



INSTITUT FÜR MATHEMATIK
FACHBEREICH MATHEMATIK
GOETHE UNIVERSITÄT

MASTERARBEIT:

Langreichweitige Seedbank-Koaleszenten

und

fraktionale Brownbewegungen

Jan Lukas Igelbrink

Betreuer:

Prof. Dr. Anton Wakolbinger, Florin Boenkost

23. März 2021

Inhaltsverzeichnis

1	Einleitung	1
2	Modell und Hauptresultate	7
2.1	Das [HS]-Modell	7
2.1.1	Verschmelzungswahrscheinlichkeit	8
2.2	Das [BGKS]-Modell	10
2.2.1	Verschmelzungswahrscheinlichkeit	13
2.3	Resultate zu [HS]	13
2.3.1	Die Asymptotik der Varianz von S_n	14
2.3.2	Funktionale Konvergenz gegen die fraktionale Brownbewegung	14
2.3.3	Couplings für alle Zeithorizonte gemeinsam	15
2.4	Ein Grenzwertsatz für das [BGKS]-Modell	15
2.4.1	Die Asymptotik der Varianz von Z_n	15
2.4.2	Konvergenz der endlichdimensionalen Verteilungen	16
2.4.3	Funktionale Konvergenz gegen die fraktionale Brownbewegung	16
2.4.4	Couplings für alle Zeithorizonte gemeinsam	16
2.4.5	Grenzwert großer Populationen	17
2.5	Ein Kriterium für Konzentration	17
3	Beweise und Diskussionen	19
3.1	Verschmelzungswahrscheinlichkeit in [HS]	19
3.1.1	Asymptotik der Verschmelzungswahrscheinlichkeit mittels Riemannsummen	20
3.1.2	Asymptotik der Verschmelzungswahrscheinlichkeit auf Basis von [HS]	28
3.2	Verschmelzungswahrscheinlichkeit in [BGKS]	28
3.3	Resultate zu [HS]	29
3.3.1	Die Asymptotik der Varianz von S_n	29
3.3.2	Herleitung der Asymptotik der Varianz von S_n aus Satz 2.3 .	33
3.3.3	Funktionale Konvergenz gegen die fraktionale Brownbewegung	36
3.3.4	Diskussion zu [HS, Lemma 4.1]	39

3.3.5	Erste Variation von [HS, Lemma 4.1]	40
3.3.6	Zweite Variation von [HS, Lemma 4.1] und alternativer Beweis der funktionalen Konvergenz	41
3.3.7	Couplings für alle Zeithorizonte gemeinsam über einen Satz von Skorochod	43
3.3.8	Couplings für alle Zeithorizonte gemeinsam über einen Satz von Strassen	45
3.4	Beweis der Grenzwertsätze für das [BGKS]-Modell	47
3.4.1	Die Asymptotik der Varianz von Z_n	47
3.4.2	Konvergenz der endlichdimensionalen Verteilungen	49
3.4.3	Funktionale Konvergenz gegen die fraktionale Brownbewegung	64
3.4.4	Couplings für alle Zeithorizonte gemeinsam	64
3.4.5	Grenzwert großer Populationen	65
3.5	Ein Kriterium für Konzentration	65
4	Ausblick	69
4.1	Das Querlegen einer Generation mit M Individuen	69
4.2	Zufällige Partition von \mathbb{Z} im [HS]-Modell	71
4.3	Skalierungslimes von Anteilsprozessen	73
4.3.1	Der Fall $0 < \alpha < \frac{1}{2}$	73
4.3.2	Der Fall $\frac{1}{2} < \alpha < 1$	73

Kapitel 1

Einleitung

Die vorliegende Arbeit beschäftigt sich mit sogenannten Seedbank-Koaleszenten. Eine erste grundlegende Arbeit dazu von 2001 stellt [KKL] dar. Kay, Krone und Lascoux betrachten in dieser Arbeit ein Modell in dem zu jedem Zeitpunkt $t \in \mathbb{Z}$ genau $N \in \mathbb{N}$ Individuen leben, die genau eine Zeiteinheit überleben. Es sei $m \in \mathbb{N}$ fixiert und μ ein Maß auf den natürlichen Zahlen mit $\mu(\{n \in \mathbb{N} : n > m\}) = 0$. Dann sucht sich jedes Individuum der zum Zeitpunkt t lebenden Population sein Elter mit Wahrscheinlichkeit $\mu(\{i\}), i = 1, \dots, m$, aus der zum Zeitpunkt $t - i$ lebenden Population, innerhalb dieser das Elter dann rein zufällig gewählt wird.

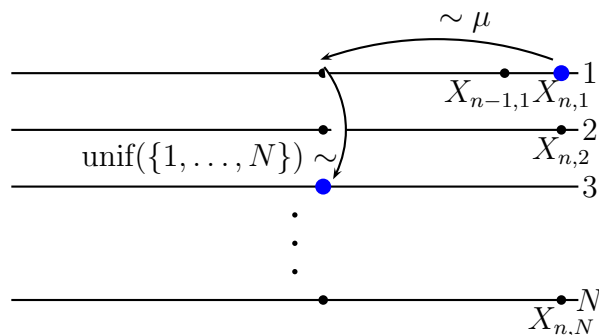


Abbildung 1.1: Aufbau der zufälligen Genealogie

Diese Vorstellung soll hier, wie auch [BGKS] dies taten, auf Maße μ ohne kompakten Träger erweitert werden. Jedes der N Individuen der zu jedem Zeitpunkt t lebenden Populationen, sucht sich also gemäß der Verteilung μ den Zeitpunkt aus zu dem sein Elter lebte und wählt dieses rein zufällig aus der Population der zu diesem Zeitpunkt lebenden Individuen.

Speziell interessieren sich [BGKS] für Maße μ , deren Schwänze nach einem Potenzgesetz abklingen, also für

$$\mu(\{n, n+1, \dots\}) := n^{-\alpha}L(n), \quad \alpha > 0, L \text{ langsam variierend.} \quad (1.1)$$

Im Falle von $\alpha > 1$ bewegt man sich für $\mathcal{L}(R) = \mu$ im Regime

$$\mathbf{E}[R] < \infty$$

und nach [BGKS, Theorem 1] liefert dies für $N \rightarrow \infty$ (mit passender Zeitskalierung) bereits die Konvergenz der Ahnenlinien der zu einem gemeinsamen Zeitpunkt lebenden Individuen gegen den Kingman-Koaleszenten.

Die vorliegende Arbeit widmet sich dem (sehr) langreichweitigen Fall von Maßen μ wie in (2.1) mit $\alpha \in (0, \frac{1}{2})$ in dem für $\mathcal{L}(R) = \mu$

$$\mathbf{E}[R] = \infty$$

und sogar

$$\mathbf{E}\left[R^{\frac{1}{2}}\right] = \infty$$

gelten. Für $\alpha \in (0, \frac{1}{2})$ zeigen [BGKS], dass die Wahrscheinlichkeit, dass zwei Individuen miteinander verwandt sind echt kleiner 1 ist. Für die Verschmelzungswahrscheinlichkeit wird in [BGKS, Lemma 3.1] eine Aussage getroffen, die dort nicht bewiesen wird. In der vorliegenden Arbeit wird der in dieser Aussage enthaltene Vorfaktor korrigiert; der Beweis findet sich in Abschnitt 3.2. Diese Behauptung entspricht der Aussage, dass

$$\mathbf{P}(0 \text{ verwandt mit } i) \sim Ci^{2\alpha-i}L(i)^{-2}, i \rightarrow \infty$$

für eine Konstante $C > 0$ gilt. Tatsächlich gilt bereits für den (nicht sehr) langreichweitigen Fall $\alpha \in (\frac{1}{2}, 1)$, dass Treffzeit der Ahnenlinien zweier verschiedener Individuen zwar fast sicher endlich ist, ihr Erwartungswert aber unendlich ist.

Das [BGKS]-Modell kann als *langreichweitiges Seedbank-Modell* gesehen werden. Im Laufe der Evolution diverser Spezies kann es vorkommen, dass Samen von Pflanzen oder auch Bakterien lange Zeit inaktiv sind und dann schließlich viele Jahre nach Beginn der Inaktivität aus dieser erwachen und sich reproduzieren. Beispielsweise ist bekannt (siehe [S2]), dass eine Dattelpalme aus 2000 Jahre altem Samen wuchs. Die Besonderheit dieser Modelle ist, dass ein Typ, der in der aktuellen Generation nicht zu beobachten ist, aber tatsächlich einmal zu beobachten war, durchaus in zukünftigen Generationen wieder beobachtet werden kann, damit kann folglich keine Fixation eintreten und genetische Typen werden nie vollständig aussterben.

Im gleichen Jahr wie [BGKS] erschien [HS], eine Arbeit auf die sich [BGKS] beziehen und die sich mit dem Spezialfall $N = 1$ befasst. Die durch unsere oben beschriebene Prozedur erzeugte zufällige Genealogie partitioniert im Falle $N = 1$ die ganzen Zahlen \mathbb{Z} zufällig. Die Ahnenlinien der Individuen können als verschmelzende Erneuerungsprozesse mit Inkrementverteilung μ modelliert werden. An dieser Stelle stellt sich bereits die Frage nach der erwarteten Größe des Schnittes einer rein zufällig aus den Komponenten mit Elementen in $\{0, \dots, n\}$ ausgewählten Komponente der zufälligen Partition mit $\{0, \dots, n\}$ für großes $n \in \mathbb{N}$. Im Hinblick auf die Frage nach *Konzentration* stellt sich dann die Frage nach der Varianz dieser Größe. Dies führt unmittelbar zur Frage nach der Asymptotik der Wahrscheinlichkeit, dass drei unabhängig und uniform aus $\{0, \dots, n\}$ herausgegriffene Individuen miteinander verwandt sind, das heißt in der gleichen Komponente liegen. Eine etwas einfachere Frage ist die Frage nach der erwarteten Anzahl an Individuen in $\{1, \dots, n\}$, die mit 0 verwandt sind. Wieder mit der Wahrscheinlichkeit von Dreierverschmelzungen verbunden ist die Berechnung der Varianz dieser Größe. Diese Fragestellungen sind allesamt noch offen und werden nur als Ausblick (siehe Kapitel 4) angerissen.

[HS] widmen sich primär einer zufällig mit Farben ± 1 eingefärbten Version der zufälligen Partition von \mathbb{Z} . Dabei wird jede Komponente mit Wahrscheinlichkeit $p \in (0, 1)$ mit $+1$ und mit Wahrscheinlichkeit $1-p$ mit -1 unabhängig voneinander eingefärbt. Auf Basis dieses gefärbten Graphen bauen [HS] eine Irrfahrt S_n auf, deren Zustand zum Zeitpunkt n genau der Summe der Farben der Individuen $1, \dots, n$ entspricht. Wie sich herausstellt, konvergieren die Reskalierungen

$$\tilde{S}_t^n := \tilde{c}n^{-\frac{1}{2}-\alpha} [S_{nt} - n(2p-1)t], \quad t \geq 0$$

gegen eine fraktionale Brownsche Bewegung mit Hurst-Parameter $H := \frac{1}{2} + \alpha$. Dabei bezeichnet S_t die lineare Interpolation zwischen $S_{\lfloor t \rfloor}$ und $S_{\lceil t \rceil}$. In Abschnitt 2.3.1 korrigieren wir die in [HS] angegebenen Vorfaktoren der Reskalierungen. Im darauf folgenden Abschnitt 3.3.2 leiten wir diese Vorfaktoren noch auf einem anderen Wege her. Diese Konvergenz gilt nicht nur in Form endlichdimensionaler Randverteilungen, sondern auch in Form funktionaler Konvergenz und darüber hinaus existieren Couplings $(C_n)_n$ einer fraktionalen Brownschen Bewegung B^H mit Hurst-Parameter $H := \frac{1}{2} + \alpha$ mit \tilde{S}^n so, dass für alle $T, \varepsilon > 0$

$$\lim_{n \rightarrow \infty} C_n \left(\|\tilde{S}^n - B^H\|_{L^\infty([0, T])} > \varepsilon \right) = 0$$

gilt. In [HS] ist der Beweis der funktionalen Konvergenz lückenhaft, siehe die Diskussion in Abschnitt 3.3.4. In Abschnitt 3.3.3 führen wir einen Beweis der funktionalen Konvergenz über ein Straffheitsargument. In Abschnitt 3.3.6 skizzieren

wir einen alternativen Beweis der funktionalen Konvergenz, der im Herbst 2020 in einer Vorlesung von Prof. Alan Hammond in Berkeley entstand, an der Prof. Wakolbinger und ich auf Einladung von Alan Hammond teilgenommen haben und der in Abschnitt 3.3.3 beschriebenen Beweis vorstellen durften. Der in Abschnitt 3.3.6 beschriebene Beweisweg ist naher an der ursprunglichen Strategie von [HS]. Er ist eine Synthese aus einem Straffheitskriterium (ahnlich dem in Abschnitt 3.3.3 verwendeten Satz 3.11) und einer von REUBEN DROGIN (einem Studenten des Berkeley Kurses) vorgeschlagenen und bewiesenen Abschwachung der problematischen Behauptung (hier als (3.49) zitiert) aus [HS].

Im [BGKS]-Modell definieren wir eine zu S analoge Irrfahrt, deren Inkremente der Mittelwert des Typs einer Generation sind:

$$Z_{n,N} - Z_{n-1,N} := \frac{1}{N} \sum_{k=1}^N X_{n,k}, \quad Z_0 := 0.$$

Wir werden zeigen, dass die endlichdimensionalen Randverteilungen der Reskalierungen $\left(\left(\tilde{Z}_t^{n,N} \right)_{t \geq 0} \right)_{n \in \mathbb{R}_+, N \in \mathbb{N}}$

$$\tilde{Z}_t^{n,N} := \tilde{c}_N n^{-\frac{1}{2} - \alpha} L(n) [Z_{nt,N} - n(2p - 1)t]$$

gegen die einer fraktionalen Brownschen Bewegung mit Hurst-Parameter $H := \frac{1}{2} + \alpha$ konvergieren. Dabei bezeichnet $Z_{t,N}$ wieder die lineare Interpolation zwischen $Z_{[t],N}$ und $Z_{[t+1],N}$. Die Beweisstrategie orientiert sich dabei an der von [HS]. Diese Konvergenz gilt wie fur \tilde{S}_t^n nicht nur in Form endlichdimensionaler Randverteilungen, sondern auch in Form funktionaler Konvergenz und daruber hinaus existieren auch hier fur festes $N \in \mathbb{N}$ Couplings C_n mit

$$\lim_{n \rightarrow \infty} C_n \left(\|\tilde{Z}^{n,N} - B^H\|_{L^\infty([0,T])} > \varepsilon \right) = 0$$

fur alle $T > 0$ und $\varepsilon > 0$. Diese beiden Konvergenzresultate lassen sich ganz analog zu denen fur das HS-Modell einsehen. Schlielich nehmen wir noch den Beweis der Konvergenz endlichdimensionaler Randverteilungen unter die Lupe und erkennen, dass Couplings C_n existieren, sodass fur festes $r > 1$ und fur alle $T > 0$ und $\varepsilon > 0$ im Grenzwert groer Populationen

$$\lim_{N \rightarrow \infty} C_N \left(\|\tilde{Z}^{N^r,N} - B^H\|_{L^\infty([0,T])} > \varepsilon \right) = 0$$

gilt.

In Abschnitt 4.1 beschrieben wir eine Idee des *Querlegens der Generationen*. In aller Kürze gesprochen entspricht diese Idee der Anschauung, im [HS]-Modell genau N Generationen als eine Generation bestehend aus N Individuen im [BGKS]-Modell zu betrachten. Diesen Gedanken haben wir für unsere Beweise allerdings nicht benötigt, auch wenn diese bereits eine gewisse Intuition dafür schaffen können, dass der Übertrag der [HS]-Resultate auf $N > 1$ funktioniert. Tatsächlich liegt es nahe, dass sich über diesen Ansatz eine zweite Beweisstrategie für diese Resultate finden lassen könnte.

Kapitel 2

Modell und Hauptresultate

Das von [HS] eingeführte Modell und dessen natürliche Erweiterung durch [BGKS] wollen wir im Folgenden kurz beschreiben. Zuerst erinnern wir an den Begriff der *langsam variierenden Funktion*:

Definition 2.1. $L : \mathbb{R}_+ \rightarrow \mathbb{R}_+$ heißt *langsam variierend*, falls

$$\lim_{u \rightarrow \infty} \frac{L(u(1+r))}{L(u)} = 1$$

für alle $r > 0$ gilt.

Betrachten wollen wir nun Verteilungen μ auf \mathbb{N} mit schweren regulär variierenden Schwänzen (*heavy-tailed*), das heißt es existiert eine langsam variierende Funktion L und ein $1 > \alpha > 0$ mit

$$\mu(\{n, n+1, \dots\}) = \frac{L(n)}{n^\alpha} \quad \forall n \in \mathbb{N}. \quad (2.1)$$

2.1 Das [HS]-Modell

[HS] betrachten einen zufälligen Graphen G_μ auf \mathbb{Z} , in welchem jeder Knoten $i \in \mathbb{Z}$ genau eine ausgehende Kante in negative Richtung hat, sodass für den Vorfahren V^i von i

$$\mathcal{L}(i - V^i) = \mu$$

gilt. Salopp ausgedrückt sucht sich also das Individuum i seinen Vater V^i indem es eine μ -verteilte Anzahl von Generationen in die Vergangenheit schaut. (An dieser Stelle ist es angebracht, einige Worte zum Begriff des Vaters oder Vorfahren sowie zum Begriff der Generation zu verlieren: Wir wollen im Folgenden ganz unfeingläubig die Begriffe von Mutter und Tochter, Vater und Sohn, Vorfahr und

Nachkomme vollkommen gleichbedeutend verwenden. Von Generationen zu sprechen ist im [HS]-Modell und dessen im nächsten Abschnitt erklärten Erweiterung etwas problematisch, denn die zu einem Zeitpunkt t lebende Population – hier der Größe $N = 1$ – kann ihre direkten Vorfahren zu jedem Zeitpunkt der Vergangenheit haben. Trotzdem wollen wir im Sinne der sprachlichen Einfachheit und auch angelehnt an die Tatsache, dass der Sprachgebrauch sich auch bei Seedbank-Modellen so einbürgerte, die zum Zeitpunkt t lebende Population auch als Generation t bezeichnen.

Jede Komponente des zufälligen Graphen G_μ färben wir nun unabhängig mit Wahrscheinlichkeit p mit $+1$ und mit Wahrscheinlichkeit $1 - p$ mit -1 ein. Aus einem populationsgenetischen Blickwinkel können die *Farben* ± 1 als Typen interpretiert werden. Für $i \in \mathbb{Z}$ bezeichne dann X_i die Farbe der Komponente zu der das Individuum i gehört. An dieser Stelle wollen wir noch eine kurze Definition einschieben:

Definition 2.2. *Wir definieren für $i, j \in \mathbb{Z}$ die Verwandtschaftsrelation*

$$i \sim j :\Leftrightarrow i \text{ und } j \text{ liegen in der gleichen Komponente von } G_\mu$$

und darauf aufbauend den Begriff die Ahnenlinien A_i von Individuen $i \in \mathbb{Z}$ durch

$$A_i := \{j \leq i : j \sim i\}.$$

[HS] zeigen, dass für $\alpha \in (0, 1/2)$ der Graph G_μ fast sicher aus unendlich vielen Komponenten besteht und dementsprechend zwei Individuen mit positiver Wahrscheinlichkeit auch in zwei verschiedenen Komponenten des Graphen liegen, also populationsgenetisch gesprochen *nicht verwandt sind*.

Wir wollen also im Folgenden den Fokus auf ein Maß μ der Form (2.1) für eine langsam variierende Funktion L und ein $\alpha \in (0, 1/2)$ legen. Für den Fall unendlich vieler Komponenten ist die durch

$$S_n - S_{n-1} := X_n, \quad S_0 := 0 \tag{2.2}$$

definierte Irrfahrt $(S_n)_{n \in \mathbb{N}_0}$ offensichtlich nicht trivial. Ein Hauptresultat aus [HS] ist die Konvergenz passender Reskalierungen dieser Irrfahrt gegen eine fraktionale Brownsche Bewegung mit Hurst-Parameter $H = \frac{1}{2} + \alpha$, siehe Satz 2.14. Dem Beweis dieser Aussage ist Abschnitt 3.4 gewidmet. Der Begriff der *Irrfahrt* ist hier weit gefasst, da die Inkremente von S_n offensichtlich nicht unabhängig sind.

2.1.1 Verschmelzungswahrscheinlichkeit

Setzen wir

$$q_n := \mathbf{P}(n \text{ ist Vorfahr von } 0), \quad n \geq 0,$$

so lässt sich für die Verschmelzungswahrscheinlichkeit zweier Ahnenlinien Folgendes zeigen:

Satz 2.3. *Es sei $\alpha \in (0, \frac{1}{2})$ und es gelte*

$$\sup_{n \geq 1} \frac{n \mathbf{P}(R = n)}{\mathbf{P}(R > n)} < \infty \quad \text{für } \mathcal{L}(R) = \mu. \quad (2.3)$$

Für $j, i \in \mathbb{Z}$ mit $j - i = k$ gilt dann

$$\mathbf{P}(A_i \cap A_j \neq \emptyset) \sim \frac{\hat{C}}{L(k)^2} \frac{k^{2\alpha-1}}{\sum_{l \geq 0} q_l^2} \quad (2.4)$$

für $k \rightarrow \infty$ sowie

$$\hat{C} := \frac{1}{2\Gamma(1-\alpha)^2\Gamma(2\alpha)\cos(\pi\alpha)}.$$

Der in (2.4) stehende Ausdruck ist wohldefiniert und größer null, denn nach [HS, Lemma 2.2.(ii)] konvergiert

$$\sum_{l \geq 0} q_l^2$$

für $\alpha \in (0, \frac{1}{2})$. Tatsächlich ist der Beweis von [HS, Lemma 2.2.(ii)] unter der Bedingung (2.3) nicht mehr notwendig, da (2.3) nach [D1] bereits liefert, dass (q_n) regulär variierend mit Exponenten $\alpha - 1$ ist. Der Beweis von Satz 2.3 findet sich in Abschnitt 3.1. [HS] zeigen außerdem das folgende Lemma für $\alpha \in (0, \frac{1}{2})$ betreffend

der Partialsummen $\sum_{k=0}^n q_k$:

Lemma 2.4 ([HS, Lemma 2.2.(i)]). *Für $\alpha \in (0, \frac{1}{2})$ gilt*

$$\sum_{l=0}^n q_l \sim (1-\alpha) \frac{n^{2\alpha} L(n)^{-2}}{\Gamma(2-\alpha)\Gamma(1+\alpha)}, \quad \text{für } n \rightarrow \infty$$

Unter Hinzunahme der Bedingung (2.3) gilt das obige Lemma nach [F, Theorem 5] aufgrund der Asymptotik (3.7).

2.2 Das [BGKS]-Modell

Eine Annahme in vielen Modellen der Populationsgenetik ist die, dass es eine Zahl $N \in \mathbb{N}$ von Individuen pro Generation gibt, die fest über alle Generationen hinweg ist. Diese Annahme treffen auch Blath, González Casanova, Kurt und Spanò in [BGKS].

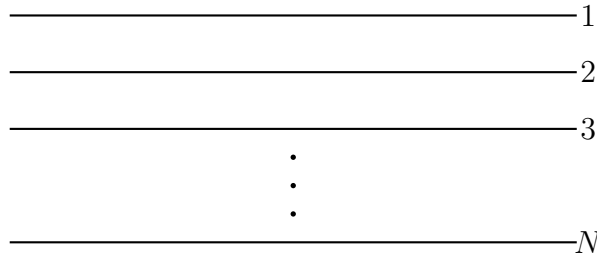


Abbildung 2.1: N Individuen pro Generation

Analog zu [HS] kopiert jedes Individuum seinen Typ ± 1 von seinem Vorfahren.

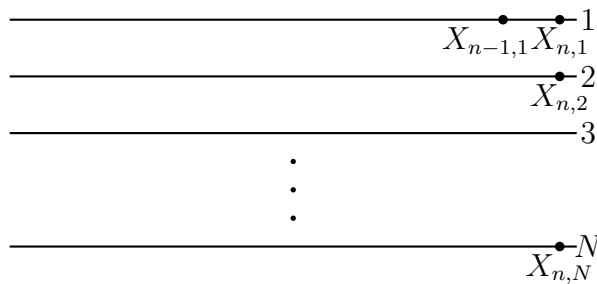


Abbildung 2.2: Jedes Individuum hat Typ ± 1

Zu einem Maß μ auf \mathbb{N} wird jetzt ein zufälliger Graphen G_μ auf $\mathbb{Z} \times \{1, \dots, N\}$ so gesampled, dass für jedes Individuum $(i, j) \in \mathbb{Z} \times \{1, \dots, N\}$ und seinen Vater $V^{(i,j)}$

$$\mathcal{L} \left(i - \left(V^{(i,j)} \right)_x \right) = \mu \quad (2.5)$$

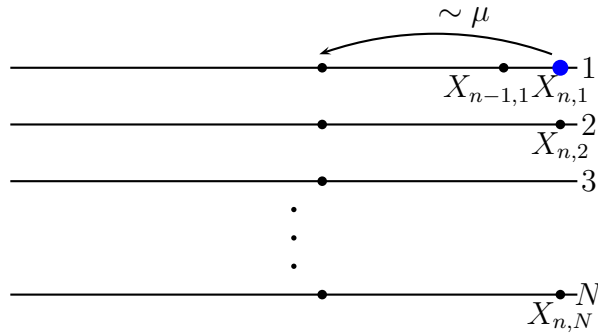


Abbildung 2.3: Die Generation des Elter-Individuums von dem jedes Individuum seinen Typ kopiert, wird gemäß μ gewählt.

und

$$\mathcal{L} \left(\left(V^{(i,j)} \right)_y \right) = \text{unif}(\{1, \dots, N\})$$

gilt. Das Modell lässt sich also informell so beschreiben, dass jedes Individuum

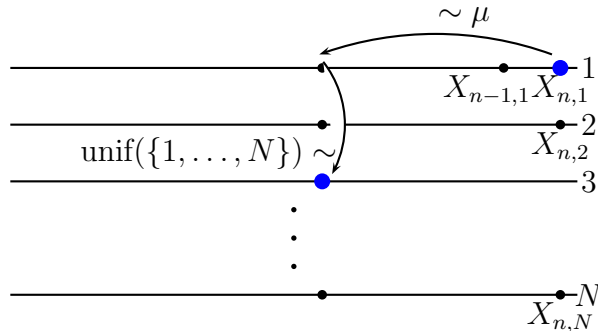


Abbildung 2.4: Innerhalb der ausgewählten Elter-Generation wird das Elter-Individuum uniform gewählt.

$(i, j) \in \mathbb{Z} \times \{1, \dots, N\}$ sich eine Vatergeneration in einer Generation, welche eine μ -verteilte Anzahl von Generationen vor i liegt, aussucht, und aus dieser rein zufällig ein Individuum auswählt.

An dieser Stelle wollen wir kurz an die in 2.1 gemachte Bemerkung erinnern, dass wir die Begriffe von Vater und Sohn, Mutter und Tochter, Nachfahr und Vorfahr vollkommen gleichbedeutend verwenden und dass der Begriff der Generation zwar problematisch aber doch allgemein anerkannt ist. Dementsprechend bezeichnen wir die zum Zeitpunkt t lebende Population auch als Generation t .

[BGKS] zeigen, dass wie in [HS] für ein μ wie in (2.1) zu einem $\alpha \in (0, 1/2)$ und einer langsam variierenden Funktion L der Graph G_μ fast sicher unendlich viele Komponenten hat.

Wie in [HS] wollen wir die Komponenten des zufälligen Graphen unabhängig voneinander mit Wahrscheinlichkeit p mit Typ +1 und mit Wahrscheinlichkeit $1 - p$ mit Typ -1 versehen. $X_{(i,j)}$ bezeichne dann für $(i, j) \in \mathbb{Z} \times \{1, \dots, N\}$ den zufälligen Typ der Komponente in der (i, j) enthalten ist.

Dieses Modell kann als Verallgemeinerung des Standard-Wright-Fisher-Modells, welches $\mu(\{1\}) = 1$ entsprechen würde, gesehen werden, wobei bei diesem nur solche Individuen als verwandt gelten, deren MRCA (*Most recent common ancestor*) zur Zeit $t \geq 0$ lebte. (Anderenfalls wären mit Wahrscheinlichkeit 1 alle Individuen miteinander verwandt.) Der Prozess $(W_{n,N})_{n \in \mathbb{N}}$

$$W_{n,N} := \frac{1}{N} \sum_{i=1}^N \mathbf{1}_{X_{n,i}=1} \quad (2.6)$$

ist dann eine Markovkette deren Skalierungslimes gegen die Wright-Fisher-Diffusion W konvergiert, den Diffusionsprozess, der Lösung der stochastischen Differentialgleichung

$$W_t = w + \int_0^t \sqrt{W_s(1 - W_s)} dB_t$$

für eine Standard Brownsche Bewegung B und $w \in (0, 1)$ ist.

Für allgemeineres μ der Form (2.1) mit $\alpha \in (0, \frac{1}{2})$ existieren auch dann mit Wahrscheinlichkeit 1 unendlich viele *Familien*, wenn man alle Individuen, die überhaupt einen MRCA besitzen, als verwandt klassifiziert. Diese kann man mit unabhängigen $p \pm 1$ -Münzwürfen einfärben. Die Typenfrequenzen

$$\mathcal{W}_{n,N} := \frac{1}{N} \sum_{i=1}^N \mathbf{1}_{X_{n,i}=1}$$

haben dann langreichweitige Korrelationen. Betrachtet man deren Summenprozess $(Z_n)_{n \in \mathbb{Z}}$ mit Inkrementen

$$Z_n - Z_{n-1} := Z_{n,N} - Z_{n-1,N} := \frac{1}{N} \sum_{i=1}^N X_{n,i}, \quad Z_0 = 0, \quad (2.7)$$

dann konvergiert dessen Skalierungslimes gegen eine fraktionale Brownsche Bewegung mit Hurst-Parameter $H := \frac{1}{2} + \alpha$. Dabei bezeichnet Z_t die lineare Interpolation zwischen $Z_{\lfloor t \rfloor}$ und $Z_{\lceil t \rceil}$.

2.2.1 Verschmelzungswahrscheinlichkeit

Die Ideen von Satz 2.3 liefern uns auch für die Verschmelzungswahrscheinlichkeit zweier um n Generationen voneinander entfernter Individuen im [BGKS]-Modell eine Asymptotik für $n \rightarrow \infty$. Dafür wollen wir noch kurz eine Definition einschleiben:

Definition 2.5. Wir definieren für $i, j \in \mathbb{Z}, l, m \in \{1, \dots, N\}$ die Verwandtschaftsrelation

$$(i, l) \sim (j, m) :\Leftrightarrow (i, l) \text{ und } (j, m) \text{ liegen in der gleichen Komponente von } G_\mu$$

und darauf aufbauend den Begriff die Ahnenlinien A_i von Individuen $i \in \mathbb{Z}$ durch

$$A_{(i,l)} := \{(j, m) : j \leq i, m \in \{1, \dots, N\}, (j, m) \sim (i, l)\}.$$

Satz 2.6. Es sei $\alpha \in (0, \frac{1}{2})$ und es gelte

$$\sup_{n \geq 1} \frac{n \mathbf{P}(R = n)}{\mathbf{P}(R > n)} < \infty \quad \text{für } \mathcal{L}(R) = \mu.$$

Für $k, l \in \{1, \dots, N\}$ sowie $i, j \in \mathbb{Z}$ mit $j - i = n$ gilt für die Verschmelzungswahrscheinlichkeit der Ahnenlinien $A_{i,k}$ und $A_{j,l}$ der Individuen (i, k) und (j, l)

$$\begin{aligned} \mathbf{P}(A_{i,k} \cap A_{j,l} \neq \emptyset) &= \frac{\sum_{m \geq 0} q_m q_{m+n}}{N + \sum_{m \geq 1} q_m^2} \\ &\sim \frac{n^{2\alpha-1}}{N + \sum_{m \geq 1} q_m^2} \cdot \frac{\hat{C}}{L(n)^2} \end{aligned}$$

für $j - i =: n \rightarrow \infty$ sowie

$$\hat{C} := \frac{1}{2\Gamma(1-\alpha)^2\Gamma(2\alpha)\cos(\pi\alpha)}.$$

Der Beweis findet sich in Abschnitt 3.2.

2.3 Resultate zu [HS]

In diesem Abschnitt wollen wir das Hauptresultat von [HS] für den Fall $\alpha \in (0, 1/2)$ und ein μ der Form (2.1), namentlich die Konvergenz passender Reskalierungen von $(S_n)_{n \in \mathbb{N}_0}$ gegen eine fraktionale Brownsche Bewegung, vorstellen und später den in [HS] dazu geführten Beweis in gewissem Sinne reparieren.

2.3.1 Die Asymptotik der Varianz von S_n

Um die passende Reskalierung zu erhalten, berechnen [HS] die asymptotische Varianz der durch (2.2) definierten Irrfahrt. Wir geben dieses Resultat im Folgenden in leicht korrigierter Form wieder (siehe [HS, Lemma 3.1] für das ursprüngliche Resultat)

Lemma 2.7. $(R^i)_{i \in \mathbb{N}}$ sei eine Familie unabhängig identisch verteilter Zufallsvariablen mit Verteilung μ und

$$R^\mu := \left\{ \sum_{i=1}^n R^i : n \in \mathbb{N} \right\} \quad (2.8)$$

der dazugehörige Erneuerungsprozess. Mit der bereits in Abschnitt 2.1.1 eingeführten Notation

$$q_0 := 1 \text{ und } q_n := \mathbf{P}(n \in R_\mu) \text{ für } n \in \mathbb{N}$$

gilt für die durch (2.2) definierte Irrfahrt $(S_n)_{n \in \mathbb{N}_0}$

$$\mathbf{Var}[S_n] \sim \frac{4p(1-p)}{2\alpha(2\alpha+1)} \frac{n^{2\alpha+1}}{\sum_{l \geq 0} q_l^2} \frac{1}{L(n)^2 \Gamma(1-\alpha)^2 \Gamma(2\alpha) \cos(\pi\alpha)} \quad (2.9)$$

2.3.2 Funktionale Konvergenz gegen die fraktionale Brownbewegung

Zur Vereinfachung der Notation setzen wir

$$\tilde{c}^{-2} := \frac{4p(1-p)}{2\alpha(2\alpha+1)} \frac{1}{\sum_{l \geq 0} q_l^2 \Gamma(1-\alpha)^2 \Gamma(2\alpha) \cos(\pi\alpha)}$$

und erhalten damit aus (2.9)

$$\mathbf{Var}[S_n] \sim \tilde{c}^{-2} n^{2\alpha+1} L(n)^{-2}.$$

[HS] zeigen zuerst, dass die Folge der reskalierten durch

$$\tilde{S}_t^n := \tilde{c} n^{-\frac{1}{2}-\alpha} L(n) [S_{nt} - n(2p-1)t]$$

definierten Prozesse im Sinne der endlichdimensionalen Randverteilungen gegen eine fraktionale Brownsche Bewegung B^H mit Hurst-Parameter $H = \alpha + \frac{1}{2}$ konvergiert. Dazu geben wir hier das entsprechende Resultat aus [HS] wieder:

Proposition 2.8 ([HS, Proposition 3]). *Im Sinne der endlichdimensionalen Randverteilungen konvergiert die Folge $\left(\tilde{S}_t^n\right)_n$ von Prozessen für $n \rightarrow \infty$ gegen eine fraktionale Brownsche Bewegung mit Hurst-Parameter $H := \alpha + \frac{1}{2}$.*

Diese Aussage erweitern wir in Proposition 2.12 auf den Fall $N \geq 1$ und führen deren Beweis in Abschnitt 3.4.2. Basierend auf Proposition 2.8 zeigen wir die folgende Erweiterung im Sinne funktionaler Konvergenz:

Proposition 2.9. *Sei $T > 0$. Bezogen auf den Pfadraum $(C([0, T]), \|\cdot\|_{L^\infty([0, T])})$ konvergiert \tilde{S}^n in Verteilung gegen B^H für eine fraktionale Brownsche Bewegung B^H mit Hurst-Parameter $H = \frac{1}{2} + \alpha$.*

Die Aussage wird in Abschnitt 3.3.3 bewiesen.

2.3.3 Couplings für alle Zeithorizonte gemeinsam

Im Folgenden sei weiterhin $H := \alpha + \frac{1}{2}$. Ziel ist es mit Proposition 2.9 den folgenden Satz zu beweisen:

Satz 2.10 ([HS, Theorem 1.1]). *Für alle $p \in (0, 1)$ existiert eine Folge von Couplings $(C_n)_n$ der Prozesse \tilde{S}^n und B^H , sodass für alle $T > 0$ und $\varepsilon > 0$*

$$\lim_{n \rightarrow \infty} C_n \left(\|\tilde{S}^n - B^H\|_{L^\infty([0, T])} > \varepsilon \right) = 0$$

gilt.

Damit wird also die Existenz einer Folge von Couplings behauptet, die nicht vom Zeithorizont T abhängt. Diese Gemeinsamkeit über die Zeithorizonte wird in [HS] nicht bewiesen. Wir beweisen sie in Abschnitt 3.3.7. In Abschnitt 3.3.8 stellen wir einen dazu alternativen Beweisweg vor.

2.4 Ein Grenzwertsatz für das [BGKS]-Modell

Wir wählen $N \in \mathbb{N}$ fest und erinnern an

$$Z_n - Z_{n-1} := Z_{n,N} - Z_{n-1,N} := \frac{1}{N} \sum_{k=1}^N X_{n,k}, \quad Z_0 := 0. \quad (2.10)$$

Dazwischen interpolieren wir wieder linear.

2.4.1 Die Asymptotik der Varianz von Z_n

Zuerst werden wir wieder die Asymptotik der Varianz unserer Irrfahrt bestimmen. Wir zeigen in Abschnitt 3.4.1, dass die Asymptotik der Varianz von Z_n der gleichen Ordnung ist wie die von S_n :

Lemma 2.11. *Für die Varianz von Z_n gilt*

$$\mathbf{Var}[Z_n] = \mathbf{Var}[S_n] \cdot \frac{\sum_{l \geq 0} q_l^2}{\left(\sum_{l \geq 1} q_l^2 + N\right)}$$

für die in (2.2) definierte Irrfahrt S_n .

2.4.2 Konvergenz der endlichdimensionalen Verteilungen

Wir werden in Abschnitt 3.4.2 basierend auf Lemma 2.11 das folgende Analogon zu [HS, Proposition 3] beweisen:

Proposition 2.12. *Für festes $N \in \mathbb{N}$ konvergieren die Reskalierungen $\left(\left(\tilde{Z}_t^n\right)_{t \geq 0}\right)_{n \in \mathbb{R}_+}$*

$$\tilde{Z}_t^n := \tilde{Z}_t^{n,N} := \left(\frac{\sum_{l \geq 0} q_l^2}{\left(\sum_{l \geq 1} q_l^2 + N\right)}\right)^{-\frac{1}{2}} \cdot \tilde{c}n^{-\frac{1}{2}-\alpha} L(n) [Z_{nt} - n(2p-1)t]$$

im Sinne der endlichdimensionalen Randverteilungen gegen eine fraktionale Brownsche Bewegung B^H mit Hurst-Parameter $H := \frac{1}{2} + \alpha$.

2.4.3 Funktionale Konvergenz gegen die fraktionale Brownbewegung

Wie auch im [HS]-Modell erweitern wir nun die Konvergenz im Sinne funktionaler Konvergenz:

Proposition 2.13. *Sei $T \geq 0$ und $N \in \mathbb{N}$ fest. Bezogen auf den Pfadraum $(C([0, T]), \|\cdot\|_{L^\infty([0, T])})$ konvergiert \tilde{Z}^n in Verteilung gegen eine fraktionale Brownsche Bewegung B^H mit Hurst-Parameter $H := \frac{1}{2} + \alpha$.*

Der Beweis wird in Abschnitt 3.4.3

2.4.4 Couplings für alle Zeithorizonte gemeinsam

Mit obigem lässt sich dann das folgende Resultat analog zu Satz 2.10 gewinnen:

Satz 2.14. *Für alle $\alpha \in (0, 1/2)$ und $p \in (0, 1)$ existiert eine fraktionale Brownsche Bewegung B^H mit Hurst-Parameter $H := \frac{1}{2} + \alpha$ und eine Folge von Couplings C_n von \tilde{Z}^n und B^H so dass für alle $T > 0$ und $\varepsilon > 0$ und $N \in \mathbb{N}$ fest*

$$\lim_{n \rightarrow \infty} C_n \left(\|\tilde{Z}^n - B^H\|_{L^\infty([0, T])} > \varepsilon \right) = 0$$

gilt.

2.4.5 Grenzwert großer Populationen

Wir werden auch die folgende Verallgemeinerung zeigen bei der die Zeit mit der Populationsgröße skaliert:

Satz 2.15. *Für alle $\alpha \in (0, 1/2)$, $r > 1$ und $p \in (0, 1)$ existiert eine fraktionale Brownsche Bewegung B^H mit Hurst-Parameter $H := \frac{1}{2} + \alpha$ und eine Folge von Couplings C_N von $\tilde{Z}^{N^r, N}$ und B^H so dass für alle $T > 0$ und $\varepsilon > 0$*

$$\lim_{N \rightarrow \infty} C_N \left(\|\tilde{Z}^{N^r, N} - B^H\|_{L^\infty([0, T])} > \varepsilon \right) = 0$$

gilt.

Den Beweis erbringen wir in Abschnitt 3.4.5

2.5 Ein Kriterium für Konzentration

Das folgende Kriterium für Konzentration, welches [HS] im Beweis der Konvergenz endlichdimensionaler Randverteilungen implizit verwenden, und von uns zum Beweis dieser Konvergenz im [BGKS]-Modell benötigt wird, wollen wir hier noch kurz vorstellen und später im Abschnitt 3.5 beweisen:

Lemma 2.16. *Es sei $(a_n)_{n \in \mathbb{N}}$ eine Folge natürlicher Zahlen mit $a_n \rightarrow \infty$, $(\varepsilon_l)_{l \in \mathbb{N}}$ eine Nullfolge und $(M_n)_{n \in \mathbb{N}}$ eine Folge von identisch verteilten Zufallsvariablen mit $\mathbf{E}[M_n] \neq 0$ und $\mathbf{Var}[M_n] < \infty$, sodass*

$$\mathbf{Cov}[M_n, M_{n+l}] = \varepsilon_l \text{ für alle } n, l \in \mathbb{N} \text{ und eine Nullfolge } (\varepsilon_l)_l$$

gilt. Für positive Gewichte $(c_m^n)_{m, n \in \mathbb{N}}$ mit

$$\frac{\max_{1 \leq m \leq a_n} c_m^n}{\sum_{m=1}^{a_n} c_m^n} \rightarrow 0 \text{ für } n \rightarrow \infty$$

gilt

$$\mathbf{Var} \left[\sum_{m=1}^{a_n} c_m^n M_m \right] = o \left(\left(\mathbf{E} \left[\sum_{m=1}^{a_n} c_m^n M_m \right] \right)^2 \right).$$

Das Lemma kann als Korollar des folgenden Satzes gesehen werden:

Satz 2.17. *Es sei $(a_n)_{n \in \mathbb{N}}$ eine Folge natürlicher Zahlen mit $a_n \rightarrow \infty$, $(\varepsilon_l)_{l \in \mathbb{N}}$ eine Nullfolge und $(M_n)_{n \in \mathbb{N}}$ eine Folge von identisch verteilten Zufallsvariablen mit $\mathbf{Var}[M_n] < \infty$, sodass*

$$\mathbf{Cov}[M_n, M_{n+l}] = \varepsilon_l \text{ für alle } n, l \in \mathbb{N} \text{ und eine Nullfolge } (\varepsilon_l)_l$$

gilt. Für Verteilungen π^n auf $\{1, \dots, a_n\}$ mit

$$\max_{1 \leq m \leq a_n} \pi_m^n \rightarrow 0 \text{ für } n \rightarrow \infty$$

gilt

$$\mathbf{Var} \left[\sum_{m=1}^{a_n} \pi_m^n M_m \right] \rightarrow 0 \text{ für } n \rightarrow \infty.$$

Der Beweis wird ebenso in Abschnitt 3.5 erbracht.

Kapitel 3

Beweise und Diskussionen

3.1 Verschmelzungswahrscheinlichkeit in [HS]

Diese Abschnitt soll die Verschmelzungswahrscheinlichkeit zweier Ahnenlinien analysieren. [BGKS] haben eine asymptotische Aussage gemacht, die die Entfernung zweier Individuen mit der Verschmelzungswahrscheinlichkeit ihrer Ahnenlinien in Verbindung bringt. Die behauptete Aussage stand unbewiesen da und wurde in Satz 2.6 korrigiert. Im Folgenden erbringen wir den Beweis.

Es sei wie gehabt und oben beschrieben $\alpha \in \left(0, \frac{1}{2}\right)$ und $R^1, R^2 \dots$ eine Folge unabhängig identisch verteilter Zufallsvariablen mit

$$\mathbf{P}(R_1 \geq n) \sim n^{-\alpha} L(n) \tag{3.1}$$

für eine langsam variierende Funktion L . Wir behalten die Notation

$$q_n := q(n) := \mathbf{P}(\exists i \geq 0 : R_1 + \dots + R_i = n)$$

aus [HS] bei, d.h. q ist die zu μ zugehörige *Erneuerungsfunktion*.

Proposition 3.1. *Setzen wir für $l \in \mathbb{N}$*

$$(|Q|^2)_l := \sum_{k \geq 0} q_k q_{k+l}$$

und seien $j, i \in \mathbb{Z}$ mit $j \geq i$, dann gilt für die Ahnenlinien A_i und A_j von i und j und ihre Verschmelzungswahrscheinlichkeit

$$\mathbf{P}(|A_i \cap A_j| \neq \emptyset) = \frac{(|Q|^2)_{j-i}}{(|Q|^2)_0}. \tag{3.2}$$

Die Notation $(|Q|^2)_l$ stammt vom engen Zusammenhang zwischen den Koeffizienten der Fourierreihe des Betragsquadrates von

$$Q(t) := \sum_{n \geq 0} q_n e^{itn}.$$

Beweis. \tilde{A}_i und \tilde{A}_j seien zwei unabhängige Erneuerungsprozesse startend in i und j mit negativen Inkrementen mit Verteilung μ . Das heißt \tilde{A}_i und \tilde{A}_j lassen sich auch als Couplings der Ahnenlinien A_i und A_j verstehen, welche nicht bei einem gemeinsamen Vorfahren verschmelzen. Für $j > i$ gilt

$$\mathbf{E} \left[|\tilde{A}_i \cap \tilde{A}_j| \right] = \sum_{k \geq 0} q_k q_{k+j-i} = (|Q|^2)_{j-i} \quad (3.3)$$

sowie

$$\begin{aligned} \mathbf{E} \left[\tilde{A}_i \cap \tilde{A}_j \right] &= \mathbf{P} \left(\tilde{A}_i \cap \tilde{A}_j \neq \emptyset \right) \underbrace{\mathbf{E} \left[|\tilde{A}_i \cap \tilde{A}_j| \mid \tilde{A}_i \cap \tilde{A}_j \neq \emptyset \right]}_{= \sum_{k \geq 0} q_k^2 = (|Q|^2)_0} \\ &= \mathbf{P} \left(\tilde{A}_i \cap \tilde{A}_j \neq \emptyset \right) \cdot (|Q|^2)_0. \end{aligned} \quad (3.4)$$

Damit folgt unmittelbar aus (3.3) und (3.4)

$$\mathbf{P} \left(\tilde{A}_i \cap \tilde{A}_j \neq \emptyset \right) = \frac{(|Q|^2)_{j-i}}{(|Q|^2)_0}.$$

Da nun

$$\mathbf{P} \left(\tilde{A}_i \cap \tilde{A}_j \neq \emptyset \right) = \mathbf{P} \left(A_i \cap A_j \neq \emptyset \right)$$

gilt, folgt die behauptete Gleichheit

$$\mathbf{P} \left(|A_i \cap A_j| \neq \emptyset \right) = \frac{(|Q|^2)_{j-i}}{(|Q|^2)_0}.$$

□

3.1.1 Asymptotik der Verschmelzungswahrscheinlichkeit mittels Riemannsummen

[BGKS, Lemma 3.c] besagt, dass für

$$\sum_{n=0}^{\infty} q_n q_{n-i} = \sum_{j=1}^{\infty} q_j q_{j+i} = (|Q|^2)_i \sim \frac{(1-\alpha)^2}{\Gamma(2-\alpha)^2 \Gamma(2\alpha)} i^{2\alpha-1} L(i) \quad \text{für } i \rightarrow \infty \quad (3.5)$$

gilt. Die Autoren verweisen für einen Beweis auf [HS], insbesondere auf das nicht existente Lemma 5.1 in diesem Artikel. Da der Beweis in diesem Artikel nicht auffindbar ist, beweisen wir hier eine ähnliche Aussage, um die Asymptotik der rechten Seite in (3.2) zu beschreiben und damit asymptotische Aussagen zur Treffwahrscheinlichkeit zweier um i voneinander entfernter Individuen zu machen.

[D1, Theorem B] besagt, dass (3.1) mit der Zusatzannahme

$$\sup_{n \geq 1} \frac{n \mathbf{P}(R = n)}{\mathbf{P}(R > n)} < \infty \quad \text{für } \mathcal{L}(R) = \mu \quad (3.6)$$

die asymptotische Beziehung

$$q_n \sim C n^{\alpha-1} \tilde{L}(n) \quad (3.7)$$

impliziert, wobei nach [GJ]

$$C := \frac{\sin(\pi\alpha)}{\pi} = \frac{1}{\Gamma(1-\alpha)\Gamma(\alpha)} \quad (3.8)$$

und

$$\tilde{L}(n) := \frac{1}{L(n)} \quad (3.9)$$

gesetzt wurde. (Tatsächlich ist die Konstante (3.8) in [D1] falsch angegeben worden. Die zweite Gleichheit in (3.8) gilt nach dem Eulerschen Reflexionsprinzip.)

An dieser Stelle lohnt es sich festzustellen, dass die langsam variierende Funktion in (3.5) nicht unbedingt die gleiche langsam variierende Funktion wie in (3.1) meint und hier schlicht in [BGKS] von Zeile zu Zeile L eine andere langsam variierende Funktion bezeichnet. Wir zeigen im Folgenden, dass für die langsam variierende Funktion $(L(i))^{-2}$ auf der rechten Seite in (3.5) die Behauptung mit einem anderen von α abhängigen Vorfaktor gilt:

Proposition 3.2. *Es sei $\alpha \in \left(0, \frac{1}{2}\right)$ und Annahme (3.7) sei erfüllt. Dann gilt*

$$\sum_{j \geq 1} q_j q_{i+j} \sim C^2 \tilde{L}(i)^2 i^{2\alpha-1} \int_0^\infty (1+x)^{\alpha-1} x^{\alpha-1} dx \quad \text{für } i \rightarrow \infty$$

für die gleiche Konstante C wie in (3.7).

Wir werden den Beweis dieser Proposition zuerst unter einer Annahme an die Karamata-Darstellung (siehe dazu [F, Korollar, p. 282] oder [BGT, Theorem 1.3.1] und insbesondere [BGT, (1.3.1')]) der langsam variierenden Funktion L erbringen und dann auf den allgemeinen Fall erweitern. Wir zeigen also zuerst das folgende Lemma:

Lemma 3.3. *Es sei $\alpha \in \left(0, \frac{1}{2}\right)$ und es gelte*

$$r_n = n^{\alpha-1}K(n) \quad (3.10)$$

für $K(n)$ von der Form

$$K(n) = \exp\left(\int_B^n \frac{l(t)}{t} dt\right) \quad (3.11)$$

für eine positive Konstante B sowie eine beschränkte messbare Funktion $l(t), t \geq B$ mit

$$\lim_{t \rightarrow \infty} l(t) = 0.$$

Damit ist $(r_n)_n$ schließlich monoton fallend und es gilt für $i \rightarrow \infty$

$$\sum_{j \geq 1} r_j r_{i+j} \sim K(i)^2 i^{2\alpha-1} \int_0^\infty (1+x)^{\alpha-1} x^{\alpha-1} dx. \quad (3.12)$$

Beweis. Wir zeigen zuerst, dass $(r_n)_n$ schließlich monoton fallend ist, was aus (3.10) zusammen mit der Karamata Darstellung (3.11) von $K(n)$ der langsam variierenden Funktion folgt: Im Sinne der Übersichtlichkeit setzen wir $\beta := 1 - \alpha$. Da $l(t)$ für $t \rightarrow \infty$ gegen 0 konvergiert, existiert $n_0 \in \mathbb{N}$ so, dass

$$l(t) < \beta \quad \text{für } t \geq n_0$$

gilt. Damit folgt

$$\begin{aligned} \frac{r_n}{r_{n+1}} &= \frac{n^{\alpha-1}}{(n+1)^{\alpha-1}} \cdot \frac{K(n)}{K(n+1)} \\ &= \frac{n^{\alpha-1}}{(n+1)^{\alpha-1}} \exp\left(-\int_n^{n+1} \frac{l(t)}{t} dt\right) \\ &= \exp\left(\beta (\ln(n+1) - \ln(n)) - \int_n^{n+1} \frac{l(t)}{t} dt\right) \\ &= \exp\left(\int_n^{n+1} \underbrace{\left(\frac{\beta}{t} - \frac{l(t)}{t}\right)}_{>0 \text{ für } t \geq n_0} dt\right) \\ &> 1 \end{aligned}$$

Damit ist $(r_n^{-1})_n$ für $n \geq n_0$ monoton wachsend und entsprechend $(r_n)_n$ für $n \geq n_0$ monoton fallend.

Wir wollen uns der in (3.12) beschriebenen Asymptotik widmen, welche mit (3.10) für $i \rightarrow \infty$ äquivalent zu

$$\frac{1}{i} \sum_{j \geq 1} \frac{r_j r_{i+j}}{r_i r_i} \rightarrow \int_0^\infty (1+x)^{\alpha-1} x^{\alpha-1} dx \quad (3.13)$$

ist. Ziel ist also jetzt, (3.13) zu beweisen. Dafür zeigen wir, dass für alle $\varepsilon > 0$ ein $N \in \mathbb{N}$ existiert, so dass für alle i groß genug

$$\frac{1}{i} \sum_{j \geq Ni} \frac{r_j r_{i+j}}{r_i r_i} < \varepsilon \quad (3.14)$$

gilt.

Nun ist jedoch $(r_n)_n$ schließlich monoton fallend, d.h. (3.14) folgt bereits, wenn ein $N \in \mathbb{N}$ existiert, sodass für alle i groß genug

$$\frac{1}{i} \sum_{j \geq Ni} \left(\frac{r_j}{r_i} \right)^2 < \varepsilon \quad (3.15)$$

gilt. Die Tatsache, dass $(r_n)_n$ schließlich monoton ist, liefert nun, dass die linke Seite von (3.15) für i groß genug von oben durch

$$\sum_{m=N}^\infty \left(\frac{r_{mi}}{r_i} \right)^2 = \sum_{m=N}^\infty m^{2\alpha-2} \left(\frac{K(mi)}{K(i)} \right)^2 \quad (3.16)$$

beschränkt ist.

Mit (3.11) folgt, dass für alle $\delta > 0$ und i so groß, dass $l(t) < \delta$ für alle $t \geq i$ gilt,

$$\frac{K(mi)}{K(i)} = \exp \left(\int_i^{mi} \frac{l(t)}{t} dt \right) \leq \exp \left(\delta (\ln(mi) - \ln(i)) \right) \leq m^\delta,$$

was mit dem Satz von der majorisierten Konvergenz für N groß genug liefert, dass die rechte Seite von (3.16) für i groß genug durch ε beschränkt wird. Damit folgt also (3.14).

Wir zeigen als Nächstes, dass für alle $\varepsilon > 0$ ein $\eta \in \mathbb{N}$ existiert, so dass für alle i groß genug

$$\frac{1}{i} \sum_{j \leq \eta i} \frac{r_j r_{i+j}}{r_i r_i} < \varepsilon \quad (3.17)$$

gilt. Dass $(r_n)_n$ schließlich monoton fallend ist, liefert

$$\frac{r_{i+j}}{r_i} \leq 1$$

für i groß genug. Es bleibt also nur zu zeigen, dass für alle $\varepsilon > 0$ ein $\eta \in \mathbb{N}$ existiert, sodass für alle i groß genug

$$\frac{1}{i} \sum_{j \leq \eta i} \frac{r_j}{r_i} < \varepsilon \quad (3.18)$$

gilt. [F, Theorem 5, p. 447] liefert

$$\sum_{j \leq \eta i} r_j \sim \frac{1}{\alpha} (\eta i)^\alpha K(\lfloor \eta i \rfloor),$$

und somit

$$\frac{1}{i} \sum_{j \leq \eta i} \frac{r_j}{r_i} \sim \frac{1}{\alpha} \eta^\alpha \frac{K(\lfloor \eta i \rfloor)}{K(i)} \sim \frac{\eta^\alpha}{\alpha}. \quad (3.19)$$

Damit ist (3.18) und somit auch (3.17) bewiesen, wobei die letzte Asymptotik in (3.19) folgt, da K langsam variierend ist.

Im Lichte von (3.10), (3.14) und (3.17) genügt für den Beweis von (3.12) nun der Beweis der folgenden Konvergenz:

$$\frac{1}{i} \sum_{\eta i \leq j \leq N i} \frac{K(j)}{K(i)} \frac{K(i+j)}{K(i)} \left(1 + \frac{j}{i}\right)^{\alpha-1} \left(\frac{j}{i}\right)^{\alpha-1} \rightarrow \int_{\eta}^N (1+x)^{\alpha-1} x^{\alpha-1} dx. \quad (3.20)$$

Aus (3.11) lässt sich

$$\lim_{i \rightarrow \infty} \sup_{\eta i \leq j \leq (N+1)i} \left| \frac{K(j)}{K(i)} - 1 \right| = 0$$

folgern. Damit reduziert sich (3.20) auf die Konvergenz von Riemann-Summen gegen das zugehörige Integral, welche wir in Proposition 3.4 zeigen. \square

Beweis von Proposition 3.2. Wir wollen nun den Beweis der Proposition komplementieren. (3.7) können wir auch als

$$q_n = C_n n^{\alpha-1} \tilde{L}(n)$$

für eine Folge von Zahlen $(C_n)_{n \in \mathbb{N}}$ mit $C_n \rightarrow C > 0$ schreiben. Wie im Beweis des vorigen Lemmas genügt es auch hier

$$\frac{1}{i} \sum_{j \geq 1} \frac{q_j}{q_i} \frac{q_{i+j}}{q_i} \rightarrow \int_0^\infty (1+x)^{\alpha-1} x^{\alpha-1} dx \quad (3.21)$$

zu zeigen. [BGT, Theorem 1.3.1] liefert die Existenz eines $K(n)$, welches (3.11) erfüllt sowie die Existenz einer Folge $(D_n)_n$ mit

$$\lim_{n \rightarrow \infty} D_n = D \in \mathbb{R},$$

sodass

$$\tilde{L}(n) = D_n K(n), \quad n = 1, 2, \dots$$

gilt. Definieren wir nun r_n wie in (3.10), so haben wir die Darstellung

$$q_n = D_n C_n r_n, \quad n = 1, 2, \dots$$

für $q_n, n \in \mathbb{N}$.

Allerdings hängt weder die Asymptotik der linken Seite von (3.13) noch die Asymptotik der linken Seite von (3.21) von einer endlichen festen Anzahl von Summanden ab, weshalb (3.21) gilt. \square

Wir liefern nun noch die behauptete Konvergenz der Riemann-Summen nach:

Proposition 3.4. *Wir setzen*

$$b_i := \frac{1}{i} \sum_{j=1}^{\infty} \left(\frac{j}{i} + 1 \right)^{\alpha-1} \left(\frac{j}{i} \right)^{\alpha-1}, \quad i = 1, 2, \dots$$

Für $0 < \alpha < \frac{1}{2}$ gilt dann

$$b_i \xrightarrow{i \rightarrow \infty} J := \int_0^{\infty} (1+x)^{\alpha-1} x^{\alpha-1} dx < \infty.$$

Beweis. Zu vorgegebenem ε sei $K := K_\varepsilon \in \mathbb{N}$ so groß, dass

$$\int_K^{\infty} x^{2(\alpha-1)} dx < \varepsilon \tag{3.22}$$

gilt. Daraus folgt für alle $i \in \mathbb{N}$

$$\frac{1}{i} \sum_{j=Ki}^{\infty} \left(\frac{j}{i} + 1 \right)^{\alpha-1} \left(\frac{j}{i} \right)^{\alpha-1} < \varepsilon.$$

Andererseits gilt die folgende Konvergenz der Riemann-Summen:

$$\tilde{b}_i := \frac{1}{i} \sum_{1 \leq j < Ki} \left(\frac{j}{i} + 1 \right)^{\alpha-1} \left(\frac{j}{i} \right)^{\alpha-1} \xrightarrow{i \rightarrow \infty} \int_0^K (1+x)^{\alpha-1} x^{\alpha-1} dx =: J_K.$$

Sei nun i_ε so groß, dass für alle $i \geq i_\varepsilon$

$$|b_i - J_K| \leq \varepsilon \quad (3.23)$$

gilt. Daraus folgt für alle $i \geq i_\varepsilon$

$$|b_i - J| \leq \underbrace{|b_i - J_K|}_{\substack{(3.23) \\ \leq \varepsilon}} + \underbrace{|J_K - J|}_{\substack{(3.22) \\ < \varepsilon}} < 2\varepsilon.$$

□

Proposition 3.1 und Proposition 3.2 liefern uns zusammen mit den folgenden beiden Lemmata nun ohne weiteren Beweis die folgende Asymptotik für die Verschmelzungswahrscheinlichkeit zweier um k voneinander entfernter Individuen j und i :

Satz 2.3. *Es sei $\alpha \in (0, \frac{1}{2})$ und es gelte*

$$\sup_{n \geq 1} \frac{n \mathbf{P}(R = n)}{\mathbf{P}(R > n)} < \infty \quad \text{für } \mathcal{L}(R) = \mu.$$

Für $j, i \in \mathbb{Z}$ mit $j - i = k$ gilt dann

$$\mathbf{P}(A_i \cap A_j \neq \emptyset) \sim \frac{\hat{C}}{L(k)^2} \frac{k^{2\alpha-1}}{\sum_{l \geq 0} q_l^2}$$

für $k \rightarrow \infty$ sowie

$$\hat{C} := \frac{1}{2\Gamma(1-\alpha)^2\Gamma(2\alpha)\cos(\pi\alpha)}.$$

Lemma 3.5. *Es gilt für $\alpha \in (0, \frac{1}{2})$*

$$\int_0^\infty x^{\alpha-1}(1+x)^{\alpha-1} dx = \frac{\Gamma(\alpha)\Gamma(1-2\alpha)}{\Gamma(1-\alpha)}$$

Beweis. Die Beta-Funktion ist durch

$$B(a, b) := \int_0^1 x^{a-1}(1-x)^{b-1} dx$$

definiert. Substituieren wir $x = \frac{t}{1+t}$, $t = \frac{x}{1-x}$, $\frac{dt}{dx} = \frac{1}{(1-x)^2}$, dann liefert das

$$\int_0^\infty \frac{t^{b-1}}{(1+t)^{a+b}} dt \quad (3.24)$$

$$= \int_0^\infty \left(\frac{t}{1+t}\right)^{b-1} \left(1 - \frac{t}{1+t}\right)^{a+1} dx \quad (3.25)$$

$$= \int_0^1 x^{b-1} (1-x)^{a+1} \frac{1}{(1-x)^2} dx \quad (3.26)$$

$$= \int_0^1 x^{b-1} (1-x)^{a-1} dx \quad (3.27)$$

$$= B(a, b), \quad (3.28)$$

also

$$B(a, b) = \int_0^\infty \frac{t^{b-1}}{(1+t)^{a+b}} dt. \quad (3.29)$$

Damit folgt

$$\int_0^\infty x^{\alpha-1} (1+x)^{\alpha-1} dx = \int_0^\infty \frac{x^{\alpha-1}}{(1+x)^{1-\alpha}} \stackrel{(3.29)}{=} B(\alpha, 1-2\alpha).$$

Das liefert also:

$$\int_0^\infty x^{\alpha-1} (1+x)^{\alpha-1} dx = B(\alpha, 1-2\alpha) = \frac{\Gamma(\alpha)\Gamma(1-2\alpha)}{\Gamma(1-\alpha)}.$$

□

Lemma 3.6. Für die in Satz 2.3 definierte Konstante \hat{C} gilt

$$\frac{\int_0^\infty x^{\alpha-1} (1+x)^{\alpha-1} dx}{(\Gamma(1-\alpha)\Gamma(\alpha))^2} = \frac{1}{2\Gamma(1-\alpha)^2\Gamma(2\alpha)\cos(\pi\alpha)}.$$

Beweis. Mit dem vorigen Lemma folgt mit dem Eulerschen Reflexionsprinzip und bekannten Additionstheoremen

$$\begin{aligned} & \frac{\int_0^\infty x^{\alpha-1} (1+x)^{\alpha-1} dx}{(\Gamma(1-\alpha)\Gamma(\alpha))^2} \\ &= \left[\frac{1}{\Gamma(\alpha)\Gamma(1-\alpha)} \right]^2 \int_0^\infty x^{\alpha-1} (1+x)^{\alpha-1} dx \\ &= \frac{\Gamma(1-2\alpha)}{\Gamma(\alpha)\Gamma(1-\alpha)^3} \\ &= \frac{1}{2\Gamma(1-\alpha)^2\Gamma(2\alpha)} \cdot \underbrace{2 \frac{\Gamma(2\alpha)\Gamma(1-2\alpha)}{\Gamma(\alpha)\Gamma(1-\alpha)}}_{= \frac{\sin(\alpha\pi)}{\sin(2\alpha\pi)^{\frac{1}{2}}} = \frac{1}{\cos(\pi\alpha)}} \\ &= \frac{1}{2\Gamma(1-\alpha)^2\Gamma(2\alpha)\cos(\pi\alpha)}. \end{aligned}$$

□

3.1.2 Asymptotik der Verschmelzungswahrscheinlichkeit auf Basis von [HS]

Mit der in Abschnitt 3.3.1 im Hinblick auf Vorfaktoren korrigierten Version von [HS, Formelzeile 1, p.702] (siehe (3.38)) gilt

$$\sum_{k=1}^n \sum_{l \geq 0} q_l q_{l+k} \sim \frac{1}{2} \frac{n^{2\alpha} L(n)^{-2}}{2\alpha \Gamma(1-\alpha)^2 \Gamma(2\alpha) \cos(\pi\alpha)}, \quad n \rightarrow \infty.$$

Angenommen nun, dass

$$(q_k^*)_k := \left(\sum_{l \geq 0} q_l q_{l+k} \right)_{k \geq 0}$$

monoton in k ist, dann liefert [F, Theorem 5]

$$\sum_{l \geq 0} q_l q_{l+n} \sim \frac{1}{2} \frac{n^{2\alpha-1} L(n)^{-2}}{\Gamma(1-\alpha)^2 \Gamma(2\alpha) \cos(\pi\alpha)}, \quad n \rightarrow \infty.$$

Dies entspricht mit Lemma 3.6 der Aussage von Proposition 3.2. Bedingung (3.6) ist dann nicht mehr notwendig. Die Monotonie von $(q_k^*)_k$ wäre bereits gegeben, forderte man Monotonie von $(q_k)_{k \geq 0}$. Diese Bedingung ist allerdings stärker als (3.6). Nach [F, Exercise 16, p. 464] wäre eine ebenso hinreichende Bedingung, dass für eine schließlich monotone Folge $(\nu_n)_n$

$$\sum_{l \geq 0} q_l q_{l+n} \sim \nu(n), \quad n \rightarrow \infty$$

gilt.

3.2 Verschmelzungswahrscheinlichkeit in [BGKS]

Wir wollen nun Satz 2.6 zur Verschmelzungswahrscheinlichkeit im [BGKS]-Modell beweisen:

Satz 2.6. *Es sei $\alpha \in (0, \frac{1}{2})$ und es gelte*

$$\sup_{n \geq 1} \frac{n \mathbf{P}(R = n)}{\mathbf{P}(R > n)} < \infty \quad \text{für } \mathcal{L}(R) = \mu$$

Für $k, l \in \{1, \dots, N\}$ sowie $i, j \in \mathbb{Z}$ mit $j - i = n$ gilt für die Verschmelzungswahrscheinlichkeit der Ahnenlinien $A_{i,k}$ und $A_{j,l}$ der Individuen (i, k) und (j, l)

$$\begin{aligned} \mathbf{P}(A_{i,k} \cap A_{j,l} \neq \emptyset) &= \frac{\sum_{m \geq 0} q_m q_{m+n}}{N + \sum_{m \geq 1} q_m^2} \\ &\sim \frac{n^{2\alpha-1}}{N + \sum_{m \geq 1} q_m^2} \cdot \frac{\hat{C}}{L(n)^2} \end{aligned}$$

für $j - i =: n \rightarrow \infty$ sowie

$$\hat{C} := \frac{1}{2\Gamma(1-\alpha)^2\Gamma(2\alpha)\cos(\pi\alpha)}.$$

Beweis. $\tilde{A}_{i,k}$ und $\tilde{A}_{j,l}$ Couplings der Ahnenlinien $A_{i,k}$ und $A_{j,l}$, welche nicht bei einem gemeinsamen Vorfahren verschmelzen. Für $j > i$ gilt

$$\mathbf{E}[|\tilde{A}_{i,k} \cap \tilde{A}_{j,l}|] = \sum_{k \geq 0} \frac{1}{N} q_k q_{k+j-i} = \frac{1}{N} (|Q|^2)_{j-i}$$

sowie

$$\begin{aligned} \mathbf{E}[|\tilde{A}_{i,k} \cap \tilde{A}_{j,l}|] &= \mathbf{P}(\tilde{A}_{i,k} \cap \tilde{A}_{j,l} \neq \emptyset) \\ &\quad \cdot \underbrace{\mathbf{E}[|\tilde{A}_{i,k} \cap \tilde{A}_{j,l}| \mid \tilde{A}_{i,k} \cap \tilde{A}_{j,l} \neq \emptyset]}_{=1 + \sum_{k \geq 1} \frac{1}{N} q_k^2} \\ &= \mathbf{P}(\tilde{A}_{i,k} \cap \tilde{A}_{j,l} \neq \emptyset) \cdot \left(1 + \sum_{k \geq 1} \frac{1}{N} q_k^2\right). \end{aligned} \quad (3.30)$$

und damit die behauptete Gleichheit

$$\mathbf{P}(|A_{i,k} \cap A_{j,l}| \neq \emptyset) = \frac{(|Q|^2)_{j-i}}{N + \sum_{k \geq 1} q_k^2} \quad (3.31)$$

und damit mit Proposition 3.2 auch die behauptete Asymptotik. \square

3.3 Resultate zu [HS]

3.3.1 Die Asymptotik der Varianz von S_n

Im Folgenden reproduzieren wir den Beweis von [HS, Lemma 3.1] und korrigieren dabei einen kleinen Rechenfehler bei der Herleitung des konstanten Faktors in

der Asymptotik der Varianz von S_n . Dafür setzen wir im Stil der Notation von Lemma 2.7

$$Q(t) := \sum_{n=0}^{\infty} q_n e^{itn}. \quad (3.32)$$

Wir zeigen zuerst das folgende auf dem Beweis von [HS, Lemma 2.2] beruhende Lemma:

Lemma 3.7. Für $Q(t)$ und $\alpha \in (0, \frac{1}{2})$ wie gehabt (3.32) gilt

$$|Q(t)|^2 \sim \Gamma(1 - \alpha)^{-2} t^{-2\alpha} L(1/t)^{-2}, \quad t \downarrow 0. \quad (3.33)$$

Beweis. Wir setzen $p_k := \mu(\{k\})$, $k \geq 0$ und

$$P(t) := \sum_{n=0}^{\infty} p_n e^{itn}.$$

[GL, Theorem 1] liefert, dass für $\alpha \in (0, \frac{1}{2})$ (tatsächlich gilt dies sogar für $\alpha \in (0, 1)$)

$$1 - \Re P(t) \sim \mu((t^{-1}, \infty)) \Gamma(1 - \alpha) \cos\left(\frac{\alpha\pi}{2}\right), \quad t \downarrow 0$$

sowie

$$\Im P(t) \sim \tan\left(\frac{\pi\alpha}{2}\right) (1 - \Re P(t)), \quad t \downarrow 0$$

gilt. Damit folgt offenbar mit $\mu((t^{-1}, \infty)) \sim t^\alpha L(t^{-1})$

$$|1 - P(t)| \sim \Gamma(1 - \alpha) t^\alpha L(t^{-1}), \quad t \downarrow 0. \quad (3.34)$$

Zerlegung nach dem ersten Sprung liefert für $n > 0$

$$q_n = \sum_{k=0}^n p_k q_{n-k}.$$

Da nun $q_0 = 1$ und $\sum_{k=0}^n p_k q_{0-k} = 0$ gilt, folgt

$$q_n = \sum_{k=0}^n p_k q_{n-k} + \delta_0(n), \quad \forall n \in \mathbb{N}_0.$$

Für die Fouriertransformierte $Q(t) = \mathfrak{F}_q$ von $q := (q_n)_{n \geq 0}$ gilt damit für $p := (p_n)_{n \geq 0}$

$$Q(t) = \mathfrak{F}_q(t)$$

$$\begin{aligned}
 &= \sum_{n \geq 0} \left(\sum_{k=0^n} p_k q_{n-k} + \delta_0(n) \right) e^{itn} \\
 &= 1 + \mathfrak{F}_{q * p}(t) \\
 &= 1 + \mathfrak{F}_q(t) \mathfrak{F}_p(t) \\
 &= 1 + Q(t)P(t).
 \end{aligned}$$

Damit folgt

$$Q(t) = (1 - P(t))^{-1}$$

und zusammen mit (3.34)

$$|Q(t)| \sim \frac{t^{-\alpha} L\left(\frac{1}{t}\right)^{-1}}{\Gamma(1 - \alpha)}$$

und durch Quadrieren die Behauptung. \square

$$\begin{aligned}
 |Q(t)|^2 &= Q(t)\overline{Q}(t) \\
 &= \sum_{n \geq 0} \sum_{m \geq 0} q_n q_m e^{itn} e^{-itm} \\
 &= \sum_{l=-\infty}^{\infty} \underbrace{e^{itl}}_{\cos(tl) + i \sin(tl)} \sum_{n \geq 0} q_n q_{n+l} \\
 &= \sum_{l=-\infty}^{\infty} \cos(tl) \sum_{n \geq 0} q_n q_{n+l} \tag{3.35} \\
 &= 2 \sum_{l=0}^{\infty} \cos(tl) \sum_{n \geq 0} q_n q_{n+l} - \sum_{n \geq 0} q_n^2.
 \end{aligned}$$

Wir bezeichnen im Stile von [HS] mit $(|Q|^2)_i$ den i -ten Fourierkoeffizienten der Fourierreihe $|Q|^2$ (siehe (3.35)). Damit gehen wir an den Beweis ganz analog zu [HS] heran:

Beweis von Lemma 2.7. Zunächst gilt für $a := 2p - 1 = \mathbf{E}[X_i]$

$$\begin{aligned}
 \mathbf{Var}[S_n] &= \mathbf{Var} \left[\sum_{i=1}^n X_i \right] \\
 &= \sum_{i=1}^n \sum_{j=1}^n \mathbf{Cov}[X_i, X_j]
 \end{aligned}$$

$$\begin{aligned}
 &= \sum_{i=1}^n \sum_{j=1}^n \mathbf{E} [(X_i - a)(X_j - a)] \\
 &= \sum_{i=1}^n \sum_{j=1}^n \mathbf{P}(A_i \cap A_j \neq \emptyset) \mathbf{E} [(X_i - a)(X_j - a) | A_i \cap A_j \neq \emptyset] \\
 &\quad + \sum_{i=1}^n \sum_{j=1}^n \mathbf{P}(A_i \cap A_j = \emptyset) \mathbf{E} [(X_i - a)(X_j - a) | A_i \cap A_j = \emptyset] \\
 &= \sum_{i=1}^n \sum_{j=1}^n \mathbf{P}(A_i \cap A_j \neq \emptyset) \mathbf{E} [(X_i - a)^2 | A_i \cap A_j \neq \emptyset] \\
 &\quad + \sum_{i=1}^n \sum_{j=1}^n \mathbf{P}(A_i \cap A_j = \emptyset) \underbrace{\mathbf{E} [(X_i - a) | A_i \cap A_j = \emptyset]}_{=0} \mathbf{E} [(X_j - a) | A_i \cap A_j = \emptyset] \\
 &= 4p(1-p) \sum_{i=1}^n \sum_{j=1}^n \mathbf{P}(A_i \cap A_j \neq \emptyset). \tag{3.36}
 \end{aligned}$$

Für nicht verschmelzende Couplings \tilde{A}_i, \tilde{A}_j der Ahnenlinien gilt hingegen

$$\begin{aligned}
 \mathbf{E} [|\tilde{A}_i \cap \tilde{A}_j|] &= \underbrace{\mathbf{P}(\tilde{A}_i \cap \tilde{A}_j \neq \emptyset)}_{\mathbf{P}(A_i \cap A_j \neq \emptyset)} \sum_{l \geq 0} q_l^2 \\
 &= \mathbf{P}(A_i \cap A_j \neq \emptyset) (|Q|^2)_0
 \end{aligned}$$

sowie

$$\mathbf{E} [|\tilde{A}_i \cap \tilde{A}_j|] = \sum_{n \geq 0} q_n q_{n+|j-i|}.$$

Damit folgt

$$\begin{aligned}
 \mathbf{Var} [S_n] &= \frac{4p(1-p)}{(|Q|^2)_0} \sum_{i=1}^n \sum_{j=1}^n (|Q|^2)_{|j-i|} \\
 &= \frac{4p(1-p)}{(|Q|^2)_0} \left[2 \sum_{i=1}^n (n-i) (|Q|^2)_i + n (|Q|^2)_0 \right]. \tag{3.37}
 \end{aligned}$$

Mit (3.33) und [BGT, Theorem 4.10.1(a)] folgt für

$$c_k := 2 \sum_{n \geq 0} q_n q_{n+k}$$

sowie $1 - \beta := 2\alpha$

$$\begin{aligned}
 \sum_{k=1}^n c_k &\sim \frac{n^{1-\beta}}{L(n)^2} \frac{1}{\Gamma(1-\alpha)^2(1-\beta)\Gamma(1-\beta) \sin(\frac{1}{2}\pi\beta)} \\
 &= \frac{n^{2\alpha}}{L(n)^2} \frac{1}{\Gamma(1-\alpha)^2(1-\beta)\Gamma(1-\beta) \sin(\frac{1}{2}\pi\beta)} \\
 &= \frac{n^{2\alpha}}{L(n)^2} \frac{1}{\Gamma(1-\alpha)^2 2\alpha \Gamma(1-\beta) \cos(\pi\alpha)} \tag{3.38}
 \end{aligned}$$

und damit

$$\begin{aligned}
 &2 \sum_{i=1}^n (n-i)(|Q|^2)_i \\
 &\sim \frac{n^{2\alpha+1}}{L(n)^2} \frac{1}{\Gamma(1-\alpha)^2 \Gamma(1-\beta) \cos(\pi\alpha)} \int_0^1 (1-x)x^{2\alpha-1} dx \\
 &= \frac{n^{2\alpha+1}}{L(n)^2} \frac{1}{\Gamma(1-\alpha)^2 \Gamma(1-\beta) \cos(\pi\alpha)} \cdot \frac{1}{2\alpha(2\alpha+1)}.
 \end{aligned}$$

Damit folgt nach (3.37) die Behauptung. \square

3.3.2 Herleitung der Asymptotik der Varianz von S_n aus Satz 2.3

Wir geben an dieser Stelle einen auf Proposition 3.2 fußenden Beweis von Lemma 2.7:

Lemma 3.8. *Es sei C wie in (3.7) und \tilde{L} wie in (3.9). Wir setzen*

$$\hat{C} := C^2 \int_0^\infty (1+x)^{\alpha-1} x^{\alpha-1} dx$$

und

$$\hat{L}(n) := \tilde{L}(n)^2,$$

dann gilt

$$\mathbf{Var}[S_n] \sim 2 \frac{4p(1-p)}{2\alpha(2\alpha+1) (|Q|^2)_0} n^{2\alpha+1} \hat{L}(n) \hat{C}.$$

Die Berechnung von

$$\int_0^\infty (1+x)^{\alpha-1} x^{\alpha-1} dx$$

mittels Substitution zeigt dann auch, dass Lemma 3.8 der Aussage von Lemma 2.7 entspricht. Wir zeigen zuerst eine Hilfsaussage ähnlich zu Lemma 3.3:

Lemma 3.9. *Es sei*

$$r_n := n^{2\alpha-1}K(n), \quad n \in \mathbb{N},$$

für ein langsam variierendes K der Form

$$K(n) := \exp\left(\int_B^n \frac{l(t)}{t} dx\right), \quad n \in \mathbb{N}, \quad (3.39)$$

für eine Konstante $B \in \mathbb{R}_{\geq 0}$ sowie eine beschränkte messbare Funktion $l(t), t \geq B$ mit

$$\lim_{t \rightarrow \infty} l(t) = 0.$$

Dann ist $(r_n)_{n \in \mathbb{N}}$ schließlich monoton fallend und es gilt

$$\sum_{k=1}^n r_k \sim n^{2\alpha} K(n) \frac{\Gamma(2\alpha)}{\Gamma(2\alpha+1)} = n^{2\alpha} K(n) \frac{1}{2\alpha}.$$

Beweis. Nach Voraussetzung existiert $n_0 \in \mathbb{N}$ mit $l(t) < \beta := -(2\alpha - 1)$ für alle $t \geq n_0$. Es sei also $n \geq n_0$, dann gilt

$$\begin{aligned} \frac{r_n}{r_{n+1}} &= \frac{n^{2\alpha-1}}{(n+1)^{2\alpha-1}} \cdot \frac{K(n)}{K(n+1)} \\ &= \left(\frac{n+1}{n}\right)^\beta \exp\left(-\int_n^{n+1} \frac{l(t)}{t} dt\right) \\ &= \exp\left(\beta(\ln(n+1) - \ln(n)) - \int_n^{n+1} \frac{l(t)}{t} dt\right) \\ &= \exp\left(\int_n^{n+1} \underbrace{\left(\frac{\beta}{t} - \frac{l(t)}{t}\right)}_{>0 \text{ für } t \geq n_0} dt\right) \\ &> 1, \end{aligned}$$

womit $(r_n)_{n \in \mathbb{N}}$ für $n \geq n_0$ monoton fallend ist. Damit liefert [F, Theorem 5, p. 447]

$$\sum_{k=1}^n r_k \sim n^{2\alpha} K(n) \frac{\Gamma(2\alpha)}{\Gamma(2\alpha+1)} = n^{2\alpha} K(n) \frac{1}{2\alpha}.$$

□

Mit dem vorigen Lemma lässt sich dann eine analoge Aussage für $(|Q|^2)_k = \sum_{j \geq 1} q_j q_{j+k}$ zeigen, ähnlich wie dies im Schritt von Lemma 3.3 nach Proposition 3.2 geschieht.

Lemma 3.10. Für die oben definierte Konstante \hat{C} und langsam variierende Funktion \hat{L} gilt

$$\sum_{k=1}^n \sum_{j \geq 1} q_j q_{j+k} \sim \frac{1}{2\alpha} \hat{C} \hat{L}(n) n^{2\alpha}.$$

Beweis. Nach [BGT, Theorem 1.3.2] existiert eine langsam variierende Funktion K , welche (3.39) erfüllt, sowie eine Folge $(D_n)_n$ mit $\lim_{n \rightarrow \infty} D_n = D \in \mathbb{R}$ so, dass

$$\hat{L}(n) = D_n K(n)$$

gilt. Nach Proposition 3.2 existiert nun eine Folge \hat{C}_n mit $\lim_{n \rightarrow \infty} \hat{C}_n = \hat{C}$ so, dass

$$\sum_{j \geq 1} q_j q_{j+n} = \hat{C}_n \hat{L}(n) n^{2\alpha-1} = \hat{C}_n D_n K(n) n^{2\alpha-1}$$

gilt. Da nun $(\hat{C}_n)_{n \in \mathbb{N}}$, $(D_n)_{n \in \mathbb{N}}$ konvergente Folgen sind sowie

$$\lim_{n \rightarrow \infty} \sum_{k=1}^n r_k = \infty$$

gilt, erhalten wir

$$\sum_{k=1}^n \sum_{j \geq 1} q_j q_{j+k} \sim n^{2\alpha} \hat{L}(n) \hat{C} \frac{\Gamma(2\alpha)}{\Gamma(2\alpha+1)} = n^{2\alpha} \hat{L}(n) \hat{C} \frac{1}{2\alpha}, \quad n \rightarrow \infty.$$

□

Analog lässt sich

$$\sum_{i=1}^n i (|Q|^2)_i \sim \frac{n^{2\alpha+1}}{2\alpha+1} \hat{L}(n) \hat{C} \quad (3.40)$$

zeigen. Damit lässt sich dann Lemma 3.8 beweisen:

Beweis von Lemma 3.8. Es gilt

$$\mathbf{Var} [S_n] = 4p(1-p) \sum_{i=1}^n \sum_{j=1}^n \mathbf{P} (A_i \cap A_j \neq \emptyset)$$

und nach (3.2)

$$\mathbf{P} (A_i \cap A_j \neq \emptyset) = \frac{(|Q|^2)_{|j-i|}}{(|Q|^2)_0}.$$

Damit erhalten wir

$$\begin{aligned} \mathbf{Var} [S_n] &= \frac{4p(1-p)}{(|Q|^2)_0} \sum_{i=1}^n \sum_{j=1}^n (|Q|^2)_{|j-i|} \\ &= \frac{4p(1-p)}{(|Q|^2)_0} \left[2 \sum_{i=1}^n (n-i) (|Q|^2)_i + n (|Q|^2)_0 \right]. \end{aligned}$$

Nach Lemma 3.10 und (3.40) gilt nun

$$\begin{aligned} &\sum_{i=1}^n (n-i) (|Q|^2)_i \\ &\sim n^{2\alpha+1} \hat{L}(n) \hat{C} \frac{1}{2\alpha(2\alpha+1)} \end{aligned}$$

Damit folgt die Behauptung. □

Nach Lemma 3.5 gilt mit dem Eulerschen Reflexionsprinzip und bekannten Additionstheoremen

$$\begin{aligned} \hat{C} &= \left[\frac{1}{\Gamma(\alpha)\Gamma(1-\alpha)} \right]^2 \int_0^\infty x^{\alpha-1} (1+x)^{\alpha-1} dx \\ &= \frac{\Gamma(1-2\alpha)}{\Gamma(\alpha)\Gamma(1-\alpha)^3} \\ &= \frac{1}{2} \frac{1}{\Gamma(1-\alpha)^2 \Gamma(2\alpha)} \cdot \underbrace{2 \frac{\Gamma(2\alpha)\Gamma(1-2\alpha)}{\Gamma(\alpha)\Gamma(1-\alpha)}}_{= \frac{\sin(\alpha\pi)}{\sin(2\alpha\pi)^{\frac{1}{2}}} = \frac{1}{\cos(\pi\alpha)}} \\ &= \frac{1}{2\Gamma(1-\alpha)^2 \Gamma(2\alpha) \cos(\pi\alpha)}, \end{aligned}$$

womit die Behauptung von Lemma 3.8 genau der von Lemma 2.7 entspricht.

3.3.3 Funktionale Konvergenz gegen die fraktionale Brownbewegung

Wir wollen aus den eben beschriebenen Gründen nun einen alternativen Beweis von Satz 2.10 vorstellen. Zuerst erweitern wir dafür die Konvergenz im Sinne der endlichdimensionalen Randverteilungen auf die Konvergenz in Verteilung bezogen auf den Pfadraum $C([0, 1])$ ausgestattet mit der Supremumsnorm:

Proposition 2.9. *Sei $T > 0$. Bezogen auf den Pfadraum $(C([0, T]), \|\cdot\|_{L^\infty([0, T])})$ konvergiert \tilde{S}^n in Verteilung gegen B^H für eine fraktionale Brownsche Bewegung B^H mit Hurst-Parameter $H = \frac{1}{2} + \alpha$.*

Mit diesem Resultat entbehrt es nun aller Notwendigkeit, eine Behauptung der Form von Behauptung 3.13, wie sie in [HS] angegeben wurde, zu zeigen, wir können den Beweis direkt führen. Dafür geben wir zuerst ein Resultat aus [B1] wieder:

Satz 3.11 ([B1, Theorem 5.16]). *Konvergiert eine Folge von stetigen Prozessen $(Y^n)_n$ auf $[0, T]$ im Sinne der endlichdimensionalen Randverteilungen gegen einen stetigen Prozess Y auf $[0, T]$ und erfüllt für alle $0 \leq s \leq t \leq u \leq T$ sowie für Wahlen von $\gamma > \frac{1}{2}$, $N \in \mathbb{N}$ und eine nichtfallende Funktion F auf $[0, T]$*

$$\mathbf{E} [|Y_t^n - Y_s^n| \cdot |Y_u^n - Y_t^n|] \leq [F(u) - F(s)]^{2\gamma} \quad \forall n \geq N,$$

so konvergiert $(Y^n)_n$ in Verteilung auf dem Pfadraum $(C([0, T]), \|\cdot\|_{L^\infty([0, T])})$ gegen Y .

Zusammen mit dem folgenden Lemma und einer leichten Folgerung aus diesem lässt sich dann der Beweis von 2.9 auch schon unbeschwert führen: Eine direkte Folgerung aus der Asymptotik der Varianz von S_n , also aus Lemma 2.7, ist, da sowohl $\mathbf{Var}[S_n]$ und die rechte Seite von (2.9) für alle $n \in \mathbb{N}$ endlich ist, das folgende Korollar:

Korollar 3.12. *Es existiert eine Konstante $c_1 > 0$, sodass*

$$\mathbf{Var} [S_n] < c_1 n^{2H} L(n)^{-2} \tag{3.41}$$

für alle $n \in \mathbb{N}$ gilt.

Wie versprochen führen wir mit diesen Aussagen nun den Beweis von 2.9.

Beweis von Proposition 2.9. Wir wählen $0 \leq s \leq t \leq u \leq T$ sowie $j, k, l \in \mathbb{N}$ mit

$$\frac{j}{n} \leq s < \frac{j+1}{n}, \quad \frac{k}{n} \leq t < \frac{k+1}{n}, \quad \frac{l}{n} \leq u < \frac{l+1}{n}.$$

Da wir S zwischen $[nt]$ und $[nt] + 1$ durch lineare Interpolation definiert haben, gilt

$$\frac{\tilde{S}_t^n - \tilde{S}_s^n}{\tilde{c}n^{-\alpha-\frac{1}{2}}L(n)} = \sum_{m=j+1}^k X_m + [1 - (sn - j)] X_{j+1} + [tn - k] X_{k+1}.$$

Dass $X_m \in \{0, 1\}$ für alle $m \in \mathbb{N}$ sowie $0 \leq [1 - (sn - j)] \leq 1$ und $0 \leq [tn - k] \leq 1$ gilt, liefert zusammen mit

$$(m + \delta)^2 = m^2 + 2\delta m + \delta^2 \leq 6m^2 + 5, \quad \text{für } \delta \in [0, 1], m \in \mathbb{Z}$$

und

$$(m - \delta)^2 = m^2 - 2\delta m + \delta^2 \leq 6m^2 + 5, \quad \text{für } \delta \in [0, 1], m \in \mathbb{Z}$$

die erste Ungleichung in der folgenden Kette:

$$\begin{aligned} \text{Var} \left[\tilde{S}_t^n - \tilde{S}_s^n \right] &\leq \left[6\text{Var} [S_{k-j}] + 5 \right] \cdot \tilde{c}^2 n^{-2\alpha-1} L(n)^2 \\ &\leq \left[6c_1(k-j)^{2\alpha+1} L(k-j)^{-2} + 5 \right] \cdot \tilde{c}^2 n^{-2\alpha-1} L(n)^2 \\ &\leq \left[c_2(k-j)^{2\alpha+1} L(k-j)^{-2} \right] \cdot \tilde{c}^2 n^{-2\alpha-1} L(n)^2 \\ &\leq c_3 \left(\frac{k-j}{n} \right)^{2\alpha+1} \left(\frac{L(n)}{L(k-j)} \right)^2. \end{aligned} \quad (3.42)$$

Die zweite Ungleichung folgt mit (3.41), die dritte folgt für ein $c_2 > 0$ und für alle $n \geq N_1$ für ein $N_1 \in \mathbb{N}$, da L langsam varriierend ist. (Man behalte im Hinterkopf, dass k, j, l für fixe s, t, u von n abhängen.) Und offensichtlich gilt die letzte Ungleichung für $c_3 := \max(1, c_2 \tilde{c}^2)$. Analog gilt

$$\text{Var} \left[\tilde{S}_u^n - \tilde{S}_t^n \right] \leq c_3 \left(\frac{l-k}{n} \right)^{2\alpha+1} \left(\frac{L(n)}{L(l-k)} \right)^2. \quad (3.43)$$

Um Lemma 3.11 anzuwenden, müssen wir den Erwartungswert

$$\mathbf{E} \left[\left| \tilde{S}_t^n - \tilde{S}_s^n \right| \cdot \left| \tilde{S}_u^n - \tilde{S}_t^n \right| \right].$$

beschränken. Für $l \leq j+1$ erhalten wir $|u-t| \leq \frac{2}{n}$ sowie $|t-s| \leq \frac{2}{n}$, was

$$(u-t)(t-s) \leq (u-s)^{1+\alpha}$$

liefert. Nach Definition von μ gilt $L(n) \leq n^{-\alpha}$ (und auch da L langsam varriierend ist, gilt das zumindest für n groß genug). Da wir S zwischen ganzzahligen Zeitpunkten linear interpolieren:

$$\mathbf{E} \left[\left| \tilde{S}_t^n - \tilde{S}_s^n \right| \cdot \left| \tilde{S}_u^n - \tilde{S}_s^n \right| \right] \quad (3.44)$$

$$\leq \left(\tilde{c} n^{-\frac{1}{2}-\alpha} L(n) \right)^2 ((t-s) \cdot (u-s)) \quad (3.45)$$

$$\leq \left(\tilde{c} n^{-\frac{1}{2}} \right)^2 (u-s)^{1+2\alpha} \quad (3.46)$$

$$= \tilde{c}^2 \frac{1}{n} (u-s)^{1+2\alpha} \quad (3.47)$$

$$\leq \tilde{c}^2 (u-s)^{1+2\alpha} = [c_0 u - c_0 s]^{1+2\alpha} \quad (3.48)$$

für eine Konstante $c_0 > 1$. Der spannendere Fall ist $l > j + 1$. Für diesen liefert Cauchy-Schwarz zusammen mit den Abschätzungen (3.42),(3.43)

$$\begin{aligned}
 & \mathbf{E} \left[\left| \tilde{S}_t^n - \tilde{S}_s^n \right| \cdot \left| \tilde{S}_u^n - \tilde{S}_t^n \right| \right] \\
 & \leq \sqrt{\mathbf{Var} \left[\tilde{S}_t^n - \tilde{S}_s^n \right]} \sqrt{\mathbf{Var} \left[\tilde{S}_u^n - \tilde{S}_t^n \right]} \\
 & \leq \sqrt{c_3 \left(\frac{k-j}{n} \right)^{2\alpha+1} \left(\frac{L(n)}{L(k-j)} \right)^2} \sqrt{c_3 \left(\frac{l-k}{n} \right)^{2\alpha+1} \left(\frac{L(n)}{L(l-k)} \right)^2} \\
 & \leq c_3 \left(\frac{l-j}{n} \right)^{2\alpha+1} \left(\frac{L(n)}{L(k-j)L(l-k)} \right)^2 \\
 & \leq 2c_3 \left(\frac{l-j}{n} \right)^{2\alpha+1} \leq (2c_3 \cdot 2(u-s))^{2\alpha+1}.
 \end{aligned}$$

(Dabei gilt die dritte Ungleichung wegen $4|k-j| \cdot |l-k| \leq |l-j|^2$ und die vierte Ungleichung für alle $n \geq N_2$ für ein $N_2 \in \mathbb{N}$, da L langsam variiierend ist.) Setzen wir nun $Y := B^H$ und $Y^n := \tilde{S}^n$, dann können wir Lemma 3.11 für $n \geq N$ mit $N := \max\{N_1, N_2\}$, $\gamma := \alpha + 1/2$ und $F(x) := 4c_3x$ anwenden. Damit folgt die Behauptung von Proposition 2.9. \square

3.3.4 Diskussion zu [HS, Lemma 4.1]

Vorgegangen wird in [HS] für den Beweis von Satz 2.10, indem mittels Proposition 2.8 an endlich vielen Stützstellen passend gekoppelt wird und die Schwankungen der Prozesse B^H und \tilde{S}^n zwischen diesen kontrolliert wird, indem einerseits der Stetigkeitsmodul der fraktionalen Brownschen Bewegung ausgenutzt wird und andererseits das Maximum (und analog das Minimum) des Prozesses S^n auf dem Intervall $[0, 1]$ kontrolliert wird. Um Letzteres zu erreichen, zielen [HS] auf den Beweis der folgenden Behauptung ab:

Behauptung 3.13 ([HS, Lemma 4.1]). *Für*

$$A_n := \max_{i \in \{0, \dots, n\}} [S_i - (2p-1)i]$$

existiert $C_0 > 0$, sodass für alle $\varepsilon > 0$ ein $N \in \mathbb{N}$ existiert, sodass

$$\mathbf{P} \left(A_n > \left(C_0 + 50\tilde{c}^{-1}\sqrt{-\ln \varepsilon} \right) n^{\frac{1}{2}+\alpha} L(n)^{-1} \right) < \varepsilon$$

für alle $n \geq N$ gilt.

Der in [HS] angegebene Beweis dieser Behauptung stellt sich leider als formallogisch falsch heraus. Im Laufe des Beweises wird argumentiert, dass die Tatsache, dass für ein festes Paar ε, K mit $2K^{-\frac{1}{2}-\alpha} < 1$ und für $n \geq N$ für ein $N \in \mathbb{N}$ eine Folge $(r_n)_{n \in \mathbb{N}}$ die Ungleichung

$$r_{Kn} \leq 2K^{-\frac{1}{2}-\alpha}r_n + 20\tilde{c}^{-1}\sqrt{-\ln \varepsilon} + 24\tilde{c}^{-1}\sqrt{-\ln \varepsilon}K^{-\frac{1}{2}-\alpha} \quad (3.49)$$

erfüllt, bereits genügt, um die Beschränktheit von $(r_n)_n$ zu zeigen. Widerlegen lässt sich dies recht einfach durch folgende Folge:

$$r_n := \begin{cases} s_n & \text{falls } K \text{ nicht } n \text{ teilt,} \\ 2K^{-\frac{1}{2}-\alpha}r_{n/K} + 20\tilde{c}^{-1}\sqrt{-\ln \varepsilon} + 24\tilde{c}^{-1}\sqrt{-\ln \varepsilon}K^{-\frac{1}{2}-\alpha} & \text{sonst,} \end{cases}$$

wobei $(s_n)_{n \in \mathbb{N}}$ eine zumindest entlang der Folge der Primzahlen unbeschränkte Folge sei. Die Folge $(r_n)_{n \in \mathbb{N}}$ erfüllt (3.49) per Definition, da es aber unendlich viele Primzahlen gibt, also auch unendlich viele $n \in \mathbb{N}$, sodass K eben nicht n teilt, existiert für alle C und N ein $n \geq \max\{N, C\}$, sodass $r_n \geq C$ gilt. Dementsprechend ist $(r_n)_{n \in \mathbb{N}}$ nicht beschränkt.

3.3.5 Erste Variation von [HS, Lemma 4.1]

Obwohl wir für den Beweis der der funktionalen Konvergenz einen anderen Weg als den über [HS, Lemma 4.1] wählten, kann man auch mit Eigenschaften der fraktionalen Brownschen Bewegung etwas in der Form von Behauptung 3.13 beweisen:

Lemma 3.14 (Borell-TIS-Ungleichung, [AT, Theorem 2.1.1]). \mathbb{T} sei ein topologischer Raum und $(Y_t)_{t \in \mathbb{T}}$ ein Gauß'scher Prozess, sodass $\sup_{t \in \mathbb{T}} |Y_t|$ fast sicher endlich ist. Dann gilt mit

$$m := \mathbf{E} \left[\sup_{t \in \mathbb{T}} |Y_t| \right], \quad v := \sup_{t \in \mathbb{T}} \mathbf{E}[|Y_t|^2]$$

für alle $u > 0$

$$\mathbf{P} \left(\sup_{t \in \mathbb{T}} |Y_t| > m + u \right) \leq \exp \left(-\frac{u^2}{2v} \right).$$

In Worten bedeutet dies nicht mehr, als dass die Schwänze der Verteilung $\mathcal{L} \left(\sup_{t \in \mathbb{T}} |Y_t| \right)$ höchstens so schwer wie die einer Normalverteilung sind.

Beweis. Wir wählen nun Y als die fraktionale Brownsche Bewegung B^H für $H := \alpha + \frac{1}{2}$ sowie $\mathbb{T} := [0, 1]$. Aufgrund der zweiten Momente der fraktionalen Bewegung liefert das

$$v = \mathbf{E} [Y_1^2] = \mathbf{E} [B_1^H] = 1.$$

Nach [VB, Theorem 1] gilt außerdem

$$\mathbf{E}[\sup_{t \in [0,1]} B_t^H] \leq \sqrt{2/\pi}.$$

Die Symmetrie der Verteilung der fraktionalen Brownschen Bewegung sowie

$$\sup_{t \in [0,1]} |B_t^H| \leq \sup_{t \in [0,1]} B_t^H + |\inf_{t \in [0,1]} B_t^H|$$

liefert

$$m = \mathbf{E}[\sup_{t \in [0,1]} |B_t^H|] \leq 2\sqrt{2/\pi} \approx 1.59.$$

Zu vorgegebenem $\varepsilon > 0$ setzen wir

$$u := \sqrt{2 \log(\varepsilon^{-1})}$$

und erhalten dank der Borell-TIS-Ungleichung

$$\mathbf{P} \left(\sup_{t \in [0,1]} B_t^H \geq 1.6 + \sqrt{2 \log(\varepsilon^{-1})} \right) \leq \varepsilon.$$

Die Abbildung $y \mapsto \sup_{t \in [0,1]} y_t$ ist bezüglich der Supremumsnorm auf $C([0,1])$ beschränkt und stetig, womit auch die Konvergenz in Verteilung von $\sup_{t \in [0,1]} S^n$ für $n \rightarrow \infty$ gegen B^H folgt. Das Portmanteau-Lemma liefert dann also die Behauptung:

$$\limsup_{n \rightarrow \infty} \mathbf{P} \left(\max_{t \in [0,1]} S_t^n \geq 1.6 + \sqrt{2 \log(\varepsilon^{-1})} \right) \leq \varepsilon.$$

□

3.3.6 Zweite Variation von [HS, Lemma 4.1] und alternativer Beweis der funktionalen Konvergenz

Diskussionen mit REUBEN DROGIN ergaben außerdem, dass ein Weg über eine schwächere Form von [HS, Lemma 4.1] durchaus zum Ziel führt. Nach unveröffentlichten Notizen von REUBEN DROGIN ist für $c_n := \tilde{c}^{-1} n^{\frac{1}{2} + \alpha} L(n)^{-1}$ die durch

$$A_n/c_n := \frac{1}{c_n} \max_{i \in \{0, \dots, n\}} S_i$$

definierte Folge straff. Insbesondere heißt das also, dass für alle $\eta > 0$ ein $K(\eta) > 0$ existiert, sodass

$$\mathbf{P} \left(\frac{A_m}{c_m} > K(\eta) \right) < \eta \quad \forall m \in \mathbb{N} \tag{3.50}$$

Diese Aussage kommt der Aussage von [HS, Lemma 4.1] bis auf die Form der Schwänze der Verteilung der A_n sehr nahe. Wir zeigen nun, dass das stark genug ist, um mit [B1, Theorem 8.1] die Behauptung von Proposition 2.9 auf einem zweiten Weg zu zeigen. Wir erinnern kurz an [B1, Theorem 8.1] :

Satz 3.15 ([B1, Theorem 8.1]). *Es sei $(Y_n)_n$ eine Folge von $C([0, T])$ -wertigen Zufallsvariablen. $(Y_n)_n$ ist genau dann straff, wenn die folgenden beiden Voraussetzungen erfüllt sind:*

1. Für alle $\eta > 0$ existiert $c > 0$, sodass

$$\mathbf{P}(|Y_n(0)| > c) \leq \eta$$

gilt.

2. Für alle $\varepsilon, \eta > 0$ existiert $\delta \in (0, 1)$ und $N \in \mathbb{N}$ so, dass für alle $n \geq N$

$$\mathbf{P}(w_{Y_n}(\delta) \geq \varepsilon) \leq \eta$$

gilt, wobei wir

$$w_y(\delta) := \sup_{|s-t| < \delta} |y(s) - y(t)|$$

schreiben.

Voraussetzung 1 ist wegen $Y_n := \tilde{S}_0^n := 0$ erfüllt. Voraussetzung 2 ist mit Reuben Drogin's Resultaten auch erfüllt, denn: Seien $\varepsilon, \eta > 0$ und $\delta = \delta(\varepsilon, \eta)$ so klein, dass für n groß genug und $m \leq \delta n$

$$\frac{c_m}{c_n} K(\eta) < \varepsilon$$

gilt, dann erhält man für alle n groß genug und m mit $\frac{m}{n} \leq \delta$

$$\begin{aligned} & \mathbf{P}\left(\frac{A_m}{c_n} > \varepsilon\right) \\ & \leq \mathbf{P}\left(\frac{A_m}{c_n} > \frac{c_m}{c_n} K(\eta)\right) \\ & = \mathbf{P}\left(\frac{A_m}{c_m} > K(\eta)\right) \\ & < \eta \end{aligned}$$

nach (3.50). Mit der analogen Argumentation für die Straffheit der Folge

$$\left(\frac{B_n}{c_n}\right)_n, \quad B_n := \min_{i \in \{0, \dots, n\}} S_i$$

ist also Voraussetzung 2 von Satz 3.15 für $Y_n := \tilde{S}^n, n = 1, 2, \dots$ erfüllt.

3.3.7 Couplings für alle Zeithorizonte gemeinsam über einen Satz von Skorochod

Basierend auf unserem Resultat zur funktionalen Konvergenz lässt sich nun die Existenz gemeinsamer Couplings zeigen:

Proposition 3.16. *Es seien die $Y^n := (Y^n(t))_{0 \leq t < \infty}$, $n = 1, 2, \dots$, und $Y := Y(t)_{0 \leq t < \infty}$ stetige Prozesse, sodass für alle T die Prozesse $(Y^n(t))_{0 \leq t \leq T}$, $n = 1, 2, \dots$, in Verteilung bezogen auf den Pfadraum $(C([0, T]), \|\cdot\|_{L^\infty([0, T])})$ gegen $(Y(t))_{0 \leq t \leq T}$ konvergieren. Dann existiert eine Folge von Couplings $(C_n)_n$, für die*

$$\lim_{n \rightarrow \infty} C_n (\|Y^n - Y\|_{L^\infty([0, T])} > \varepsilon) = 0 \quad (3.51)$$

für alle $\varepsilon, T > 0$ gilt.

Beweis. Skorochods Darstellungssatz ([B2, Theorem 6.7]) liefert, dass für alle $T > 0$ ein Wahrscheinlichkeitsraum $(\Omega^T, \mathcal{F}^T, Q^T)$ und darauf definierte Prozesse

$$(Y^{n,T}(t))_{0 \leq t \leq T}, (Y^T(t))_{0 \leq t \leq T}$$

existieren, sodass

$$\left(Y^{n,T}(t) \right)_{0 \leq t \leq T} \stackrel{(d)}{=} (Y^n(t))_{0 \leq t \leq T}, \quad \left(Y^T(t) \right)_{0 \leq t \leq T} \stackrel{(d)}{=} (Y(t))_{0 \leq t \leq T}$$

gilt, und sodass $(Y^{n,T}(t))_{0 \leq t \leq T}$, $n = 1, 2, \dots$, betrachtet als Zufallsvariablen mit Werten in

$$(C([0, T]), \|\cdot\|_{L^\infty([0, T])}),$$

in Wahrscheinlichkeit gegen $(Y^T(t))_{0 \leq t \leq T}$ konvergieren. Unser erstes Ziel ist es, das Coupling auf den Zeithorizont $[0, \infty)$ zu erweitern. Es gilt also den Wahrscheinlichkeitsraum $(\Omega^T, \mathcal{F}^T, Q^T)$ passend zu erweitern. Dafür sei Π^T eine reguläre Desintegration der Verteilung von $(Y(t))_{0 \leq t < \infty}$ bezüglich $(Y(t))_{0 \leq t \leq T}$ und analog Π_n^T eine reguläre Desintegration der Verteilung von $(Y^n(t))_{0 \leq t < \infty}$ bezüglich $(Y^n(t))_{0 \leq t \leq T}$. Gegeben das Paar

$$((Y^{n,T}(t))_{0 \leq t \leq T}, (Y^T(t))_{0 \leq t \leq T})$$

setzen wir den Pfad $(Y^{n,T}(t))_{0 \leq t \leq T}$ gemäß der Verteilung

$$\Pi^T((Y^{n,t}(t))_{0 \leq t \leq T}, \cdot)$$

fort und den Pfad $(Y^T(t))_{0 \leq t \leq T}$ gemäß

$$\Pi^T((Y^T(t))_{0 \leq t \leq T}, \cdot).$$

Dabei sind diese zwei Fortsetzungen unabhängig gegeben das Paar

$$((Y^{n,T}(t))_{0 \leq t \leq T}, (Y^T(t))_{0 \leq t \leq T}).$$

Damit erhalten wir also Couplings C_n^T von Y^n und Y , welche

$$\lim_{n \rightarrow \infty} C_n^T (\|Y^n - Y\|_{L^\infty([0,T])} > \varepsilon) = 0 \quad (3.52)$$

für alle $\varepsilon > 0$ erfüllen. Wir finden also für alle $\varepsilon > 0$ und alle $T > 0$ eine natürliche Zahl $N(T, \varepsilon)$, sodass für alle $n \geq N(T, \varepsilon)$

$$C_n^T (\|Y^n - Y\|_{L^\infty([0,T])} > \varepsilon) < \varepsilon \quad (3.53)$$

gilt. Induktiv definieren wir nun die streng monotone Folge (m_n) durch $m_1 := N(1, 1)$ und $m_n := \min(m_{n-1} + 1, N(n, 1/n))$. Damit definieren wir die nichtsteigende Folge $(\varepsilon_k)_{k \geq m_1}$ durch

$$\varepsilon_k := \frac{1}{n} \quad \text{für } n \text{ mit } m_n \leq k < m_{n+1}.$$

Setzen wir $C_n := C_n^n$, dann gilt

$$C_n (\|Y^n - Y\|_{L^\infty([0,n])} > \varepsilon_n) < \varepsilon_n$$

für alle $n \geq m_1$. Nach Konstruktion gilt $m_n \rightarrow \infty$ für $n \rightarrow \infty$ und damit

$$\lim_{n \rightarrow \infty} \varepsilon_n = 0.$$

Das heißt für alle $\varepsilon, T > 0$ gilt

$$\lim_{n \rightarrow \infty} C_n (\|Y^n - Y\|_{L^\infty([0,T])} > \varepsilon) = 0.$$

□

Damit ist Satz 2.10 nun eine direkte Folgerung aus Proposition 2.9 und Proposition 3.16.

Der Aufbau des Beweises von Proposition 3.16 suggeriert nun bereits, dass in diesen Gedanken allgemeinere Resultate stecken und sich dies auf einen breiteren Kontext anwenden lässt. Auf den folgenden Seiten wollen wir dies auch machen. Gleichzeitig wollen wir das recht abstrakte Resultat von Skorochod durch ein greifbareres Resultat von Strassen ersetzen.

3.3.8 Couplings für alle Zeithorizonte gemeinsam über einen Satz von Strassen

Das folgende Lemma, das ein direktes Korollar aus dem Satz von Strassen (siehe z.B. [P, Theorem 46]) ist, wird für unseren alternativen Beweis von Proposition 3.16 absolut essentiell sein.

Lemma 3.17 ([D2, Corollary 11.6.4 und Theorem 11.3.5] oder [P, Theorem 47]). *Zu zwei Wahrscheinlichkeitsmaße μ und ν auf einem vollständigen separablen metrischen Raum (E, d) existiert ein Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbf{P})$ und auf diesem definierte E -wertige Zufallsvariablen X und Y für die*

$$\mathcal{L}(X) = \mu, \quad \mathcal{L}(Y) = \nu$$

sowie

$$\alpha(X, Y) = \rho(\nu, \mu),$$

gilt, wenn wir mit

$$\rho(\mu, \nu) := \inf \left\{ \delta > 0 : \mu(A) \leq \nu(A^\delta) + \delta \text{ für alle } A \in \mathfrak{B}(E) \right\}$$

den Prohorov-Abstand von μ und ν sowie mit

$$\alpha(X, Y) := \inf \left\{ \delta \geq 0 : \mathbf{P}(d(X, Y) > \delta) \leq \delta \right\}$$

den Abstand von X und Y in der Ky-Fan-Metrik bezeichnen.

Die bekannte Tatsache, dass der Prohorov-Abstand die durch die schwache Topologie induzierte Topologie auf dem Raum der Wahrscheinlichkeitsmaße auf E metrisiert, liefert damit sofort das folgende Korollar:

Korollar 3.18. *Es sei wieder (E, d) ein vollständiger separabler metrischer Raum und X, X_1, X_2, \dots seien E -wertige Zufallsvariablen, sodass*

$$X_n \xrightarrow{(d)} X$$

gilt. (Anders gesagt: $\mathcal{L}(X_n)$ konvergiert schwach gegen $\mathcal{L}(X)$.) Dann existiert eine Nullfolge $(\varepsilon_n)_n$ sowie Couplings C^n von X_n und X , sodass für alle $n \in \mathbb{N}$

$$C^n(d(X_n, X) > \varepsilon_n) \leq \varepsilon_n$$

gilt.

Bisher wäre es uns möglich für jeden Zeithorizont T einzeln Couplings zu definieren, um die vollständige Behauptung zu beweisen, muss man diese Couplings jedoch erweitern können: Das folgende Lemma hat also zum Ziel, für zwei stetige Prozesse $X = (X_t)_{t \geq 0}$ und $Y = (Y_t)_{t \geq 0}$, für die wir ein Coupling von $(X_t)_{0 \leq t \leq T}$ und $(Y_t)_{0 \leq t \leq T}$ haben, auch ein Coupling von $(X_t)_{t \geq 0}, (Y_t)_{t \geq 0}$ zu finden, sodass die gemeinsame Verteilung von $(X_t)_{0 \leq t \leq T}, (Y_t)_{0 \leq t \leq T}$ sowie die Randverteilungen von X und Y respektiert werden.

Lemma 3.19. *E und E' seien Standardborelräume sowie (X^1, Y^1) eine Zufallsvariable mit Wertebereich $E \times E'$. Setzt man nun $\mu_1 := \mathcal{L}(X^1)$, $\nu_1 := \mathcal{L}(Y^1)$ und sind μ, ν zwei Verteilungen auf $E \times E'$, deren Projektionen auf E wir μ_1 und ν_1 nennen, dann existiert ein Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, \mathbf{P})$ mit Zufallsvariablen $X = (X_1, X_2), Y = (Y_1, Y_2)$ mit Wertebereich $E \times E'$, sodass*

$$\mathcal{L}(X) = \mu, \quad \mathcal{L}(Y) = \nu$$

sowie

$$\mathcal{L}(X_1, Y_1) = \mathcal{L}(X^1, Y^1)$$

gelten.

Der Beweis kommt mit sehr elementaren Mitteln aus und anschaulich ist das Lemma sowieso überzeugend:

Beweis. ρ bezeichne die gemeinsame Verteilung von (X^1, Y^1) und Π_1 sei eine reguläre Desintegration der Verteilung μ bezüglich μ_1 sowie Π_2 eine reguläre Desintegration von ν bezüglich ν_1 .

Das kanonische Modell $E \times E' \times E \times E'$ mit Wahrscheinlichkeitsmaß

$$\rho(dx_1, dy_1) \Pi_1(x_1, dx_2) \Pi_2(y_1, dy_2), \quad x_1, y_1 \in E, x_2, y_2 \in E',$$

erfüllt dann die Behauptung. □

Damit ist es nun auf Basis eines leichten *inversen Diagonalfolgenarguments* möglich Proposition 3.16 schnell zu beweisen.

Beweis von Proposition 3.16. Nach Annahme konvergieren für alle $T > 0$ die Zufallsvariablen $(Y^n(t))_{0 \leq t \leq T}$ in Verteilung in $(C([0, T]), \|\cdot\|_{L^\infty([0, T])})$ für $n \rightarrow \infty$ gegen $(Y(t))_{0 \leq t \leq T}$. Für

$$X_n := (Y^n(t))_{0 \leq t \leq T}, \quad X = (Y(t))_{0 \leq t \leq T}$$

und

$$(E, d) = (C([0, T]), \|\cdot\|_{L^\infty([0, T])}) \quad \text{und} \quad E' = C([T, \infty])$$

liefern Korollar 3.18 und Lemma 3.19 also zusammen für alle $T > 0$ die Existenz von Couplings C_n^T von Y^n und Y sowie einer Nullfolge $(\varepsilon_n^T)_n$, sodass für alle $n \in \mathbb{N}$

$$C_n^T \left(\|Y^n - Y\|_{L^\infty([0,T])} > \varepsilon_n^T \right) \leq \varepsilon_n^T.$$

gilt. Wir setzen nun

$$N(T, \varepsilon) := \min \left\{ k \in \mathbb{N} : \varepsilon_m^T < \varepsilon \text{ für alle } m \geq k \right\}. \quad (3.54)$$

und definieren damit die strikt monoton steigende Folge $(m_n)_{n \in \mathbb{N}}$ durch

$$m_n := \max \left\{ m_{n-1} + 1, N(n, 1/n) \right\}, \quad m_1 = N(1, 1).$$

Wiederum basierend auf dieser definieren wir die nichtsteigende Folge $(\delta_n)_{n \geq m_1}$ durch

$$\delta_n := \frac{1}{k} \quad \text{für das größtes } k \text{ mit } m_k < n \leq m_{k+1}. \quad (3.55)$$

Zusammen liefert das

$$C_n^n \left(\|Y^n - Y\|_{L^\infty([0,n])} > \delta_n \right) \leq \delta_n.$$

Nach Konstruktion gilt für $n \rightarrow \infty$ also $m_n \rightarrow \infty$ und $\delta_n \rightarrow 0$, was für alle $\varepsilon > 0$

$$\lim_{n \rightarrow \infty} C_n^n \left(\|Y^n - Y\|_{L^\infty([0,n])} > \varepsilon \right) = 0$$

liefert. Insbesondere erfüllen die Couplings $C^n := C_n^n$ also (3.51). \square

3.4 Beweis der Grenzwertsätze für das [BGKS]-Modell

3.4.1 Die Asymptotik der Varianz von Z_n

Zum Beweis von Proposition 2.12 betrachten wieder zuerst die Asymptotik der Varianz von Z_n (siehe (2.10)) für $N \in \mathbb{N}$ fest:

Lemma 2.11. *Für die Varianz von Z_n gilt*

$$\mathbf{Var}[Z_n] = \mathbf{Var}[S_n] \cdot \frac{\sum_{l \geq 0} q_l^2}{\left(\sum_{l \geq 1} q_l^2 + N \right)}$$

für die in (2.2) beschriebene Irrfahrt S .

Damit liefert [HS, Lemma 3.1]:

$$\mathbf{Var}[Z_n] \sim \frac{n^{2\alpha+1}L(n)^{-2}}{\tilde{c}^2} \cdot \frac{\sum_{l \geq 0} q_l^2}{\left(\sum_{l \geq 1} q_l^2 + N\right)} \quad (3.56)$$

Beweis von Lemma 2.11. Wir bezeichnen mit den $A_{i,k}$ für $i \in \mathbb{Z}$ und $k \in \{1, \dots, N\}$ die Ahnenlinien des Individuums k in Generation i , die so gekoppelt sind, dass $A_{i,k}$ und $A_{j,l}$ für $(i, k) \neq (j, l)$ beim Treffen nicht verschmelzen. Sei ohne Beschränkung der Allgemeinheit $j > i$. Dann gilt:

$$\mathbf{Var}[Z_n] = 4p(1-p) \frac{1}{N^2} \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^N \sum_{m=1}^N \mathbf{P}(A_{i,k} \cap A_{j,m} \neq \emptyset), \quad (3.57)$$

sowie

$$\mathbf{E}[|A_{i,k} \cap A_{j,m}|] = \sum_{s \geq 0} \frac{q_s q_{j-i+s}}{N} =: \frac{1}{N} (|Q|^2)_{j-i} \quad (3.58)$$

und

$$\begin{aligned} \mathbf{E}[|A_{i,k} \cap A_{j,m}|] &= \mathbf{P}(A_{i,k} \cap A_{j,m} \neq \emptyset) \left(\sum_{l \geq 1} q_l^2 \frac{1}{N} + 1 \right) \\ &= \mathbf{P}(A_{i,k} \cap A_{j,m} \neq \emptyset) \left(\frac{1}{N} (|Q|^2)_0 + \frac{N-1}{N} \right) \end{aligned} \quad (3.59)$$

(3.58) und (3.59) liefern zusammen

$$\mathbf{P}(A_{i,k} \cap A_{j,m} \neq \emptyset) = \frac{(|Q|^2)_{j-i}}{\sum_{l \geq 1} q_l^2 + N}.$$

Damit erhalten wir aus (3.57)

$$\begin{aligned} \mathbf{Var}[Z_n] &= 4p(1-p) \frac{1}{N^2} \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^N \sum_{m=1}^N \mathbf{P}(A_{i,j} \cap A_{j,m} \neq \emptyset) \\ &= 4p(1-p) \frac{1}{N^2} \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^N \sum_{m=1}^N \frac{(|Q|^2)_{j-1}}{\sum_{l \geq 1} q_l^2 + N} \\ &= \frac{4p(1-p)}{N^2 \left(\sum_{l \geq 1} q_l^2 + N\right)} \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^N \sum_{m=1}^N (|Q|^2)_{j-i} \\ &\stackrel{q_0=1}{=} \frac{4p(1-p)N^2}{N^2 \left(\sum_{l \geq 1} q_l^2 + N\right)} \sum_{i=1}^n \sum_{j=1}^n (|Q|^2)_{j-i} \end{aligned}$$

$$= \frac{4p(1-p)}{\left(\sum_{l \geq 1} q_l^2 + N\right)} \left[2 \sum_{i=1}^n (n-i) (|Q|^2)_i + n (|Q|^2)_0 \right], \quad (3.60)$$

womit nach den Rechnungen im Beweis von [HS, Lemma 3.1] bereits (3.56) folgt. \square

3.4.2 Konvergenz der endlichdimensionalen Verteilungen

Um eine Konvergenz der endlichdimensionalen Randverteilungen des in Proposition 2.12 definierten Prozesses \tilde{Z}^n gegen die einer fraktionalen Brownschen Bewegung zu zeigen, ist also nur noch zu zeigen, dass die Verteilung von Z_n asymptotisch Gauß'sch ist. Dafür bedienen wir uns der gleichen Hilfsmittels wie [HS].

Definition 3.20. Sei $(k_n)_n$ eine Folge von natürlichen Zahlen und σ -Algebren $\mathcal{F}_{n,m}$ gegeben, so dass für alle $n \in \mathbb{N}$

$$\mathcal{F}_{n,0} \subseteq \mathcal{F}_{n,2} \subseteq \dots \subseteq \mathcal{F}_{n,k_n+n+1}$$

gilt. Wir bezeichnen $(Y_{n,m}, \mathcal{F}_{n,m})_{n \in \mathbb{N}_0, 0 \leq m \leq k_n+n+1}$ als Martingal-Differenz-Liste (oder auch als ein Dreiecksfeld von Martingaldifferenzen), falls

1. $Y_{n,m}$ für alle $n \in \mathbb{N}, 1 \leq m \leq k_n + n + 1$ messbar bezüglich $\mathcal{F}_{n,m}$ ist und
2. $\mathbf{E}[Y_{n,m} | \mathcal{F}_{n,m-1}] = 0$ für alle $n \geq 1, 1 \leq m \leq k_n + n + 1$ sowie
3. $Y_{n,0}$ ist $\mathcal{F}_{n,0}$ messbar für alle $n \in \mathbb{N}$ gilt.

Satz 3.21 ([HS, Theorem 3.1]). Es sei $(Y_{n,m}, \mathcal{F}_{n,m})_{n \in \mathbb{N}_0, 0 \leq m \leq k_n+n+1}$ eine Martingal-Differenz-Liste. Wir setzen

$$\xi_n := \sum_{j=0}^{k_n+n+1} Y_{n,j}$$

und

$$V_n := \sum_{i=1}^{k_n+n+1} \mathbf{E} \left[Y_{n,i}^2 | \mathcal{F}_{n,i-1} \right].$$

Gilt für eine Nullfolge $(\varepsilon_n)_n$

1. $\frac{|Y_{n,m}|^2}{\mathbf{E}[V_n]} \leq \varepsilon_n$ fast sicher für alle $m \in \{1, \dots, k_n + n + 1\}$,

$$2. \frac{\mathbf{Var} Y_{n,0}}{\mathbf{E}[V_n]} \leq \varepsilon_n \text{ für } n \text{ groß genug,}$$

und

$$3. \frac{V_n}{\mathbf{E}[V_n]} \xrightarrow{\mathbf{P}} 1,$$

dann gilt

$$\frac{\xi_n}{\sqrt{\mathbf{E}[V_n]}} \xrightarrow{(d)} \mathcal{N}(0, 1).$$

Intuitiv können die Bedingungen 2 und 3 über die Zerlegung der Varianz von ξ_n in $\mathbf{E}[V_n]$ und $\mathbf{Var}[Y_{n,0}]$ verstanden werden. Voraussetzung 1 stellt eine Forderung an die gleichmäßige Kleinheit der Summanden $1, \dots, k_n + n + 1$ von ξ_n dar. Voraussetzung 2 stellt dann eine Bedingung an die Kleinheit der *Startvarianz* $\mathbf{Var}[Y_{n,0}]$ dar, Voraussetzung 3 verlangt eine Form von Konzentration von V_n und besagt damit intuitiv, dass *gleichmäßig viel Varianz produziert wird*.

Wir erinnern daran, dass $X_{k,l}$ die zufällige Farbe beziehungsweise den zufälligen Typ $\in \{\pm 1\}$ von Individuum $l \in \{1, \dots, N\}$ in Generation $k \in \mathbb{Z}$ bezeichnet und setzen

$$Z_n^m := \mathbf{E} \left[Z_n | \sigma \left((X_{k,l})_{l \in \{1, \dots, N\}} : k \in \mathbb{Z}, k < m \right) \right]$$

und

$$X_{n,l}^m := \mathbf{E} \left[X_{n,l} | \sigma \left((X_{k,l})_{l \in \{1, \dots, N\}} : k \in \mathbb{Z}, k < m \right) \right]$$

sowie

$$Y_{n,0} := Z_n^{-k_n}, \quad Y_{n,i} := Z_n^{-k_n+i} - Z_n^{-k_n+i-1}, \quad n \geq 1, 1 \leq i \leq k_n + n + 1$$

und für $n \geq 1, 1 \leq i \leq k_n + n + 1$

$$\mathcal{F}_{n,0} := \sigma_{-k_n} := \sigma \left((X_{k,l})_{l \in \{1, \dots, N\}} : k \in \mathbb{Z}, k < -k_n \right),$$

$$\mathcal{F}_{n,i} := \sigma_{-k_n+i} := \sigma \left((X_{k,l})_{l \in \{1, \dots, N\}} : k \in \mathbb{Z}, k < -k_n + i \right).$$

Zuerst bemerken wir, dass nach [BGKS, Beweis von Lemma 5]

$$\sigma_{-\infty} := \bigcap_{n \in \mathbb{Z}} \sigma_n \tag{3.61}$$

trivial ist, das heißt für $A \in \sigma_{-\infty}$ gilt $\mathbf{P}(A) \in \{0, 1\}$. Die $Y_{n,0}, \dots, Y_{n,k_n+n+1}$ stellen sukzessive Prognosedifferenzen von Z_n auf Basis der Färbungen der Individuen bis

zur Zeit $-k_n - 1, \dots, -(k_n + n + 1) - 1$ dar. Insbesondere werden wir später $(k_n)_n$ so wählen, dass zu einer gegebenen Nullfolge $(\varepsilon_n)_n$

$$\mathbf{Var} \left[Z^{-k_n} \right] \leq \frac{\varepsilon_n}{2} \mathbf{Var}[Z_n]$$

gilt. Nach Definition gilt außerdem

$$\begin{aligned} \xi_n &:= \sum_{j=0}^{k_n+n+1} Y_{n,j} \\ &= Z_n^{-k_n} + \sum_{j=1}^{k_n+n+1} \left[Z_n^{-k_n+j} - Z_n^{-k_n+j-1} \right] \\ &= Z_n^{n+1} \\ &= Z_n. \end{aligned} \tag{3.62}$$

Zuerst wollen wir nun die Erwartung V_n darstellen durch die Varianz einer bedingten Erwartung von Z_n sowie Z_n selbst und dann die Martingaldifferenzen $Z_n^{m+1} - Z_n^m$ in ihre Grundbausteine aus $X_{k,l}$ sowie bedingte Erwartungen von $X_{k,l}$ zerlegen.

$$\begin{aligned} V_n &= \sum_{i=1}^{k_n+n+1} \mathbf{E} \left[Y_{n,i}^2 | \mathcal{F}_{n,i-1} \right] \\ &= \sum_{i=1}^{k_n+n+1} \mathbf{E} \left[\left(Z_n^{-k_n+i} - Z_n^{k_n+i-1} \right)^2 | \mathcal{F}_{n,i-1} \right] \\ &= \sum_{i=1}^{k_n+n+1} \mathbf{Var} \left[\left(Z_n^{-k_n+i} - Z_n^{k_n+i-1} \right) | \mathcal{F}_{n,i-1} \right]. \end{aligned}$$

also wegen der Unkorreliertheit der $Y_{n,m}, Y_{n,m'}$ für $m \neq m'$ und $\mathbf{E} [Y_{n,m}] = 0$

$$\begin{aligned} \mathbf{E} [V_n] &= \sum_{i=1}^{k_n+n+1} \mathbf{E} \left[\mathbf{Var} \left[\left(Z_n^{-k_n+i} - Z_n^{k_n+i-1} \right) | \mathcal{F}_{n,i-1} \right] \right] \\ &= \sum_{i=1}^{k_n+n+1} \mathbf{Var} \left[\left(Z_n^{-k_n+i} - Z_n^{k_n+i-1} \right) \right] \\ &= \mathbf{Var} \left[\sum_{i=1}^{k_n+n+1} \left(Z_n^{-k_n+i} - Z_n^{k_n+i-1} \right) \right] \end{aligned}$$

$$= \mathbf{Var} \left[Z_n - Z_n^{-k_n} \right]. \quad (3.63)$$

Es gilt für $M < n$ und alle $l \in \{1, \dots, N\}$

$$\begin{aligned} & X_{n,l}^{M+1} - X_{n,l}^M \\ = & q_{n-M} \cdot \left[\underbrace{\mathbf{E} \left[X_{n,l} | \sigma_{M+1}, (n, l) \text{ hat Vorfahr in Generation } M \right]}_{\frac{1}{N} \sum_{l=1}^N X_{M,l}} \right. \\ & \left. - \mathbf{E} \left[X_{n,l} | \sigma_M, (n, l) \text{ hat Vorfahr in Generation } M \right] \right] \\ & + (1 - q_{n-M}) \\ & \cdot \left[\underbrace{\mathbf{E} \left[X_{n,l} | \sigma_{M+1}, (n, l) \text{ hat keinen Vorfahr in Generation } M \right]}_{=\mathbf{E} \left[X_{n,l} | \sigma_M, (n, l) \text{ hat keinen Vorfahr in Generation } M \right]} \right. \\ & \left. - \mathbf{E} \left[X_{n,l} | \sigma_M, (n, l) \text{ hat keinen Vorfahr in Generation } M \right] \right] \\ = & q_{n-M} \left[\frac{1}{N} \sum_{l=1}^N X_{M,l} - \mathbf{E} \left[X_{M,l} | \sigma_M \right] \right] \\ = & q_{n-M} \left[\frac{1}{N} \sum_{l=1}^N X_{M,l} - X_{M,l}^M \right] \end{aligned}$$

Da gegeben σ_M die $X_{M,1}, \dots, X_{M,k}$ identisch verteilt sind, gilt

$$X_{M,l}^M = \frac{1}{N} \sum_{k=1}^N X_{M,k}^M$$

und damit folgt dann insbesondere für $M < n$ und alle $l \in \{1, \dots, N\}$

$$X_{n,l}^{M+1} - \frac{1}{N} \sum_{k=1}^N X_{n,k}^M = q_{n-M} \left[\frac{1}{N} \sum_{k=1}^N X_{M,k} - \frac{1}{N} \sum_{k=1}^N X_{M,k}^M \right] = q_{n-M} \zeta_M$$

für

$$\zeta_M := \frac{1}{N} \sum_{k=1}^N X_{M,k} - \frac{1}{N} \sum_{k=1}^N X_{M,k}^M,$$

wobei die Differenz ζ_m sich wie eben vorgerechnet im Ereignis, dass (n, l) seinen Ahnen in Generationen M hat, begründet. Damit erhalten wir für $n \geq 1, m < 0$

$$Z_n^{m+1} - Z_n^m = \sum_{i=1}^n \left[\frac{1}{N} \sum_{k=1}^N X_{i,k}^{m+1} - \frac{1}{N} \sum_{k=1}^N X_{i,k}^m \right]$$

$$\begin{aligned}
 &= \sum_{i=1}^n \left[\frac{1}{N} \sum_{k=1}^N X_{i,k}^{m+1} \right] - \sum_{i=1}^n \left[\frac{1}{N} \sum_{k=1}^N X_{i,k}^m \right] \\
 &= \frac{1}{N} \sum_{k=1}^N \sum_{i=1}^n X_{i,k}^{m+1} - \frac{1}{N} \sum_{k=1}^N \sum_{i=1}^n \frac{1}{N} \sum_{l=1}^N X_{i,l}^m \\
 &= \frac{1}{N} \sum_{k=1}^N \sum_{i=1}^n \left[X_{i,k}^{m+1} - \frac{1}{N} \sum_{l=1}^N X_{i,l}^m \right] \\
 &= \frac{1}{N} \sum_{k=1}^N \sum_{i=1}^n q_{i-m} \zeta_m \\
 &= \zeta_m \sum_{i=1}^n q_{i-m} = \zeta_m F_n^m \tag{3.64}
 \end{aligned}$$

für

$$F_n^m := \sum_{i=1}^n q_{i-m}. \tag{3.65}$$

Eine ähnliche Rechnung liefert eine ähnliche Form für $Z_n^{m+1} - Z_n^m$ für $n \geq 1$ und $m \in \{0, \dots, n-1\}$:

$$\begin{aligned}
 Z_n^{m+1} - Z_n^m &= \sum_{i=1}^n \left[\frac{1}{N} \sum_{k=1}^N X_{i,k}^{m+1} - \frac{1}{N} \sum_{k=1}^N X_{i,k}^m \right] \\
 &= \sum_{i=1}^n \left[\frac{1}{N} \sum_{k=1}^N X_{i,k}^{m+1} \right] - \sum_{i=1}^n \left[\frac{1}{N} \sum_{k=1}^N X_{i,k}^m \right] \\
 &= \frac{1}{N} \sum_{k=1}^N \sum_{i=1}^n X_{i,k}^{m+1} - \frac{1}{N} \sum_{k=1}^N \sum_{i=1}^n \frac{1}{N} \sum_{l=1}^N X_{i,l}^m \\
 &= \frac{1}{N} \sum_{k=1}^N \sum_{i=1}^n \left[X_{i,k}^{m+1} - \frac{1}{N} \sum_{l=1}^N X_{i,l}^m \right] \\
 &= \frac{1}{N} \sum_{i=m}^n \sum_{k=1}^N \zeta_m q_{i-m} \\
 &= \sum_{i=m}^n \zeta_m q_{i-m}
 \end{aligned}$$

$$= \zeta_m \sum_{i=0}^{n-m} q_i. \quad (3.66)$$

$X_{m,l}$ ist für $l \in \{1, \dots, N\}$ nach Definition $\{\pm 1\}$ -wertig und $\frac{1}{N} \sum_{k=1}^N X_{m,k}^m$ definitionsgemäß σ_m -messbar, womit wir

$$\mathbf{Var} \left[X_{m,l} - \frac{1}{N} \sum_{k=1}^N X_{m,k}^m \middle| \sigma_m \right] = 4P_m(1 - P_m)$$

für

$$P_m := \mathbf{P}(X_{m,1} = 1 | \sigma_m) = \dots = \mathbf{P}(X_{m,N} = 1 | \sigma_m) \quad (3.67)$$

erhalten.

Wir haben damit die Komponenten des Konvergenzsatzes schon in eine leicht verwertbare und analysierbare Form gebracht. Nach und nach wollen wir nun die Voraussetzungen von Satz 3.21 nachrechnen.

Lemma 3.22. *Zu jeder gegebenen Nullfolge $(\varepsilon_n)_n$ erfüllt $(Y_{n,m}, \mathcal{F}_{n,m})_{n \in \mathbb{N}_0, 0 \leq m \leq k_n + n + 1}$ Bedingung 2 von Satz 3.21 für eine passende Folge $(k_n)_n$.*

Beweis. Für $k \rightarrow \infty$ gilt wegen $|Z_n| \leq n$ nach dem Rückwärtsmartingalkonvergenzsatz

$$Z_n^{-k} = \mathbf{E}[Z_n | \sigma_{-k}] \rightarrow \mathbf{E}[Z_n | \sigma_{-\infty}] \text{ f.s. und in } \mathcal{L}_2 \quad (3.68)$$

und da $\sigma_{-\infty}$ trivial ist (siehe (3.61)), gilt

$$\mathbf{E}[Z_n | \sigma_{-\infty}] = \mathbf{E}[Z_n] \text{ f.s. für } m \rightarrow \infty.$$

Damit gilt

$$\mathbf{Var} \left[\frac{\mathbf{E}[Z_n | \sigma_{-k_m}]}{\sqrt{\mathbf{Var}[Z_n]}} \right] \rightarrow 0 \text{ für } m \rightarrow \infty.$$

Damit existiert also zu einer gegebenen Nullfolge $(\varepsilon_n)_n$ eine Folge $(k_n)_{n \in \mathbb{N}}$ mit

$$\mathbf{Var}[Z_n^{-k_n}] \leq \frac{\varepsilon_n}{2} \mathbf{Var}[Z_n] \quad (3.69)$$

für n groß genug. Sei ab jetzt

$$(\varepsilon)_n \text{ fest aber beliebig und } (k_n)_n \text{ so dass (3.69) gilt.} \quad (3.70)$$

Nach (3.63) gilt dafür dann

$$\mathbf{Var}[Z_n^{-k_n}] \leq \frac{\varepsilon_n}{2} \left[\mathbf{Var}[Z_n^{-k_n}] + \mathbf{E}[V_n] \right],$$

und damit

$$\mathbf{Var} \left[Z_n^{-k_n} \right] \leq \frac{\varepsilon_n}{2} \mathbf{Var} [Z_n] = \frac{\varepsilon_n}{2} \left[\mathbf{E}[V_n] + \mathbf{Var} \left[Z_n^{-k_n} \right] \right],$$

also

$$\left(1 - \frac{\varepsilon_n}{2} \right) \mathbf{Var} \left[Z_n^{-k_n} \right] \leq \frac{\varepsilon_n}{2} \mathbf{E}[V_n]$$

für n groß genug. Damit existiert ein $N_0 \in \mathbb{N}$ so, dass für $n \geq N_0$

$$\mathbf{Var} \left[Z_n^{-k_n} \right] \leq \varepsilon_n \mathbf{E}[V_n]$$

gilt. Nach Definition von $Y_{n,0}$ (mit $Y_{n,0} := Z_n^{-k_n}$) folgt damit die Behauptung. \square

Lemma 3.23. $(Y_{n,m}, \mathcal{F}_{n,m})_{n \in \mathbb{N}_0, 0 \leq m \leq k_n + n + 1}$ erfüllt Bedingung 3 von Satz 3.21 für eine Folge $(k_n)_{n \in \mathbb{N}}$ mit $k_n \geq Cn$ für ein $C > 1$.

Beweis. Zur Analyse von V_n stellen wir diese Zufallsvariable etwas anders und mittels der F_m aus (3.67) dar, dabei sei F_n^m weiterhin definiert wie in (3.65):

$$\begin{aligned} V_n &= \sum_{i=1}^{k_n+n+1} \mathbf{E} \left[Y_{n,i}^2 | \mathcal{F}_{n,i-1} \right] \\ &= \sum_{i=1}^{k_n+n+1} \mathbf{Var} \left[Y_{n,i} | \mathcal{F}_{n,i-1} \right] \\ &= \sum_{i=1}^{k_n+n+1} \mathbf{Var} \left[Z_n^{-k_n+i} - Z_n^{-k_n+i-1} | \sigma_{-k_n+i-1} \right] \\ &= \sum_{i=1}^{k_n} \mathbf{Var} \left[Z_n^{-k_n+i} - Z_n^{-k_n+i-1} | \sigma_{-k_n+i-1} \right] \\ &\quad + \sum_{i=k_n+1}^{k_n+n+1} \mathbf{Var} \left[Z_n^{-k_n+i} - Z_n^{-k_n+i-1} | \sigma_{-k_n+i-1} \right] \\ &= \sum_{m=-k_n}^{-1} \mathbf{Var} \left[Z_n^{m+1} - Z_n^m | \sigma_m \right] + \sum_{m=0}^n \mathbf{Var} \left[Z_n^{m+1} - Z_n^m | \sigma_m \right] \\ &= \sum_{m=-k_n}^{-1} \mathbf{Var} \left[\zeta_m F_n^m | \sigma_m \right] + \sum_{m=0}^n \mathbf{Var} \left[Z_n^{m+1} - Z_n^m | \sigma_m \right] \\ &= \sum_{m=-k_n}^{-1} \mathbf{Var} \left[\zeta_m F_n^m | \sigma_m \right] + \sum_{m=0}^n \mathbf{Var} \left[\zeta_m \sum_{i=0}^{n-m} q_i | \sigma_m \right] \end{aligned}$$

$$\begin{aligned}
 &= \sum_{m=-k_n}^{-1} (F_n^m)^2 \underbrace{\mathbf{Var} \left[\frac{1}{N} \sum_{k=1}^N X_{m,k} - \frac{1}{N} \sum_{k=1}^N X_{m,k}^m \middle| \sigma_m \right]}_{= \frac{1}{N^2} N 4 P_m (1 - P_m) = \frac{4}{N} P_m (1 - P_m)} \\
 &+ \sum_{m=0}^n \left(\sum_{i=0}^{n-m} q_i \right)^2 \underbrace{\mathbf{Var} \left[\frac{1}{N} \sum_{k=1}^N X_{m,k} - \frac{1}{N} \sum_{k=1}^N X_{m,k}^m \middle| \sigma_m \right]}_{= \frac{4}{N} P_m (1 - P_m)} \\
 &= \frac{4}{N} \sum_{m=-k_n}^{-1} (F_n^m)^2 P_m (1 - P_m) \\
 &+ \frac{4}{N} \sum_{m=0}^n \left(\sum_{i=0}^{n-m} q_i \right)^2 P_m (1 - P_m) \tag{3.71}
 \end{aligned}$$

Um das weiter zu analysieren, stellen wir zuerst für $l \in \{1, \dots, N\}$

$$\mathbf{Var} \left[X_{i,l} - \frac{1}{N} \sum_{k=1}^N X_{i,k}^i \middle| \sigma_i \right] = 4P_i(1 - P_i)$$

fest und setzen

$$\rho_{i,j} := \mathbf{E} [P_i(1 - P_i)P_j(1 - P_j)] - \mathbf{E}[P_i(1 - P_i)] \mathbf{E}[P_j(1 - P_j)]$$

und zeigen, dass für alle $\varepsilon > 0$ ein $K_0 \in \mathbb{N}$ existiert, sodass

$$|i - j| \geq K_0 \Rightarrow |\rho_{i,j}| < \varepsilon \tag{3.72}$$

gilt. Für $i \in \mathbb{Z}, k \in \mathbb{N}$ setzen wir

$$G_{i,k} := \sum_{j=1}^k \frac{\mu(\{j\})}{N} \left[\sum_{k=1}^N X_{i-j,k} \right]$$

und approximieren damit im folgenden Sinne $X_{i,l}^i$ für $l \in \{1, \dots, N\}$:

$$|X_{i,l}^i - G_{i,k}| \leq \sum_{j \geq k+1} \mu(\{j\}) = (k+1)^{-\alpha} L(k+1).$$

Wir können dann zeigen, dass für (3.72) bereits ausreichend ist, dass für alle $\varepsilon > 0, k \in \mathbb{N}$ ein $n_0 \in \mathbb{N}$ existiert, sodass für alle $i, j \in \mathbb{Z}$ für das durch

$$R_{i,j,k} := \left\{ \left(\bigcup_{m=i-k}^{i-1} \bigcup_{l=1}^N A_{m,l} \right) \cap \left(\bigcup_{n=j-k}^{j-1} \bigcup_{l=1}^N A_{n,l} \right) = \emptyset \right\}$$

definierte Ereignis

$$|i - j| \geq n_0 \Rightarrow \mathbf{P} \left(R_{i,j,k}^c \right) < \varepsilon \quad (3.73)$$

gilt, wobei die $A_{m,l}$ für $m \in \mathbb{Z}$ und $l \in \{1, \dots, N\}$ die Ahnenlinie von (m, l) bezeichnen. $R_{i,j,k}$ ist damit das Ereignis, dass die Ahnenlinien der Generationen $i - k$ bis $i - 1$ keine der Ahnenlinien der Generationen $j - k$ bis $j - 1$ treffen. Wir zeigen zunächst, dass (3.73) ausreichend für (3.72) ist. Angenommen, dass (3.73) gilt. Seien $i, j \in \mathbb{Z}, k \in \mathbb{N}$ und $|i - j| \geq k$ und

Y eine bezüglich $\sigma(X_{i-k,1}, \dots, X_{i-k,N}, \dots, X_{i-1,1}, \dots, X_{i-1,N})$ messbare ZV

sowie

Z eine bezüglich $\sigma(X_{j-k,1}, \dots, X_{j-k,N}, \dots, X_{j-1,1}, \dots, X_{j-1,N})$ messbare ZV,

dann gilt

$$\mathbf{E}[YZ|R_{i,j,k}] = \mathbf{E}[Y|R_{i,j,k}] \cdot \mathbf{E}[Z|R_{i,j,k}].$$

Damit und mit

$$P_i = \mathbf{P}(X_{i,1} = 1|\sigma_i) = \dots = \mathbf{P}(X_{i,N} = 1|\sigma_i) = \frac{X_{i,1}^i + 1}{2} = \dots = \frac{X_{i,N}^i + 1}{2}$$

folgt für $|i - j| \geq n_0$ unter Beachtung von $k^{-\alpha}L(k) \geq (k + 1)^{-\alpha}L(k + 1)$

$$\begin{aligned} & \mathbf{E}[P_i(1 - P_i)P_j(1 - P_j)] \\ &= \mathbf{E}\left[\frac{X_{i,1}^i + 1}{2} \cdot \frac{-X_{i,1}^i + 1}{2} \cdot \frac{X_{j,1}^j + 1}{2} \cdot \frac{-X_{j,1}^j + 1}{2}\right] \\ &= \mathbf{E}\left[\frac{1 + G_{i,k}}{2} \cdot \frac{1 - G_{i,k}}{2} \cdot \frac{1 + G_{j,k}}{2} \cdot \frac{1 - G_{j,k}}{2}\right] + O(k^{-\alpha}L(k)) \\ &= \mathbf{E}\left[\frac{1 + G_{i,k}}{2} \cdot \frac{1 - G_{i,k}}{2} \cdot \frac{1 + G_{j,k}}{2} \cdot \frac{1 - G_{j,k}}{2} \middle| R_{i,j,k}\right] \\ & \quad + O\left(\mathbf{P}\left(R_{i,j,k}^c\right)\right) + O(k^{-\alpha}L(k)) \\ &= \mathbf{E}\left[\frac{1 + G_{i,k}}{2} \cdot \frac{1 - G_{i,k}}{2} \middle| R_{i,j,k}\right] \cdot \mathbf{E}\left[\frac{1 + G_{j,k}}{2} \cdot \frac{1 - G_{j,k}}{2} \middle| R_{i,j,k}\right] \\ & \quad + O\left(\mathbf{P}\left(R_{i,j,k}^c\right)\right) + O(k^{-\alpha}L(k)) \\ &= \left[\mathbf{E}\left[\frac{1 + G_{i,k}}{2} \cdot \frac{1 - G_{i,k}}{2}\right] + O\left(\mathbf{P}\left(R_{i,j,k}^c\right)\right)\right] \end{aligned}$$

$$\begin{aligned}
 & \cdot \left[\mathbf{E} \left[\frac{1 + G_{j,k}}{2} \cdot \frac{1 - G_{j,k}}{2} \right] + O \left(\mathbf{P} \left(R_{i,j,k}^c \right) \right) \right] \\
 & + O \left(\mathbf{P} \left(R_{i,j,k}^c \right) \right) + O \left(k^{-\alpha} L(k) \right) \\
 = & \mathbf{E} \left[\frac{1 + G_{i,k}}{2} \cdot \frac{1 - G_{i,k}}{2} \right] \cdot \mathbf{E} \left[\frac{1 + G_{j,k}}{2} \cdot \frac{1 - G_{j,k}}{2} \right] \\
 & + O \left(\mathbf{P} \left(R_{i,j,k}^c \right) \right) + O \left(k^{-\alpha} L(k) \right).
 \end{aligned}$$

Damit existiert also für alle $\varepsilon > 0$ ein $n_0 \in \mathbb{N}$ so, dass für alle $i, j \in \mathbb{Z}$ mit $|i - j| \geq n_0$

$$\mathbf{Cov} [P_i(1 - P_i), P_j(1 - P_j)] < \varepsilon$$

gilt. Wir müssen also noch (3.73) zeigen. Dafür genügt es zu zeigen, dass für alle $\varepsilon > 0$ ein $n_0 \in \mathbb{N}$ existiert so dass für alle $i, j \in \mathbb{Z}$ mit $|i - j| \geq n_0$

$$\mathbf{P} (A_{i,r_1} \cap A_{j,r_2} \neq \emptyset) < \varepsilon \quad \forall r_1, r_2 \in \{1, \dots, N\}$$

gilt. Seien also $i < j$ und $r_1, r_2 \in \{1, \dots, N\}$, dann gilt ($A_{l,m}$ seien wieder die Couplings der Ahnenlinien, die beim Zusammenlaufen nicht verschmelzen)

$$\begin{aligned}
 & \mathbf{P} (A_{i,r_1} \cap A_{j,r_2} \neq \emptyset) \\
 & \leq \mathbf{E} [|A_{i,r_1} \cap A_{j,r_2}|] \\
 = & \sum_{n=0}^{\infty} \frac{1}{N} q_n q_{n+i-j} \leq \frac{1}{N} \left(\sum_{n \geq 0} q_n^2 \right)^{\frac{1}{2}} \left(\sum_{n \geq 0} q_{n+i-j}^2 \right)^{\frac{1}{2}} \tag{3.74}
 \end{aligned}$$

und (3.74) hängt offenbar nicht von k ab. Wir können also (3.74) für $|i - j|$ groß genug durch $\frac{\varepsilon}{k^2 N^3}$ abschätzen, da

$$\left(\sum_{n \geq 0} q_n^2 \right)^{\frac{1}{2}} < \infty$$

und somit für $|i - j| \rightarrow \infty$

$$\left(\sum_{n \geq 0} q_{n+i-j}^2 \right)^{\frac{1}{2}} \rightarrow 0$$

gilt. Das liefert uns

$$\begin{aligned}
 \mathbf{P} \left(R_{i,j,k}^c \right) &\leq \sum_{m_1=1}^N \sum_{m_2=1}^N \sum_{r_1=i-k}^{i-1} \sum_{r_2=j-k}^{j-1} \mathbf{P} \left(A_{r_1,m_1} \cap A_{r_2,m_2} \neq \emptyset \right) \\
 &< N^2 k^2 \cdot \frac{\varepsilon}{k^2 N^3} \\
 &\leq \frac{\varepsilon}{N}.
 \end{aligned} \tag{3.75}$$

Damit gilt also (3.73), also nach dem eben gezeigten auch (3.72) und wir können den letzten Schritt zum Beweis des Lemmas machen. Dazu setzen wir

$$\beta_n := \sum_{M=-k_n}^{-1} P_M (1 - P_M) (F_M^n)^2.$$

Wir wollen Lemma 2.16 mit

$$a_n := k_n, \quad M_i := P_{i-1} (1 - P_{i-1}), \quad c_i^n := (F_{i-1}^n)^2 \quad \text{für } i \in \{0, \dots, k_n - 1\}$$

anwenden. Dafür bemerken wir, dass nach [HS, Gleichung (3.15)] für eine Konstante $C > 0$

$$\frac{\max_{-k_n < M \leq 0} (F_{i-1}^n)^2}{\sum_{M=-\lfloor Cn \rfloor}^0 (F_{i-1}^n)^2} \leq \frac{C^2}{n} \tag{3.76}$$

gilt und damit für eine Folge $k_n \geq Cn$ die Voraussetzungen von Lemma 2.16 erfüllt sind (siehe [HS, Gleichung (3.14)]). In den obigen Rechnungen haben wir keine Annahme an das Wachstum von k_n gemacht, wir können also ohne Beschränkung der Allgemeinheit von jetzt an $k_n \geq Cn$ wählen. Bis zum Ende des Beweises sei also

$$(k_n)_{n \in \mathbb{N}} \text{ eine Folge mit } k_n \geq Cn. \tag{3.77}$$

Lemma 2.16 lässt sich folglich anwenden. Damit erhalten wir also

$$\mathbf{Var} [\beta_n] = o \left((\mathbf{E}[\beta_n])^2 \right) \quad \text{für } n \rightarrow \infty$$

und dementsprechend

$$\mathbf{P} \left((\beta_n - \mathbf{E}[\beta_n])^2 \geq \varepsilon (\mathbf{E}[\beta_n])^2 \right) \rightarrow 0 \quad \text{für } n \rightarrow \infty \quad \text{für alle } \varepsilon > 0,$$

also

$$\mathbf{P} \left(\frac{(\beta_n - \mathbf{E}[\beta_n])^2}{(\mathbf{E}[\beta_n])^2} \geq \varepsilon \right) \rightarrow 0 \quad \text{für } n \rightarrow \infty \quad \text{für alle } \varepsilon > 0,$$

Das heißt $\frac{(\beta_n - \mathbf{E}[\beta_n])^2}{(\mathbf{E}[\beta_n])^2}$ konvergiert in Wahrscheinlichkeit gegen 0. Für

$$c_0 := \mathbf{E}[P_0(1 - P_0)]$$

liefert das also für eine Folge von Zufallsvariablen $(E_n^{(1)})_{n \in \mathbb{N}}$ mit

$$E_n^{(1)} \xrightarrow{\mathbf{P}} 0 \text{ für } n \rightarrow \infty$$

nach Definition

$$\beta_n = (1 + E_n^{(1)})c_0 \sum_{M=-k_n}^{-1} (F_M^n)^2. \quad (3.78)$$

Analog verfahren wir für

$$\gamma_n := \sum_{m=0}^n \left(\sum_{i=0}^{n-m} q_i \right)^2 P_m(1 - P_m),$$

setzen

$$\tilde{F}_m^n := \sum_{i=0}^{n-m} q_i$$

und wollen erneut Lemma 2.16 mit

$$a_n := n, \quad c_m^n := \tilde{F}_m^n, \quad M_m := P_m(1 - P_m) \text{ für } m \in \{0, \dots, n\}$$

anwenden. Die Voraussetzungen von Lemma 2.16 sind in der Tat nach einer zu (3.76) analogen Rechnung (siehe [HS, Gleichung (3.14)]) erfüllt. Wir können das Lemma also erneut anwenden und erhalten analog zu vorhin

$$\mathbf{Var}[\gamma_n] = o\left(\left(\mathbf{E}[\gamma_n]\right)^2\right) \text{ für } n \rightarrow \infty$$

und dementsprechend

$$\mathbf{P}\left(\left(\gamma_n - \mathbf{E}[\gamma_n]\right)^2 \geq \varepsilon \left(\mathbf{E}[\gamma_n]\right)^2\right) \rightarrow 0 \text{ für alle } \varepsilon > 0 \text{ und } n \rightarrow \infty,$$

also

$$\mathbf{P}\left(\frac{(\gamma_n - \mathbf{E}[\gamma_n])^2}{(\mathbf{E}[\gamma_n])^2} \geq \varepsilon\right) \rightarrow 0 \text{ für alle } \varepsilon > 0 \text{ und } n \rightarrow \infty.$$

$\frac{(\gamma_n - \mathbf{E}[\gamma_n])^2}{(\mathbf{E}[\gamma_n])^2}$ konvergiert dementsprechend in Wahrscheinlichkeit gegen null. Für

$$c_0 := \mathbf{E} [P_0(1 - P_0)]$$

liefert das also für eine Folge von Zufallsvariablen $(E_n^{(2)})_{n \in \mathbb{N}}$ mit

$$E_n^{(2)} \xrightarrow{\mathbf{P}} 0 \text{ für } n \rightarrow \infty$$

ebenso wie vorhin

$$\gamma_n = (1 + E_n^{(2)})c_0 \sum_{M=0}^n \left(\tilde{F}_M^n \right)^2. \quad (3.79)$$

Nach (3.71) erhalten wir also

$$\begin{aligned} V_n &= \frac{4}{N} \sum_{m=-k_n}^{-1} (F_n^m)^2 P_m(1 - P_m) + \frac{4}{N} \sum_{m=0}^n \left(\sum_{i=0}^{n-m} q_i \right)^2 P_m(1 - P_m) \\ &= \frac{4}{N} \beta_n + \frac{4}{N} \gamma_n \\ &= \frac{4}{N} (1 + E_n^{(1)})c_0 \sum_{M=-k_n}^{-1} (F_M^n)^2 + \frac{4}{N} (1 + E_n^{(2)})c_0 \sum_{M=0}^n \left(\tilde{F}_M^n \right)^2, \end{aligned} \quad (3.80)$$

womit offenbar

$$\frac{V_n}{\mathbf{E}[V_n]} \xrightarrow{\mathbf{P}} 1$$

gilt. Damit ist Voraussetzung 3 wie behauptet erfüllt. \square

Wir widmen uns der letzten verbliebenen Bedingung:

Lemma 3.24. $(Y_{n,m}, \mathcal{F}_{n,m})_{n \in \mathbb{N}_0, 0 \leq m \leq k_n + n + 1}$ erfüllt Bedingung 1 von Satz 3.21 zu einer Nullfolge $(\tilde{\varepsilon}_n)_n$ und einer Folge $(k_n)_{n \in \mathbb{N}}$ mit $k_n \rightarrow \infty$.

Beweis. Noch zu zeigen ist also, dass für alle $m \in \{1, \dots, k_n + n\}$

$$\frac{|Y_{n,m}|^2}{\mathbf{E}[V_n]} \leq \tilde{\varepsilon}_n$$

für eine Nullfolge $(\tilde{\varepsilon}_n)_n$ gilt. Für $-k_n < M < 0$ gilt nach (3.78), (3.79) und (3.66)

$$|Z_n^{M+1} - Z_n^M|^2 = |\zeta_M(F_M^n)|^2$$

$$\begin{aligned}
 &\leq (F_M^n)^2 \underbrace{\left| \frac{1}{N} \sum_{k=1}^N X_{M,k} - \frac{1}{N} \sum_{k=1}^N X_{M,k}^M \right|}_{\leq 4}^2 \\
 &\leq 4(F_M^n)^2 \\
 &= o(\mathbf{E}[V_n]).
 \end{aligned} \tag{3.81}$$

Für $0 \leq M < n$ gilt nach (3.78) und Lemma 2.4 für eine Konstante c

$$\begin{aligned}
 |Z_n^{M+1} - Z_n^M|^2 &= \left| \zeta_m \sum_{l=0}^{n-M} q_l \right|^2 \\
 &\leq 4 \left(\sum_{l=0}^{n-M} q_l \right)^2 \\
 &\leq 4c^2 n^{2\alpha} L(n)^{-2}.
 \end{aligned} \tag{3.82}$$

Sei

$$k_n := \max \left\{ k_n^{(1)}, k_n^{(2)} \right\} \tag{3.83}$$

für die in (3.70) und (3.77) definierten divergenten Folgen $k_n^{(1)}, k_n^{(2)}$. Nach (3.69) und Lemma 2.11 zusammen mit der Asymptotik aus Lemma 2.7 existiert eine Konstante c_0 , sodass für die Nullfolge $(\varepsilon_n)_n$ aus (3.70)

$$\begin{aligned}
 \mathbf{E}[V_n] &= \mathbf{Var}[Z_n - Z_n^{-k_n}] \\
 &= \mathbf{Var}[Z_n] - \mathbf{Var}[Z_n^{-k_n}] \\
 &\geq (1 - \varepsilon_n) \mathbf{Var}[Z_n] \\
 &\geq c_0(1 - \varepsilon_n) n^{2\alpha+1} L(n)^{-2}
 \end{aligned} \tag{3.84}$$

gilt. Damit folgt für eine Konstante c_1

$$\begin{aligned}
 \frac{[Y_{n,m}]^2}{\mathbf{E}[V_n]} &= \frac{|Z_n^{-k_n+m} - Z_n^{-k_n+m-1}|^2}{\mathbf{E}[V_n]} \\
 &\leq \frac{4c^2 n^{2\alpha} L(n)^{-2}}{\mathbf{E}[V_n]} \\
 &\leq \frac{c_1}{n}.
 \end{aligned} \tag{3.85}$$

Und da $\frac{c_1}{n} \rightarrow 0$ für $n \rightarrow \infty$ gilt, gilt für

$$\tilde{\varepsilon}_n := \max \left\{ \varepsilon_n, \frac{c_1}{n} \right\} \quad (3.86)$$

und k_n aus (3.83)

$$\frac{|Y_{n,m}|^2}{\mathbf{E}[V_n]} \leq \tilde{\varepsilon}_n \quad \forall m \in \{1, \dots, k_n + n + 1\}. \quad (3.87)$$

□

Nun können wir den Beweis von Proposition 2.12 erbringen.

Beweis von Proposition 2.12. Um Satz 3.21 anzuwenden, verwenden wir die Nullfolge $(\tilde{\varepsilon}_n)_n$ aus (3.86) sowie $(k_n)_n$ aus (3.83). Damit sind die Bedingungen 1, 2 und 3 erfüllt. Es gilt nach (3.62)

$$\xi_n = Z_n$$

und damit dann nach Satz 3.21

$$\frac{Z_n}{\sqrt{\mathbf{E}[V_n]}} = \frac{Z_n}{\sqrt{\mathbf{Var}[Z_n - Z_n^{-k_n}]}} \xrightarrow{(d)} \mathcal{N}(0, 1).$$

Aus (3.68) wissen wir

$$\mathbf{Var}[Z_n^{-k}] = \mathbf{E} \left[\left(Z_n^{-k} - \mathbf{E}[Z_n] \right)^2 \right] \rightarrow 0 \text{ für } k \rightarrow \infty.$$

Damit gilt entlang einer Teilfolge $(\hat{k}_n)_n$ von $(k_n)_n$

$$\mathbf{Var}[Z_n^{-\hat{k}_n}] = \mathbf{E} \left[\left(Z_n^{-\hat{k}_n} - \mathbf{E}[Z_n] \right)^2 \right] \rightarrow 0 \text{ für } n \rightarrow \infty$$

und damit

$$\frac{\mathbf{E}[V_n]}{\mathbf{Var}[Z_n]} = \frac{\mathbf{E}[V_n]}{\mathbf{Var}[Z_n^{-k_n}] + \mathbf{E}[V_n]} \rightarrow 1.$$

Also folgt

$$\frac{Z_n}{\sqrt{\mathbf{Var}[Z_n]}} \xrightarrow{(d)} \mathcal{N}(0, 1).$$

Und da nach Lemma 2.11

$$\frac{\mathbf{Var}[Z_{nt}]}{\sqrt{\mathbf{Var}[Z_n]}} \rightarrow t^{2\alpha+1}$$

gilt, folgt insbesondere

$$\tilde{Z}_t^n \xrightarrow{(d)} \mathcal{N}(0, t^{2\alpha+1})$$

und damit die Konvergenz der endlichdimensionalen Randverteilungen der Prozessfolge \tilde{Z}^n gegen die einer fraktionalen Brownschen Bewegung B^H mit Hurst-Parameter $H = \frac{1}{2} + \alpha$, womit die Behauptung folgt. \square

3.4.3 Funktionale Konvergenz gegen die fraktionale Brownbewegung

Wir erinnern an das Analogon zu Proposition 2.9 an:

Proposition 2.13. *Sei $T \geq 0$. Bezogen auf den Pfadraum $(C([0, T]), \|\cdot\|_{L^\infty([0, T])})$ konvergiert \tilde{Z}^n in Verteilung gegen eine fraktionale Brownsche Bewegung B^H mit Hurst-Parameter $H := \frac{1}{2} + \alpha$.*

Beweis. Wie wir bereits festgestellt haben gilt

$$\mathbf{Var}[\tilde{Z}_t^n] = \mathbf{Var}[\tilde{S}_t^n]$$

und dementsprechend analog zu (3.43) für die dort definierte Konstante c_3

$$\mathbf{Var} \left[\tilde{Z}_u^n - \tilde{Z}_t^n \right] \leq c_3 \left(\frac{l-k}{n} \right)^{2\alpha+1} \left(\frac{L(n)}{L(l-k)} \right)$$

für $0 \leq s \leq t \leq u \leq T$ und

$$\frac{j}{n} \leq s \leq \frac{j+1}{n}, \quad \frac{k}{n} \leq t \leq \frac{k+1}{n}, \quad \frac{l}{n} \leq u \leq \frac{l+1}{n}.$$

Damit folgt die Behauptung sofort mit der gleichen Rechnung wie im Beweis von Proposition 2.9. \square

3.4.4 Couplings für alle Zeithorizonte gemeinsam

Damit lässt sich dann der zu Satz 2.10 analoge Satz 2.14 beweisen:

Satz 2.14. *Für alle $\alpha \in (0, 1/2)$ und $p \in (0, 1)$ existiert eine fraktionale Brownsche Bewegung B^H mit Hurst-Parameter $H := \frac{1}{2} + \alpha$ und eine Folge von Couplings C_n von \tilde{Z}^n und B^H so dass für alle $T > 0$ und $\varepsilon > 0$*

$$\lim_{n \rightarrow \infty} C_n \left(\|\tilde{Z}^n - B^H\|_{L^\infty([0, T])} > \varepsilon \right) = 0$$

gilt.

Beweis. Die Prozesse \tilde{Z}_n und B^H aus Proposition 2.13 übernehmen die Rolle der Prozesses Y^n und Y aus Proposition 3.16. Damit folgt sofort die Behauptung. \square

3.4.5 Grenzwert großer Populationen

Dass der Beweis von Proposition 2.13 nur an einer Stelle von N abhängt, erlaubt es uns, den folgenden Satz schnell zu zeigen:

Satz 2.15. *Für alle $\alpha \in (0, 1/2)$, $r > 1$ und $p \in (0, 1)$ existiert eine fraktionale Brownsche Bewegung B^H mit Hurst-Parameter $H := \frac{1}{2} + \alpha$ und eine Folge von Couplings C_N von $\tilde{Z}^{N^r, N}$ und B^H so dass für alle $T > 0$ und $\varepsilon > 0$*

$$\lim_{N \rightarrow \infty} C_N \left(\|\tilde{Z}^{N^r, N} - B^H\|_{L^\infty([0, T])} > \varepsilon \right) = 0$$

gilt.

Beweis. Nach Lemma 2.11 gilt

$$\mathbf{Var}[Z_n] = \mathbf{Var}[S_n] \cdot \frac{\sum_{l \geq 0} q_l^2}{\sum_{l \geq 1} q_l^2 + N}$$

und damit liefern (3.84) und (3.85), dass (3.87) für $(n_k, N_k)_{k \in \mathbb{N}}$, $n_k, N_k \rightarrow \infty$ gilt, falls

$$\lim_{k \rightarrow \infty} \frac{N_k}{n_k} = 0$$

gilt. Insbesondere ist dies für $n_k := N_k^r$ für $r > 1$ erfüllt. (3.81) bleibt in jedem Fall erfüllt. Damit bleibt Proposition 2.12 erfüllt, da nur die Überprüfung von Bedingung 1 von N abhängig ist. Die funktionale Konvergenz sowie die Existenz gemeinsamer Couplings folgen dann mit Proposition 2.13 sowie Satz 2.14, womit die Behauptung folgt. \square

3.5 Ein Kriterium für Konzentration

Wir beweisen im Folgenden Lemma 2.16, welches auch als Korollar des folgenden Satzes gesehen werden kann:

Satz 2.17. *Es sei $(a_n)_{n \in \mathbb{N}}$ eine Folge natürlicher Zahlen mit $a_n \rightarrow \infty$, $(\varepsilon_l)_{l \in \mathbb{N}}$ eine Nullfolge und $(M_n)_{n \in \mathbb{N}}$ eine Folge von identisch verteilten Zufallsvariablen mit $\mathbf{Var}[M_n] < \infty$, sodass*

$$\mathbf{Cov}[M_n, M_{n+l}] = \varepsilon_l \text{ für alle } n, l \in \mathbb{N} \text{ und eine Nullfolge } (\varepsilon_l)_l$$

gilt. Für Verteilungen π^n auf $\{1, \dots, a_n\}$ mit

$$\max_{1 \leq m \leq a_n} \pi_m^n \rightarrow 0 \text{ für } n \rightarrow \infty$$

gilt

$$\mathbf{Var} \left[\sum_{m=1}^{a_n} \pi_m^n M_m \right] \rightarrow 0 \text{ für } n \rightarrow \infty.$$

Beweis. Sei $\varepsilon > 0$ beliebig. Da $(\varepsilon_l)_{l \in \mathbb{N}}$ eine Nullfolge ist, existiert $N_0 \in \mathbb{N}$, sodass für $k \geq N_0$ $\varepsilon_k < \varepsilon/2$ gilt. Um notationelle Konsistenz zu erzeugen setzen wir

$$\varepsilon_0 := \mathbf{Var} [M_1].$$

Für $n > N_0$ gilt

$$\begin{aligned} & \sum_{i=1}^{a_n} \sum_{j=1}^{a_n} \mathbf{Cov} [M_i, M_j] \pi_i^n \pi_j^n \\ &= 2 \sum_{i=0}^{a_n} \sum_{l=0}^{a_n-i} \underbrace{\mathbf{Cov} [M_i, M_{i+l}]}_{=\varepsilon_l} \pi_i^n \pi_{i+l}^n \\ &= 2 \sum_{i=1}^{a_n} \sum_{l=0}^{N_0-1} \underbrace{\varepsilon_l}_{\leq \max_{l \leq N_0-1} \varepsilon_l =: \delta} \pi_i^n \pi_{i+l}^n + 2 \sum_{i=0}^{a_n} \sum_{l=N_0}^{a_n-i} \varepsilon_l \pi_i^n \pi_{i+l}^n \\ &\leq 2\delta \sum_{i=1}^{a_n} \pi_i^n \underbrace{\sum_{l=0}^{N_0-1} \pi_{i+l}^n}_{\leq N_0 \frac{\max_{j \leq a_n} c_j^n}{\sum_{j=1}^{a_n} c_j^n}} + \varepsilon \sum_{i=1}^{a_n} \pi_i^n \underbrace{\sum_{l=N_0}^{a_n-i} \pi_{i+l}^n}_{\leq 1} \\ &\leq 2\delta N_0 \frac{\max_{j \leq a_n} c_j^n}{\sum_{j=1}^{a_n} c_j^n} \underbrace{\sum_{i=0}^{a_n} \pi_i^n}_{\leq 1} + \varepsilon \underbrace{\sum_{i=1}^{a_n} \pi_i^n}_{\leq 1} \\ &\leq 2\delta N_0 \frac{\max_{j \leq a_n} c_j^n}{\sum_{j=1}^{a_n} c_j^n} + \varepsilon, \\ &\quad \rightarrow 0 \text{ für } n \rightarrow \infty \end{aligned}$$

womit

$$\lim_{n \rightarrow \infty} \mathbf{Var} \left[\sum_{m=1}^{a_n} \pi_m^n M_m \right] < \varepsilon$$

für alle $\varepsilon > 0$ gilt, also wie behauptet

$$\mathbf{Var} \left[\sum_{m=1}^{a_n} \pi_m^n M_m \right] \rightarrow 0 \text{ für } n \rightarrow \infty.$$

□

Wir erinnern an Lemma 2.16:

Lemma 2.16. *Es sei $(a_n)_{n \in \mathbb{N}}$ eine Folge natürlischer Zahlen mit $a_n \rightarrow \infty$, $(\varepsilon_l)_{l \in \mathbb{N}}$ eine Nullfolge und $(M_n)_{n \in \mathbb{N}}$ eine Folge von identisch verteilten Zufallsvariablen mit $\mathbf{E}[M_n] \neq 0$ und $\mathbf{Var}[M_n] < \infty$, sodass*

$$\mathbf{Cov}[M_n, M_{n+l}] = \varepsilon_l \text{ für alle } n, l \in \mathbb{N} \text{ und eine Nullfolge } (\varepsilon_l)_l$$

gilt. Für positive Gewichte $(c_m^n)_{m, n \in \mathbb{N}}$ mit

$$\frac{\max_{1 \leq m \leq a_n} c_m^n}{\sum_{m=1}^{a_n} c_m^n} \rightarrow 0 \text{ für } n \rightarrow \infty$$

gilt

$$\mathbf{Var} \left[\sum_{m=1}^{a_n} c_m^n M_m \right] = o \left(\left(\mathbf{E} \left[\sum_{m=1}^{a_n} c_m^n M_m \right] \right)^2 \right).$$

Beweis. Setzen wir

$$\pi_i^n := \frac{c_i^n}{\sum_{i=1}^{a_n} c_i^n},$$

dann sind die Voraussetzungen von Satz 2.17 erfüllt. Damit folgt

$$\begin{aligned} \frac{\mathbf{Var} \left[\sum_{m=1}^{a_n} c_m^n M_m \right]}{\left(\mathbf{E} \left[\sum_{m=1}^{a_n} c_m^n M_m \right] \right)^2} &= \frac{\mathbf{Var} \left[\sum_{m=1}^{a_n} c_m^n M_m \right]}{\mathbf{E}[M_1]^2 \left(\sum_{m=1}^{a_n} c_m^n \right)^2} \\ &= \frac{1}{\mathbf{E}[M_1]^2} \mathbf{Var} \left[\sum_{m=1}^{a_n} \pi_m^n M_m \right] \\ &\rightarrow 0 \quad \text{für } n \rightarrow \infty. \end{aligned}$$

Damit folgt die Behauptung. □

Kapitel 4

Ausblick

Wir wollen an dieser Stelle noch einen kurzen Ausblick auf weitere sich in Arbeit befindende Ideen werfen.

4.1 Das Querlegen einer Generation mit M Individuen

Wir betrachten erneut das [BGKS]-Modell. Für $i \in \mathbb{Z}$ und $l \in \{1, \dots, N\}$ setzen wir

$$T_{(i,l)} := \{(j, m) \in \mathbb{Z} \times \{1, \dots, N\} : (i, l) \sim (j, m)\}.$$

Damit ist

$$\mathcal{P} := \{T_{(i,l)} : (i, l) \in \mathbb{Z} \times \{1, \dots, N\}\}$$

die durch \sim (siehe Definition 2.5) induzierte zufällige Partition von $\mathbb{Z} \times \{1, \dots, N\}$. Das [HS]-Modell entspricht $N = 1$ und kann damit als zufällige Partitionierung von \mathbb{Z} gesehen werden. Wir haben für die aus zufälligen Färbungen dieser Partitionen entstehenden Irrfahrten S . respektive $Z_{\cdot, N}$ für $N \in \mathbb{N}$, die wir in (2.2) beziehungsweise (2.7) definiert haben, beobachtet, dass deren passende Reskalierungen gegen eine fraktionale Brownsche Bewegung mit Hurst-Parameter $H := \alpha + \frac{1}{2}$ konvergieren. Wir werden in Satz 4.1 und Satz 4.2 zeigen, dass die durch

$$Z_0^M := 0 \text{ sowie } Z_n^M - Z_{n-1}^M := \frac{1}{M} \sum_{i=(n-1)M+1}^{nM} X_i, \quad n \geq 1$$

definierte Irrfahrt das gleiche Grenzwertverhalten aufweist, wobei wir zwischen $\lfloor t \rfloor$ und $\lceil t \rceil$ wieder linear interpolieren.

Die Idee hinter diesem Konstrukt ist die Folgende: M Individuen, die zuvor M Generationen im [HS]-Modell darstellten, werden nun als eine Generation aufgefasst. Wir berufen uns also vollständig auf das [HS]-Modell, aber fassen Individuum 1 als das Individuum mit Nummer 1 in Generation 0 des [BGKS]-Modells auf, Individuum M als das Individuum mit Nummer M in Generation 1, und allgemeiner gesprochen Individuum $(n - 1)M + i$ als das Individuum Nummer i in Generation n des [BGKS]-Modells. Gegeben das Ereignis, dass ein Individuum k seinen Vater in einer solchen *Pseudogeneration* l wählt, ist die Wahl des Vaters innerhalb dieser Generation offensichtlich nicht mehr uniform!

An dieser Stelle liegt es allerdings trotzdem nahe, dass sich Partitionen der in Blöcke der Länge M unterteilten Menge \mathbb{Z} auf einem *grobkörnigen Level* nicht zu sehr von Partitionen der Menge $\mathbb{Z} \times \{1, \dots, M\}$ unterscheiden.

Dementsprechend ist es naheliegend, dass man für ein sehr langsam variierendes Wahrscheinlichkeitsmaß μ gegeben das Ereignis, dass ein Individuum k seinen Vater in einer *Pseudogeneration* l wählt, beinahe uniform aus dieser wählt. Diese Vermutung werden wir im Folgenden nicht beweisen, sondern uns lediglich des Grenzwertverhaltens von \mathcal{Z}_n^M versichern.

Zunächst gilt

$$\mathcal{Z}_n^M = \frac{1}{M} S_{nM}, \quad (4.1)$$

also

$$\mathbf{Var} \left[\mathcal{Z}_n^M \right] = \frac{1}{M^2} \mathbf{Var} [S_{nM}]. \quad (4.2)$$

(4.2) liefert damit zusammen mit Lemma 2.7 bereits die Asymptotik

$$\mathbf{Var} \left[\mathcal{Z}_n^M \right] \sim \frac{1}{M^2} \frac{4p(1-p)}{2\alpha(2\alpha+1) \sum_{l \geq 0} q_l^2} \frac{n^{2\alpha+1}}{L(n)^2} \frac{1}{\Gamma(1-\alpha)^2 \Gamma(2\alpha) \cos(\pi\alpha)}.$$

Damit erhalten wir den folgenden Satz:

Satz 4.1. Für $\alpha \in (0, 1/2)$, $\mu \in \Gamma_\alpha$ sowie

$$\tilde{c}^{-2} := \frac{4p(1-p)K_\alpha}{\sum_{j \geq 0} q_j^2}$$

konvergieren die endlichdimensionalen Marginalverteilungen der Prozesse

$$\tilde{\mathcal{Z}}_t^{n,M} := \tilde{c}(nM)^{-\frac{1}{2}-\alpha} L(nM) \left(\mathcal{Z}_{nt} - nt(2p-1) \right) M, \quad t \geq 0$$

für $n \rightarrow \infty$ gegen eine fraktionale Brownsche Bewegung mit Hurst Parameter $\alpha + \frac{1}{2}$.

Beweis. Nach (4.1) gilt

$$\tilde{c}(nM)^{-\frac{1}{2}-\alpha} L(nM) \left[\mathcal{Z}_{nt} - nt(2p-1) \right] \cdot M$$

$$= \tilde{c}(nM)^{-\frac{1}{2}-\alpha} L(nM) [S_{nMt} - nMt(2p-1)].$$

[HS, Proposition 3] besagt nun, dass für $\tilde{n} := nM \rightarrow \infty$ die durch

$$[0, \infty) \rightarrow \mathbb{R}, \quad t \mapsto \tilde{c}(\tilde{n})^{-\frac{1}{2}-\alpha} L(\tilde{n}) [S_{\tilde{n}t} - \tilde{n}t(2p-1)]$$

definierten Prozesse im Sinne der endlichdimensionalen Randverteilungen gegen eine fraktionale Brownsche Bewegung mit Hurst-Parameter $\alpha + \frac{1}{2}$ konvergieren. \square

Ganz im Sinne von [HS, Theorem 1.1] beziehungsweise Satz 2.10 verbessern wir das Resultat noch:

Satz 4.2. *Unter den bisher gemachten Annahmen existiert eine Folge von Couplings $(C_n)_{n \in \mathbb{N}}$ und eine fraktionale Brownsche Bewegung B^H mit Hurst-Parameter $H := \alpha + \frac{1}{2}$, sodass für alle $\varepsilon > 0, M \in \mathbb{N}, T \geq 0$*

$$\lim_{n \rightarrow \infty} C_n \left(\|\tilde{\mathcal{Z}}^{n,M} - B^H\|_{L^\infty([0,T])} > \varepsilon \right) = 0$$

gilt.

Beweis. Wegen

$$\tilde{\mathcal{Z}}_t^{n,M} = \tilde{c}(nM)^{-\frac{1}{2}-\alpha} L(nM) [\mathcal{Z}_{nt} - nt(2p-1)] M \tag{4.3}$$

$$= \tilde{c}(nM)^{-\frac{1}{2}-\alpha} L(nM) \left[\frac{1}{M} S_{nMt} - nt(2p-1) \right] M \tag{4.4}$$

$$= \tilde{c}(nM)^{-\frac{1}{2}-\alpha} L(nM) [S_{nMt} - nMt(2p-1)] \tag{4.5}$$

$$= \tilde{\mathcal{S}}_t^{nM} \tag{4.6}$$

folgt die Behauptung sofort mit Satz 2.10. \square

4.2 Zufällige Partition von \mathbb{Z} im [HS]-Modell

Der Beweis von Satz 2.10 fußt in seiner aktuellen Form und auch in der Erweiterung auf das [BGKS]-Modell auf einem Zentralen Grenzwertsatz für Martingale, angewandt auf (sukzessive Prognosedifferenzen) von $(S_n)_n$ beziehungsweise $(Z_n)_n$. Tatsächlich ist allerdings anzunehmen, dass die asymptotische Normalität der Irrfahrt sich in der Baumstruktur selbst begründet, da das einzige zusätzliche Element, das die Irrfahrt dem zufälligen Graphen G_μ voraus hat, die unabhängig identisch verteilten $\{\pm 1\}$ -Färbungen der einzelnen Komponenten sind. Es liegt also nahe an dieser Stelle die Zufallsvariablen $(H^n)_{n \in \mathbb{N}}$, welche die Zahl der disjunkten Komponenten des Graphen G_μ auf $\{0, \dots, n\}$ angibt, und $(K_i^n)_{0 \leq i \leq n, n \in \mathbb{N}}$,

welche die Zahl der mit i verwandten Individuen in $\{0, \dots, n\}$ angibt, zu betrachten. Formal lassen sich die beiden Zufallsvariablen beispielsweise durch

$$H^n := \sum_{k=0}^n \prod_{j=0}^{k-1} (1 - \mathbf{1}_{k \sim j}), \quad K_i^n := \sum_{j=0}^n \mathbf{1}_{j \sim i}$$

ausdrücken. Dabei sieht man die Form von H^n leicht ein, indem man sich vor Augen führt, dass ein Individuum k eine neue Komponente eröffnet, wenn es mit keinem Individuum $0, \dots, k-1$ verwandt ist.

Eine Analyse der Varianz von H^n scheint auf den ersten Blick aufgrund der Korrelationen zwischen den $\mathbf{1}_{k \sim j}$ und den Produkten von bis zu einer Anzahl n dieser Indikatoren unangenehm. Um K_i^n zu analysieren ist es naheliegend die Komponente der 0 zuerst zu betrachten. Naheliegend wäre es an dieser Stelle, Konzentration von K_0^n zu zeigen. Geschehen könnte dies mittels einer Analyse der Varianz, welche allerdings Rechnungen der Form

$$\mathbf{Var}[K_0^n] = \sum_{j=0}^n \mathbf{Var}[\mathbf{1}_{j \sim 0}] + 2 \sum_{0 \leq l < j \leq n} \mathbf{Cov}[\mathbf{1}_{l \sim 0}, \mathbf{1}_{j \sim 0}]$$

erfordert. Die Asymptotik der Varianzterme entspricht der von Verschmelzungswahrscheinlichkeiten und ist uns bereits aus Satz 2.3 bekannt. Um die Asymptotik von $\mathbf{Cov}[\mathbf{1}_{j \sim 0}, \mathbf{1}_{l \sim 0}]$ zu bestimmen, bedarf es jedoch der Überlegungen zur derer von

$$\mathbf{E}[\mathbf{1}_{j \sim 0} \mathbf{1}_{l \sim 0}] = \mathbf{P}(0, j, l \text{ liegen in der gleichen Komponente von } G_\mu).$$

Die Berechnung oder wenigstens die Beschreibung des asymptotischen Verhaltens dieser Wahrscheinlichkeit ist kombinatorisch herausfordernd, da beispielsweise im Falle $0 < j < l$ unterschieden werden muss, ob zuerst die Ahnenlinien j und l , 0 und j oder zuerst die Ahnenlinien von 0 und l verschmelzen, was wiederum zu der Problematik führt, dass bedingte Wahrscheinlichkeiten der Form

$$\begin{aligned} & \mathbf{P}(\text{Ahnenlinien von } j, l \text{ treffen sich} \\ & \quad | \text{Ahnenlinie von } 0, l \text{ treffen sich vor jenen von } j \text{ und } l) \end{aligned}$$

kontrolliert werden müssten. Diese lassen sich nicht auf einfache Art über die Erneuerungsfunktion ausdrücken.

4.3 Skalierungslimes von Anteilsprozessen

4.3.1 Der Fall $0 < \alpha < \frac{1}{2}$

Im [BGKS]-Modell wird ein der Standard-Wright-Fisher-Diffusion analoger Prozess als Grenzwert des Anteilsprozesses, der zum Zeitpunkt n den Anteil der Typ-1-Individuen in der zum Zeitpunkt n lebende Population angibt, nicht zu finden sein, denn: Generation n , die zum Zeitpunkt n lebende Population, stammt nicht nur von Generation $n-1$ sondern potentiell von beliebig lange vergangenen Generationen ab. Uns ist allerdings bekannt, dass passende Reskalierungen des Prozesses $Z_{n,N}$ (siehe (2.7)) mit Inkrementen

$$\mathcal{W}_{n,N} := \frac{1}{N} \sum_{i=1}^N \mathbf{1}_{X_{n,i}=1}$$

gegen eine fraktionale Brownsche Bewegung mit Hurst-Parameter $H := \alpha + \frac{1}{2}$ konvergieren. Es liegt also nahe, dass passende Reskalierungen von W_n gegen eine *passend definierte Form von Ableitung* der fraktionalen Brownschen Bewegung konvergieren.

4.3.2 Der Fall $\frac{1}{2} < \alpha < 1$

Für $\alpha \in (\frac{1}{2}, 1)$ hat man in der zufälligen Partitionierung von \mathbb{Z} fast sicher nur eine Komponente des Graphen G_μ . Betrachtet man allerdings nur solche Individuen als verwandt, deren MRCA zur Zeit $-f(N, \alpha)$ oder später gelebt hat (mit passend gewähltem $f(N, \alpha) \in \mathbb{N}$), dann hat man N verschiedene Komponenten, die man per Münzwurf einfärben kann. Eine Möglichkeit, die Vererbungsdynamik konsistent zu machen, besteht dann darin, die Verteilung μ bei $f(N, \alpha)$ abzuschneiden. So bekommt man eine Folge von Maßen $\mu^{(N)}$. Die Typenfrequenzen

$$\mathcal{W}_{n,N} := \frac{1}{N} \sum_{i=1}^N \mathbf{1}_{X_{n,i}=1}$$

haben dann für wachsendes N immer langreichweitigere Korrelationen. Es wäre spannend, das Skalierungsverhalten der Dynamik von $\mathcal{W}_{n,N}$ mit wachsendem N zu untersuchen und einen Skalierungslimes zu finden. Es wäre zum Beispiel denkbar, dass für eine Abbildung g und einen Startwert \tilde{w} zum Zeitpunkt null ein Skalierungslimes von $\mathcal{W}_{n,N}$ eine stochastische Differentialgleichung der Form

$$\tilde{W}_t = \tilde{w} + \int_0^t \left[\int_{-\infty}^s g(\tilde{W}_\xi) d\xi \right] dB_s$$

oder

$$\tilde{\mathcal{W}}_t = \tilde{w} + \int_0^t \left[\int_{-\infty}^s g(\tilde{\mathcal{W}}_\xi) d\xi \right] dB_s^H$$

für eine Brownsche Bewegung B respektive eine fraktionale Brownsche Bewegung B^H erfüllt. In gewisser Hinsicht könnte dieser Prozess als *fraktionale Wright-Fisher-Diffusion betrachtet werden* (siehe (2.6) zum Vergleich).

Literaturverzeichnis

- [AT] R. J. Adler and J. E. Taylor. *Random fields and geometry*. Springer Monographs in Mathematics. Springer, New York, 2007.
- [B1] P. Billingsley. *Convergence of probability measures*. John Wiley & Sons, Inc., New York-London-Sydney, 1968.
- [B2] P. Billingsley. *Convergence of probability measures 2nd edition*. John Wiley & Sons, Inc., New York-London-Sydney, 1968.
- [BGKS] J. Blath, A. González Casanova, N. Kurt, and D. Spanò. The ancestral process of long-range seed bank models. *Journal of Applied Probability*, 50(3):741–759, 2013. URL <http://www.jstor.org/stable/43283498>.
- [BGT] N. H. Bingham, C. M. Goldie, and J. L. Teugels. *Regular variation*. Cambridge University Press Cambridge [Cambridgeshire] ; New York, 1987. URL <http://www.loc.gov/catdir/toc/cam027/86028422.html>.
- [D1] R. Doney. One-sided local large deviation and renewal theorems in the case of infinite mean. *Probability Theory and Related Fields*, 107:451–465, 04 1997. <http://dx.doi.org/10.1007/s004400050093>.
- [D2] R. M. Dudley. *Real analysis and probability*. The Wadsworth & Brooks/Cole Mathematics Series. Wadsworth & Brooks/Cole Advanced Books & Software, Pacific Grove, CA, 1989.
- [F] W. Feller. *An introduction to probability theory and its applications. Vol. II*. Second edition. John Wiley & Sons Inc., New York, 1971.
- [GJ] A. Garsia and J. Lamperti. A discrete renewal theorem with infinite mean. *Commentarii Mathematici Helvetici*, 37(1):221–234, 1962. <http://dx.doi.org/10.1007/BF02566974>.
- [GL] J. Geluk and L. Haan. Stable probability distributions and their domains of attraction. *Tinbergen Institute, Tinbergen Institute Discussion Papers*, 20, 01 1997.

- [HS] A. Hammond and S. Sheffield. Power law Pólya's urn and fractional brownian motion. *Probability Theory and Related Fields*, 157(3):691–719, 2013. <http://dx.doi.org/10.1007/s00440-012-0468-6>.
- [KKL] I. Kaj, S. Krone, and M. Lascoux. Coalescent theory for seed bank models. *Journal of Applied Probability*, 38:285–300, 06 2001. <http://dx.doi.org/10.1017/S0021900200019860>.
- [LS] W. Li and Q.-M. Shao. Gaussian processes: Inequalities, small ball probabilities and applications. In *Stochastic Processes: Theory and Methods*, volume 19 of *Handbook of Statistics*, pages 533 – 597. Elsevier, 2001. [http://dx.doi.org/https://doi.org/10.1016/S0169-7161\(01\)19019-X](http://dx.doi.org/https://doi.org/10.1016/S0169-7161(01)19019-X).
- [P] D. Panchenko. Lecture notes on 18.175 theory of probability, Fall 2008. URL https://dspace.mit.edu/bitstream/handle/1721.1/96865/18-175-fall-2008/contents/lecture-notes/all_notes.pdf.
- [S1] T. Sottinen. Fractional Brownian motion, random walks and binary market models. *Finance Stoch.*, 5(3):343–355, 2001. <http://dx.doi.org/10.1007/PL00013536>.
- [S2] S. Sallon, E. Solowey, Y. Cohen, R. Korchinsky, M. Egli, I. Woodhatch, O. Simchoni, and M. Kislev. Germination, genetics, and growth of an ancient date seed. *Science*, 320(5882):1464–1464, 2008, <http://arxiv.org/abs/https://science.sciencemag.org/content/320/5882/1464.full.pdf>. <http://dx.doi.org/10.1126/science.1153600>.
- [VB] C. Vardar-Acar and H. Bulut. Bounds on the expected value of maximum loss of fractional brownian motion. *Statistics & Probability Letters*, 104:117 – 122, 2015. <http://dx.doi.org/https://doi.org/10.1016/j.spl.2015.05.001>.

Selbstständigkeitserklärung

Hiermit erkläre ich, dass ich die Masterarbeit selbständig verfasst und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe und die aus fremden Quellen direkt oder indirekt übernommenen Gedanken als solche kenntlich gemacht habe. Die Arbeit habe ich bisher keinem anderen Prüfungsamt in gleicher oder vergleichbarer Form vorgelegt. Sie wurde bisher nicht veröffentlicht.

Frankfurt, den 23. März 2021