

Konvergenzraten für Pólya-Urnen

Masterarbeit im Studiengang Mathematik

vorgelegt von
ANDREA KUNTSCHIK

eingereicht bei
PROF. DR. RALPH NEININGER

Fachbereich Informatik und Mathematik
Institut für Mathematik
Goethe-Universität
Frankfurt am Main

Frankfurt am Main, den 1. September 2014

Eidesstattliche Erklärung

Hiermit erkläre ich, dass ich die vorliegende Arbeit selbstständig abgefasst und keine anderen Hilfsmittel als die angegebenen benutzt habe. Ich erkläre ferner, dass diejenigen Stellen der Arbeit, die anderen Werken wörtlich oder dem Sinne nach entnommen sind, in jedem Falle unter Angabe der Quellen kenntlich gemacht sind.

Andrea Kuntschik
Frankfurt am Main, den 01.09.2014

Inhaltsverzeichnis

1	Einführung	3
2	Von Urnen und Bäumen zu Rekursionen und Fixpunkten	5
3	Technische Vorbemerkungen	12
4	Existenz und Eindeutigkeit eines Fixpunkts	15
5	Konvergenzrate für die Wasserstein-Metrik ℓ_2^V	18
6	Konvergenzrate für die Wasserstein-Metrik ℓ_p^V mit $p \geq 2$	28
7	Konvergenzrate für die Kolmogorov-Metrik ϱ^V	36
8	Ausblick	37

1 Einführung

Wir betrachten eine Urne, in der zu Beginn eine schwarze oder eine weiße Kugel liegt, und definieren einen Urnenprozess mithilfe der folgenden Dynamik: In jedem Schritt wird zufällig eine Kugel aus der Urne gezogen. Anschließend wird diese Kugel zusammen mit einer weiteren Kugel in die Urne zurückgelegt. Welche Farbe die neue Kugel hat, wird zuvor per Münzwurf bestimmt, dessen Erfolgswahrscheinlichkeit von der Farbe der gezogenen Kugel abhängt. Wir betrachten also eine balancierte Pólya-Urne mit zwei Farben und einer zufälligen Rücklegematrix.

Dieses Urnenmodell findet Anwendung in der „play-the-winner-rule“, welche ein randomisiertes adaptives Modell für klinische Studien beschreibt. In diesem Modell soll von zwei zu vergleichenden Behandlungsformen diejenige häufiger an Probanden vergeben werden, die sich im Verlauf der Studie als erfolgreicher herausstellt. Die beiden Farben stellen die beiden Behandlungsformen dar und der Anteil der Kugeln einer Farbe spiegelt den Erfolg der zugehörigen Behandlungsform im Verlauf der Studie wider.

In diesem Zusammenhang trat das oben beschriebene Urnenmodell in Erscheinung und wurde in [15, 16, 14, 13, 12, 2, 3, 7] untersucht und verallgemeinert.

Nun interessieren wir uns für die Anzahl der Kugeln einer bestimmten Farbe nach n Schritten in diesem Prozess. Zentriert und skaliert man diese Größe, so konvergiert ihre Verteilung bekanntermaßen gegen eine Grenzverteilung. In dieser Arbeit wollen wir erstmals eine obere Schranke für die zugehörige Konvergenzrate bestimmen.

Wir verfolgen zunächst den Ansatz der Kontraktionsmethode und erstellen in Kapitel 2 eine Rekursion für die Verteilung der Anzahl der Kugeln einer Farbe, indem wir das Urnenmodell in einen durch Knappe und Neininger [9, Abschnitt 2] vorgestellten Zusammenhang mit einer passenden Baumstruktur bringen. Aus einem System von Rekursionsgleichungen für diese Verteilung können wir nach Zentrieren und Skalieren ein System von Fixpunktgleichungen herleiten. Nach ein paar technischen Vorbemerkungen in Kapitel 3 werden wir diese Fixpunktgleichungen in Kapitel 4 als eine Abbildung zwischen Maßräumen auffassen und dann ein Kriterium nennen, mithilfe dessen wir überprüfen können, ob eine Kontraktion vorliegt. Nachdem wir so die Existenz eines eindeutigen Fixpunkts in einem passenden Maßraum nachgewiesen haben werden, beginnen wir in Kapitel 5 damit, eine obere Schranke für die Konvergenzrate der Verteilung der zentrierten und skalierten Größe gegen diesen Fixpunkt zu bestimmen. Der Beweis erfolgt zunächst in der Wasserstein-Metrik ℓ_2^\vee auf Produkträumen mit vollständiger Induktion und wird anschließend in Kapitel 6 auf Wasserstein-Metriken ℓ_p^\vee mit $p \geq 2$ verbessert, ebenfalls durch Induktion. In Kapitel 7 werden wir als letzten Schritt mithilfe der oberen Schranke in den Wasserstein-Metriken auf eine obere Schranke für die Konvergenzrate in der gängigen und insbesondere in der Statistik relevanten Kolmogorov-Metrik ρ^\vee auf Produkträumen schließen.

Ein Zugang zur Analyse von Pólya-Urnen von Athreya und Karlin [1], der auf einer Einbettung

des Urnenprozesses in zeitstetige Verzweigungsprozesse beruht, wurde von Svante Janson in [7, 8] genutzt, um hinsichtlich von Grenzwertsätzen sehr allgemeine Nachlegeschemata universell zu behandeln. In [7, Remark 4.7] schreibt Janson über Konvergenzraten zu diesem Zugang allerdings: „Our methods give no information on the rate of convergence.“

Wir liefern hier eine obere Schranke für die Konvergenzrate in einem speziellen, jedoch nichttrivialen Fall: Mit B_n^b bezeichnen wir die Anzahl schwarzer Kugeln nach n Schritten bei Start mit einer schwarzen Kugel (ein analoges Resultat ergibt sich aber auch bei Start mit einer weißen Kugel) und $X_n := \frac{B_n^b - \mathbb{E}[B_n^b]}{n^\lambda}$ bezeichne die entsprechende normalisierte Größe. Es sei $\alpha \in (0, 1)$ die Erfolgswahrscheinlichkeit der Münze, die zur Bestimmung der Farbe der neuen Kugel geworfen wird, falls die gezogene Kugel schwarz ist. Gleichermäßen sei $\beta \in (0, 1)$ diese Erfolgswahrscheinlichkeit, falls die gezogene Kugel weiß ist. Ein „Erfolg“ liegt dann vor, wenn die Kugel, die mit der gezogenen Kugel in die Urne zurückgelegt wird, dieselbe Farbe hat wie diese. Weiter sei $\lambda := \alpha + \beta - 1$. Gilt $\lambda \leq \frac{1}{2}$, so führt dies zu normalverteilten Grenzwerten, welche in dieser Arbeit nicht betrachtet werden, andernfalls zu nicht-normalverteilten Grenzwerten, vgl. Knappe und Neininger [9, Abschnitt 6.2]. Schließlich bezeichne $\mathcal{L}(X_{\alpha,\beta})$ den zugehörigen Grenzwert der Folge $(\mathcal{L}(X_n))_{n \in \mathbb{N}}$ wie in Kapitel 4 und es sei $\lambda > \frac{1}{2}$. Das Hauptresultat der vorliegenden Arbeit lautet dann:

Zu jedem $\varepsilon > 0$ gibt es eine Konstante $0 < K < \infty$, sodass für alle $n \in \mathbb{N}$ gilt:

$$\varrho(X_n, X_{\alpha,\beta}) := \sup_{x \in \mathbb{R}} |\mathbb{P}(X_n \leq x) - \mathbb{P}(X_{\alpha,\beta} \leq x)| \leq \frac{K}{n^{\lambda - \frac{1}{2} - \varepsilon}}.$$

Notationen

Die verwendeten Bezeichnungen für die auftretenden Größen sind meist von Knappe und Neininger [9] übernommen worden, um eine bessere Lesbarkeit im Zusammenspiel mit dieser Arbeit zu gewährleisten. Für eine Zufallsvariable X bezeichnet $\mathcal{L}(X)$ ihre Verteilung. Falls $\mathcal{L}(X) = \mathcal{L}(Y)$ für zwei Zufallsvariablen X und Y gilt, so schreiben wir $X \stackrel{d}{=} Y$. Außerdem werden die Bachmann-Landau-Symbole für asymptotische Aussagen benutzt.

2 Von Urnen und Bäumen zu Rekursionen und Fixpunkten

Wir werden den Urnenprozess nun, wie in Knappe und Neininger [9, Abschnitt 2], mit einem kombinatorischen zufälligen Baum, den wir als *zugehörigen Baum* bezeichnen werden, identifizieren und diesen an der Wurzel zerlegen. Das Wachstum der Teilbäume der Wurzel entspricht dann in Verteilung dem Wachstum des ganzen Baums. Diese Selbstähnlichkeit ausnutzend können wir ein System von Rekursionsgleichungen für die Verteilung der Anzahl der Kugeln einer Farbe aufstellen und nach Zentrierung und Skalierung zu einem System von Fixpunktgleichungen für diese Verteilung gelangen.

In einer Urne liege zu Beginn eine schwarze oder eine weiße Kugel. In jedem Schritt des Urnenprozesses wird nach bestimmten Regeln, welche mit der zufälligen Rücklegematrix

$$(2.1) \quad R = \begin{bmatrix} F_\alpha & 1 - F_\alpha \\ 1 - F_\beta & F_\beta \end{bmatrix}$$

beschrieben werden, eine schwarze oder eine weiße Kugel hinzugefügt. Dabei seien F_α und F_β Bernoulli-Zufallsvariablen mit Erfolgsparametern $0 < \alpha, \beta < 1$. Da die Zeilensummen von R fast sicher gleich 1 sind, ist die Urne balanciert, was später für das Aufstellen der Fixpunktgleichung relevant sein wird. Weiter sei $\lambda := \alpha + \beta - 1$. Ab Kapitel 4 werden wir ausschließlich den Fall $\lambda > \frac{1}{2}$, der zu nicht-normalverteilten Grenzwerten führt, betrachten; vgl. Knappe und Neininger [9, Abschnitt 6.2].

Der Urnenprozess unterliegt der folgenden Dynamik: Wurde eine schwarze Kugel aus der Urne gezogen, so wird mit Wahrscheinlichkeit α eine neue schwarze Kugel mit der gezogenen Kugel in die Urne zurückgelegt und mit Wahrscheinlichkeit $1 - \alpha$ eine weiße. Falls eine weiße Kugel gezogen wurde, so wird diese mit Wahrscheinlichkeit $1 - \beta$ mit einer schwarzen Kugel und mit Wahrscheinlichkeit β mit einer weiteren weißen Kugel in die Urne zurückgelegt.

Es bezeichne B_n^b die Anzahl schwarzer Kugeln nach n -maligem Ziehen aus der Urne bei Start mit einer schwarzen Kugel und gleichermaßen B_n^w die Anzahl schwarzer Kugeln bei Start mit einer weißen Kugel. Offenbar gilt $B_0^b = 1$ und $B_0^w = 0$.

Wir werden nun den Urnenprozess auf einen Baum übertragen, der mit jedem Ziehen aus der Urne wächst – der sich also in diskreten Zeitschritten entwickelt – und bei dem ein schwarzer bzw. ein weißer Knoten einer schwarzen bzw. einer weißen Kugel in der Urne entspricht.

In der Urne liege zu Beginn eine schwarze Kugel. Die Urne zu dieser Zeit soll durch einen Baum zur Zeit null, der nur aus einer schwarzen Wurzel besteht, dargestellt werden. Nun findet der erste Zug statt: Die schwarze Kugel wird gezogen und eine Münze mit Erfolgswahrscheinlichkeit α entscheidet, welche Farbe die Kugel hat, die mit der gezogenen schwarzen Kugel in die Urne zurückgelegt wird. Im zugehörigen Baum vollzieht sich diese Dynamik folgendermaßen: Die schwarze Wurzel erhält zwei Kinder und verliert ihre Farbe. Mit Wahrscheinlichkeit α sind beide Kinder schwarz und mit Wahrscheinlichkeit $1 - \alpha$ erhält sie ein weißes und ein schwarzes Kind.

Eines der beiden Kinder entspricht der gezogenen Kugel und das andere der neuen Kugel, die mit der gezogenen Kugel in die Urne zurückgelegt wurde. O.B.d.A. soll das linke Kind der gezogenen Kugel entsprechen, das rechte Kind der neuen Kugel. Nach dem ersten Ziehen besteht der zugehörige Baum aus drei Knoten, von denen die beiden Blätter gerade die Kugeln in der Urne symbolisieren.

Iteriert man diesen Vorgang, so erhält man nach dem n -ten Ziehen einen Baum mit $2n+1$ Knoten, von denen n farblos sind, und, dessen $n+1$ Blätter den Kugeln in der Urne nach dem n -ten Ziehen entsprechen. Jedes Blatt hat dabei die Farbe der Kugel, die es im Baum repräsentiert. Zieht man im nächsten Schritt wieder aus der Urne, so ist dies gleichbedeutend damit, ein Blatt des Baums zufällig auszuwählen. Dieses Blatt verliert daraufhin seine Farbe und erhält zwei Kinder. War das zufällig ausgewählte Blatt schwarz, so sind, wie oben bereits beschrieben, mit Wahrscheinlichkeit α beide Kinder schwarz und andernfalls ist, mit Wahrscheinlichkeit $1-\alpha$, ein Kind schwarz und eines weiß. Falls das zufällig ausgewählte Blatt weiß war, so bekommt es mit Wahrscheinlichkeit β zwei weiße Kinder und mit Wahrscheinlichkeit $1-\beta$ ein weißes und ein schwarzes Kind. Alle anderen Blätter des Baums bleiben wie die nicht gezogenen Kugeln in der Urne unverändert. Die Anzahl der Kugeln in der Urne wächst wie die Anzahl der Blätter des Baums mit jedem Ziehen bzw. mit jedem Zeitschritt um eins. Weiter gilt für die Knoten des Baums offenbar, dass sie entweder zwei Kinder haben oder ein Blatt sind.

Abbildung 2.1 illustriert beispielhaft einen möglichen Verlauf des Urnenprozesses zusammen mit dem zugehörigen Baum.

Wir könnten den Urnenprozess also simulieren, indem wir den zugehörigen Baum in diskreten Zeitschritten wachsen lassen, wobei in jedem Schritt ein Blatt des Baums gleichverteilt unter allen Blättern ausgewählt und weiter die oben beschriebene Dynamik durchgeführt wird. Nach n Schritten hat der zugehörige Baum $n+1$ Blätter und, falls die Wurzel schwarz war, ist die Anzahl schwarzer Blätter verteilt wie B_n^b , andernfalls wie B_n^w .

Weiter bezeichnen wir einen zugehörigen Baum, wenn er aus einer schwarzen Wurzel hervorgeht, als zugehörigen Baum vom Typ b , andernfalls als zugehörigen Baum vom Typ w . Die beiden Teilbäume des zugehörigen Baums, die von den Kindern der Wurzel ausgehen, haben eine zufällige Anzahl von Blättern, die davon abhängt, wie oft ein Blatt aus dem jeweiligen Teilbaum zufällig ausgewählt wurde. Wir wollen nun einen zugehörigen Baum nach n Schritten betrachten und darauf bedingen, dass $i+1$ der Blätter im linken Teilbaum und $j+1$ im rechten Teilbaum mit $i, j \in \mathbb{N}_0$ sind. Die Zahlen i und j beschreiben dann, wie oft ein Blatt des jeweiligen Teilbaums zufällig ausgewählt wurde, und es gilt $i+j = n-1$, da die Teilbäume beim ersten Ziehen erzeugt werden und erst die darauffolgenden $n-1$ Schritte innerhalb der Teilbäume stattfinden.

Bedingt darauf, dass der linke Teilbaum $i+1$ und der rechte Teilbaum $j+1$ Blätter hat, sind die beiden Teilbäume unabhängig und verteilt wie ein zugehöriger Baum nach i bzw. j Schritten vom Typ der Farbe ihrer Wurzel.

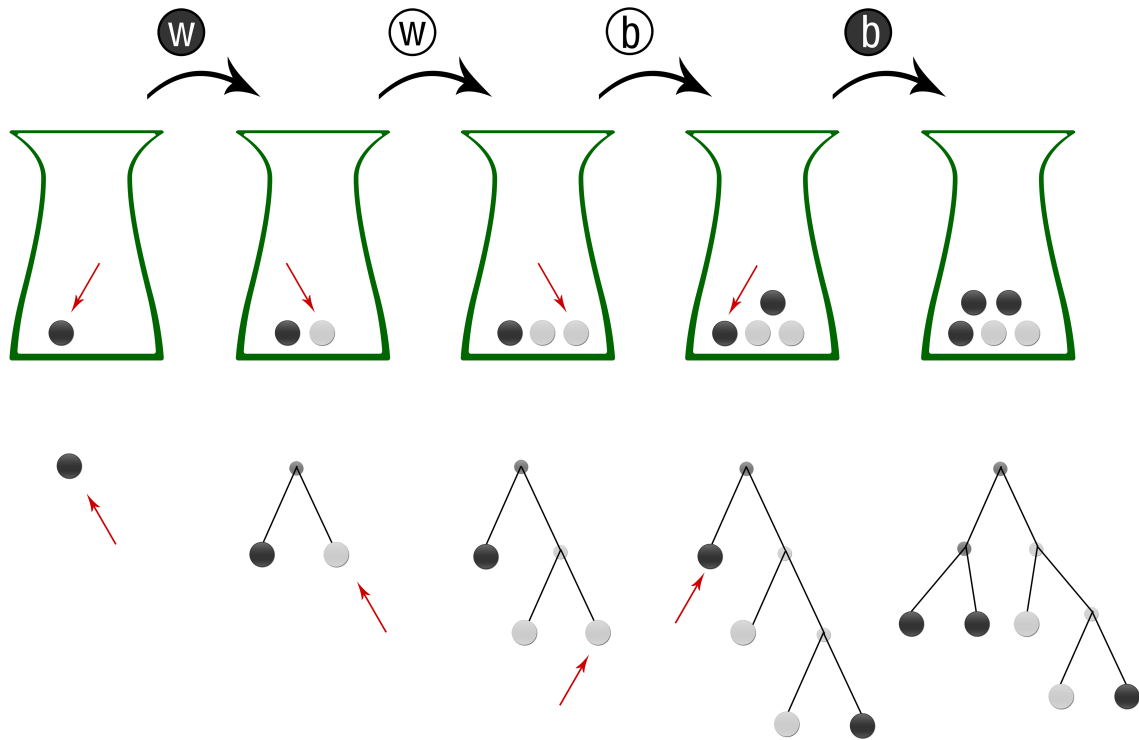


Abbildung 2.1: Illustration eines möglichen Verlaufs des oben beschriebenen Urnenprozesses bei Start mit einer schwarzen Kugel. Unter den Urnen befindet sich der jeweilige zugehörige Baum. Die roten Pfeile zeigen auf die Kugeln bzw. Blätter, die zufällig gezogen wurden. Die Pfeile zwischen den Urnen symbolisieren den Münzwurf in Abhängigkeit von der Farbe der gezogenen Kugel. Die Blätter der zugehörigen Bäume entsprechen den Kugeln in der Urne; andere Knoten stellen keine Kugeln mehr dar, ihre Farbe wird jedoch für die rekursive Zerlegung benötigt und deswegen verkleinert und schattiert dargestellt.

Diese Selbstähnlichkeit zwischen dem ganzen Baum und den Teilbäumen führt nun zu einem System von Rekursionsgleichungen für die Verteilungen von B_n^b bzw. B_n^w : B_n^b beschreibt die Anzahl schwarzer Kugeln nach n -maligem Ziehen aus der Urne, wenn mit einer schwarzen Kugel begonnen wurde, und entspricht mittels obiger Überlegungen der Anzahl schwarzer Blätter eines zugehörigen Baums vom Typ b nach n Schritten. Nun ist B_n^b gerade die Summe der Anzahlen der schwarzen Blätter der Teilbäume der Wurzel und diese sind bedingt auf den Vektor (I_n, J_n) unabhängig und wie zugehörige Bäume vom Typ der Farbe ihrer Wurzel verteilt. Der Vektor (I_n, J_n) mit $J_n := n - 1 - I_n$ beschreibt dabei, wie oft ein Blatt aus dem linken bzw. dem rechten Teilbaum zufällig ausgewählt wird. Da per Münzwurf entschieden wird, ob die Wurzel des zugehörigen Baums zwei schwarze oder ein schwarzes und ein weißes Kind bekommt, erhalten

wir für $n \geq 1$

$$(2.2) \quad B_n^b \stackrel{d}{=} B_{I_n}^{b,(1)} + F_\alpha B_{J_n}^{b,(2)} + (1 - F_\alpha) B_{J_n}^w.$$

Ebenso erhalten wir

$$(2.3) \quad B_n^w \stackrel{d}{=} B_{I_n}^{w,(1)} + F_\beta B_{J_n}^{w,(2)} + (1 - F_\beta) B_{J_n}^b, \quad n \geq 1.$$

Dabei gilt, dass

$$(2.4) \quad \left(B_k^{b,(1)} \right)_{0 \leq k \leq n-1}, \left(B_k^{b,(2)} \right)_{0 \leq k \leq n-1}, (B_k^w)_{0 \leq k \leq n-1}, F_\alpha \text{ und } I_n \text{ sowie}$$

$$(2.5) \quad \left(B_k^{w,(1)} \right)_{0 \leq k \leq n-1}, \left(B_k^{w,(2)} \right)_{0 \leq k \leq n-1}, \left(B_k^b \right)_{0 \leq k \leq n-1}, F_\beta \text{ und } I_n \text{ unabhängig sind und}$$

$$(2.6) \quad B_k^{b,(1)} \stackrel{d}{=} B_k^b \stackrel{d}{=} B_k^{b,(2)} \text{ sowie } B_k^{w,(1)} \stackrel{d}{=} B_k^w \stackrel{d}{=} B_k^{w,(2)} \text{ für } k = 0, \dots, n-1 \text{ gilt.}$$

Die Bedingungen (2.4) und (2.5) formalisieren, dass die Teilbäume bedingt darauf, wie oft aus einem Teilbaum Blätter zufällig ausgewählt wurden, unabhängig sind. Die Bedingung (2.6) hingegen formalisiert die Selbstähnlichkeit zwischen dem gesamten zugehörigen Baum und den von den beiden Kindern der Wurzel ausgehenden Teilbäumen, in die dieser zerlegt wird.

Die obige Rekursion definiert zusammen mit der Startverteilung $(B_0^b, B_0^w) = (1, 0)$ eine Folge $(\mathcal{L}(B_n^b), \mathcal{L}(B_n^w))_{n \geq 0}$ von Verteilungen.

Nun wollen wir von dem obigen System von Rekursionsgleichungen (2.2) und (2.3) zu einem System von Rekursionsgleichungen für zentrierte und skalierte Versionen von B_n^b und B_n^w übergehen, um daraus auf eine Fixpunktgleichung schließen zu können.

Dazu brauchen wir zunächst eine Entwicklung für die Erwartungswerte von B_n^b und B_n^w , die wir mit $\mu_b(n) := \mathbb{E}[B_n^b]$ und $\mu_w(n) := \mathbb{E}[B_n^w]$ bezeichnen. Für diese gilt nach Knappe und Neininger [9, Lemma 6.3]:

$$\begin{aligned} \mu_b(n) &= \frac{1 - \beta}{2 - \alpha - \beta} n + \frac{1 - \alpha}{2 - \alpha - \beta} \frac{\Gamma(n + \alpha + \beta)}{\Gamma(\alpha + \beta)\Gamma(n + 1)} + \frac{1 - \beta}{2 - \alpha - \beta}, \\ \mu_w(n) &= \frac{1 - \beta}{2 - \alpha - \beta} n - \frac{1 - \beta}{2 - \alpha - \beta} \frac{\Gamma(n + \alpha + \beta)}{\Gamma(\alpha + \beta)\Gamma(n + 1)} + \frac{1 - \beta}{2 - \alpha - \beta}. \end{aligned}$$

Wir werden jedoch die asymptotischen Entwicklungen der Erwartungswerte zur Zentrierung nutzen.

2.1 Lemma

Für die Erwartungswerte $\mu_b(n)$ und $\mu_w(n)$ gilt für $n \rightarrow \infty$:

$$\begin{aligned} \mu_b(n) &= c'_b n + d'_b n^\lambda + \mathcal{O}(1), \\ \mu_w(n) &= c'_w n + d'_w n^\lambda + \mathcal{O}(1) \end{aligned}$$

mit den Konstanten

$$c'_b := \frac{1-\beta}{1-\lambda} =: c'_w, \quad d'_b := \frac{1-\alpha}{(1-\lambda)\Gamma(\lambda+1)} \quad \text{und} \quad d'_w := -\frac{1-\beta}{(1-\lambda)\Gamma(\lambda+1)}.$$

Beweis. Seien $n \in \mathbb{N}$ und $\varphi, \xi > 0$. Mit der asymptotischen Entwicklung der Gamma-Funktion folgt:

$$\begin{aligned} & \frac{\Gamma(n+\varphi+\xi)}{\Gamma(n+\varphi)} \\ = & \frac{\left(\frac{n+\varphi+\xi}{e}\right)^{n+\varphi+\xi} \left[(n+\varphi+\xi)^{-\frac{1}{2}} + \frac{1}{12} (n+\varphi+\xi)^{-\frac{3}{2}} + \mathcal{O}\left(n^{-\frac{5}{2}}\right) \right]}{\left(\frac{n+\varphi}{e}\right)^{n+\varphi} \left[(n+\varphi)^{-\frac{1}{2}} + \frac{1}{12} (n+\varphi)^{-\frac{3}{2}} + \mathcal{O}\left(n^{-\frac{5}{2}}\right) \right]} \\ = & e^{-\xi} (n+\varphi+\xi)^\xi \left(1 + \frac{\xi}{n+\varphi}\right)^{n+\varphi} (1+\delta) \quad \text{mit } \delta \rightarrow 0 \text{ f\"ur } n \rightarrow \infty \\ = & (n+\varphi+\xi)^\xi (1+\delta') \quad \text{mit } \delta' \rightarrow 0 \text{ f\"ur } n \rightarrow \infty. \end{aligned}$$

Wir betrachten nun die Taylorentwicklung der Funktion $f : x \mapsto x^\xi$ um n , dann gilt:

$$\begin{aligned} f(n+\varphi+\xi) &= f(n) + f'(n) \cdot (\varphi+\xi) + \frac{f''(n)}{2} \cdot (\varphi+\xi)^2 + \mathcal{O}\left(n^{\xi-3}\right) \\ &= n^\xi + \xi \cdot n^{\xi-1} (\varphi+\xi) + \frac{\xi(\xi-1)n^{\xi-2}}{2} (\varphi+\xi)^2 + \mathcal{O}\left(n^{\xi-3}\right) \\ &= n^\xi + \xi(\varphi+\xi) \cdot n^{\xi-1} + \mathcal{O}\left(n^{\xi-2}\right). \end{aligned}$$

Setzt man nun $\varphi = 1$ und $\xi = \lambda$, so ergibt sich die Behauptung. \square

Da die Varianz von der Größenordnung $n^{2\lambda}$ ist, vgl. Matthews und Rosenberger [12, Abschnitt 4], scheint die folgende Skalierung zweckmäßig zu sein:

Wir setzen $X_0 := 0 =: Y_0$ und für $n \geq 1$

$$(2.7) \quad X_n := \frac{B_n^b - \mu_b(n)}{n^\lambda},$$

$$(2.8) \quad Y_n := \frac{B_n^w - \mu_w(n)}{n^\lambda}.$$

Für diese ergibt sich durch die obigen Rekursionen (2.2) und (2.3) für $n \geq 1$:

$$(2.9) \quad X_n \stackrel{d}{=} \left(\frac{I_n}{n}\right)^\lambda X_{I_n}^{(1)} + F_\alpha \left(\frac{J_n}{n}\right)^\lambda X_{J_n}^{(2)} + (1 - F_\alpha) \left(\frac{J_n}{n}\right)^\lambda Y_{J_n} + b'_b(n),$$

$$(2.10) \quad Y_n \stackrel{d}{=} \left(\frac{I_n}{n}\right)^\lambda Y_{I_n}^{(1)} + F_\beta \left(\frac{J_n}{n}\right)^\lambda Y_{J_n}^{(2)} + (1 - F_\beta) \left(\frac{J_n}{n}\right)^\lambda X_{J_n} + b'_w(n)$$

mit

$$(2.11) \quad b'_b(n) = d'_b \left(\left(\left(\frac{I_n}{n} \right)^\lambda + F_\alpha \left(\frac{J_n}{n} \right)^\lambda - 1 \right) + d'_w (1 - F_\alpha) \left(\frac{J_n}{n} \right)^\lambda + \mathcal{O}(n^{-\lambda}) \right),$$

$$(2.12) \quad b'_w(n) = d'_w \left(\left(\frac{I_n}{n} \right)^\lambda + F_\beta \left(\frac{J_n}{n} \right)^\lambda - 1 \right) + d'_b (1 - F_\beta) \left(\frac{J_n}{n} \right)^\lambda + \mathcal{O}(n^{-\lambda})$$

mit den folgenden Bedingungen an Unabhängigkeit und Verteilung, wie aus (2.4), (2.5) und (2.6) folgt:

$$\begin{aligned} & \left(X_k^{(1)} \right)_{0 \leq k \leq n-1}, \left(X_k^{(2)} \right)_{0 \leq k \leq n-1}, (Y_k)_{0 \leq k \leq n-1}, F_\alpha \text{ und } I_n \text{ seien unabhängig, ebenso} \\ & \left(Y_k^{(1)} \right)_{0 \leq k \leq n-1}, \left(Y_k^{(2)} \right)_{0 \leq k \leq n-1}, (X_k)_{0 \leq k \leq n-1}, F_\beta \text{ und } I_n, \end{aligned}$$

und es gilt

$$X_j^{(1)} \stackrel{d}{=} X_j \stackrel{d}{=} X_j^{(2)} \text{ sowie } Y_j^{(1)} \stackrel{d}{=} Y_j \stackrel{d}{=} Y_j^{(2)} \text{ für alle } j = 0, \dots, n-1.$$

Um nun eine Fixpunktgleichung aufstellen zu können, müssen wir herausfinden, welches Verhalten $\frac{I_n}{n}$ und $\frac{J_n}{n}$ für $n \rightarrow \infty$ aufweisen. Wir wissen, dass $I_n + 1$ und $J_n + 1$ die Anzahlen der Blätter im linken Teilbaum und im rechten Teilbaum des zugehörigen Baums sind. Diese beiden Größen können wir ebenfalls mithilfe einer Pólya-Urne analysieren, die wir die *induzierte Urne* nennen:

Wir verteilen im zugehörigen Baum zwei neue Farben an die Blätter: Die Blätter im linken Teilbaum erhalten die Farbe 1, die Blätter im rechten Teilbaum die Farbe 2. Wir wissen weiter, dass im Falle des Ziehens aus dem linken Teilbaum dessen Anzahl von Blättern um eins wächst. Ebenso verhält es sich, wenn aus dem rechten Teilbaum ein Blatt ausgewählt wird, da die Urne balanciert ist. Im Rahmen der induzierten Urne bedeutet dies, dass beim Ziehen einer Kugel der Farbe 1 eine weitere Kugel der Farbe 1 mit der gezogenen Kugel in die Urne zurückgelegt wird. Gleichermaßen wird beim Ziehen einer Kugel der Farbe 2 verfahren. Die induzierte Urne ist also gerade eine Pólya-Urne mit zwei Farben und der zweidimensionalen Einheitsmatrix als Rücklegematrix, in der sich zu Beginn je eine Kugel pro Farbe befindet. Die Größe $I_n + 1$ beschreibt die Anzahl der Kugeln der Farbe 1 nach $(n-1)$ -maligem Ziehen aus der Urne, die Größe $J_n + 1$ die Anzahl der Kugeln der Farbe 2 nach $(n-1)$ -maligem Ziehen. Damit erhalten wir analog zu Knappe und Neiningger [9, Lemma 2.1]:

Die Größen I_n und J_n sind uniform auf $\{0, \dots, n-1\}$ verteilt mit $J_n = n-1 - I_n$. Weiter gilt

$$\left(\frac{I_n}{n}, \frac{J_n}{n} \right) \rightarrow (U, 1-U) \text{ für } n \rightarrow \infty$$

fast sicher und in L_p für alle $p \geq 1$, wobei die Zufallsvariable U uniform auf $[0, 1]$ verteilt ist.

Für einen Grenzwert (X, Y) von $(X_n, Y_n)_{n \geq 1}$ ergäben sich nun also die folgenden Fixpunktgleichungen:

$$(2.13) \quad X \stackrel{d}{=} U^\lambda X^{(1)} + F_\alpha (1-U)^\lambda X^{(2)} + (1-F_\alpha) (1-U)^\lambda Y^{(1)} + b'_b,$$

$$(2.14) \quad Y \stackrel{d}{=} U^\lambda Y^{(1)} + F_\beta (1-U)^\lambda Y^{(2)} + (1-F_\beta) (1-U)^\lambda X^{(1)} + b'_w$$

mit

$$(2.15) \quad b'_b = d'_b \left(U^\lambda + F_\alpha (1 - U)^\lambda - 1 \right) + d'_w (1 - F_\alpha) (1 - U)^\lambda,$$

$$(2.16) \quad b'_w = d'_w \left(U^\lambda + F_\beta (1 - U)^\lambda - 1 \right) + d'_b (1 - F_\beta) (1 - U)^\lambda,$$

wobei $X^{(1)}, X^{(2)}, Y^{(1)}, Y^{(2)}, F_\alpha, F_\beta$ und U unabhängig sind und $X^{(1)} \stackrel{d}{=} X^{(2)} \stackrel{d}{=} X$ sowie $Y^{(1)} \stackrel{d}{=} Y^{(2)} \stackrel{d}{=} Y$ gilt. Wegen $U^\lambda \stackrel{d}{=} (1 - U)^\lambda$ gilt mit den Bezeichnungen aus Lemma 2.1

$$\begin{aligned} \mathbb{E}[b'_b] &= d'_b \left(\mathbb{E}[U^\lambda] + \alpha \mathbb{E}[(1 - U)^\lambda] - 1 \right) + d'_w (1 - \alpha) \mathbb{E}[(1 - U)^\lambda] \\ &= \mathbb{E}[U^\lambda] \left((1 + \alpha) d'_b + (1 - \alpha) d'_w \right) - d'_b \\ &= \frac{1}{\lambda + 1} \frac{(1 + \alpha)(1 - \alpha) - (1 - \alpha)(1 - \beta)}{(1 - \lambda) \Gamma(\lambda + 1)} - d'_b \\ &= \frac{1}{\alpha + \beta} \frac{(1 - \alpha)(\alpha + \beta)}{(1 - \lambda) \Gamma(\lambda + 1)} - d'_b = 0 \end{aligned}$$

und ebenso $\mathbb{E}[b'_w] = 0$.

3 Technische Vorbemerkungen

Räume

Es bezeichne \mathcal{M} den Raum aller Wahrscheinlichkeitsmaße auf \mathbb{R} mit der Borel- σ -Algebra. Weiter betrachten wir den Teilraum $\mathcal{M}_p := \{\mathcal{L}(X) \in \mathcal{M} \mid \mathbb{E}[|X|^p] < \infty\}$ mit $p > 0$ von \mathcal{M} sowie

$$\mathcal{M}_p(m) := \{\mathcal{L}(X) \in \mathcal{M}_p \mid \mathbb{E}[X] = m\}, p \geq 1, m \in \mathbb{R}.$$

Für $d \in \mathbb{N}$ bezeichnen wir das d -fache kartesische Produkt von \mathcal{M} bzw. \mathcal{M}_p und $\mathcal{M}_p(m)$ mit

$$\mathcal{M}^{\times d} := \underbrace{\mathcal{M} \times \dots \times \mathcal{M}}_{d\text{-mal}}$$

bzw. ebenso $(\mathcal{M}_p)^{\times d}$ und $(\mathcal{M}_p(m))^{\times d}$.

Wasserstein-Metrik

Für zwei Wahrscheinlichkeitsmaße $\mu, \nu \in \mathcal{M}_p$, $1 \leq p < \infty$ ist die Wasserstein-Metrik ℓ_p definiert durch

$$\ell_p(\mu, \nu) := \inf \left\{ \|X - Y\|_p \mid \mathcal{L}(X) = \mu, \mathcal{L}(Y) = \nu \right\},$$

wobei $\|X - Y\|_p := (\mathbb{E}[|X - Y|^p])^{\frac{1}{p}}$ die L_p -Norm bezeichnet. Sind V, W reelle Zufallsvariablen, so schreiben wir $\ell_p(V, W) := \ell_p(\mathcal{L}(V), \mathcal{L}(W))$. Zwei Zufallsvariablen V' bzw. W' mit Verteilungen μ bzw. $\nu \in \mathcal{M}_p$, für die $\|V' - W'\|_p = \ell_p(\mu, \nu)$ gilt, heißen optimale ℓ_p -Couplings von μ und ν .

Die folgende Aussage wird in den Beweisen zur Konvergenzrate eine grundlegende Rolle spielen und geht zurück auf Major [11, Theorem (8.1)] sowie Bickel und Freedman [4, Lemma 8.2].

3.1 Lemma

Für alle $1 \leq p < \infty$ und jede Familie $P \subset \mathcal{M}_p$ gilt: Es existieren Zufallsvariablen W_μ für $\mu \in P$, sodass $\mathcal{L}(W_\mu) = \mu$ gilt und für alle $\mu, \nu \in P$ das Paar (W_μ, W_ν) ein optimales ℓ_p -Coupling von μ und ν ist; insbesondere gibt es ein solches Coupling, das für alle $p \geq 1$ optimal ist.

Für Wahrscheinlichkeitsmaße $(\mu_1, \dots, \mu_d), (\nu_1, \dots, \nu_d) \in (\mathcal{M}_p)^{\times d}$ wird die Wasserstein-Metrik ℓ_p^\vee auf dem Produktraum $(\mathcal{M}_p)^{\times d}$ definiert durch

$$\ell_p^\vee((\mu_1, \dots, \mu_d), (\nu_1, \dots, \nu_d)) := \max_{1 \leq i \leq d} \ell_p(\mu_i, \nu_i).$$

Für Zufallsvektoren (V_1, \dots, V_d) und (W_1, \dots, W_d) schreiben wir analog zu oben

$$\ell_p^\vee((V_1, \dots, V_d), (W_1, \dots, W_d)) := \ell_p^\vee((\mathcal{L}(V_1), \dots, \mathcal{L}(V_d)), (\mathcal{L}(W_1), \dots, \mathcal{L}(W_d))).$$

Weiter gilt der folgende Zusammenhang:

3.2 Lemma

Sei $1 \leq q < p < \infty$ und seien X, Y zwei reellwertige Zufallsvariablen. Dann gilt

$$\ell_q(X, Y) \leq \ell_p(X, Y).$$

Beweis. Man setze $r := \frac{p}{q}$ und $s := \frac{p}{p-q}$. Dann gilt offenbar $\frac{q}{p} + \frac{p-q}{p} = 1$ und wir erhalten mit der Hölderschen Ungleichung für eine Zufallsvariable Z mit $\mathbb{E}[|Z|^p] < \infty$ (andernfalls gilt die Ungleichung trivialerweise):

$$\|Z\|_q^q = \mathbb{E}[|Z|^q] = \| |Z|^q \|_1 \leq \| |Z|^q \|_r \cdot \|1\|_s = \mathbb{E}\left[|Z|^{q \cdot \frac{p}{q}}\right]^{\frac{q}{p}} = \|Z\|_p^q.$$

Daraus folgt

$$(3.1) \quad \|Z\|_q \leq \|Z\|_p.$$

Dies liefert weiter

$$\begin{aligned} \ell_q(X, Y) &= \inf \left\{ \|X' - Y'\|_q \mid \mathcal{L}(X') = \mathcal{L}(X), \mathcal{L}(Y') = \mathcal{L}(Y) \right\} \\ &\leq \inf \left\{ \|X' - Y'\|_p \mid \mathcal{L}(X') = \mathcal{L}(X), \mathcal{L}(Y') = \mathcal{L}(Y) \right\} = \ell_p(X, Y). \end{aligned}$$

□

Schließlich benötigen wir, um den Banachschen Fixpunktsatz im folgenden Abschnitt anwenden zu können, die Vollständigkeit der betrachteten Räume:

3.3 Lemma

Die Paare (\mathcal{M}_p, ℓ_p) , $(\mathcal{M}_p(m), \ell_p)$ und $\left((\mathcal{M}_p(m))^{\times d}, \ell_p^\vee\right)$ mit $1 \leq p < \infty$, $m \in \mathbb{R}$ und $d \in \mathbb{N}$ sind vollständige metrische Räume.

Beweis. Seien also $1 \leq p < \infty$, $m \in \mathbb{R}$ und $d \in \mathbb{N}$. Dass (\mathcal{M}_p, ℓ_p) ein metrischer Raum ist, liefern Bickel und Freedman in [4, Lemma 8.1]. Damit ist $(\mathcal{M}_p(m), \ell_p)$ als Teilraum von (\mathcal{M}_p, ℓ_p) ebenfalls ein metrischer Raum. Positive Definitheit, Symmetrie und die Dreiecksungleichung folgen für ℓ_p^\vee direkt daraus, dass ℓ_p eine Metrik ist; also ist auch $\left((\mathcal{M}_p(m))^{\times d}, \ell_p^\vee\right)$ ein metrischer Raum.

Zur Vollständigkeit von (\mathcal{M}_p, ℓ_p) : Sei $(\mu_n)_{n \geq 1}$ eine Cauchyfolge in (\mathcal{M}_p, ℓ_p) . Dann gibt es wegen Lemma 3.1 eine Folge von Zufallsvariablen $(W_n)_{n \geq 1}$, welche auf einem gemeinsamen Wahrscheinlichkeitsraum $(\Omega, \mathcal{A}, \mathbb{P})$ definiert sind, sodass das Paar (W_n, W_m) ein optimales ℓ_p -Coupling von μ_n und μ_m für alle $n, m \geq 1$ ist. Wegen $\|W_n - W_m\|_p = \ell_p(\mu_n, \mu_m)$ ist $(W_n)_{n \geq 1}$ eine Cauchyfolge in $(L_p(\Omega, \mathcal{A}, \mathbb{P}), \|\cdot\|_p)$. Der Satz von Riesz-Fischer liefert nun die Existenz eines Grenzwerts $W \in L_p(\Omega, \mathcal{A}, \mathbb{P})$ mit $\|W_n - W\|_p \rightarrow 0$ für $n \rightarrow \infty$. Mit $\mu := \mathcal{L}(W)$ gilt nun weiter: $\ell_p(\mu_n, \mu) \leq \|W_n - W\|_p \rightarrow 0$ für $n \rightarrow \infty$. Damit folgt die Behauptung.

Der Teilraum $\mathcal{M}_p(m)$ von (\mathcal{M}_p, ℓ_p) ist abgeschlossen und somit vollständig: Sei $(\mu_n)_{n \geq 1}$ eine Cauchyfolge in $\mathcal{M}_p(m)$ mit Grenzwert $\mu \in \mathcal{M}_p$. Mit Lemma 3.1 können wir wieder Zufallsvariablen X_n und X auf einem gemeinsamen Wahrscheinlichkeitsraum mit $\mathcal{L}(X_n) = \mu_n$ für $n \in \mathbb{N}$ und $\mathcal{L}(X) = \mu$ konstruieren, sodass X_n und X optimale Couplings von μ_n und μ für alle $n \in \mathbb{N}$ sind. Dann gilt mit der Jensenschen Ungleichung

$$\begin{aligned} |\mathbb{E}[X_n] - \mathbb{E}[X]| &\leq \mathbb{E}[|X_n - X|] = \|X_n - X\|_1 \\ &\stackrel{(3.1)}{\leq} \|X_n - X\|_p = \ell_p(\mu_n, \mu) \rightarrow 0 \quad \text{für } n \rightarrow \infty. \end{aligned}$$

Für alle $n \in \mathbb{N}$ gilt außerdem $\mathbb{E}[X_n] = m$ und damit folgt $\mathbb{E}[X] = m$. Also gilt $\mu \in \mathcal{M}_p(m)$.

Auch $((\mathcal{M}_p(m))^{\times d}, \ell_p^\vee)$ ist vollständig: Sei $(\mu_n)_{n \geq 1}$ mit $\mu_n = (\mu_n^{(1)}, \dots, \mu_n^{(d)})$ eine Cauchyfolge in $((\mathcal{M}_p(m))^{\times d}, \ell_p^\vee)$. Dann sind die Folgen $(\mu_n^{(1)})_{n \geq 1}, \dots, (\mu_n^{(d)})_{n \geq 1}$ jeweils Cauchyfolgen in $(\mathcal{M}_p(m), \ell_p)$. Demzufolge existieren $\mu^{(1)}, \dots, \mu^{(d)} \in \mathcal{M}_p(m)$ mit $\ell_p(\mu_n^{(i)}, \mu^{(i)}) \rightarrow 0$ für $n \rightarrow \infty$ und $i = 1, \dots, d$. Mit $\mu := (\mu^{(1)}, \dots, \mu^{(d)}) \in (\mathcal{M}_p(m))^{\times d}$ gilt dann auch $\ell_p^\vee(\mu_n, \mu) \rightarrow 0$ für $n \rightarrow \infty$ und damit folgt die Behauptung. \square

Kolmogorov-Metrik

Seien $\mu, \nu \in \mathcal{M}$. Die Kolmogorov-Metrik ϱ zwischen μ und ν ist definiert durch

$$\varrho(\mu, \nu) := \sup_{x \in \mathbb{R}} |\mu((-\infty, x]) - \nu((-\infty, x])|.$$

Für $(\mu_1, \dots, \mu_d), (\nu_1, \dots, \nu_d) \in \mathcal{M}^{\times d}$ ist die Kolmogorov-Metrik ϱ^\vee auf dem Produktraum definiert durch

$$\varrho^\vee((\mu_1, \dots, \mu_d), (\nu_1, \dots, \nu_d)) := \max_{1 \leq i \leq d} \varrho(\mu_i, \nu_i).$$

Sind X und Y reelle Zufallsvariablen mit Verteilungen μ und ν , so schreiben wir auch

$$\varrho(X, Y) := \varrho(\mu, \nu)$$

und verfahren gleichermaßen für reellwertige Zufallsvektoren.

Zwischen der Kolmogorov-Metrik und der Wasserstein-Metrik besteht der folgende, aus Fill und Janson [6, Lemma 5.1] bekannte Zusammenhang:

3.4 Satz

Seien X, Y reelle Zufallsvariablen, wobei Y eine beschränkte Lebesgue-Dichte f habe. Dann gilt für alle $1 \leq p < \infty$

$$\varrho(X, Y) \leq (p+1)^{\frac{1}{p+1}} (\|f\|_\infty \ell_p(X, Y))^{\frac{p}{p+1}}.$$

Dieser Satz wird in Kapitel 7 das entscheidende Werkzeug sein, um die Konvergenzrate von den Wasserstein-Metriken auf die Kolmogorov-Metrik zu übertragen.

4 Existenz und Eindeutigkeit eines Fixpunkts

Das System von Fixpunktgleichungen, bestehend aus (2.13) und (2.14), werden wir nun als Abbildung zwischen Maßräumen auffassen. Auf diesem Wege können wir das System (2.13) – (2.14) hinsichtlich der Existenz und Eindeutigkeit eines Fixpunkts untersuchen.

Seien dazu $d, d' \in \mathbb{N}$, $(A_{ir})_{ir}$ eine $d \times d'$ -Matrix aus reellwertigen Zufallsvariablen, $(b_1, \dots, b_{d'})$ ein reellwertiger Zufallsvektor und $(\pi(i, r))_{ir}$ eine $d \times d'$ -Matrix mit $\pi(i, r) \in \{1, \dots, d\}$ für $1 \leq i \leq d$ und $1 \leq r \leq d'$. Wir betrachten nun die Abbildung

$$(4.1) \quad T : \mathcal{M}^{\times d} \rightarrow \mathcal{M}^{\times d} \\ (\mu_1, \dots, \mu_d) \mapsto (T_1(\mu_1, \dots, \mu_d), \dots, T_d(\mu_1, \dots, \mu_d)) \text{ mit}$$

$$(4.2) \quad T_i((\mu_1, \dots, \mu_d)) := \mathcal{L} \left(\sum_{r=1}^{d'} A_{ir} Z_{ir} + b_i \right),$$

wobei $\mathcal{L}(Z_{ir}) = \mu_{\pi(i,r)}$ gilt und $(A_{i1}, \dots, A_{id'}, b_i), Z_{i1}, \dots, Z_{id'}$ unabhängig sind für $r = 1, \dots, d'$ und alle Komponenten $i = 1, \dots, d$.

4.1 Satz

Die Zufallsvariablen A_{ir} und b_i aus (4.1) – (4.2) seien quadratintegrierbar mit $\mathbb{E}[b_i] = 0$ für alle $1 \leq i \leq d$ und $1 \leq r \leq d'$ und es gelte

$$\max_{1 \leq i \leq d} \sum_{r=1}^{d'} \mathbb{E}[A_{ir}^2] < 1.$$

Dann hat die Abbildung T eingeschränkt auf $(\mathcal{M}_2(0))^{\times d}$ einen eindeutigen Fixpunkt.

Beweis. Sei $\mu \in (\mathcal{M}_2(0))^{\times d}$, dann gilt $T(\mu) \in (\mathcal{M}_2(0))^{\times d}$ und für X mit $\mathcal{L}(X) = \mu$ gilt für $1 \leq i \leq d$

$$\mathbb{E}[T_i(\mu)] = \mathbb{E} \left[\sum_{r=1}^{d'} A_{ir} X_r + b_i \right] = \sum_{r=1}^{d'} \mathbb{E}[A_{ir}] \underbrace{\mathbb{E}[X_r]}_{=0} + \underbrace{\mathbb{E}[b_i]}_{=0} = 0.$$

Also folgt $T(\mu) \in (\mathcal{M}_2(0))^{\times d}$.

Wir zeigen nun, dass T bezüglich ℓ_2^\vee eine Kontraktion ist, d.h.

$$\ell_2^\vee(T(\mu), T(\nu)) \leq q \cdot \ell_2^\vee(\mu, \nu) \text{ mit } q \in (0, 1) \text{ und } \mu, \nu \in (\mathcal{M}_2(0))^{\times d}.$$

Dann folgt, da $((\mathcal{M}_2(0))^{\times d}, \ell_2^\vee)$ ein vollständiger metrischer Raum ist, die Behauptung aus dem Banachschen Fixpunktsatz.

Seien dazu $\mu, \nu \in (\mathcal{M}_2(0))^{\times d}$. Nun existieren mit Lemma 3.1 optimale ℓ_2 -Couplings (W_{ir}, V_{ir}) von $\mu_{\pi(i,r)}$ und $\nu_{\pi(i,r)}$ für $i = 1, \dots, d$ und $r = 1, \dots, d'$, sodass

$$(W_{i1}, V_{i1}), \dots, (W_{id'}, V_{id'}), (A_{i1}, \dots, A_{id'}, b_i)$$

für $i = 1, \dots, d$ unabhängig sind.

Dann gilt für $1 \leq i \leq d$

$$\begin{aligned}
(\ell_2(T_i(\mu), T_i(\nu)))^2 &\leq \left\| \sum_{r=1}^{d'} A_{ir} (W_{ir} - V_{ir}) \right\|_2^2 = \mathbb{E} \left[\left(\sum_{r=1}^{d'} A_{ir} (W_{ir} - V_{ir}) \right)^2 \right] \\
&= \sum_{r=1}^{d'} \mathbb{E} \left[(A_{ir} (W_{ir} - V_{ir}))^2 \right] \\
&\quad + \sum_{r,s=1, r \neq s}^{d'} \mathbb{E} [(A_{ir} (W_{ir} - V_{ir})) (A_{is} (W_{is} - V_{is}))] \\
&= \sum_{r=1}^{d'} \mathbb{E} [A_{ir}^2] \mathbb{E} [(W_{ir} - V_{ir})^2] \\
&\quad + \sum_{r,s=1, r \neq s}^{d'} \mathbb{E} [A_{ir} A_{is}] \underbrace{\mathbb{E} [(W_{ir} - V_{ir})]}_{=0} \underbrace{\mathbb{E} [(W_{is} - V_{is})]}_{=0} \\
&= \sum_{r=1}^{d'} \mathbb{E} [A_{ir}^2] \mathbb{E} [(W_{ir} - V_{ir})^2] = \sum_{r=1}^{d'} \mathbb{E} [A_{ir}^2] (\ell_2(W_{ir}, V_{ir}))^2 \\
&\leq \sum_{r=1}^{d'} \mathbb{E} [A_{ir}^2] (\ell_2^\vee(\mu, \nu))^2 \leq \max_{1 \leq i \leq d} \sum_{r=1}^{d'} \mathbb{E} [A_{ir}^2] (\ell_2^\vee(\mu, \nu))^2.
\end{aligned}$$

Daraus folgt

$$\ell_2^\vee(T(\mu), T(\nu)) \leq \sqrt{\max_{1 \leq i \leq d} \sum_{r=1}^{d'} \mathbb{E} [A_{ir}^2] \cdot \ell_2^\vee(\mu, \nu)}.$$

Da nach Voraussetzung $\max_{1 \leq i \leq d} \sum_{r=1}^{d'} \mathbb{E} [A_{ir}^2] < 1$ gilt, folgt die Behauptung. \square

Nun ist zu überprüfen, ob die Voraussetzungen von Satz 4.1 für das System (2.13) und (2.14) erfüllt sind. Wir betrachten dabei die folgende Matrix

$$\begin{pmatrix} U^\lambda & F_\alpha (1 - U)^\lambda & (1 - F_\alpha) (1 - U)^\lambda \\ U^\lambda & F_\beta (1 - U)^\lambda & (1 - F_\beta) (1 - U)^\lambda \end{pmatrix}$$

mit $\frac{1}{2} < \lambda < 1$ und den Zufallsvektor (b'_b, b'_w) , dessen Komponenten durch (2.15) und (2.16) gegeben sind. Offenbar sind die Einträge der Matrix und die Komponenten des Vektors quadratintegrierbar. Wir wissen bereits, dass $\mathbb{E} [b'_b] = \mathbb{E} [b'_w] = 0$ erfüllt ist. Weiter gilt

$$\begin{aligned}
&\mathbb{E} [U^{2\lambda}] + \mathbb{E} [F_\alpha^2 (1 - U)^{2\lambda}] + \mathbb{E} [(1 - F_\alpha)^2 (1 - U)^{2\lambda}] \\
&= \mathbb{E} [U^{2\lambda}] + \mathbb{E} [(1 - U)^{2\lambda}] = \frac{2}{2\lambda + 1} < 1,
\end{aligned}$$

da $\lambda > \frac{1}{2}$. Ebenso folgt

$$\mathbb{E} [U^{2\lambda}] + \mathbb{E} [F_\beta^2 (1 - U)^{2\lambda}] + \mathbb{E} [(1 - F_\beta)^2 (1 - U)^{2\lambda}] < 1.$$

Das System bestehend aus (2.13) und (2.14) charakterisiert also eine Verteilung. Wir werden diesen eindeutigen Fixpunkt in $\mathcal{M}_2(0) \times \mathcal{M}_2(0)$ im weiteren Verlauf der Arbeit mit $(\mathcal{L}(\Lambda'_b), \mathcal{L}(\Lambda'_w))$ bezeichnen. Es gilt also

$$(4.3) \quad \Lambda'_b \stackrel{d}{=} U^\lambda \Lambda_b'^{(1)} + F_\alpha (1-U)^\lambda \Lambda_b'^{(2)} + (1-F_\alpha) (1-U)^\lambda \Lambda_w'^{(1)} + b'_b,$$

$$(4.4) \quad \Lambda'_w \stackrel{d}{=} U^\lambda \Lambda_w'^{(1)} + F_\beta (1-U)^\lambda \Lambda_w'^{(2)} + (1-F_\beta) (1-U)^\lambda \Lambda_b'^{(1)} + b'_w$$

mit b'_b und b'_w wie in (2.15) und (2.16). Weiter gilt $\Lambda_b'^{(1)} \stackrel{d}{=} \Lambda'_b \stackrel{d}{=} \Lambda_b'^{(2)}$ sowie $\Lambda_w'^{(1)} \stackrel{d}{=} \Lambda'_w \stackrel{d}{=} \Lambda_w'^{(2)}$ und, dass $\Lambda_b'^{(1)}, \Lambda_b'^{(2)}, \Lambda_w'^{(1)}, \Lambda_w'^{(2)}, F_\alpha, F_\beta$ und U unabhängig sind.

5 Konvergenzrate für die Wasserstein-Metrik ℓ_2^\vee

Wir wollen nun mittels Induktion eine obere Schranke für die Konvergenzrate in der Wasserstein-Metrik ℓ_2^\vee beweisen. Dazu werden wir zunächst eine rekursive Darstellung für den Abstand in dieser Metrik herleiten, anschließend bestimmte Terme, die in dieser Rekursion auftauchen, abschätzen und schließlich den Induktionsbeweis führen.

5.1 Satz: Obere Schranke für die Konvergenzrate in ℓ_2^\vee

Wir betrachten die Pólya-Urne mit zufälliger Rücklegematrix (2.1) mit $\alpha, \beta \in (0, 1)$ und $\lambda := \alpha + \beta - 1$. Es sei $\lambda > \frac{1}{2}$ und die Zufallsvariablen X_n und Y_n mit $n \in \mathbb{N}$ seien wie in (2.7) und (2.8) die zentrierten und skalierten Anzahlen schwarzer Kugeln nach n Schritten. Weiter bezeichne $(\mathcal{L}(\Lambda'_b), \mathcal{L}(\Lambda'_w))$ den eindeutigen Fixpunkt von (2.13) – (2.14) in $\mathcal{M}_2(0) \times \mathcal{M}_2(0)$ und es sei $\varepsilon > 0$. Dann existiert eine Konstante $0 < C < \infty$, sodass für alle $n \in \mathbb{N}$ gilt:

$$\ell_2^\vee((X_n, Y_n), (\Lambda'_b, \Lambda'_w)) \leq \frac{C}{n^{\lambda - \frac{1}{2} - \varepsilon}}.$$

Beweis. Wir führen zunächst die folgenden Bezeichnungen für $n \in \mathbb{N}$ ein:

$$\begin{aligned} \Delta(n) &:= \ell_2^\vee((X_n, Y_n), (\Lambda'_b, \Lambda'_w)), & \Delta^2(n) &:= (\Delta(n))^2, \\ \Delta_b(n) &:= \ell_2(X_n, \Lambda'_b), & \Delta_b^2(n) &:= (\Delta_b(n))^2, \\ \Delta_w(n) &:= \ell_2(Y_n, \Lambda'_w), & \Delta_w^2(n) &:= (\Delta_w(n))^2 \text{ sowie} \\ M &:= \max\{\|\Lambda'_b\|_2, \|\Lambda'_w\|_2\}. \end{aligned}$$

Mit Lemma 3.1 können wir nun die Zufallsvariablen $X_j^{(1)}, X_j^{(2)}, X_j, Y_j^{(1)}, Y_j^{(2)}, Y_j$ für $j = 0, \dots, n-1$ sowie F_α und F_β aus (2.9) und (2.10) und $\Lambda_b^{(1)}, \Lambda_b^{(2)}, \Lambda_w^{(1)}, \Lambda_w^{(2)}$ und U aus (4.3) und (4.4) auf einem gemeinsamen Wahrscheinlichkeitsraum so wählen, dass für $n \in \mathbb{N}$ und $j = 0, \dots, n-1$ gilt:

1. $I_n \stackrel{d}{=} \lfloor nU \rfloor$.
2. Die Paare $(X_j^{(1)}, \Lambda_b^{(1)})$, $(X_j^{(2)}, \Lambda_b^{(2)})$ und $(X_j, \Lambda_b^{(1)})$ sind optimale ℓ_2 -Couplings von $\mathcal{L}(X_j)$ und $\mathcal{L}(\Lambda'_b)$.
3. Die Paare $(Y_j^{(1)}, \Lambda_w^{(1)})$, $(Y_j^{(2)}, \Lambda_w^{(2)})$ und $(Y_j, \Lambda_w^{(1)})$ sind optimale ℓ_2 -Couplings von $\mathcal{L}(Y_j)$ und $\mathcal{L}(\Lambda'_w)$.
4. $(X_j^{(1)}, \Lambda_b^{(1)})$, $(X_j^{(2)}, \Lambda_b^{(2)})$, $(X_j, \Lambda_b^{(1)})$, $(Y_j^{(1)}, \Lambda_w^{(1)})$, $(Y_j^{(2)}, \Lambda_w^{(2)})$, $(Y_j, \Lambda_w^{(1)})$, F_α , F_β und U sind unabhängig.

Wir widmen uns zunächst einer rekursiven Abschätzung für den quadrierten Abstand $\Delta^2(n)$, um mithilfe dieser schließlich per Induktion eine obere Schranke für die Konvergenzrate herleiten zu können.

5.2 Lemma

Sei $n \in \mathbb{N}$, dann gilt:

$$\begin{aligned} \Delta^2(n) &\leq 2\mathbb{E} \left[\left(\frac{I_n}{n} \right)^{2\lambda} \Delta^2(I_n) \right] + 2M^2 \left\| \left(\frac{I_n}{n} \right)^\lambda - U^\lambda \right\|_2^2 \\ &\quad + 4M\mathbb{E} \left[\left| \left(\frac{I_n}{n} \right)^\lambda \left(\left(\frac{I_n}{n} \right)^\lambda - U^\lambda \right) \right| \cdot \Delta(I_n) \right] \\ &\quad + \max \left\{ \|b'_b(n) - b'_b\|_2^2, \|b'_w(n) - b'_w\|_2^2 \right\}. \end{aligned}$$

Beweis. Für $n \in \mathbb{N}$ gilt:

$$\begin{aligned} \Delta_b^2(n) &\leq \|X_n - \Lambda'_b\|_2^2 \\ &= \mathbb{E} \left[\left| \left(\frac{I_n}{n} \right)^\lambda X_{I_n}^{(1)} - U^\lambda \Lambda_b'^{(1)} + F_\alpha \left(\left(\frac{J_n}{n} \right)^\lambda X_{J_n}^{(2)} - (1-U)^\lambda \Lambda_b'^{(2)} \right) \right. \right. \\ &\quad \left. \left. + (1-F_\alpha) \left(\left(\frac{J_n}{n} \right)^\lambda Y_{J_n} - (1-U)^\lambda \Lambda_w'^{(1)} \right) + b'_b(n) - b'_b \right|^2 \right] \\ (5.1) \quad &= \alpha \left\| \left(\frac{I_n}{n} \right)^\lambda X_{I_n}^{(1)} - U^\lambda \Lambda_b'^{(1)} + \left(\frac{J_n}{n} \right)^\lambda X_{J_n}^{(2)} - (1-U)^\lambda \Lambda_b'^{(2)} + b'_b(n) - b'_b \right\|_2^2 \\ &\quad + (1-\alpha) \left\| \left(\frac{I_n}{n} \right)^\lambda X_{I_n}^{(1)} - U^\lambda \Lambda_b'^{(1)} + \left(\frac{J_n}{n} \right)^\lambda Y_{J_n} - (1-U)^\lambda \Lambda_w'^{(1)} + b'_b(n) - b'_b \right\|_2^2. \end{aligned}$$

Wir betrachten nun den ersten Summanden von (5.1):

$$\begin{aligned} &\left\| \left(\frac{I_n}{n} \right)^\lambda X_{I_n}^{(1)} - U^\lambda \Lambda_b'^{(1)} + \left(\frac{J_n}{n} \right)^\lambda X_{J_n}^{(2)} - (1-U)^\lambda \Lambda_b'^{(2)} + b'_b(n) - b'_b \right\|_2^2 \\ (5.2) \quad &= \left\| \left(\frac{I_n}{n} \right)^\lambda X_{I_n}^{(1)} - U^\lambda \Lambda_b'^{(1)} + \left(\frac{J_n}{n} \right)^\lambda X_{J_n}^{(2)} - (1-U)^\lambda \Lambda_b'^{(2)} \right\|_2^2 + \|b'_b(n) - b'_b\|_2^2 \\ &\quad + 2\mathbb{E} \left[\left(\left(\frac{I_n}{n} \right)^\lambda X_{I_n}^{(1)} - U^\lambda \Lambda_b'^{(1)} + \left(\frac{J_n}{n} \right)^\lambda X_{J_n}^{(2)} - (1-U)^\lambda \Lambda_b'^{(2)} \right) (b'_b(n) - b'_b) \right], \end{aligned}$$

wobei der letzte Summand wegen der oben gestellten Unabhängigkeitsbedingungen und

$$\mathbb{E} \left[X_{I_n}^{(1)} \right] = \mathbb{E} \left[X_{J_n}^{(2)} \right] = \mathbb{E} \left[\Lambda_b'^{(1)} \right] = \mathbb{E} \left[\Lambda_b'^{(2)} \right] = 0$$

gleich null ist.

Weiter ergibt sich für den ersten Summanden von (5.2):

$$\begin{aligned}
& \left\| \left(\frac{I_n}{n} \right)^\lambda X_{I_n}^{(1)} - U^\lambda \Lambda_b'^{(1)} + \left(\frac{J_n}{n} \right)^\lambda X_{J_n}^{(2)} - (1-U)^\lambda \Lambda_b'^{(2)} \right\|_2^2 \\
&= \left\| \left(\frac{I_n}{n} \right)^\lambda X_{I_n}^{(1)} - U^\lambda \Lambda_b'^{(1)} \right\|_2^2 + \left\| \left(\frac{J_n}{n} \right)^\lambda X_{J_n}^{(2)} - (1-U)^\lambda \Lambda_b'^{(2)} \right\|_2^2 \\
&+ 2\mathbb{E} \left[\left(\left(\frac{I_n}{n} \right)^\lambda X_{I_n}^{(1)} - U^\lambda \Lambda_b'^{(1)} \right) \left(\left(\frac{J_n}{n} \right)^\lambda X_{J_n}^{(2)} - (1-U)^\lambda \Lambda_b'^{(2)} \right) \right],
\end{aligned}$$

wobei auch hier der letzte Summand aus denselben Gründen wie bei (5.2) gleich null ist.

Für den ersten Summanden aus (5.1) gilt also:

$$\begin{aligned}
(5.3) \quad & \left\| \left(\frac{I_n}{n} \right)^\lambda X_{I_n}^{(1)} - U^\lambda \Lambda_b'^{(1)} + \left(\frac{J_n}{n} \right)^\lambda X_{J_n}^{(2)} - (1-U)^\lambda \Lambda_b'^{(2)} + b'_b(n) - b'_b \right\|_2^2 \\
&= \left\| \left(\frac{I_n}{n} \right)^\lambda X_{I_n}^{(1)} - U^\lambda \Lambda_b'^{(1)} \right\|_2^2 + \left\| \left(\frac{J_n}{n} \right)^\lambda X_{J_n}^{(2)} - (1-U)^\lambda \Lambda_b'^{(2)} \right\|_2^2 + \|b'_b(n) - b'_b\|_2^2.
\end{aligned}$$

Analog erhält man für den zweiten Summanden aus (5.1):

$$\begin{aligned}
(5.4) \quad & \left\| \left(\frac{I_n}{n} \right)^\lambda X_{I_n}^{(1)} - U^\lambda \Lambda_b'^{(1)} + \left(\frac{J_n}{n} \right)^\lambda Y_{J_n} - (1-U)^\lambda \Lambda_w'^{(1)} + b'_b(n) - b'_b \right\|_2^2 \\
&= \left\| \left(\frac{I_n}{n} \right)^\lambda X_{I_n}^{(1)} - U^\lambda \Lambda_b'^{(1)} \right\|_2^2 + \left\| \left(\frac{J_n}{n} \right)^\lambda Y_{J_n} - (1-U)^\lambda \Lambda_w'^{(1)} \right\|_2^2 + \|b'_b(n) - b'_b\|_2^2.
\end{aligned}$$

Insgesamt liefern (5.3) und (5.4):

$$\begin{aligned}
(5.5) \quad \Delta_b^2(n) &\leq \left\| \left(\frac{I_n}{n} \right)^\lambda X_{I_n}^{(1)} - U^\lambda \Lambda_b'^{(1)} \right\|_2^2 + \alpha \left\| \left(\frac{J_n}{n} \right)^\lambda X_{J_n}^{(2)} - (1-U)^\lambda \Lambda_b'^{(2)} \right\|_2^2 \\
&\quad + (1-\alpha) \left\| \left(\frac{J_n}{n} \right)^\lambda Y_{J_n} - (1-U)^\lambda \Lambda_w'^{(1)} \right\|_2^2 + \|b'_b(n) - b'_b\|_2^2.
\end{aligned}$$

Wir betrachten nun weiter den ersten Summanden von (5.5) und fügen im zweiten Schritt eine gemischte Folge ein:

$$\begin{aligned}
& \left\| \left(\frac{I_n}{n} \right)^\lambda X_{I_n}^{(1)} - U^\lambda \Lambda_b'^{(1)} \right\|_2^2 \\
&= \left\| \left(\frac{I_n}{n} \right)^\lambda X_{I_n}^{(1)} - U^\lambda \Lambda_b'^{(1)} - \left(\frac{I_n}{n} \right)^\lambda \Lambda_b'^{(1)} + \left(\frac{I_n}{n} \right)^\lambda \Lambda_b'^{(1)} \right\|_2^2 \\
&= \left\| \left(\frac{I_n}{n} \right)^\lambda \left(X_{I_n}^{(1)} - \Lambda_b'^{(1)} \right) + \Lambda_b'^{(1)} \left(\left(\frac{I_n}{n} \right)^\lambda - U^\lambda \right) \right\|_2^2
\end{aligned}$$

$$\begin{aligned}
&= \underbrace{\left\| \left(\frac{I_n}{n} \right)^\lambda \left(X_{I_n}^{(1)} - \Lambda_b'^{(1)} \right) \right\|_2^2}_{(A)} + \underbrace{\left\| \Lambda_b'^{(1)} \left(\left(\frac{I_n}{n} \right)^\lambda - U^\lambda \right) \right\|_2^2}_{(B)} \\
&\quad + 2 \underbrace{\mathbb{E} \left[\left(\frac{I_n}{n} \right)^\lambda \Lambda_b'^{(1)} \left(X_{I_n}^{(1)} - \Lambda_b'^{(1)} \right) \left(\left(\frac{I_n}{n} \right)^\lambda - U^\lambda \right) \right]}_{(C)}.
\end{aligned}$$

Zu (A):

$$\begin{aligned}
\left\| \left(\frac{I_n}{n} \right)^\lambda \left(X_{I_n}^{(1)} - \Lambda_b'^{(1)} \right) \right\|_2^2 &= \mathbb{E} \left[\left(\frac{I_n}{n} \right)^{2\lambda} \left(X_{I_n}^{(1)} - \Lambda_b'^{(1)} \right)^2 \right] \\
&= \sum_{j=0}^{n-1} \frac{1}{n} \left(\frac{j}{n} \right)^{2\lambda} \mathbb{E} \left[\left(X_j^{(1)} - \Lambda_b'^{(1)} \right)^2 \right] \\
&= \sum_{j=0}^{n-1} \frac{1}{n} \left(\frac{j}{n} \right)^{2\lambda} \Delta_b^2(j) = \mathbb{E} \left[\left(\frac{I_n}{n} \right)^{2\lambda} \Delta_b^2(I_n) \right],
\end{aligned}$$

da $X_j^{(1)}$ und $\Lambda_b'^{(1)}$ für $j = 0, \dots, n-1$ optimale ℓ_2 -Couplings sind.

Zu (B):

$$\begin{aligned}
\left\| \Lambda_b'^{(1)} \left(\left(\frac{I_n}{n} \right)^\lambda - U^\lambda \right) \right\|_2^2 &= \mathbb{E} \left[\left(\Lambda_b'^{(1)} \right)^2 \left(\left(\frac{I_n}{n} \right)^\lambda - U^\lambda \right)^2 \right] \\
&= \mathbb{E} \left[\left(\Lambda_b'^{(1)} \right)^2 \right] \mathbb{E} \left[\left(\left(\frac{I_n}{n} \right)^\lambda - U^\lambda \right)^2 \right] \\
&= \|\Lambda_b'\|_2^2 \left\| \left(\frac{I_n}{n} \right)^\lambda - U^\lambda \right\|_2^2,
\end{aligned}$$

wobei das zweite Gleichheitszeichen aus der Unabhängigkeit der Zufallsvariablen folgt.

Zu (C):

$$\begin{aligned}
&\mathbb{E} \left[\left(\frac{I_n}{n} \right)^\lambda \Lambda_b'^{(1)} \left(X_{I_n}^{(1)} - \Lambda_b'^{(1)} \right) \left(\left(\frac{I_n}{n} \right)^\lambda - U^\lambda \right) \right] \\
&\leq \mathbb{E} \left[\left(\frac{I_n}{n} \right)^\lambda \left| \Lambda_b'^{(1)} \right| \left| X_{I_n}^{(1)} - \Lambda_b'^{(1)} \right| \cdot \left| \left(\frac{I_n}{n} \right)^\lambda - U^\lambda \right| \right] \\
&= \int_0^1 \left(\frac{\lfloor nu \rfloor}{n} \right)^\lambda \mathbb{E} \left[\left| \Lambda_b'^{(1)} \right| \left| X_{\lfloor nu \rfloor}^{(1)} - \Lambda_b'^{(1)} \right| \right] \left| \left(\frac{\lfloor nu \rfloor}{n} \right)^\lambda - u^\lambda \right| du \\
&= \sum_{j=0}^{n-1} \int_{\frac{j}{n}}^{\frac{j+1}{n}} \left(\frac{j}{n} \right)^\lambda \mathbb{E} \left[\left| \Lambda_b'^{(1)} \right| \left| X_j^{(1)} - \Lambda_b'^{(1)} \right| \right] \left| \left(\frac{j}{n} \right)^\lambda - u^\lambda \right| du
\end{aligned}$$

$$\begin{aligned}
&\leq \sum_{j=0}^{n-1} \int_{\frac{j}{n}}^{\frac{j+1}{n}} \left(\frac{j}{n}\right)^\lambda \left(\mathbb{E} \left[\left| \Lambda_b'^{(1)} \right|^2 \right]\right)^{\frac{1}{2}} \left(\mathbb{E} \left[\left| X_j^{(1)} - \Lambda_b'^{(1)} \right|^2 \right]\right)^{\frac{1}{2}} \left| \left(\frac{j}{n}\right)^\lambda - u^\lambda \right| du \\
&= \sum_{j=0}^{n-1} \int_{\frac{j}{n}}^{\frac{j+1}{n}} \left(\frac{j}{n}\right)^\lambda \|\Lambda_b'\|_2 \Delta_b(j) \left| \left(\frac{j}{n}\right)^\lambda - u^\lambda \right| du \\
&= \mathbb{E} \left[\left| \left(\frac{I_n}{n}\right)^\lambda \left(\left(\frac{I_n}{n}\right)^\lambda - U^\lambda \right) \right| \cdot \Delta_b(I_n) \right] \|\Lambda_b'\|_2.
\end{aligned}$$

Für die zweite Ungleichung wurde die Höldersche Ungleichung verwendet und anschließend wurde wieder ausgenutzt, dass $X_j^{(1)}$ und $\Lambda_b'^{(1)}$ für $j = 0, \dots, n-1$ optimale ℓ_2 -Couplings sind.

Man erhält nun insgesamt für den ersten Summanden von (5.5):

$$\begin{aligned}
\left\| \left(\frac{I_n}{n}\right)^\lambda X_{I_n}^{(1)} - U^\lambda \Lambda_b'^{(1)} \right\|_2^2 &\leq \mathbb{E} \left[\left(\frac{I_n}{n}\right)^{2\lambda} \Delta_b^2(I_n) \right] + \|\Lambda_b'\|_2^2 \left\| \left(\frac{I_n}{n}\right)^\lambda - U^\lambda \right\|_2^2 \\
&\quad + 2\|\Lambda_b'\|_2 \mathbb{E} \left[\left| \left(\frac{I_n}{n}\right)^\lambda \left(\left(\frac{I_n}{n}\right)^\lambda - U^\lambda \right) \right| \cdot \Delta_b(I_n) \right] \\
&\leq \mathbb{E} \left[\left(\frac{I_n}{n}\right)^{2\lambda} \Delta^2(I_n) \right] + M^2 \left\| \left(\frac{I_n}{n}\right)^\lambda - U^\lambda \right\|_2^2 \\
&\quad + 2M \mathbb{E} \left[\left| \left(\frac{I_n}{n}\right)^\lambda \left(\left(\frac{I_n}{n}\right)^\lambda - U^\lambda \right) \right| \cdot \Delta(I_n) \right].
\end{aligned}$$

Analog kann man

$$\left\| \left(\frac{J_n}{n}\right)^\lambda X_{J_n}^{(2)} - (1-U)^\lambda \Lambda_b'^{(2)} \right\|_2^2 \quad \text{und} \quad \left\| \left(\frac{J_n}{n}\right)^\lambda Y_{J_n} - (1-U)^\lambda \Lambda_w'^{(1)} \right\|_2^2$$

abschätzen, sodass für (5.5) folgt:

$$\begin{aligned}
\Delta_b^2(n) &\leq \mathbb{E} \left[\left(\frac{I_n}{n}\right)^{2\lambda} \Delta^2(I_n) \right] + M^2 \left\| \left(\frac{I_n}{n}\right)^\lambda - U^\lambda \right\|_2^2 \\
&\quad + 2M \mathbb{E} \left[\left| \left(\frac{I_n}{n}\right)^\lambda \left(\left(\frac{I_n}{n}\right)^\lambda - U^\lambda \right) \right| \cdot \Delta(I_n) \right] \\
&\quad + \mathbb{E} \left[\left(\frac{J_n}{n}\right)^{2\lambda} \Delta^2(J_n) \right] + M^2 \left\| \left(\frac{J_n}{n}\right)^\lambda - (1-U)^\lambda \right\|_2^2 \\
&\quad + 2M \mathbb{E} \left[\left| \left(\frac{J_n}{n}\right)^\lambda \left(\left(\frac{J_n}{n}\right)^\lambda - (1-U)^\lambda \right) \right| \cdot \Delta(J_n) \right] + \|b'_b(n) - b_b\|_2^2.
\end{aligned}$$

Wegen $I_n \stackrel{d}{=} J_n$ und $U \stackrel{d}{=} 1 - U$ gilt

$$\begin{aligned}\mathbb{E} \left[\left(\frac{I_n}{n} \right)^{2\lambda} \Delta^2(I_n) \right] &= \mathbb{E} \left[\left(\frac{J_n}{n} \right)^{2\lambda} \Delta^2(J_n) \right], \\ \left\| \left(\frac{I_n}{n} \right)^\lambda - U^\lambda \right\|_2^2 &= \left\| \left(\frac{J_n}{n} \right)^\lambda - (1 - U)^\lambda \right\|_2^2\end{aligned}$$

sowie

$$\mathbb{E} \left[\left| \left(\frac{I_n}{n} \right)^\lambda \left(\left(\frac{I_n}{n} \right)^\lambda - U^\lambda \right) \right| \cdot \Delta(I_n) \right] = \mathbb{E} \left[\left| \left(\frac{J_n}{n} \right)^\lambda \left(\left(\frac{J_n}{n} \right)^\lambda - (1 - U)^\lambda \right) \right| \cdot \Delta(J_n) \right].$$

Somit ergibt sich die folgende Abschätzung:

$$\begin{aligned}\Delta_b^2(n) &\leq 2\mathbb{E} \left[\left(\frac{I_n}{n} \right)^{2\lambda} \Delta^2(I_n) \right] + 2M^2 \left\| \left(\frac{I_n}{n} \right)^\lambda - U^\lambda \right\|_2^2 \\ &\quad + 4M\mathbb{E} \left[\left| \left(\frac{I_n}{n} \right)^\lambda \left(\left(\frac{I_n}{n} \right)^\lambda - U^\lambda \right) \right| \cdot \Delta(I_n) \right] \\ &\quad + \|b'_b(n) - b'_b\|_2^2.\end{aligned}$$

Ebenso erhält man eine entsprechende Abschätzung für $\Delta_w(n)$, sodass die obigen Abschätzungen insgesamt die folgende rekursive Darstellung für $\Delta^2(n)$ liefern:

$$\begin{aligned}\Delta^2(n) &\leq 2\mathbb{E} \left[\left(\frac{I_n}{n} \right)^{2\lambda} \Delta^2(I_n) \right] + 2M^2 \left\| \left(\frac{I_n}{n} \right)^\lambda - U^\lambda \right\|_2^2 \\ &\quad + 4M\mathbb{E} \left[\left| \left(\frac{I_n}{n} \right)^\lambda \left(\left(\frac{I_n}{n} \right)^\lambda - U^\lambda \right) \right| \cdot \Delta(I_n) \right] \\ &\quad + \max \left\{ \|b'_b(n) - b'_b\|_2^2, \|b'_w(n) - b'_w\|_2^2 \right\}.\end{aligned}$$

□

Bevor wir den Induktionsbeweis für $\Delta(n)$ durchführen, wollen wir an dieser Stelle die Terme $\left\| \left(\frac{I_n}{n} \right)^\lambda - U^\lambda \right\|_p$ und $\max \left\{ \|b'_b(n) - b'_b\|_p, \|b'_w(n) - b'_w\|_p \right\}$, die mit $p = 2$ in der obigen Rekursion vorkommen, für $p \geq 2$ betrachten, da wir diese auch anschließend in Kapitel 6 antreffen werden.

5.3 Lemma

Für alle $p \geq 2$ und $n \in \mathbb{N}$ gilt:

$$\left\| \left(\frac{I_n}{n} \right)^\lambda - U^\lambda \right\|_p \leq \frac{2^{\frac{1}{p}}}{n^\lambda}.$$

Beweis. Sei $p \geq 2$ und $n \in \mathbb{N}$:

$$\begin{aligned}
\left\| \left(\frac{I_n}{n} \right)^\lambda - U^\lambda \right\|_p^p &= \mathbb{E} \left[\left| \left(\frac{\lfloor nU \rfloor}{n} \right)^\lambda - U^\lambda \right|^p \right] = \int_0^1 \left| \left(\frac{\lfloor nu \rfloor}{n} \right)^\lambda - u^\lambda \right|^p du \\
&= \sum_{j=0}^{n-1} \int_{\frac{j}{n}}^{\frac{j+1}{n}} \left| \left(\frac{j}{n} \right)^\lambda - u^\lambda \right|^p du \\
&= \int_0^{\frac{1}{n}} u^{p\lambda} du + \sum_{j=1}^{n-1} \int_{\frac{j}{n}}^{\frac{j+1}{n}} \left| \left(\frac{j}{n} \right)^\lambda - u^\lambda \right|^p du \\
&= \frac{1}{p\lambda + 1} \frac{1}{n^{p\lambda+1}} + \sum_{j=1}^{n-1} \int_{\frac{j}{n}}^{\frac{j+1}{n}} \left| \left(\frac{j}{n} \right)^\lambda - u^\lambda \right|^p du.
\end{aligned}$$

Für $u \in \left[\frac{j}{n}, \frac{j+1}{n} \right]$, $j = 1, \dots, n-1$ ist die Abbildung $\left[\frac{j}{n}, u \right] \rightarrow \mathbb{R}$, $x \mapsto x^\lambda$ stetig und auf dem offenen Intervall differenzierbar. Mit dem Mittelwertsatz der Differentialrechnung ergibt sich, dass ein $\xi \in \left(\frac{j}{n}, u \right)$ existiert, sodass

$$(5.6) \quad \left| u^\lambda - \left(\frac{j}{n} \right)^\lambda \right| = \left| \lambda \xi^{\lambda-1} \left(u - \frac{j}{n} \right) \right| = \lambda \xi^{\lambda-1} \left(u - \frac{j}{n} \right) \leq \left(\frac{j}{n} \right)^{\lambda-1} \frac{1}{n} = j^{\lambda-1} \frac{1}{n^\lambda}$$

gilt, da $0 < \lambda < 1$ gilt.

Damit gilt weiter

$$\begin{aligned}
\left\| \left(\frac{I_n}{n} \right)^\lambda - U^\lambda \right\|_p^p &\leq \frac{1}{p\lambda + 1} \frac{1}{n^{p\lambda+1}} + \sum_{j=1}^{n-1} \int_{\frac{j}{n}}^{\frac{j+1}{n}} \underbrace{\left| j^{\lambda-1} \frac{1}{n^\lambda} \right|}_{\leq 1}^p du \\
&\leq \frac{1}{p\lambda + 1} \frac{1}{n^{p\lambda+1}} + \sum_{j=1}^{n-1} \int_{\frac{j}{n}}^{\frac{j+1}{n}} \frac{1}{n^{p\lambda}} du = \frac{1}{p\lambda + 1} \frac{1}{n^{p\lambda+1}} + \sum_{j=1}^{n-1} \frac{1}{n^{p\lambda+1}} \\
&\leq \frac{1}{p\lambda + 1} \frac{1}{n^{p\lambda+1}} + \frac{1}{n^{p\lambda}} \stackrel{p\lambda+1 > p\lambda}{\leq} \left(\frac{1}{p\lambda + 1} + 1 \right) \frac{1}{n^{p\lambda}} \stackrel{p\lambda+1 \geq 1}{\leq} \frac{2}{n^{p\lambda}}.
\end{aligned}$$

□

5.4 Lemma

Für alle $p \geq 2$ existiert eine Konstante $0 < A < \infty$, sodass für alle $n \in \mathbb{N}$ gilt:

$$\max \left\{ \|b'_b(n) - b'_b\|_p, \|b'_w(n) - b'_w\|_p \right\} \leq \frac{A}{n^\lambda}.$$

Beweis. Sei $p \geq 2$ und $n \in \mathbb{N}$, dann gilt wegen (2.11) und (2.15) mit einer passenden Konstanten $0 < c < \infty$:

$$\begin{aligned}
& \|b'_b(n) - b'_b\|_p \\
&= \left\| d'_b \left(\left(\frac{I_n}{n} \right)^\lambda - U^\lambda + F_\alpha \left(\left(\frac{J_n}{n} \right)^\lambda - (1-U)^\lambda \right) \right) \right. \\
&\quad \left. + d'_w (1 - F_\alpha) \left(\left(\frac{J_n}{n} \right)^\lambda - (1-U)^\lambda \right) + \frac{c}{n^\lambda} \right\|_p \\
&\leq \left\| d'_b \left(\left(\frac{I_n}{n} \right)^\lambda - U^\lambda \right) \right\|_p + \left\| F_\alpha d'_b \left(\left(\frac{J_n}{n} \right)^\lambda - (1-U)^\lambda \right) \right\|_p \\
&\quad + \left\| (1 - F_\alpha) d'_w \left(\left(\frac{J_n}{n} \right)^\lambda - (1-U)^\lambda \right) \right\|_p + \frac{c}{n^\lambda} \\
&= d'_b \left\| \left(\frac{I_n}{n} \right)^\lambda - U^\lambda \right\|_p + \alpha^{\frac{1}{p}} d'_b \left\| \left(\frac{J_n}{n} \right)^\lambda - (1-U)^\lambda \right\|_p \\
&\quad + (1 - \alpha)^{\frac{1}{p}} |d'_w| \left\| \left(\frac{J_n}{n} \right)^\lambda - (1-U)^\lambda \right\|_p + \frac{c}{n^\lambda} \\
&\leq k_1 \left\| \left(\frac{I_n}{n} \right)^\lambda - U^\lambda \right\|_p + \frac{k_2}{n^\lambda}
\end{aligned}$$

mit passenden Konstanten $0 < k_1, k_2 < \infty$.

Eine analoge Rechnung für $\|b'_w(n) - b'_w\|_p$ mit (2.12) und (2.16) liefert mit passenden Konstanten $0 < K_1, K_2 < \infty$

$$\max \left\{ \|b'_b(n) - b'_b\|_p, \|b'_w(n) - b'_w\|_p \right\} \leq K_1 \left\| \left(\frac{I_n}{n} \right)^\lambda - U^\lambda \right\|_p + \frac{K_2}{n^\lambda}.$$

Mit Lemma 5.3 folgt nun die Behauptung. □

Wir fahren nun mit dem Beweis von Satz 5.1 fort und betrachten die übrigen in Lemma 5.2 auftretenden Terme.

Es gilt

$$\mathbb{E} \left[\left(\frac{I_n}{n} \right)^{2\lambda} \Delta^2(I_n) \right] = \sum_{j=0}^{n-1} \frac{1}{n} \left(\frac{j}{n} \right)^{2\lambda} \Delta^2(j) = \frac{1}{n^{2\lambda+1}} \sum_{j=1}^{n-1} j^{2\lambda} \Delta^2(j)$$

sowie

$$\begin{aligned}
& \mathbb{E} \left[\left| \left(\frac{I_n}{n} \right)^\lambda \left(\left(\frac{I_n}{n} \right)^\lambda - U^\lambda \right) \right| \cdot \Delta(I_n) \right] \\
&= \mathbb{E} \left[\left| \left(\frac{\lfloor nU \rfloor}{n} \right)^\lambda \left(\left(\frac{\lfloor nU \rfloor}{n} \right)^\lambda - U^\lambda \right) \right| \cdot \Delta(\lfloor nU \rfloor) \right]
\end{aligned}$$

$$\begin{aligned}
&= \int_0^1 \left| \left(\frac{\lfloor nu \rfloor}{n} \right)^\lambda \left(\left(\frac{\lfloor nu \rfloor}{n} \right)^\lambda - u^\lambda \right) \right| \cdot \Delta(\lfloor nu \rfloor) du \\
&= \sum_{j=0}^{n-1} \int_{\frac{j}{n}}^{\frac{j+1}{n}} \left| \left(\frac{j}{n} \right)^\lambda \left(\left(\frac{j}{n} \right)^\lambda - u^\lambda \right) \right| \cdot \Delta(j) du \\
&= \sum_{j=1}^{n-1} \left(\frac{j}{n} \right)^\lambda \Delta(j) \int_{\frac{j}{n}}^{\frac{j+1}{n}} \left| \left(\frac{j}{n} \right)^\lambda - u^\lambda \right| du \\
&\stackrel{(5.6)}{\leq} \sum_{j=1}^{n-1} \left(\frac{j}{n} \right)^\lambda \Delta(j) \int_{\frac{j}{n}}^{\frac{j+1}{n}} j^{\lambda-1} \frac{1}{n^\lambda} du \\
&= \sum_{j=1}^{n-1} j^{2\lambda-1} \frac{1}{n^{2\lambda+1}} \Delta(j) = \frac{1}{n^{2\lambda+1}} \sum_{j=1}^{n-1} j^{2\lambda-1} \Delta(j).
\end{aligned}$$

Für den quadrierten Abstand $\Delta^2(n)$ gilt also mit der Rekursion aus Lemma 5.2 und mit den Lemmata 5.3 und 5.4:

$$\begin{aligned}
\Delta^2(n) &\leq \frac{2}{n^{2\lambda+1}} \sum_{j=1}^{n-1} j^{2\lambda} \Delta^2(j) + 2 \cdot M^2 \cdot 2 \frac{1}{n^{2\lambda}} \\
&\quad + 4 \cdot M \frac{1}{n^{2\lambda+1}} \sum_{j=1}^{n-1} j^{2\lambda-1} \Delta(j) + \frac{A}{n^{2\lambda}} \\
&= \frac{2}{n^{2\lambda+1}} \sum_{j=1}^{n-1} j^{2\lambda} \Delta^2(j) + \frac{4M}{n^{2\lambda+1}} \sum_{j=1}^{n-1} j^{2\lambda-1} \Delta(j) + \frac{B}{n^{2\lambda}}
\end{aligned}$$

mit einer passenden Konstanten $0 < B < \infty$.

Sei nun $\varepsilon > 0$. Mit der Induktionsvoraussetzung

$$\Delta^2(j) \leq \frac{C}{j^{2\lambda-1-2\varepsilon}} \quad \text{für } j \leq n-1$$

erhält man im Induktionsschritt $n-1 \rightarrow n$:

$$\Delta^2(n) \leq \frac{2}{n^{2\lambda+1}} \sum_{j=1}^{n-1} j^{2\lambda} \frac{C}{j^{2\lambda-1-2\varepsilon}} + \frac{4M}{n^{2\lambda+1}} \sum_{j=1}^{n-1} j^{2\lambda-1} \frac{\sqrt{C}}{j^{\lambda-\frac{1}{2}-\varepsilon}} + \frac{B}{n^{2\lambda}}.$$

Der erste Summand liefert den folgenden Beitrag:

$$\begin{aligned}
\frac{2C}{n^{2\lambda+1}} \sum_{j=1}^{n-1} j^{1+2\varepsilon} &\stackrel{1+2\varepsilon>0}{\leq} \frac{2C}{n^{2\lambda+1}} \int_1^n x^{1+2\varepsilon} dx = \frac{2C}{n^{2\lambda+1}} \left(\frac{1}{2+2\varepsilon} n^{2+2\varepsilon} - \frac{1}{2+2\varepsilon} \right) \\
&\leq \frac{2}{2+2\varepsilon} \frac{C}{n^{2\lambda-1-2\varepsilon}} = \frac{C}{n^{2\lambda-1-2\varepsilon}} - \frac{2\varepsilon}{2+2\varepsilon} \frac{1}{n^{2\lambda-1-2\varepsilon}}.
\end{aligned}$$

Der zweite Summand liefert:

$$\begin{aligned}
\frac{4M\sqrt{C}}{n^{2\lambda+1}} \sum_{j=1}^{n-1} j^{\lambda-\frac{1}{2}+\varepsilon} &\stackrel{\lambda-\frac{1}{2}+\varepsilon>0}{\leq} \frac{4M\sqrt{C}}{n^{2\lambda+1}} \int_1^n x^{\lambda-\frac{1}{2}+\varepsilon} dx \\
&= \frac{4M\sqrt{C}}{n^{2\lambda+1}} \left(\underbrace{\frac{1}{\lambda+\frac{1}{2}+\varepsilon}}_{<1} n^{\lambda+\frac{1}{2}+\varepsilon} - \underbrace{\frac{1}{\lambda+\frac{1}{2}+\varepsilon}}_{\geq 0} \right) \\
&\leq \frac{4M\sqrt{C}}{n^{2\lambda+1}} n^{\lambda+\frac{1}{2}+\varepsilon} = \frac{4M\sqrt{C}}{n^{\lambda+\frac{1}{2}-\varepsilon}}.
\end{aligned}$$

Dies führt insgesamt zu der folgenden Abschätzung:

$$\Delta^2(n) \leq \frac{C}{n^{2\lambda-1-2\varepsilon}} - \frac{2\varepsilon}{2+2\varepsilon} \frac{1}{n^{2\lambda-1-2\varepsilon}} + \frac{4M\sqrt{C}}{n^{\lambda+\frac{1}{2}-\varepsilon}} + \frac{B}{n^{2\lambda}}.$$

Wegen $2\lambda - 1 - 2\varepsilon < \lambda + \frac{1}{2} - \varepsilon < 2\lambda$ gibt es ein $n_0 \in \mathbb{N}$, sodass für alle $n \geq n_0$

$$-\frac{2\varepsilon}{2+2\varepsilon} \frac{1}{n^{2\lambda-1-2\varepsilon}} + \frac{4M\sqrt{C}}{n^{\lambda+\frac{1}{2}-\varepsilon}} + \frac{B}{n^{2\lambda}} < 0$$

gilt.

Mit einer Konstanten

$$C \geq \max_{1 \leq j \leq n_0} \Delta(j) \cdot j^{\lambda-\frac{1}{2}-\varepsilon}$$

folgt

$$\Delta^2(n) \leq \frac{C^2}{n^{2\lambda-1-2\varepsilon}}$$

und damit folgt die Behauptung. □

6 Konvergenzrate für die Wasserstein-Metrik ℓ_p^\vee mit $p \geq 2$

Unser Ziel ist es nun, die obere Schranke für die Konvergenzrate aus dem vorigen Abschnitt auf die Wasserstein-Metrik ℓ_p^\vee mit $p \geq 2$ zu übertragen.

6.1 Satz: Obere Schranke für die Konvergenzrate in ℓ_p^\vee mit $p \geq 2$

Wir betrachten wie in Satz 5.1 die Pólya-Urne mit zufälliger Rücklegematrix (2.1) mit $\alpha, \beta \in (0, 1)$ und $\lambda := \alpha + \beta - 1$. Es sei $\lambda > \frac{1}{2}$ und die Zufallsvariablen X_n und Y_n mit $n \in \mathbb{N}$ seien wie in (2.7) und (2.8) die zentrierten und skalierten Anzahlen schwarzer Kugeln nach n Schritten. Weiter bezeichne $(\mathcal{L}(\Lambda'_b), \mathcal{L}(\Lambda'_w))$ den eindeutigen Fixpunkt von (2.13) – (2.14) in $\mathcal{M}_2(0) \times \mathcal{M}_2(0)$. Seien $\varepsilon > 0$ und $p \geq 2$. Dann existiert eine Konstante $0 < C_p < \infty$, sodass für alle $n \in \mathbb{N}$ gilt:

$$\ell_p^\vee((X_n, Y_n), (\Lambda'_b, \Lambda'_w)) \leq \frac{C_p}{n^{\lambda - \frac{1}{2} - \varepsilon}}.$$

Das folgende Lemma wird das Grundgerüst für den Beweis von Satz 6.1 sein.

6.2 Lemma

Seien W, X, Y, Z unabhängige reelle Zufallsvariablen, $p \in \mathbb{N}$ und $p \geq 2$. Dann gilt:

$$\begin{aligned} \mathbb{E}[|W + X + Y + Z|^p] &\leq \mathbb{E}[|W|^p] + \mathbb{E}[|X|^p] + \mathbb{E}[|Y|^p] \\ &\quad + \left(\|W\|_{p-1} + \|X\|_{p-1} + \|Y\|_{p-1} + \|Z\|_{p-1} \right)^p. \end{aligned}$$

Beweis. Mit dem Multinomialtheorem, wobei die Summe im 3., 4. und 5. Schritt über $i, j, k, l \in \{0, \dots, p\}$ mit $i + j + k + l = p$ läuft, und wegen der Unabhängigkeit der Zufallsvariablen gilt

$$\begin{aligned} \mathbb{E}[|W + X + Y + Z|^p] &\leq \mathbb{E}[(|W| + |X| + |Y| + |Z|)^p] \\ &= \mathbb{E} \left[\sum_{i,j,k,l} \binom{p}{i,j,k,l} |W|^i |X|^j |Y|^k |Z|^l \right] \\ &= \sum_{i,j,k,l} \binom{p}{i,j,k,l} \mathbb{E}[|W|^i] \mathbb{E}[|X|^j] \mathbb{E}[|Y|^k] \mathbb{E}[|Z|^l] \\ &\leq \mathbb{E}[|W|^p] + \mathbb{E}[|X|^p] + \mathbb{E}[|Y|^p] \\ &\quad + \sum_{i,j,k,l} \binom{p}{i,j,k,l} \|W\|_{p-1}^i \|X\|_{p-1}^j \|Y\|_{p-1}^k \|Z\|_{p-1}^l \\ &= \mathbb{E}[|W|^p] + \mathbb{E}[|X|^p] + \mathbb{E}[|Y|^p] \\ &\quad + \left(\|W\|_{p-1} + \|X\|_{p-1} + \|Y\|_{p-1} + \|Z\|_{p-1} \right)^p, \end{aligned}$$

wobei sich der vorletzte Schritt folgendermaßen ergibt: Für $m = 1, \dots, p-1$ gilt nach (3.1) die folgende Abschätzung $\mathbb{E}[|W|^m] = \|W\|_m^m \leq \|W\|_{p-1}^m$ und ebenfalls für $m = 0$ trivialerweise $\mathbb{E}[|W|^m] \leq \|W\|_{p-1}^m$. Ebenso gilt dies für die Zufallsvariablen X und Y und analog gilt $\mathbb{E}[|Z|^m] \leq \|Z\|_p^m$ für $m = 0, \dots, p$. \square

Beweis von Satz 6.1. Wir werden Satz 6.1 mittels vollständiger Induktion über p und n beweisen. Für $p = 2$ gilt die Aussage mit dem vorigen Abschnitt. Wegen Lemma 3.2 genügt es, ganzzahlige p zu betrachten.

Wir können nun, ähnlich wie zu Beginn des Kapitels 5, wegen Lemma 3.1 die Zufallsvariablen $X_j^{(1)}, X_j^{(2)}, X_j, Y_j^{(1)}, Y_j^{(2)}, Y_j$ für $j = 0, \dots, n-1$ sowie F_α und F_β aus (2.9) und (2.10) und $\Lambda_b^{(1)}, \Lambda_b^{(2)}, \Lambda_w^{(1)}, \Lambda_w^{(2)}$ und U aus (4.3) und (4.4) auf einem gemeinsamen Wahrscheinlichkeitsraum so wählen, dass für $n \in \mathbb{N}$ und $j = 0, \dots, n-1$ gilt:

1. $I_n \stackrel{d}{=} \lfloor nU \rfloor$.
2. Die Paare $(X_j^{(1)}, \Lambda_b^{(1)})$, $(X_j^{(2)}, \Lambda_b^{(2)})$ und $(X_j, \Lambda_b^{(1)})$ sind optimale ℓ_p -Couplings von $\mathcal{L}(X_j)$ und $\mathcal{L}(\Lambda_b')$ für alle $p \geq 2$.
3. Die Paare $(Y_j^{(1)}, \Lambda_w^{(1)})$, $(Y_j^{(2)}, \Lambda_w^{(2)})$ und $(Y_j, \Lambda_w^{(1)})$ sind optimale ℓ_p -Couplings von $\mathcal{L}(Y_j)$ und $\mathcal{L}(\Lambda_w')$ für alle $p \geq 2$.
4. $(X_j^{(1)}, \Lambda_b^{(1)})$, $(X_j^{(2)}, \Lambda_b^{(2)})$, $(X_j, \Lambda_b^{(1)})$, $(Y_j^{(1)}, \Lambda_w^{(1)})$, $(Y_j^{(2)}, \Lambda_w^{(2)})$, $(Y_j, \Lambda_w^{(1)})$, F_α , F_β und U sind unabhängig.

Sei $\varepsilon > 0$. Wir beginnen mit der Induktion nach p :

$$\begin{aligned}
\ell_p(X_n, \Lambda_b') &\leq \|X_n - \Lambda_b'\|_p \\
&= \left\| \left(\left(\frac{I_n}{n} \right)^\lambda X_{I_n}^{(1)} - U^\lambda \Lambda_b^{(1)} \right) + F_\alpha \left(\left(\frac{J_n}{n} \right)^\lambda X_{J_n}^{(2)} - (1-U)^\lambda \Lambda_b^{(2)} \right) \right. \\
&\quad \left. + (1-F_\alpha) \left(\left(\frac{J_n}{n} \right)^\lambda Y_{J_n} - (1-U)^\lambda \Lambda_w^{(1)} \right) + b'_b(n) - b'_b \right\|_p \\
&=: \|V_1^b + V_2^b + V_3^b + V_4^b\|_p, \text{ sowie} \\
\ell_p(Y_n, \Lambda_w') &\leq \|Y_n - \Lambda_w'\|_p \\
&= \left\| \left(\left(\frac{I_n}{n} \right)^\lambda Y_{I_n}^{(1)} - U^\lambda \Lambda_w^{(1)} \right) + F_\beta \left(\left(\frac{J_n}{n} \right)^\lambda Y_{J_n}^{(2)} - (1-U)^\lambda \Lambda_w^{(2)} \right) \right. \\
&\quad \left. + (1-F_\beta) \left(\left(\frac{J_n}{n} \right)^\lambda X_{J_n} - (1-U)^\lambda \Lambda_b^{(1)} \right) + b'_w(n) - b'_w \right\|_p \\
&=: \|V_1^w + V_2^w + V_3^w + V_4^w\|_p
\end{aligned}$$

Bedingt auf $\{U = u\}$ mit $u \in (0, 1)$ sind $V_1^b, V_2^b, V_3^b, V_4^b$ sowie $V_1^w, V_2^w, V_3^w, V_4^w$ unabhängige Zufallsvariablen, sodass sich Lemma 6.2 anwenden lässt.

Dann ergibt sich

$$\begin{aligned}
(6.1) \quad & \mathbb{E} \left[\left| V_1^b + V_2^b + V_3^b + V_4^b \right|^p \middle| U = u \right] \\
& \leq \mathbb{E} \left[\left| V_1^b \right|^p \middle| U = u \right] + \mathbb{E} \left[\left| V_2^b \right|^p \middle| U = u \right] + \mathbb{E} \left[\left| V_3^b \right|^p \middle| U = u \right] \\
& \quad + \left(\left(\mathbb{E} \left[\left| V_1^b \right|^{p-1} \middle| U = u \right] \right)^{\frac{1}{p-1}} + \left(\mathbb{E} \left[\left| V_2^b \right|^{p-1} \middle| U = u \right] \right)^{\frac{1}{p-1}} \right. \\
& \quad \left. + \left(\mathbb{E} \left[\left| V_3^b \right|^{p-1} \middle| U = u \right] \right)^{\frac{1}{p-1}} + \left(\mathbb{E} \left[\left| V_4^b \right|^p \middle| U = u \right] \right)^{\frac{1}{p}} \right)^p.
\end{aligned}$$

Wir schätzen zunächst die einzelnen Summanden aus (6.1) ab. Sei dabei $j := \lfloor nu \rfloor$. Für $q \geq 2$ gilt:

$$\begin{aligned}
\left(\mathbb{E} \left[\left| V_1^b \right|^q \middle| U = u \right] \right)^{\frac{1}{q}} &= \left\| \left(\frac{j}{n} \right)^\lambda X_j^{(1)} - u^\lambda \Lambda_b'^{(1)} \right\|_q \\
&= \left\| \left(\frac{j}{n} \right)^\lambda X_j^{(1)} - \left(\frac{j}{n} \right)^\lambda \Lambda_b'^{(1)} + \left(\frac{j}{n} \right)^\lambda \Lambda_b'^{(1)} - u^\lambda \Lambda_b'^{(1)} \right\|_q \\
&\leq \left\| \left(\frac{j}{n} \right)^\lambda (X_j^{(1)} - \Lambda_b'^{(1)}) \right\|_q + \left\| \Lambda_b'^{(1)} \left(\left(\frac{j}{n} \right)^\lambda - u^\lambda \right) \right\|_q \\
&\leq \begin{cases} \frac{1}{n^\lambda} \|\Lambda_b'\|_q, & \text{falls } u < \frac{1}{n} \\ \left(\frac{j}{n} \right)^\lambda \ell_q(X_j, \Lambda_b') + \frac{j^{\lambda-1}}{n^\lambda} \|\Lambda_b'\|_q, & \text{falls } u \geq \frac{1}{n} \end{cases} \\
&\leq \left(\frac{j}{n} \right)^\lambda \ell_q(X_j, \Lambda_b') + \frac{1}{n^\lambda} \|\Lambda_b'\|_q.
\end{aligned}$$

In der vorletzten Ungleichung wurde ausgenutzt, dass $X_j^{(1)}$ und $\Lambda_b'^{(1)}$ optimale ℓ_p -Couplings für $j = 0, \dots, n-1$ sind. Zusätzlich wurde für den Fall $u \geq \frac{1}{n}$ wieder die Abschätzung aus (5.6) verwendet.

Weiter gilt

$$\begin{aligned}
& \left(\mathbb{E} \left[\left| V_2^b \right|^q \middle| U = u \right] \right)^{\frac{1}{q}} \\
&= \left(\mathbb{E} \left[\left| F_\alpha \left(\left(\frac{n-1-j}{n} \right)^\lambda X_{n-1-j}^{(2)} - (1-u)^\lambda \Lambda_b'^{(2)} \right) \right|^q \right] \right)^{\frac{1}{q}} \\
&= \left(\alpha \cdot \mathbb{E} \left[\left| \left(\left(\frac{n-1-j}{n} \right)^\lambda X_{n-1-j}^{(2)} - (1-u)^\lambda \Lambda_b'^{(2)} \right) \right|^q \right] \right)^{\frac{1}{q}} \\
&= \alpha^{\frac{1}{q}} \left\| \left(\frac{n-1-j}{n} \right)^\lambda X_{n-1-j}^{(2)} - (1-u)^\lambda \Lambda_b'^{(2)} \right\|_q \\
&\leq \alpha^{\frac{1}{q}} \left(\left\| \left(\frac{n-1-j}{n} \right)^\lambda (X_{n-1-j}^{(2)} - \Lambda_b'^{(2)}) \right\|_q \right)
\end{aligned}$$

$$\begin{aligned}
& + \left\| \Lambda_b^{(2)} \left(\left(\frac{n-1-j}{n} \right)^\lambda - (1-u)^\lambda \right) \right\|_q \\
& \leq \begin{cases} \alpha^{\frac{1}{q}} \left(\left(\frac{n-1-j}{n} \right)^\lambda \ell_q(X_{n-1-j}, \Lambda'_b) + \frac{(n-1-j)^{\lambda-1}}{n^\lambda} \|\Lambda'_b\|_q \right), & \text{falls } u < \frac{n-1}{n} \\ \alpha^{\frac{1}{q}} \frac{1}{n^\lambda} \|\Lambda'_b\|_q, & \text{falls } u \geq \frac{n-1}{n} \end{cases} \\
& \leq \alpha^{\frac{1}{q}} \left(\left(\frac{n-1-j}{n} \right)^\lambda \ell_q(X_{n-1-j}, \Lambda'_b) + \frac{1}{n^\lambda} \|\Lambda'_b\|_q \right).
\end{aligned}$$

In der vorletzten Ungleichung wurde wie zuvor ausgenutzt, dass $X_{n-1-j}^{(2)}$ und $\Lambda_b^{(2)}$ optimale ℓ_p -Couplings für $j = 0, \dots, n-1$ sind. Weiter wurde im Fall $u < \frac{n-1}{n}$ die zu (5.6) analoge, folgende Abschätzung für $\left| \left(\frac{n-1-j}{n} \right)^\lambda - (1-u)^\lambda \right|$ verwendet: Für $u \in \left[\frac{j}{n}, \frac{j+1}{n} \right]$ mit $j = 0, \dots, n-2$ ist die Abbildung $\left[u, \frac{j+1}{n} \right] \rightarrow \mathbb{R}$, $x \mapsto (1-x)^\lambda$ stetig und auf dem offenen Intervall differenzierbar. Mit dem Mittelwertsatz der Differentialrechnung ergibt sich nun, dass $\xi \in \left(u, \frac{j+1}{n} \right)$ existiert, sodass folgt:

$$\begin{aligned}
\left| \left(1 - \frac{j+1}{n} \right)^\lambda - (1-u)^\lambda \right| &= \left| -\lambda (1-\xi)^{\lambda-1} \left(\frac{j+1}{n} - u \right) \right| \\
&= \lambda (1-\xi)^{\lambda-1} \left(\frac{j+1}{n} - u \right) \leq \frac{1}{n} \left(1 - \frac{j+1}{n} \right)^{\lambda-1} \\
&= (n-1-j)^{\lambda-1} \frac{1}{n^\lambda},
\end{aligned}$$

da $0 < \lambda < 1$ gilt.

Ebenso erhält man

$$\begin{aligned}
& \left(\mathbb{E} \left[\left| V_3^b \right|^q \middle| U = u \right] \right)^{\frac{1}{q}} \\
& \leq (1-\alpha)^{\frac{1}{q}} \left(\left(\frac{n-1-j}{n} \right)^\lambda \ell_q(Y_{n-1-j}, \Lambda'_w) + \frac{1}{n^\lambda} \|\Lambda'_w\|_q \right).
\end{aligned}$$

Sei im Folgenden $\Delta_q(j) := \ell_q((X_j, Y_j), (\Lambda'_b, \Lambda'_w))$ und $M_q := \max \left\{ \|\Lambda'_b\|_q, \|\Lambda'_w\|_q \right\}$ für $q \geq 2$ und $j \in \mathbb{N}$. Dann lassen sich die obigen Abschätzungen folgendermaßen umschreiben:

$$(6.2) \quad \left(\mathbb{E} \left[\left| V_1^b \right|^q \middle| U = u \right] \right)^{\frac{1}{q}} \leq \left(\frac{j}{n} \right)^\lambda \Delta_q(j) + \frac{1}{n^\lambda} M_q,$$

$$(6.3) \quad \left(\mathbb{E} \left[\left| V_2^b \right|^q \middle| U = u \right] \right)^{\frac{1}{q}} \leq \alpha^{\frac{1}{q}} \left(\left(\frac{n-1-j}{n} \right)^\lambda \Delta_q(n-1-j) + \frac{1}{n^\lambda} M_q \right),$$

$$(6.4) \quad \left(\mathbb{E} \left[\left| V_3^b \right|^q \middle| U = u \right] \right)^{\frac{1}{q}} \leq (1-\alpha)^{\frac{1}{q}} \left(\left(\frac{n-1-j}{n} \right)^\lambda \Delta_q(n-1-j) + \frac{1}{n^\lambda} M_q \right).$$

Analog ergibt sich

$$(6.5) \quad (\mathbb{E}[|V_1^w|^q | U = u])^{\frac{1}{q}} \leq \left(\frac{j}{n}\right)^\lambda \Delta_q(j) + \frac{1}{n^\lambda} M_q,$$

$$(6.6) \quad (\mathbb{E}[|V_2^w|^q | U = u])^{\frac{1}{q}} \leq \beta^{\frac{1}{q}} \left(\left(\frac{n-1-j}{n}\right)^\lambda \Delta_q(n-1-j) + \frac{1}{n^\lambda} M_q \right),$$

$$(6.7) \quad (\mathbb{E}[|V_3^w|^q | U = u])^{\frac{1}{q}} \leq (1-\beta)^{\frac{1}{q}} \left(\left(\frac{n-1-j}{n}\right)^\lambda \Delta_q(n-1-j) + \frac{1}{n^\lambda} M_q \right).$$

Sei nun $q = p-1$, dann ergibt sich mit der Induktionsannahme $\Delta_{p-1}(j) \leq C_{p-1} \frac{1}{j^{\lambda-\frac{1}{2}-\varepsilon}}$ aus (6.2), (6.3), (6.4) und (6.5), (6.6), (6.7):

$$(6.8) \quad \begin{aligned} \left(\mathbb{E} \left[|V_1^b|^{p-1} \mid U = u \right] \right)^{\frac{1}{p-1}} &\leq \left(\frac{j}{n}\right)^\lambda C_{p-1} \frac{1}{j^{\lambda-\frac{1}{2}-\varepsilon}} + \frac{1}{n^\lambda} M_{p-1} \\ &= C_{p-1} \frac{1}{n^\lambda} j^{\frac{1}{2}+\varepsilon} + \frac{1}{n^\lambda} M_{p-1} \\ &\leq C_{p-1} \frac{1}{n^\lambda} n^{\frac{1}{2}+\varepsilon} + \frac{1}{n^\lambda} M_{p-1}, \quad \text{da } 0 \leq j \leq n-1, \\ &= C_{p-1} \frac{1}{n^{\lambda-\frac{1}{2}-\varepsilon}} + \frac{1}{n^\lambda} M_{p-1} \\ &\leq a_1 \frac{1}{n^{\lambda-\frac{1}{2}-\varepsilon}} \end{aligned}$$

mit einer passenden Konstanten $0 < a_1 < \infty$, da $\lambda > \lambda - \frac{1}{2} - \varepsilon$.

Weiter ergibt sich

$$(6.9) \quad \begin{aligned} \left(\mathbb{E} \left[|V_2^b|^{p-1} \mid U = u \right] \right)^{\frac{1}{p-1}} &\leq \alpha^{\frac{1}{p-1}} \left(\left(\frac{n-1-j}{n}\right)^\lambda \frac{C_{p-1}}{(n-1-j)^{\lambda-\frac{1}{2}-\varepsilon}} + \frac{1}{n^\lambda} M_{p-1} \right) \\ &\leq \alpha^{\frac{1}{p-1}} a_2 \frac{1}{n^{\lambda-\frac{1}{2}-\varepsilon}} \end{aligned}$$

sowie

$$(6.10) \quad \left(\mathbb{E} \left[|V_3^b|^{p-1} \mid U = u \right] \right)^{\frac{1}{p-1}} \leq (1-\alpha)^{\frac{1}{p-1}} a_3 \frac{1}{n^{\lambda-\frac{1}{2}-\varepsilon}}$$

mit passenden Konstanten $0 < a_2, a_3 < \infty$, sowie

$$(6.11) \quad \left(\mathbb{E} \left[|V_1^w|^{p-1} \mid U = u \right] \right)^{\frac{1}{p-1}} \leq a_1 \frac{1}{n^{\lambda-\frac{1}{2}-\varepsilon}},$$

$$(6.12) \quad \left(\mathbb{E} \left[|V_2^w|^{p-1} \mid U = u \right] \right)^{\frac{1}{p-1}} \leq \beta^{\frac{1}{p-1}} a_2 \frac{1}{n^{\lambda-\frac{1}{2}-\varepsilon}},$$

$$(6.13) \quad \left(\mathbb{E} \left[|V_3^w|^{p-1} \mid U = u \right] \right)^{\frac{1}{p-1}} \leq (1-\beta)^{\frac{1}{p-1}} a_3 \frac{1}{n^{\lambda-\frac{1}{2}-\varepsilon}}.$$

Bedingt auf $\{U = u\}$ erhalt man mit Lemma 5.4, dass $|V_4^b|, |V_4^w| \leq a_4 \frac{1}{n^\lambda}$ mit $0 < a_4 < \infty$ gilt. Zusammen mit (6.8), (6.9), (6.10) und (6.11), (6.12), (6.13) ergibt sich mit einer passenden Konstanten $0 < a_5 < \infty$

$$(6.14) \quad \left(\mathbb{E} \left[|V_1^b|^{p-1} \middle| U = u \right] \right)^{\frac{1}{p-1}} + \left(\mathbb{E} \left[|V_2^b|^{p-1} \middle| U = u \right] \right)^{\frac{1}{p-1}} \\ + \left(\mathbb{E} \left[|V_3^b|^{p-1} \middle| U = u \right] \right)^{\frac{1}{p-1}} + \left(\mathbb{E} \left[|V_4^b|^p \middle| U = u \right] \right)^{\frac{1}{p}} \leq a_5 \frac{1}{n^{\lambda - \frac{1}{2} - \varepsilon}}$$

sowie

$$(6.15) \quad \left(\mathbb{E} \left[|V_1^w|^{p-1} \middle| U = u \right] \right)^{\frac{1}{p-1}} + \left(\mathbb{E} \left[|V_2^w|^{p-1} \middle| U = u \right] \right)^{\frac{1}{p-1}} \\ + \left(\mathbb{E} \left[|V_3^w|^{p-1} \middle| U = u \right] \right)^{\frac{1}{p-1}} + \left(\mathbb{E} \left[|V_4^w|^p \middle| U = u \right] \right)^{\frac{1}{p}} \leq a_5 \frac{1}{n^{\lambda - \frac{1}{2} - \varepsilon}}.$$

Schlielich konnen wir nun die Abschatzungen (6.2), (6.3) und (6.4) sowie (6.5), (6.6) und (6.7) mit $q = p$ einsetzen und erhalten zusammen mit (6.14) und (6.15) fur (6.1):

$$\mathbb{E} \left[|V_1^b + V_2^b + V_3^b + V_4^b|^p \middle| U = u \right] \\ \leq \left(\left(\frac{j}{n} \right)^\lambda \Delta_p(j) + \frac{1}{n^\lambda} M_p \right)^p \\ + \left(\alpha^{\frac{1}{p}} \left(\left(\frac{n-1-j}{n} \right)^\lambda \Delta_p(n-1-j) + \frac{1}{n^\lambda} M_p \right) \right)^p \\ + \left((1-\alpha)^{\frac{1}{p}} \left(\left(\frac{n-1-j}{n} \right)^\lambda \Delta_p(n-1-j) + \frac{1}{n^\lambda} M_p \right) \right)^p \\ + a_5^p \frac{1}{n^{p(\lambda - \frac{1}{2} - \varepsilon)}} \\ = \left(\left(\frac{j}{n} \right)^\lambda \Delta_p(j) + \frac{1}{n^\lambda} M_p \right)^p + \left(\left(\frac{n-1-j}{n} \right)^\lambda \Delta_p(n-1-j) + \frac{1}{n^\lambda} M_p \right)^p \\ + a_5^p \frac{1}{n^{p(\lambda - \frac{1}{2} - \varepsilon)}}$$

sowie

$$\mathbb{E} \left[|V_1^w + V_2^w + V_3^w + V_4^w|^p \middle| U = u \right] \\ \leq \left(\left(\frac{j}{n} \right)^\lambda \Delta_p(j) + \frac{1}{n^\lambda} M_p \right)^p + \left(\left(\frac{n-1-j}{n} \right)^\lambda \Delta_p(n-1-j) + \frac{1}{n^\lambda} M_p \right)^p \\ + a_5^p \frac{1}{n^{p(\lambda - \frac{1}{2} - \varepsilon)}}$$

Damit folgt

$$\begin{aligned}
(\ell_p(X_n, \Lambda_b))^p &\leq \mathbb{E} \left[\left| V_1^b + V_2^b + V_3^b + V_4^b \right|^p \right] \\
&= \mathbb{E} \left[\mathbb{E} \left[\left| V_1^b + V_2^b + V_3^b + V_4^b \right|^p \middle| U \right] \right] \\
&\leq \sum_{j=0}^{n-1} \frac{1}{n} \left(\left(\left(\frac{j}{n} \right)^\lambda \Delta_p(j) + \frac{1}{n^\lambda} M_p \right)^p \right. \\
&\quad \left. + \left(\left(\frac{n-1-j}{n} \right)^\lambda \Delta_p(n-1-j) + \frac{1}{n^\lambda} M_p \right)^p + a_5^p \frac{1}{n^{p(\lambda-\frac{1}{2}-\varepsilon)}} \right) \\
&= \frac{2}{n} \sum_{j=0}^{n-1} \left(\left(\frac{j}{n} \right)^\lambda \Delta_p(j) + \frac{1}{n^\lambda} M_p \right)^p + a_5^p \frac{1}{n^{p(\lambda-\frac{1}{2}-\varepsilon)}}
\end{aligned}$$

und ebenso

$$(\ell_p(Y_n, \Lambda'_w))^p \leq \frac{2}{n} \sum_{j=0}^{n-1} \left(\left(\frac{j}{n} \right)^\lambda \Delta_p(j) + \frac{1}{n^\lambda} M_p \right)^p + a_5^p \frac{1}{n^{p(\lambda-\frac{1}{2}-\varepsilon)}}.$$

Schließlich liefert eine weitere Induktion, diesmal nach n , mit der Induktionsvoraussetzung

$$\Delta_p(j) \leq C_p \frac{1}{j^{\lambda-\frac{1}{2}-\varepsilon}} \text{ für } 1 \leq j \leq n-1 :$$

$$\begin{aligned}
(\Delta_p(n))^p &\leq \frac{2}{n} \sum_{j=0}^{n-1} \left(\left(\frac{j}{n} \right)^\lambda C_p \frac{1}{j^{\lambda-\frac{1}{2}-\varepsilon}} + \frac{1}{n^\lambda} M_p \right)^p + a_5^p \frac{1}{n^{p(\lambda-\frac{1}{2}-\varepsilon)}} \\
&= \frac{2}{n} \sum_{j=0}^{n-1} \frac{1}{n^{p\lambda}} \left(C_p \cdot j^{\frac{1}{2}+\varepsilon} + M_p \right)^p + a_5^p \frac{1}{n^{p(\lambda-\frac{1}{2}-\varepsilon)}} \\
&= \frac{2}{n^{p\lambda+1}} M_p^p + \frac{2}{n^{p\lambda+1}} \sum_{j=1}^{n-1} \left(C_p \cdot j^{\frac{1}{2}+\varepsilon} + M_p \right)^p + a_5^p \frac{1}{n^{p(\lambda-\frac{1}{2}-\varepsilon)}} \\
&= \frac{2}{n^{p\lambda+1}} M_p^p + \frac{2}{n^{p\lambda+1}} \sum_{j=1}^{n-1} \sum_{k=0}^p \binom{p}{k} C_p^k j^{k(\frac{1}{2}+\varepsilon)} M_p^{p-k} + a_5^p \frac{1}{n^{p(\lambda-\frac{1}{2}-\varepsilon)}} \\
&\leq \frac{2}{n^{p\lambda+1}} M_p^p + \frac{2}{n^{p\lambda+1}} \sum_{j=1}^{n-1} j^{p(\frac{1}{2}+\varepsilon)} \sum_{k=0}^p \binom{p}{k} C_p^k M_p^{p-k} + a_5^p \frac{1}{n^{p(\lambda-\frac{1}{2}-\varepsilon)}} \\
&= \frac{2}{n^{p\lambda+1}} M_p^p + \frac{2}{n^{p\lambda+1}} (C_p + M_p)^p \sum_{j=1}^{n-1} j^{p(\frac{1}{2}+\varepsilon)} + a_5^p \frac{1}{n^{p(\lambda-\frac{1}{2}-\varepsilon)}} \\
&\leq \frac{2}{n^{p\lambda+1}} M_p^p + \frac{2}{n^{p\lambda+1}} (C_p + M_p)^p \int_1^n x^{p(\frac{1}{2}+\varepsilon)} dx + a_5^p \frac{1}{n^{p(\lambda-\frac{1}{2}-\varepsilon)}} \\
&\leq \frac{2}{n^{p\lambda+1}} M_p^p + \frac{2}{n^{p\lambda+1}} (C_p + M_p)^p \frac{n^{p(\frac{1}{2}+\varepsilon)+1}}{p(\frac{1}{2}+\varepsilon)+1} + a_5^p \frac{1}{n^{p(\lambda-\frac{1}{2}-\varepsilon)}}
\end{aligned}$$

$$\begin{aligned}
&= \frac{2}{n^{p\lambda+1}} M_p^p + \frac{2(C_p + M_p)^p}{p\left(\frac{1}{2} + \varepsilon\right) + 1} \frac{1}{n^{p(\lambda - \frac{1}{2} - \varepsilon)}} + a_5^p \frac{1}{n^{p(\lambda - \frac{1}{2} - \varepsilon)}} \\
&\leq \left(\frac{2(C_p + M_p)^p}{p\left(\frac{1}{2} + \varepsilon\right) + 1} + a_6 \right) \frac{1}{n^{p(\lambda - \frac{1}{2} - \varepsilon)}}.
\end{aligned}$$

mit einer passenden Konstanten $0 < a_6 < \infty$. Man beachte, dass die Konstanten a_1, a_2, a_3, a_4, a_5 und a_6 ausschließlich von p abhängen. Nun gilt $p \geq 3$ und somit

$$\frac{2}{p\left(\frac{1}{2} + \varepsilon\right) + 1} \leq \frac{2}{3 \cdot \frac{1}{2} + 1} = \frac{4}{5} < 1$$

und dementsprechend gilt für $0 < C_p < \infty$ ausreichend groß gewählt

$$\frac{2}{p\left(\frac{1}{2} + \varepsilon\right) + 1} (C_p + M_p)^p + a_6 \leq C_p^p.$$

Für solche C_p folgt also $\ell_p^\vee((X_n, Y_n), (\Lambda'_b, \Lambda'_w)) \leq C_p \frac{1}{n^{\lambda - \frac{1}{2} - \varepsilon}}$. Damit sind beide Induktionen abgeschlossen und es folgt die Behauptung. \square

7 Konvergenzrate für die Kolmogorov-Metrik ϱ^\vee

Mit den Erkenntnissen aus dem vorangegangenen Kapitel können wir nun mittels Satz 3.4 auf eine obere Schranke für die Konvergenzrate in der Kolmogorov-Metrik schließen.

7.1 Satz: Obere Schranke für die Konvergenzrate in ϱ^\vee

Wie in Satz 5.1 und Satz 6.1 betrachten wir die Pólya-Urne mit zufälliger Rücklegematrix (2.1) mit $\alpha, \beta \in (0, 1)$ und $\lambda := \alpha + \beta - 1$. Es sei $\lambda > \frac{1}{2}$ und die Zufallsvariablen X_n und Y_n mit $n \in \mathbb{N}$ seien wie in (2.7) und (2.8) die zentrierten und skalierten Anzahlen schwarzer Kugeln nach n Schritten. Weiter bezeichne $(\mathcal{L}(\Lambda'_b), \mathcal{L}(\Lambda'_w))$ den eindeutigen Fixpunkt von (2.13) – (2.14) in $\mathcal{M}_2(0) \times \mathcal{M}_2(0)$. Sei weiter $\varepsilon > 0$. Dann existiert eine Konstante $0 < K < \infty$, sodass für alle $n \in \mathbb{N}$ gilt:

$$\varrho^\vee((X_n, Y_n), (\Lambda'_b, \Lambda'_w)) \leq \frac{K}{n^{\lambda - \frac{1}{2} - \varepsilon}}.$$

Beweis. Die Lebesgue-Dichten von Λ'_b und Λ'_w sind beschränkt, wie man mithilfe der jeweiligen charakteristischen Funktionen und einer Technik von Fill und Janson [5] und Liu [10] zeigen kann: Wir bezeichnen mit φ_b bzw. φ_w die charakteristischen Funktionen von Λ'_b bzw. Λ'_w . Man zeigt nun, dass die charakteristischen Funktionen von Λ'_b und Λ'_w integrierbar sind, indem man folgende Aussage beweist: Für alle $q > 0$ gilt für $t \rightarrow \pm\infty$

$$|\varphi_b(t)| = \mathcal{O}(|t|^{-q}) \text{ sowie } |\varphi_w(t)| = \mathcal{O}(|t|^{-q}).$$

Damit sind φ_b und φ_w integrierbar und die Fourier-Inversionsformel liefert die Existenz und Beschränktheit der Lebesgue-Dichten von Λ'_b und Λ'_w . Die Details hierzu wurden von Kevin Leckey überprüft.

Die Voraussetzungen von Satz 3.4 sind also erfüllt. Wir bezeichnen mit f_b bzw. f_w die Lebesgue-Dichten von Λ'_b bzw. Λ'_w . Sei $D := \max\{\|f_b\|_\infty, \|f_w\|_\infty, 1\}$ und seien weiter $0 < \varepsilon' < \varepsilon$. Der Term $(p+1)^{\frac{1}{p+1}}$ ist durch die eulersche Zahl e nach oben beschränkt. Mit Satz 6.1 existiert zu $p \geq 2$ eine Konstante $0 < C_p < \infty$, sodass für $n \in \mathbb{N}$ gilt:

$$\begin{aligned} \varrho(X_n, \Lambda'_b) &\leq (p+1)^{\frac{1}{p+1}} (D \ell_p(X_n, \Lambda'_b))^{\frac{p}{p+1}} \leq (p+1)^{\frac{1}{p+1}} \left(D \frac{C_p}{n^{\lambda - \frac{1}{2} - \varepsilon'}} \right)^{\frac{p}{p+1}} \\ &\leq e \cdot D \cdot C_p \cdot \frac{1}{n^{\frac{p}{p+1}(\lambda - \frac{1}{2} - \varepsilon')}} = \frac{e \cdot D \cdot C_p}{n^{(1 - \frac{1}{p+1})(\lambda - \frac{1}{2} - \varepsilon')}}. \end{aligned}$$

Ebenso gilt

$$\varrho(Y_n, \Lambda'_w) \leq \frac{e \cdot D \cdot C_p}{n^{(1 - \frac{1}{p+1})(\lambda - \frac{1}{2} - \varepsilon')}}.$$

Man wähle nun p so groß, dass $\lambda - \frac{1}{2} - \varepsilon' - \frac{\lambda - \frac{1}{2} - \varepsilon'}{p+1} > \lambda - \frac{1}{2} - \varepsilon$ gilt. Dann folgt die Behauptung. \square

8 Ausblick

Von diesem Spezialfall ausgehend könnte man nun irreduzible, balancierte Pólya-Urnen mit zwei Farben und einer gegebenen Anfangsbelegung der Urne sowie einer Rücklegematrix der Form

$$R = \begin{pmatrix} a & b \\ c & d \end{pmatrix} \text{ mit } a, d \in \mathbb{N} \cup \{-1\} \text{ und } b, c \in \mathbb{N}_0 \text{ mit } bc > 0 \text{ sowie } a + b = c + d =: K - 1$$

betrachten.

Wieder kann man die in Knappe und Neininger [9, Abschnitt 2] vorgestellte zeitdiskrete Einbettung solcher Urnen in eine kombinatorische Baumstruktur ausnutzen, um zu einem System von Verteilungsrekursionen zu gelangen. Die Anzahl schwarzer Kugeln nach n -maligem Ziehen entspricht dann wieder der Anzahl schwarzer Blätter im zugehörigen Baum nach n Schritten und setzt sich aus den Beiträgen der K Teilbäume der Wurzel zusammen.

In Abhängigkeit von dem Parameter $\lambda := \frac{a-c}{a+b}$ ergeben sich analog zu der hier betrachteten Urne zwei Fälle: Der Fall $\lambda \leq \frac{1}{2}$ führt wieder zu normalverteilten Grenzwerten, der Fall $\lambda > \frac{1}{2}$ liefert nicht-normalverteilte Grenzwerte, vgl. Knappe und Neininger [9, Abschnitt 6.1].

Für die Zukunft wäre es wünschenswert, die hier vorgestellte Vorgehensweise auf solche Pólya-Urnen im Falle nicht-normalverteilter Grenzwerte auszuweiten und im Falle normalverteilter Grenzwerte auf die Zolotarev-Metrik zu übertragen.

Literaturverzeichnis

- [1] KRISHNA B. ATHREYA UND SAMUEL KARLIN: Embedding of urn schemes into continuous time Markov branching processes and related limit theorems. In: *Ann. Math. Statist.* **39** (1968), S. 1801–1817
- [2] ZHI-DONG BAI UND FEIFANG HU: Asymptotic theorems for urn models with nonhomogeneous generating matrices. In: *Stochastic Process. Appl.* **80** (1999), S. 87–101.
- [3] ZHI-DONG BAI, FEIFANG HU UND LI-XIN ZHANG: Gaussian Approximation Theorems for Urn Models and Their Applications. In: *Ann. Appl. Probab.* **12**(4) (2002), S. 1149–1173.
- [4] PETER J. BICKEL UND DAVID A. FREEDMAN: Some Asymptotic Theory for the Bootstrap. In: *Ann. Statist.* **9**(6) (1981), S.1196–1217.
- [5] JAMES ALLEN FILL UND SVANTE JANSON: Smoothness and decay properties of the limiting Quicksort density function. In: D. Gardy und A. Mokkadem (Hg.): *Mathematics and computer science (Versailles, 2000) Trends Math.* Basel: Birkhäuser 2000, S. 53–64.
- [6] JAMES ALLEN FILL UND SVANTE JANSON: Quicksort asymptotics. In: *J. Algorithms* **44** (2002), S. 4–28.
- [7] SVANTE JANSON: Functional limit theorems for multitype branching processes and generalized Pólya urns. In: *Stochastic Process. Appl.* **110**(2) (2004), S. 177–245.
- [8] SVANTE JANSON: Limit theorems for triangular urn schemes. In: *Probab. Theory Related Fields* **134**(3) (2006), S. 417–452.
- [9] MARGARETE KNAPE UND RALPH NEININGER: Pólya urns via the contraction method. Erscheint in *Comb. Probab. Comput.*, online verfügbar unter:
<http://arxiv.org/abs/1301.3404v3> (Stand 18.08.2014) (2013)
- [10] QUANSHENG LIU: Asymptotic properties and absolute continuity of laws stable by random weighted mean. In: *Stochastic Process. Appl.* **95** (2001), S. 83–107.
- [11] PÉTER MAJOR: On the Invariance Principle for Sums of Independent Identically Distributed Random Variables In: *Jour. of Multivariate Anal.* **8** (1978), S. 487–501.
- [12] PETER C. MATTHEWS UND WILLIAM F. ROSENBERGER: Variance in randomized play-the-winner clinical trials. In: *Statistics & Probability Letters* **35** (1997), S. 233–240.
- [13] ROBERT T. SMYTHE: Central limit theorems for urn models. In: *Stochastic Process. Appl.* **65**(1) (1996), S. 115–137.

- [14] ROBERT T. SMYTHE UND WILLIAM F. ROSENBERGER: Play-the-winner designs, generalized Pólya urns, and Markov branching processes. In: N. Flournoy und W. F. Rosenberger (Hg.): *Adaptive designs. IMS Lecture Notes Monograph Ser., Vol. 25* Inst. Math. Statist., Hayward, CA: 1995, S. 13–22.
- [15] L. J. WEI UND S. DURHAM: The randomized play-the-winner-rule in medical trials. In: *J. Amer. Statist. Assoc.* **73** (1978), S. 840–843.
- [16] L. J. WEI, ROBERT T. SMYTHE, D. Y. LIN UND T. S. PARK: Statistical inference with datadependent treatment allocation rules. In: *J. Amer. Statist. Assoc.* **85** (1990), S. 156–162.