Grenzwertsätze für empirische Funktionen von Partialsummen zur Untersuchung von DNA- und Proteinsequenzen

Diplomarbeit

ausgeführt unter der Leitung von
Prof. Dr. Matthias Löwe
am Institut für Mathematische Statistik der Universität Münster

Münster, November 2005

Sarah Behrens

Inhaltsverzeichnis

Symbolverzeichnis

$\mathbb{N}$	Menge der positiven natürlichen Zahlen 1,2,...
$\mathbb{N}_0$	Menge der natürlichen Zahlen 0,1,...
$\mathbb{Z}$	Menge der ganzen Zahlen
$\mathbb{R}$	Menge der reellen Zahlen
$\mathbb{B}$	Borelsche σ-Algebra über $\mathbb{R}$
$\mathbb{I}_A$	Indikatorfunktion der Menge A
$A^{\complement}$	Komplement der Menge A
$\mathcal{N}(a, \sigma^2)$	Normalverteilung mit den Parametern $a \in \mathbb{R}$ und $\sigma^2 \in (0, \infty)$
$\mathcal{N}(a, \Sigma)$	mehrdimensionale Normalverteilung mit Parametern $a \in \mathbb{R}^n$ und Kovarianzmatrix Σ
$[x]$	$\max\{n \in \mathbb{Z} : n \leq x\}$
$\lVert x \rVert$	$\sqrt{x_1^2 + ... + x_n^2}$ für $x = (x_1, ..., x_n) \in \mathbb{R}^n$ (Vektornorm)
$\langle x, y \rangle$	$x_1 y_1 + ... + x_n y_n$ für $x,\ y \in \mathbb{R}^n$ (Skalarprodukt)
$f(x) = O(g(x))$ $(x \to \infty)$	es existieren Konstanten $C > 0$ und x_0, so dass $\lvert f(x)\rvert \leq C \lvert g(x)\rvert$ für alle $x \geq x_0$ (Landau-Symbol)
$f(x) = O(g(x))$ $(x \to a)$	es existieren Konstanten $C > 0$ und $\epsilon > 0$, so dass $\lvert f(x)\rvert \leq C \lvert g(x)\rvert$ für alle x mit $\lvert x - a\rvert < \epsilon$ (Landau-Symbol)
$\xrightarrow[(V)]{}$	verteilungskonvergent
$\xrightarrow[P-f.s.]{}$	P-fast sicher konvergent
$A \geq 0$	A nichtnegative Matrix, d.h. alle Komponenten von A sind ≥ 0
$A \gg 0$	A positive Matrix, d.h. alle Komponenten von A sind > 0
$x \geq 0$	x nichtnegativer Vektor, d.h. alle Komponenten von x sind ≥ 0
$x \gg 0$	x positiver Vektor, d.h. alle Komponenten von x sind > 0

Kapitel 1

Einführung

Im Rahmen dieser Diplomarbeit wird es darum gehen, Buchstabensequenzen im Hinblick auf Länge und Zusammensetzung zu charakterisieren. Diese Charakterisierung wird mittels zweier starker als auch mittels eines zentralen Grenzwertsatzes erfolgen. Aussagen hinsichtlich der Länge und Zusammensetzung von Sequenzstücken mit hohem Score (zur Erläuterung des Begriffs „Score“ siehe Abschnitt 1.2) werden sowohl unter der Hypotheses eines iid-Modells (d.h. Buchstaben in der Sequenz iid; siehe Kapitel 2) als auch unter der Hypothese eines Markov-Modells (Buchstaben in der Sequenz hängen vom Vorgänger oder allgemeiner von m vorhergehenden Buchstaben ab; siehe Kapiel 3 bzw. Ausblick, Seite 101) getroffen werden.
Motiviert werden diese mathematischen Überlegungen durch Anwendungen in der Molekularbiologie bzw. Genetik - und zwar möchte man Ähnlichkeiten zweier oder mehrerer DNA- bzw. Proteinsequenzen bewerten, um so z.B. Aufschluss über den Verwandtschaftsgrad dieser Sequenzen zu gewinnen.
Die Aussagen dieser Diplomarbeit gehen auf drei Artikel von Amir Dembo und Samuel Karlin (siehe [4], [5], [6]) zurück, alle Theoreme, Propositionen und Lemmata, die nicht anderweitig gekennzeichnet sind, stammen also aus diesen drei Artikeln.
Zunächst werden nun im Rahmen der Einführung die biologischen Grundlagen erläutert, sowie der Begriff einer Scorefunktion motiviert und anschließend die aus den Anwendungen resultierenden Fragen formuliert, auf welche diese Diplomarbeit eingeht.

1.1 Biologische Grundlagen

1.1.1 DNA- und Proteinsequenzen

Die Desoxyribonucleinsäure oder kurz DNA (desoxyribonucleic acid) ist der Träger der Erbinfomation eines Organismus. Sie besteht aus einer Verkettung von Nucleotiden, welche aus Nucleosiden (Zucker gebunden an Phosphat) assoziiert mit Basen bestehen. In der DNA befinden sich vier verschiedene Nucleotide, verschieden im Hinblick auf die gebundenen Basen Adenin, Guanin, Cytosin und Thymin (im Folgenden abgekürzt als A, G, C bzw. T). DNA liegt zwar stets als Doppelstrang vor, da jedoch die komplementäre Basenpaarung über Wasserstoffbrücken eindeutig (A-T und C-G) ist, lässt sich ein DNA-Strang eindeutig durch eine Buchstabensequenz (Buchstabe=Base) charakterisieren, z.B. durch AGAAACTTT.

Die Information einer DNA wird in Form von Triplets gespeichert, d.h. drei Nucleotide codieren für eine Aminosäure. Bei der Proteinbiosynthese wird die DNA abgelesen und über mehrere komplexe Prozesse eine Kette von Aminosäuren gebildet, welche modifiziert und gefaltet schlussendlich ein Protein ergibt. Eine Abfolge von Aminosäuren charakterisiert also ein Protein und wird dementsprechend Proteinsequenz genannt. Insgesamt sind 20 Aminosäuren bekannt (z.B. Histidin, Arginin, Lysin, Aspartat, Glutamat usw.).

In der Molekularbiologie ist es möglich, DNA- oder Proteinsequenzen zu entschlüsseln. Diese sogenannten Primärstrukturen geben allerdings nur bedingt Informationen über die räumliche Struktur, welche über Faltung entsteht, oder gar über die Funktionen des Proteins wieder.

Dennoch liegt eine Reihe von Proteinen vor, die bezüglich ihrer Struktur und Funktion charakterisiert wurden (über andere experimentelle Methoden). So existieren beispielsweise Datenbanken von Proteinen, in welchen neben den entschlüsselten Sequenzen auch Informationen über Funktionen und Struktur der Proteine gespeichert sind.
Im Bereich der Erforschung des HI-Virus (human immunodeficiency virus) wird zur Zeit z.B. an der Erstellung von Datenbanken, welche das sequenzierte Genom von HI-Viren und weitere Informationen über die Patienten (z.B. Art der bisherigen Therapie usw.) enthalten, gearbeitet.

Ein naheliegender Ansatz bei der Untersuchung von Funktionen von Proteinen bzw. Teilen des Genoms ist der Vergleich von den zu untersuchenden Sequenzen mit solchen bekannter Struktur und Funktion. Dieser Vergleich kann dann z.B. mittels einer DNA-Datenbank geschehen. Man versucht also, Aussagen über den Verwandtschaftsgrad zweier oder mehrerer Sequenzen zu gewinnen.

1.1.2 Alignments von Sequenzen

Ein erster Schritt in die Richtung Sequenzen miteinander zu vergleichen besteht darin, ein sogenanntes Alignment, d.h. eine Aneinanderlegung dieser Sequenzen, durchzuführen - in der Abbildung (1.1) sehen wir beispielsweise ein Alignment zweier Sequenzabschnitte aus dem Genom eines Wales und eines Menschen. Es ist möglich, zwei oder

Abbildung 1.1: Alignment eines Abschnitts aus den geschlechtsbestimmenden Genen des Wals und des Menschen; siehe [1]

auch mehrere Sequenzen (multiples Alignment) zu alignieren. Bei einem Alignment, welches die komplette Länge von Sequenzen umfasst, spricht man von einem globalen Alignment, bei Alignments, die nur einen Abschnitt beinhalten, von einem lokalen Alignment.

Um ein Alignment vorzunehmen, ist es zunächst wichtig, sich die Ursachen genetischer Variation vor Auge zu führen. Im Verlauf der Evolution entwickelte sich die genetische Vielfalt über verschiedene Veränderungs-Mechanismen: über Insertion (Einfügung einzelner Nucleotide), über Deletion (Entfernung einzelner Nucleotide), über Mutation (punktueller Nucleotid-Austausch) oder über größere Umstrukturierungen. Ein Alignment sollte also eigentlich keine reine Aneinanderlegung von zwei Sequenzen sein, sondern sollte berücksichtigen, dass zwei verwandte Sequenzen aufgrund von Deletionen und Insertionen identische Sequenzsegmente (sogenannte Motive) an verschiedenen Positionen enthalten. Ein gutes Alignment ist also ein solches mit Lücken, sogenannten 'gaps', beispielsweise eines der Form:

```
A G - C C C A G
A C A C C - - G
```

Aus dieser Tatsache wird die Komplexität ein vernünftiges Alignment durchzuführen deutlich. Dies sei hier aber nur am Rande erwähnt, da es kein Thema der Diplomarbeit ist.

Im nächsten Schritt wird es nun darum gehen, ein Maß für die Ähnlichkeit bzw. Verwandtschaft von Sequenzen zu erhalten, d.h. Ähnlichkeit zu beurteilen. Ferner möchte man überprüfen, ob zwischen Sequenzen eine evolutionäre Beziehung vorliegt oder ob es sich um eine rein zufällige Ähnlichkeit handelt.

1.2 Scorefunktionen

1.2.1 Beispiele für Scorefunktionen

Als Maß für den Ähnlichkeitsgrad zweier (oder auch mehrerer) Sequenzen führt man nun den sogenannten Score-Wert ein.
Dem Wort nach bedeutet 'Score' in etwa 'Punktestand'/'Spielstand', das 'Oxford Dictionary of Statistical Terms' (siehe [9]) erläutert den Begriff 'Score' folgendermaßen:

> **Score**: In der Regel ein numerischer Wert, der einer Beobachtung zugeordnet wird (üblicherweise wenn keine messbaren Daten existieren) als Ersatz für eine Variable, die über einer Skala variiert. Der Begriff wird auch benutzt, um die erste Ableitung des natürlichen Logarithmus der Likelihood-Funktion zu beschreiben.

In dieser konkreten Situation ordnet eine Scorefunktion also einem Paar bzw. Tupel von Sequenzen mit Einträgen aus einem Buchstabenalphabet $\mathcal{A}$ (also z.B. bei DNA $\mathcal{A} = \{\text{A, G, C, T}\}$) in möglichst sinnvoller Weise einen numerischen Wert zu, mathematisch formuliert:

Sei $\mathcal{A} = \{a_1, ..., a_r\}$ ein Alphabet von Buchstaben. Beim Vergleich von k Sequenzen ist eine Score-Funktion (oder kurz Score) eine Abbildung $s : \mathcal{A}^k \to \mathbb{R}$ mit

$$s(a_{i_1}, ..., a_{i_k}) = s_{i_1,...,i_k} \text{ für alle } i_1, ..., i_k \in \{1, ..., r\}$$

(Anm.: später bezeichne $\mathcal{A}$ bereits ein Buchstabenalphabet aus k-Tupeln, so dass eine Score-Funktion eine Abbildung $s : \mathcal{A} \to \mathbb{R},\ s(a_i) = s_i$ ist)

Im Folgenden sei $k = 2$ (für zwei Sequenzen).

Den Gesamt-Score erhält man, indem man beim Vergleich zweier oder mehrerer Sequenzen der Länge n jeder Buchstabenposition $i,\ 1 \leq i \leq n$ einen Score zuordnet und diese n Score-Werte aufsummiert. Ein Score-Wert entspricht also wirklich einer Art 'Punktestand'.
Wie lässt sich nun eine sinnvolle Score-Zuordung vornehmen? Dazu seien an dieser Stelle zwei Beispiele angeführt:

Beispiel 1.2.1. (triviale Zuordnung)
a) Gegeben sei ein Buchstabenalphabet $\mathcal{A} = \{a_1, ..., a_r\}$, sowie ein Paar von zwei alignierten Sequenzen mit Einträgen aus diesem Alphabet:

$$((a_i, a_j))_{i,j \in \{1,...,r\}}$$

Eine naheliegende Zuteilung von Scores ist die triviale Zuordnung

$$s(a_i, a_j) := \begin{cases} 1 & \text{falls } i = j \\ 0 & \text{sonst} \end{cases}.$$

Der Scorewert macht hier also nur eine Aussage darüber, ob die Sequenzen an einer bestimmten Position übereinstimmen oder nicht.

b) Betrachtet man Alignments mit 'gaps', ist es üblich für Lücken im Alignment sogenannte 'gap penalties' aufzuschlagen, d.h. Lücken wirken sich negativ auf den Gesamtscore aus. Dies ist sinnvoll, da ja der Grad an Ähnlichkeit geringer ist, sobald identische Sequenzsegmente in den beiden Sequenzen anders positioniert sind.
In diesem Fall könnte eine Zuordnung von Scores folgendermaßen aussehen:

$$s(a_i, a_j) := \begin{cases} 1 & \text{falls } i = j \\ 0 & \text{falls } i \neq j \\ -1 & \text{falls } a_i = \text{gap oder } a_j = \text{gap} \end{cases}.$$

□

So anschaulich diese Score-Zuteilungen sind, so wenig nützlich sind sie bei der Beurteilung von Verwandtschaft zweier Sequenzen. Beispielsweise gibt es Aminosäuren mit sehr großer funktioneller Ähnlichkeit (z.B. die beiden basischen Aminosäuren Arginin und Lysin), wohingegen andere sich sehr stark unterscheiden. Entsprechend sollte ein Paar funktionell ähnlicher Aminosäuren einen höheren Score als ein solches geringer funktioneller Ähnlichkeit erhalten.
Ein Austausch funktionell ähnlicher Aminosäuren wird häufiger beobachtet, wird insgesamt also als wahrscheinlicher erachtet. Eine Beurteilung von Ähnlichkeit, d.h. eine Zuteilung von Scores, sollte also auch unter dem Gesichtspunkt der Wahrscheinlichkeit erfolgen:
Wie wahrscheinlich ist das Auftreten einer Aminosäure unter der Hypothese, dass die Sequenzen nicht miteinander verwandt sind?
Wie wahrscheinlich ist das Auftreten eines Aminosäurenpaares unter der Hypothese, dass die Sequenzen miteinander verwandt sind?
Diese Überlegungen führen zum sogenannten log-Likelihood-Ratio-Score:

Beispiel 1.2.2. (log-Likelihood-Ratio-Score) (siehe auch [18])
Gegeben sei ein Buchstabenalphabet $\mathcal{A} = \{a_1, ..., a_r\}$, sowie ein Paar von zwei alignierten Sequenzen mit Einträgen aus diesem Buchstabenalphabet

$$\begin{gathered} A_1, ..., A_n \\ A'_1, ..., A'_n. \end{gathered}$$

Der Buchstabe a_α trete in der 1. Sequenz an der Position i für alle $1 \leq i \leq n$ mit Wahrscheinlichkeit

$$P(A_i = a_\alpha) = p_{a_\alpha} =: p_\alpha,$$

der Buchstabe a_β in der 2. Sequenz mit der Wahrscheinlichkeit

$$P(A'_i = a_\beta) = p'_{a_\beta} =: p'_\beta$$

auf. Ferner trete das Buchstabenpaar (a_α, a_β) unter der Hypothese, dass beide Sequenzen miteinander verwandt sind, an der Position i für alle $1 \leq i \leq n$ mit der Wahrscheinlichkeit $q_{\alpha,\beta}$ (bei Aminosäuren: Aminosäure-Austauschwahrscheinlichkeit) auf:

$$P(A_i = a_\alpha, A'_i = a_\beta) = q_{a_\alpha, a_\beta} =: q_{\alpha,\beta}.$$

Ein bestimmtes Alignment hat also unter der Hypothese, dass die Sequenzen nicht miteinander verwandt sind (unabhängiges Auftreten), die Wahrscheinlichkeit

$$p := \prod_{i=1}^{n} p_{A_i} p'_{A'_i},$$

sowie unter der Hypothese, dass sie verwandt sind, die Wahrscheinlichkeit:

$$q := \prod_{i=1}^{n} q_{A_i,A'_i}.$$

Das Verhältnis der beiden Wahrscheinlichkeiten $\frac{q}{p}$ wird auch mit *odds ratio* bezeichnet. Logarithmiert man diesen *odds ratio* (um zu einem additiven Score-System zu gelangen), erhält man den log-Likelihood-Ratio-Score

$$\sum_{i=1}^{n} \log\left(\frac{q_{A_i,A'_i}}{p_{A_i}p'_{A'_i}}\right),$$

bzw. den Einzelscore

$$s_{\alpha,\beta} = \log\left(\frac{q_{\alpha,\beta}}{p_\alpha p'_\beta}\right).$$

Der Score ist hier somit ein Maß für die Wahrscheinlichkeit, dass zwei DNA- bzw. Proteinsequenzen ähnlich, also evolutiv auseinander hervorgegangen sind, relativ zur Wahrscheinlichkeit, dass sie lediglich zufällig miteinander in Beziehung gesetzt wurden.

□

1.2.2 Resultierende Fragestellung

Es stellt sich die Frage, ob die Benutzung des log-Likelihood-Ratio-Scores mathematisch gerechtfertigt ist, d.h. ob es vielleicht möglich ist, alle Scores eines dem Modell nach verwandten (also hochscorigen) Sequenzenpaares als log-Likelihood-Ratio-Scores mit geeigneter Basis des Logarithmus darzustellen.
Eine interessante Aufgabenstellung ist es also, die Art der Buchstabenverteilung in einem hochscorigen Sequenzenpaar zu untersuchen.
Am Rande interessiert ferner, wie denn nun die Wahrscheinlichkeiten $q_{\alpha,\beta}$ (auch *target frequencies* genannt) sowie die Wahrscheinlichkeiten p_α bzw. p'_β (auch *background frequencies* genannt) aussehen. Diese Charakterisierung kann allerdings nur mittels empirischer Daten geschehen und wird im späteren Abschnitt Anwendungen (siehe Abschnitt 2.6) kurz angerissen.
Von besonderem Interesse ist außerdem das Sequenzenpaarsegment mit maximalem Verwandtschaftsgrad, d.h. mit maximalem Scorewert. Dieses Segment (bzw. allgemeiner ein hochscoriges Sequenzenpaarsegment) möchte man z.B. hinsichtlich seiner Länge charakterisieren.

Im Weiteren wir es also darum gehen, folgende Punkte zu bearbeiten:

(i) **Welche Länge haben hochscorige Sequenzsegmente?**
(bzw. solche mit maximalem Score)

(ii) **Welche Zusammensetzung haben hochscorige Sequenzsegmente? (bzw. solche mit maximalem Score)**

Diese Fragen werden zunächst unter der Annahme eines iid-Modells (siehe Kapitel 2), also unter der Hypothese, dass die Aminosäuren (bzw. Nucleotide) einer Sequenz unabhängig voneinander auftreten, und danach unter der Annahme eines Markov-Modells (siehe Kapitel 3), also unter der Hypothese, dass das Auftreten einer Aminosäure (bzw. eines Nucleotids) von der vorhergehenden Aminosäure abhängt, angegangen.

Anmerkung 1.2.3. Der Einfachheit bzw. der Verallgemeinerung halber wird im Folgenden nicht mehr von einem 'Sequenzenpaar' (oder 'Sequenzentupel'), sondern nur noch von einer 'Sequenz' gesprochen, da die Elemente einer Sequenz auch als Tupel aufgefasst werden können. Dementsprechend erfolgt die Score-Zuteilung (vergleiche Seite 3) nicht bezogen auf Paare von Buchstaben, sondern gemäß:

$$s : \mathcal{A} \to \mathbb{R},\ s(a_i) = s_i$$

Formal wird also $\mathcal{A}^2$ durch $\mathcal{A}$ ersetzt, $q_{\alpha,\beta}$ durch q_i, sowie $p_\alpha p'_\beta$ durch p_i.

Kapitel 2

Grenzwertsätze im iid-Fall

Um Aussagen über die Länge und Zusammensetzung von Sequenzen mit hohem Score zu treffen (siehe Kapitel 1), sei zunächst davon ausgegangen, dass die Buchstaben (also Aminosäuren bzw. Nucleotide) einer Sequenz unabhängig voneinander auftreten und identisch verteilt sind.
Man geht also erst einmal von einem iid-Modell aus.
Es interessiert also hinsichtlich des assoziierten Testprobelms unter der Nullhypothese 'die X_i sind iid', wie wahrscheinlich ein hoher Score, als auch wie lang ein Segment mit hohem Score ist.
Dazu werden im Folgenden zwei starke Grenzwertsätze (siehe Abschnitt 2.2) als auch ein zentraler Grenzwertsatz (siehe Abschnitt 2.4) bewiesen.

2.1 Notationen und Voraussetzungen

Es sei $(X_n)_{n\in\mathbb{N}}$ eine Folge von iid Zufallsgrößen, welche jedem Buchstaben aus einem Buchstabenalphabet $\mathcal{A} = \{a_1, ..., a_r\}$ einen Score zuordnen:

$$X_n := \mathcal{A} \to \{s_1, ..., s_r\} \text{ mit}$$

$$P(X_n = s_i) = p_i, \quad 1 \leq i \leq r, \quad \sum_{i=1}^{r} p_i = 1, \quad p_i > 0.$$

Diese Scores summiert man zu einem Gesamt-Score $S_n = \sum_{i=1}^{n} X_i$ auf.
Da man ferner die Zusammensetzung von Sequenzstücken mit hohem Score untersuchen möchte, interessiert man sich für die Zufallsgröße W_n, welche die Häufigkeit des Auftretens eines Buchstabens a_α in einem Sequenzensegment angibt, also für

$$W_n = \sum_{i=1}^{n} U_i \quad \text{mit} \quad U_i = \begin{cases} 1 & \text{falls } X_i = s_\alpha \\ 0 & \text{sonst} \end{cases}.$$

Für die Beweise liegt allgemeiner die folgende Situation vor (es genügt, dass die X_i beschränkt sind):

Es sei $(X_n, U_n)_{n\in\mathbb{N}}$ eine Folge von iid und beschränkten Zufallsgrößen, dabei können die U_i von den X_i abhängig sein, seien aber unabhängig von den X_j, $i \neq j$. Im Folgenden gelte also $|X_i| \leq K$ sowie $|U_i| \leq K'$.

Man betrachtet den Partialsummenprozess $(S_n)_{n\in\mathbb{N}_0}$ mit $S_0 = 0$ und $S_n = \sum_{i=1}^n X_i$, sowie $(W_n)_{n\in\mathbb{N}_0}$ mit $W_0 = 0$ und $W_n = \sum_{i=1}^n U_i$.

Von besonderem Interesse ist das Segment mit maximalem Scorewert $M(n)$, welcher durch

$$M(n) := \sup_{0\leq k<l\leq n} (S_l - S_k)$$

gegeben ist.
Man setzt voraus, dass

(i) $E(X_1) < 0$

(ii) $P(X_1 > 0) > 0$

gilt, so dass der Prozess eine negative Drift erhält, aber eine positive Wahrscheinlichkeit, frühzeitig einen positiven Score anzunehmen. Die Voraussetzung $E(X_1) < 0$ macht Sinn, da im Falle $E(X_1) > 0$ das Segment mit maximalem Score, welches gerade genauer untersucht werden soll, in vielen Fällen schon die ganze bzw. ein Großteil der Sequenz sein würde. Daraus würden sich keine interessanten Aussagen ergeben.

Nun definiert man folgende Stoppzeiten:

$$K_0 = 0, \quad K_\nu = \min\{k \geq K_{\nu-1} + 1 : S_k - S_{K\nu-1} \leq 0\}, \quad \nu \in \mathbb{N} \tag{2.1}$$

Durch die negative Drift von (S_m) sind die K_ν P-f.s. endlich, denn:

$$\begin{aligned}
P(K_1 = \infty) &= P(\inf\{k \geq 1 : S_k \leq 0\} = \infty) \\
&= P(\forall k \geq 1 : S_k > 0) \\
&\leq P(\exists I \subset \mathbb{N}, |I| = \infty\ \forall i \in I : X_i > 0) \\
&= P(\bigcup_{I\subset\mathbb{N},\ |I|=\infty} \bigcap_{i\in I} \{X_i > 0\} \\
&\leq \sum_{I\subset\mathbb{N},\ |I|=\infty} \prod_{i\in I} P(X_i > 0),\ \text{da die } (X_i) \text{ stochastisch unabhängig sind} \\
&= 0, \quad \text{da } E(X_i) < 0 \text{ und somit } P(X_i < 0) < 1 \text{ für alle } i.
\end{aligned}$$

(Die Aussage für $K_\nu, \nu > 1$ folgt induktiv.)

Der Zeitraum $K_{\nu-1} + 1$ bis K_ν gibt dabei gerade die sogenannte νte Epoche wieder. Die 1. Epoche startet also im Punkt 0 und endet, sobald ein nichtpositiver Gesamtscore erreicht wird. Die darauffolgende 2. Epoche beginnt, sobald wieder positive Scores auftreten und endet wiederum, wenn der Gesamtscore vom Startpunkt der 2. Epoche aus gerechnet unter 0 fällt usw.

Jetzt definiert man ferner für $y > 0$ die Stoppzeit

$$T_1(y) = \min\{m > 0 : S_m \geq y \text{ oder } S_m \leq 0\}.$$

$T_1(y)$ gibt also den Zeitpunkt an, zu dem der Prozess (S_k) das erste Mal das Intervall $(0, y)$ verlässt. Sukzessiv definiert man weiterhin:

$$T_\nu(y) = \min\{m > K_{\nu-1} : S_m - S_{K_{\nu-1}} \leq 0 \text{ oder } S_m - S_{K_{\nu-1}} \geq y\}, \quad \nu \in \mathbb{N} \tag{2.2}$$

Man setzt:

$$I_\nu(y) := \begin{cases} 1 & \text{falls } S_{T_\nu(y)} - S_{K_{\nu-1}} \geq y \\ 0 & \text{falls } S_{T_\nu(y)} - S_{K_{\nu-1}} \leq 0 \end{cases}. \tag{2.3}$$

Im Weiteren interessiert nur der Fall $I_\nu(y) = 1$, denn man möchte ja gerade hochscorige Sequenzsegmente betrachten.

Die Länge des νten Segmentes $(K_{\nu-1}, T_\nu(y))$ ist gegeben durch:

$$L_\nu(y) := T_\nu(y) - K_{\nu-1}$$

Ferner sei $W_\nu(y)$ definiert durch

$$W_\nu(y) := W_{T_\nu(y)} - W_{K_{\nu-1}}.$$

Da die (X_i, U_i) iid sind, sind ebenso die Zufallsgrößen $(I_\nu(y))_{\nu\in\mathbb{N}}$ $(L_\nu(y))_{\nu\in\mathbb{N}}$ und $(W_\nu(y))_{\nu\in\mathbb{N}}$ iid mit den Verteilungen von $L_1(y) := L(y)$, $I_1(y) = I(y)$ und $W_1(y) := W(y)$.

Da im Folgenden nur der Fall interessant ist, bei dem der Prozess einen Wert $\geq y$ annimmt, also der Fall $I_\nu(y) = 1$, seien jetzt noch alternative Bezeichnungen eingeführt, die es einem hinterher ersparen werden mit bedingten Verteilungen zu rechnen.

Sei $(\kappa_\nu(y), \tau_\nu(y), \sigma_\nu(y))_{\nu\in\mathbb{N}}$ eine Folge von Epochen mit

$$\begin{aligned}
\kappa_1(y) &= \inf\{k \geq 0 : \exists m \geq k+1 : S_j - S_k > 0 \text{ für alle } k+1 \leq j \leq m-1 \\
&\qquad\qquad \text{und } S_m - S_k \geq y\} \\
\tau_1(y) &= \inf\{k \geq \kappa_1(y) : S_k - S_{\kappa_1(y)} \geq y\} \\
\sigma_1(y) &= \inf\{k \geq \tau_1(y) : S_k - S_{\kappa_1(y)} \leq 0\}
\end{aligned}$$

$\kappa_1(y)$ bezeichnet also den Anfangspunkt einer Exkursion, welche beginnt, sobald positive Werte auftreten, die zusammen addiert mindestens den Wert y erreichen.

$\tau_1(y)$ gibt die Wartezeit auf der positiven Achse an, bis der Prozess zum ersten Mal diesen Wert y erreicht bzw. überschreitet.
$\sigma_1(y)$ bezeichnet wiederum den Zeitpunkt, bei welchem (S_k) das erste Mal zur nichtpositiven Achse zurückkehrt, also den Endpunkt der Exkursion $(\kappa_1(y), \tau_1(y), \sigma_1(y))_\nu$.

Sukzessiv definiert man weiterhin

$$\begin{aligned} \kappa_\nu(y) &= \inf\{k \geq \sigma_{\nu-1}(y) : \exists m \geq k+1 : S_j - S_k > 0 \text{ für alle } k+1 \leq j \leq m-1 \\ &\qquad \text{und } S_m - S_k \geq y\} \\ \tau_\nu(y) &= \inf\{k \geq \kappa_\nu(y) : S_k - S_{\kappa_\nu(y)} \geq y\} \\ \sigma_\nu(y) &= \inf\{k \geq \tau_\nu(y) : S_k - S_{\kappa_\nu(y)} \leq 0\} \end{aligned} \tag{2.4}$$

Man erhält also eine Folge von Epochen, die bei $\kappa_\nu(y)$ starten, einen Wert $\geq y$ zum Zeitpunkt $\tau_\nu(y)$ erreichen und zum Zeitpunkt $\sigma_\nu(y)$ die positive Achse verlassen. Falls also ν der erste Index mit $I_\nu(y) = 1$ ist, d.h. $\nu = \min\{k \geq 1 : I_k(y) = 1\}$, so gilt:

$$\tau_1(y) - \kappa_1(y) = L_\nu(y).$$

Im Gegensatz zur Folge $(K_{\nu-1}, T_\nu(y))_{\nu\in\mathbb{N}}$ spart man sich diejenigen Segmente der Sequenz aus, bei denen der Prozess negativ wird, also $I_\nu(y) = 0$ gilt, was für die folgenden Untersuchungen irrelevant ist, man möchte ja gerade Segmente mit sehr hohen Scorewerten charakterisieren. Damit ergibt sich folgende Proposition:

Proposition 2.1.1. *Für alle $\nu \in \mathbb{N}$ gilt: $P^{\tau_\nu(y)-\kappa_\nu(y)} = P^{L(y)|I(y)=1}$.*

Beweis. Falls $I_\nu(y) = 1$ für ein ν, folgt, da $L_\nu(y) = T_\nu(y) - K_{\nu-1}$:

$$\begin{aligned} S_j - S_{K_{\nu-1}} &> 0 \quad \text{für alle } K_{\nu-1} + 1 \leq j \leq T_\nu(y) - 1, \quad \text{sowie} \\ S_{T_\nu}(y) - S_{K_{\nu-1}} &\geq y \end{aligned} .$$

Somit gilt, dass die Aussagen $L_\nu(y) = k$ unter $I_\nu(y) = 1$ für ein ν und $\tau_{\tilde\nu}(y) - \kappa_{\tilde\nu}(y) = k$ für ein $\tilde\nu$ für alle $k \in \mathbb{N}$ äquivalent sind. (Wähle $T_\nu = \tau_{\tilde\nu}(y)$ bzw. $K_{\nu-1} = \kappa_{\tilde\nu}(y)$.)

Da die Zufallsgrößen $(L_\nu(y))_{\nu\in\mathbb{N}}$ und $(I_\nu(y))_{\nu\in\mathbb{N}}$ iid sind mit den Verteilungen von $L(y)$ bzw. $I(y)$, erhält man für alle $k \in \mathbb{N}$ und für alle $\tilde\nu \in \mathbb{N}$:

$$P(L(y) = k | I(y) = 1) = P(\tau_{\tilde\nu}(y) - \kappa_{\tilde\nu}(y) = k).$$

□

Auf diese Bezeichnungen wird später zurückgegriffen.

2.2 Zwei starke Grenzwertsätze

Unter den Voraussetzungen aus dem vorhergehenden Abschnitt werden nun zwei starke Grenzwertsätze formuliert, welche im nächsten Abschnitt bewiesen werden.
Das erste Theorem macht eine Aussage über die Länge von Segmenten mit hohem Score.

Theorem 2.2.1. *Mit obigen Bezeichnungen sei ν der erste Index, für den $I_\nu(y) = 1$ gilt, d.h. $\nu = \min\{k \geq 1 : I_k(y) = 1\}$. Dann folgt*

$$\frac{L_\nu(y)}{y} \xrightarrow[y \to \infty]{} \frac{1}{\omega^*} \quad \textit{fast sicher,}$$

wobei $\omega^* = E(X_1 e^{\theta^* X_1})$ und θ^* die eindeutig bestimmte positive Lösung der Gleichung $E(e^{\theta X_1}) = 1$ ist.

Anmerkung 2.2.2. (Lemma von Wald)
Es existiert eine eindeutig bestimmte positive Lösung θ^* von $E(e^{\theta X_1}) = 1$ und es gilt $\omega^* := E(X_1 e^{\theta^* X_1}) > 0$.

Beweis der Anmerkung. Nach Voraussetzung gilt $|X_1| \leq K$. Aufgrund der Analytizität und strengen Konvexität der Exponentialfunktion und nach Anwendung des Satzes von der dominierten Konvergenz, sowie der Linearität des Erwartungswertes gilt, dass die Funktion

$$\phi(\theta) := E(e^{\theta X_1})$$

analytisch, d.h. in allen Punkten in eine Potenzreihe entwickelbar, und streng konvex ist.
Somit erhält man

(i) ϕ ist unendlich oft stetig differenzierbar

(ii) $\phi''(\theta) > 0$ für alle $\theta \in \mathbb{R}$

(iii) $\phi(0) = 1$

(iv) $\phi'(0) = E(X_1 e^{\theta X_1})|_{\theta=0} = E(X_1) < 0$ nach Voraussetzung

(v)

$$\phi(\theta) = \int_{\{X_1 \leq 0\}} e^{\theta X_1} dP + \int_{\{X_1 > 0\}} e^{\theta X_1} dP \geq \int_{\{X_1 > 0\}} e^{\theta X_1} dP \xrightarrow[\theta \to +\infty]{} +\infty,$$

da $P(X_1 > 0) > 0$ nach Voraussetzung.
Somit existiert nach dem Mittelwertsatz genau ein $\theta^* > 0$ mit $\phi(\theta^*) = 1$ und $\phi'(\theta^*) = E(X_1 e^{\theta^* X_1}) := \omega^* > 0$. $\square$

Jetzt soll eine Aussage über die Zusammensetzung von hochscorigen Segmenten getroffen werden.

Theorem 2.2.3. *Mit obigen Bezeichnungen sei* ν *der erste Index, für den* $I_\nu(y) = 1$ *gilt, d.h.* $\nu = \min\{k \geq 1 : I_k(y) = 1\}$. *Dann folgt*

$$\frac{W_\nu(y)}{L_\nu(y)} \quad \xrightarrow[y\to\infty]{} \quad u^* \quad \textit{fast sicher,}$$

wobei $u^* = E(U_1 e^{\theta^* X_1})$ *und* θ^* *die eindeutig bestimmte positive Lösung der Gleichung* $E(e^{\theta X_1}) = 1$ *ist.*

Mit Hilfe dieses Theorems lassen sich nun wie schon angekündigt Aussagen über die Buchstabenhäufigkeiten in einem Segment (also z.B. über die relative Häufigkeit des Auftretens eines Aminosäurepaares in einem Paar verwandter Sequenzen; siehe auch Beispiel 1.2.2) treffen.
Dazu sei

$$U_k := U(X_k) = \mathbb{I}_{\{s_i\}}(X_k) = \begin{cases} 1 & \text{falls } X_k = s_i \\ 0 & \text{sonst} \end{cases}.$$

Korollar 2.2.4. *Die empirische Häufigkeitsverteilung* $\mu_\nu(y)$ *der Buchstaben* $\{a_1, ..., a_r\}$ *beobachtet während der* ν*ten Epoche (wobei* ν *wie in den vorhergehenden Theoremen definiert ist) konvergiert für* $y \longrightarrow \infty$ *fast sicher gegen das Häufigkeitsmaß* μ^*, *welches den Wert* a_i *mit Wahrscheinlichkeit*

$$q_i = p_i e^{\theta^* s_i}, \quad 1 \leq i \leq r,$$

annimmt.

Beweis. Sei $U_k = \mathbb{I}_{\{s_i\}}(X_k)$, dann gibt

$$\mu_\nu(y) := \frac{W_\nu(y)}{L_\nu(y)} = \frac{\sum_{k=K_{\nu-1}+1}^{T_\nu(y)} U_k}{L_\nu(y)} = \frac{\sum_{k=K_{\nu-1}+1}^{T_\nu(y)} \mathbb{I}_{\{s_i\}}(X_k)}{T_\nu(y) - K_{\nu-1}}$$

die relative Häufigkeit des Auftretens des Buchstabens a_i während einer Exkursion an.

Nach Theorem 2.2.3 gilt also (sofern ein Score $\geq y$ in $T_\nu(y)$ erreicht wird):

$$\mu_\nu(y) \quad \xrightarrow[y\to\infty]{} \quad E(\mathbb{I}_{\{s_i\}}(X_1) e^{\theta^* X_1}) = P(X_1 = s_i) e^{\theta^* s_i} = p_i e^{\theta^* s_i} \quad \text{fast sicher.}$$

$\square$

Die sogenannten 'target frequencies' q_i (siehe Seite 5) in einem hochscorigen Segment einer zufälligen Sequenz können also auch mit Hilfe der sogenannten 'background frequencies' p_i ausgedrückt werden. D.h. alle Scores s_i sind als log-Likelihood-Ratio-Scores $\log(\frac{q_i}{p_i})$ mit geeigneter Basis des Logarithmus darstellbar. Da die 'background frequencies' in der Praxis einfach als relative Buchstabenhäufigkeiten in der jeweiligen Sequenz wählbar sind, reduziert sich die Suche nach einem optimalen Score-System also auf eine optimale Charakterisierung dieser 'target frequencies'.

Anmerkungen 2.2.5.
a) (q_i) ist gemäß der Definition von θ^* eine Wahrscheinlichkeitsdichte.
b) Allgemeiner gilt für die Zufallsgröße

$$U(X_k) = \mathbb{I}_A(X_k) = \begin{cases} 1 & \text{falls } X_k \in A \\ 0 & \text{sonst} \end{cases}, \; A \in \mathbb{B}, \; X_k \text{ nichtdiskret, aber beschränkt:}$$

$$\mu(y) \underset{y\to\infty}{\longrightarrow} E(\mathbb{I}_{\{A\}}(X_1)e^{\theta X_1}) \quad \text{fast sicher.}$$

2.3 Beweis der starken Grenzwertsätze

In diesem Abschnitt wird es darum gehen, die beiden Theoreme (siehe 2.2.1 und 2.2.3) zu beweisen. Dazu werden insgesamt acht Lemmata benötigt. Die Vorgehensweise dabei ist folgende:
Unter Ausnutzung einiger Eigenschaften des sogenannten Waldmartingals (siehe Lemma 2.3.2) wird gezeigt, dass die Wahrscheinlichkeit, dass $I_\nu(y) = 1$ ist, also ein Score $\geq y$ in der ν-ten Epoche erreicht wird, exponentiell klein ist (siehe Lemma 2.3.3). Mit Hilfe dieser Aussage lassen sich zwei Abschätzungen für die Zufallsgrößen $\frac{L_\nu(y)}{y}$ und $\frac{W_\nu(y)}{L_\nu(y)}$ gewinnen (siehe Lemma 2.3.4 und 2.3.5), mit welchen man die Aussagen der Theoreme für den Fall $y = n$, $n \in \mathbb{N}$, $\nu \in \{1, ..., j_n\}$ (j_n wird später charakterisiert) zeigen kann (siehe Lemma 2.3.6 und 2.3.7)
Um auf den Fall zu schließen, bei dem $y > 0$ beliebig wählbar und $\nu = \min\{k \geq 1 : I_k(y) = 1\}$ ist, werden zwei weitere Lemmata benötigt (siehe Lemma 2.3.8 und 2.3.9). Diese enthalten Aussagen über das Verhältnis von $\kappa_\nu(n)$ zu $\kappa_1(y)$ bzw. $\tau_\nu(n)$ zu $\tau_1(y)$, so dass man unter Ausnutzung der Beziehung $\tau_1(y) - \kappa_1(y) = L_\nu(y)$ falls $\nu = \min\{k \geq 1 : I_k(y) = 1\}$ die Aussagen der Theoreme auch für den Fall, bei dem $y > 0$ beliebig ist, beweisen kann.
Genauere Erläuterungen dazu erfolgen später.

Da die (X_i) iid sind und somit auch die Zufallsgrößen $(I_\nu(y))_{\nu\in\mathbb{N}}$ bzw. $(L_\nu(y))_{\nu\in\mathbb{N}}$ mit den Verteilungen von $I_1(y) =: I(y)$ und $L_1(y) =: L(y)$ (manchmal auch als L abgekürzt), werden in den folgenden Beweisen häufig die Indizes weggelassen.

Im Folgenden seien θ^*, ω^* und u^* bestimmt wie in den Theoremen 2.2.1 und 2.2.3.

Zunächst sei nun also auf die Eigenschaften des Waldmartingals eingegangen. Dazu benötigt man das Optional-Sampling-Theorem.

Theorem 2.3.1. ***(Optional-Sampling-Theorem)*** *(abgeändert nach [19], S.377)*
$(X_{\mathbb{N}}, \mathcal{S}_{\mathbb{N}})$ sei ein Martingal, τ eine Stoppregel, für die $E(X_\tau)$ existiert und es gelte

$$\lim_{n\to\infty} \int_{\{\tau>n\}} |X_n| \, dP = 0.$$

Dann folgt $E(X_\tau) = E(X_1)$.

Zum *Beweis*: siehe z.B. Schmitz „Vorlesungen über Wahrscheinlichkeitstheorie“ ([19]), Seite 377.

Lemma 2.3.2. ***(Eigenschaften des Waldmartingals)***
Die durch

$$Y_m := \frac{e^{\theta S_m + tW_m}}{[\psi(\theta, t)]^m}, \quad m \in \mathbb{N}_0, \quad \textit{wobei } \psi(\theta, t) = E(e^{\theta X_1 + tU_1})$$

gegebene Zufallsgröße erfüllt folgende Eigenschaften:

(i) $(Y_m, \mathcal{S}_m)_{m\in\mathbb{N}_0}$ *ist ein Martingal.*

(ii) (Y_m) *und* $L := L_1(y)$ *genügen für genügend kleines* t *und* $\theta \geq \theta^*$ *den Voraussetzungen des Optional-Sampling-Theorems.*

(iii) $E(Y_L) = 1$ *für genügend kleines* t *und* $\theta \geq \theta^*$.

Beweis. zu (i): zu zeigen ist: $E(Y_m | \mathcal{S}_n) = Y_n$ für alle n, m mit $n < m$, wobei

$$\mathcal{S}_n = \sigma\left(\bigcup_{i,j=1}^{n} (X_i, U_j)^{-1}(\mathbb{B}^2)\right).$$

$$\begin{aligned}
E\left(\frac{e^{\theta S_m+tW_m}}{[\psi(\theta,t)]^m}\Big|\ \mathcal{S}_n\right) &= \frac{1}{[\psi(\theta,t)]^m}E(e^{\theta S_n+tW_n}e^{\theta\sum_{i=n+1}^m X_i+t\sum_{i=n+1}^m U_i}|\ \mathcal{S}_n) \\
&= \frac{e^{\theta S_n+tW_n}}{[\psi(\theta,t)]^m}E(e^{\theta\sum_{i=n+1}^m X_i+t\sum_{i=n+1}^m U_i}|\ \mathcal{S}_n) \\
&\quad \text{(da } S_n \text{ und } W_n \text{ messbar bzgl. } \mathcal{S}_n\text{)} \\
&= \frac{e^{\theta S_n+tW_n}}{[\psi(\theta,t)]^m}E(e^{\theta\sum_{i=n+1}^m X_i+t\sum_{i=n+1}^m U_i}) \\
&\quad \text{(nach Weglassen stochastisch unabhängiger Bedingung)} \\
&= \frac{e^{\theta S_n+tW_n}}{[\psi(\theta,t)]^m}[E(e^{\theta X_1+tU_1})]^{m-n} \\
&\quad \text{(da } (X_i) \text{ und } (U_i) \text{ identisch verteilt)} \\
&= \frac{e^{\theta S_n+tW_n}}{[\psi(\theta,t)]^n}.
\end{aligned}$$

zu (ii): $L = L_1(y)$ ist nach Definition eine Stoppzeit;
nach (i) gilt, dass $(Y_m, \mathcal{S}_m)_{m\in\mathbb{N}_0}$ ein Martingal ist;
$E(Y_L)$ existiert, denn:

$$\begin{aligned}
E|Y_L| = E\left|\frac{e^{\theta S_L+tW_L}}{[\psi(\theta,t)]^L}\right| &\le E\left(\frac{e^{|\theta||S_L|+|t||W_L|}}{|\ [\psi(\theta,t)]^L|}\right) \\
&\le e^{|\theta|(y+K)}\sum_{l=1}^{\infty}E\left(\frac{e^{|t||W_l|}}{|\ [\psi(\theta,t)]^l|}\ \mathbb{I}_{\{L=l\}}\right) \\
&\le e^{|\theta|(y+K)}\sum_{l=1}^{\infty}\left(\frac{e^{K'|t|}}{|\ [\psi(\theta,t)]|}\right)^l P(L=l) \quad < \infty
\end{aligned}$$

für $\dfrac{e^{K'|t|}}{|\psi(\theta,t)|} \le 1$, also für $|t|$ genügend klein und $\theta \ge \theta^*$, da

$$|S_L| \le |S_{L-1}| + |X_L| \le y + K, \quad \text{sowie} \quad |W_L| = |W_1(y)| \le \sum_{i=1}^{L(y)}|U_i| \le L(y)K'$$

(denn es gilt: $|X_i| \le K, \quad |U_i| \le K'$ und
$L = L_1(y) = T_1(y) = \min\{m > 0 : S_m \ge y \text{ oder } S_m \le 0\}$).

Ferner gilt für genügend kleines $|t|$ und $\theta \ge \theta^*$:

$$\begin{aligned}
\int_{\{L>m\}}|Y_m|\ dP = \int_{\{L>m\}}\left|\frac{e^{\theta S_m+tW_m}}{[\psi(\theta,t)]^m}\right|\ dP \\
&\le \frac{e^{|\theta|y+|t|mK'}}{|\ [\psi(\theta,t)]^m|}\int_{\{L>m\}}dP \\
&= \frac{e^{|\theta|y+|t|mK'}}{|\ [\psi(\theta,t)]^m|}P(L>m) \underset{m\to\infty}{\longrightarrow} 0,
\end{aligned}$$

denn nach dem Schwachen Gesetz der großen Zahlen gilt:

$$P(L > m) \leq P(S_m \in (0, y)) \leq P(S_m > 0) \underset{m\to\infty}{\longrightarrow} 0.$$

zu (iii): Mit dem Optional-Sampling-Theorem folgt für genügend kleines t und $\theta \geq \theta^*$: $E(Y_L) = E(Y_1)$.
Daher gilt:

$$E\left(\frac{e^{\theta S_L + tW_L}}{[\psi(\theta,t)]^L}\right) = E\left(\frac{e^{\theta X_1 + tU_1}}{E(e^{\theta X_1 + tU_1})}\right) = 1.$$

□

Formuliert man die Aussage (iii) aus Lemma 2.3.2 um, so erhält man für genügend kleines t und $\theta \geq \theta^*$ folgende Gleichheit:

$$E(e^{\theta S_L + tW_L - L\zeta(\theta,t)}) = 1, \quad \text{wobei } \zeta(\theta,t) = \log(E(e^{\theta X_1 + tU_1})) \tag{2.5}$$

Im Folgenden wird nun wie angekündigt bewiesen, dass $P(I_\nu(y) = 1)$ exponentiell klein ist.

Lemma 2.3.3. *Für jedes $\nu \geq 1$ ist die Wahrscheinlichkeit $P(I_\nu(y) = 1)$ exponentiell klein, d.h. es gilt:*

$$0 < \delta \leq P(I_\nu(y) = 1)e^{\theta^* y} \leq 1 \qquad \textit{für alle } y > 0 \textit{ und geeignetes } \delta > 0.$$

Beweis. Wählt man $\theta = \theta^*$ und $t = 0$, so gilt nach Definition von θ^* :

$$\zeta(\theta^*, 0) = \log(E(e^{\theta^* X_1})) = 1.$$

Somit folgt mit (2.5):

$$\begin{aligned} 1 \equiv E(e^{\theta^* S_L}) &= P(I(y) = 1)e^{\theta^* y} E(e^{\theta^*(S_L - y)} | I(y) = 1) \\ &\quad + P(I(y) = 0)E(e^{\theta^* S_L} | I(y) = 0). \end{aligned} \tag{2.6}$$

Ferner gilt:

(i) $\theta^* > 0$ nach Lemma 2.2.2

(ii) $y \leq S_L \leq y + K$ falls $I(y) = 1$ (siehe oben)

(iii) $-L(y)K \leq S_L \leq 0$ falls $I(y) = 0$ (siehe oben)

Die Wahrscheinlichkeit $P(I(y)=1)$ lässt sich also nach (2.6) folgendermaßen nach oben abschätzen:

$$\begin{aligned}1 &= P(I(y)=1)e^{\theta^* y}E(e^{\theta^*(S_L-y)}|I(y)=1) + P(I(y)=0)E(e^{\theta^* S_L}|I(y)=0)\\ &\geq P(I(y)=1)e^{\theta^* y} \quad \text{nach (i) und (ii).}\end{aligned}$$

Umgekehrt folgt aber auch:

$$\begin{aligned}1 &= P(I(y)=1)e^{\theta^* y}E(e^{\theta^*(S_L-y)}|I(y)=1) + P(I(y)=0)E(e^{\theta^* S_L}|I(y)=0)\\ &\leq P(I(y)=1)e^{\theta^* y}e^{\theta^* K} + P(I(y)=0)E(e^{\theta^* S_L}|I(y)=0) \quad \text{nach (ii).}\end{aligned}$$

Da zusätzlich die Teilmengenbeziehung $\{X_1 < 0\} \subset \{I(y)=0\}$ (falls $X_1 < 0$, so gilt auch $S_1 < 0$ und somit $I_1(y) = 0$) erfüllt ist, gilt

$$\begin{aligned}P(I(y)=1)e^{\theta^* y}e^{\theta^* K} &\geq 1 - P(I(y)=0)E(e^{\theta^* S_L}|I(y)=0)\\ &= 1 - (P(I(y)=0)P(X_1<0|I(y)=0)E(e^{\theta^* S_L}|X_1<0, I(y)=0)\\ &\quad + P(I(y)=0)P(X_1\geq 0|I(y)=0)E(e^{\theta^* S_L}|X_1\geq 0, I(y)=0))\\ &= 1 - P(X_1<0, I(y)=0)E(e^{\theta^* S_L}|X_1<0, I(y)=0)\\ &\quad - P(X_1\geq 0, I(y)=0)E(e^{\theta^* S_L}|X_1\geq 0, I(y)=0)\\ &\geq 1 - P(X_1<0)E(e^{\theta^* X_1}|X_1<0)\\ &\quad - P(X_1\geq 0)E(e^{\theta^* S_L}|X_1\geq 0, I(y)=0)\\ &\geq 1 - P(X_1<0)E(e^{\theta^* X_1}|\ X_1<0) - P(X_1\geq 0) \quad \text{(nach (iii))}\\ &= P(X_1<0)E(1-e^{\theta^* X_1}|\ X_1<0).\end{aligned}$$

Also erhält man

$$P(X_1<0)E(1-e^{\theta^* X_1}|\ X_1<0) \leq P(I(y)=1)e^{\theta^* y}e^{\theta^* K}.$$

Setzt man nun

$$\delta := P(X_1<0)E(1-e^{\theta^* X_1}|\ X_1<0)e^{-\theta^* K},$$

so ergibt sich

$$\delta \leq P(I(y)=1)e^{\theta^* y}.$$

Es sei angemerkt, dass $\delta > 0$ ist, denn es gilt $E(X_1) < 0$, also $P(X_1 < 0) > 0$ und da $\theta^* > 0$, folgt $E(e^{\theta^* X_1}|\ X_1 < 0) < 1$ bzw. $E(1 - e^{\theta^* X_1}|\ X_1 < 0) > 0$.

□

Ein nächster Schritt in die Richtung, den Beweis der fast sicheren Konvergenz von $\frac{L(y)}{y}$ bzw. $\frac{W(y)}{L(y)}$ in Angriff zu nehmen, sind wie schon oben erwähnt die folgenden Abschätzungen.

Lemma 2.3.4. *Für $y \longrightarrow \infty$ gilt:*

$$E\left(\left(\frac{L(y)}{y} - \frac{1}{\omega^*}\right)^4 \Bigm| I(y)=1\right) = O\left(\frac{1}{y^2}\right).$$

Beweis. Die linke Seite des Ausdrucks (2.5) $E(e^{\theta S_L + tW_L - L\zeta(\theta,t)}) = 1$ ist analytisch und somit unendlich oft stetig differenzierbar für alle θ und $|t|$ genügend klein.
Setzt man $t = 0$ und leitet an der Stelle $\theta = \theta^*$ ab, so ergibt sich mit $\phi(\theta) = E(e^{\theta X_1})$:

$$\begin{aligned}
&\frac{d}{d\theta} E(e^{\theta S_L - L\log(\phi(\theta))})\Big|_{\theta=\theta^*} = 0\\
\Leftrightarrow\ & E\left(e^{\theta S_L - L\log(\phi(\theta))}\left(S_L + (-L)\frac{\phi'(\theta)}{\phi(\theta)}\right)\right)\Big|_{\theta=\theta^*} = 0\\
\Leftrightarrow\ & E(e^{\theta^* S_L}(S_L - \phi'(\theta^*)L)) = 0 \quad (\text{da } \phi(\theta^*) = 1)
\end{aligned}$$

Man erhält also:

$$E((S_L - \omega^* L)e^{\theta^* S_L}) = 0, \quad \text{wobei } \omega^* = \phi'(\theta^*) = E(X_1 e^{\theta^* X_1}) > 0. \tag{2.7}$$

Weiteres Differenzieren an der Stelle $\theta = \theta^*$ ergibt:

$$\begin{aligned}
&\frac{d^2}{d\theta^2} E(e^{\theta S_L - L\log(\phi(\theta))})\Big|_{\theta=\theta^*}\\
&= E\left(e^{\theta S_L - L\log(\phi(\theta))}\left(S_L\left(S_L - \frac{L\phi'(\theta)}{\phi(\theta)}\right) - L\frac{d}{d\theta}\left[\frac{\phi'(\theta)}{\phi(\theta)}\right] - L\frac{\phi'(\theta)}{\phi(\theta)}\left(S_L - \frac{\phi'(\theta)}{\phi(\theta)}\right)\right)\right)\Big|_{\theta=\theta^*}\\
&= E\left(e^{\theta^* S_L}(S_L - \omega^* L)^2 - \frac{d}{d\theta}\left[\frac{\phi'(\theta)}{\phi(\theta)}\right]\Big|_{\theta=\theta^*} E(Le^{\theta^* S_L})\right) = 0.
\end{aligned}$$

Es gilt also:

$$E((S_L - \omega^* L)^2 e^{\theta^* S_L}) = k(\theta^*)E(Le^{\theta^* S_L}), \quad \text{wobei } k(\theta) = \frac{d}{d\theta}\left[\frac{\phi'(\theta)}{\phi(\theta)}\right]. \tag{2.8}$$

Nach dreimaligem Ableiten erhält man schließlich:

$$\begin{aligned}
E((S_L - \omega^* L)^4 e^{\theta^* S_L}) &= 6k(\theta^*)E((S_L - \omega^* L)^2 L e^{\theta^* S_L})\\
&\quad + 4k'(\theta^*)E((S_L - \omega^* L)Le^{\theta^* S_L})\\
&\quad - 3(k(\theta^*))^2 E(L^2 e^{\theta^* S_L}) + k''(\theta^*)E(Le^{\theta^* S_L}).
\end{aligned} \tag{2.9}$$

Im Weiteren werden nun schrittweise die folgenden Behauptungen bewiesen:

(i) $E(Le^{\theta^* S_L}) = O(y)$

(ii) $\xi^2 := E(L^2 e^{\theta^* S_L}) = O(y^2)$

(iii) $\eta^2 := E((S_L - \omega^* L)^4 e^{\theta^* S_L}) = O(y^2)$

(iv) $E((y - \omega^* L(y))^4 |\ I(y) = 1) = O(y^2)$

zu (i): Mit (2.7) folgt, dass

$$\begin{aligned}\omega^* E(Le^{\theta^* S_L}) &= E(S_L e^{\theta^* S_L}) \\ &= P(I(y)=1)e^{\theta^* y} E(S_L e^{\theta^*(S_L-y)} | \, I(y)=1) \\ &\quad + P(I(y)=0)E(S_L e^{\theta^* S_L} | \, I(y)=0)\end{aligned}$$

gilt. Somit erhält man für genügend großes y und geeignete C_1 und C_2:

$$\begin{aligned}E(Le^{\theta^* S_L}) &= \frac{1}{\omega^*}[P(I(y)=1)e^{\theta^* y} E(S_L e^{\theta^*(S_L-y)} | \, I(y)=1) \\ &\quad + P(I(y)=0)E(S_L e^{\theta^* S_L} | \, I(y)=0)] \\ &\in \; [C_1 y, C_2 y].\end{aligned}$$

Dies folgt aus Lemma 2.3.3, denn es gilt $P(I(y)=1)e^{\theta^* y} \in [\delta, 1]$, sowie $y \leq S_L \leq y+K$ unter $I(y)=1$ bzw. $-KL(y) \leq S_L \leq 0$ unter $I(y)=0$.

zu (ii): Mit (2.8) und Behauptung (i) folgt, dass

$$E((S_L - \omega^* L)^2 e^{\theta^* S_L}) = k(\theta^*) E(Le^{\theta^* S_L}) = O(y)$$

gilt. Ferner ist

$$\begin{aligned}D^2 := E(S_L^2 e^{\theta^* S_L}) &= P(I(y)=1)e^{\theta^* y} E(S_L^2 e^{\theta^*(S_L-y)} | \, I(y)=1) \\ &\quad + P(I(y)=0)E(S_L^2 e^{\theta^* S_L} | \, I(y)=0) \\ &\in \; [C_1 y^2, C_2 y^2] \quad \text{(analog zu (i))}.\end{aligned}$$

Man erhält also, dass $D^2 = O(y^2)$.
Mit Hilfe dieser beiden Aussagen lässt sich folgern:

$$\omega^{*2} E(L^2 e^{\theta^* S_L}) - 2\omega^* E(S_L L e^{\theta^* S_L}) + E(S_L^2 e^{\theta^* S_L}) = O(y).$$

Weiterhin folgt:

$$\begin{aligned}\omega^{*2}\xi^2 &= 2\omega^* E(S_L L e^{\theta^* S_L}) - O(y^2) + O(y) \\ &\leq 2\omega^* \xi D + E.\end{aligned}$$

Hierbei ist gerade $D(y) = O(y)$ und $E = O(y^2)$.
Dies erhält man aus der Cauchy-Schwarzschen Ungleichung und aus $E(e^{\theta^* S_L}) = 1$, sowie aus der Beschränktheit von S_L:

$$E(S_L L e^{\theta^* S_L}) \leq (E(S_L^2 e^{\theta^* S_L}) E(L^2 e^{\theta^* S_L}))^{1/2} = D\xi.$$

Die größte Wurzel aus dieser quadratischen Ungleichung in ω^* ist sicherlich durch Cy für eine geeignete Konstante $C > 0$ beschränkt, also $\xi \leq Cy$, da ξ positiv ist, folgt $\xi^2 = O(y^2)$.

zu (iii): Nach Cauchy-Schwarzscher Ungleichung gilt:

$$E((S_L - \omega^* L)^2 L e^{\theta^* S_L}) \leq (E((S_L - \omega^* L)^4 e^{\theta^* S_L}) E(L^2 e^{\theta^* S_L}))^{1/2} = \eta\xi$$
$$E((S_L - \omega^* L) L e^{\theta^* S_L}) \leq (E((S_L - \omega^* L)^2 e^{\theta^* S_L}) E(L^2 e^{\theta^* S_L}))^{1/2} \leq \sqrt{\eta\xi}\xi.$$

Damit folgt zusammen mit (2.9):

$$\begin{aligned}
\eta^2 &= 6k(\theta^*)E((S_L - \omega^* L)^2 L e^{\theta^* S_L}) \\
&\quad +4k'(\theta^*)E((S_L - \omega^* L) L e^{\theta^* S_L}) \\
&\quad -3(k(\theta^*))^2 E(L^2 e^{\theta^* S_L}) + k''(\theta^*)E(L e^{\theta^* S_L}) \\
&\leq 6k(\theta^*)\eta\xi + 4k'(\theta^*)\sqrt{\eta\xi}\xi - 3(k(\theta^*))^2\xi^2 + k''(\theta^*)E(L e^{\theta^* S_L}) \\
&= O(y)\eta + O(y^2) \text{ nach (i) und (ii).}
\end{aligned}$$

Somit erhält man analog zur obigen Argumentation im Beweis zur Behauptung (ii): $\eta^2 = O(y^2)$.

zu (iv): Da nach Lemma 2.3.3

$$0 < \delta \leq P(I_\nu(y) = 1)e^{\theta^* y}$$

gilt, folgt mit (iii) unmittelbar, dass

$$E((S_L - \omega^* L(y))^4 e^{\theta^*(S_L - y)} | I(y) = 1) = O(y^2).$$

Somit folgt mit der Minkowskischen Ungleichung:

$$\begin{aligned}
&E((y - \omega^* L(y))^4 | I(y) = 1) \\
&= E((S_L - \omega^* L(y) + y - S_L)^4 | I(y) = 1) \\
&\leq E((S_L - \omega^* L(y))^4 | I(y) = 1) + E((y - S_L)^4 | I(y) = 1) \\
&\leq E((S_L - \omega^* L(y))^4 e^{\theta^*(S_L - y)} | I(y) = 1) + E((y - S_L)^4 | I(y) = 1) \\
&= O(y^2)
\end{aligned}$$

da $0 \leq S_L - y \leq K$ unter $I(y) = 1$.

Zusammen mit (iv) erhält man schließlich die gewünschte Aussage:

$$E\left(\left(\frac{L(y)}{y} - \frac{1}{\omega^*}\right)^4 \Bigg| \, I(y) = 1\right) = E\left(\frac{(y - \omega^* L(y))^4}{\omega^{*4} y^4} \Bigg| \, I(y) = 1\right) = O\left(\frac{1}{y^2}\right).$$

□

Lemma 2.3.5. *Für $y \longrightarrow \infty$ gilt:*

$$E\left(\left(\frac{W(y)}{L(y)} - u^*\right)^4 \Bigg| \, I(y) = 1\right) = O\left(\frac{1}{y^2}\right).$$

Beweis. Der Beweis erfolgt ganz analog zum Beweis von Lemma 2.3.4.
Der Ausdruck (2.5)

$$E(e^{\theta S_L + tW_L - L\zeta(\theta,t)}) = 1, \quad \text{wobei } \zeta(\theta,t) = \log(E(e^{\theta X_1 + tU_1}))$$

wird nun dreimal nach t an der Stelle $t = 0$ abgeleitet, wobei man $\theta = \theta^*$ setzt. Man erhält:

$$E((W_L - u^*L)e^{\theta^* S_L}) = 0 \text{ mit } u^* = u(\theta^*, 0) = E(U_1 e^{\theta^* X_1}),\ u(\theta,t) = \frac{d\zeta}{dt}(\theta,t) \tag{2.10}$$

$$E((W_L - u^*L)^2 e^{\theta^* S_L}) = \kappa(\theta^*, 0)E(Le^{\theta^* S_L}) \quad \text{mit } \kappa(\theta,t) = \frac{d^2\zeta}{dt^2}(\theta,t) \tag{2.11}$$

$$\begin{aligned} E((W_L - u^*L)^4 e^{\theta^* S_L}) = {} & 6\kappa(\theta^*, 0)E((W_L - u^*L)^2 L e^{\theta^* S_L}) \\ & + 4\frac{d\kappa}{dt}(\theta^*, 0)E((W_L - u^*L)Le^{\theta^* S_L}) \\ & - 3(\kappa(\theta^*, 0))^2 E(L^2 e^{\theta^* S_L}) + \frac{d^2\kappa}{dt^2}(\theta^*, 0)E(Le^{\theta^* S_L}). \end{aligned} \tag{2.12}$$

Nun gelten folgende Aussagen:

(i) $E(Le^{\theta^* S_L}) = O(y)$

(ii) $\xi^2 := E(L^2 e^{\theta^* S_L}) = O(y^2)$

(iii) $\zeta_0^2 := E((W_L - u^*L)^4 e^{\theta^* S_L}) = O(y^2)$

(iv) $E((W_L - u^*L)^4 |\ I(y) = 1) = O(y^2)$

Die Aussagen (i) und (ii) wurden schon in Lemma 2.3.4 bewiesen.

zu (iii): Nach Cauchy-Schwarzscher Ungleichung gilt:

$$\begin{aligned} E((W_L - u^*L)^2 Le^{\theta^* S_L}) &\le (E((W_L - u^*L)^4 e^{\theta^* S_L})E(L^2 e^{\theta^* S_L}))^{1/2} = \zeta_0\xi \\ E((W_L - u^*L)Le^{\theta^* S_L}) &\le (E((W_L - u^*L)^2 e^{\theta^* S_L})E(L^2 e^{\theta^* S_L}))^{1/2} \le \sqrt{\zeta_0\xi}\xi. \end{aligned}$$

Damit folgt zusammen mit (2.12):

$$\begin{aligned}\zeta_0^2 &= 6\kappa(\theta^*,0)E((W_L-u^*L)^2Le^{\theta^*S_L})\\ &\quad +4\frac{d\kappa}{dt}(\theta^*,0)E((W_L-u^*L)Le^{\theta^*S_L})\\ &\quad -3(\kappa(\theta^*,0))^2E(L^2e^{\theta^*S_L})+\frac{d^2\kappa}{dt^2}(\theta^*,0)E(Le^{\theta^*S_L}).\\ &\leq 6\kappa(\theta^*,0)\zeta_0\xi+4\frac{d\kappa}{dt}(\theta^*,0)\sqrt{\zeta_0\xi}\xi-3(\kappa(\theta^*,0))^2\xi^2+\frac{d^2\kappa}{dt^2}(\theta^*,0)E(Le^{\theta^*S_L})\\ &= O(y)\zeta_0+O(y^2) \text{ nach (i) und (ii).}\end{aligned}$$

Die größte Wurzel aus dieser quadratischen Ungleichung ist sicherlich $O(y)$, also da ζ_0 positiv ist, folgt $\zeta_0^2 = O(y^2)$.

zu (iv): Da nach Lemma 2.3.3

$$0 < \delta \leq P(I_\nu(y)=1)e^{\theta^*y}$$

gilt, folgt mit (iii) unmittelbar, dass

$$E((W_L-u^*L(y))^4e^{\theta^*(S_L-y)}|\ I(y)=1)=O(y^2).$$

Ferner ist

$$E((W_L-u^*L)^4|\ I(y)=1)\leq E((W_L-u^*L)^4e^{\theta^*(S_L-y)}|\ I(y)=1),$$

da $0 \leq S_L-y\leq K$ unter $I(y)=1$.
Somit erhalten wir

$$E((W_L-u^*L)^4|\ I(y)=1)=O(y^2).$$

Zusammen mit (iv) erhält man nun die gewünschte Aussage:

$$E\left(\left(\frac{W_{L(y)}}{L(y)}-u^*\right)^4\Bigg|\ I(y)=1\right)=E\left(\frac{1}{(L(y))^4}\ (W_{L(y)}-u^*L)^4\Big|\ I(y)=1\right)=O\left(\frac{1}{y^2}\right),$$

denn unter $I(y)=1$ gilt $L(y)\geq \frac{y}{K}$, da $y\leq S_L\leq L(y)K$.

$\square$

Sei nun $y=n$. Man geht wie angekündigt in der Beweisführung weiter vor, indem man zunächst zeigt, dass für alle $\nu\in\{1,...,j_n\}$ (für ein geeignetes j_n) unter $I_\nu(n)=1$ gilt:

(i)

$$\frac{L_\nu(n)}{n}\ \xrightarrow[n\to\infty]{}\ \frac{1}{\omega^*}\quad \text{fast sicher}$$

(ii)
$$\frac{W_\nu(n)}{L_\nu(n)} \xrightarrow[n\to\infty]{} u^* \quad \text{fast sicher}$$

Oder anders gesagt, da $P^{L_\nu(y)|I_\nu(y)=1} = P^{\tau_\nu(y)-\kappa_\nu(y)}$ (vergleiche Proposition 2.1.1), dass für alle $\nu \in \{1, ..., j_n\}$ gilt:

(i)
$$\frac{\tau_\nu(n) - \kappa_\nu(n)}{n} \xrightarrow[n\to\infty]{} \frac{1}{\omega^*} \quad \text{fast sicher}$$

(ii)
$$\frac{\sum_{i=\kappa_\nu(n)+1}^{\tau_\nu(n)} U_i}{\tau_\nu(n) - \kappa_\nu(n)} \xrightarrow[n\to\infty]{} u^* \quad \text{fast sicher}$$

Dabei wählt man $j_n = [A \log n]$, wobei A eine feste Konstante ist mit $-A\log(1 - P(0 < S_1 < S_2 < ... < S_{m_0-1} < 1 < S_{m_0})) > 1$ (diese Eigenschaft benötigt man im Beweis zu Lemma 2.3.8). Um die obigen Aussagen zu beweisen, stehen Lemma 2.3.4 und 2.3.5 zur Verfügung.

Lemma 2.3.6. $\lim_{n\to\infty} \max_{\nu=1,...,[A\log n]} \left| \frac{\tau_\nu(n) - \kappa_\nu(n)}{n} - \frac{1}{\omega^*} \right| = 0$ *fast sicher.*

Beweis. Aus Lemma 2.3.4 folgt, dass

$$E\left(\left(\frac{\tau_\nu(y) - \kappa_\nu(y)}{y} - \frac{1}{\omega^*}\right)^4\right) = O\left(\frac{1}{y^2}\right),$$

also gilt insbesondere für $y = n$:

$$E\left(\left(\frac{\tau_\nu(n) - \kappa_\nu(n)}{n} - \frac{1}{\omega^*}\right)^4\right) \le \frac{C}{n^2} \quad \text{für geeignetes } C < \infty, \text{ für alle } n \in \mathbb{N}$$

Mit der Markov-Ungleichung folgt somit für alle $\varepsilon > 0$, dass

$$P\left(\left|\frac{\tau_\nu(n) - \kappa_\nu(n)}{n} - \frac{1}{\omega^*}\right| > \varepsilon\right) \le \frac{E\left(\left(\frac{\tau_\nu(n) - \kappa_\nu(n)}{n} - \frac{1}{\omega^*}\right)^4\right)}{\varepsilon^4} \le \frac{C}{\varepsilon^4 n^2} =: \frac{C_\varepsilon}{n^2}$$

Somit ergibt sich

$$\sum_{n=1}^{\infty} \sum_{\nu=1}^{[A\log n]} P\left(\left|\frac{\tau_\nu(n) - \kappa_\nu(n)}{n} - \frac{1}{\omega^*}\right| > \varepsilon\right) \le \sum_{n=1}^{\infty} \sum_{\nu=1}^{[A\log n]} \frac{C_\varepsilon}{n^2} \le C_\varepsilon A \sum_{n=1}^{\infty} \frac{\log n}{n^2} < \infty.$$

Damit folgt zusammen mit dem Lemma von Borel-Cantelli, dass

$$P\left(\limsup_{n\to\infty}\sum_{\nu=1}^{[A\log n]}\left\{\left|\frac{\tau_\nu(n)-\kappa_\nu(n)}{y}-\frac{1}{\omega^*}\right|>\varepsilon\right\}\right)=0 \qquad \text{für alle } \varepsilon>0$$

gilt. Also ist auch

$$P\left(\lim_{n\to\infty}\max_{\nu=1,\dots,[A\log n]}\left|\frac{\tau_\nu(n)-\kappa_\nu(n)}{n}-\frac{1}{\omega^*}\right|>\varepsilon\right)=0 \qquad \text{für alle } \varepsilon>0.$$

Man erhält also schließlich:

$$P\left(\lim_{n\to\infty}\max_{\nu=1,\dots,[A\log n]}\left|\frac{\tau_\nu(n)-\kappa_\nu(n)}{n}-\frac{1}{\omega^*}\right|=0\right)=1.$$

□

Lemma 2.3.7. $\lim_{n\to\infty}\max_{\nu=1,\dots,[A\log n]}\left|\frac{\sum_{i=\kappa_\nu(n)}^{\tau_\nu(n)}U_i}{\tau_\nu(n)-\kappa_\nu(n)}-u^*\right|=0$ *fast sicher.*

Beweis. Der Beweis verläuft analog zum Beweis von Lemma 2.3.6. Nach Lemma 2.3.5 gilt

$$E\left(\left(\frac{\sum_{i=\kappa_\nu(n)+1}^{\tau_\nu(n)}U_i}{\tau_\nu(n)-\kappa_\nu(n)}-u^*\right)^4\right)\le\frac{C}{n^2} \qquad \text{für geeignetes } C<\infty,\ \text{für alle } n\in\mathbb{N}$$

Damit erhält man zusammen mit der Markov-Ungleichung für alle $\varepsilon>0$:

$$\sum_{n=1}^{\infty}\sum_{\nu=1}^{[A\log n]}P\left(\left|\frac{\sum_{i=\kappa_\nu(n)+1}^{\tau_\nu(n)}U_i}{\tau_\nu(n)-\kappa_\nu(n)}-u^*\right|>\varepsilon\right)\le\sum_{n=1}^{\infty}\sum_{\nu=1}^{[A\log n]}\frac{E\left(\left(\frac{\sum_{i=\kappa_\nu(n)+1}^{\tau_\nu(n)}U_i}{\tau_\nu(n)-\kappa_\nu(n)}-u^*\right)^4\right)}{\varepsilon^4}$$
$$\le\sum_{n=1}^{\infty}\sum_{\nu=1}^{[A\log n]}\frac{C_\varepsilon}{n^2}\le C_\varepsilon A\sum_{n=1}^{\infty}\frac{\log n}{n^2}<\infty.$$

Also folgt analog zum Beweis von Lemma 2.3.6 mit dem Lemma von Borel-Cantelli, dass

$$P\left(\lim_{n\to\infty}\max_{\nu=1,\dots,[A\log n]}\left|\frac{\sum_{i=\kappa_\nu(n)+1}^{\tau_\nu(n)}U_i}{\tau_\nu(n)-\kappa_\nu(n)}-u^*\right|=0\right)=1$$

gilt.

□

Man hat nun also die Aussagen der Theoreme für ein Level $y = n$ zur Verfügung, wobei $\nu \in \{1, ..., [A \log n]\}$. Um dies zu verallgemeinern auf den Fall, bei dem $y > 0$ beliebig, sowie $\nu = \min \{k \geq 1 : I_k(y) = 1\}$ ist, beweist man nun die folgenden zwei Lemmata.

Lemma 2.3.8. *Die erste Exkursion zum Level $n+1$ stimmt für genügend großes n fast sicher mit einer der ersten $[A \log n]$ Exkursionen zum Level n überein, d.h.*

$$P(\kappa_1(n+1) = \kappa_\nu(n) \textit{ für ein } \nu \in \{1, ..., [A \log n]\}) = 1$$

Beweis. Sei $\mathcal{E}_n$ gegeben durch

$$\begin{aligned}\mathcal{E}_n =& \{\text{keine der } j_n = [A \log n] \text{ Exkursionen, die durch } (\kappa_\nu(n), \tau_\nu(n), \sigma_\nu(n))_{\nu \in \{1,...,j_n\}} \\ & \text{gegeben sind, erreichen ein Level} \geq n+1\}\end{aligned}$$

Es gilt $P(\limsup\limits_{n\to\infty} \mathcal{E}_n) = 0$, denn:

Nach Voraussetzung gilt $P(X_1 > 0) > 0$. Also existiert ein $m_0 \in \mathbb{N}$, derart dass

$$\begin{aligned}& P(0 < S_1 < S_2 < ... < S_{m_0-1} < 1 < S_{m_0}) \\ =& P(0 < X_1 < X_1 + X_2 < ... < X_1 + ... + X_{m_0-1} < X_1 + ... + X_{m_0}) =: a > 0.\end{aligned}$$

Wähle nun A genügend groß mit $-A \log(1-a) := \gamma > 1$, dann gilt:

$$\begin{aligned}P(\mathcal{E}_n) =& P(S_{\tau_\nu(n)} - S_{\kappa_\nu(n)} < n+1, S_{\tau_\nu(n)+1} - S_{\kappa_\nu(n)} < n+1, ..., S_{\sigma_\nu(n)} - S_{\kappa_\nu(n)} < n+1 \\ & \text{für alle } \nu \in \{1, ..., j_n\}) \\ & (\text{da } \max_{\sigma_{\nu-1}(n) < k \leq l \leq \tau_\nu(n)} (S_l - S_k) < n \text{ nach Definition der Stoppzeiten}) \\ =& P(\bigcap_{\nu=1}^{j_n} \{S_{\tau_\nu(n)} - S_{\kappa_\nu(n)} < n+1, ..., S_{\sigma_\nu(n)} - S_{\kappa_\nu(n)} < n+1\}) \\ =& \prod_{\nu=1}^{j_n} P(S_{\tau_\nu(n)} - S_{\kappa_\nu(n)} < n+1, ..., S_{\sigma_\nu(n)} - S_{\kappa_\nu(n)} < n+1) \\ & (\text{wegen stochastischer Unabhängigkeit}) \\ \leq& \prod_{\nu=1}^{j_n} P(X_{\tau_\nu(n)+1} < 1, X_{\tau_\nu(n)+1} + X_{\tau_\nu(n)+2} < 1, ..., X_{\tau_\nu(n)+1} + ... + X_{\sigma_\nu(n)} < 1) \\ & (\text{da } S_{\tau_\nu(n)} - S_{\kappa_\nu(n)} \geq n \text{ nach Definition}) \\ \leq& \prod_{\nu=1}^{j_n} P(\{0 < X_1 < X_1 + X_2 < ... < X_1 + ... X_{m_0-1} < 1 < X_1 + ... + X_{m_0}\}^{\mathsf{C}}) \\ =& (1-a)^{j_n} = (1-a)^{[A \log n]} \leq \frac{C}{n^\gamma} \quad \text{für geeignetes } C < \infty,\end{aligned}$$

denn $(1-a)^{[A\log n]} \leq C(1-a)^{A\log n} = Cn^{A\log(1-a)} = \dfrac{C}{n^\gamma}$.

Es gilt also

$$\sum_{n=1}^{\infty} P(\mathcal{E}_n) \leq \sum_{n=1}^{\infty} \frac{C}{n^\gamma} < \infty, \quad \text{da } \gamma > 1.$$

Damit folgt zusammen mit dem Lemma von Borel-Cantelli, dass $P(\limsup_{n\to\infty} \mathcal{E}_n) = 0$ gilt.

Also gilt für genügend großes n mit Wahrscheinlichkeit 1, dass mindestens eine der ersten $[A\log n]$ Exkursionen zum Level n auch das Level $n+1$ erreicht.

□

Aus Lemma 2.3.8 können wir somit schließen, dass auch die erste Exkursion zum Level y mit $n \leq y \leq n+1$, fast sicher mit einer der ersten j_n Exkursionen zum Level n übereinstimmt, also die „Startpunkte" der Exkursionen fast sicher gleich sind:

$$\kappa_1(y) = \kappa_\nu(n) \text{ fast sicher für geeignetes } \nu \text{ falls } n \leq y \leq n+1.$$

Aber wie sieht es mit den „Endpunkten" $\tau_1(y)$ und $\tau_\nu(n)$ aus?
Um also die Aussagen der Theoreme

(i)

$$\frac{\tau_1(y) - \kappa_1(y)}{y} \xrightarrow[y\to\infty]{} \frac{1}{\omega^*} \quad \text{fast sicher}$$

(ii)

$$\frac{\sum_{i=\kappa_1(y)+1}^{\tau_1(y)} U_i}{\tau_1(y) - \kappa_1(y)} \xrightarrow[y\to\infty]{} u^* \quad \text{fast sicher}$$

(es gilt $\tau_1(y) - \kappa_1(y) = L_\nu(y)$, falls $\nu = \min\{k \geq 1 : I_k(y) = 1\}$)
zu zeigen, benötigt man noch eine Aussage über das Verhältnis von $\tau_\nu(n)$ zu $\tau_1(y)$. Dazu sei

$$T_n^* := \sup_{n \leq y < n+1} \min_{\substack{1\leq \nu \leq j_n \\ \tau_\nu(n) \leq \tau_1(y)}} (\tau_1(y) - \tau_\nu(n)).$$

Jetzt liegt es nahe zu zeigen, dass $\frac{T_n^*}{n} \xrightarrow[n\to\infty]{} 0$ fast sicher, denn dann steht eine Konvergenzaussage für $\tau_\nu(n)$ gegen $\tau_1(y)$ zur Verfügung, die es einem ermöglicht, die Aussagen der Theoreme 2.2.1 und 2.2.3 vom Fall $y = n$ auf den Fall $y > 0$ beliebig zu verallgemeinern.

Lemma 2.3.9.

$$\frac{T_n^*}{n} \xrightarrow[n\to\infty]{} 0 \quad \textit{fast sicher.}$$

Beweis. Es sei $M := \max_k S_k$ (M ist fast sicher wohldefiniert, da $E(X_1) < 0$) und J der Zeitpunkt, bei dem M das erste Mal angenommen wird, also

$$J = \inf\{k \geq 1 : S_k = M\}.$$

Dann gilt für alle $k \geq 1$ für ein geeignetes $b > 0$ nach der Markov-Ungleichung, da $\phi(\theta) = E(e^{\theta X_1}) < 1$ für $0 < \theta < \theta^*$ (vergleiche Beweis zur Anmerkung 2.2.2):

$$P(J = k) \leq P(S_k > 0) \leq \frac{E(e^{\theta S_k})}{e^{\theta 0}} = (E(e^{\theta X_1}))^k = (\phi(\theta))^k = e^{-bk}.$$

Also erhält man

$$P(J \geq k) = \sum_{i=k}^{\infty} P(J = i) \leq \sum_{i=k}^{\infty} e^{-bk} = e^{-bk} \sum_{i=k}^{\infty} e^{-bi} = e^{-bk}\tilde{C} \tag{2.13}$$

für ein geeignetes $\tilde{C}$, d.h. J hat einen exponentiellen Zerfall.

Für alle Exkursionen zum Level n sei nun $\tilde{J}_\nu(n)$ die Zeitspanne nach $\tau_\nu(n)$, die benötigt wird, bis $\max_{\tau_\nu(n) \leq k} (S_k - S_{\tau_\nu(n)})$ zuerst angenommen wird, also

$$\tilde{J}_\nu(n) = \inf\{j \geq 1 : S_{j+\tau_\nu(n)} - S_{\tau_\nu(n)} = \max_{\tau_\nu(n) \leq k} (S_k - S_{\tau_\nu(n)})\}.$$

Betrachte nun einen beliebigen Wert y mit $n \leq y \leq n + 1$. Nach Lemma 2.3.8 gilt für genügend großes n, dass die erste Exkursion zum Level y mit einer der Exkursionen aus $(\kappa_\nu(n), \tau_\nu(n), \sigma_\nu(n))_{\nu \in \{1,\dots,j_n\}}$ übereinstimmt, d.h. $\kappa_\nu(n) = \kappa_1(y)$ und somit $\tau_\nu(n) \leq \tau_1(y)$ (da $n \leq y$) für mindestens ein $\nu \in \{1, ..., j_n\}$. Für dieses ν und alle y mit $n \leq y \leq n + 1$ gilt also:

$$\max_{k \geq \tau_\nu(n)} (S_k - S_{\tau_\nu(n)}) \geq S_{\tau_1(y)} - S_{\tau_\nu(n)}$$

und somit

$$T_n^* = \sup_{n \leq y < n+1} \min_{\substack{1 \leq \nu \leq j_n \\ \tau_\nu(n) \leq \tau_1(y)}} (\tau_1(y) - \tau_\nu(n)) \leq \max\{\tilde{J}_1(n), \tilde{J}_2(n), ..., \tilde{J}_{j_n}(n)\} =: J_n^*.$$

Also gilt für alle $\varepsilon > 0$, da J und $\tilde{J}_\nu(n)$ dieselbe Verteilung besitzen:

$$\begin{aligned}
\sum_{n=1}^{\infty} P\left(\frac{T_n^*}{n} > \varepsilon\right) &\leq \sum_{n=1}^{\infty} P\left(\frac{J_n^*}{n} > \varepsilon\right) \\
&= \sum_{n=1}^{\infty} P(\tilde{J}_\nu(n) > n\varepsilon \text{ für ein } \nu \in \{1, ..., j_n\}) \\
&= \sum_{n=1}^{\infty} \sum_{\nu=1}^{j_n} P(J > n\varepsilon) \\
&\leq \tilde{C} \sum_{n=1}^{\infty} A \log(n) e^{-bn\varepsilon} < \infty \quad \text{nach (2.13).}
\end{aligned}$$

Damit folgt mit dem Lemma von Borel-Cantelli, dass

$$P\left(\limsup_{n\to\infty}\,\{\frac{T_n^*}{n} > \varepsilon\}\right) = P\left(\liminf_{n\to\infty}\,\{\frac{T_n^*}{n} > \varepsilon\}\right) = 0 \quad \text{für alle } \varepsilon > 0.$$

Also gilt fast sicher $\frac{T_n^*}{n} \longrightarrow 0$.

□

Jetzt steht alles zur Verfügung, um die Grenzwertsätze 2.2.1 und 2.2.3 zu beweisen.

Beweis von Theorem 2.2.1.
Folgende Aussagen stehen zur Verfügung:

(i)

$$\frac{\tau_\nu(n) - \kappa_\nu(n)}{n} \xrightarrow[n\to\infty]{} \frac{1}{\omega^*} \quad \text{fast sicher für alle } \nu \in \{1, ..., j_n\}$$

nach Lemma 2.3.6

(ii)

$$\max_{\substack{1\le\nu\le j_n \\ \tau_\nu(n)\le\tau_1(y)}} \kappa_\nu(n) = \kappa_1(y) \quad \text{fast sicher für genügend großes } n \text{ für alle } y \in [n, n+1)$$

(denn falls $\tau_\nu(n) \le \tau_1(y)$, so gilt auch $\kappa_\nu(n) \le \kappa_1(y)$ und nach Lemma 2.3.8 gilt für mindestens ein $\nu \in \{1, ..., j_n\} : \kappa_\nu(n) = \kappa_1(y)$)

(iii)

$$\min_{\substack{1\le\nu\le j_n \\ \tau_\nu(n)\le\tau_1(y)}} \left(\frac{\tau_1(y) - \tau_\nu(n)}{n}\right) \xrightarrow[n\to\infty]{} 0 \quad \text{fast sicher für alle } y \in [n, n+1)$$

nach Lemma 2.3.9

Somit können wir folgern

$$\begin{aligned}
\lim_{y\to\infty} \frac{\tau_1(y) - \kappa_1(y)}{y} &= \lim_{y\to\infty} \frac{\tau_1(y) - \max_{1\le\nu\le j_n, \tau_\nu(n)\le\tau_1(y)} \kappa_\nu(n)}{y} \qquad \text{(nach (ii))}\\
&= \lim_{y\to\infty} \min_{\substack{1\le\nu\le j_n \\ \tau_\nu(n)\le\tau_1(y)}} \frac{\tau_1(y) - \kappa_\nu(n)}{y}\\
&= \lim_{y\to\infty} \min_{\substack{1\le\nu\le j_n \\ \tau_\nu(n)\le\tau_1(y)}} \left(\left(\frac{\tau_1(y) - \tau_\nu(n)}{n}\right) + \left(\frac{\tau_\nu(n) - \kappa_\nu(n)}{n}\right)\right)\\
&= \frac{1}{\omega^*} \quad \text{fast sicher (nach (i) und (iii)).}
\end{aligned}$$

Also erhält man schließlich unter Ausnutzung der Tatsache, dass $\tau_1(y) - \kappa_1(y) = L_\nu(y)$, falls $\nu = \min\{k \geq 1 : I_k(y) = 1\}$, die gewünschte Aussage:

$$\frac{L_\nu(y)}{y} \xrightarrow[y\to\infty]{} \frac{1}{\omega^*} \quad \text{fast sicher,} \quad \text{falls } \nu = \min\{k \geq 1 : I_k(y) = 1\}.$$

□

Beweis von Theorem 2.2.3.

$$\text{Sei} \quad W_n^* := \sup_{n \leq y < n+1} \min_{1 \leq \nu \leq j_n} \left| \sum_{i=\kappa_1(y)+1}^{\tau_1(y)} U_i - \sum_{i=\kappa_\nu(n)+1}^{\tau_\nu(n)} U_i \right|.$$

Dann gilt

$$\frac{W_n^*}{n} \leq \sup_{n \leq y < n+1} \min_{\substack{1 \leq \nu \leq j_n \\ \tau_\nu(n) \leq \tau(y)}} \left(\max |U_i| \left| \frac{\tau_1(y) - \tau_\nu(y)}{n} \right| \right) = \max |U_i| \frac{T_n^*}{n} \xrightarrow[n\to\infty]{} 0$$

fast sicher nach Lemma 2.3.9.
Man hat also

(i)

$$\min_{1 \leq \nu \leq j_n} \left| \frac{\sum_{i=\kappa_1(y)+1}^{\tau_1(y)} U_i - \sum_{i=\kappa_\nu(n)+1}^{\tau_\nu(n)} U_i}{n} \right| \xrightarrow[n\to\infty]{} 0 \quad \text{fast sicher für alle } y \in [n, n+1)$$

(ii)

$$\left| \frac{\sum_{i=\kappa_\nu(n)+1}^{\tau_\nu(n)} U_i}{\tau_\nu(n) - \kappa_\nu(n)} \right| \xrightarrow[n\to\infty]{} u^* \quad \text{fast sicher für alle } \nu \in \{1, ..., j_n\} \text{ nach Lemma 2.3.7.}$$

Somit ergibt sich

$$\lim_{y\to\infty} \left(\frac{1}{\tau_1(y) - \kappa_1(y)} \sum_{i=\kappa_1(y)+1}^{\tau_1(y)} U_i \right)$$

$$= \lim_{y\to\infty} \min_{1 \leq \nu \leq j_n} \left(\frac{1}{\tau_\nu(n) - \kappa_\nu(n)} \sum_{i=\kappa_\nu(n)+1}^{\tau_\nu(n)} U_i + \frac{1}{\tau_\nu(n) - \kappa_\nu(n)} \left(\sum_{i=\kappa_1(y)+1}^{\tau_1(y)} U_i - \sum_{i=\kappa_\nu(n)+1}^{\tau_\nu(n)} U_i \right) \right)$$

(da nach Theorem 2.2.1 $\lim\limits_{y\to\infty} \dfrac{\tau_1(y) - \kappa_1(y)}{y} = \lim\limits_{n\to\infty} \dfrac{\tau_\nu(n) - \kappa_\nu(n)}{n}$ gilt)

$$= \lim_{y\to\infty} \min_{1 \leq \nu \leq j_n} \left(\frac{1}{\tau_\nu(n) - \kappa_\nu(n)} \sum_{i=\kappa_\nu(n)+1}^{\tau_\nu(n)} U_i + \frac{1}{\frac{\tau_\nu(n) - \kappa_\nu(n)}{n}} \left(\frac{\sum_{i=\kappa_1(y)+1}^{\tau_1(y)} U_i - \sum_{i=\kappa_\nu(n)+1}^{\tau_\nu(n)} U_i}{n} \right) \right)$$

$= u^*$ fast sicher nach (i) und (ii) und nach Theorem 2.2.1 ($\dfrac{\tau_\nu(n) - \kappa_\nu(n)}{n} \to \dfrac{1}{\omega^*}$ f.s.).

Da unter $\nu = \min\{k \geq 1 : I_k(y) = 1\}$

$$\frac{W_\nu(y)}{L_\nu(y)} = \frac{\sum_{i=\kappa_1(y)+1}^{\tau_1(y)} U_i}{\tau_1(y) - \kappa_1(y)}$$

gilt, erhält man schließlich die gewünschte Aussage:

$$\frac{W_\nu(y)}{L_\nu(y)} \xrightarrow[y\to\infty]{} u^* \quad \text{fast sicher,} \quad \text{falls } \nu = \min\{k \geq 1 : I_k(y) = 1\}.$$

□

2.4 Ein zentraler Grenzwertsatz

In diesem Abschnitt werden nun Aussagen über Länge und Zusammensetzung von Sequenzstücken mit hohem Score in Form von Folgerungen (siehe Korollar 2.4.2-2.4.4) aus einem zentralen Grenzwertsatz (siehe Theorem 2.4.1) getroffen.

Dazu seien die Notationen und Voraussetzungen wie bisher (siehe Abschnitt 2.1), mit der Ausnahme, dass die Zufallsgrößen (U_i) hier Zufallsvektoren sein sollen.

Ferner sei vorausgesetzt, dass die (X_i) entweder alle Gitterverteilungen der Spanne d besitzen (d.h. $d = \sup\{\delta \in [0,\infty] : P^{X_1}(\delta\mathbb{Z}) = 1\}$) oder stetige Zufallsgrößen sind.

Um Verwechselungen zu vermeiden (im Beweis wird später zunächst der eindimensionale Fall gezeigt), seien diese Zufallsvektoren und alle übrigen Vektoren fett gedruckt, d.h. gegeben sei eine Folge von iid und beschränkten Zufallsvektoren $\boldsymbol{U_1}, \boldsymbol{U_2}, ...$, wobei die $\boldsymbol{U_i} = (U_{i1}, ..., U_{im})$ von den X_i abhängig sein können, aber unabhängig von den $X_j,\ i \neq j$, seien.
Dementsprechend sei im Folgenden

$$\zeta(\theta, \boldsymbol{t}) = \log E(e^{\theta X_1 + \langle \boldsymbol{t}, \boldsymbol{U_1} \rangle}),$$

wobei $\langle \boldsymbol{t}, \boldsymbol{U_1} \rangle = \sum_{i=1}^m t_i U_{1i}$ das Skalarprodukt der Vektoren $\boldsymbol{t} = (t_1, ..., t_m)$ und $\boldsymbol{U_1}$ bezeichne.

θ^* sei wie bisher die eindeutig bestimmte positive Lösung von $E(e^{\theta X_1}) = 1$, bzw. von $\zeta(\theta, 0) = 0$ (Existenz und Eindeutigkeit siehe Anmerkung 2.2.2) und

$$\omega^* = \frac{d\zeta}{d\theta}(\theta^*, 0) = \frac{E(X_1 e^{\theta^* X_1})}{E(e^{\theta^* X_1})} = E(X_1 e^{\theta^* X_1}) > 0, \quad \text{sowie}$$

$$\begin{aligned}\boldsymbol{u^*} &= (u_1^*, ..., u_m^*) = \left(\frac{d\zeta}{dt_1}(\theta^*, 0), ..., \frac{d\zeta}{dt_m}(\theta^*, 0)\right) = \left(\frac{E(U_{11} e^{\theta^* X_1})}{E(e^{\theta^* X_1})}, ..., \frac{E(U_{1m} e^{\theta^* X_1})}{E(e^{\theta^* X_1})}\right) \\ &= (E(U_{11} e^{\theta^* X_1}), ..., E(U_{1m} e^{\theta^* X_1})) =: E(\boldsymbol{U_1} e^{\theta^* X_1}).\end{aligned}$$

Aus dem folgenden zentralen Grenzwertsatz ergeben sich in den Anwendungen nützliche Aussagen über hochscorige Sequenzsegmente.

Theorem 2.4.1. *Es sei ν der erste Index, für welchen ein Score y angenommen bzw. überschritten wird, d.h. $\nu = \min\{k \geq 1 : I_k(y) = 1\}$. Dann gilt, dass*

$$\sqrt{\frac{\omega^*}{y}} \sum_{k=K_{\nu-1}+1}^{T_\nu(y)} (\boldsymbol{U_k} - \boldsymbol{u^*})$$

in Verteilung ($y \to \infty$) gegen eine mehrdimensionale Normalverteilung mit Erwartungsvektor $\mathbf{0}$ und Kovarianzmatrix $\Sigma = (\sigma_{ij})_{i,j \in \{1,\ldots,m\}}$ konvergiert, wobei

$$\sigma_{ij} = \frac{d^2\zeta}{dt_i dt_j}(\theta^*, 0) = E((\boldsymbol{U_1} - \boldsymbol{u^*})_i (\boldsymbol{U_1} - \boldsymbol{u^*})_j e^{\theta^* X_1}).$$

Bezeichnung im Folgenden: $\sqrt{\frac{\omega^*}{y}} \sum_{k=K_{\nu-1}+1}^{T_\nu(y)} (\boldsymbol{U_k} - \boldsymbol{u^*}) \underset{(V)}{\longrightarrow} \mathcal{N}(\mathbf{0}, \Sigma)$.

Als erste Anwendung dieses Grenzwertsatzes erhält man für den Fall $\boldsymbol{U_k} = X_k$ eine Aussage über die Länge $L_\nu(y) = T_\nu(y) - K_{\nu-1}$ von Sequenzstücken mit hohem Score.

Korollar 2.4.2. *Sei $\nu = \min\{k \geq 1 : I_k(y) = 1\}$, dann gilt:*

$$\sqrt{\frac{\omega^*}{y}} (y - \omega^* L_\nu(y)) \underset{(V)}{\longrightarrow} \mathcal{N}(0, \sigma^2) \quad \textit{falls } y \to \infty,$$

wobei $\sigma^2 = E((X_1 - \omega^*)^2 e^{\theta^* X_1}) = E(X_1^2 e^{\theta^* X_1}) - E(X_1 e^{\theta^* X_1})^2$.

Beweis. Sei $\boldsymbol{U_k} = X_k$. Dann gilt nach Theorem 2.4.1 falls $y \to \infty$:

$$\sqrt{\frac{\omega^*}{y}} (S_{L_\nu(y)} - \omega^* L_\nu(y)) \underset{(V)}{\longrightarrow} \mathcal{N}(0, \sigma^2)$$

mit $\sigma^2 = E((X_1 - \omega^*)^2 e^{\theta^* X_1})$. Ferner gilt unter $I_\nu(y) = 1$ auch $0 \leq S_{L_\nu(y)} - y \leq K$, denn nach Voraussetzung gilt $|X_i| \leq K$ und somit $y \leq S_{L_\nu(y)} = S_{L_\nu(y)-1} + X_{L_\nu(y)} \leq y + K$. Damit erhält man für $y \to \infty$

$$\sqrt{\frac{\omega^*}{y}} (S_{L_\nu(y)} - (S_{L_\nu(y)} - y) - \omega^* L_\nu(y)) = \sqrt{\frac{\omega^*}{y}} (y - \omega^* L_\nu(y)) \underset{(V)}{\longrightarrow} \mathcal{N}(0, \sigma^2).$$

□

Im Folgenden interessieren wir uns wieder für die empirische Häufigkeitsverteilung von Buchstaben in einem Sequenzenpaar hoher Ähnlichkeit (also mit hohem Score) bzw. allgemeiner für die Häufigkeit des Eintretens eines Ereignisses A. Dazu sei

$$\boldsymbol{U_k} = \mathbb{I}_A(X_k) = \begin{cases} 1 & \text{falls } X_k \in A \\ 0 & \text{falls } X_k \notin A \end{cases}, \quad A \in \mathbb{B}.$$

Korollar 2.4.3. *Es sei A eine Borelmenge aus dem Wertebereich von X_1 und $\nu = \min\{k \geq 1 : I_k(y) = 1\}$. Dann gilt für die empirische Häufigkeitsverteilung $\mu_\nu(A; y) = \frac{\sum_{k=K_{\nu-1}+1}^{T_\nu(y)} \mathbb{I}_A(X_k)}{L_\nu(y)}$ des Eintretens von A:*

$$\sqrt{L_\nu(y)}(\mu_\nu(A; y) - \mu^*(A)) \underset{(V)}{\longrightarrow} \mathcal{N}(0, c^*) \quad \textit{für } y \to \infty,$$

wobei $\mu^*(A) = E(\mathbb{I}_A(X_1)e^{\theta^* X_1})$ und $c^* = \mu^*(A) - \mu^*(A)^2$.

Beweis. Sei $\boldsymbol{U_k} = \mathbb{I}_A(X_k)$, dann gilt $\mu_\nu(A; y) = \frac{\sum_{k=K_{\nu-1}+1}^{T_\nu(y)} \boldsymbol{U_k}}{L_\nu(y)}$ und $\mu^*(A) = \boldsymbol{u^*}$. Damit ergibt sich

$$\sqrt{L_\nu(y)}(\mu_\nu(A; y) - \mu^*(A)) = \sqrt{\frac{\omega^*}{y}} \sum_{k=K_{\nu-1}+1}^{T_\nu(y)} (\boldsymbol{U_k} - \boldsymbol{u^*}) \sqrt{\frac{y}{\omega^*}} \frac{1}{\sqrt{L_\nu(y)}}.$$

Nun gilt aber nach Theorem 2.2.1, dass $\frac{L_\nu(y)}{y} \underset{y\to\infty}{\longrightarrow} \frac{1}{\omega^*}$ P-fast sicher unter $I_\nu(y) = 1$ und somit

$$\sqrt{\frac{y}{\omega^*}} \frac{1}{\sqrt{L_\nu(y)}} \underset{y\to\infty}{\longrightarrow} 1 \quad P\text{-fast sicher.}$$

Dementsprechend gilt nach Theorem 2.4.1: $\sqrt{L_\nu(y)}(\mu_\nu(A; y) - \mu^*(A)) \underset{(V)}{\longrightarrow} \mathcal{N}(0, c^*)$ mit

$$\begin{aligned} c^* &= E((\mathbb{I}_A(X_1) - \mu^*(A))^2 e^{\theta^* X_1}) \\ &= E(\mathbb{I}_A^2(X_1)e^{\theta^* X_1}) - 2\mu^*(A)E(\mathbb{I}_A(X_1)e^{\theta^* X_1}) + \mu^*(A)^2 E(e^{\theta^* X_1}) \\ &= \mu^*(A) - \mu^*(A)^2, \quad \text{da } E(e^{\theta^* X_1}) = 1 \text{ und } \mathbb{I}_A^2 = \mathbb{I}_A. \end{aligned}$$

□

Genauso ergibt sich, falls X_1 diskret verteilt ist, für $A = \{s_i\}$, also die relative Häufigkeit $\mu_\nu(y)$ des Auftretens eines Buchstabens a_i (z.B. eines Aminosäurenpaares) in einem Segment mit hohem Score folgende Aussage:

$$\sqrt{L_\nu(y)}(\mu_\nu(y) - \mu^*) \underset{(V)}{\longrightarrow} \mathcal{N}(0, c^*) \quad \text{für } y \to \infty$$

mit $\mu^* = p_i e^{\theta^* s_i}$ und $c^* = p_i e^{\theta^* s_i} - (p_i e^{\theta^* s_i})^2$ (siehe auch Korollar 2.2.4).

Sei nun wieder auf die alternativen Bezeichnungen $\tau_\nu(y)$ und $\kappa_\nu(y)$ zurückgegriffen (siehe (2.4)), für welche gilt:

$$\tau_1(y) - \kappa_1(y) = L_\nu(y) = T_\nu(y) - K_{\nu-1}, \quad \text{falls} \quad \nu = \min\{k \geq 1 : I_k(y) = 1\}.$$

Korollar 2.4.4. *Für $y \to \infty$ gilt:*

$$\frac{1}{\sqrt{L_\nu(y)}} \sum_{k=\kappa_1(y)+1}^{\tau_1(y)} (\boldsymbol{U_k} - \boldsymbol{u^*}) \underset{(V)}{\longrightarrow} \mathcal{N}(\boldsymbol{0}, \Sigma),$$

wobei $\Sigma = (\sigma_{ij}),\ \sigma_{ij} = E((\boldsymbol{U_1} - \boldsymbol{u^*})_i (\boldsymbol{U_1} - \boldsymbol{u^*})_j e^{\theta^* X_1})$.

Beweis. Es gilt

$$\frac{1}{\sqrt{L_\nu(y)}} \sum_{k=\kappa_1(y)+1}^{\tau_1(y)} (\boldsymbol{U_k} - \boldsymbol{u^*}) = \sqrt{\frac{\omega^*}{y}} \sum_{k=K_{\nu-1}+1}^{T_\nu(y)} (\boldsymbol{U_k} - \boldsymbol{u^*}) \sqrt{\frac{y}{\omega^*} \frac{1}{L_\nu(y)}},$$

somit folgt die gewünschte Aussage mit Theorem 2.2.1 und Theorem 2.4.1. □

2.5 Beweis des zentralen Grenzwertsatzes

Der zentrale Grenzwertsatz (siehe Theorem 2.4.1) aus dem vorhergehenden Abschnitt wird zunächst für den eindimensionalen Fall mit $u^* = 0$ gezeigt, d.h. es wird gezeigt, dass für $\nu = \min\{k \geq 1 : I_k(y) = 1\}$ gilt:

$$\sqrt{\frac{\omega^*}{y}} \sum_{k=K_{\nu-1}+1}^{T_\nu(y)} U_k \underset{(V)}{\longrightarrow} \mathcal{N}(0, v^*) \quad \text{mit } v^* = E(U_1^2 e^{\theta^* X_1}). \tag{2.14}$$

Anschließend wendet man diese Aussage auf die Folge von Zufallsgrößen (U_i) mit $U_i = \langle \boldsymbol{z}, \boldsymbol{U_i} - \boldsymbol{u^*} \rangle$ an, für welche gerade $u^* = 0$ gilt. Mit Hilfe des Satzes von Cramér-Wold (siehe Proposition 2.5.9) erhält man damit die Aussage des zentralen Grenzwertsatzes auch für den mehrdimensionalen Fall.
Um die Aussage (2.14) zu erhalten, genügt es nach dem Stetigkeitssatz für Laplace-Transformierte (siehe Proposition 2.5.8) zu zeigen, dass die Laplace-Transformierte von $\sqrt{\frac{\omega^*}{y}} \sum_{k=K_{\nu-1}+1}^{T_\nu(y)} U_k$ gegen die Laplace-Transformierte einer $\mathcal{N}(0, v^*)$-Verteilung konvergiert (vergleiche Lemma 2.5.7) bzw. dass für alle $t \in \mathbb{R}$ gilt:

$$\lim_{y \to \infty} E(e^{t\sqrt{\frac{\omega^*}{y}} \sum_{k=K_{\nu-1}+1}^{T_\nu(y)} U_k} | I_\nu(y) = 1) = e^{\frac{1}{2}t^2 v^*}. \tag{2.15}$$

Dazu werden insgesamt sechs Lemmata benötigt. Wie auch schon im Beweis der starken Grenzwertsätze werden dabei Eigenschaften des Waldmartingals (siehe Lemma 2.3.2) eingehen.
Da die (X_i) und (U_i) iid sind, sind es ebenfalls die Zufallsgrößen $(I_\nu(y))$ und $(L_\nu(y))$, d.h. es genügt, sämtliche Aussagen für die Zufallsgrößen $I_1(y) := I(y)$ und $L_1(y) := L(y)$ (manchmal auch als L abgekürzt) und damit für $K_{\nu-1} = 0$ und $T_\nu(y) = L(y)$ zu zeigen.

Um zu der Aussage (2.15) zu gelangen, wird zunächst folgendes Lemma benötigt:

Lemma 2.5.1. *Es existiert ein $\theta(t)$ nahe bei θ^* mit $\zeta(\theta(t),t) = 0$ und unter der Voraussetzung $u^* = \frac{d\zeta}{dt}(\theta^*,0) = E(U_1 e^{\theta^* X_1}) = 0$ gilt für dieses $\theta(t)$:*

$$\theta(t) = \theta^* - \frac{1}{2}\frac{v^*}{\omega^*}t^2 + O(t^3),$$

wobei $v^* = \frac{d^2\zeta}{dt^2}(\theta^*,0) = E(U_1^2 e^{\theta^* X_1})$ und $\omega^* = \frac{d\zeta}{d\theta}(\theta^*,0) = E(X_1 e^{\theta^* X_1})$.

Beweis. Da die Funktion $\zeta(\theta,t) = \log E(e^{\theta X_1 + tU_1})$ für genügend kleines t analytisch in t und da $\zeta(\theta^*,0) = \log E(e^{\theta^* X_1}) = 0$, sowie $\frac{d\zeta}{d\theta}(\theta^*,0) = \omega^* > 0$ (also $\frac{d\zeta}{d\theta}(\theta^*,0)$ invertierbar) ist, existiert nach dem Satz über implizite Funktionen ein $\theta(t)$ nahe bei θ^*, welches analytisch in t ist, derart, dass $\zeta(\theta(t),t) = 0$ gilt.
Nimmt man nun eine Taylorentwicklung um den Punkt $(\theta^*,0)$ vor, so erhält man unter der Voraussetzung $u^* = 0$ folgende Darstellung von $\zeta(\theta(t),t)$:

$$\begin{aligned}
0 &= \zeta(\theta(t),t)\\
&= \zeta\left((\theta(t)-\theta^*,t) + (\theta^*,0)\right)\\
&= \zeta(\theta^*,0) + \left(\frac{d\zeta}{d\theta}(\theta^*,0), \frac{d\zeta}{dt}(\theta^*,0)\right)(\theta(t)-\theta^*,t)\\
&\quad + \frac{1}{2}\left(\frac{d^2\zeta}{dt^2}(\theta^*,0)t^2 + 2\frac{d^2\zeta}{dtd\theta}(\theta^*,0)t(\theta(t)-\theta^*) + \frac{d^2\zeta}{d\theta^2}(\theta^*,0)(\theta(t)-\theta^*)^2\right)\\
&\quad + f(\theta(t)-\theta^*,t)\,\|(\theta(t)-\theta^*,t)\|^2 \quad \text{mit } f(\theta(t)-\theta^*,t) \underset{(\theta(t)-\theta^*,t)\to 0}{\longrightarrow} 0\\
&= 0 + (\omega^*,u^*)(\theta(t)-\theta^*,t) + \frac{1}{2}v^*t^2 + O((\theta(t)-\theta^*)^2 + t|\theta(t)-\theta^*| + t^3)\\
&= \omega^*(\theta(t)-\theta^*) + \frac{1}{2}v^*t^2 + O((\theta(t)-\theta^*)^2 + t|\theta(t)-\theta^*| + t^3).
\end{aligned}$$

Daraus folgt aber auch, dass

$$|\omega^*(\theta(t)-\theta^*)| = \left|\frac{1}{2}v^*t^2 + O((\theta(t)-\theta^*)^2 + t|\theta(t)-\theta^*| + t^3)\right|,$$

und somit, da $(\theta(t)-\theta^*)^2 + t|\theta(t)-\theta^*| + t^3 = O(1)$ für $(\theta(t)-\theta^*,t) \longrightarrow 0$, dass $\theta(t)-\theta^* = O(t^2)$ gelten muss. Wir erhalten also für t nahe bei 0:

$$\begin{aligned}
0 &= \omega^*(\theta(t)-\theta^*) + \frac{1}{2}v^*t^2 + O(t^3+t^4) \quad \text{bzw.}\\
\theta(t) &= \theta^* - \frac{1}{2}\frac{v^*}{\omega^*}t^2 + O(t^3)
\end{aligned}$$

Berechnung von v^* :

$$v^* = \frac{d^2\zeta}{dt^2}(\theta^*, 0) = \frac{E(U_1^2 e^{\theta^* X_1})E(e^{\theta^* X_1}) - (E(U_1 e^{\theta^* X_1}))^2}{(E(e^{\theta^* X_1}))^2} = E(U_1^2 e^{\theta^* X_1}),$$

da $E(e^{\theta^* X_1}) = 1$, sowie $E(U_1 e^{\theta^* X_1}) = u^* = 0$. □

Im Folgenden sei $\theta(t)$ wie in Lemma 2.5.1 gewählt. Man betrachtet nun $\theta(t_y)$ mit $t_y := t\sqrt{\frac{\omega^*}{y}}$.
Mit Hilfe von Lemma 2.5.1 lässt sich nun die folgende Aussage beweisen.

Lemma 2.5.2. *Unter der Voraussetzung* $u^* = E(U_1 e^{\theta^* X_1}) = 0$ *gilt:*

$$\lim_{y\to\infty} E(e^{\theta(t_y)S_L + t_y W_L} - e^{\theta^* S_L} | I(y) = 1)P(I(y) = 1) = 0.$$

Beweis. Nach Lemma 2.3.2 („Eigenschaften des Waldmartingals") gilt für genügend kleines t und θ nahe bei θ^*:

$$E(e^{\theta S_L + tW_L - L\zeta(\theta(t),t)}) = 1.$$

Insbesondere gilt für $t = 0$ und $\theta = \theta^*$:

$$E(e^{\theta^* S_L}) = 1.$$

Ferner folgt mit Lemma 2.5.1, dass ein $\theta(t)$ in der Nähe von θ^* existiert mit $\zeta(\theta(t), t) = 0$. Dieses $\theta(t)$ wählen wir nun, wobei wir $t_y = t\sqrt{\frac{\omega^*}{y}}$ einsetzen. Wenn wir y genügend groß wählen, können wir Lemma 2.3.2 anwenden (denn dann liegt $(\theta(t_y), t_y)$ nahe bei $(\theta^*, 0)$). Wir erhalten also, da $\zeta(\theta(t_y), t_y) = 0$:

$$E(e^{\theta^* S_L}) = 1 = E(e^{\theta(t_y)S_L + t_y W_L})$$

und somit

$$\begin{aligned} 0 &= E(e^{\theta(t_y)S_L + t_y W_L} - e^{\theta^* S_L}) \\ &= E(e^{\theta(t_y)S_L + t_y W_L} - e^{\theta^* S_L} | I(y) = 1)P(I(y) = 1) \\ &\quad + E(e^{\theta(t_y)S_L + t_y W_L} - e^{\theta^* S_L} | I(y) = 0)P(I(y) = 0). \end{aligned}$$

Es genügt also zu zeigen, dass

$$\lim_{y\to\infty} E(e^{\theta(t_y)S_L + t_y W_L} - e^{\theta^* S_L} | I(y) = 0)P(I(y) = 0) = 0.$$

Dieser Ausdruck lässt sich auch folgendermaßen darstellen

$$\begin{aligned} &E(e^{\theta(t_y)S_L + t_y W_L} - e^{\theta^* S_L} | I(y) = 0)P(I(y) = 0) \\ &= E(e^{\theta(t_y)S_L} - e^{\theta^* S_L} | I(y) = 0)P(I(y) = 0) \\ &\quad + E(e^{\theta(t_y)S_L + t_y W_L}(1 - e^{-t_y W_L}) | I(y) = 0)P(I(y) = 0). \end{aligned}$$

Unter $I(y) = 0$ gilt $-LK \leq S_L \leq 0$ und somit existiert ein C derart, dass

$$e^{\theta(t_y)S_L} - e^{\theta^* S_L} \leq C|\theta(t_y) - \theta^*|.$$

Da $C\left|\theta(t\sqrt{\frac{\omega^*}{y}}) - \theta^*\right| \underset{y\to\infty}{\longrightarrow} 0$ (denn $\theta(0) = \theta^*$), gilt also

$$E(e^{\theta(t_y)S_L} - e^{\theta^* S_L}|I(y) = 0)P(I(y) = 0) \underset{y\to\infty}{\longrightarrow} 0.$$

Zu zeigen bleibt, dass auch der zweite Term aus obiger Gleichung gegen 0 konvergiert. Dazu schätzen wir diesen folgendermaßen ab:

$$\begin{aligned}
&E(e^{\theta(t_y)S_L + t_y W_L}(1 - e^{-t_y W_L})|I(y) = 0)P(I(y) = 0)\\
\leq\ & E(e^{t_y W_L}(1 - e^{-t_y W_L})|I(y) = 0)P(I(y) = 0) \quad (\text{da } e^{\theta(t_y)S_L} \leq 1)\\
=\ & \sum_{n=0}^{\infty} E(e^{t_y W_L}(1 - e^{-t_y W_L})|I(y) = 0, L = n+1)P(L = n+1|I(y) = 0)P(I(y) = 0)\\
\leq\ & \sum_{n=0}^{\infty} e^{|t|\sqrt{\frac{\omega^*}{y}}K'(n+1)} P(L = n+1|I(y) = 0)\\
=\ & C_{N,t}\frac{1}{\sqrt{y}} + \sum_{n=N}^{\infty} e^{|t|\sqrt{\frac{\omega^*}{y}}K'(n+1)} P(L = n+1|I(y) = 0) \qquad (2.16)
\end{aligned}$$

für alle festen N und für y genügend groß, wobei $C_{N,t}$ eine passende Konstante ist, die von N und t abhängt.
Nun folgt mit der Markov-Ungleichung, dass

$$P(L = n+1|I(y) = 0) \leq P(S_n > 0) \leq \frac{E(e^{\theta S_n})}{e^{\theta 0}} = (E(e^{\theta X_1}))^n = e^{-bn} \qquad (2.17)$$

für ein geeignetes $b > 0$ (nach Anmerkung 2.2.2 für $0 < \theta < \theta^*$). Durch geeignete Wahl von N lässt sich also erreichen, dass der Reihenterm in (2.16) beliebig klein wird und somit, dass

$$E(e^{\theta(t_y)S_L + t_y W_L}(1 - e^{-t_y W_L})|I(y) = 0)P(I(y) = 0) \underset{y\to\infty}{\longrightarrow} 0.$$

Damit folgt die gewünschte Aussage. □

Ein nächster Schritt in die Richtung, die Konvergenz

$$\lim_{y\to\infty} E(e^{t\sqrt{\frac{\omega^*}{y}}\sum_{k=K_{\nu-1}+1}^{T_\nu(y)} U_k}|I_\nu(y) = 1) = e^{\frac{1}{2}t^2 v^*}$$

zu etablieren, ist der Beweis des folgenden Lemmas.

Lemma 2.5.3. *Sei $u^* = E(U_1 e^{\theta^* X_1}) = 0$. Dann gilt für alle $t \in \mathbb{R}$:*

$$\lim_{y\to\infty} \frac{E(e^{t\sqrt{\frac{\omega^*}{y}}W_L} e^{\theta^*(S_L - y)}|I(y) = 1)}{E(e^{\theta^*(S_L - y)}|I(y) = 1)} = e^{\frac{1}{2}t^2 v^*}, \quad \textit{wobei} \quad v^* = \frac{d^2\zeta}{dt^2}(\theta^*, 0) = E(U_1^2 e^{\theta^* X_1}).$$

Beweis. Nach Lemma 2.5.2 gilt:

$$\left(E(e^{(\theta(t_y)-\theta^*)S_L}e^{\theta^*(S_L-y)}e^{t\sqrt{\frac{\omega^*}{y}}W_L}|I(y)=1) - E(e^{\theta^*(S_L-y)}|I(y)=1)\right)e^{\theta^* y}P(I(y)=1)$$
$$= E(e^{\theta(t_y)S_L+t\sqrt{\frac{\omega^*}{y}}W_L} - e^{\theta^* S_L}|I(y)=1)P(I(y)=1) \underset{y\to\infty}{\longrightarrow} 0.$$

Per Division durch $E(e^{\theta^*(S_L-y)}|I(y)=1) > 0$ ergibt sich also

$$\left(\frac{E(e^{(\theta(t_y)-\theta^*)S_L}e^{\theta^*(S_L-y)}e^{t\sqrt{\frac{\omega^*}{y}}W_L}|I(y)=1)}{E(e^{\theta^*(S_L-y)}|I(y)=1)} - 1\right)e^{\theta^* y}P(I(y)=1) \underset{y\to\infty}{\longrightarrow} 0.$$

Ferner gilt nach Lemma 2.3.3, dass $e^{\theta^* y}P(I(y)=1) \in [\delta,1]$ für ein $\delta > 0$, somit erhält man

$$\frac{E(e^{(\theta(t_y)-\theta^*)S_L}e^{\theta^*(S_L-y)}e^{t\sqrt{\frac{\omega^*}{y}}W_L}|I(y)=1)}{E(e^{\theta^*(S_L-y)}|I(y)=1)} \underset{y\to\infty}{\longrightarrow} 1.$$

Nun besitzt $\theta(t_y)$ nach Lemma 2.5.1 folgende Darstellung:

$$\theta(t_y) = \theta^* - \frac{v^*}{2\omega^*}\left(t\sqrt{\frac{\omega^*}{y}}\right)^2 + O\left(t^3\left(\frac{\omega^*}{y}\right)^{3/2}\right) = \theta^* - \frac{t^2}{2y}v^* + O\left(\frac{1}{y^{3/2}}\right),$$

setzen wir also

$$\theta(t_y) - \theta^* = -\frac{t^2}{2y}v^* + O\left(\frac{1}{y^{3/2}}\right)$$

in obige Konvergenzaussage ein, so erhalten wir

$$\frac{E(e^{(-\frac{t^2}{2y}v^*+O(y^{-3/2}))S_L}e^{\theta^*(S_L-y)}e^{t\sqrt{\frac{\omega^*}{y}}W_L}|I(y)=1)}{E(e^{\theta^*(S_L-y)}|I(y)=1)} \underset{y\to\infty}{\longrightarrow} 1.$$

Nun gilt aber, dass

$$e^{(-\frac{t^2}{2y}v^*+O(y^{-3/2}))S_L} \underset{y\to\infty}{\longrightarrow} e^{-\frac{1}{2}t^2 v^*},$$

denn aufgrund der Beschränktheit ($y \le S_L \le y+K$) unter $I(y)=1$ gilt $\left|\frac{S_L}{y}\right| \to 1$. Damit folgt die gewünschte Aussage:

$$\frac{E(e^{t\sqrt{\frac{\omega^*}{y}}W_L}e^{\theta^*(S_L-y)}|I(y)=1)}{E(e^{\theta^*(S_L-y)}|I(y)=1)} \underset{y\to\infty}{\longrightarrow} e^{\frac{1}{2}t^2 v^*}.$$

□

Um nun zu zeigen, dass

$$E(e^{t\sqrt{\frac{\omega^*}{y}}W_L}|I(y)=1) \underset{y\to\infty}{\longrightarrow} e^{\frac{1}{2}t^2v^*},$$

brauchen wir nach Lemma 2.5.3 also nur noch die asymptotische Unabhängigkeit von W_L und $S_L - y$ unter $I(y)=1$ zu zeigen. Dazu werden zwei weitere Lemmata benötigt. Außerdem wird uns folgende Proposition hilfreich sein, welche z.B. in einem Artikel von Iglehart („Extreme values in the GI/G/1 queue“; siehe [12]) bewiesen wird.

Proposition 2.5.4. *(siehe [12], Lemma 1 und Theorem 1)*
Es sei X_1 eine Zufallsgröße und es existiere eine Zahl $\gamma \neq 0$ mit $E(e^{\gamma X_1}) = 1$, $E(X_1 e^{\gamma X_1}) = \mu_\gamma < \infty$ und es gelte $E(X_1) < 0$.
Falls X_1 keine Gitterverteilung besitzt, so gilt

(i) $\lim\limits_{y\to\infty} P(\max\limits_{k\geq 0} S_k > y)e^{\gamma y} = e^*$ *für ein e^* mit $0 < e^* < \infty$*

(ii) $\lim\limits_{y\to\infty} P(\max\limits_{0\leq k\leq\sigma} S_k > y)e^{\gamma y} = (1 - E(e^{\gamma S_\sigma}))e^*$ *für ein e^* mit $0 < e^* < \infty$.*

Falls X_1 eine Gitterverteilung der Spanne d besitzt, d.h.
$d = \sup\{\delta \in [0,\infty] : P^{X_1}(\delta\mathbb{Z}) = 1\}$*, so gilt*

(i) $\lim\limits_{n\to\infty} P(\max\limits_{k\geq 0} S_k > nd)e^{\gamma nd} = \tilde{e}^*$ *für ein $\tilde{e}^*$ mit $0 < \tilde{e}^* < \infty$*

(ii) $\lim\limits_{n\to\infty} P(\max\limits_{0\leq k\leq\sigma} S_k > nd)e^{\gamma nd} = (1 - E(e^{\gamma S_\sigma}))\tilde{e}^*$ *für ein $\tilde{e}^*$ mit $0 < \tilde{e}^* < \infty$.*

Für den Beweis des zentralen Grenzwertsatzes wurde bereits vorausgesetzt (siehe Abschnitt 2.4), dass die (X_i) entweder alle Gitterverteilungen der Spanne d besitzen oder stetige Zufallsgrößen sind, so dass wir diese Proposition im Folgenden anwenden können. Es sei $-a \leq 0 < b$. Wir benötigen die folgende Notation:

$$L(-a,b) := \inf\{k \geq 1 : S_k \notin (-a,b)\}$$
$$I(-a,b) := \begin{cases} 1 & \text{falls } S_{L(-a,b)} \geq b \\ 0 & \text{falls } S_{L(-a,b)} \leq -a \end{cases}$$

Lemma 2.5.5. *Es sei Z^- die erste nichtpositive Partialsumme von (S_m) und σ der erste Index, bei dem der Prozess die positive Achse verlässt, d.h. $\sigma = \inf\{k \geq 1 : S_k \leq 0\}$ und $Z^- = S_\sigma$. Ferner sei $F(y)$ die Verteilungsfunktion von $\max\limits_{0\leq k\leq\sigma} S_k$, sowie $M(y)$ die Verteilungsfunktion von $\max\limits_{k\geq 0} S_k$. Dann gilt:*

(i) $\lim\limits_{y\to\infty} E(e^{\theta^*(S_L - y)}|I(y)=1) = \frac{1-E(e^{\theta^* Z^-})}{\lim\limits_{y\to\infty} e^{\theta^* y}(1-F(y))} = \frac{1}{e^*}$ *für ein e^*, $0 < e^* < \infty$*

(ii) $\lim\limits_{\substack{y\to\infty \\ a(y)\to\infty}} E(e^{\theta^*(S_{L(-a(y),y)}-y)}|I(-a(y),y)=1) = \frac{1}{\lim\limits_{y\to\infty} e^{\theta^* y}(1-M(y))} = \frac{1}{e^*}$ *für ein* e^*, $0 < e^* < \infty$, *für alle* $a(y)$.

Falls X_1 *eine Gitterverteilung besitzt, so gelten diese Aussagen ebenso. Dabei sind* y *und* $a(y)$ *Punkte des Gitterfeldes.*

Beweis. Die (X_i) seien zunächst stetige Zufallsgrößen. Die Voraussetzungen für Proposition 2.5.4 sind mit $\gamma = \theta^*$ und $\mu_\gamma = \omega^*$ erfüllt. Es gilt also, da $Z^- = S_\sigma$, dass

$$\begin{aligned} &\lim_{y\to\infty}(1-M(y))e^{\theta^* y} = e^* \quad \text{sowie} \\ &\lim_{y\to\infty}(1-F(y))e^{\theta^* y} = (1-E(e^{\theta^* Z^-}))e^* \quad \text{für ein } e^*,\ 0 < e^* < \infty. \end{aligned} \tag{2.18}$$

zu (i): Nach Anwendung des Optional-Sampling-Theorems (vergleiche Lemma 2.3.2: „Eigenschaften des Waldmartingals") erhält man

$$\begin{aligned} 1 &= E(e^{\theta^* S_L}) \\ &= P(I(y)=1)e^{\theta^* y}E(e^{\theta^*(S_L-y)}|I(y)=1) + P(I(y)=0)E(e^{\theta^* S_L}|I(y)=0) \\ &= (1-F(y))e^{\theta^* y}E(e^{\theta^*(S_L-y)}|I(y)=1) + F(y)E(e^{\theta^* S_L}|I(y)=0), \end{aligned}$$

denn es gilt (da S_σ die erste nichtpositive Partialsumme ist):

$$P(I(y)=0) = P(\max_{0\le k\le\sigma} S_k < y) = F(y).$$

Stellt man diesen Ausdruck um, so ergibt sich

$$\begin{aligned} \lim_{y\to\infty} E(e^{\theta^*(S_L-y)}|I(y)=1) &= \lim_{y\to\infty} \frac{1-F(y)E(e^{\theta^* S_L}|I(y)=0)}{e^{\theta^* y}(1-F(y))} \\ &= \frac{1-E(e^{\theta^* Z^-})}{\lim\limits_{y\to\infty} e^{\theta^* y}(1-F(y))} \\ &\quad (\text{da } \lim_{y\to\infty} P^{L(y)|I(y)=0} = P^\sigma,\ \sigma = \inf\{k\ge 1 : S_k \le 0\}) \\ &= \frac{1-E(e^{\theta^* Z^-})}{e^*(1-E(e^{\theta^* Z^-}))} = \frac{1}{e^*} \quad (\text{nach } 2.18). \end{aligned}$$

zu (ii): Die zweite Aussage lässt sich völlig analog beweisen. Nach Anwendung des Optional-Sampling-Theorems ergibt sich

$$\begin{aligned} 1 &= E(e^{\theta^* S_{L(-a(y),y)}}) \\ &= P(I(-a(y),y)=1)e^{\theta^* y}E(e^{\theta^*(S_{L(-a(y),y)}-y)}|I(-a(y),y)=1) \\ &\quad + P(I(-a(y),y)=0)E(e^{\theta^* S_{L(-a(y),y)}}|I(-a(y),y)=0) \end{aligned}$$

Nun gilt:

$$\lim_{a(y)\to\infty} P(I(-a(y),y)=1) = P(\max_{k\geq 0} S_k \geq y) = 1-M(y),$$

denn falls $I(-a(y),y)=1$, so gilt $S_{L(-a(y),y)} \geq y$ und somit $\max\limits_{k\geq 0} S_k \geq y$. Andersherum folgt aus $\max\limits_{k\geq 0} S_k \geq y$ auch $\lim\limits_{a(y)\to\infty} I(-a(y),y)=1$.
Somit erhält man

$$\begin{aligned}
&\lim_{\substack{y\to\infty\\ a(y)\to\infty}} E(e^{\theta^*(S_{L(-a(y),y)}-y)}|I(-a(y),y)=1)\\
=\;&\lim_{\substack{y\to\infty\\ a(y)\to\infty}} \frac{1-M(y)E(e^{\theta^* S_{L(-a(y),y)}}|I(-a(y),y)=0)}{e^{\theta^* y}(1-M(y))}\\
=\;&\frac{1}{\lim\limits_{y\to\infty} e^{\theta^* y}(1-M(y))} \quad (\text{da } S_{L(-a(y),y)} \leq -a(y) \text{ unter } I(-a(y),y)=0)\\
=\;&\frac{1}{e^*} \quad (\text{nach 2.18}).
\end{aligned}$$

Falls die X_i nun gitterverteilt sind mit der Spanne d, lässt sich Proposition 2.5.4 analog anwenden mit $y=nd$. Die Aussagen des Lemmas folgen also ebenso für den Gittertyp, dabei sind y und $a(y)$ Gitterpunkte. □

Der Übersichtlichkeit halber werden im weiteren Verlauf folgende Bezeichnungen verwendet:

$$\begin{aligned}
\Gamma(y) &:= E(e^{t\sqrt{\frac{\omega^*}{y}}W_{L(y)}} e^{\theta^*(S_{L(y)}-y)}|I(y)=1)\\
\hat{\Gamma}(y) &:= E(e^{t\sqrt{\frac{\omega^*}{y}}W_{L(a(y))}} e^{\theta^*(S_{L(y)}-y)}|I(y)=1).
\end{aligned}$$

Mit diesen Bezeichnungen stehen bereits folgende Aussagen zur Verfügung:

(i) $\lim\limits_{y\to\infty} \frac{\Gamma(y)}{E(e^{\theta^*(S_L-y)}|I(y)=1)} = e^{\frac{1}{2}t^2 v^*}$ (siehe Lemma 2.5.3)

(ii) $\lim\limits_{y\to\infty} E(e^{\theta^*(S_L-y)}|I(y)=1) = \frac{1}{e^*}$ (siehe Lemma 2.5.5 (ii)).

Um also zu der gewünschten Aussage

$$\lim_{y\to\infty} E(e^{t\sqrt{\frac{\omega^*}{y}}\sum_{k=K_{\nu-1}+1}^{T_\nu(y)} U_k}|I_\nu(y)=1) = e^{\frac{1}{2}t^2 v^*}$$

zu gelangen, müssen wir noch zeigen, dass

$$\lim_{y\to\infty} \Gamma(y) = \frac{1}{e^*}\lim_{y\to\infty} E(e^{t\sqrt{\frac{\omega^*}{y}}W_{L(y)}}|I_\nu(y)=1)$$

gilt. Dazu werden wir zunächst beweisen, dass die Grenzwerte von $\Gamma(y)$ und $\hat{\Gamma}(y)$ übereinstimmen, und anschließend, dass

$$\lim_{y\to\infty} \hat{\Gamma}(y) = \frac{1}{e^*}\lim_{y\to\infty} E(e^{t\sqrt{\frac{\omega^*}{y}}W_{L(a(y))}}|I_\nu(y)=1) \quad \text{gilt.}$$

Im Folgenden sei $a(y) = y - \log y$ und y hinreichend groß gewählt, so dass $\log y > K$ gilt (K derart, dass $|X_i| \leq K$).

Lemma 2.5.6. *Es gilt:* $\lim\limits_{y\to\infty} \Gamma(y) = \lim\limits_{y\to\infty} \hat{\Gamma}(y)$.

Beweis. Man zeigt zunächst, dass

$$|\Gamma(y) - \hat{\Gamma}(y)| \leq e^{\theta^* K} D(y) \hat{\Gamma}(y)$$

für ein geeignetes $D(y)$, welches später genauer angegeben wird, und danach, dass dieses $D(y)$ für $y \to \infty$ gegen 0 konvergiert. Die Aussage des Lemmas folgt dann aufgrund der Existenz und Endlichkeit von $\lim\limits_{y\to\infty} \hat{\Gamma}(y)$.

Nach Voraussetzung sind die X_i und U_i beschränkt, d.h. $|X_i| \leq K$, $|U_i| \leq K'$, also $|W_L| \leq K'L$. Ferner gilt unter $I(y) = 1$, dass $S_L - y \in [0, K]$ und somit:

$$\begin{aligned} |\Gamma(y) - \hat{\Gamma}(y)| &= |E(e^{t\sqrt{\frac{\omega^*}{y}} W_{L(a(y))}} (e^{t\sqrt{\frac{\omega^*}{y}}(W_{L(y)} - W_{L(a(y))})} - 1) e^{\theta^*(S_{L(y)} - y)} | I(y) = 1)| \\ &\leq e^{\theta^* K} E(e^{t\sqrt{\frac{\omega^*}{y}} W_{L(a(y))}} (e^{|t|\sqrt{\frac{\omega^*}{y}} K'(L(y) - L(a(y)))} - 1) | I(y) = 1). \end{aligned} \tag{2.19}$$

Man betrachtet nun die σ-Algebra

$$\mathcal{S} := \sigma(S_1, ..., S_{L(a(y))}, W_1, ..., W_{L(a(y))}, L(a(y)), I(y) = 1).$$

Ferner bezeichne G_y die Verteilungsfunktion von $S_{L(a(y))}$ unter $I(y) = 1$, d.h. $G_y(\xi) = P(S_{L(a(y))} \leq \xi | I(y) = 1)$. Nach Voraussetzung gilt $a(y) = y - \log y < y$, da $\log y > K$, somit impliziert $I(y) = 1$ auch $I(a(y)) = 1$, also liegt $S_{L(a(y))}$ unter $I(y) = 1$ im Intervall $[a(y), a(y) + K]$. Da ferner $(e^{|t|\sqrt{\frac{\omega^*}{y}} K'(L(y) - L(a(y)))} - 1)$ messbar bezüglich der σ-Algebra $\mathcal{S}$ ist, gilt

$$\begin{aligned} &E(e^{t\sqrt{\frac{\omega^*}{y}} W_{L(a(y))}} (e^{|t|\sqrt{\frac{\omega^*}{y}} K'(L(y) - L(a(y)))} - 1) | I(y) = 1) \\ &= E\left(e^{t\sqrt{\frac{\omega^*}{y}} W_{L(a(y))}} E(e^{|t|\sqrt{\frac{\omega^*}{y}} K'(L(y) - L(a(y)))} - 1 |\ \mathcal{S}) \Big| I(y) = 1\right) \\ &= E\Big(e^{t\sqrt{\frac{\omega^*}{y}} W_{L(a(y))}} \\ &\qquad \cdot \int_{a(y)}^{a(y)+K} E(e^{|t|\sqrt{\frac{\omega^*}{y}} K'(L(y) - L(a(y)))} - 1 | I(y) = 1, S_{L(a(y))} \leq \xi) dG_y(\xi) \Big| I(y) = 1\Big) \\ &= E\left(e^{t\sqrt{\frac{\omega^*}{y}} W_{L(a(y))}} \int_{a(y)}^{a(y)+K} E(e^{|t|\sqrt{\frac{\omega^*}{y}} K' L(-\xi, y-\xi)} - 1 | I(-\xi, y - \xi) = 1) dG_y(\xi) \Big| I(y) = 1\right). \end{aligned} \tag{2.20}$$

Mit der Bezeichnung

$$D(y) := \sup_{a(y)\le\xi\le a(y)+K} E(e^{|t|\sqrt{\frac{\omega^*}{y}}K'L(-\xi,y-\xi)} - 1|I(-\xi,y-\xi)=1)$$

folgt nun mit (2.19) und (2.20):

$$\begin{aligned}
|\Gamma(y)-\hat{\Gamma}(y)| &\le e^{\theta^* K} D(y) E(e^{t\sqrt{\frac{\omega^*}{y}}W_{L(a(y))}}|I(y)=1) \\
&\le e^{\theta^* K} D(y) E(e^{t\sqrt{\frac{\omega^*}{y}}W_{L(a(y))}} e^{\theta^*(S_{L(y)}-y)}|I(y)=1) \\
&\quad (\text{da } S_L \ge y \text{ unter } I(y)=1) \\
&= e^{\theta^* K} D(y)\hat{\Gamma}(y)
\end{aligned}$$

Wie schon oben angemerkt, genügt es nun zu zeigen, dass $D(y) \underset{y\to\infty}{\longrightarrow} 0$.
$Q_{\xi,y}$ bezeichne die Verteilungsfunktion von $L(-\xi,y-\xi)$ unter $I(-\xi,y-\xi)=1$. Damit gilt:

$$\begin{aligned}
D(y) &= \sup_{a(y)\le\xi\le a(y)+K} E(e^{|t|\sqrt{\frac{\omega^*}{y}}K'L(-\xi,y-\xi)} - 1|I(-\xi,y-\xi)=1) \\
&= \sup_{a(y)\le\xi\le a(y)+K} \left(\int_0^\infty e^{|t|\sqrt{\frac{\omega^*}{y}}K'x} dQ_{\xi,y}(x)\right) - 1 \\
&= \sup_{a(y)\le\xi\le a(y)+K} \Big(\int_0^{y^{1/4}} e^{|t|\sqrt{\frac{\omega^*}{y}}K'x} dQ_{\xi,y}(x) + \int_{y^{1/4}}^\infty e^{|t|\sqrt{\frac{\omega^*}{y}}K'x} dQ_{\xi,y}(x)\Big) - 1 \\
&\le e^{|t|\sqrt{\frac{\omega^*}{y}}K'y^{1/4}} + \sup_{a(y)\le\xi\le a(y)+K} \left(\int_{y^{1/4}}^\infty e^{|t|\sqrt{\frac{\omega^*}{y}}K'x} dQ_{\xi,y}(x)\right) - 1
\end{aligned}$$

Es gilt $(e^{|t|\sqrt{\omega^*}K'y^{-1/4}} - 1) \underset{y\to\infty}{\longrightarrow} 0$, es bleibt also zu zeigen, dass

$$\sup_{a(y)\le\xi\le a(y)+K} \left(\int_{y^{1/4}}^\infty (e^{|t|\sqrt{\frac{\omega^*}{y}}K'x}) dQ_{\xi,y}(x)\right) \underset{y\to\infty}{\longrightarrow} 0.$$

Folgende Teilmengenbeziehung ist erfüllt:

$$\{L(-\xi,y-\xi)\ge x,\ I(-\xi,y-\xi)=1\} \subset \bigcup_{n\ge x}\{S_n \ge 0\}, \tag{2.21}$$

denn unter $I(-\xi,y-\xi)=1$ gilt $L(-\xi,y-\xi)=\inf\{k\ge 1: S_k \ge y-\xi\}$; falls also $L(-\xi,y-\xi)\ge x$, so existiert ein $n\ge x$ mit $S_n \ge y-\xi$; da ξ im Intervall $[a(y),a(y)+K]$ liegt und nach Voraussetzung $y-a(y)=\log y \ge K$ für alle y gilt, ergibt sich $S_n \ge y-\xi \ge y-a(y) = \log y \ge K > 0$.

Falls $I(y-\xi)=1$ ist, gilt $S_{L(y-\xi)} \ge y-\xi$ und $S_k > 0$ für alle $1\le k\le L(y-\xi)$ und somit auch $S_k > -\xi$ für alle $1\le k\le L(y-\xi)$, also $I(-\xi,y-\xi)=1$, folglich gilt

$$\{I(y-\xi)=1\} \subset \{I(-\xi,y-\xi)=1\}. \tag{2.22}$$

Mit diesen beiden Teilmengenbeziehungen lässt sich $1 - Q_{\xi,y}(x)$ folgendermaßen abschätzen:

$$\begin{aligned}
1 - Q_{\xi,y}(x) &= P(L(-\xi, y-\xi) \geq x | I(-\xi, y-\xi) = 1) \\
&= \frac{P(L(-\xi, y-\xi) \geq x,\ I(-\xi, y-\xi) = 1)}{P(I(-\xi, y-\xi) = 1)} \\
&\leq \frac{P(\bigcup_{n \geq x} \{S_n \geq 0\}}{P(I(y-\xi) = 1)} \quad \text{(nach (2.21) und (2.22))} \\
&\leq \frac{\sum_{n=x}^{\infty} P(S_n \geq 0)}{P(I(y-\xi) = 1)}.
\end{aligned}$$

Nach Lemma 2.3.3 gilt $P(I(y-\xi) = 1) \geq \delta e^{-\theta^*(y-\xi)}$ für ein $\delta > 0$, ferner folgt mit der Aussage (2.17), dass $P(S_n \geq 0) \leq e^{-nb}$ für ein $b > 0$.
Also erhält man:

$$1 - Q_{\xi,y}(x) \leq \frac{\sum_{n=x}^{\infty} e^{-nb}}{\delta e^{-\theta^*(y-\xi)}} = \frac{e^{-bx}}{\delta e^{-\theta^*(y-\xi)}} \frac{1}{1-e^{-b}} = C_1 y^{\theta^*} e^{-bx} \tag{2.23}$$

mit $C_1 := \frac{1}{\delta(1-e^{-b})}$, da $y - \xi \leq \log y$.

Mit Hilfe einer partiellen Integration (wobei $f(x) = e^{|t|\sqrt{\frac{\omega^*}{y}}K'x}$ und $g'(x) = \frac{dQ_{\xi,y}}{dx}(x)$) erhält man für genügend großes y

$$\begin{aligned}
&\int_{y^{1/4}}^{\infty} e^{|t|\sqrt{\frac{\omega^*}{y}}K'x} dQ_{\xi,y}(x) \\
&= e^{|t|\sqrt{\frac{\omega^*}{y}}K'x} Q_{\xi,y}(x) \Big|_{y^{1/4}}^{\infty} - |t|\sqrt{\frac{\omega^*}{y}}K' \int_{y^{1/4}}^{\infty} e^{|t|\sqrt{\frac{\omega^*}{y}}K'x} Q_{\xi,y}(x) dx \\
&= e^{|t|\sqrt{\frac{\omega^*}{y}}K'y^{1/4}} - e^{|t|\sqrt{\frac{\omega^*}{y}}K'y^{1/4}} + e^{|t|\sqrt{\frac{\omega^*}{y}}K'x} Q_{\xi,y}(x) \Big|_{y^{1/4}}^{\infty} - |t|\sqrt{\frac{\omega^*}{y}}K' \int_{y^{1/4}}^{\infty} e^{|t|\sqrt{\frac{\omega^*}{y}}K'x} Q_{\xi,y}(x) dx \\
&= e^{|t|\sqrt{\frac{\omega^*}{y}}K'y^{1/4}} (1 - Q_{\xi,y}(y^{1/4}) + e^{|t|\sqrt{\frac{\omega^*}{y}}K'x} \Big|_{y^{1/4}}^{\infty} - |t|\sqrt{\frac{\omega^*}{y}}K' \int_{y^{1/4}}^{\infty} e^{|t|\sqrt{\frac{\omega^*}{y}}K'x} Q_{\xi,y}(x) dx \\
&= e^{|t|\sqrt{\omega^*}K'y^{-1/4}} (1 - Q_{\xi,y}(y^{1/4}) + |t|\sqrt{\frac{\omega^*}{y}}K' \int_{y^{1/4}}^{\infty} e^{|t|\sqrt{\frac{\omega^*}{y}}K'x} (1 - Q_{\xi,y}(x)) dx \\
&\leq e^{|t|\sqrt{\omega^*}K'y^{-1/4}} C_1 y^{\theta^*} e^{-by^{1/4}} + |t|\sqrt{\frac{\omega^*}{y}}K' \int_{y^{1/4}}^{\infty} e^{|t|\sqrt{\frac{\omega^*}{y}}K'x} C_1 y^{\theta^*} e^{-bx} dx \quad \text{(nach (2.23))} \\
&\leq C_2 e^{|t|\sqrt{\omega^*}K'y^{-1/4}} y^{\theta^*} e^{-by^{1/4}} \\
&\leq C_3 e^{-\frac{1}{2}by^{1/4}}.
\end{aligned}$$

Es ergibt sich also schließlich

$$D(y) = \sup_{a(y) \leq \xi \leq a(y)+K} \left(\int_{y^{1/4}}^{\infty} (e^{|t|\sqrt{\frac{\omega^*}{y}}K'x}) dQ_{\xi,y}(x) \right) \leq C_3 e^{-\frac{1}{2}by^{1/4}} \underset{y \to \infty}{\longrightarrow} 0$$

und somit folgt die gewünschte Aussage

$$|\Gamma(y) - \hat{\Gamma}(y)| \underset{y\to\infty}{\longrightarrow} 0.$$

$\square$

Mit den bisher gewonnenen Aussagen sind wir nun in der Lage, unser schon formuliertes „Zwischenziel“ (siehe (2.15)) zu beweisen:

Lemma 2.5.7. *Sei $u^* = E(U_1 e^{\theta^* X_1}) = 0$. Dann gilt für alle $t \in \mathbb{R}$:*

$$\lim_{y\to\infty} E(e^{t\sqrt{\frac{\omega^*}{y}}\sum_{k=K_{\nu-1}+1}^{T_\nu(y)} U_k} | I_\nu(y) = 1) = e^{\frac{1}{2}t^2 v^*}, \quad \textit{wobei } v^* = E(U_1^2 e^{\theta^* X_1}).$$

Beweis. Wie im Beweis von Lemma 2.5.6 sei die σ-Algebra $\mathcal{S}$ durch

$$\mathcal{S} = \sigma(S_1, ..., S_{L(a(y))}, W_1, ..., W_{L(a(y))}, L(a(y)), I(y) = 1)$$

und die Verteilungsfunktion von $S_{L(a(y))}$ unter $I(y) = 1$ durch G_y gegeben. Da $e^{\theta^*(S_L - y)}$ messbar bezüglich $\mathcal{S}$ ist, lässt sich $\hat{\Gamma}(y)$ folgendermaßen darstellen:

$$\begin{aligned}
\hat{\Gamma}(y) &= E(e^{t\sqrt{\frac{\omega^*}{y}}W_{L(a(y))}} e^{\theta^*(S_L - y)} | I(y) = 1) \\
&= E\left(e^{t\sqrt{\frac{\omega^*}{y}}W_{L(a(y))}} E(e^{\theta^*(S_L - y)} |\ \mathcal{S}) \Big| I(y) = 1\right) \\
&= E\left(e^{t\sqrt{\frac{\omega^*}{y}}W_{L(a(y))}} E(e^{\theta^*(S_L - S_{L(a(y))} - (y - S_{L(a(y))}))} |\ \mathcal{S}) \Big| I(y) = 1\right) \\
&= E\Big(e^{t\sqrt{\frac{\omega^*}{y}}W_{L(a(y))}} \int_{a(y)}^{a(y)+K} E(e^{\theta^*(S_{L(-\xi, y-\xi)} - (y-\xi))} | I(-\xi, y-\xi) = 1) dG_y(\xi) \Big| I(y) = 1\Big)
\end{aligned}$$

Nach Lemma 2.5.5 (ii) gilt nun, dass der Integrand für alle $a(y) < \xi < a(y) + K$ gegen $\frac{1}{e^*}$ konvergiert, d.h.

$$E(e^{\theta^*(S_{L(-\xi, y-\xi)} - (y-\xi))} | I(-\xi, y-\xi) = 1) \underset{y\to\infty}{\longrightarrow} \frac{1}{e^*}.$$

Falls die X_i eine Gitterverteilung besitzen, gilt eine entsprechende Aussage, wobei das Integral einer Summe entspricht.
Wir erhalten also

$$\lim_{y\to\infty} \hat{\Gamma}(y) = \frac{1}{e^*} \lim_{y\to\infty} E(e^{t\sqrt{\frac{\omega^*}{y}}W_{L(a(y))}} | I(y) = 1).$$

Ferner gilt

$$\begin{aligned}
&|E(e^{t\sqrt{\frac{\omega^*}{y}}W_{L(y)}}|I(y)=1) - E(e^{t\sqrt{\frac{\omega^*}{y}}W_{L(a(y))}}|I(y)=1)| \\
\leq\ &|E((e^{t\sqrt{\frac{\omega^*}{y}}W_{L(y)}} - e^{t\sqrt{\frac{\omega^*}{y}}W_{L(a(y))}})e^{\theta^*(S_L-y)}|I(y)=1)| \\
&(\text{da } e^{\theta^*(S_L-y)} \geq 1 \text{ unter } I(y)=1) \\
=\ &|\Gamma(y) - \hat{\Gamma}(y)| \underset{y\to\infty}{\longrightarrow} 0 \quad (\text{nach Lemma 2.5.6}),
\end{aligned}$$

also

$$\lim_{y\to\infty} E(e^{t\sqrt{\frac{\omega^*}{y}}W_L}|I(y)=1) = \lim_{y\to\infty} E(e^{t\sqrt{\frac{\omega^*}{y}}W_{L(a(y))}}|I(y)=1).$$

Somit ergibt sich mit Lemma 2.5.6 die Aussage

$$\lim_{y\to\infty} \Gamma(y) = \frac{1}{e^*} \lim_{y\to\infty} E(e^{t\sqrt{\frac{\omega^*}{y}}W_L}|I(y)=1).$$

Da ferner nach Lemma 2.5.3

$$\lim_{y\to\infty} \frac{\Gamma(y)}{E(e^{\theta^*(S_L-y)}|I(y)=1)} = e^{\frac{1}{2}t^2v^*}$$

und nach Lemma 2.5.5 (i)

$$\lim_{y\to\infty} E(e^{\theta^*(S_L-y)}|I(y)=1) = \frac{1}{e^*}$$

gilt, folgt

$$\lim_{y\to\infty} E(e^{t\sqrt{\frac{\omega^*}{y}}W_L}|I(y)=1) = e^{\frac{1}{2}t^2v^*}.$$

Da die (X_i, U_i) iid sind und somit auch die Zufallsgrößen $(I_\nu(y))$ und $(W_{L_\nu(y)})$ (es gilt $W_{L_\nu(y)} = \sum_{k=K_{\nu-1}+1}^{T_\nu(y)} U_k$) ist die Aussage des Lemmas bewiesen. □

Um nun schlussendlich den zentralen Grenzwertsatz dieses Kapitels (siehe Theorem 2.4.1) beweisen zu können, werden an dieser Stelle noch zwei Propositionen zitiert: der Stetigkeitssatz für Laplace-Transformierte (bzw. eine Folgerung daraus) und der Satz von Cramér-Wold. Der Stetigkeitssatz für Laplace-Transformierte wird z.B. durch Alsmeyer (siehe [2], S.233-234) bewiesen, der Beweis vom Satz von Cramér-Wold ist z.B. bei Schmitz nachzulesen (siehe [19], S.248).
Die Laplace-Transformierte einer nichtnegativen Zufallsgröße X ist für $t \geq 0$ gegeben durch

$$\varphi_X(t) := \int_0^\infty e^{-tx} dP^X(x) = E(e^{-tX}).$$

Proposition 2.5.8. ***(Stetigkeitssatz für Laplace-Transformierte)***
(siehe [2], Korollar 45.8, S.235)
Für eine Folge $(P_n)_{n\in\mathbb{N}_0}$ endlicher Maße auf $([0,\infty), \mathbb{B}_{|[0,\infty)})$ mit Laplace-Transformierten $(\varphi_n)_{n\in\mathbb{N}_0}$ gilt: $(P_n)_{n\in\mathbb{N}}$ konvergiert genau dann schwach (d.h. mit unseren Bezeichnungen in Verteilung) gegen P_0, wenn die Folge $(\varphi_n)_{n\in\mathbb{N}}$ auf $[0,\infty)$ gegen φ_0 konvergiert.

Proposition 2.5.9. ***(Satz von Cramér-Wold)*** *(siehe [19], S.248)*
Es sei $(\boldsymbol{X_n})_{n\in\mathbb{N}_0}$ *eine Folge von m-dimensionalen Zufallsvektoren. Dann gilt* $P^{\boldsymbol{X_n}} \underset{(V)}{\longrightarrow} P^{\boldsymbol{X_0}}$ *genau dann, wenn*

$$P^{\langle \boldsymbol{z},\boldsymbol{X_n}\rangle} \underset{(V)}{\longrightarrow} P^{\langle \boldsymbol{z},\boldsymbol{X_0}\rangle} \quad \textit{für alle } \boldsymbol{z} \in \mathbb{R}^m \textit{ mit } |\boldsymbol{z}| = 1.$$

Nun sind wir in der Lage, den zentralen Grenzwertsatz 2.4.1 zu beweisen.

Beweis von Theorem 2.4.1. Nach Lemma 2.5.7 gilt, falls $u^* = 0$, für alle $t \in \mathbb{R}$, dass

$$\lim_{y\to\infty} E(e^{t\sqrt{\frac{\omega^*}{y}}\sum_{k=K_{\nu-1}+1}^{T_\nu(y)} U_k}|I_\nu(y) = 1) = e^{\frac{1}{2}t^2v^*}.$$

Insbesondere gilt also für alle $t > 0$:

(i) $\lim_{y\to\infty} E(e^{-t\sqrt{\frac{\omega^*}{y}}\sum_{k=K_{\nu-1}+1}^{T_\nu(y)} U_k}|I_\nu(y) = 1) = e^{\frac{1}{2}t^2v^*}$

(ii) $\lim_{y\to\infty} E(e^{(-t)(-\sqrt{\frac{\omega^*}{y}}\sum_{k=K_{\nu-1}+1}^{T_\nu(y)} U_k)}|I_\nu(y) = 1) = e^{\frac{1}{2}t^2v^*}$.

Falls die Zufallsgröße $Y := \sqrt{\frac{\omega^*}{y}}\sum_{k=K_{\nu-1}+1}^{T_\nu(y)} U_k$ nichtnegativ ist, so ist der linke Ausdruck von (i) gerade die Laplace-Transformierte von Y unter $I_\nu(y) = 1$, sowie $e^{\frac{1}{2}t^2v^*}$ die Laplace-Transformierte einer $\mathcal{N}(0, v^*)$-Verteilung. Es gilt also für alle $t > 0$:

$$\varphi_{\sqrt{\frac{\omega^*}{y}}\sum_{k=K_{\nu-1}+1}^{T_\nu(y)} U_k}(t) \underset{y\to\infty}{\longrightarrow} \varphi_{\mathcal{N}(0,v^*)}(t).$$

Falls die Zufallsgröße Y negativ ist, so gilt, dass der linke Ausdruck von (ii) die Laplace-Transformierte der (positiven) Zufallsgröße $-Y$ ist, also gilt für alle $t > 0$:

$$\varphi_{-\sqrt{\frac{\omega^*}{y}}\sum_{k=K_{\nu-1}+1}^{T_\nu(y)} U_k}(t) \underset{y\to\infty}{\longrightarrow} \varphi_{\mathcal{N}(0,v^*)}(t).$$

Mit dem Stetigkeitssatz für Laplace-Transformierte (siehe Proposition 2.5.8) folgt also (unabhängig davon, ob Y nichtnegativ oder negativ ist), falls $u^* = 0$:

$$\sqrt{\frac{\omega^*}{y}}\sum_{k=K_{\nu-1}+1}^{T_\nu(y)} U_k \underset{(V)}{\longrightarrow} \mathcal{N}(0, v^*) \quad \text{unter } I_\nu(y) = 1 \quad \text{mit } v^* = E(U_1^2 e^{\theta^* X_1}). \tag{2.24}$$

Seien nun $\boldsymbol{U_1}, \boldsymbol{U_2}, ...$ wie im Theorem iid und beschränkte Zufallsvektoren und $\boldsymbol{z} = (z_1, ..., z_m)$ ein reeller Vektor mit $|\boldsymbol{z}| = 1$.
Wähle nun $U_i = \langle \boldsymbol{z}, \boldsymbol{U_i} - \boldsymbol{u^*}\rangle = \sum_{k=1}^m z_k(U_{ik} - u_k^*),\ i \in \mathbb{N}$
ζ ist also hier durch $\zeta(\theta, t) = \log E(e^{\theta X_1 + t\langle \boldsymbol{z},\boldsymbol{U_1}-\boldsymbol{u^*}\rangle})$ gegeben. Somit gilt

$$\omega^* = \frac{d\zeta}{d\theta}(\theta^*, 0) = E(X_1 e^{\theta^* X_1})$$
$$u^* = \frac{d\zeta}{dt}(\theta^*, 0) = E(\sum_k z_k(U_{ik} - u_k^*)e^{\theta^* X_1}) = \sum_k z_k(E(U_{ik}e^{\theta^* X_1}) - u_k^* E(e^{\theta^* X_1})) = 0,$$

da $E(U_{ik}e^{\theta^* X_1}) = u_k^*,\ E(e^{\theta^* X_1}) = 1$.
Mit (2.24) erhalten wir daher

$$\sqrt{\frac{\omega^*}{y}} \sum_{k=K_{\nu-1}+1}^{T_\nu(y)} \langle \boldsymbol{z}, \boldsymbol{U_i} - \boldsymbol{u^*} \rangle \underset{(V)}{\longrightarrow} \mathcal{N}(0, v^*)$$

unter $I_\nu(y) = 1$ mit $v^* = E(\langle \boldsymbol{z}, \boldsymbol{U_1} - \boldsymbol{u^*} \rangle^2 e^{\theta^* X_1})$. Abschließend folgt mit dem Satz von Cramér-Wold (siehe Proposition 2.5.9) die gewünschte Aussage:

$$\sqrt{\frac{\omega^*}{y}} \sum_{k=K_{\nu-1}+1}^{T_\nu(y)} (\boldsymbol{U_i} - \boldsymbol{u^*}) \underset{(V)}{\longrightarrow} \mathcal{N}(0, \Sigma) \quad \text{unter } I_\nu(y) = 1$$

mit $\Sigma = (\sigma_{ij})_{i,j \in \{1,\dots,m\}}$, wobei $\sigma_{ij} = E((\boldsymbol{U_1} - \boldsymbol{u^*})_i (\boldsymbol{U_1} - \boldsymbol{u^*})_j e^{\theta^* X_1})$. □

2.6 Anwendungen

Wie in der Einführung (Kapitel 1) motiviert, geht es in den biologischen Anwendungen unter anderem darum, DNA- bzw. Proteinsequenzen miteinander zu vergleichen, um so Aufschluß über den Verwandtschafts- bzw. Ähnlichkeitsgrad der Sequenzen zu gewinnen. Wir ordnen nun also wie zu Beginn nicht mehr einer Sequenz, sondern einem Sequenzenpaar einen gemeinsamen Score zu. Formal ersetzt man also bei zwei Squenzen $p_i,\ q_i$ und $\mathcal{A}$ durch $p_\alpha p'_\beta,\ q_{\alpha,\beta}$ und $\mathcal{A}^2$ (vergleiche Anmerkung 1.2.3) bzw. bei drei Sequenzen durch $p_\alpha^{(i)} p_\beta^{(j)},\ i,j \in \{1,2,3\},\ q_{\alpha,\beta}$ und $\mathcal{A}^2$ (man vergleicht hier je zwei der drei Sequenzen).
Die Resultate dieses Kapitels seien nun zunächst auf auf den Fall übertragen, bei welchem man zwei Sequenzen miteinander vergleicht, sowie auf den Fall, bei dem drei Sequenzen miteinander verglichen werden.

2.6.1 Vergleich zweier Sequenzen

Wir haben die gleiche Ausgangssituation wie in Beispiel 1.2.2.
Gegeben sei ein Buchstabenalphabet $\mathcal{A} = \{a_1, \dots, a_r\}$, sowie ein Paar von zwei unabhängigen, identisch verteilten Sequenzen mit Einträgen aus diesem Buchstabenalphabet

$$A_1, \dots, A_n$$
$$A'_1, \dots, A'_n.$$

Sei nun $(X_i)_{i \in \{1,\dots,n\}}$ eine Folge von Zufallsgrößen mit $X_i := X(A_i, A'_i)$, $1 \le i \le n$, wobei $X : \mathcal{A}^2 \to \{s_{\alpha,\beta} \mid \alpha, \beta \in \{1,\dots,r\}\}$ jedem Buchstabenpaar (a_α, a_β) einen Score $s_{\alpha,\beta}$ mit der Wahrscheinlichkeit $p_\alpha p'_\beta$ zuordnet, also

$$P(X(A_i, A'_i) = s_{\alpha,\beta}) = p_\alpha p'_\beta, \quad 1 \le i \le n.$$

Nun stellt sich in den Anwendungen die Frage nach einem geeigneten Scoresystem. Wieviele verschiedene Scorewerte benötigt man eigentlich? X kann theoretisch r^2 Werte annehmen, aber da es in den Anwendungen in der Regel keinen Unterschied macht, ob ein Paar (a_i, a_j) oder ein Paar (a_j, a_i) von Aminosäuren bzw. Nucleotiden auftritt, wird es bei einer Scorezuteilung meist nur $\frac{r^2+r}{2}$ unterschiedliche Scores geben, d.h. bei Aminosäuren ($r = 20$) bräuchte man maximal 210 verschiedene Scorewerte.
Aber welche Scorewerte eignen sich? Im Beispiel 1.2.2 wurde bereits die Verwendung des log-Likelihood-Ratio-Scores anschaulich motiviert. Nach Theorem 2.2.3 bzw. genauer nach Korollar 2.2.4 wissen wir nun aber explizit, dass sich in hochscorigen Segmenten jeder Scorewert darstellen lässt als

$$s_{\alpha,\beta} = \frac{\ln(\frac{q_{\alpha,\beta}}{p_\alpha p'_\beta})}{\theta^*},$$

bzw. bei Wahl einer geeigneten Basis des Logarithmus als $\log(\frac{q_{\alpha,\beta}}{p_\alpha p'_\beta})$, also gerade als log-Likelihood-Ratio-Score.
Die p_α und p'_β lassen sich als relative Häufigkeiten der Aminosäuren bzw. der Nucleotide in der jeweiligen Sequenz wählen. Die Frage nach einem geeigneten Scoresystem reduziert sich also auf die empirische Ermittlung der 'target frequencies' $q_{\alpha,\beta}$, also der Aminosäure- bzw. Nucleotid-Austauschwahrscheinlichkeiten.

In den Anwendungen benutzt man Scorematrizen, sogenannnte Substitutionsmatrizen, die sich dieses log-Likelihood-Ratio-Systems bedienen, d.h. ihre Scorewerte aus empirischen Daten über die Aminosäure-Austauschwahrscheinlichkeiten $q_{\alpha,\beta}$ errechnen. Gängige Substitutionsmatrizen sind bei Aminosäuren etwa sogenannte PAM-Matrizen („Percent of Accepted Mutations“) oder BLOSUM-Matrizen („Blocks Substitution Matrix“). In der Abbildung (2.1) sehen wir beispielsweise eine BLOSUM50 Substitutionsmatrix, eine symmetrische 20×20-Matrix. Die log-Likelihood-Ratio-Scores sind dabei skaliert und zu ganzen Zahlen gerundet.
Wie zu erwarten, sind in dieser Matrix die Diagonal-Scorewerte am höchsten, da dies am ehesten eine Verwandtschaft zweier Sequenzen nahelegt. Dennoch wird nicht jede identische Paarung gleich bewertet, da einige Aminosäuren eher Funktionsträger sind als andere - eine identische Paarung von funktionstragenden Aminosäuren wird also mit einem höheren Score belohnt als eine solche funktionsärmerer Aminosäuren.
Da ferner Austausche funktionell ähnlicher Aminosäuren häufiger zu beobachten sind als Austausche völlig verschiedener Aminosäuren, werden Paarungen außerhalb der Diagonalen völlig unterschiedlich bewertet. Je funktionell ähnlicher sich zwei Aminosäuren sind, desto höher fällt der gemeinsame Score aus.
Welche der vielen gängigen Matrizen (und damit welche empirischen Daten über die 'target frequencies') man nun beim Vergleich zweier Sequenzen zugrunde legt, hängt von den Sequenzen ab, die man untersuchen möchte. Vermutet man ein hoch divergentes Sequenzenpaar, also ein solches geringer Verwandtschaft, so sollte man mit einer anderen Matrix arbeiten als beim Vergleich zweier vermutlich eng verwandter Sequenzen.
Eine PAM1-Matrix beispielsweise bezieht ihre empirischen Daten (also die Aminosäure-Austauschwahrscheinlichkeiten $q_{\alpha,\beta}$) aus der Untersuchung eines Sequenzenalignments,

	A	R	N	D	C	Q	E	G	H	I	L	K	M	F	P	S	T	W	Y	V
A	**5**	-2	-1	-2	-1	-1	-1	0	-2	-1	-2	-1	-1	-3	-1	1	0	-3	-2	0
R	-2	**7**	-1	-2	-4	1	0	-3	0	-4	-3	3	-2	-3	-3	-1	-1	-3	-1	-3
N	-1	-1	**7**	2	-2	0	0	0	1	-3	-4	0	-2	-4	-2	1	0	-4	-2	-3
D	-2	-2	2	**8**	-4	0	2	-1	-1	-4	-4	-1	-4	-5	-1	0	-1	-5	-3	-4
C	-1	-4	-2	-4	**13**	-3	-3	-3	-3	-2	-2	-3	-2	-2	-4	-1	-1	-5	-3	-1
Q	-1	1	0	0	-3	**7**	2	-2	1	-3	-2	2	0	-4	-1	0	-1	-1	-1	-3
E	-1	0	0	2	-3	2	**6**	-3	0	-4	-3	1	-2	-3	-1	-1	-1	-3	-2	-3
G	0	-3	0	-1	-3	-2	-3	**8**	-2	-4	-4	-2	-3	-4	-2	0	-2	-3	-3	-4
H	-2	0	1	-1	-3	1	0	-2	**10**	-4	-3	0	-1	-1	-2	-1	-2	-3	2	-4
I	-1	-4	-3	-4	-2	-3	-4	-4	-4	**5**	2	-3	2	0	-3	-3	-1	-3	-1	4
L	-2	-3	-4	-4	-2	-2	-3	-4	-3	2	**5**	-3	3	1	-4	-3	-1	-2	-1	1
K	-1	3	0	-1	-3	2	1	-2	0	-3	-3	**6**	-2	-4	-1	0	-1	-3	-2	-3
M	-1	-2	-2	-4	-2	0	-2	-3	-1	2	3	-2	**7**	0	-3	-2	-1	-1	0	1
F	-3	-3	-4	-5	-2	-4	-3	-4	-1	0	1	-4	0	**8**	-4	-3	-2	1	4	-1
P	-1	-3	-2	-1	-4	-1	-1	-2	-2	-3	-4	-1	-3	-4	**10**	-1	-1	-4	-3	-3
S	1	-1	1	0	-1	0	-1	0	-1	-3	-3	0	-2	-3	-1	**5**	2	-4	-2	-2
T	0	-1	0	-1	-1	-1	-1	-2	-2	-1	-1	-1	-1	-2	-1	2	**5**	-3	-2	0
W	-3	-3	-4	-5	-5	-1	-3	-3	-3	-3	-2	-3	-1	1	-4	-4	-3	**15**	2	-3
Y	-2	-1	-2	-3	-3	-1	-2	-3	2	-1	-1	-2	0	4	-3	-2	-2	2	**8**	-1
V	0	-3	-3	-4	-1	-3	-3	-4	-4	4	1	-3	1	-1	-3	-2	0	-3	-1	**5**

Abbildung 2.1: BLOSUM50 Substitutionsmatrix für Aminosäuren (siehe [18])

welches sich in 1% der Positionen unterscheidet. Über Extrapolation der PAM1-Matrix wurden weitere PAM-Matrizen erstellt, wie z.B. die PAM250-Matrix für hochdivergente Sequenzen (250 Mutationen pro 100 Aminosäuren).
Bei den BLOSUM-Matrizen, welche mit Hilfe von Daten aus der sogenannten Blocks-Datenbank erstellt wurden, ist es dabei genau umgekehrt. So bedeutet BLOSUM60 beispielsweise, dass ein Sequenzenalignment zu 60% übereinstimmt.
Übliche Programme zum Vergleich einer Sequenz mit Sequenzen aus einer Datenbank (wie z.B. das Programm BLAST, "Basic Local Alignment Search Tool") bedienen sich voreingestellter Substitutionsmatrizenmatrizen, die durch den Benutzer geändert werden können.

So viel zur Komplexität, repräsentative Daten über die 'target frequencies', d.h. Aminosäure-Austauschwahrscheinlichkeiten, zur Bestimmung geeigneter Scores auszuwählen.

Wie schon in Abschnitt 2.1 angedeutet, interessiert man sich in den Anwendungen auch häufig für die Zufallsgröße $M(n)$, die den maximalen Scorewert eines Sequenzenpaares (also den Score des Abschnitts mit maximalen Verwandtschaftsgrad) angibt.
Um Aussagen über $M(n)$ zu erlangen, setzen wir wie bisher eine negative Drift voraus, also

$$E(X_1) = E(X(A_1, A_1')) = \sum_{\alpha,\beta} s_{\alpha,\beta} p_\alpha p_\beta' < 0,$$

und $P(X(A_1, A_1') > 0) > 0$. Der Gesamtscore eines Segmentes der Länge m entspricht hier der Summe

$$S_m := X(A_1, A_1') + ... + X(A_m, A_m'), \quad m \geq 1, \quad S_0 = 0.$$

$$M(n) := \max_{0 \leq k \leq l \leq n} (S_l - S_k)$$

gibt also den maximalen Score eines Segmentes des kompletten Alignments an.
Nun gilt, dass $M(n)$ eine asymptotische Wachstumsrate aufweist:

$$\frac{\theta^* M(n)}{\log(n)} \underset{n\to\infty}{\longrightarrow} 1 \text{ f.s.},$$

was man z.B. bei Dembo und Zeitouni (1993) (siehe [8], Seite 69, Theorem 3.2.1; siehe auch Erdös-Rényi, [10]) nachlesen kann. Also lassen sich die Theoreme 2.2.1 und 2.4.1 (hier Korollar 2.4.2) auch auf den Fall $y = M(n)$ anwenden, d.h. es gilt

(i) $\dfrac{L(M(n))}{M(n)} \underset{n\to\infty}{\longrightarrow} \dfrac{1}{E(X_1 e^{\theta^* X_1})} \quad (I(M(n)) = 1 \text{ trifft natürlich zu}),$

(ii) $\sqrt{\frac{\omega^*}{M(n)}}(M(n) - \omega^* L(M(n))) \underset{(V)}{\longrightarrow} \mathcal{N}(0, \sigma^2) \quad \text{für } n \to \infty,$

wobei $\sigma^2 = E(X_1^2 e^{\theta^* X_1}) - E(X_1 e^{\theta^* X_1})^2$, $\omega^* = E(X_1 e^{\theta^* X_1})$ und θ^* die eindeutig bestimmte positive Lösung von

$$E(e^{\theta X_1}) = \sum_{\alpha,\beta} p_\alpha p'_\beta e^{\theta s_{\alpha,\beta}} = 1 \quad \text{ist.}$$

Eine weitere interessante Aussage über die Zufallsgröße $M(n)$ findet man in einem Artikel von Karlin und Dembo (siehe [14], Theorem A) und zwar gilt unter den bisherigen Voraussetzungen, falls X_1 keine Gitterverteilung besitzt, dass

$$P(M(n) - \frac{\ln n}{\theta^*} > x) \approx 1 - e^{-K^* e^{-\theta^* x}},$$

wobei

$$K^* = \frac{\exp(-2\sum_{k=1}^{\infty} \frac{1}{k}(E(e^{\theta^* S_k} | S_k < 0) P(S_k < 0) + P(S_k \geq 0)))}{\theta^* E(X_1 e^{\theta^* X_1})}.$$

Diese Approximation erweist sich in den Anwendungen als sehr nützlich zur Überprüfung der Signifikanz von hohen Scorewerten (beim Vergleich zweier Sequenzen, aber auch bei der Untersuchung einzelner Sequenzen z.B. auf chemische Eigenschaften), dies sei an dieser Stelle jedoch nur am Rande erwähnt.

Allgemeiner möchte man manchmal auch Aussagen über den maximalen Score eines Sequenzenpaarsegmentes machen, bei welchem das Alignment nicht an den gleichen Positionen stattfinden muss, sondern die Paarung der Sequenzen an verschiedenen Positionen erfolgen kann. Dies ist z.B. der Fall, wenn man in zwei Sequenzen nach sogenannten Motiven sucht, d.h. nahezu identischen Sequenzstücken (siehe auch Abschnitt 1.1.2: Alignments von Sequenzen). Man möchte also im Unterschied zu $M(n)$ die Größe

$$\tilde{M}(n) := \max_{k,m,n} \sum_{\substack{i=m \\ j=n}}^{\substack{m+k \\ n+k}} X(A_i, A'_j)$$

untersuchen. Dazu sei auf einen Artikel von Dembo, Karlin und Zeitouni (siehe [7]) verwiesen, in welchem gezeigt wird, dass die Grenzverteilung von $\tilde{M}(n)$ unter bestimmten Voraussetzungen dieselbe wie diejenige von $M(n^2)$ ist.

2.6.2 Vergleich dreier Sequenzen

Nun habe man drei stochastisch unabhängige, identisch verteilte Buchstabensequenzen

$$(A_i^{(1)})_{i\in\{1,\ldots,n\}}, \quad (A_i^{(2)})_{i\in\{1,\ldots,n\}}, \quad (A_i^{(3)})_{i\in\{1,\ldots,n\}}$$

mit zugehörigen Wahrscheinlichkeiten

$$(p_i^{(1)})_{i\in\{1,\ldots,n\}}, \quad (p_i^{(2)})_{i\in\{1,\ldots,n\}}, \quad (p_i^{(3)})_{i\in\{1,\ldots,n\}}$$

und beim Vergleich von je zwei Sequenzen die zugehörigen Scorewerte

$$(s_{\alpha,\beta}^{(1,2)}), \quad (s_{\alpha,\beta}^{(1,3)}), \quad (s_{\alpha,\beta}^{(2,3)})$$

gegeben. Den maximalen Score erhält man, indem man jeweils zwei von den drei Sequenzen miteinander vergleicht:

$$M(n) := \max_{0\le k\le l\le n}((S_l^{(1,2)} - S_k^{(1,2)}), (S_l^{(1,3)} - S_k^{(1,3)}), (S_l^{(2,3)} - S_k^{(2,3)})).$$

$M(n)$ soll charakterisiert werden. Bisher wurde gezeigt, dass

$$P(M(n) \ge y) \le P(I(y) = 1) \le e^{-\theta^* y}$$

gilt (vergleiche Lemma 2.3.3), es gilt aber sogar

$$P(T(y)e^{-\theta^* y} \le t) \underset{y\to\infty}{\longrightarrow} 1 - e^{-K^* t}, \quad t > 0,$$

wobei $T(y) = \inf\{n : M(n) > y\}$ und K^* eine Konstante ist.
Dies kann man z.B. bei Karlin und Dembo (1992) ([14], Theorem A) nachlesen.
In diesem Fall hat man nun $\theta_{1,2}^*, \theta_{1,3}^*, \theta_{2,3}^*$ als die eindeutig bestimmten Lösungen von

$$E(e^{\theta X_1}) = \sum_{\alpha,\beta} p_\alpha^{(i)} p_\beta^{(j)} e^{\theta s_{\alpha,\beta}^{(i,j)}} = 1, \quad i,j \in \{1,2,3\}, \tag{2.25}$$

und die zugehörigen Parameter $K_{1,2}^*, K_{1,3}^*, K_{2,3}^*$.

Dabei kann man drei Fälle betrachten:

(i) $\theta_{1,2}^* < \min(\theta_{1,3}^*, \theta_{2,3}^*)$:
Hierbei tritt der maximale Scorewert mit Wahrscheinlichkeit 1 im Sequenzenpaar 1 und 2 auf (vergleiche (2.25): je kleiner θ, desto größer ist $s_{\alpha,\beta}^{(i,j)}$), man braucht also nur noch die Sequenz 1 mit der Sequenz 2 zu vergleichen.

(ii) $\theta_{1,2}^* = \theta_{1,3}^* < \theta_{2,3}^*$:
Hierbei tritt der maximale Scorewert mit Wahrscheinlichkeit $\frac{K_{1,2}^*}{K_{1,2}^*+K_{1,3}^*}$ im Sequenzenpaar 1 und 2 und mit Wahrscheinlichkeit $\frac{K_{1,3}^*}{K_{1,2}^*+K_{1,3}^*}$ im Sequenzenpaar 1 und 3 auf; die Sequenzen 2 und 3 brauchen nicht miteinander verglichen zu werden.

(iii) $\theta^*_{1,2} = \theta^*_{1,3} = \theta^\star_{2,3}$:

Hierbei tritt der maximale Scorewert mit Wahrscheinlichkeit $\frac{K^*_{i,j}}{K^*_{1,2}+K^*_{1,3}+K^*_{2,3}}$ im Sequenzenpaar i und j auf; alle Sequenzen müssen miteinander verglichen werden.

Der dritte Fall tritt z.B. ein, wenn man als Scores die natürlichen log-Likelihood-Ratio-Scores (siehe Beispiel 1.2.2) $\ln(\frac{q_i}{p_i})$ wählt, die hier durch

$$s^{(i,j)}_{\alpha,\beta} = \ln\left(\frac{q_{\alpha,\beta}}{p^{(i)}_\alpha p^{(j)}_\beta}\right)$$

gegeben sind.

Nach Korollar 2.2.4 gilt nämlich, dass

$$q_{\alpha,\beta} = p^{(i)}_\alpha p^{(j)}_\beta e^{\theta^*_{i,j} s^{(i,j)}_{\alpha,\beta}} \quad \text{und somit dass} \quad s^{(i,j)}_{\alpha,\beta} = \ln\left(\frac{p^{(i)}_\alpha p^{(j)}_\beta e^{\theta^*_{i,j} s^{(i,j)}_{\alpha,\beta}}}{p^{(i)}_\alpha p^{(j)}_\beta}\right).$$

Also folgt notwendigerweise, dass

$$\theta^*_{1,2} = \theta^*_{1,3} = \theta^\star_{2,3} = 1$$

gelten muss. In diesem Fall müssen also alle Sequenzen miteinander verglichen werden.

Kapitel 3

Grenzwertsätze im Markov-Fall

Während im Kapitel 2 auf der Grundlage eines iid-Modells Aussagen über die Länge und Zusammensetzung von Sequenzen mit hohem Score getroffen wurden, sei nun allgemeiner davon ausgegangen, dass das Auftreten eines Buchstabens (also z.B. eines Aminosäurepaares) vom vorherigen Buchstaben abhängt.
Im Weiteren ordnet man also nicht länger nur einzelnen Nucleotiden, sondern zwei aufeinanderfolgenden Nucleotiden, sogenannten Dinucleotiden, einen Score zu. Beim Vergleich zweier Sequenzen bedeutet dies, dass zwei aufeinanderfolgenden Paaren von Aminosäuen bzw. Basen ein Scorewert zugeordnet wird.
Dies lässt sich mit Hilfe eines Markov-Modells modellieren.

3.1 Notationen und Voraussetzungen

Es sei $\mathcal{A} = \{a_1, ..., a_r\}$ ein Buchstabenalphabet, sowie $A_1, A_2, ...A_n$ eine Sequenz bestehend aus Buchstaben dieses Alphabetes, wobei das Auftreten eines Buchstabens a_β vom vorhergehenden Buchstaben a_α abhänge, d.h. man betrachte die bedingten Wahrscheinlichkeiten (welche in der Praxis als relative Dinucleotidhäufigkeiten interpretiert werden)

$$P(A_{i+1} = a_\beta |\ A_i = a_\alpha) = p_{\alpha\beta}.$$

Diesen Dinucleotiden wird ein gemeinsamer Score $s_{\alpha\beta}$ zugeordnet.
Gegeben sei also eine Folge von Zufallsgrößen (X_i), $X_i := \mathcal{A}^2 \to \mathbb{R}$ mit

$$P(X_i = s_{\alpha,\beta}) = P(A_{i-1} = a_\alpha, A_i = a_\beta) = P(A_{i-1} = a_\alpha)p_{\alpha\beta}, \quad 1 \leq i \leq r,$$

d.h. die Zufallsgröße X_i hängt von A_{i-1} und von A_i ab.
Man betrachtet den Gesamt-Score $S_{n\alpha} = \sum_{i=1}^n X_i$, wobei $A_0 = \alpha$ den Anfangsbuchstaben der Sequenz bezeichne.
Da man die Zusammensetzung hochscoriger Segmente untersuchen möchte, interessiert man sich auch für die Zufallsgröße $W_{n\alpha}$, welche die Häufigkeit des Auftretens von Dinucleotiden (a_α, a_β) in einer Sequenz angibt, d.h. für

$$W_{n\alpha} = \sum_{i=1}^n U_i \quad \text{mit} \quad U_i = \begin{cases} 1 & \text{falls } X_i = s_{\alpha,\beta} \\ 0 & \text{sonst} \end{cases}.$$

Allgemeiner modelliert man die vorliegende Situation folgendermaßen:
Es sei $(A_n)_{n\in\mathbb{N}_0}$ eine homogene Markov-Kette mit endlichem Zustandsraum $\Omega = \{1, ..., r\}$, d.h. für alle $n \geq 0$ und alle $i_0, ..., i_{n-1}, i, j \in \Omega$ gelte

$$P(A_{n+1} = j|\ A_0 = i_0, ..., A_{n-1} = i_{n-1}, A_n = i) = P(A_{n+1} = j|\ A_n = i) =: p_{ij}\ P\text{-f.s.},$$

welche irreduzibel sei, es gelte also

$$\sup_{n\geq 0} p_{ij}^{(n)} > 0 \quad \text{für alle } 1 \leq i, j \leq r,$$

wobei die n-fachen Übergangswahrscheinlichkeiten $p_{ij}^{(n)}$ für $n \geq 0$ rekursiv durch

$$p_{ij}^{(0)} = \delta_{ij}, \quad p_{ij}^{(n)} = \sum_{k\in\Omega} p_{ik}^{(n-1)} p_{kj}$$

gegeben sind.

Ferner sei $(X_n, U_n)_{n\in\mathbb{N}}$ eine Folge von unter $(A_n)_{n\in\mathbb{N}_0}$ unabhängigen Zufallsgrößen. Die Verteilung von (X_n, U_n) hänge nur von den Werten von A_{n-1} und A_n ab und besitze einen beschränkten Träger. $(A_n, (X_n, U_n))_{n\in\mathbb{N}}$ ist also eine homogene Markov-Kette mit Zustandsraum $\Omega \times \mathbb{R}^2$.

Im Folgenden betrachtet man den Partialsummenprozess $(S_{ni})_{n\in\mathbb{N}_0}$ mit
$S_0 = 0$ und $S_{ni} = \sum_{j=1}^n X_j$, wobei $A_0 = i$ den Anfangszustand der Sequenz bezeichne, sowie $(W_{ni})_{n\in\mathbb{N}_0}$ mit $W_0 = 0,\ W_{ni} = \sum_{j=1}^n U_j$.
Das Paar $(A_n, (S_n, W_n))_{n\in\mathbb{N}_0}$ wird auch Markov-additiver Prozess genannt.
(Im Folgenden wird der Anfanszustand i meist weggelassen, so dass beispielsweise statt S_{ni} die Bezeichnung S_n benutzt wird.)

Von besonderem Interesse ist wie im iid-Fall das Segment der Sequenz $(S_m)_{m\in\{1,...,n\}}$ (Anfangszustand $A_0 = i$) mit maximalem Scorewert $M_i(n)$, welcher hier durch

$$M_i(n) := \sup_{0\leq k<l\leq n} (S_l - S_k)$$

gegeben ist.

Man setzt folgende Bedingungen voraus:

(i) $E(X_1) = \sum_{i,j=1}^r P(A_0 = i) p_{ij} E(X_1 | A_0 = i, A_1 = j) < 0$

(ii) Es existiere ein endlicher Zyklus $\mathcal{C} = \{A_0 = i_0, ..., A_k = i_k = i_0\}$, derart dass $P(\{\min_{m=1,...,k} \sum_{i=1}^m X_i > 0\} \cap \mathcal{C}) > 0$,

so dass der Prozess eine negative Drift erhält, aber eine positive Wahrscheinlichkeit, frühzeitig einen positiven Score anzunehmen.

Bei vorgegebenem Anfanszustand definiert man analog zum iid-Fall (vergleiche (2.1)-(2.3)) die Stoppzeiten $(K_{\nu-1})_{\nu\in\mathbb{N}}$, $(T_\nu)_{\nu\in\mathbb{N}}$, sowie $(I_\nu(y))_{\nu\in\mathbb{N}}$ durch

$$\begin{aligned}
&K_0 = 0,\ K_\nu = \min\{k \geq K_{\nu-1}+1 : S_k - S_{K\nu-1} \leq 0\},\\
&T_\nu(y) = \min\{m > K_{\nu-1} : S_m - S_{K_{\nu-1}} \leq 0 \text{ oder } S_m - S_{K_{\nu-1}} \geq y\},\\
&I_\nu(y) := \begin{cases} 1 & \text{falls } S_{T_\nu(y)} - S_{K_{\nu-1}} \geq y \\ 0 & \text{falls } S_{T_\nu(y)} - S_{K_{\nu-1}} \leq 0 \end{cases}, \quad \nu \in \mathbb{N}.
\end{aligned}$$

Aufgrund der negativen Drift von (S_m) sind die (K_ν) endlich (vergleiche (2.1)). Für festes i sind die Zufallsgrößen $(K_\nu - K_{\nu-1})$ unter $A_{K_{\nu-1}} = i,\ \nu \in \mathbb{N}$ iid.
Die Länge des ν-ten Segmentes $(K_{\nu-1}, T_\nu(y))$ ist wie in Kapitel 2 durch

$$L_\nu(y) = T_\nu(y) - K_{\nu-1}$$

gegeben, wobei die Verteilung von $L_\nu(y)$ hier allerdings vom Zustand von $A_{_{\nu-1}}$ abhängt. Für festes i sind die Zufallsgrößen $L_\nu(y)$ unter $A_{K_{\nu-1}} = i,\ \nu \in \mathbb{N}$ iid. Außerdem bezeichne wie bisher

$$W_\nu(y) = \sum_{m=K_{\nu-1}+1}^{T_\nu(y)} U_m.$$

Ferner führt man bei vorgegebenem Anfangszustand (siehe (2.4)) folgende Stoppzeiten ein:

$$\begin{aligned}
\kappa_\nu(y) &= \inf\{k \geq \sigma_{\nu-1}(y) : \exists m \geq k+1 : S_j - S_k > 0 \text{ für alle } k+1 \leq j \leq m-1\\
&\qquad \text{und } S_m - S_k \geq y\}\\
\tau_\nu(y) &= \inf\{k \geq \kappa_\nu(y) : S_k - S_{\kappa_\nu(y)} \geq y\}\\
\sigma_\nu(y) &= \inf\{k \geq \tau_\nu(y) : S_k - S_{\kappa_\nu(y)} \leq 0\}.
\end{aligned} \tag{3.1}$$

Mit diesen Bezeichnungen gilt für alle $i \in \Omega,\ \nu \in \mathbb{N}$ (siehe Proposition 2.1.1):

$$P^{\tau_\nu(y)-\kappa_\nu(y)|A_{K_{\nu-1}}=i} = P^{L_\nu(y)|I_\nu(y)=1, A_{K_{\nu-1}}=i}.$$

Dies ermöglicht es, ohne die Bedingung $I_\nu(y) = 1$ weiterzurechnen.

Um die Grenzwertsätze (siehe Abschnitt 3.3) dieses Kapitels zu formulieren und zu beweisen, werden im Folgenden Aussagen aus der sogenannten Spektral- oder Frobenius-Theorie nicht-negativer Matrizen bzw. in unserem Fall irreduzibler Matrizen benötigt. Diese seien nun in einem seperaten Abschnitt vorweggenommen.

3.2 Auszüge aus der Spektraltheorie nicht-negativer Matrizen

Es sei $A = (a_{ij})_{i,j\in\{1,2,\ldots,n\}}$ eine $n \times n$-Matrix. Falls alle Einträge a_{ij} von A nicht-negativ sind, schreiben wir $A \geq 0$, falls alle Einträge positiv sind $A \gg 0$.

Im Folgenden wird diese Notation ebenso für Vektoren benutzt, dabei bezeichnen wir einen Vektor x mit $x \gg 0$ als streng positiv.

Eine Zahl λ heißt Eigenwert der Matrix A, falls ein Vektor $\pi \neq 0$ existiert mit $A\pi = \lambda\pi$. π heißt der zu λ gehörige Rechtseigenvektor von A. Ein Vektor $\psi \neq 0$ heißt zu λ gehöriger Linkseigenvektor von A, falls $\psi A = \lambda\psi$ gilt. Der Spektralradius ρ der Matrix A ist der betragsmäßig größte Eigenwert von A.

Das folgende Theorem macht eine Aussage über die Existenz des Spektralradius' einer irreduziblen Matrix und über die sich daraus ergebenden Eigenschaften der zugehörigen Rechts- und Linkseigenvektoren.

Theorem 3.2.1. ***(Perron-Frobenius-Theorem für irreduzibele Matrizen)***
(siehe [20], S.20)
Es sei $A \geq 0$ eine irreduzible $n \times n$-Matrix. Dann existiert ein Eigenwert ρ, welcher folgende Eigenschaften erfüllt:

a) $\rho > 0$

b) zu ρ existieren streng positive Rechts- und Linkseigenvektoren

c) $\rho \geq |\lambda|$ *für jeden Eigenwert λ von A*

d) die zu ρ gehörigen Rechts- und Linkseigenvektoren sind eindeutig bis auf Multiplikation mit einer Konstanten.

Zum *Beweis*: siehe z.B. Seneta „Non-Negative Matrices“, [20], Seite 20-22.

In den Beweisen (siehe Abschnitt 3.4) wird uns die folgende Darstellung des Spektralradius' einer Matrix hilfreich sein.

Proposition 3.2.2. *Es sei $A \geq 0$ eine irreduzible $n \times n$-Matrix, ρ der Spektralradius von A und $\pi = (\pi(1), ..., \pi(n))$ sowie $\psi = (\psi(1), ..., \psi(n))$ die zugehörigen Rechts- bzw. Linkseigenvektoren. Dann gilt für alle $1 \leq i \leq n$:*

$$\rho = \sum_j \frac{a_{ij}\pi(j)}{\pi(i)}$$

und für alle $1 \leq j \leq n$:

$$\rho = \sum_i \frac{a_{ij}\psi(i)}{\psi(j)}.$$

Beweis. Die Existenz von $\rho > 0$ ist nach dem Perron-Frobenius-Theorem (siehe Theorem 3.2.1) sichergestellt, π und ψ können als streng positiv angenommen werden.

Per Definition gilt $A\pi = \rho\pi$, also

$$\begin{pmatrix} a_{11}\pi(1) + ... + a_{1r}\pi(n) \\ \vdots \\ a_{r1}\pi(1) + ... + a_{rr}\pi(n) \end{pmatrix} = \begin{pmatrix} \rho\pi(1) \\ \vdots \\ \rho\pi(n) \end{pmatrix}.$$

Die erste Aussage ergibt sich somit per Division durch $\pi(i) > 0,\ 1 \leq i \leq n$, die zweite Aussage analog aus $\psi A = \rho\psi$ und per Division durch $\psi(i) > 0,\ 1 \leq i \leq n$. □

Schließlich wird für die spätere Beweisführung eine Abschätzung des Spektralradius' des sogenannten Schur-Produktes zweier Matrizen benötigt.

Definition 3.2.3. *Das Schur-Produkt (oder Hadamard-Produkt) zweier Matrizen $A = (a_{ij})_{ij}$ und $B = (b_{ij})_{ij}$ wird definiert durch*

$$A \circ B = (a_{ij}b_{ij})_{ij}.$$

Theorem 3.2.4. *(siehe [16]) A und B seien zwei Matrizen mit $A, B \geq 0$. Dann gilt für den Spektralradius des Schurproduktes von A und B:*

$$\rho(A \circ B) \leq \sqrt{\rho(A \circ A)}\sqrt{\rho(B \circ B)}.$$

Wenn A und B zusätzlich irreduzibel sind, tritt Gleichheit genau dann ein, wenn $A = DBD^{-1}$ für eine positive Diagonalmatrix D gilt.

Zum *Beweis*: siehe z.B. Karlin, Ost „Some Monotonicity Properties of Schur Powers of Matrices and Related Inequalities“ ([16]).

3.3 Zwei starke Grenzwertsätze

Nun seien die Hauptresultate dieses Kapitels formuliert: die Grenzwertsätze. Der erste Grenzwertsatz macht eine Aussage über die Länge von Segmenten mit hohem Score, diesmal unter der Hypothese eines Markov-Modells.

Dazu sei $\rho(\theta)$ der Spektralradius der Matrix

$$P_\theta = (p_{ij}E(e^{\theta X_1}|A_0 = i, A_1 = j))_{i,j\in\{1,...,r\}}$$

sowie $\pi_\theta = (\pi_\theta(1), ..., \pi_\theta(r)) \gg 0$ und $\psi_\theta \gg 0$ die zugehörigen Rechts- und Linkseigenvektoren. Diese werden o.B.d.A. normiert, so dass $\langle \pi_\theta, \psi_\theta \rangle = 1$ und $\langle \pi_\theta, e \rangle = \sum_{j=1}^{r} \pi_\theta(j) = 1$ gilt.
Die Existenz derartiger Eigenvektoren und des Spektralradius' ist nach dem Perron-Frobenius-Theorem für irreduzible Matrizen (siehe Theorem 3.2.1) sichergestellt, da P_θ aufgrund der Irreduzibilität der Matrix $P = (p_{ij})_{i,j\in\{1,...r\}}$ ebenfalls irreduzibel ist.

Theorem 3.3.1. *Mit den Bezeichnungen aus Abschnitt 3.1 sei ν der erste Index, für den $I_\nu(y) = 1$ gilt, d.h. $\nu = \min\{k \geq 1 : I_k(y) = 1\}$. Dann folgt*

$$\frac{L_\nu(y)}{y} \xrightarrow[y\to\infty]{} \frac{1}{\omega^*} \quad \textit{fast sicher,}$$

für jeden Wert von A_0. Dabei ist

$$\omega^* = \sum_{i,j=1}^{r} p_{ij} E(X_1 e^{\theta^* X_1} | A_0 = i, A_1 = j) \pi_{\theta^*}(j) \psi_{\theta^*}(i)$$

und θ^* die eindeutig bestimmte positive Lösung der Gleichung $\rho(\theta) = 1$.

Anmerkung 3.3.2. Es existiert eine eindeutig bestimmte positive Lösung θ^* von $\rho(\theta) = 1$ und es gilt

$$\omega^* = \sum_{i,j=1}^{r} p_{ij} E(X_1 e^{\theta^* X_1} | A_0 = i, A_1 = j) \pi_{\theta^*}(j) \psi_{\theta^*}(i) = \rho'(\theta^*) > 0.$$

Beweis. Nach Proposition 3.2.2 gilt, dass $\rho(\theta)$ folgende Darstellungen besitzt:

$$\rho(\theta) = \sum_{j=1}^{r} p_{ij} E(e^{\theta X_1} | A_0 = i, A_1 = j) \frac{\pi_\theta(j)}{\pi_\theta(i)} \ , \quad 1 \leq i \leq n \tag{3.2}$$

$$\rho(\theta) = \sum_{i=1}^{r} p_{ij} E(e^{\theta X_1} | A_0 = i, A_1 = j) \frac{\psi_\theta(i)}{\psi_\theta(j)} \ , \quad 1 \leq j \leq n, \tag{3.3}$$

wobei π_θ und ψ_θ die zugehörigien Rechts- bzw. Linkseigenvektoren sind.
Im Folgenden werden nun die folgenden Behauptungen schrittweise bewiesen:

(i) ρ ist unendlich oft stetig differenzierbar mit Ableitung $\rho'(\theta) = \sum_{i,j=1}^{r} p_{ij} E(X_1 e^{\theta X_1} | A_0 = i, A_1 = j) \pi_\theta(j) \psi_\theta(i)$

(ii) ρ ist streng konvex in θ

(iii) $\rho(0) = 1$

(iv) $\rho'(0) < 0$

(v) $\rho(\theta) \xrightarrow[\theta\to+\infty]{} +\infty$.

zu (i): $\rho(\theta)$ ist analytisch, da alle Einträge der Matrix P_θ analytisch sind, also unendlich oft stetig differenzierbar.
Um die erste Ableitung von ρ zu bestimmen, erweitert man den Ausdruck (3.2) auf beiden Seiten mit $\pi_\theta(i)$, differenziert ihn nach θ, erweitert ihn mit $\psi_\theta(i)$ und summiert anschließend über i auf, so dass man die Gleichheit

$$\begin{aligned}\rho'(\theta)\sum_{i=1}^{r}\pi_\theta(i)\psi_\theta(i)+\rho(\theta)\sum_{i=1}^{r}\pi'_\theta(i)\psi_\theta(i) &= \sum_{i,j=1}^{r} p_{ij}E(X_1e^{\theta X_1}|A_0=i,A_1=j)\pi_\theta(j)\psi_\theta(i)\\ &\quad+\sum_{j=1}^{r}\pi'_\theta(j)\sum_{i=1}^{r}p_{ij}E(e^{\theta X_1}|A_0=i,A_1=j)\psi_\theta(i).\end{aligned}$$

erhält. Aufgrund der Normierung der Eigenvektoren gilt aber nun $\sum_i \pi_\theta(i)\psi_\theta(i)=1$, ferner folgt nach Anwendung von (3.3), dass

$$\sum_{j=1}^{r}\pi'_\theta(j)\sum_{i=1}^{r}p_{ij}E(e^{\theta X_1}|A_0=i,A_1=j)\psi_\theta(i)=\sum_{j=1}^{r}\pi'_\theta(j)\rho(\theta)\psi_\theta(j).$$

Damit folgt die gewünschte Aussage.

zu (ii): Es genügt zu zeigen, dass $\log\rho(\theta)$ streng konvex ist, also dass für alle θ_1,θ_2 mit $\theta_1\neq\theta_2$

$$\log\left(\rho\left(\frac{\theta_1+\theta_2}{2}\right)\right)<\frac{1}{2}\left(\log(\rho(\theta_1))+\log(\rho(\theta_2))\right)\quad\text{bzw.}\quad \rho\left(\frac{\theta_1+\theta_2}{2}\right)<\sqrt{\rho(\theta_1)}\sqrt{\rho(\theta_2)}$$

gilt, denn dann folgt mit der Ungleichung zwischen dem geometrischen und arithmetischen Mittel ($\sqrt{\rho(\theta_1)}\sqrt{\rho(\theta_2)}\leq\frac{1}{2}(\rho(\theta_1)+\rho(\theta_2))$) auch die Konvexität von $\rho(\theta)$.

Um dies zu beweisen seien nun zwei $r\times r$-Matrizen

$$A:=(\sqrt{p_{ij}}\;E(e^{\frac{\theta_1X_1}{2}}|A_0=i,A_1=j))_{i,j}$$

und

$$B:=(\sqrt{p_{ij}}\;E(e^{\frac{\theta_2X_1}{2}}|A_0=i,A_1=j))_{i,j}$$

gegeben. A und B sind nichtnegativ und aufgrund der Irreduzibilität von P_θ irreduzibel. Um das Theorem 3.2.4 über den Spektralradius des Schurproduktes zweier Matrizen anwenden zu können, bestimmen wir das Schurprodukt von $A\circ B$, $A\circ A$ bzw. $B\circ B$:

$$\begin{aligned}a_{ij}b_{ij}&=p_{ij}\sqrt{E(e^{\theta_1X_1}|A_0=i,A_1=j)}\sqrt{E(e^{\theta_2X_1}|A_0=i,A_1=j)}\\ &\geq p_{ij}E(e^{\frac{(\theta_1+\theta_2)X_1}{2}}|A_0=i,A_1=j)\quad\text{(nach Cauchy-Schwarzscher Ungleichung)}\\ a_{ij}a_{ij}&=p_{ij}E(e^{\theta_1X_1}|A_0=i,A_1=j)\\ b_{ij}b_{ij}&=p_{ij}E(e^{\theta_2X_1}|A_0=i,A_1=j)\end{aligned}$$

Man erhält also nach Theorem 3.2.4 die folgende Aussage:

$$\rho\left(\frac{\theta_1+\theta_2}{2}\right)\leq\rho(A\circ B)\leq\sqrt{\rho(A\circ A)\rho(B\circ B)}=\sqrt{\rho(\theta_1)}\sqrt{\rho(\theta_2)},\tag{3.4}$$

dabei tritt Gleichheit in (3.4) genau dann ein, wenn X_1 unter der Bedingung $A_0 = i, A_1 = j$ konstant ist (vergleiche oben: Cauchy-Schwarzsche Ungleichung) und wenn $A = D^{-1}BD$ für eine positive Diagonalmatrix D gilt.
In (3.4) gelte also Gleichheit, zu zeigen ist nun, dass $\theta_1 = \theta_2$ gilt.

Sei also $E(e^{\theta X_1}|A_0 = i, A_1 = j) = s_{ij}$ und $A = D^{-1}BD$ mit $D = \begin{pmatrix} \lambda_1 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \lambda_r \end{pmatrix}$, d.h. es gilt $\sqrt{p_{ij}}e^{\frac{\theta_1 s_{ij}}{2}} = \lambda_i^{-1}\sqrt{p_{ij}}e^{\frac{\theta_2 s_{ij}}{2}}\lambda_j, \quad 1 \leq i, j \leq r.$
Falls $p_{ij} > 0$ für alle i, j folgt unmittelbar, dass $\theta_1 = \theta_2$ gelten muss. Falls i, j existieren mit $p_{ij} = 0$ lässt sich diese Aussage aus der Irreduzibiliät folgern, denn dann gilt, dass ein n existiert, derart dass $p_{ij}^{(n)} > 0$ für alle i, j.
Damit folgt die gewünschte Aussage.

zu (iii) Es gilt: $\rho(0) = \sum_{j=1}^{r} p_{ij} = 1$.

zu (iv): Nach (3.2) gilt

$$\rho(\theta)\pi_\theta(i) = \sum_{j=1}^{r} p_{ij}E(e^{\theta X_1}|A_0 = i, A_1 = j)\pi_\theta(j).$$

Leitet man diesen Ausdruck nach θ an der Stelle $\theta = 0$ ab, erweitert dann auf beiden Seiten mit $P(A_0 = i)$ und summiert über i auf, so erhält man:

$$\rho'(0) = \sum_{i,j=1}^{r} P(A_0 = i)p_{ij}E(X_1|A_0 = i, A_1 = j) = E(X_1) < 0 \quad \text{nach Voraussetzung.}$$

zu (v): Nach (3.2) gilt

$$\rho(\theta) = E\left(e^{\theta X_1}\frac{\pi_\theta(A_1)}{\pi_\theta(A_0)}\Big|A_0 = i\right) \geq \int_{\{X_1>0, A_0=i\}} e^{\theta X_1}\frac{\pi_\theta(A_1)}{\pi_\theta(A_0)}\, dP \quad \underset{\theta\to\infty}{\longrightarrow} +\infty,$$

da nach Voraussetzung ein endlicher Zyklus $\mathcal{C} = \{A_0 = i_0, ..., A_{i_{k_0}} = i_{k_0} = i_0\}$ existiert mit $P(\sum_{i=1}^{m} X_i > 0, m = 1, ..., k_0,\ \mathcal{C}\ ) > 0$.

Mit den Aussagen (i)-(v) und dem Mittelwertsatz folgt also, dass ein eindeutig bestimmtes $\theta^* > 0$ mit $\rho(\theta^*) = 1$ existiert und dass
$\rho'(\theta^*) = \sum_{i,j} p_{ij}E(X_1 e^{\theta^* X_1}|A_0 = i, A_1 = j)\pi_{\theta^*}(j)\psi_{\theta^*}(i) = \omega^* > 0$ gilt.

□

Jetzt wird analog zum iid-Fall eine Aussage über die Zusammensetzung hochscoriger Segmente getroffen werden.

Theorem 3.3.3. *Mit den Bezeichnungen aus Abschnitt 3.1 sei ν der erste Index, für den $I_\nu(y) = 1$ gilt, d.h. $\nu = \min\{k \geq 1 : I_k(y) = 1\}$. Dann folgt*

$$\frac{W_\nu(y)}{L_\nu(y)} \quad \underset{y\to\infty}{\longrightarrow} \quad u^* \quad \textit{fast sicher,}$$

für jeden Wert von A_0. Dabei ist

$$u^* = \sum_{i,j=1}^{r} p_{ij} E(U_1 e^{\theta^* X_1} | A_0 = i, A_1 = j) \pi_{\theta^*}(j) \psi_{\theta^*}(i)$$

und θ^* die eindeutig bestimmte positive Lösung der Gleichung $\rho(\theta) = 1$.

3.4 Beweis der starken Grenzwertsätze

Die Beweisführung entspricht derjenigen im iid-Fall. Es werden insgesamt acht Lemmata bewiesen, welche die Aussagen des iid-Falls (siehe Lemma 2.3.2 - 2.3.9) auf den Markov-Fall verallgemeinern.

Zunächst werden also Eigenschaften der Markov-Ketten-Erweiterung des Waldmartingals benötigt, die man u.a. durch Anwendung des Optional-Sampling-Theorems (siehe Theorem 2.3.1) erhält.

Im Folgenden seien θ^*, ω^* und u^* bestimmt wie in den Theoremen 3.3.1 und 3.3.3 und $\rho(\theta, t)$ der Spektralradius der Matrix

$$(p_{ij} E(e^{\theta X_1 + tU_1} | A_0 = i, A_1 = j))_{i,j\in\{1,\dots r\}}.$$

Lemma 3.4.1. ***(Eigenschaften der Markov-Ketten-Erweiterung des Waldmartingals)***
Die durch

$$Y_m := \frac{e^{\theta S_m + tW_m}}{\rho(\theta, t)^m} \frac{\pi_{\theta,t}(A_m)}{\pi_{\theta,t}(A_0)}, \quad m \in \mathbb{N}_0$$

gegebene Zufallsgröße erfüllt folgende Eigenschaften:

(i) $(Y_m, \mathcal{S}_m)_{m\in\mathbb{N}_0}$ *ist ein Martingal bezüglich $A_0, A_1, \dots$.*

(ii) (Y_m) *und $L := L_1(y)$ genügen für jeden Wert von A_0 und genügend kleines t und θ nahe bei θ^* den Voraussetzungen des Optional-Sampling-Theorems.*

(iii) $E(Y_L | A_0 = i) = 1,\ 1 \leq i \leq r$ *für genügend kleines t und θ nahe bei θ^*.*

Beweis. zu (i): zu zeigen ist: $E(Y_{n+1}| \mathcal{S}_n) = Y_n$ für alle $n \in \mathbb{N}$, wobei

$$\mathcal{S}_n = \sigma(A_0, ..., A_n, X_1, ..., X_n, U_1, ..., U_n).$$

$$\begin{aligned}
& E\left(\frac{e^{\theta S_{n+1}+tW_{n+1}}}{\rho(\theta,t)^{n+1}} \frac{\pi_{\theta,t}(A_{n+1})}{\pi_{\theta,t}(A_0)} \Big| \ \mathcal{S}_n \right) \\
= & \frac{e^{\theta S_n+tW_n}}{\rho(\theta,t)^{n+1}} \ \frac{1}{\pi_{\theta,t}(A_0)} \ E(e^{\theta X_{n+1}+tU_{n+1}} \pi_{\theta,t}(A_{n+1})| \ \mathcal{S}_n) \\
& (\text{ da } S_n, W_n \text{ und } A_0 \text{ messbar bzgl. } \mathcal{S}_n) \\
= & \frac{e^{\theta S_n+tW_n}}{\rho(\theta,t)^{n+1}} \ \frac{1}{\pi_{\theta,t}(A_0)} \ E(e^{\theta X_{n+1}+tU_{n+1}} \pi_{\theta,t}(A_{n+1})| \ A_n) \\
& (\text{da } (X_i, U_i) \text{ stochastisch unabhängig unter } (A_i) \text{ und nach Markov-Eigenschaft}) \\
= & \frac{e^{\theta S_n+tW_n}}{\rho(\theta,t)^{n+1}} \ \frac{1}{\pi_{\theta,t}(A_0)} \ \sum_j p_{A_n j} E(e^{\theta X_{n+1}+tU_{n+1}} | \ A_n, A_{n+1} = j) \pi_{\theta,t}(j) \\
= & \frac{e^{\theta S_n+tW_n}}{\rho(\theta,t)^{n+1}} \ \frac{1}{\pi_{\theta,t}(A_0)} \ \rho(\theta,t)\pi_{\theta,t}(A_n) \\
& (\text{da nach Proposition 3.2.2 für alle } i: \\
& \rho(\theta,t)\pi_{\theta,t}(i) = \sum_j p_{ij} E(e^{\theta X_1+tU_1} | A_0 = i, A_1 = j)\pi_{\theta,t}(j)) \\
= & \frac{e^{\theta S_n+tW_n}}{\rho(\theta,t)^{n+1}} \ \frac{\pi_{\theta,t}(A_n)}{\pi_{\theta,t}(A_0)},
\end{aligned}$$

zu (ii): L ist nach Definition eine Stoppzeit; nach (i) gilt, dass $(Y_m, \mathcal{S}_m)_{m \in \mathbb{N}_0}$ ein Martingal bezüglich $A_0, A_1, ...$ ist;
$E(Y_L|A_0 = i)$ existiert, denn mit der Bezeichnung $c(\theta,t) = \min_j \pi_{\theta,t}(j)$ gilt, da $\pi_{\theta,t}(i) \leq 1$:

$$\begin{aligned}
E(|Y_L| \ |A_0 = i) = & E\left(\left| \frac{e^{\theta S_L+tW_L}}{\rho(\theta,t)^L} \ \frac{\pi_{\theta,t}(A_L)}{\pi_{\theta,t}(A_0)} \right| \Big| A_0 = i \right) \\
\leq & \ \frac{e^{|\theta|(y+K)}}{c(\theta,t)} \sum_{l=1}^{\infty} \left(\frac{e^{K'|t|}}{|\rho(\theta,t)|} \right)^l P(L = l|A_0 = i) \quad < \infty
\end{aligned}$$

für $\dfrac{e^{K'|t|}}{|\rho(\theta,t)|} \leq 1$, also für $|t|$ genügend klein und θ nahe bei θ^*, da $|S_L| \leq y + K$, sowie $|W_L| \leq L(y)K'$.
Ferner gilt für $|t|$ genügend klein und θ nahe bei θ^*:

$$\begin{aligned}
\int_{\{L>m, A_0=i\}} |Y_m| \ dP = & \int_{\{L>m, A_0=i\}} \left| \frac{e^{\theta S_m+tW_m}}{\rho(\theta,t)^m} \frac{\pi_{\theta,t}(A_m)}{\pi_{\theta,t}(A_0)} \right| \ dP \\
\leq & \ \frac{e^{|\theta|y+|t|mK'}}{| \ \rho(\theta,t)^m| \ c(\theta,t)} P(L > m, A_0 = i) \underset{m \to \infty}{\longrightarrow} 0,
\end{aligned}$$

denn es gilt nach dem Schwachen Gesetz der großen Zahlen, da S_m nach Voraussetzung eine negative Drift aufweist, dass:

$$P(L > m) \leq P(S_m \in (0, y)) \leq P(S_m > 0) = P(\frac{S_m}{m} > 0) \underset{n \to \infty}{\longrightarrow} 0.$$

zu (iii): Mit dem Optional-Sampling-Theorem folgt für genügend kleines t und θ nahe bei θ^* für alle i: $E(Y_L|A_0 = i) = E(Y_1|A_0 = i)$.
Daher gilt:

$$E\left(\frac{e^{\theta S_L + tW_L}}{\rho(\theta,t)^L} \frac{\pi_{\theta,t}(A_L)}{\pi_{\theta,t}(A_0)}\Big| A_0 = i\right) = E\left(\frac{e^{\theta X_1 + tU_1}}{\rho(\theta,t)} \frac{\pi_{\theta,t}(A_1)}{\pi_{\theta,t}(A_0)}\Big| A_0 = i\right)$$
$$= \frac{1}{\rho(\theta,t)} \sum_j \frac{p_{ij} E(e^{\theta X_1 + tU_1}\Big| A_0 = i, A_1 = j)\pi_{\theta,t}(j)}{\pi_{\theta,t}(i)} = \frac{\rho(\theta,t)}{\rho(\theta,t)} = 1$$

nach Proposition 3.2.2. □

Nach Lemma 3.4.1 (iii) gilt also mit $\zeta(\theta,t) := \log \rho(\theta,t)$ für genügend kleines t und θ nahe bei θ^*:

$$E\left(e^{\theta S_L + tW_L - L\zeta(\theta,t)} \frac{\pi_{\theta,t}(A_L)}{\pi_{\theta,t}(A_0)}\Big| A_0 = i\right) = 1. \tag{3.5}$$

Sei nun

$$\begin{aligned} &\Omega_+ := \{i \in \{1,...,r\} : \text{für alle } y > 0 \text{ gilt } P(I_1(y) = 1|A_0 = i) > 0\} \quad \text{und} \\ &\Omega_- = \{i \in \{1,...,r\} : \text{es existiert ein } y_0 > 0, \text{ so dass } P(I_1(y_0) = 1|A_0 = i) = 0\} \end{aligned} \tag{3.6}$$

das Komplement von Ω_+ in Ω.
Analog zum iid-Fall soll gezeigt werden, dass die Wahrscheinlichkeit $P(I_1(y) = 1|A_0 = i)$ echt größer 0 und exponentiell klein ist. Diese Aussage lässt sich allerdings nicht für beliebiges $i \in \Omega$ nachweisen.
Sei nämlich $i \in \Omega_-$. Aufgrund der Tatsache, dass $I_1(y)$ monoton fallend in y ist, existiert ein y_0, so dass $P(I_1(y) = 1|A_0 = i) = 0$ für alle $y \geq y_0$ gilt.
Für $i \in \Omega_+$ kann jedoch eine zum iid-Fall analoge Aussage getroffen werden (vergleiche Lemma 2.3.3).

Lemma 3.4.2. *Für alle $y > 0$ und $i \in \Omega_+$ gilt:*

$$0 < \delta \leq P(I_1(y) = 1|A_0 = i)e^{\theta^* y} \leq C,$$

wobei C und δ nicht von i abhängen.

Beweis. Zunächst wird die Wahrscheinlichkeit $P(I_1(y) = 1|A_0 = i)$ nach oben abgeschätzt.
Dazu sei $t = 0$ und $\theta = \theta^*$, so dass die Aussagen von Lemma 3.4.2 zutreffen und $\log \rho(\theta^*, 0) = 0$ gilt (nach Definition von θ^*). Mit (3.5) und der Bezeichnung $c(\theta^*, 0) := \min_j \pi_{\theta^*,0}(j) > 0$ folgt:

$$\begin{aligned}
1 &= E\left(e^{\theta^* S_L}\frac{\pi_{\theta^*,0}(A_L)}{\pi_{\theta^*,0}(A_0)}\Big|A_0 = i\right)\\
&= P(I_1(y) = 1|A_0 = i)e^{\theta^* y}E\left(e^{\theta^*(S_L-y)}\frac{\pi_{\theta^*,0}(A_L)}{\pi_{\theta^*,0}(A_0)}\Big|A_0 = i, I_1(y) = 1\right)\\
&\quad + P(I_1(y) = 0|A_0 = i)E\left(e^{\theta^* S_L}\frac{\pi_{\theta^*,0}(A_L)}{\pi_{\theta^*,0}(A_0)}\Big|A_0 = i, I_1(y) = 0\right)\\
&\geq P(I_1(y) = 1|A_0 = i)e^{\theta^* y}E\left(\frac{\pi_{\theta^*,0}(A_L)}{\pi_{\theta^*,0}(A_0)}\Big|A_0 = i, I_1(y) = 1\right)\\
&\quad \text{(da } \pi_{\theta^*,0}(i) > 0 \text{ für alle } i \text{ und } S_L \geq y \text{ unter } I_1(y) = 1)\\
&\geq P(I_1(y) = 1|A_0 = i)e^{\theta^* y}c(\theta^*, 0) \quad \text{(da } \pi_{\theta^*,0}(i) \leq 1 \text{ für alle } i).
\end{aligned}$$

Also erhält man mit $C := (c(\theta^*, 0))^{-1} > 0$, welches unabhängig von i ist, dass

$$P(I_1(y) = 1|A_0 = i)e^{\theta^* y} \leq C \quad \text{gilt.}$$

Um die Wahrscheinlichkeit $P(I_1(y) = 1|A_0 = i)$ nach unten abzuschätzen, wählt man nun $a_- < 0 < a_+$ und definiert

$$L(a_-, a_+) := \inf\{k \geq 1 : S_k \notin (a_-, a_+)\} \quad \text{sowie}$$

$$I(a_-, a_+) := \begin{cases} 1 & \text{falls } S_{L(a_-,a_+)} \geq a_+ \\ 0 & \text{falls } S_{L(a_-,a_+)} \leq a_- \end{cases}.$$

$I(a_-, a_+)$ ist sowohl in a_- als auch in a_+ monoton fallend, denn:

(i) sei $b_+ > a_+$, dann impliziert $I(a_-, b_+) = 1$ auch $I(a_-, a_+) = 1$, da $S_{L(a_-,b_+)} \geq b_+ > a_+$ und $S_k > a_-$ für alle $1 \leq k \leq L(a_-, b_+)$; also gilt $I(a_-, b_+) \leq I(a_-, a_+)$

(ii) sei $b_- > a_-$, dann impliziert $I(b_-, a_+) = 1$ auch $I(a_-, a_+) = 1$, da $S_{L(b_-,a_+)} \geq a_+$ und $S_k > b_- > a_-$ für alle $1 \leq k \leq L(b_-, a_+)$; also gilt $I(b_-, a_+) \leq I(a_-, a_+)$.

Aufgrund dieser Monotonie gilt für alle $y_0 \in (0, y - K)$ (wobei K derart, dass $|X_i| \leq K$ für alle i):

$$P(I_1(y) = 1|A_0 = i) \geq P(I_1(y_0) = 1|A_0 = i)\min_j P(I(-y_0, y - y_0) = 1|A_0 = j). \quad (3.7)$$

Nun wendet man das Optional-Sampling-Theorem (siehe Theorem 2.3.1) auf die Stoppzeit $\tilde{L} := L(-y_0, y - y_0)$ und das Martingal

$$Y_m := e^{\theta^* S_m}\frac{\pi_{\theta^*,0}(A_m)}{\pi_{\theta^*,0}(A_0)}, \quad m \in \mathbb{N}_0$$

an (dies ist die Markov-Ketten-Erweiterung des Waldmartingals für $t = 0$ und $\theta = \theta^*$; Nachweis der Martingaleigenschaft siehe Lemma 3.4.1). Dass die Voraussetzungen dazu erfüllt sind, wird ganz analog zu Lemma 2.3.2 (ii) gezeigt.
Folglich gilt:

$$\begin{aligned}
1 &= E\left(e^{\theta^* S_{\tilde{L}}}\frac{\pi_{\theta^*}(A_{\tilde{L}})}{\pi_{\theta^*}(A_0)}|A_0 = j\right)\\
&= E\left(e^{\theta^* S_{\tilde{L}}}\frac{\pi_{\theta^*}(A_{\tilde{L}})}{\pi_{\theta^*}(A_0)}|A_0 = j, I(-y_0, y-y_0) = 1\right)P(I(y_0, y-y_0) = 1|A_0 = j)\\
&\quad + E\left(e^{\theta^* S_{\tilde{L}}}\frac{\pi_{\theta^*}(A_{\tilde{L}})}{\pi_{\theta^*}(A_0)}|A_0 = j, I(-y_0, y-y_0) = 0\right)P(I(y_0, y-y_0) = 0|A_0 = j)\\
&\leq \frac{e^{\theta^*(y-y_0+K)}}{c(\theta^*,0)}P(I(y_0, y-y_0) = 1|A_0 = j)\\
&\quad + \frac{e^{-\theta^* y_0}}{c(\theta^*,0)}(1 - P(I(y_0, y-y_0) = 1|A_0 = j)).
\end{aligned}$$

(da $S_{\tilde{L}} \leq y - y_0 + K$ unter $I(y_0, y-y_0) = 1$ und $S_{\tilde{L}} \leq -y_0$ unter $I(y_0, y-y_0) = 0$)

Es gilt demnach also

$$\begin{aligned}
\frac{c(\theta^*,0) - e^{-\theta^* y_0}}{e^{\theta^*(y-y_0+K)} - e^{-\theta^* y_0}} &\leq P(I(-y_0, y-y_0) = 1|A_0 = j), \quad \text{was äquivalent ist zu}\\
\frac{c(\theta^*,0)e^{\theta^* y_0} - 1}{e^{\theta^* y}(e^{\theta^* K} - e^{-\theta^* y})} &\leq P(I(-y_0, y-y_0) = 1|A_0 = j) \quad \text{für alle } j.
\end{aligned}$$

Also erhält man mit (3.7), dass für alle $y > y_0 + K$ folgende Aussage erfüllt ist:

$$\begin{aligned}
e^{\theta^* y}P(I_1(y) = 1|A_0 = j) &\geq P(I_1(y_0) = 1|A_0 = j)\frac{c(\theta^*,0)e^{\theta^* y_0} - 1}{e^{\theta^* K} - e^{-\theta^* y}}\\
&\geq \frac{\varepsilon}{e^{\theta^* K} - 1} =: \delta > 0,
\end{aligned}$$

denn es gilt für alle $i \in \Omega_+$, da $I_1(y)$ monoton fallend in y und da $y_0 < y - K$, sowie $X_1 \geq -K$, dass

$$\begin{aligned}
P(I_1(y_0) = 1|A_0 = i) &\geq P(I_1(y-K) = 1|A_0 = i)\\
&\geq P(\sum_{j=2}^{L(y-k)} X_j \geq y|A_0 = i) =: \varepsilon > 0 \quad (\varepsilon \text{ unabhängig von } i).
\end{aligned}$$

Also gilt aufgrund der Monotonie von $I_1(y)$ die Aussage

$$e^{\theta^* y}P(I_1(y) = 1|A_0 = j) \geq \delta > 0$$

auch für alle $y \leq y_0 + K$.

$\square$

Es folgen nun analog zum iid-Fall zwei Abschätzungen für die beiden Zufallsgrößen $\frac{L_1(y)}{y}$ und $\frac{W_1(y)}{L_1(y)}$, die sich aus Lemma 3.4.2 herleiten lassen.

Lemma 3.4.3. *Für $y \longrightarrow \infty$ und alle $i \in \Omega_+$ gilt:*

$$E\left(\left(\frac{L_1(y)}{y} - \frac{1}{\omega^*}\right)^4 \Big| \; I_1(y) = 1, A_0 = i\right) = O\left(\frac{1}{y^2}\right),$$

wobei die Schranke in $O(\frac{1}{y^2})$ nicht von i abhängt.

Beweis. Der Beweis verläuft analog zum Beweis von Lemma 2.3.4 im iid-Fall. Sei

$$G_n(j,i) := \frac{d^n}{d\theta^n} \log\left(\frac{\pi_{\theta,0}(i)}{\pi_{\theta,0}(j)}\right)\Big|_{\theta=\theta^*},$$

$$\hat{S}_L := S_L - G_1(A_L, i) = S_L - \frac{d}{d\theta}\log\left(\frac{\pi_{\theta,0}(i)}{\pi_{\theta,0}(A_L)}\right)\Big|_{\theta=\theta^*} \quad \text{und} \quad k(\theta) = \frac{d^2 \log\rho}{d\theta^2}(\theta, 0).$$

Man leitet nun den Ausdruck

$$E\left(e^{\theta S_L + tW_L - L\zeta(\theta,t)} \frac{\pi_{\theta,t}(A_L)}{\pi_{\theta,t}(A_0)}\Big| A_0 = i\right) = 1$$

für $t = 0$ nach θ an der Stelle $\theta = \theta^*$ ab.
Einmaliges Ableiten ergibt:

$$\begin{aligned}
0 &= \frac{d}{d\theta} E\left(e^{\theta S_L - L\log\rho(\theta,0)} \frac{\pi_{\theta,0}(A_L)}{\pi_{\theta,0}(A_0)}\Big| A_0 = i\right)\Big|_{\theta=\theta^*} \\
&= E\left(e^{\theta S_L - L\log\rho(\theta,0)} \left(\left(S_L - L\frac{\rho'(\theta,0)}{\rho(\theta,0)}\right)\frac{\pi_{\theta,0}(A_L)}{\pi_{\theta,0}(A_0)} + \frac{d}{d\theta}\frac{\pi_{\theta,0}(A_L)}{\pi_{\theta,0}(A_0)}\right)\Big| A_0 = i\right)\Big|_{\theta=\theta^*} \\
&= E\left(e^{\theta S_L - L\log\rho(\theta,0)} \frac{\pi_{\theta,0}(A_L)}{\pi_{\theta,0}(A_0)}\left(S_L - L\frac{\rho'(\theta,0)}{\rho(\theta,0)} + \left(\frac{d}{d\theta}\frac{\pi_{\theta,0}(A_L)}{\pi_{\theta,0}(A_0)}\right)\frac{\pi_{\theta,0}(A_0)}{\pi_{\theta,0}(A_L)}\right)\Big| A_0 = i\right)\Big|_{\theta=\theta^*} \\
&= E\left(e^{\theta S_L - L\log\rho(\theta,0)} \frac{\pi_{\theta,0}(A_L)}{\pi_{\theta,0}(A_0)}\left(S_L - L\frac{\rho'(\theta,0)}{\rho(\theta,0)} - \frac{d}{d\theta}\log\left(\frac{\pi_{\theta,0}(A_0)}{\pi_{\theta,0}(A_L)}\right)\right)\Big| A_0 = i\right)\Big|_{\theta=\theta^*} \\
&= E\left(e^{\theta^* S_L}\frac{\pi_{\theta^*,0}(A_L)}{\pi_{\theta^*,0}(A_0)}(\hat{S}_L - L\omega^*)\Big| A_0 = i\right), \quad \text{da } \rho(\theta^*,0) = 1 \text{ und } \frac{d\rho}{d\theta}(\theta^*,0) = \omega^*.
\end{aligned}$$

Also ergibt sich nach Multiplikation mit $\pi_{\theta^*,0}(i) > 0$, dass

$$E((\hat{S}_L - L\omega^*)e^{\theta^* S_L}\pi_{\theta^*,0}(A_L)|A_0 = i) = 0 \quad \text{gilt.} \tag{3.8}$$

Ein zweites Ableiten nach θ an der Stelle $\theta = \theta^*$ ergibt:

$$
\begin{aligned}
0 &= \frac{d^2}{d\theta^2} E\left(e^{\theta S_L - L\log\rho(\theta,0)} \frac{\pi_{\theta^*,0}(A_L)}{\pi_{\theta^*,0}(A_0)} \Big| A_0 = i \right) \Big|_{\theta=\theta^*} \\
&= \frac{d}{d\theta} E\left(e^{\theta S_L - L\log\rho(\theta,0)} \frac{\pi_{\theta,0}(A_L)}{\pi_{\theta,0}(A_0)} \left(S_L - L\frac{\rho'(\theta,0)}{\rho(\theta,0)} - \frac{d}{d\theta}\log\left(\frac{\pi_{\theta,0}(A_0)}{\pi_{\theta,0}(A_L)}\right)\right) \Big| A_0 = i \right) \Big|_{\theta=\theta^*} \\
&= E\left(e^{\theta S_L - L\log\rho(\theta,0)} \frac{\pi_{\theta,0}(A_L)}{\pi_{\theta,0}(A_0)} \left(S_L - L\frac{\rho'(\theta,0)}{\rho(\theta,0)} - \frac{d}{d\theta}\log\left(\frac{\pi_{\theta,0}(A_0)}{\pi_{\theta,0}(A_L)}\right)\right)^2 \Big| A_0 = i \right) \Big|_{\theta=\theta^*} \\
&\quad + E\left(e^{\theta S_L - L\log\rho(\theta,0)} \frac{\pi_{\theta,0}(A_L)}{\pi_{\theta,0}(A_0)} \left((-L)k(\theta) - \frac{d^2}{d\theta^2}\log\left(\frac{\pi_{\theta,0}(A_0)}{\pi_{\theta,0}(A_L)}\right)\right) \Big| A_0 = i \right) \Big|_{\theta=\theta^*} \\
&= E\left(e^{\theta^* S_L} \frac{\pi_{\theta^*,0}(A_L)}{\pi_{\theta^*,0}(A_0)} (\hat{S}_L - L\omega^*)^2 \Big| A_0 = i \right) \\
&\quad - E\left(e^{\theta^* S_L} \frac{\pi_{\theta^*,0}(A_L)}{\pi_{\theta^*,0}(A_0)} (Lk(\theta^*) + G_2(A_L,i)) \Big| A_0 = i \right).
\end{aligned}
$$

Man erhält also nach Multiplikation mit $\pi_{\theta^*,0}(i)$ folgende Aussage:

$$
\begin{aligned}
&E((\hat{S}_L - L\omega^*)^2 e^{\theta^* S_L} \pi_{\theta^*,0}(A_L) | A_0 = i) \\
&= E((k(\theta^*)L + G_2(A_L,i)) e^{\theta^* S_L} \pi_{\theta^*,0}(A_L) | A_0 = i).
\end{aligned}
\tag{3.9}
$$

Ein drittes Mal Ableiten führt zu folgender Gleichung:

$$
\begin{aligned}
&E((\hat{S}_L - L\omega^*)^4 e^{\theta^* S_L} \pi_{\theta^*,0}(A_L) | A_0 = i) \\
&= 6E((\hat{S}_L - L\omega^*)^2 (k(\theta^*)L + G_2(A_L,i)) e^{\theta^* S_L} \pi_{\theta^*,0}(A_L) | A_0 = i) \\
&\quad + 4E((\hat{S}_L - L\omega^*)(k'(\theta^*)L + G_3(A_L,i)) e^{\theta^* S_L} \pi_{\theta^*,0}(A_L) | A_0 = i) \\
&\quad + E((k''(\theta^*)L + G_4(A_L,i)) e^{\theta^* S_L} \pi_{\theta^*,0}(A_L) | A_0 = i) \\
&\quad - 3E((k(\theta^*)L + G_2(A_L,i))^2 e^{\theta^* S_L} \pi_{\theta^*,0}(A_L) | A_0 = i).
\end{aligned}
\tag{3.10}
$$

$\pi_{\theta,0}$ ist nach Voraussetzung streng positiv (dies folgt aus dem Perron-Frobenius-Theorem (siehe Theorem 3.2.1)), ferner ist $\pi_{\theta,0}$ analytisch in θ, da alle Einträge der zugehörigen Matrix P_θ analytisch sind.
Daher gilt

$$
\max_{j,i} |G_n(j,i)| \le C, \quad n \in \{1, ..., 4\}.
\tag{3.11}
$$

Nun werden analog zum Beweis von Lemma 2.3.4 im iid-Fall schrittweise die folgenden fünf Behauptungen bewiesen. Für alle $i \in \Omega_+$ gilt:

(i) $E(Le^{\theta^* S_L} \pi_{\theta^*,0}(A_L) | A_0 = i) = O(y)$

(ii) $E((\hat{S}_L - \omega^* L)^2 e^{\theta^* S_L} \pi_{\theta^*,0}(A_L) | A_0 = i) = O(y)$

(iii) $\xi^2 := E(L^2 e^{\theta^* S_L} \pi_{\theta^*,0}(A_L) | A_0 = i) = O(y^2)$

(iv) $\eta^2 := E((\hat{S}_L - \omega^* L)^4 e^{\theta^* S_L} \pi_{\theta^*,0}(A_L) | A_0 = i) = O(y^2)$

(v) $E((y - \omega^* L_1(y))^4 | I_1(y) = 1, A_0 = i) = O(y^2)$

zu (i): Da nach Lemma 3.4.2 für alle $i \in \Omega_+$

$$P(I_1(y) = 1 | A_0 = i) e^{\theta^* y} \in [\delta, C] \quad \text{und nach (3.11)}$$

$$\begin{aligned} &y \leq S_L \leq y + K \quad \text{sowie} \quad y - C \leq \hat{S}_L \leq y + K - C \quad \text{unter} \quad I_1(y) = 1 \quad \text{bzw.} \\ &-KL \leq S_L \leq 0 \quad \text{sowie} \quad -KL - C \leq \hat{S}_L \leq -C \quad \text{unter} \quad I_1(y) = 0 \end{aligned}$$

gilt, folgt mit (3.8) für genügend großes y und geeignete C_1 und C_2, sowie für alle $i \in \Omega_+$:

$$\begin{aligned} &E(L e^{\theta^* S_L} \pi_{\theta^*,0}(A_L) | A_0 = i) \\ &= \frac{1}{\omega^*} [P(I_1(y) = 1 | A_0 = i) e^{\theta^* y} E(\hat{S}_L e^{\theta^*(S_L - y)} \pi_{\theta^*,0}(A_L) | I_1(y) = 1, A_0 = i) \\ &\quad + P(I_1(y) = 0 | A_0 = i) E(\hat{S}_L e^{\theta^* S_L} \pi_{\theta^*,0}(A_L) | I_1(y) = 0, A_0 = i)] \\ &\in [C_1 y, C_2 y]. \end{aligned}$$

zu (ii): $\pi_{\theta^*,0}(i)$ sowie S_L sind beschränkt. Also folgt mit (3.9), (3.11) und Behauptung (i), dass

$$\begin{aligned} &E((\hat{S}_L - \omega^* L)^2 e^{\theta^* S_L} \pi_{\theta^*,0}(A_L) | A_0 = i) \\ &= E\left((k(\theta^*) L + G_2(A_L, i)) e^{\theta^* S_L} \pi_{\theta^*,0}(A_L) \middle| A_0 = i\right) \\ &= k(\theta^*) E(L e^{\theta^* S_L} \pi_{\theta^*,0}(A_L) | A_0 = i) + E(G_2(A_L, i) e^{\theta^* S_L} \pi_{\theta^*,0}(A_L) | A_0 = i) = O(y). \end{aligned}$$

zu (iii): Es gilt

$$\begin{aligned} D^2 &:= E(\hat{S}_L e^{\theta^* S_L} \pi_{\theta^*,0}(A_L) | A_0 = i) \\ &= P(I_1(y) = 1 | A_0 = i) e^{\theta^* y} E(\hat{S}_L^2 e^{\theta^*(S_L - y)} \pi_{\theta^*,0}(A_L) | I_1(y) = 1, A_0 = i) \\ &\quad + P(I_1(y) = 0 | A_0 = i) E(\hat{S}_L^2 e^{\theta^* S_L} \pi_{\theta^*,0}(A_L) | I_1(y) = 0, A_0 = i) \\ &\in [C_1 y^2, C_2 y^2] \quad \text{(analog zu (i))}. \end{aligned}$$

Daraus lässt sich mit Hilfe von (ii) folgern, dass mit den Bezeichnungen $D = O(y)$ und $E = O(y^2)$

$$\begin{aligned} \omega^{*2} \xi^2 &= 2\omega^* E(\hat{S}_L L e^{\theta^* S_L} \pi_{\theta^*,0}(A_L) | A_0 = i) - O(y^2) + O(y) \\ &\leq 2\omega^* \xi D + E, \end{aligned}$$

gilt, da nach Cauchy-Schwarzscher Ungleichung:

$$E(\hat{S}_L L e^{\theta^* S_L} \pi_{\theta^*,0}(A_L) | A_0 = i) \leq D\xi.$$

Die größte Wurzel aus dieser quadratischen Ungleichung in ω^* ist durch Cy für eine geeignete Konstante $C > 0$ beschränkt, also gilt, da ξ positiv ist, dass $\xi^2 = O(y^2)$.

zu (iv): Mit Hilfe der Cauchy-Schwarzschen Ungleichung folgt:

$$\begin{aligned}
&E((\hat{S}_L - \omega^* L)^2 L e^{\theta^* S_L} \pi_{\theta^*,0}(A_L) | A_0 = i) \leq \eta\xi \\
&E((\hat{S}_L - \omega^* L) L e^{\theta^* S_L} \pi_{\theta^*,0}(A_L) | A_0 = i) \leq \sqrt{\eta\xi}\xi.
\end{aligned}$$

Somit ergibt sich unter Ausnutzung von (3.10):

$$\begin{aligned}
\eta^2 &= 6k(\theta^*)E((\hat{S}_L - L\omega^*)^2 L e^{\theta^* S_L} \pi_{\theta^*,0}(A_L) | A_0 = i) \\
&\quad + 6E((\hat{S}_L - L\omega^*)^2 e^{\theta^* S_L} G_2(A_L, i) \pi_{\theta^*,0}(A_L) | A_0 = i) \\
&\quad + 4k'(\theta^*)E((\hat{S}_L - L\omega^*) L e^{\theta^* S_L} \pi_{\theta^*,0}(A_L) | A_0 = i) \\
&\quad + 4E((\hat{S}_L - L\omega^*) G_3(A_L, i) e^{\theta^* S_L} \pi_{\theta^*,0}(A_L) | A_0 = i) \\
&\quad + k''(\theta^*)E(L e^{\theta^* S_L} \pi_{\theta^*,0}(A_L) | A_0 = i) \\
&\quad + E(G_4(A_L, i) e^{\theta^* S_L} \pi_{\theta^*,0}(A_L) | A_0 = i) \\
&\quad - 3(k(\theta^*))^2 E(L^2 e^{\theta^* S_L} \pi_{\theta^*,0}(A_L) | A_0 = i) \\
&\quad + 6k(\theta^*)E(L G_2(A_L, i) e^{\theta^* S_L} \pi_{\theta^*,0}(A_L) | A_0 = i) \\
&\quad - 3E((G_2(A_L, i))^2 e^{\theta^* S_L} \pi_{\theta^*,0}(A_L) | A_0 = i). \\
&\leq 6k(\theta^*)\eta\xi + 4k'(\theta^*)\sqrt{\eta\xi}\xi + k''(\theta^*)E(L e^{\theta^* S_L} \pi_{\theta^*,0}(A_L) | A_0 = i) \\
&\quad - 3(k(\theta^*))^2 \xi^2 + O(y) \\
&= O(y)\eta + O(y^2) \\
&\quad \text{nach (i), (ii), (iii), sowie (3.8) und (3.11) und da } S_L \text{ beschränkt ist.}
\end{aligned}$$

Somit erhält man analog zur obigen Argumentation im Beweis zur Behauptung (iii): $\eta^2 = O(y^2)$.

zu (v): Da nach Lemma 2.3.3

$$0 < \delta \leq P(I_1(y) = 1 | A_0 = i) e^{\theta^* y}$$

für alle $i \in \Omega_+$ gilt, folgt mit (iv) unmittelbar, dass

$$\begin{aligned}
O(y^2) &= E((\hat{S}_L - \omega^* L)^4 e^{\theta^*(S_L - y)} \pi_{\theta^*,0}(A_L) | I_1(y) = 1, A_0 = i) \\
&= E((\hat{S}_L - y + y - \omega^* L)^4 e^{\theta^*(S_L - y)} \pi_{\theta^*,0}(A_L) | I_1(y) = 1, A_0 = i).
\end{aligned}$$

Da $\hat{S}_L - y$ und $e^{\theta^*(S_L - y)} \pi_{\theta^*,0}(A_L)$ unter $I_1(y) = 1$ beschränkt sind, ergibt sich

$$E((y - \omega^* L_1(y))^4 | I_1(y) = 1, A_0 = i) = O(y^2).$$

Zusammen mit (v) erhält man schließlich die gewünschte Aussage:

$$
\begin{aligned}
E\left(\left(\frac{L_1(y)}{y}-\frac{1}{\omega^*}\right)^4 \Big| I(y)=1, A_0=i\right) &= E\left(\frac{(y-\omega^* L_1(y))^4}{\omega^{*4} y^4}\Big| I(y)=1, A_0=i\right) \\
&= O\left(\frac{1}{y^2}\right).
\end{aligned}
$$

□

Lemma 3.4.4. *Für $y \longrightarrow \infty$ und alle $i \in \Omega_+$ gilt:*

$$
E\left(\left(\frac{W_1(y)}{L_1(y)}-u^*\right)^4 \Big| I_1(y)=1, A_0=i\right) = O\left(\frac{1}{y^2}\right)
$$

wobei die Schranke in $O(\frac{1}{y^2})$ nicht von i abhängt.

Beweis. Der Beweis erfolgt analog zum Beweis von Lemma 3.4.3.
Zunächst wird der Ausdruck (3.5)

$$
E\left(e^{\theta S_L + tW_L - L\zeta(\theta,t)} \frac{\pi_{\theta,t}(A_L)}{\pi_{\theta,t}(A_0)}\Big| A_0 = i\right) = 1
$$

nun dreimal nach t an der Stelle $t=0$ abgeleitet, wobei man $\theta = \theta^*$ setzt. Man erhält analog zum vorhergehenden Beweis mit den Bezeichnungen

$$
H_n(j,i) := \frac{d^n}{dt^n} \log\left(\frac{\pi_{\theta^*,t}(i)}{\pi_{\theta^*,t}(j)}\right)\Big|_{t=0},
$$

$$
\hat{W}_L := W_L - H_1(A_L, i) = W_L - \frac{d}{dt}\log\left(\frac{\pi_{\theta^*,t}(i)}{\pi_{\theta^*,t}(A_L)}\right)\Big|_{t=0} \quad \text{und}
$$

$$
\kappa(\theta^*,0) = \frac{d^2}{dt^2}\left(\log(\rho(\theta^*,t))\right)\Big|_{t=0}
$$

die folgenden drei Ableitungen:

$$
E((\hat{W}_L - u^* L)e^{\theta^* S_L}\pi_{\theta^*,0}(A_L)|A_0=i) = 0, \quad \text{da } u^* = \frac{d\rho}{dt}(\theta^*,0) = \frac{d\log\rho}{dt}(\theta^*,0) \tag{3.12}
$$

(wobei sich die Aussage $u^* = \frac{d\rho}{dt}(\theta^*,0)$ völlig analog zur Aussage $\omega^* = \frac{d\rho}{d\theta}(\theta^*,0)$, welche im Beweis zur Anmerkung 3.3.2 gezeigt wird, beweisen lässt)

$$
\begin{aligned}
&E((\hat{W}_L - u^* L)^2 e^{\theta^* S_L}\pi_{\theta^*,0}(A_L)|A_0=i) \\
&= E((\kappa(\theta^*,0)L + H_2(A_L,i))e^{\theta^* S_L}\pi_{\theta^*,0}(A_L)|A_0=i)
\end{aligned} \tag{3.13}
$$

$$
\begin{aligned}
&E((\hat{W}_L - u^*L)^4 e^{\theta^* S_L} \pi_{\theta^*,0}(A_L)|A_0 = i) \\
&= 6E((\hat{W}_L - u^*L)^2(\kappa(\theta^*,0)L + H_2(A_L,i))e^{\theta^* S_L} \pi_{\theta^*,0}(A_L)|A_0 = i) \\
&\quad + 4E((\hat{W}_L - u^*L)(\frac{d\kappa}{dt}(\theta^*,0)L + H_3(A_L,i))e^{\theta^* S_L} \pi_{\theta^*,0}(A_L)|A_0 = i) \\
&\quad + E((\frac{d^2\kappa}{dt^2}(\theta^*,0)L + H_4(A_L,i))e^{\theta^* S_L} \pi_{\theta^*,0}(A_L)|A_0 = i). \\
&\quad - 3E((\kappa(\theta^*,0)L + H_2(A_L,i))^2 e^{\theta^* S_L} \pi_{\theta^*,0}(A_L)|A_0 = i).
\end{aligned} \tag{3.14}
$$

Dabei erfüllt $H_n(j,i)$ analog zu Lemma 3.4.3 die Eigenschaft

$$\max_{j,i} |H_n(j,i)| \leq C, \quad n \in \{1,...,4\}. \tag{3.15}$$

Es werden nun wiederum die folgenden Aussagen bewiesen:

(i) $E(Le^{\theta^* S_L} \pi_{\theta^*,0}(A_L)|A_0 = i) = O(y)$

(ii) $E((\hat{W}_L - u^*L)^2 e^{\theta^* S_L} \pi_{\theta^*,0}(A_L)|A_0 = i) = O(y)$

(iii) $\xi^2 := E(L^2 e^{\theta^* S_L} \pi_{\theta^*,0}(A_L)|A_0 = i) = O(y^2)$

(iv) $\zeta_0^2 := E((\hat{W}_L - u^*L)^4 e^{\theta^* S_L} \pi_{\theta^*,0}(A_L)|A_0 = i) = O(y^2)$

(v) $E((\hat{W}_L - u^*L)^4 |A_0 = i, I_1(y) = 1) = O(y^2)$

Die Aussagen (i) und (iii) wurden schon in Lemma 3.4.3 bewiesen.

zu (ii): $\pi_{\theta^*,0}(i)$ sowie S_L sind beschränkt. Also folgt mit (3.13), (3.15) und Behauptung (i) analog zum Beweis von Lemma 3.4.3 (Behauptung (ii)), dass gilt:

$$
\begin{aligned}
&E((\hat{W}_L - u^*L)^2 e^{\theta^* S_L} \pi_{\theta^*,0}(A_L)|A_0 = i) \\
&= E((\kappa(\theta^*,0)L + H_2(A_L,i))e^{\theta^* S_L} \pi_{\theta^*,0}(A_L)|A_0 = i) = O(y).
\end{aligned}
$$

zu (iv): Erneut wendet man die Cauchy-Schwarzsche Ungleichung an, so dass man folgende Aussagen erhält:

$$
\begin{aligned}
&E((\hat{W}_L - u^*L)^2 L e^{\theta^* S_L} \pi_{\theta^*,0}(A_L)|A_0 = i) \leq \zeta_0 \xi \\
&E((\hat{W}_L - u^*L) L e^{\theta^* S_L} \pi_{\theta^*,0}(A_L)|A_0 = i) \leq \sqrt{\zeta_0 \xi} \xi.
\end{aligned}
$$

Daraus und aus (3.14) lässt sich nun herleiten, dass

$$\begin{aligned}
\zeta_0^2 &= 6\kappa(\theta^*,0)E((\hat{W}_L - u^*L)^2 Le^{\theta^* S_L}\pi_{\theta^*,0}(A_L)|A_0 = i) \\
&\quad + 6E((\hat{W}_L - u^*L)^2 H_2(A_L,i)e^{\theta^* S_L}\pi_{\theta^*,0}(A_L)|A_0 = i) \\
&\quad + 4\frac{d\kappa}{dt}(\theta^*,0)E((\hat{W}_L - u^*L)Le^{\theta^* S_L}\pi_{\theta^*,0}(A_L)|A_0 = i) \\
&\quad + 4E((\hat{W}_L - u^*L)H_3(A_L,i)e^{\theta^* S_L}\pi_{\theta^*,0}(A_L)|A_0 = i) \\
&\quad + \frac{d^2\kappa}{dt^2}(\theta^*,0)E(Le^{\theta^* S_L}\pi_{\theta^*,0}(A_L)|A_0 = i). \\
&\quad + E(H_4(A_L,i)e^{\theta^* S_L}\pi_{\theta^*,0}(A_L)|A_0 = i). \\
&\quad - 3(\kappa(\theta^*,0))^2 E(L^2 e^{\theta^* S_L}\pi_{\theta^*,0}(A_L)|A_0 = i). \\
&\quad + 6\kappa(\theta^*,0)E(LH_2(A_L,i)^2 e^{\theta^* S_L}\pi_{\theta^*,0}(A_L)|A_0 = i). \\
&\quad - 3E((H_2(A_L,i))^2 e^{\theta^* S_L}\pi_{\theta^*,0}(A_L)|A_0 = i). \\
&\le 6\kappa(\theta^*,0)\zeta_0\xi + 4\frac{d\kappa}{dt}(\theta^*,0)\sqrt{\zeta_0\xi}\xi + \frac{d^2\kappa}{dt^2}(\theta^*,0)E(Le^{\theta^* S_L}\pi_{\theta^*,0}(A_L)|A_0 = i) \\
&\quad - 3(\kappa(\theta^*,0))^2\xi^2 + O(y) \\
&= O(y)\zeta_0 + O(y^2) \\
&\quad \text{nach (i), (ii), (iii), sowie (3.12) und (3.15) und da } W_L \text{ beschränkt ist.}
\end{aligned}$$

Die größte Wurzel aus dieser quadratischen Ungleichung ist $O(y)$, also da ζ_0 positiv ist, folgt $\zeta_0^2 = O(y^2)$.

zu (v): Da nach Lemma 3.4.2

$$0 < \delta \le P(I_1(y) = 1|A_0 = i)e^{\theta^* y}$$

für alle $i \in \Omega_+$ gilt, folgt mit (iv) unmittelbar, dass

$$E((\hat{W}_L - u^*L(y))^4 e^{\theta^*(S_L - y)}\pi_{\theta^*,0}(A_L)|I_1(y) = 1, A_0 = i) = O(y^2)$$

Da $e^{\theta^*(S_L-y)}\pi_{\theta^*,0}(A_L)$ unter $I_1(y) = 1$ beschränkt ist, erhält man

$$E((\hat{W}_L - u^*L)^4|I_1(y) = 1, A_0 = i) = O(y^2).$$

Zusammen mit (v) erhält man nun die gewünschte Aussage:

$$\begin{aligned}
&E\left(\left(\frac{\hat{W}_{L(y)}}{L(y)} - u^*\right)^4 \middle| I_1(y) = 1, A_0 = i\right) \\
&= E\left(\frac{1}{(L(y))^4}(\hat{W}_{L(y)} - u^*L)^4 \middle| I_1(y) = 1, A_0 = i\right) = O\left(\frac{1}{y^2}\right),
\end{aligned}$$

denn unter $I_1(y) = 1$ gilt $L(y) \ge \frac{y}{K}$, da $y \le S_L \le L(y)K$. □

Analog zum iid-Fall zeigt man nun zunächst die Aussagen der beiden Grenzwertsätze (siehe Theorem 3.3.1 und Theorem 3.3.3) für den Fall $y = n,\ n \in \mathbb{N},\ \nu \in \{1, ...j_n\}$, d.h. man zeigt, dass für jeden Wert von A_0 und für alle $\nu \in \{1, ..., j_n\}$ gilt:

(i)
$$\frac{\tau_\nu(n) - \kappa_\nu(n)}{n} \xrightarrow[n\to\infty]{} \frac{1}{\omega^*} \quad \text{fast sicher}$$

(ii)
$$\frac{\sum_{i=\kappa_\nu(n)+1}^{\tau_\nu(n)} U_i}{\tau_\nu(n) - \kappa_\nu(n)} \xrightarrow[n\to\infty]{} u^* \quad \text{fast sicher}$$

Dabei wählt man $j_n = [A \log n]$, wobei A eine feste Konstante ist mit $-A \log(1 - P(-n_0 < S_1, ..., S_{m_0-1} < 1 < S_{m_0} | A_0 = i)) > 1$.

Lemma 3.4.5. *Für jeden Wert von A_0 gilt:*

$$\lim_{n\to\infty} \max_{\nu=1,...,[A\log n]} \left| \frac{\tau_\nu(n) - \kappa_\nu(n)}{n} - \frac{1}{\omega^*} \right| = 0 \quad \textit{fast sicher.}$$

Beweis. Ω_+ und Ω_- (vergleiche (3.6)) besitzen die Darstellungen:

$$\Omega_+ = \{i : \text{für alle } y > 0 \text{ gilt: } P(I_\nu(y) = 1 | A_{K_\nu - 1} = i) > 0\},$$
$$\Omega_- = \{i : \text{es existiert ein } y_0 > 0, \text{ so dass } P(I_\nu(y_0) = 1 | A_{K_\nu - 1} = i) = 0\}.$$

Aus der Monotonie von $I_\nu(y)$ folgt, dass ein y_0 existiert, so dass für alle $y \geq y_0$ und für alle $\nu \in \mathbb{N}$ $P(I_\nu(y) = 1 | A_{K_\nu - 1} \in \Omega_-) = 0$ gilt (vergleiche Aussage nach (3.6)).
Falls also $I_\nu(y) = 1$ gilt, muss notwenigerweise auch $A_{K_\nu - 1} \in \Omega_+$ gelten.
Mit anderen Worten:

Es existiert ein y_0, so dass für alle $y \geq y_0$, für alle ν und für jeden Anfangzustand $I_\nu(y) = 1$ auch $A_{K_\nu - 1} = i \in \Omega_+$ impliziert. (3.16)

Aus Lemma 3.4.3 folgt insbesondere für $y = n$, dass

$$E\left(\left(\frac{L_\nu(n)}{n} - \frac{1}{\omega^*}\right)^4 \middle| I_\nu(n) = 1, A_{K_\nu - 1} \in \Omega_+\right) = O\left(\frac{1}{n^2}\right),$$

bzw. mit Hilfe der alternativen Bezeichnungen, dass

$$E\left(\left(\frac{\tau_\nu(n) - \kappa_\nu(n)}{n} - \frac{1}{\omega^*}\right)^4 \middle| A_{K_\nu - 1} \in \Omega_+\right) = O\left(\frac{1}{n^2}\right) \quad \text{gilt.} \tag{3.17}$$

Es gilt also für alle $\varepsilon > 0$:

$$
\begin{aligned}
& P\left(\left|\frac{\tau_\nu(n)-\kappa_\nu(n)}{n} - \frac{1}{\omega^*}\right| > \varepsilon \middle| A_0 = i\right) \\
\leq\ & \frac{E\left(\left(\frac{\tau_\nu(n)-\kappa_\nu(n)}{n} - \frac{1}{\omega^*}\right)^4 \middle| A_0 = i\right)}{\varepsilon^4} \quad \text{(nach Markov-Ungleichung)} \\
\leq\ & \frac{E\left(\left(\frac{\tau_\nu(n)-\kappa_\nu(n)}{n} - \frac{1}{\omega^*}\right)^4 \middle| A_{K_\nu-1} = i \in \Omega_+\right)}{\varepsilon^4} \quad \text{(nach (3.16) für genügend großes } n\text{)} \\
=\ & \frac{E\left(\left(\frac{\tau_\nu(n)-\kappa_\nu(n)}{n} - \frac{1}{\omega^*}\right)^4 \middle| A_{K_\nu-1} \in \Omega_+\right)}{\varepsilon^4 P(A_{K_\nu-1} = i | A_{K_\nu-1} \in \Omega_+)} \\
\leq\ & \frac{C}{n^2\varepsilon^4}\frac{1}{P(A_{K_\nu-1} = i | A_{K_\nu-1} \in \Omega_+)} =: \frac{C_\varepsilon}{n^2\tilde{p}_i}, \quad \text{(nach (3.17))}
\end{aligned}
$$

wobei $\tilde{p}_i = P(A_{K_\nu-1} = i | A_{K_\nu-1} \in \Omega_+)$. Somit ergibt sich für alle $\varepsilon > 0$:

$$
\sum_{n=1}^{\infty} \sum_{\nu}^{[A \log n]} P\left(\left|\frac{\tau_\nu(n)-\kappa_\nu(n)}{n} - \frac{1}{\omega^*}\right| > \varepsilon \middle| A_0 = i\right) \leq \frac{C_\varepsilon}{\tilde{p}_i} A \sum_{n=1}^{\infty} \frac{\log n}{n^2} < \infty.
$$

Damit ergibt sich analog zur Beweisführung im iid-Fall (siehe Lemma 2.3.6) mit dem Lemma von Borel-Cantelli, dass für alle $i \in \Omega$

$$
P\left(\lim_{n\to\infty} \max_{\nu=1,\dots,[A\log n]} \left|\frac{\tau_\nu(n)-\kappa_\nu(n)}{n} - \frac{1}{\omega^*}\right| = 0 \middle| A_0 = i\right) = 1 \quad \text{gilt.}
$$

□

Lemma 3.4.6. *Für jeden Wert von A_0 gilt:*

$$
\lim_{n\to\infty} \max_{\nu=1,\dots,[A\log n]} \left|\frac{\sum_{i=\kappa_\nu(n)}^{\tau_\nu(n)} U_i}{\tau_\nu(n)-\kappa_\nu(n)} - u^*\right| = 0 \quad \textit{fast sicher.}
$$

Beweis. Aus Lemma 3.4.4 folgt für $y = n$, dass

$$
E\left(\left(\frac{W_\nu(n)}{L_\nu(n)} - u^*\right)^4 \middle| I_\nu(n) = 1, A_{K_\nu-1} \in \Omega_+\right) = O\left(\frac{1}{n^2}\right),
$$

bzw. mit Hilfe der alternativen Bezeichnungen, dass

$$
E\left(\left(\frac{\sum_{i=\kappa_\nu(n)}^{\tau_\nu(n)} U_i}{\tau_\nu(n)-\kappa_\nu(n)} - u^*\right)^4 \middle| A_{K_\nu-1} \in \Omega_+\right) = O\left(\frac{1}{n^2}\right) \quad \text{gilt.} \tag{3.18}
$$

Analog zum vorhergehenden Beweis von Lemma 3.4.5 ergibt sich somit für alle $\varepsilon > 0$ nach Markov-Ungleichung sowie (3.16) und (3.18):

$$P\left(\left|\frac{\sum_{i=\kappa_\nu(n)}^{\tau_\nu(n)} U_i}{\tau_\nu(n)-\kappa_\nu(n)} - u^*\right| > \varepsilon \middle| A_0 = i\right) \leq \frac{C}{n^2\varepsilon^4}\frac{1}{P(A_{K_\nu-1} = i|A_{K_\nu-1} \in \Omega_+)} =: \frac{C_\varepsilon}{n^2\tilde{p}_i}$$

und somit:

$$\sum_{n=1}^{\infty}\sum_{\nu}^{[A\log n]} P\left(\left|\frac{\sum_{i=\kappa_\nu(n)}^{\tau_\nu(n)} U_i}{\tau_\nu(n)-\kappa_\nu(n)} - u^*\right| > \varepsilon \middle| A_0 = i\right) \leq \frac{C_\varepsilon}{\tilde{p}_i} A \sum_{n=1}^{\infty} \frac{\log n}{n^2} < \infty.$$

Nach dem Lemma von Borel-Cantelli gilt also (vergleiche iid-Fall: Lemma 2.3.7) für alle $i \in \Omega$

$$P\left(\lim_{n\to\infty}\max_{\nu=1,\dots,[A\log n]}\left|\frac{\sum_{i=\kappa_\nu(n)}^{\tau_\nu(n)} U_i}{\tau_\nu(n)-\kappa_\nu(n)} - u^*\right| = 0 \middle| A_0 = i\right) = 1.$$

□

Um nun die Aussagen der beiden Grenzwertsätze vom Fall $y = n,\ \nu \in \{1, ..., j_n\}$ auf den Fall auszuweiten, bei dem $y > 0$ beliebig wählbar, sowie $\nu = \min\{k \geq 1 : I_k(y) = 1\}$ ist, werden analog zum iid-Fall zwei weitere Lemmata benötigt, welche Aussagen über das Verhältnis von $\kappa_\nu(n)$ zu $\kappa_1(y)$ bzw. von $\tau_\nu(n)$ zu $\tau_1(y)$ treffen.

Lemma 3.4.7. *Für jeden Wert von A_0 stimmt die erste Exkursion zum Level $n+1$ für genügend großes n fast sicher mit einer der ersten $[A\log n]$ Exkursionen zum Level n überein, d.h. es gilt*

$$P(\kappa_1(n+1) = \kappa_\nu(n) \textit{ für ein } \nu \in \{1, ..., [A\log n]\}|A_0 = i) = 1 \quad \textit{für alle } i \in \Omega.$$

Beweis. Es existieren $m_0, n_0 \in \mathbb{N}$ (unabhängig von i), derart dass für alle $i \in \Omega$

$$P(-n_0 < S_1, S_2, ..., S_{m_0-1} < 1 < S_{m_0}|A_0 = i) =: a > 0$$

gilt, wobei a unabhängig von i ist, denn:

Da die Markov-Kette irreduzibel ist, existiert ein $m_1 \in \mathbb{N}$ und ein $n_0 > m_1 K$ (K derart, dass $|X_i| \leq K$), so dass

(i) $P(A_{m_1} = i_0|A_0 = i) = p_{ii_0}^{(m_1)} > 0$

(ii) $\min\{S_1, ..., S_{m_1}\} > -n_0 \quad$ (da $|S_i| \leq iK < n_0,\ 1 \leq i \leq m_1$)

gilt. i_0 soll hier den Anfangszustand eines Zyklus $\mathcal{C}$ der Länge k mit $P(\sum_{i=1}^m X_i \geq \delta, m = 1, ..., k,\ \mathcal{C}) > 0$ für geeignetes δ bezeichnen. Die Existenz eines derartigen Zyklus ist sichergestellt, da nach Voraussetzung $P(\sum_{i=1}^m X_i > 0, m = 1, ..., k,\ \mathcal{C}) > 0$ für einen endlichen Zyklus $\mathcal{C}$ gilt. Wird dieser Zyklus von S_{m_1} ausgehend $\frac{n_0+1}{\delta}$ mal durchlaufen, erreicht der Prozess also einen Wert größer $-n_0+\delta\frac{1+n_0}{\delta} = 1$ (da nach (ii): $S_{m_1} > -n_0$), d.h. mit positiver Wahrscheinlichkeit existiert ein $m_0 \in \mathbb{N}$ mit $S_{m_0} > 1$. Es gilt also:

$$\begin{aligned}
&P(-n_0 < S_1, ..., S_{m_0-1} < 1 < S_{m_0}|A_0 = i)\\
\geq\ & P(-n_0 < S_1, ..., S_{m_0-1} < 1 < S_{m_0}|A_0 = i, A_{m_1} = i_0)P(A_{m_1} = i_0|A_0 = i)\\
\geq\ & P(-n_0 < S_1, ..., S_{m_0-1} < 1 < S_{m_0},\ \mathcal{C} \text{ wird } \frac{1+n_0}{\delta}\text{-mal durchlaufen}|A_0 = i, A_{m_1} = i_0)\\
& P(A_{m_1} = i_0|A_0 = i)\\
=:\ & a > 0.
\end{aligned}$$

Sei nun analog zum iid-Fall $\mathcal{E}_n$ gegeben durch:

$\mathcal{E}_n$ ={für jeden Wert von A_0 erreicht keine der $j_n = [A\log n]$ Exkursionen, die durch $(\kappa_\nu(n), \tau_\nu(n), \sigma_\nu(n))_{\nu\in\{1,...,j_n\}}$ gegeben sind, ein Level $\geq n+1$}

Es gilt $P(\limsup_{n\to\infty} \mathcal{E}_n) = 0$, denn:

Wählt man A genügend groß mit $-A\log(1-a) := \gamma > 1$, so gilt:

$$\begin{aligned}
&P(\mathcal{E}_n)\\
\leq\ & \prod_{\nu=1}^{j_n} P(X_{\tau_\nu(n)+1} < 1, X_{\tau_\nu(n)+1} + X_{\tau_\nu(n)+2} < 1, ..., X_{\tau_\nu(n)+1} + ... + X_{\sigma_\nu(n)} < 1|A_0 = i)\\
& \text{(analog zum iid-Fall, siehe Beweis zu Lemma 2.3.8)}\\
\leq\ & \prod_{\nu=1}^{j_n} P(\{-n_0 < X_1 < ... < X_1 + ... + X_{m_0-1} < 1 < X_1 + ... + X_{m_0}\}^{\complement}|A_0 = i)\\
=\ & (1 - P(-n_0 < S_1, ..., S_{m_0-1} < 1 < S_{m_0}|A_0 = i))^{j_n}\\
=\ & (1-a)^{j_n} = (1-a)^{[A\log n]} \leq \frac{C}{n^\gamma} \quad \text{für geeignetes } C < \infty,
\end{aligned}$$

denn $(1-a)^{[A\log n]} \leq C(1-a)^{A\log n} = Cn^{A\log(1-a)} = \dfrac{C}{n^\gamma}$.

Damit folgt, da $\gamma > 1$ ist, dass

$$\sum_{n=1}^\infty P(\mathcal{E}_n) \leq \sum_{n=1}^\infty \frac{C}{n^\gamma} < \infty \quad \text{gilt.}$$

Also ergibt sich zusammen mit dem Lemma von Borel-Cantelli die gewünschte Aussage: $P(\limsup_{n\to\infty} \mathcal{E}_n) = 0$.

Für jeden Wert von A_0 erreicht also die erste Exkursion zum Level n für genügend großes n P-f.s. auch das Level $n+1$.

$\square$

Nach Lemma 3.4.7 gilt also auch, dass die erste Exkursion zum Level y mit $n \leq y \leq n+1$, für jeden Wert von A_0 fast sicher mit einer der ersten j_n Exkursionen zum Level n übereinstimmt, also die „Startpunkte“ fast sicher gleich sind:

$$P(\kappa_1(y) = \kappa_\nu(n) \text{ für ein } \nu \in \{1, ..., [A \log n]\} | A_0 = i) = 1 \quad \text{für } n \leq y \leq n+1.$$

Das folgende Lemma liefert nun eine Konvergenz für $\tau_\nu(n)$ gegen $\tau(y)$, die es einem analog zum iid-Fall ermöglicht, die Aussagen der beiden Grenzwertsätze vom Fall $y = n,\ \nu \in \{1, ..., j_n\}$ auf den Fall $y > 0,\ \nu = \min\{k \geq 1 : I_k(y) = 1\}$ zu verallgemeinern. Dazu sei

$$T_n^* = \sup_{n \leq y < n+1} \min_{\substack{1 \leq \nu \leq j_n \\ \tau_\nu(n) \leq \tau_1(y)}} (\tau_1(y) - \tau_\nu(n)).$$

Lemma 3.4.8. *Für jeden Wert von A_0 gilt:*

$$\frac{T_n^*}{n} \xrightarrow[n \to \infty]{} 0 \quad \textit{fast sicher.}$$

Beweis. Der Beweis verläuft völlig analog zum iid-Fall (vergleiche Lemma 2.3.9). Man betrachtet das Maximum von $(S_m)_{m \in \mathbb{N}}$, welches durch $M := \max_k S_k$ gegeben ist. Dieses ist fast sicher wohldefiniert, da nach Voraussetzung $E(X_1) < 0$ gilt. Ferner sei J der Zeitpunkt, bei dem M das erste Mal angenommen wird, also

$$J = \inf\{k \geq 1 : S_k = M\}.$$

Nach Ney und Nummelin (siehe [17], Theorem 5.3) existieren Konstanten c und b, so dass

$$P(J = k | A_0 = i) \leq P\left(\frac{S_k}{k} > 0 | A_0 = i\right) \leq ce^{-bk}$$

gilt, J weist also einen exponentiellen Zerfall auf. Damit gilt auch

$$P(J \geq k | A_0 = i) = \sum_{i=k}^{\infty} P(J = i | A_0 = i) \leq \sum_{i=k}^{\infty} ce^{-bk} = ce^{-bk} \sum_{i=k}^{\infty} e^{-bi} = e^{-bk} \tilde{C} \quad (3.19)$$

für ein geeignetes $\tilde{C}$.

Analog zum iid-Fall sei $\tilde{J}_\nu(n)$ definiert durch

$$\tilde{J}_\nu(n) = \inf\{j \geq 1 : S_{j+\tau_\nu(n)} - S_{\tau_\nu(n)} = \max_{\tau_\nu(n) \leq k} (S_k - S_{\tau_\nu(n)})\}.$$

Betrachtet man nun einen beliebigen Wert y mit $n \leq y \leq n+1$, so gilt nach Lemma 3.4.7 für jeden Wert von A_0 für genügend großes n, dass die erste Exkursion zum Level y mit einer der Exkursionen aus $(\kappa_\nu(n), \tau_\nu(n), \sigma_\nu(n))_{\nu \in \{1,\dots,j_n\}}$ übereinstimmt. Es folgt also genau wie im Beweis zu Lemma 2.3.9, dass

$$T_n^* = \sup_{n \leq y < n+1} \min_{\substack{1 \leq \nu \leq j_n \\ \tau_\nu(n) \leq \tau_1(y)}} (\tau_1(y) - \tau_\nu(n)) \leq \max\{\tilde{J}_1(n), \tilde{J}_2(n), \dots, \tilde{J}_{j_n}(n)\}$$

gilt. Damit ergibt sich völlig analog zum iid-Fall für alle $\varepsilon > 0$:

$$\begin{aligned}
\sum_{n=1}^{\infty} P\left(\frac{T_n^*}{n} > \varepsilon | A_0 = i\right) &\leq \sum_{n=1}^{\infty} P(\tilde{J}_\nu(n) > n\varepsilon \text{ für ein } \nu \in \{1, \dots, j_n\} | A_0 = i) \\
&= \sum_{n=1}^{\infty} \sum_{\nu=1}^{j_n} P(J > n\varepsilon | A_0 = i) \\
&\leq \tilde{C} \sum_{n=1}^{\infty} A \log(n) e^{-bn\varepsilon} < \infty \quad \text{nach (3.19).}
\end{aligned}$$

Zusammen mit dem Lemma von Borel-Cantelli folgt die gewünsche Aussage. □

Damit stehen nun zum iid-Fall völlig analoge Aussagen zur Verfügung, mit Hilfe derer man die beiden Grenzwertsätze beweisen kann.

Beweis von Theorem 3.3.1. Für jeden Wert von A_0 gilt:

$$\begin{aligned}
\lim_{y \to \infty} \frac{\tau_1(y) - \kappa_1(y)}{y} &= \lim_{y \to \infty} \frac{\tau_1(y) - \max_{1 \leq \nu \leq j_n, \tau_\nu(n) \leq \tau_1(y)} \kappa_\nu(n)}{y} \quad \text{(nach Lemma 3.4.7)} \\
&= \lim_{y \to \infty} \min_{\substack{1 \leq \nu \leq j_n \\ \tau_\nu(n) \leq \tau_1(y)}} \left(\left(\frac{\tau_1(y) - \tau_\nu(n)}{n} \right) + \left(\frac{\tau_\nu(n) - \kappa_\nu(n)}{n} \right) \right) \\
&= \frac{1}{\omega^*} \quad \text{fast sicher (nach Lemma 3.4.5 und Lemma 3.4.8)}
\end{aligned}$$

Also erhält man schließlich unter Ausnutzung der Tatsache $\tau_1(y) - \kappa_1(y) = L_\nu(y)$, falls $\nu = \min\{k \geq 1 : I_k(y) = 1\}$, die gewünschte Aussage. □

Beweis von Theorem 3.3.3.

$$\text{Sei} \quad W_n^* := \sup_{n \leq y < n+1} \min_{1 \leq \nu \leq j_n} \left| \sum_{i=\kappa_1(y)+1}^{\tau_1(y)} U_i - \sum_{i=\kappa_\nu(n)+1}^{\tau_\nu(n)} U_i \right|.$$

Dann gilt für jeden Wert von A_0

$$\frac{W_n^*}{n} \leq \sup_{n \leq y < n+1} \min_{\substack{1 \leq \nu \leq j_n \\ \tau_\nu(n) \leq \tau(y)}} \left(\max |U_i| \left| \frac{\tau_1(y) - \tau_\nu(y)}{n} \right| \right) = \max |U_i| \frac{T_n^*}{n} \underset{n \to \infty}{\longrightarrow} 0$$

fast sicher nach Lemma 3.4.8.

Es gilt also für jeden Wert von A_0:

$$\begin{aligned}
&\lim_{y \to \infty} \left(\frac{1}{\tau_1(y) - \kappa_1(y)} \sum_{i=\kappa_1(y)+1}^{\tau_1(y)} U_i \right) \\
&= \lim_{y \to \infty} \min_{1 \leq \nu \leq j_n} \left(\frac{1}{\tau_\nu(n) - \kappa_\nu(n)} \sum_{i=\kappa_\nu(n)+1}^{\tau_\nu(n)} U_i + \frac{1}{\tau_\nu(n) - \kappa_\nu(n)} \left(\sum_{i=\kappa_1(y)+1}^{\tau_1(y)} U_i - \sum_{i=\kappa_\nu(n)+1}^{\tau_\nu(n)} U_i \right) \right) \\
&\quad \left(\text{da nach Theorem 3.3.1} \quad \lim_{y \to \infty} \frac{\tau_1(y) - \kappa_1(y)}{y} = \lim_{n \to \infty} \frac{\tau_\nu(n) - \kappa_\nu(n)}{n} \quad \text{gilt}\right) \\
&= \lim_{y \to \infty} \min_{1 \leq \nu \leq j_n} \left(\frac{1}{\tau_\nu(n) - \kappa_\nu(n)} \sum_{i=\kappa_\nu(n)+1}^{\tau_\nu(n)} U_i + \frac{1}{\frac{\tau_\nu(n) - \kappa_\nu(n)}{n}} \left(\frac{\sum_{i=\kappa_1(y)+1}^{\tau_1(y)} U_i - \sum_{i=\kappa_\nu(n)+1}^{\tau_\nu(n)} U_i}{n} \right) \right) \\
&= u^*
\end{aligned}$$

fast sicher nach Lemma 3.4.6, nach Theorem 3.3.1 ($\frac{\tau_\nu(n) - \kappa_\nu(n)}{n} \to \frac{1}{\omega^*}$ f.s.) und da $\frac{W_n^*}{n} \to 0$ (s.o.).

Da unter $\nu = \min \{k \geq 1 : I_k(y) = 1\}$

$$\frac{\sum_{\kappa_1(y)+1}^{\tau_1(y)} U_i}{\tau_1(y) - \kappa_1(y)} = \frac{L_\nu(y)}{W_\nu(y)}$$

gilt, erhält man somit die gewünschte Aussage. □

3.5 Ein zentraler Grenzwertsatz

Wie auch schon im iid-Fall (siehe Abschnitt 2.4) sollen nun Aussagen über hochscorige Sequenzsegmente in Form eines zentralen Grenzwertsatzes getroffen werden. Konkrete Aussagen bezüglich Länge und Zusammensetzung von Sequenzen mit hohem Score sind in den Korollaren (siehe Korollar 3.5.2 - Korollar 3.5.4) zu finden.

Die Notationen und Voraussetzungen seien wie bisher (siehe Abschnitt 3.1) mit dem Unterschied, dass die (U_i) hier Zufallsvektoren sein sollen.

Ferner setzen wir zusätzlich voraus, dass X_1 unter $A_0 = i$ und $A_1 = j$ entweder eine Gitterverteilung der Spanne d besitzt oder eine stetige Zufallsgröße ist.

Des Weiteren seien für alle Paare (i, j) die strengen Ungleichungen

$$P(X_1 > 0 | A_0 = i, A_1 = j) > 0 \quad \text{und} \quad P(X_1 < 0 | A_0 = i, A_1 = j) > 0$$

erfüllt und es gelte $p_{ij} > 0$.

Gegeben seien also beschränkte und unter $A_1, A_2, ...$ unabhängige Zufallsvektoren $\boldsymbol{U_1}, \boldsymbol{U_2}, ...$ Die Verteilung von $\boldsymbol{U_n} = (U_{n1}, ..., U_{nm})$ hänge von nur X_n, sowie von den Werten von A_{n-1} und A_n ab.
(Anmerkung: Um Verwechslungen zu vermeiden, werden im Folgenden Vektoren fett gedruckt.)

$\rho(\theta, \boldsymbol{t})$ sei der Spektralradius der Matrix

$$P_{\theta,t} := (p_{ij} E(e^{\theta X_1 + \langle \boldsymbol{t}, \boldsymbol{U_1} \rangle} | A_0 = i, A_1 = j))_{i,j \in \{1,...,r\}},$$

$\pi_{\boldsymbol{\theta,t}} = (\pi_{\theta,t}(1), ..., \pi_{\theta,t}(r)) \gg 0$ und $\psi_{\boldsymbol{\theta,t}} \gg 0$ seien die zugehörigen Rechts- bzw. Linkseigenvektoren, welche man ohne Einschränkung normiert, d.h. es gelte $\langle \pi_{\boldsymbol{\theta,t}}, \psi_{\boldsymbol{\theta,t}} \rangle = 1$, sowie $\langle \pi_{\boldsymbol{\theta,t}}, \boldsymbol{e} \rangle = 1$ (die Existenz derartiger Eigenvektoren ist nach dem Perron-Frobenius-Theorem (siehe Theorem 3.2.1) sichergestellt). Man definiert ferner $\zeta(\theta, \boldsymbol{t})$ durch

$$\zeta(\theta, \boldsymbol{t}) = \log \rho(\theta, \boldsymbol{t}).$$

θ^* sei wie bisher die eindeutig bestimmte positive Lösung von $\rho(\theta, 0) = 1$ bzw. von $\zeta(\theta, 0) = 0$ (zur Existenz und Eindeutigkeit sei auf Anmerkung 2.2.2 verwiesen) und es sei

$$\begin{aligned}
\omega^* &= \frac{d\zeta}{d\theta}(\theta^*, 0) = \sum_{i,j=1}^{r} p_{ij} E(X_1 e^{\theta^* X_1} | A_0 = i, A_1 = j) \pi_{\theta^*,0}(j) \psi_{\theta^*,0}(i) > 0, \quad \text{sowie} \\
\boldsymbol{u^*} &= (u_1^*, ..., u_m^*) = \left(\frac{d\zeta}{dt_1}(\theta^*, 0), ..., \frac{d\zeta}{dt_m}(\theta^*, 0) \right) \qquad (3.20) \\
&= \sum_{i,j=1}^{r} p_{ij} E(\boldsymbol{U_1} e^{\theta^* X_1} | A_0 = i, A_1 = j) \pi_{\theta^*,0}(j) \psi_{\theta^*,0}(i) \quad \text{wobei}
\end{aligned}$$

$$\begin{aligned}
&E(\boldsymbol{U_1} e^{\theta^* X_1} | A_0 = i, A_1 = j) := \\
&\qquad (E(U_{11} e^{\theta^* X_1} | A_0 = i, A_1 = j), ..., E(U_{1m} e^{\theta^* X_1} | A_0 = i, A_1 = j)).
\end{aligned}$$

(ω^* und $\boldsymbol{u^*}$ besitzen dabei diese Summendarstellungen, da $\frac{d\zeta}{d\theta}(\theta^*, 0) = \frac{\frac{d\rho}{d\theta}(\theta^*,0)}{\rho(\theta^*,0)} = \frac{d\rho}{d\theta}(\theta^*, 0)$ und da nach Anmerkung 3.3.2

$$\frac{d\rho}{d\theta}(\theta^*, 0) = \sum_{ij} p_{ij} E(X_1 e^{\theta^* X_1} | A_0 = i, A_1 = j) \pi_{\theta^*,0}(j) \psi_{\theta^*,0}(i)$$

gilt. Völlig analog zum Beweis dieser Anmerkung lässt sich zeigen, dass auch

$$\frac{d\rho}{dt_k}(\theta^*, 0) = \sum_{ij} p_{ij} E(U_{1k} e^{\theta^* X_1} | A_0 = i, A_1 = j) \pi_{\theta^*,0}(j) \psi_{\theta^*,0}(i) \quad \text{gilt.)}$$

Mit diesen Notationen gilt nun folgender zentraler Grenzwertsatz:

Theorem 3.5.1. *Es sei ν der erste Index mit $I_\nu(y) = 1$, d.h. $\nu = \min\{k \geq 1 : I_k(y) = 1\}$. Dann gilt für jeden Wert von A_0 für $y \to \infty$:*

$$\sqrt{\frac{\omega^*}{y}} \sum_{k=K_{\nu-1}+1}^{T_\nu(y)} (\boldsymbol{U_k} - \boldsymbol{u^*}) \underset{(V)}{\longrightarrow} \mathcal{N}(\boldsymbol{0}, \Sigma) \quad mit \quad \Sigma = (\sigma_{kl})_{k,l \in \{1,\dots,m\}},$$

$$\sigma_{kl} = \frac{d^2\zeta}{dt_k dt_l}(\theta^*, 0) = \sum_{i,j=1}^{r} p_{ij} E((\boldsymbol{U_1} - \boldsymbol{u^*})_k (\boldsymbol{U_1} - \boldsymbol{u^*})_l e^{\theta^* X_1} | A_0 = i, A_1 = j) \pi_{\theta^*,0}(j) \psi_{\theta^*,0}(i).$$

Analog zum iid-Fall lassen sich daraus Aussagen bezüglich der Länge $L_\nu(y)$ von Sequenstücken mit hohem Score (siehe Korollar 3.5.2), über die Zusammensetzung von hochscorigen Sequenzstücken (siehe Korollar 3.5.3) und bezüglich der Zufallsgröße $\frac{1}{L_\nu(y)} \sum_{k=\kappa_1(y)}^{\tau_1(y)} (\boldsymbol{U_k} - \boldsymbol{u^*})$ (alternative Bezeichnungen siehe (3.1)) treffen.
Die in den Beweisen zu den Korollaren verwendete Aussage $\frac{L_\nu(y)}{y} \underset{y \to \infty}{\longrightarrow} \frac{1}{\omega^*}$ P-fast sicher, falls $\nu = \min\{k \geq 1 : I_k(y) = 1\}$, gilt nach Theorem 3.3.1 auch im Markov-Fall für jeden Wert von A_0. Da die Beweise von Korollar 3.5.2 und 3.5.4 somit völlig analog zum iid-Fall verlaufen, wird an dieser Stelle lediglich Korollar 3.5.3 bewiesen. Für die beiden anderen Beweise sei auf Abschnitt 2.4 verwiesen.

Korollar 3.5.2. *Sei $\nu = \min\{k \geq 1 : I_k(y) = 1\}$, dann gilt für jeden Wert von A_0 für $y \to \infty$:*

$$\sqrt{\frac{\omega^*}{y}} (y - \omega^* L_\nu(y)) \underset{(V)}{\longrightarrow} \mathcal{N}(0, \sigma^2),$$

wobei $\sigma^2 = \sum_{i,j=1}^r p_{ij} E((X_1 - \omega^*)^2 e^{\theta^* X_1} | A_0 = i, A_1 = j) \pi_{\theta^*,0}(j) \psi_{\theta^*,0}(i)$.

Korollar 3.5.3. *Es sei A eine Borelmenge aus dem Wertebereich von X_1 und $\nu = \min\{k \geq 1 : I_k(y) = 1\}$. Dann gilt für die empirische Häufigkeitsverteilung $\mu_\nu(A; y) = \frac{\sum_{k=K_{\nu-1}+1}^{T_\nu(y)} \mathbb{I}_A(X_k)}{L_\nu(y)}$ des Eintretens von A für $y \to \infty$:*

$$\sqrt{L_\nu(y)} (\mu_\nu(A; y) - \mu^*(A)) \underset{(V)}{\longrightarrow} \mathcal{N}(0, c^*),$$

für jeden Wert von A_0, wobei $c^* = \mu^*(A) - \mu^*(A)^2$ und

$$\mu^*(A) = \sum_{i,j=1}^{r} p_{ij} E(\mathbb{I}_A(X_1) e^{\theta^* X_1} | A_0 = i, A_1 = j) \pi_{\theta^*,0}(j) \psi_{\theta^*,0}(i).$$

Beweis. Analog zum Beweis von Korollar 2.4.3 ergibt sich mit $\boldsymbol{U_k} = \mathbb{I}_A(X_k)$ nach Theorem 3.5.1, dass $\mu_\nu(A;y) = \frac{\sum_{k=K_{\nu-1}+1}^{T_\nu(y)} \boldsymbol{U_k}}{L_\nu(y)}$, $\mu^*(A) = \boldsymbol{u^*}$ und somit

$$\sqrt{L_\nu(y)}(\mu_\nu(A;y) - \mu^*(A)) \underset{(V)}{\longrightarrow} \mathcal{N}(0, c^*).$$

Es bleibt lediglich zu zeigen, dass $c^* = \mu^*(A) - \mu^*(A)^2$ gilt:

$$\begin{aligned} c^* &= \sum_{ij} p_{ij} E((\mathbb{I}_A - \mu^*(A))^2 e^{\theta^* X_1} | A_0 = i, A_1 = j) \pi_{\theta^*,0}(j) \psi_{\theta^*,0}(i) \\ &= \sum_{ij} p_{ij} \Big(E(\mathbb{I}_A^2 e^{\theta^* X_1} | A_0 = i, A_1 = j) - 2\mu^*(A) E(\mathbb{I}_A e^{\theta^* X_1} | A_0 = i, A_1 = j) \\ &\quad + \mu^*(A)^2 E(e^{\theta^* X_1} | A_0 = i, A_1 = j) \Big) \pi_{\theta^*,0}(j) \psi_{\theta^*,0}(i) \\ &= \mu^*(A) - 2\mu^*(A)^2 + \mu^*(A) \sum_{ij} p_{ij} E(e^{\theta^* X_1} | A_0 = i, A_1 = j) \pi_{\theta^*,0}(j) \psi_{\theta^*,0}(i) \\ &= \mu^*(A) - \mu^*(A)^2, \end{aligned}$$

nach Definition von $\mu^*(A)$ und da nach Proposition 3.2.2 gilt:

$$1 = \rho(\theta^*, 0) = \sum_j p_{ij} E(e^{\theta^* X_1} | A_0 = i, A_1 = j) \frac{\pi_{\theta^*,0}(j)}{\pi_{\theta^*,0}(i)},$$

was aufgrund der Normiertheit der Eigenvektoren ($\sum_{ij} \pi_{\theta^*,0}(i) \psi_{\theta^*,0}(i) = 1$) äquivalent ist zu

$$1 = \sum_{ij} p_{ij} E(e^{\theta^* X_1} | A_0 = i, A_1 = j) \pi_{\theta^*,0}(j) \psi_{\theta^*,0}(i).$$

□

Korollar 3.5.4. *Für jeden Wert von A_0 gilt für $y \to \infty$:*

$$\frac{1}{\sqrt{L_\nu(y)}} \sum_{k=\kappa_1(y)+1}^{\tau_1(y)} (\boldsymbol{U_k} - \boldsymbol{u^*}) \underset{(V)}{\longrightarrow} \mathcal{N}(\boldsymbol{0}, \Sigma), \quad \textit{wobei} \quad \Sigma = (\sigma_{kl})_{k,l \in \{1,\dots,m\}} \ \textit{und}$$

$\sigma_{kl} = \sum_{i,j=1}^r p_{ij} E((\boldsymbol{U_1} - \boldsymbol{u^*})_k (\boldsymbol{U_1} - \boldsymbol{u^*})_l e^{\theta^* X_1} | A_0 = i, A_1 = j) \pi_{\theta^*,0}(j) \psi_{\theta^*,0}(i).$

3.6 Beweis des zentralen Grenzwertsatzes

Der Beweis verläuft weitgehend analog zum iid-Fall (siehe Kapitel 2.5). Die Aussage des Theorems 3.5.1 wird also zunächst für den eindimensionalen Fall mit $u^* = 0$ gezeigt. Dazu benötigen wir sechs Lemmata. Zunächst sei an dieser Stelle ein Analogon zu Lemma 2.5.1 angeführt.

Lemma 3.6.1. *Es existiert ein $\theta(t)$ nahe bei θ^* mit $\zeta(\theta(t),t)=0$ und unter der Voraussetzung $u^* = \frac{d\zeta}{dt}(\theta^*,0) = \sum_{ij} p_{ij} E(U_1 e^{\theta^* X_1}|A_0=i, A_1=j)\pi_{\theta^*,0}(j)\psi_{\theta^*,0}(i) = 0$ gilt für dieses $\theta(t)$:*

$$\theta(t) = \theta^* - \frac{1}{2}\frac{v^*}{\omega^*}t^2 + O(t^3),$$

$$\textit{wobei} \quad v^* = \frac{d^2\zeta}{dt^2}(\theta^*,0) = \sum_{ij} p_{ij} E(U_1^2 e^{\theta^* X_1}|A_0=i, A_1=j)\pi_{\theta^*,0}(j)\psi_{\theta^*,0}(i) \quad \textit{und}$$

$$\omega^* = \frac{d\zeta}{d\theta}(\theta^*,0) = \sum_{ij} p_{ij} E(X_1 e^{\theta^* X_1}|A_0=i, A_1=j)\pi_{\theta^*,0}(j)\psi_{\theta^*,0}(i).$$

Beweis. Da die Funktion $\zeta(\theta,t) = \log\rho(\theta,t)$ für genügend kleines t analytisch in t und da $\zeta(\theta^*,0)=0$ ist, lässt sich dieses Lemma völlig analog zum iid-Fall beweisen.
Wir wollen an dieser Stelle nur zeigen, dass

$$v^* = \frac{d\zeta^2}{dt^2}(\theta^*,0) = \sum_{ij} p_{ij} E(U_1^2 e^{\theta^* X_1}|A_0=i, A_1=j)\pi_{\theta^*,0}(j)\psi_{\theta^*,0}(i)$$

gilt (zu den Darstellungen von u^* und ω^* siehe (3.20)). Nach Proposition 3.2.2 besitzt ρ folgende Darstellungen:

$$\rho(\theta,t) = \sum_{j=1}^{r} p_{ij} E(e^{\theta X_1 + tU_1}|A_0=i, A_1=j)\frac{\pi_{\theta,t}(j)}{\pi_{\theta,t}(i)}\ , \quad 1 \le i \le n \tag{3.21}$$

$$\rho(\theta,t) = \sum_{i=1}^{r} p_{ij} E(e^{\theta X_1 + tU_1}|A_0=i, A_1=j)\frac{\psi_{\theta,t}(i)}{\psi_{\theta,t}(j)}\ , \quad 1 \le j \le n, \tag{3.22}$$

Da $\rho(\theta^*,0)=1$ und $\frac{d\rho(\theta^*,0)}{dt} = u^* = 0$, gilt

$$\frac{d^2\zeta}{dt^2}(\theta^*,0) = \frac{d^2\log\rho}{dt^2}(\theta^*,0) = \frac{\frac{d^2\rho}{dt}(\theta^*,0)\rho(\theta^*,0) - (\frac{d\rho(\theta^*,0)}{dt})^2}{(\rho(\theta^*,0))^2} = \frac{d^2\rho}{dt^2}(\theta^*,0).$$

Um die Ableitung $\frac{d^2\rho}{dt^2}(\theta,t)$ zu bestimmen, erweitern wir den Ausdruck (3.21) auf beiden Seiten mit $\pi_{\theta,t}(i)$, differenzieren ihn zweimal nach θ, erweitern ihn mit $\psi_{\theta,t}(i)$ und summieren anschließend über i auf, so dass wir die Gleichheit

$$\begin{aligned}
&\frac{d^2\rho}{dt^2}(\theta,t)\sum_i \pi_{\theta,t}(i)\psi_{\theta,t}(i) + 2\frac{d\rho}{dt}(\theta,t)\sum_i \frac{d\pi_{\theta,t}}{dt}(i)\psi_{\theta,t}(i) + \rho(\theta,t)\sum_i \frac{d^2\pi_{\theta,t}}{dt^2}(i)\psi_{\theta,t}(i)\\
= &\sum_{ij} p_{ij} E(U_1^2 e^{\theta X_1 + tU_1}|A_0=i, A_1=j)\pi_{\theta,t}(j)\psi_{\theta,t}(i)\\
&+ 2\sum_{ij} p_{ij} E(U_1 e^{\theta X_1 + tU_1}|A_0=i, A_1=j)\frac{d\pi_{\theta,t}}{dt}(j)\psi_{\theta,t}(i)\\
&+ \sum_{ij} p_{ij} E(e^{\theta X_1 + tU_1}|A_0=i, A_1=j)\frac{d^2\pi_{\theta,t}}{dt^2}(j)\psi_{\theta,t}(i)
\end{aligned}$$

erhalten. Nun gilt aufgrund der Normiertheit der Eigenvektoren: $\sum_i \pi_{\theta,t}(i)\psi_{\theta,t}(i) = 1$, ferner gilt, dass

$$2\sum_j \frac{d\pi_{\theta,t}}{dt}(j)\sum_i p_{ij}E(U_1 e^{\theta X_1+tU_1}|A_0=i,A_1=j)\psi_{\theta,t}(i) = 2\frac{d\rho}{dt}(\theta,t)\sum_j \frac{d\pi_{\theta,t}}{dt}(j)\psi_{\theta,t}(j),$$

sowie nach (3.22), dass

$$\sum_j \frac{d^2\pi_{\theta,t}}{dt^2}(j)\sum_i p_{ij}E(e^{\theta X_1+tU_1}|A_0=i,A_1=j)\psi_{\theta,t}(i) = \rho(\theta,t)\sum_j \frac{d^2\pi_{\theta,t}}{dt^2}(j)\psi_{\theta,t}(j).$$

Also erhält man

$$\frac{d^2\rho}{dt^2}(\theta,t) = \sum_{ij} p_{ij}E(U_1^2 e^{\theta X_1+tU_1}|A_0=i,A_1=j)\pi_{\theta,t}(j)\psi_{\theta,t}(i).$$

□

Wie im iid-Fall betrachtet man im weiteren Verlauf $\theta(t_y)$ mit $t_y = t\sqrt{\frac{\omega^*}{y}}$.
Mit Lemma 3.6.1 lässt sich folgende Aussage beweisen.

Lemma 3.6.2. *Unter der Voraussetzung $u^* = 0$ gilt für alle $i \in \Omega$:*

$$\lim_{y\to\infty} E((e^{\theta(t_y)S_L+t_yW_L} - e^{\theta^* S_L})\pi_{\theta^*,0}(A_L)|I(y)=1,A_0=i)P(I(y)=1|A_0=i) = 0.$$

Beweis. Da $\pi_{\theta,t}$ in θ und t analytisch ist, gilt

$$\sup_i \left|\log\frac{\pi_{\theta(t_y),t_y}(i)}{\pi_{\theta^*,0}(i)}\right| \le O(|\theta(t_y)-\theta^*|) = O\left(\left|\theta\left(t\sqrt{\frac{\omega^*}{y}}\right)-\theta^*\right|\right) \le \frac{C}{\sqrt{y}} \underset{y\to\infty}{\longrightarrow} 0,$$

somit erhalten wir für alle i:

$$\pi_{\theta(t_y),t_y}(i) \underset{y\to\infty}{\longrightarrow} \pi_{\theta^*,0}(i). \tag{3.23}$$

Demnach gilt also

$$\begin{aligned}
&\lim_{y\to\infty} E((e^{\theta(t_y)S_L+t_yW_L} - e^{\theta^* S_L})\pi_{\theta^*,0}(A_L)|I(y)=1,A_0=i)P(I(y)=1|A_0=i)\\
&= \pi_{\theta^*,0}(i)\lim_{y\to\infty}\left(E\left(e^{\theta(t_y)S_L+t_yW_L}\frac{\pi_{\theta^*,0}(A_L)}{\pi_{\theta(t^*),0}(A_0)} - e^{\theta^* S_L}\frac{\pi_{\theta^*,0}(A_L)}{\pi_{\theta^*,0}(A_0)}\middle|I(y)=1,A_0=i\right)\right.\\
&\qquad\left.\cdot P(I(y)=1|A_0=i)\right)\\
&=\pi_{\theta^*,0}(i)\lim_{y\to\infty}\left(E\left(e^{\theta(t_y)S_L+t_yW_L}\frac{\pi_{\theta(t_y),t_y}(A_L)}{\pi_{\theta(t_y),t_y}(A_0)} - e^{\theta^* S_L}\frac{\pi_{\theta^*,0}(A_L)}{\pi_{\theta^*,0}(A_0)}\middle|I(y)=1,A_0=i\right)\right.\\
&\qquad\left.\cdot P(I(y)=1|A_0=i)\right).
\end{aligned}$$

Damit genügt es zu zeigen, dass

$$\lim_{y\to\infty}\Big(E\left(e^{\theta(t_y)S_L+t_yW_L}\frac{\pi_{\theta(t_y),t_y}(A_L)}{\pi_{\theta(t_y),t_y}(A_0)}-e^{\theta^*S_L}\frac{\pi_{\theta^*,0}(A_L)}{\pi_{\theta^*,0}(A_0)}\Big|I(y)=1,A_0=i\right)$$
$$\cdot P(I(y)=1|A_0=i)\Big)=0.$$

Dazu wenden wir Lemma 3.4.1 („Eigenschaften der Markov-Ketten-Erweiterung des Waldmartingals") an: für genügend kleines $|t|$ und $\theta(t)$ nahe bei θ^* gilt:

$$E\left(e^{\theta(t)S_L+tW_L-L\zeta(\theta(t),t)}\frac{\pi_{\theta(t),t}(A_L)}{\pi_{\theta(t),t}(A_0)}\Big|A_0=i\right)=1,$$

insbesondere gilt für $t=0$ und $\theta(0)=\theta^*$:

$$E\left(e^{\theta^*S_L}\frac{\pi_{\theta^*,0}(A_L)}{\pi_{\theta^*,0}(A_0)}\Big|A_0=i\right)=1.$$

Ferner folgt mit Lemma 3.6.1, dass ein $\theta(t)$ in der Nähe von θ^* existiert mit $\zeta(\theta(t),t)=0$. Dieses $\theta(t)$ wählen wir nun, wobei wir $t_y=t\sqrt{\frac{\omega^*}{y}}$ einsetzen. Wenn wir y genügend groß wählen, so können wir Lemma 3.4.1 anwenden (denn dann liegt $(\theta(t_y),t_y)$ nahe bei $(\theta^*,0)$). Wir erhalten also, da $\zeta(\theta(t_y),t_y)=0$:

$$\begin{aligned}0=E&\left(e^{\theta(t_y)S_L+t_yW_L}\frac{\pi_{\theta(t_y),t_y}(A_L)}{\pi_{\theta(t_y),t_y}(A_0)}-e^{\theta^*S_L}\frac{\pi_{\theta^*,0}(A_L)}{\pi_{\theta^*,0}(A_0)}\Big|I(y)=1,A_0=i\right)\\&\cdot P(I(y)=1|A_0=i)\\+E&\left(e^{\theta(t_y)S_L+t_yW_L}\frac{\pi_{\theta(t_y),t_y}(A_L)}{\pi_{\theta(t_y),t_y}(A_0)}-e^{\theta^*S_L}\frac{\pi_{\theta^*,0}(A_L)}{\pi_{\theta^*,0}(A_0)}\Big|I(y)=0,A_0=i\right)\\&\cdot P(I(y)=0|A_0=i).\end{aligned}$$

Wie auch im iid-Fall genügt es also zu zeigen, dass

$$\Big(\lim_{y\to\infty}E\left(e^{\theta(t_y)S_L+t_yW_L}\frac{\pi_{\theta(t_y),t_y}(A_L)}{\pi_{\theta(t_y),t_y}(A_0)}-e^{\theta^*S_L}\frac{\pi_{\theta^*,0}(A_L)}{\pi_{\theta^*,0}(A_0)}\Big|I(y)=0,A_0=i\right)$$
$$\cdot P(I(y)=0|A_0=i)\Big)=0.$$

Nun gilt nach (3.23), dass

$$\begin{aligned}&\lim_{y\to\infty}E\left(e^{\theta(t_y)S_L+t_yW_L}\frac{\pi_{\theta(t_y),t_y}(A_L)}{\pi_{\theta(t_y),t_y}(A_0)}-e^{\theta^*S_L}\frac{\pi_{\theta^*,0}(A_L)}{\pi_{\theta^*,0}(A_0)}\Big|I(y)=0,A_0=i\right)P(I(y)=0|A_0=i)\\&=\lim_{y\to\infty}E\left((e^{\theta(t_y)S_L+t_yW_L}-e^{\theta^*S_L})\frac{\pi_{\theta^*,0}(A_L)}{\pi_{\theta^*,0}(A_0)}\Big|I(y)=0,A_0=i\right)P(I(y)=0|A_0=i).\end{aligned}$$

Ebenso gilt, da $\pi_{\theta^*,0}(i)\leq 1$ für alle i, dass

$$\min_j \pi_{\theta^*,0}(j)\leq\frac{\pi_{\theta^*,0}(A_L)}{\pi_{\theta^*,0}(A_0)}\leq\frac{1}{\min_j\pi_{\theta^*,0}(j)},$$

somit bleibt lediglich zu zeigen, dass

$$\lim_{y\to\infty} E(e^{\theta(t_y)S_L+t_yW_L} - e^{\theta^* S_L}|I(y) = 0, A_0 = i)P(I(y) = 0|A_0 = i) = 0.$$

Dieser Beweis verläuft völlig analog zum iid-Fall, deshalb sei er an dieser Stelle abgekürzt und auf Lemma 2.5.2 verwiesen. Es sei nur angemerkt, dass die Abschätzung $P(L = n+1|I(y) = 0) \leq e^{-bn}$ auch im Markov-Fall gültig ist, denn nach Ney und Nummelin (siehe [17], Theorem 5.3) existieren Konstanten c und b, so dass

$$P(L = n+1|I(y) = 0, A_0 = i) \leq P(\frac{S_n}{n} > 0|A_0 = i) \leq ce^{-bk}.$$

□

Im Folgenden wollen wir eine analoge Aussage zu Lemma 2.5.3 für den Markov-Fall beweisen. An dieser Stelle tritt jedoch eine kleine Schwierigkeit auf, da die im iid-Fall verwendete Aussage von Lemma 2.3.3 ($e^{\theta^* y}P(I(y) = 1) \in [\delta, 1]$ für ein $\delta > 0$) im Markov-Fall nicht für beliebigen Anfangzustand i erfüllt ist. Nach Lemma 3.4.2 gilt nämlich ausschließlich für

$$i \in \Omega_+ = \{i \in \{1, ..., r\} : \text{für alle } y > 0 \text{ gilt } P(I_1(y) = 1|A_0 = i) > 0\},$$

dass ein C und ein $\delta > 0$ existieren, derart dass $P(I_1(y) = 1|A_0 = i)e^{\theta^* y} \in [\delta, C]$ (vergleiche (3.6)). Für

$$i \in \Omega_- = \{i \in \{1, ..., r\} : \text{es existiert ein } y_0 > 0, \text{ so dass } P(I_1(y_0) = 1|A_0 = i) = 0\}$$

gilt dagegen, dass ein y_0 existiert, so dass $P(I_1(y) = 1|A_0 = i) = 0$ für alle $y \geq y_0$ gilt. Um den Beweis im Weiteren nicht allzu sehr zu verkomplizieren, wurde bereits am Anfang des Abschnittes 3.5 vorausgesetzt, dass die strengen Ungleichungen $P(X_1 > 0|A_0 = i, A_1 = j) > 0$, $P(X_1 < 0|A_0 = i, A_1 = j) > 0$ und $p_{ij} > 0$ erfüllt sind. Unter diesen Voraussetzungen gilt nämlich $\Omega_- = \varnothing$, denn für alle $y > 0$ gilt:

$$\begin{aligned}
&P(I(y) = 1|A_0 = i) \\
=& \sum_{(i_1,...,i_{L(y)})} p_{ii_1}p_{i_1i_2}...p_{i_{L(y)-1}i_{L(y)}} P(I(y) = 1|A_0 = i, A_1 = i_1, ...A_{L(y)} = i_{L(y)}) \\
\geq& \Big(\sum_{(i_1,...,i_{L(y)})} p_{ii_1}p_{i_1i_2}...p_{i_{L(y)-1}i_{L(y)}} P(I(y) = 1|A_0 = i, ...A_{L(y)} = i_{L(y)}, X_1 > 0, ..., X_{L(y)} > 0) \\
&\cdot P(X_1 > 0|A_0 = i, A_1 = i_1)...P(X_{L(y)} > 0|A_{L(y)-1}i_{L(y)-1}, A_{L(y)} = i_{L(y)})\Big) > 0.
\end{aligned}$$

Unter den obigen Voraussetzungen sind wir nun in der Lage, das folgende Lemma zu beweisen.

Lemma 3.6.3. *Sei $u^* = 0$. Dann gilt für alle $i \in \Omega$ und für alle $t \in \mathbb{R}$:*

$$\lim_{y\to\infty} \frac{E(e^{t\sqrt{\frac{\omega^*}{y}}W_L} e^{\theta^*(S_L-y)}\pi_{\theta^*,0}(A_L)|I(y)=1, A_0=i)}{E(e^{\theta^*(S_L-y)}\pi_{\theta^*,0}(A_L)|I(y)=1, A_0=i)} = e^{\frac{1}{2}t^2v^*},$$

wobei $v^* = \frac{d^2\zeta}{dt^2}(\theta^*,0) = \sum_{ij} p_{ij} E(U_1^2 e^{\theta^* X_1}|A_0=i, A_1=j)\pi_{\theta^*,0}(j)\psi_{\theta^*,0}(i)$.

Beweis. Der Beweis verläuft analog zum iid-Fall (siehe Beweis zu Lemma 2.5.3), deshalb sei er hier etwas verkürzt dargestellt. Nach dem vorhergehenden Lemma gilt

$$\Big(E(e^{(\theta(t_y)-\theta^*)S_L} e^{\theta^*(S_L-y)} e^{t\sqrt{\frac{\omega^*}{y}}W_L}\pi_{\theta^*,0}(A_L)|I(y)=1, A_0=i)$$
$$- E(e^{\theta^*(S_L-y)}\pi_{\theta^*,0}(A_L)|I(y)=1, A_0=i)\Big)e^{\theta^* y}P(I(y)=1|A_0=i) \underset{y\to\infty}{\longrightarrow} 0.$$

Dividiert man durch $E(e^{\theta^*(S_L-y)}\pi_{\theta^*,0}(A_L)|I(y)=1, A_0=i) > 0$ und wendet die Aussage $e^{\theta^* y}P(I(y)=1|A_0=i) \in [\delta, C]$ für geeignete $\delta, C > 0$ aus Lemma 3.4.2 und der Eingangsüberlegung (s.o.) an, so ergibt sich

$$\frac{E(e^{(\theta(t_y)-\theta^*)S_L} e^{\theta^*(S_L-y)} e^{t\sqrt{\frac{\omega^*}{y}}W_L}\pi_{\theta^*,0}(A_L)|I(y)=1, A_0=i)}{E(e^{\theta^*(S_L-y)}\pi_{\theta^*,0}(A_L)|I(y)=1, A_0=i)} \underset{y\to\infty}{\longrightarrow} 1.$$

Analog zum iid-Fall erhält man ferner mit Hilfe von Lemma 3.6.1 folgende Darstellung von $\theta(t_y) - \theta^*$:

$$\theta(t_y) - \theta^* = -\frac{t^2}{2y}v^* + O\left(\frac{1}{y^{3/2}}\right).$$

Setzt man diesen Term ein, so erhält man aufgrund der Konvergenz

$$e^{(-\frac{t^2}{2y}v^* + O(y^{-3/2}))S_L} \underset{y\to\infty}{\longrightarrow} e^{-\frac{1}{2}t^2v^*}$$

schließlich die Aussage des Lemmas. □

Um zu der Aussage

$$\lim_{y\to\infty} E(e^{t\sqrt{\frac{\omega^*}{y}}\sum_{k=K_{\nu-1}+1}^{T_\nu(y)} U_k}|I_\nu(y)=1, A_0=i) = e^{\frac{1}{2}t^2v^*}$$

zu gelangen, benötigen wir nun ein Analogon für den Markov-Fall zur Proposition 2.5.4:

Proposition 3.6.4. *(siehe [14], Lemma 4.3 und Lemma 4.4)*
Es gelte $E(X_1) < 0$ und es existiere ein endlicher Zyklus $\mathcal{C} = \{A_0 = i_0, ..., A_k = i_k = i_0\}$, derart dass $P(\{\min_{m=1,...,k}\sum_{i=1}^m X_i > 0\} \cap \mathcal{C}) > 0$.
Falls X_1 unter $A_0 = i$ und $A_1 = j$ keine Gitterverteilung besitzt, so gilt

(i) $\lim\limits_{y\to\infty} P(\max\limits_{k\geq 0} S_k > y|A_0 = i)e^{\theta^* y} = e^*\pi_{\theta^*,0}(i)$ *für ein* e^* *mit* $0 < e^* < \infty$

(ii) $\lim\limits_{y\to\infty} P(\max\limits_{0\leq k\leq\sigma} S_k > y|A_0 = i)e^{\theta^* y} = (\pi_{\theta^*,0}(i) - E(e^{\theta^* S_\sigma}\pi_{\theta^*,0}(A_\sigma)|A_0 = i))e^*$ *für ein* e^* *mit* $0 < e^* < \infty$.

Falls X_1 *unter* $A_0 = i$ *und* $A_1 = j$ *eine Gitterverteilung der Spanne* d *besitzt, so gilt*

(i) $\lim\limits_{n\to\infty} P(\max\limits_{k\geq 0} S_k > nd|A_0 = i)e^{\theta^* nd} = e^*\pi_{\theta^*,0}(i)$ *für ein* e^* *mit* $0 < e^* < \infty$

(ii) $\lim\limits_{n\to\infty} P(\max\limits_{0\leq k\leq\sigma} S_k > nd|A_0 = i)e^{\theta^* nd} = (\pi_{\theta^*,0}(i) - E(e^{\theta^* S_\sigma}\pi_{\theta^*,0}(A_\sigma)|A_0 = i))e^*$ *für ein* e^* *mit* $0 < e^* < \infty$.

Beweis. zu (i): Die Aussage ergibt sich als eine Anwendung der Markov-Erneuerungstheorie (siehe z.B. [3], Kapitel 10).

zu (ii): Es gilt:

$$\begin{aligned}
P(\max_{k\geq 0} S_k > y|A_0 = i) &= P(\max_{k\geq 0} S_k > y, \max_{0\leq k\leq\sigma} S_k > y|A_0 = i) \\
&\quad + P(\max_{k\geq 0} S_k > y, \max_{0\leq k\leq\sigma} S_k \leq y|A_0 = i) \\
&= P(\max_{0\leq k\leq\sigma} S_k > y|A_0 = i) \\
&\quad + \sum_j \int_{-\infty}^0 P(\max_{k\geq 0} S_k > y - x|A_0 = j)dK_{i,j}^{(y)}(x),
\end{aligned}$$

$$\begin{aligned}
\text{wobei} \quad K_{i,j}^{(y)}(x) &= P(A_\sigma = j, S_\sigma \leq x, \max_{0\leq k\leq\sigma} S_k \leq y|A_0 = i) \\
&= P(S_\sigma \leq x, \max_{0\leq k\leq\sigma} S_k \leq y|A_0 = i, A_\sigma = j)P(A_\sigma = j|A_0 = i).
\end{aligned}$$

Somit erhält man

$$\begin{aligned}
P(\max_{k\geq 0} S_k > y|A_0 = i)e^{\theta^* y} &= P(\max_{0\leq k\leq\sigma} S_k > y|A_0 = i)e^{\theta^* y} \\
&\quad + \sum_j \int_{-\infty}^0 e^{\theta^*(y-x)} P(\max_{k\geq 0} S_k > y - x|A_0 = j)e^{\theta^* x} dK_{i,j}^{(y)}(x).
\end{aligned}$$

Wendet man schließlich die Aussage (i) dieser Proposition an, so ergibt sich mit

$K_{i,j}(x) = P(S_\sigma \leq x | A_0 = i, A_\sigma = j) P(A_\sigma = j | A_0 = i)$:

$$
\begin{aligned}
\lim_{y\to\infty} P(\max_{0\leq k\leq\sigma} S_k > y | A_0 = i) e^{\theta^* y} &= e^* \pi_{\theta^*,0}(i) - \sum_j \int_{-\infty}^0 e^* \pi_{\theta^*,0}(i) e^{\theta^* x} dK_{i,j}(x) \\
&= e^* \left(\pi_{\theta^*,0}(i) - \sum_j \int_{-\infty}^0 \pi_{\theta^*,0}(i) e^{\theta^* x} dK_{i,j}(x) \right) \\
&= e^* \Big(\pi_{\theta^*,0}(i) - \sum_j \pi_{\theta^*,0}(i) E(e^{\theta^* S_\sigma} | A_0 = i, A_\sigma = j) \\
&\qquad\qquad \cdot P(A_\sigma = j | A_0 = i) \Big) \\
&= e^* (\pi_{\theta^*,0}(i) - E(e^{\theta^* S_\sigma} \pi_{\theta^*,0}(A_\sigma) | A_0 = i)).
\end{aligned}
$$

Falls X_1 unter $A_0 = i$ und $A_1 = j$ eine Gitterverteilung der Spanne d besitzt, so lässt sich analog herleiten, dass

$$
\lim_{n\to\infty} P(\max_{0\leq k\leq\sigma} S_k > nd | A_0 = i) e^{\theta^* nd} = (\pi_{\theta^*,0}(i) - E(e^{\theta^* S_\sigma} \pi_{\theta^*,0}(A_\sigma) | A_0 = i)) e^*.
$$

$\square$

Um diese Proposition anwenden zu können, wurde bereits zu Anfang des Abschnitts 3.5 vorausgesetzt, dass X_1 unter $A_0 = i$ und $A_1 = j$ entweder eine Gitterverteilung der Spanne d besitzt oder eine stetige Zufallsgrößen ist.

Für $-a \leq 0 < b$ seien $L(-a,b)$ und $I(-a,b)$ definiert durch

$$
L(-a,b) := \inf\{k \geq 1 : S_k \notin (-a,b)\}, \quad \text{sowie} \quad I(-a,b) := \begin{cases} 1 & \text{falls } S_{L(-a,b)} \geq b \\ 0 & \text{falls } S_{L(-a,b)} \leq -a \end{cases}.
$$

Lemma 3.6.5. *Es sei Z^- die erste nichtpositive Partialsumme von (S_m) und σ der erste Index, bei dem der Prozess die positive Achse verlässt, d.h. $\sigma = \inf\{k \geq 1 : S_k \leq 0\}$ und $Z^- = S_\sigma$. Ferner sei $F_i(y)$ die Verteilungsfunktion von $\max_{0\leq k\leq\sigma} S_k$ unter $A_0 = i$, sowie $M_i(y)$ die Verteilungsfunktion von $\max_{k\geq 0} S_k$ unter $A_0 = i$. Dann gilt für alle $i \in \Omega$:*

(i)

$$
\begin{aligned}
&\lim_{y\to\infty} E(e^{\theta^*(S_L - y)} \pi_{\theta^*,0}(A_L) | I(y) = 1, A_0 = i) \\
&= \frac{\pi_{\theta^*,0}(i) - E(e^{\theta^* Z^-} \pi_{\theta^*,0}(A_\sigma) | A_0 = i)}{\lim_{y\to\infty} e^{\theta^* y}(1 - F_i(y))} = \frac{1}{e^*} \quad \textit{für ein } e^*,\ 0 < e^* < \infty
\end{aligned}
$$

(ii)

$$\lim_{\substack{y\to\infty \\ a(y)\to\infty}} E(e^{\theta^*(S_{L(-a(y),y)}-y)}\pi_{\theta^*,0}(A_{L(-a(y),y)})|I(-a(y),y)=1, A_0=i)$$

$$= \frac{\pi_{\theta^*,0}(i)}{\lim\limits_{y\to\infty} e^{\theta^* y}(1-M_i(y))} = \frac{1}{e^*} \quad \textit{für ein } e^*,\ 0<e^*<\infty,\ \textit{für alle } a(y).$$

Falls X_1 unter $A_0=i$ und $A_1=j$ eine Gitterverteilung besitzt, so gelten die Aussagen ebenso. Dabei sind y und $a(y)$ Punkte des Gitterfeldes.

Beweis. Der Beweis verläuft analog zum iid-Fall. Deshalb sei er an dieser Stelle etwas verkürzt dargestellt, für ausführlichere Begründungen siehe Beweis zu Lemma 2.5.5.
zu (i): Nach Anwendung des Optional-Sampling-Theorems (vergleiche Lemma 3.4.1: Eigenschaften der Markov-Ketten-Erweiterung des Waldmartingals) erhält man, da $P(I(y)=0|A_0=i)=F_i(y)$:

$$\begin{aligned} 1 &= (1-F_i(y))e^{\theta^* y}E(e^{\theta^*(S_L-y)}\frac{\pi_{\theta^*,0}(A_L)}{\pi_{\theta^*,0}(A_0)}|I(y)=1, A_0=i) \\ &\quad + F_i(y)E(e^{\theta^* S_L}\frac{\pi_{\theta^*,0}(A_L)}{\pi_{\theta^*,0}(A_0)}|I(y)=0, A_0=i). \end{aligned}$$

Multipliziert man auf beiden Seiten mit $\pi_{\theta^*,0}(i)>0$ und stellt den Ausdruck um, so erhält man

$$\begin{aligned} &\lim_{y\to\infty} E(e^{\theta^*(S_L-y)}\pi_{\theta^*,0}(A_L)|I(y)=1, A_0=i) \\ &= \lim_{y\to\infty} \frac{\pi_{\theta^*,0}(i)-F_i(y)E(e^{\theta^* S_L}\pi_{\theta^*,0}(A_L)|I(y)=0, A_0=i)}{e^{\theta^* y}(1-F_i(y))} \\ &= \frac{\pi_{\theta^*,0}(i)-E(e^{\theta^* Z^-}\pi_{\theta^*,0}(A_\sigma)|A_0=i)}{\lim\limits_{y\to\infty} e^{\theta^* y}(1-F_i(y))} \\ &\quad (\text{da } \lim_{y\to\infty} P^{L(y)|I(y)=0} = P^\sigma,\ \sigma=\inf\{k\geq 1: S_k\leq 0\}) \\ &= \frac{1}{e^*} \quad (\text{nach Proposition 3.6.4 (ii), da } 1-F_i(y)=P(\max_{0\leq k\leq\sigma} S_k>y|A_0=i)). \end{aligned}$$

zu (ii): Völlig analog zu (i) erhält man

$$\begin{aligned} &\lim_{\substack{y\to\infty \\ a(y)\to\infty}} E(e^{\theta^*(S_{L(-a(y),y)}-y)}\pi_{\theta^*,0}(A_{L(-a(y),y)})|I(-a(y),y)=1, A_0=i) \\ &= \lim_{\substack{y\to\infty \\ a(y)\to\infty}} \frac{\pi_{\theta^*,0}(i)-M_i(y)E(e^{\theta^* S_{L(-a(y),y)}}\pi_{\theta^*,0}(A_{L(-a(y),y)})|I(-a(y),y)=0, A_0=i)}{e^{\theta^* y}(1-M_i(y))} \\ &= \frac{\pi_{\theta^*,0}(i)}{\lim\limits_{y\to\infty} e^{\theta^* y}(1-M_i(y))} \\ &\quad (\text{da } S_{L(-a(y),y)}\leq -a(y) \text{ unter } I(-a(y),y)=0,\ \pi_{\theta^*,0}(A_{L(-a(y),y)})\leq 1) \\ &= \frac{1}{e^*} \quad (\text{nach Proposition 3.6.4 (i), da } 1-M_i(y)=P(\max_{k\geq 0} S_k>y|A_0=i)). \end{aligned}$$

Für den Gittertyp lässt sich Proposition 3.6.4 analog anwenden mit $y = nd$, falls y und $a(y)$ Gitterpunkte sind. $\square$

Im Folgenden benutzen wir folgende Bezeichnungen

$$\begin{aligned}\Gamma_i(y) &:= E(e^{t\sqrt{\frac{\omega^*}{y}}W_{L(y)}}e^{\theta^*(S_{L(y)}-y)}\pi_{\theta^*,0}(A_L)|I(y)=1, A_0=i)\\ \hat{\Gamma}_i(y) &:= E(e^{t\sqrt{\frac{\omega^*}{y}}W_{L(a(y))}}e^{\theta^*(S_{L(y)}-y)}\pi_{\theta^*,0}(A_L)|I(y)=1, A_0=i),\end{aligned}$$

ferner sei $a(y) = y - \log y$ und y hinreichend groß, so dass $\log y > K$ (K derart, dass $|X_i| \leq K$).
Analog zum iid-Fall wird folgendes Lemma benötigt:

Lemma 3.6.6. *Für alle $i \in \Omega$ gilt:* $\lim_{y\to\infty}\Gamma_i(y) = \lim_{y\to\infty}\hat{\Gamma}_i(y)$.

Beweis. Auch hier sei der Beweis gegenüber dem iid-Fall etwas kürzer dargestellt, für genauere Erklärungen siehe Beweis zu Lemma 2.5.6.

$$\begin{aligned}&|\Gamma_i(y) - \hat{\Gamma}_i(y)|\\ =\ &\Big|E\Big(e^{t\sqrt{\frac{\omega^*}{y}}W_{L(a(y))}}\pi_{\theta^*,0}(A_L)(e^{t\sqrt{\frac{\omega^*}{y}}(W_{L(y)}-W_{L(a(y))})}-1)e^{\theta^*(S_{L(y)}-y)}\Big|I(y)=1, A_0=i\Big)\Big|\\ \leq\ &e^{\theta^*K}E\Big(e^{t\sqrt{\frac{\omega^*}{y}}W_{L(a(y))}}\pi_{\theta^*,0}(A_L)(e^{|t|\sqrt{\frac{\omega^*}{y}}K'(L(y)-L(a(y)))}-1)\Big|I(y)=1, A_0=i\Big).\end{aligned}\tag{3.24}$$

Wir betrachten nun die σ-Algebra

$$\mathcal{S} := \sigma(S_1, ..., S_{L(a(y))}, W_1, ..., W_{L(a(y))}, A_0, ..., A_{L(a(y))}, L(a(y)), I(y)=1).$$

Ferner bezeichne $G_{y,i}$ die Verteilungsfunktion von $S_{L(a(y))}$ unter $I(y) = 1$ und $A_0 = i$, d.h.

$$G_{y,i}(\xi) = P(S_{L(a(y))} \leq \xi | I(y) = 1, A_0 = i).$$

Mit diesen Bezeichnungen gilt

$$\begin{aligned}&E\left(e^{t\sqrt{\frac{\omega^*}{y}}W_{L(a(y))}}\pi_{\theta^*,0}(A_L)\left(e^{|t|\sqrt{\frac{\omega^*}{y}}K'(L(y)-L(a(y)))}-1\right)\Big|I(y)=1, A_0=i\right)\\ &= E\left(e^{t\sqrt{\frac{\omega^*}{y}}W_{L(a(y))}}\pi_{\theta^*,0}(A_L)E\left(e^{|t|\sqrt{\frac{\omega^*}{y}}K'(L(y)-L(a(y)))}-1\Big|\,\mathcal{S}\right)\Big|I(y)=1, A_0=i\right)\\ &= E\Big(e^{t\sqrt{\frac{\omega^*}{y}}W_{L(a(y))}}\pi_{\theta^*,0}(A_L)\\ &\quad\cdot\int_{a(y)}^{a(y)+K}E\Big(e^{|t|\sqrt{\frac{\omega^*}{y}}K'L(-\xi,y-\xi)}-1\Big|I(-\xi,y-\xi)=1, A_0=i\Big)dG_{y,i}(\xi)\Big|I(y)=1, A_0=i\Big).\end{aligned}\tag{3.25}$$

Sei nun

$$D_i(y) := \sup_{a(y)\le\xi\le a(y)+K} E(e^{|t|\sqrt{\frac{\omega^*}{y}}K'L(-\xi,y-\xi)} - 1|I(-\xi, y-\xi) = 1, A_0 = i).$$

Damit erhält man nun mit (3.24) und (3.25) folgende Abschätzung:

$$\begin{aligned}|\Gamma_i(y) - \hat{\Gamma}_i(y)| &\le e^{\theta^* K} D_i(y) E(e^{t\sqrt{\frac{\omega^*}{y}}W_{L(a(y))}}\pi_{\theta^*,0}(A_L)|I(y) = 1, A_0 = i)\\ &\le e^{\theta^* K} D_i(y)\hat{\Gamma}_i(y).\end{aligned}$$

Es bleibt also zu zeigen, dass $D_i(y)$ für $y \to \infty$ gegen 0 konvergiert.
Die Zufallsgröße $D_i(y)$ aus dem Markov-Fall unterscheidet sich nur in der Bedingung $A_0 = i$ von der Zufallsgröße $D(y)$ aus dem iid-Fall. Falls also $Q_{\xi,y,i}$ die Verteilungsfunktion von $L(-\xi, y-\xi)$ unter $I(-\xi, y-\xi) = 1$ und $A_0 = i$ bezeichnet, lässt sich völlig analog zum iid-Fall herleiten, dass

$$D_i(y) \le e^{|t|\sqrt{\frac{\omega^*}{y}}K'y^{1/4}} + \sup_{a(y)\le\xi\le a(y)+K} \left(\int_{y^{1/4}}^{\infty} e^{|t|\sqrt{\frac{\omega^*}{y}}K'x} dQ_{\xi,y,i}(x)\right) - 1.$$

Da $(e^{|t|\sqrt{\omega^*}K'y^{-1/4}} - 1) \underset{y\to\infty}{\longrightarrow} 0$, bleibt zu zeigen, dass

$$\sup_{a(y)\le\xi\le a(y)+K} \left(\int_{y^{1/4}}^{\infty} (e^{|t|\sqrt{\frac{\omega^*}{y}}K'x}) dQ_{\xi,y,i}(x)\right) \underset{y\to\infty}{\longrightarrow} 0.$$

Genau wie im iid-Fall gelten ferner - hier für jeden Wert von A_0 - folgende Teilmengenbeziehungen:

$$\begin{aligned}\{L(-\xi, y-\xi) \ge x,\ I(-\xi, y-\xi) = 1\} &\subset \bigcup_{n\ge x} \{S_n \ge 0\}\\ \{I(y-\xi) = 1\} &\subset \{I(-\xi, y-\xi) = 1\}.\end{aligned}$$

und somit erhält man analog zum Beweis von Lemma 2.5.6:

$$1 - Q_{\xi,y,i}(x) = P(L(-\xi, y-\xi) \ge x|I(-\xi, y-\xi) = 1, A_0 = i) \le \frac{\sum_{n=x}^{\infty} P(S_n \ge 0|A_0 = i)}{P(I(y-\xi) = 1|A_0 = i)}.$$

Nach Lemma 3.4.2 gilt $P(I(y-\xi) = 1|A_0 = i) \ge \delta e^{-\theta^*(y-\xi)}$ für ein $\delta > 0$, ferner existieren nach Ney und Nummelin (siehe [17], Theorem 5.3) Konstanten c und b, so dass $P(\frac{S_n}{n} > 0|A_0 = i) \le ce^{-bn}$.
Also erhält man wie im iid-Fall, hier mit $C_1 := \frac{c}{\delta(1-e^{-n})}$:

$$1 - Q_{\xi,y,i}(x) \le C_1 y^{\theta^*} e^{-bx}.$$

Mit Hilfe dieser Abschätzung kann man analog zum iid-Fall mittels partieller Integration zeigen, dass für genügend großes y und für eine geeignete Konstante C_3 folgende Abschätzung erfüllt ist:

$$\int_{y^{1/4}}^{\infty} e^{|t|\sqrt{\frac{\omega^*}{y}}K'x} dQ_{\xi,y,i}(x) \le C_3 e^{-\frac{1}{2}by^{1/4}}.$$

Es ergibt sich also schließlich

$$D_i(y) = \sup_{a(y)\le\xi\le a(y)+K} \left(\int_{y^{1/4}}^{\infty} (e^{|t|\sqrt{\frac{\omega^*}{y}}K'x}) dQ_{\xi,y}(x) \right) \le C_3 e^{-\frac{1}{2}by^{1/4}} \underset{y\to\infty}{\longrightarrow} 0$$

und somit folgt die gewünschte Aussage

$$|\Gamma_i(y) - \hat{\Gamma}_i(y)| \underset{y\to\infty}{\longrightarrow} 0.$$

□

Nun können wir eine zu Lemma 2.5.7 analoge Aussage beweisen.

Lemma 3.6.7. *Sei $u^* = 0$. Dann gilt für alle $i \in \Omega$ und für alle $t \in \mathbb{R}$:*

$$\lim_{y\to\infty} E(e^{t\sqrt{\frac{\omega^*}{y}}\sum_{k=K_{\nu-1}+1}^{T_\nu(y)} U_k} | I_\nu(y) = 1, A_0 = i) = e^{\frac{1}{2}t^2 v^*},$$

wobei $v^* = \frac{d^2\zeta}{dt^2}(\theta^*, 0) = \sum_{ij} p_{ij} E(U_1^2 e^{\theta^* X_1} | A_0 = i, A_1 = j)\pi_{\theta^*,0}(j)\psi_{\theta^*,0}(i)$.

Beweis. Die σ-Algebra $\mathcal{S}$ sei wie im vorhergehenden Beweis durch

$$\mathcal{S} = \sigma(S_1, ..., S_{L(a(y))}, W_1, ..., W_{L(a(y))}, A_0, ..., A_{L(a(y))}, L(a(y)), I(y) = 1)$$

gegeben. Ferner bezeichne $H_{y,i}$ die gemeinsame Verteilungsfunktion von $S_{L(a(y))}$ und $A_{L(a(y))}$ unter $I(y) = 1$, $A_0 = i$, d.h.

$$H_{y,i}(\xi, j) = P(S_{L(a(y))} \le \xi, A_{L(a(y))} \le j | I(y) = 1, A_0 = i).$$

Analog zum iid-Fall erhalten wir

$$\begin{aligned}
\hat{\Gamma}_i(y) &= E(e^{t\sqrt{\frac{\omega^*}{y}}W_{L(a(y))}} \pi_{\theta^*,0}(A_L) e^{\theta^*(S_L - y)} | I(y) = 1, A_0 = i) \\
&= E\left(e^{t\sqrt{\frac{\omega^*}{y}}W_{L(a(y))}} E(e^{\theta^*(S_L - y)} \pi_{\theta^*,0}(A_L) e^{\theta^*(S_L - y)} |\ \mathcal{S}) \Big| I(y) = 1, A_0 = i \right) \\
&= E\Big(e^{t\sqrt{\frac{\omega^*}{y}}W_{L(a(y))}} \sum_j \int_{a(y)}^{a(y)+K} E(e^{\theta^*(S_{L(-\xi,y-\xi)} - (y-\xi))} \pi_{\theta^*,0}(A_{L(-\xi,y-\xi)}) \\
&\qquad | I(-\xi, y-\xi) = 1, A_0 = j) dH_{y,i}(\xi, j) \Big| I(y) = 1, A_0 = i \Big).
\end{aligned}$$

Nach Lemma 3.6.5 (ii) gilt nun, dass der Integrand für alle $a(y) < \xi < a(y) + K$ gegen $\frac{1}{e^*}$ konvergiert (beim Gittertyp gilt eine entsprechende Aussage, das Integral entspricht dann einer Summe).
Daher gilt für alle i:

$$\lim_{y\to\infty} \hat{\Gamma}_i(y) = \frac{1}{e^*} \lim_{y\to\infty} E(e^{t\sqrt{\frac{\omega^*}{y}}W_{L(a(y))}} | I(y) = 1, A_0 = i).$$

Nach Lemma 3.6.6 gilt also auch für alle i:

$$\lim_{y\to\infty} \Gamma_i(y) = \frac{1}{e^*} \lim_{y\to\infty} E(e^{t\sqrt{\frac{\omega^*}{y}}W_L} | I(y) = 1, A_0 = i).$$

Da ferner nach Lemma 3.6.3

$$\lim_{y\to\infty} \frac{\Gamma_i(y)}{E(e^{\theta^*(S_L-y)}\pi_{\theta^*,0}(A_L)|I(y) = 1, A_0 = i)} = e^{\frac{1}{2}t^2 v^*}$$

und nach Lemma 3.6.5 (i)

$$\lim_{y\to\infty} E(e^{\theta^*(S_L-y)}\pi_{\theta^*,0}(A_L)|I(y) = 1, A_0 = i) = \frac{1}{e^*}$$

gilt, folgt

$$\lim_{y\to\infty} E(e^{t\sqrt{\frac{\omega^*}{y}}W_L} | I(y) = 1, A_0 = i) = e^{\frac{1}{2}t^2 v^*}.$$

□

Jetzt steht alles zur Verfügung, um den zentralen Grenzwertsatz dieses Kapitels zu beweisen. Die Vorgehensweise entspricht dabei genau derjenigen des iid-Falls.

Beweis von Theorem 3.5.1. Analog zum iid-Fall erhalten wir mit Hilfe von Lemma 3.6.7 und dem Stetigkeitssatz für Laplace-Transformierte (siehe Proposition 2.5.8), dass, falls $u^* = 0$, für jeden Wert von A_0 unter $I_\nu(y) = 1$ gilt:

$$\sqrt{\frac{\omega^*}{y}} \sum_{k=K_{\nu-1}+1}^{T_\nu(y)} U_k \underset{(V)}{\longrightarrow} \mathcal{N}(0, v^*) \quad \text{mit} \tag{3.26}$$

$$v^* = \sum_{ij} p_{ij} E(U_1^2 e^{\theta^* X_1} | A_0 = i, A_1 = j)\pi_{\theta^*,0}(j)\psi_{\theta^*,0}(i).$$

Um die Aussage nun auch für den mehrdimensionalen Fall zu beweisen, betrachten wir die Zufallsvektoren $\boldsymbol{U_1}, \boldsymbol{U_2}, ...$ aus dem Theorem. $\boldsymbol{z} = (z_1, ..., z_m)$ sei ein reeller Vektor mit $|\boldsymbol{z}| = 1$.
Wähle nun $U_i = \langle \boldsymbol{z}, \boldsymbol{U_i} - \boldsymbol{u^*} \rangle = \sum_{k=1}^m z_k(U_{ik} - u_k^*),\ i \in \mathbb{N}$
Dann ist ζ hier durch $\zeta(\theta, t) = \log \rho(\theta, t)$ gegeben, wobei $\rho(\theta, t)$ der Spektralradius der Matrix $P_\theta = (p_{ij} E(e^{\theta X_1 + t\langle \boldsymbol{z}, \boldsymbol{U_1} - \boldsymbol{u^*}\rangle} | A_0 = i, A_1 = j))_{i,j\in\{1,...r\}}$ ist. Somit gilt

$$
\begin{aligned}
\omega^* = \frac{d\zeta}{d\theta}(\theta^*, 0) &= \sum_{ij} p_{ij} E(X_1 e^{\theta^* X_1} | A_0 = i, A_1 = j) \pi_{\theta^*,0}(j) \psi_{\theta^*,0}(i) \\
u^* = \frac{d\zeta}{dt}(\theta^*, 0) &= \sum_{ij} p_{ij} E\left(\sum_k z_k (U_{ik} - u_k^*) e^{\theta^* X_1} | A_0 = i, A_1 = j\right) \pi_{\theta^*,0}(j) \psi_{\theta^*,0}(i) \\
&= \sum_k z_k \Big(\sum_{i,j} p_{ij} E(U_{ik} e^{\theta^* X_1} | A_0 = i, A_1 = j) \pi_{\theta^*,0}(j) \psi_{\theta^*,0}(i) \\
&\qquad - u_k^* \sum_{ij} p_{ij} E(e^{\theta^* X_1} | A_0 = i, A_1 = j) \pi_{\theta^*,0}(j) \psi_{\theta^*,0}(i) \Big) \\
&= \sum_k z_k (u_k^* - u_k^* \rho(\theta^*, 0)) = 0 \quad (\text{da } \rho(\theta^*, 0) = 1).
\end{aligned}
$$

Mit (3.26) erhalten wir daher für jeden Wert von A_0

$$
\sqrt{\frac{\omega^*}{y}} \sum_{k=K_{\nu-1}+1}^{T_\nu(y)} \langle \boldsymbol{z}, \boldsymbol{U_i} - \boldsymbol{u^*} \rangle \underset{(V)}{\longrightarrow} \mathcal{N}(0, v^*) \quad \text{unter } I_\nu(y) = 1
$$

mit

$$
v^* = \sum_{ij} p_{ij} E(\langle \boldsymbol{z}, \boldsymbol{U_1} - \boldsymbol{u^*} \rangle^2 e^{\theta^* X_1} | A_0 = i, A_1 = j) \pi_{\theta^*,0}(j) \psi_{\theta^*,0}(i)
$$

Mit dem Satz von Cramér-Wold (siehe Proposition 2.5.9) folgt schließlich die gewünschte Aussage:
Für jeden Wert von A_0 gilt unter $I_\nu(y) = 1$

$$
\sqrt{\frac{\omega^*}{y}} \sum_{k=K_{\nu-1}+1}^{T_\nu(y)} (\boldsymbol{U_i} - \boldsymbol{u^*}) \underset{(V)}{\longrightarrow} \mathcal{N}(0, \Sigma)
$$

mit $\Sigma = (\sigma_{kl})_{k,l \in \{1,\ldots,m\}}$, wobei

$$
\sigma_{lk} = \sum_{ij} p_{ij} E((\boldsymbol{U_1} - \boldsymbol{u^*})_l (\boldsymbol{U_1} - \boldsymbol{u^*})_k e^{\theta^* X_1} | A_0 = i, A_1 = j) \pi_{\theta^*,0}(j) \psi_{\theta^*,0}(i).
$$

□

3.7 Anwendungen

Wie auch im iid-Fall (siehe Abschnitt 2.6) lassen sich die Resultate dieses Kapitels zum Vergleich zweier oder mehrerer DNA- bzw. Proteinsequenzen anwenden, um so z.B. Aufschluß über den Verwandtschafts- bzw. Ähnlichkeitsgrad von Sequenzen zu gewinnen.
Beim Markov-Modell gehen wir nun im Gegensatz zum iid-Fall allgemeiner davon aus, dass Nucleotide bzw. einzelne Aminosäuren nicht unabhängig voneinander auftreten,

sondern dass sie jeweils vom vorhergehenden Nucleotid bzw. von der vorhergehenden Aminosäure abhängen. Im Blickfeld stehen also sogenannte Dinucleotide bzw. zwei aufeinanderfolgende Aminosäuren. Dementsprechend wird beim Vergleich zweier Sequenzen das mathematische Modell komplizierter - so ersetzen wir beispielsweise p_i durch $p_{\alpha,\beta}p'_{\gamma,\delta}$ (miteinander multiplizierte Dinucleotidwahrscheinlichkeiten in den jeweiligen Sequenzen) und q_i durch $q_{\alpha,\beta,\gamma,\delta}$ (Dinucleotidaustauschwahrscheinlichkeit). Dennoch lassen sich die Theoreme völlig analog anwenden, so dass für den biologischen Bezug auf die Abschnitte 2.6.1 und 2.6.2 verwiesen sei. Stattdessen sei an dieser Stelle ein Beispiel angeführt, bei welchem man Sequenzen betrachtet, die Einträge aus einem Alphabet mit nur zwei Buchstaben enthalten.

3.7.1 Sequenzen aus zwei Buchstaben

Gegeben sei eine Sequenz (oder ein Sequenzenpaar), deren Elemente nur zwei Zustände annehmen können, d.h. $\Omega = \{0, 1\}$, dabei hänge jeder Buchstabe vom vorherigen ab, es liege also eine Markov-Kette vor. Jeweils zwei aufeinanderfolgenden Buchstaben wird ein Score der Form $s_{\alpha,\beta} = 1$ oder -1 zugeordnet, d.h. gegeben sei eine Zufallsgröße $X_1 : \Omega^2 \to \{-1, 1\}$, deren Werte von A_0 und A_1 abhängen, sowie die Übergangswahrscheinlichkeits-Matrix

$$\begin{pmatrix} \alpha & 1-\alpha \\ 1-\beta & \beta \end{pmatrix},$$

mit $\alpha = p_{00} = P(A_1 = 0|A_0 = 0)$ und $\beta = p_{11} = P(A_1 = 1|A_0 = 1)$.
Um Erwartungswerte ausrechnen zu können, benötigen wir die Startverteilung, also die Verteilung von A_0. Dazu bestimmen wir zunächst $P(A_1 = 0)$:

$$\begin{aligned} P(A_1 = 0) &= P(A_1 = 0, A_0 = 0) + P(A_1 = 0, A_0 = 1) \\ &= P(A_0 = 0)\alpha + P(A_0 = 1)(1-\beta) \\ &= P(A_0 = 0)(\alpha - (1-\beta)) + (1-\beta). \end{aligned}$$

Damit erhält man auch

$$\begin{aligned} P(A_2 = 0) &= P(A_1 = 0)(\alpha - (1-\beta)) + (1-\beta) \\ &= P(A_0 = 0)(\alpha - (1-\beta))^2 + (1-\beta)(\alpha - (1-\beta)) + (1-\beta) \\ &= P(A_0 = 0)(\alpha - (1-\beta))^2 + (1-\beta)\sum_{j=0}^{1}(\alpha - (1-\beta))^j. \end{aligned}$$

Induktiv folgt weiterhin für alle $n \in \mathbb{N}$:

$$\begin{aligned} P(A_n = 0) &= P(A_0 = 0)(\alpha - (1-\beta))^n + (1-\beta)\sum_{j=0}^{n-1}(\alpha - (1-\beta))^j \\ &= P(A_0 = 0)(\alpha - (1-\beta))^n + (1-\beta)\frac{1-(\alpha-(1-\beta))^n}{1-(\alpha-(1-\beta))} \\ &\quad \text{(endliche geometrische Reihe)} \\ &= \frac{1-\beta}{(1-\alpha)+(1-\beta)} + (\alpha - (1-\beta))^n\left(P(A_0 = 0) - \frac{1-\beta}{(1-\alpha)+(1-\beta)}\right). \end{aligned}$$

Wir möchten aber nun, dass $P(A_n = 0)$ nicht von n abhängt, daher erhalten wir:

$$P(A_0 = 0) = \frac{1-\beta}{(1-\alpha)+(1-\beta)}.$$

Scorezuteilungen können nun auf drei Weisen erfolgen:

(i) X_1 hängt nur von einem Zustand ab, also z.B. $X_1 = (-1)^{A_1}$, d.h. $s_{01} = s_{11} = -1$ und $s_{10} = s_{00} = 1$; falls $E(X_1) = \sum_{i,j\in\{0,1\}} P(A_0 = i)p_{ij}s_{ij} < 0$ gelten soll, folgt $\beta > \alpha$, denn:

$$\begin{aligned} 0 &> P(A_0 = 0)(s_{00}p_{00} + s_{01}p_{01}) + (1 - P(A_0 = 0))(s_{10}p_{10} + s_{11}p_{11}) \\ &= P(A_0 = 0)(s_{00}p_{00} + s_{01}p_{01} - s_{10}p_{10} - s_{11}p_{11}) + s_{10}p_{10} + s_{11}p_{11} \\ &= \frac{1-\beta}{(1-\alpha)+(1-\beta)}\big(\alpha - (1-\alpha) - (1-\beta) + \beta\big) + (1-\beta) - \beta \\ &= \frac{1-\beta}{(1-\alpha)+(1-\beta)}(2\alpha + 2\beta - 2) + 1 - 2\beta, \end{aligned}$$

was äquivalent ist zu

$$\begin{aligned} 0 &> (1-\beta)(2\alpha + 2\beta - 2) + (1-2\beta)((1-\alpha)+(1-\beta)) \\ &= 2\alpha + 2\beta - 2 - 2\alpha\beta - 2\beta^2 + 2\beta + 1 - \alpha + 1 - \beta - 2\beta + 2\alpha\beta - 2\beta + 2\beta^2 \\ &= \alpha - \beta; \end{aligned}$$

(ii) X_1 misst Übergänge von einem Zustand in den anderen, also z.B. $X_1 = (-1)^{A_0+A_1}$, d.h. $s_{01} = s_{10} = -1$ und $s_{00} = s_{11} = 1$; falls $E(X_1) < 0$ muss also gelten

$$\begin{aligned} 0 &> \frac{1-\beta}{(1-\alpha)+(1-\beta)}\big(\alpha - (1-\alpha) + (1-\beta) - \beta\big) - (1-\beta) + \beta \\ &= \frac{1-\beta}{(1-\alpha)+(1-\beta)}(2\alpha - 2\beta) + 2\beta - 1, \end{aligned}$$

dies ist äquivalent zu

$$\begin{aligned} 0 &> (1-\beta)(2\alpha - 2\beta) + (2\beta - 1)((1-\alpha)+(1-\beta)) \\ &= 3\alpha + 3\beta - 4\alpha\beta - 2, \end{aligned}$$

was sich umformen lässt zu

$$2 - 2\alpha - 2\beta + 2\alpha\beta > \alpha - 2\alpha\beta + \beta$$

und was wiederum äquivalent ist zu

$$2(1-\alpha)(1-\beta) > \alpha(1-\beta) + \beta(1-\alpha) \quad \text{bzw.}$$

$$2 > \frac{\alpha}{1-\alpha} + \frac{\beta}{1-\beta};$$

(iii) X_1 misst die Länge von Runs desselben Zustandes, also z.B. $X_1 = (-1)^{\sup(A_0, A_1)}$, d.h. $s_{00} = 1$ und $s_{01} = s_{10} = s_{11} = -1$; falls $E(X_1) < 0$ muss also gelten

$$\begin{aligned} 0 &> \frac{1-\beta}{(1-\alpha)+(1-\beta)}\big(\alpha - (1-\alpha) + (1-\beta) + \beta\big) - (1-\beta) - \beta \\ &= \frac{1-\beta}{(1-\alpha)+(1-\beta)} 2\alpha - 1, \end{aligned}$$

also

$$0 > (1-\beta)2\alpha - ((1-\alpha)+(1-\beta)) = 3\alpha + \beta - 2\alpha\beta - 2,$$

was äquivalent ist zu

$$2(1-\alpha)(1-\beta) > \alpha - \beta = \alpha(1-\beta) - \beta(1-\alpha) \quad \text{bzw.}$$

$$2 > \frac{\alpha}{1-\alpha} - \frac{\beta}{1-\beta}.$$

Um nun in der Praxis verwertbare Aussagen aus den starken Grenzwertsätzen (siehe Abschnitt 3.3) zu gewinnen, ist es erforderlich die Parameter ω^* und u^* genauer zu bestimmen. Dazu wiederum benötigt man die eindeutig bestimmte Lösung θ^* der Gleichung $\rho(\theta) = 1$, wobei $\rho(\theta)$ den Spektralradius der Matrix

$$P_\theta = (p_{ij} E(e^{\theta X_1} | A_0 = i, A_1 = j))_{i,j=0,1} = (p_{ij} e^{\theta s_{ij}})_{i,j \in \{0,1\}}$$

bezeichnet.

Die zu den obigen drei Kategorien gehörigen Matrizen P_θ sind hier also durch

(i) $P_\theta^{(1)} = \begin{pmatrix} \alpha e^\theta & (1-\alpha)e^{-\theta} \\ (1-\beta)e^\theta & \beta e^{-\theta} \end{pmatrix}$

(ii) $P_\theta^{(2)} = \begin{pmatrix} \alpha e^\theta & (1-\alpha)e^{-\theta} \\ (1-\beta)e^{-\theta} & \beta e^{\theta} \end{pmatrix}$

(iii) $P_\theta^{(3)} = \begin{pmatrix} \alpha e^\theta & (1-\alpha)e^{-\theta} \\ (1-\beta)e^{-\theta} & \beta e^{-\theta} \end{pmatrix}$.

gegeben. Nun müssen aber alle drei Übergangsmatrizen die Gestalt $\begin{pmatrix} \gamma & 1-\gamma \\ 1-\delta & \delta \end{pmatrix}$ haben. Dabei gilt $1 = \gamma + \delta - \gamma\delta + (1-\gamma)(1-\delta)$.

Sei nun $\xi := e^\theta$. Bestimmt werden sollen also die Lösungen $\xi^* = e^{\theta^*}$ der folgenden drei Gleichungen

$$\begin{aligned} 1 &= \alpha\xi + \beta\xi^{-1} - \alpha\xi\beta\xi^{-1} + (1-\alpha)\xi^{-1}(1-\beta)\xi = \alpha\xi + \beta\xi^{-1} + 1 - (\alpha+\beta) \quad \text{für } P_\theta^{(1)} \\ 1 &= \alpha\xi + \beta\xi - \alpha\beta\xi^2 + (1-\alpha)(1-\beta)\xi^{-2} \quad \text{für } P_\theta^{(2)} \\ 1 &= \alpha\xi + \beta\xi^{-1} - \alpha\beta + (1-\alpha)(1-\beta)\xi^{-2} \quad \text{für } P_\theta^{(3)}. \end{aligned}$$

Die erste Gleichung hat die Lösung $\xi = 1$, welche irrelevant ist, denn $\xi^* = e^{\theta^*} = 1$ ergäbe die Lösung $\theta^* = 0$ (aber gesucht wird die eindeutig bestimmte positive Lösung von $\rho(\theta^*) = 1$), sowie die erforderliche Lösung $\xi^* = \frac{\beta}{\alpha}$. Somit ergibt sich, dass die Übergangsmatrix, mit Hilfe derer man die Werte ω^* und u^* bestimmen kann, folgende Gestalt hat:

$$P_\theta^{(1)} = \begin{pmatrix} \beta & 1-\beta \\ 1-\alpha & \alpha \end{pmatrix}.$$

Die zweite Gleichung lässt sich mittels einiger Umformungen auf die Form

$$\xi\left(\frac{\xi}{\hat{\alpha}} - 1\right)\left(\frac{\xi}{\hat{\beta}} - 1\right) + \left(1 - \frac{\xi}{\sqrt{\hat{\alpha}\hat{\beta}}}\right)\left(1 + \frac{\xi}{\sqrt{\hat{\alpha}\hat{\beta}}}\right) = 0$$

bringen, wobei $\hat{\alpha} = \frac{1}{\alpha} - 1$ und $\hat{\beta} = \frac{1}{\beta} - 1$. Die Gleichung hat vier Lösungen, nämlich $\xi = 1$, eine negative Lösung (beide sind irrelevant, da $\theta^* = 0$ als triviale Lösung uninteressant ist (s.o.) und da $\xi^* < 0$ im Widerspruch zu $\xi^* = e^{\theta^*}$ stünde) und zwei positive Lösungen. Die größere der beiden positiven Lösungen liegt im Intervall $(\max(\hat{\alpha}, \hat{\beta}) + 1, +\infty)$, für diese erhält man allerdings, dass entweder $\alpha e^\theta > 1$ oder $\beta e^\theta > 1$ gelten muss. Es kommt also nur noch die kleinere der beiden positiven Lösungen in Frage. Diese liegt im Intervall $(\min(\hat{\alpha}, \hat{\beta}), \sqrt{\hat{\alpha}\hat{\beta}})$ und ergibt die korrekte Matrix

$$P_\theta^{(2)} = \begin{pmatrix} \xi^*\alpha & 1-\xi^*\alpha \\ 1-\xi^*\beta & \xi^*\beta \end{pmatrix}.$$

Die dritte Gleichung hat drei Lösungen, nämlich $\xi = 1$, eine negative Lösung (beide sind wieder irrelevant) und eine relevante Lösung

$$\xi = \frac{1}{2\alpha}(1 - \alpha + \alpha\beta + \sqrt{(1-\alpha+\alpha\beta)^2 + 4\alpha(1-\alpha)(1-\beta)}),$$

welche die Matrix

$$P_\theta^{(3)} = \begin{pmatrix} \xi^*\alpha & 1-\xi^*\alpha \\ 1-\frac{\beta}{\xi^*} & \frac{\beta}{\xi^*} \end{pmatrix}$$

liefert, mit Hilfe derer die Größen aus den starken Grenzwertsätzen ω^* und u^* bestimmbar sind.

Ausblick

Ein nächster Schritt unser Modell zu verallgemeinern, bestünde nun darin, davon auszugehen, dass ein Nucleotid (bzw. eine Aminosäure) nicht nur vom Vorgänger, sondern von m vorhergehenden Nucleotiden abhängt und erst vom $(m+1)$ten vorhergehenden Nucleotid unabhängig ist.
Mathematisch liegt dann eine Markov-Kette der Ordnung m vor.

Gegeben sei also eine homogene Markov-Kette $(A_{n-m+1})_{n\in\mathbb{N}_0}$ der Ordnung m mit endlichem Zustandsraum $\Omega = \{1, ..., r\}$ und Anfangsverteilung $P^{(A_{-m+1}, A_{-m+2}, ..., A_0)}$, d.h. für alle $n \geq 0$ und alle $i_{-m+1}, ..., i_0, i_1, ..., i_{n+1} \in \Omega$ gelte

$$\begin{aligned}&P(A_{n+1} = i_{n+1} | A_{-m+1} = i_{-m+1}, ..., A_0 = i_0, A_1 = i_1, ..., A_n = i_n)\\&= P(A_{n+1} = i_{n+1} | A_{n-m+1} = i_{n-m+1}, ..., A_n = i_n) \quad P\text{-f.s.},\end{aligned}$$

wobei diese Wahrscheinlichkeit nicht von n abhängt.
(Unsere Markov-Kette aus Kapitel 3 ist somit eine Markov-Kette der Ordnung 1.)

Formal erhalten wir hier in diesem Modell eine Markov-Kette der Ordnung 1, indem wir

$$B_n := (A_{n-m+1}, ..., A_n),\ n \in \mathbb{N}_0$$

setzen, denn mit dieser Bezeichnung gilt

$$\begin{aligned}&P(B_n = (i_{n-m+1}, ..., i_n) | B_0 = (i_{-m+1}, ..., i_0), ..., B_{n-1} = (i_{n-m}, ..., i_{n-1}))\\&= P(A_{n-m+1} = i_{n-m+1}, ..., A_n = i_n | A_{-m+1} = i_{-m+1}, ..., A_{n-1} = i_{n-1})\\&= P(A_{n-m+1} = i_{n-m+1}, ..., A_n = i_n | A_{n-m} = i_{n-m}, ..., A_{n-1} = i_{n-1})\\&= P(B_n = (i_{n-m+1}, ..., i_n) | B_{n-1} = (i_{n-m}, ..., i_{n-1})).\end{aligned}$$

$(B_n)_{n\in\mathbb{N}_0}$ ist also eine Markov-Kette der Ordnung 1 mit Zustandsraum Ω^m und Übergangswahrscheinlichkeiten

$$p_{(i_1,...,i_m)(j_1,...,j_m)} := \begin{cases} p_{i_m j_m} & \text{falls } i_{k+1} = j_k \text{ für alle } 1 \leq k \leq m-1 \\ 0 & \text{sonst} \end{cases}.$$

Für die Scorezuteilungen betrachten wir wieder die Folge $(X_n)_{n\in\mathbb{N}}$ oder allgemeiner die Folge $(X_n, U_n)_{n\in\mathbb{N}}$, um die Zusammensetzung hochscoriger Sequenzsegmente zu untersuchen (mit $U_n = \mathbb{I}_A(X_n)$). Da die Buchstaben in der Sequenz von je m vorhergehenden Buchstaben abhängen sollen, ordnen wir nun also $(m+1)$-Tupeln einen Scorewert

zu. Die Verteilung von (X_n, U_n) hängt also von $A_{n-m}, ..., A_n$ oder anders gesagt von $B_{n-1} = (A_{n-m}, ..., A_{n-1})$ und $B_n = (A_{n-m+1}, ..., A_n)$ ab.
Wir haben nun also die gleiche Ausgangssituation wie in Kapitel 3 mit der Ausnahme, dass der Zustandsraum hier m-dimensional ist (Ω^m statt Ω) und wir formal A_n durch B_n ersetzen.

Diese Überlegungen könnten bei der Verallgemeinerung der Grenzwertsätze auf den Fall, bei welchem eine Markov-Kette der Ordnung m vorliegt, behilflich sein. Es ist sogar zu vermuten, dass die Beweise der starken Grenzwertsätze (siehe Abschnitt 3.4) eins zu eins übertragbar sind (indem man A_n durch B_n ersetzt und Ω durch Ω^m). Dahingegen würde der Beweis des zentralen Grenzwertsatzes (siehe Abschnitt 3.6) weitere Untersuchungen erfordern, da beispielsweise die Voraussetzung $p_{ij} > 0$ für alle i, j, welche wir bisher zum Beweis des zentralen Grenzwertsatzes benötigt hatten, in diesem Fall nicht gestellt werden kann, da nach Definition der Überganswahrscheinlichkeiten bei Zustandsraum Ω^m (siehe oben) immer $i, j \in \Omega^m$ existieren mit $p_{ij} > 0$.
Eine genauere Untersuchung der Gültigkeit der Grenzwertsätze bei Vorliegen einer Markov-Kette der Ordnung m würde jedoch den Rahmen dieser Diplomarbeit sprengen, deshalb seien diese Überlegungen lediglich als ein Ausblick formuliert.

Literaturverzeichnis

[1] ALBERTS, B., BRAY, D., JOHNSON, A., LEWIS, J., RAFF, M., ROBERTS, K., WALTER, P. (1999). Lehrbuch der molekularen Zellbiologie. Wiley-VCH.

[2] ALSMEYER, G. (1998). Wahrscheinlichkeitstheorie. Skripten zur Mathematischen Statistik, Nr. 30, Universität Münster.

[3] ÇINLAR, E. (1975). Introduction to Stochastic Processes. Prentice-Hall, Englewood Cliffs, New Jersey.

[4] DEMBO, A., KARLIN, S. (1991). Strong limit theorems of empirical functionals for large exceedances of partial sums of i.i.d. variables. *Ann. Prob.* 19 1737-1755.

[5] DEMBO, A., KARLIN, S. (1991). Strong limit theorems of empirical distributions for large segmental exceedances of partial sums of Markov variables. *Ann. Prob.* 19 1756-1767.

[6] DEMBO, A., KARLIN, S. (1993). Central limit theorems of partial sums for large segmental values. *Stochastic Processes Appl.* 45 259-271.

[7] DEMBO, A., KARLIN, S., ZEITOUNI, O. (1994). Limit distribution of maximal non-aligned two-sequence segmental score. *Ann. Appl. Prob.* 22 2022-2039.

[8] DEMBO, A., ZEITOUNI, (1993). Large deviations techniques and applications. Jones and Barlett book in mathematics, Boston.

[9] DODGE, Y. (2003) Oxford Dictionary of Statistical Terms. Oxford University Press.

[10] ERDÖS, P., RÉNYI, A. (1970). On a new law of large numbers. *J. Anal. Math.* 23 103-111.

[11] HIGGINS, D., TAYLOR, W. (2000). Bioinformatics: Sequence, structure and databanks. Oxford University Press.

[12] IGLEHART, D. (1972). Extreme values in the GI/G/1 queue. *Ann. Math. Statist.* 43 627-635.

[13] KARLIN, S., ALTSCHUL, S.F. (1990). New methods for assessing the statistical significance of molecular sequence features using general scoring schemes. *Proc. Nat. Acad. Sci. USA* 87 2264-2268.

[14] Karlin, S., Dembo, A. (1992). Limit distributions of maximal segmental score among Markov dependent partial sums. *Adv. in Appl. Prob.* 24 113-140.

[15] Karlin, S., Dembo, A., Kawabata, T. (1990). Statistical composition of high scoring segments from molecular sequences. *Ann. Statist.* 18 571-581.

[16] Karlin, S., Ost, F. (1985). Some monotonicity properties of Schur powers of matrices and related inequalities. *Linear Algebra Appl.* 68 47-65.

[17] Ney, P., Nummelin, E. (1987). Markov additive processes. I. Eigenvalue properties and limit theorems. *Ann. Prob.* 15 561-592.

[18] Rauhut, R. (2001). Bioinformatik: Sequenz, Struktur, Funktion. Wiley-VCH.

[19] Schmitz, N. (1996). Vorlesungen über Wahrscheinlichkeitstheorie. Teubner Studienbücher Mathematik.

[20] Seneta, E.(1981). Non-Negative Matrices. Springer Verlag.

Hiermit versichere ich, dass ich diese Arbeit selbständig verfasst und keine anderen als die angegebenen Hilfsmittel benutzt habe. Alle Stellen, die den Ausführungen anderer Autoren wörtlich oder sinngemäß entnommen worden sind, habe ich durch Angabe der Quellen als solche kenntlich gemacht.

Münster, den 10.11.2005