt{6}}{\sqrt{n}}$. Dann folgt aus (3.53), der Rechnung (3.54) und der Induktionsvoraussetzung:

$$\begin{aligned} \delta(n) &\leq \frac{404}{\sqrt{n}} + \frac{\delta(n-1)}{\sqrt{6}} \leq \frac{404}{\sqrt{n}} + \frac{808}{\sqrt{6}\sqrt{n-1}} \\ &\leq \frac{404}{\sqrt{n}} + \sqrt{\frac{3}{2}} \frac{808}{\sqrt{6}\sqrt{n}} = \frac{404}{\sqrt{n}} + \frac{404}{\sqrt{n}} \\ &= \frac{808}{\sqrt{n}} \end{aligned}$$

□

Literaturverzeichnis

- [Bar88] A. D. Barbour. Stein's method and Poisson process convergence. *J. Appl. Probab.*, (Special Vol. 25A):175–184, 1988. A celebration of applied probability.
- [BC05] A. D. Barbour and Louis H. Y. Chen, editors. *An introduction to Stein's method*, volume 4 of *Lecture Notes Series. Institute for Mathematical Sciences. National University of Singapore*. Singapore University Press, Singapore, 2005. Lectures from the Meeting on Stein's Method and Applications: a Program in Honor of Charles Stein held at the National University of Singapore, Singapore, July 28–August 31, 2003.
- [Bil95] Patrick Billingsley. *Probability and measure*. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons Inc., New York, third edition, 1995. A Wiley-Interscience Publication.
- [Bol84] E. Bolthausen. An estimate of the remainder in a combinatorial central limit theorem. *Z. Wahrsch. Verw. Gebiete*, 66(3):379–386, 1984.
- [DG89] P. Diaconis and C. Greene. Applications of Murphy's elements. *Stanford University Technical Report*, 335, 1989.
- [Dia88] Persi Diaconis. *Group representations in probability and statistics*. Institute of Mathematical Statistics Lecture Notes—Monograph Series, 11. Institute of Mathematical Statistics, Hayward, CA, 1988.
- [DS81] Persi Diaconis and Mehrdad Shahshahani. Generating a random permutation with random transpositions. *Z. Wahrsch. Verw. Gebiete*, 57(2):159–179, 1981.
- [Els05] Jürgen Elstrodt. *Maß- und Integrationstheorie*. Springer-Lehrbuch. [Springer Textbook]. Springer-Verlag, Berlin, fourth edition, 2005. Grundwissen Mathematik. [Basic Knowledge in Mathematics].
- [Ful05] Jason Fulman. Stein's method and Plancherel measure of the symmetric group. *Trans. Amer. Math. Soc.*, 357(2):555–570 (electronic), 2005.

- [Ful06a] Jason Fulman. An inductive proof of the Berry-Esseen theorem for character ratios. *Ann. Comb.*, 10(3):319–332, 2006.
- [Ful06b] Jason Fulman. Martingales and character ratios. *Trans. Amer. Math. Soc.*, 358(10):4533–4552 (electronic), 2006.
- [Göt91] F. Götze. On the rate of convergence in the multivariate CLT. *Ann. Probab.*, 19(2):724–739, 1991.
- [GS77] Peter Gänsler and Winfried Stute. *Wahrscheinlichkeitstheorie*. Springer-Verlag, 1977.
- [HS75] Edwin Hewitt and Karl Stromberg. *Real and abstract analysis*. Springer-Verlag, New York, 1975. A modern treatment of the theory of functions of a real variable, Third printing, Graduate Texts in Mathematics, No. 25.
- [Jam78] G. D. James. *The representation theory of the symmetric groups*, volume 682 of *Lecture Notes in Mathematics*. Springer, Berlin, 1978.
- [Ker93a] S. V. Kerov. Transition probabilities of continual Young diagrams and the Markov moment problem. *Funktsional. Anal. i Prilozhen.*, 27(2):32–49, 96, 1993.
- [Ker93b] Serguei Kerov. Gaussian limit for the Plancherel measure of the symmetric group. *C. R. Acad. Sci. Paris Sér. I Math.*, 316(4):303–308, 1993.
- [Ker96] S. Kerov. The boundary of Young lattice and random Young tableaux. In *Formal power series and algebraic combinatorics (New Brunswick, NJ, 1994)*, volume 24 of *DIMACS Ser. Discrete Math. Theoret. Comput. Sci.*, pages 133–158. Amer. Math. Soc., Providence, RI, 1996.
- [Kle06] Achim Klenke. *Wahrscheinlichkeitstheorie*. Springer-Verlag, 2006.
- [Las04] Michel Lassalle. Jack polynomials and some identities for partitions. *Trans. Amer. Math. Soc.*, 356(9):3455–3476 (electronic), 2004.
- [Man94] B. Mann. Bolthausen’s proof of Berry-Esséen. *unpublished manuscript*, 1994.
- [Mur81] G. E. Murphy. A new construction of Young’s seminormal representation of the symmetric groups. *J. Algebra*, 69(2):287–297, 1981.
- [Sag01] Bruce E. Sagan. *The symmetric group*, volume 203 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, second edition, 2001. Representations, combinatorial algorithms, and symmetric functions.

- [Ser77] Jean-Pierre Serre. *Linear representations of finite groups*. Springer-Verlag, New York, 1977. Translated from the second French edition by Leonard L. Scott, Graduate Texts in Mathematics, Vol. 42.
- [SS06] Qi-Man Shao and Zhong-Gen Su. The Berry-Esseen bound for character ratios. *Proc. Amer. Math. Soc.*, 134(7):2153–2159 (electronic), 2006.
- [Ste86] Charles Stein. *Approximate computation of expectations*. Institute of Mathematical Statistics Lecture Notes—Monograph Series, 7. Institute of Mathematical Statistics, Hayward, CA, 1986.
- [Ste97] J. Michael Steele. *Probability theory and combinatorial optimization*, volume 69 of *CBMS-NSF Regional Conference Series in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1997.

Literaturverzeichnis

Selbständigkeitserklärung:

Hiermit erkläre ich, dass ich diese Diplomarbeit selbständig und nur mit den angegebenen Quellen und Hilfsmitteln verfasst habe.

Bochum, den 13. Oktober 2008