Funktionale von n-Coalescents

Dissertation
der Mathematisch-Naturwissenschaftlichen Fakultät
der Eberhard Karls Universität Tübingen
zur Erlangung des Grades eines
Doktors der Naturwissenschaften
(Dr. rer. nat.)

vorgelegt von
Fabian Freund
aus Stuttgart

Tübingen
2011

Tag der mündlichen Qualifikation:	09.06.2011
Dekan:	Prof. Dr. Wolfgang Rosenstiel
1. Berichterstatter	Prof. Dr. Martin Möhle
2. Berichterstatter	Prof. Dr. Martin Zerner

Inhaltsverzeichnis

Zusammenfassung

In dieser Arbeit werden die Eigenschaften verschiedener Funktionale von ($n-$)Coalescents untersucht. Coalescents sind stochastische Prozesse, die Werte in der Menge $\mathbb{E}$ der Partitionen der natürlichen Zahlen annehmen, wobei hier die Teilmengen, aus denen eine Partition besteht, Blöcke genannt werden. Ein stochastischer Prozess $\Pi = (\Pi_t)_{t\geq 0}$ mit Zuständen in $\mathbb{E}$ heißt Coalescent, falls er folgende Eigenschaften besitzt:

- Die Pfade von Π sind càdlàg.
- Übergänge von Π sind nur möglich durch Vereinigung von Blöcken des derzeitigen Zustandes.
- Für jedes $n \in \mathbb{N}$ betrachte $\Pi^{(n)} = (\Pi_t^{(n)})_{t\geq 0}$, die Restriktion von Π auf die Menge $\mathbb{E}_n$ der Partitionen von $\{1, \ldots, n\}$. $\Pi^{(n)}$ ist ein Markov-Prozess. Jeder Übergang von $\eta \in \mathbb{E}_n$ nach $\xi \in \mathbb{E}_n$, der durch Vereinigen von (disjunkten) Mengen von $k_1 \geq \ldots \geq k_m \geq 2$ der b Blöcke von η zu jeweils einem Block von ξ und dem Beibehalten der restlichen $b - \sum_{i\in m} k_i$ Blöcke von η entsteht, besitzt die identische infinitesimale Rate $\lambda(b; k_1, \ldots, k_m)$.
- Π startet in der Partition $(\{i\})_{i\in\mathbb{N}}$ der natürlichen Zahlen.

Der Prozess $\Pi^{(n)}$ wird n-Coalescent genannt, ebenso jeder càdlàg Prozess mit Werten in $\mathbb{E}_n$, der identisch verteilt wie $\Pi^{(n)}$ ist. Die Vereinigung von einer Menge von Blöcken zu einem neuen Block in einem Übergang von Π (bzw. $\Pi^{(n)}$) wird Kollision genannt.
Das erste Beispiel eines solchen Prozesses wurde 1982 von Kingman (siehe [56]) vorgestellt. Der Kingman-Coalescent hat die Eigenschaft, dass für $n \in \mathbb{N}$ die einzigen möglichen Übergänge in $\Pi^{(n)}$ diejenigen sind, die genau zwei Blöcke des Zustandes zu einem vereinigen und alle anderen Blöcke unverändert lassen. Der Kingman-Coalescent lässt also nur binäre Kollisionen zu. Ein weiterer Coalescent, der Bolthausen-Sznitman-Coalescent, wurde 1998 von Bolthausen und Sznitman (siehe [18]) im Zusammenhang mit der Spin-Glass-Theorie in der Physik eingeführt. Beide Prozesse haben die besondere Eigenschaft, dass zu jedem Zeitpunkt $t \geq 0$ nur genau eine Kollision stattfinden kann. Solche Coalescent-Prozesse nennt man auch Coalescents mit multiplen Kollisionen. Unabhängig voneinander führten Pitman und Sagitov 1999 die Klasse aller Coalescents mit multiplen Kollisionen ein (siehe

[71] und [75]). Pitman zeigte, dass die Verteilung eines Coalescents mit multiplen Kollisionen eindeutig durch ein endliches Maß Λ auf $[0,1]$ charakterisiert werden kann. Man nennt Coalescents mit multiplen Kollisionen deshalb auch Λ-Coalescents. Dieses Maß Λ ist für den Kingman-Coalescent das Dirac-Maß in 0 und für den Bolthausen-Sznitman-Coalescent die Gleichverteilung auf $[0,1]$. Alle Coalescents, die mehr als eine Kollision pro Zeitpunkt zulassen, nennt man Coalescents mit simultanen multiplen Kollisionen.
Die Gesamtklasse aller Coalescent-Prozesse, die sowohl die Coalescents mit multiplen Kollisionen als auch die Coalescents mit simultanen multiplen Kollisionen beinhaltet, wurde unabhängig voneinander von Möhle und Sagitov und von Schweinsberg eingeführt (siehe [69] und [79]). Schweinsberg zeigte, dass die Verteilung eines Coalescents eindeutig durch ein endliches Maß Ξ auf dem unendlich dimensionalen Simplex

$$\Delta := \left\{ (x_i)_{i\in\mathbb{N}} \in \mathbb{R}^{\mathbb{N}} \middle| x_i \geq x_{i+1} \geq 0 \ \forall \ i \in \mathbb{N}, \sum_{i\in\mathbb{N}} x_i \leq 1 \right\}$$

charakterisiert wird. Deswegen werden die Coalescent-Prozesse auch Ξ-Coalescents genannt.
Eine Klasse der Ξ-Coalescents (bzw. der Λ-Coalescents) sind die Coalescents mit Staub. Dies sind all diejenigen Coalescents, deren charakterisierendes Maß Ξ bzw. Λ

$$\Xi(\{0\}) = 0 \ \text{ und } \ \mu_{-1} := \int_{\Delta\setminus\{0\}} \sum_{i\in\mathbb{N}} x_i \frac{\Xi(dx)}{\sum_{i\in\mathbb{N}} x_i^2} < \infty.$$

respektive

$$\Lambda(\{0\}) = 0 \ \text{ und } \ \mu_{-1} := \int_{(0,1]} x^{-1}\Lambda(dx) < \infty.$$

erfüllt. Dies ist äquivalent dazu, dass ein Coalescent Π genau dann ein Coalescent mit Staub ist, falls $P(S_t > 0) > 0$ für alle $t \geq 0$ gilt, wobei $S_t = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} 1_{\{\{i\} \text{ ist Block von } \Pi_t\}}$ der Anteil der natürlichen Zahlen ist, die in Π bis zum Zeitpunkt $t \geq 0$ nicht an einer Kollision teilgenommen haben (diese Größe existiert nach der Kingman-Darstellung austauschbarer Partitionen fast sicher). Sowohl der Kingman-Coalescent als auch der Bolthausen-Sznitman-Coalescent sind keine Coalescents mit Staub.
Eine weitere Klasse der Coalescents, die eine Unterklasse der Coalescents mit Staub sind, sind die Simple Coalescents, die von Bertoin und LeGall

in [12] benannt wurden. Ein Coalescent ist ein Simple Coalescent, falls das charakterisierende Maß Ξ bzw. Λ

$$\Xi(\{0\}) = 0 \quad \text{und} \quad \mu_{-2} := \int_{\Delta\setminus\{0\}} \frac{\Xi(dx)}{\sum_{i\in\mathbb{N}} x_i^2} < \infty.$$

respektive

$$\Lambda(\{0\}) = 0 \quad \text{und} \quad \mu_{-2} := \int_{(0,1]} x^{-2}\Lambda(dx) < \infty.$$

erfüllt.
Man kann einen beliebigen n-Coalescent als einen zufälligen Baum mit n Blättern und zufälligen Kantenlängen auffassen. Diese n-Coalescent-Bäume werden in der zumeist biologischen Anwendung als approximatives Modell des Stammbaums einer Stichprobe von n Individuen einer sehr großen eingeschlechtlichen Population mit nichtüberlappenden Generationen aufgefasst, die sich im biologischen Gleichgewicht befindet, d.h. deren Populationsgröße über die Zeit konstant bleibt. Die n Individuen der Stichprobe entsprechen gerade den Blättern des Baumes, die Vorfahren den Knoten und der jüngste gemeinsame Urahn (most recent common ancestor, kurz MRCA) entspricht der Wurzel des Baumes. Für biologische Fragestellungen wird auf den n-Coalescent-Baum eine neutrale Mutationsstruktur gesetzt, d.h. die Mutationen beeinflussen das Reproduktionsverhalten der Individuen nicht. Man setzt diese Mutationsstruktur auf den n-Coalescent-Baum, indem man auf dessen Zweigen gemäß eines unabhängigen homogenen Poisson-Punktprozesses mit Intensität $r > 0$ Mutationen (also Punkte des Poisson-Punktprozesses) setzt. Die Mutationen werden im Sinne des Infinitely-Many-Alleles-Modell interpretiert. Dies bedeutet, dass zwei Individuen der Stichprobe genau dann dieselbe genetische Struktur besitzen, falls im Stammbaum auf den jeweils direkten Wegen zum jüngsten Urahn exakt dieselben Mutationen liegen. Die genetische Struktur eines Individuums wird auch Typ genannt.
Im Rahmen dieser Arbeit werden bestimmte Kenngrößen von n-Coalescents untersucht, sogenannte Funktionale. Ein Funktional ordnet jedem Pfad eines n-Coalescents eine Kennzahl zu. Funktionale sind derzeit von starkem Interesse in der (mathematischen) Coalescent-Theorie. Hier werden insbesondere die Funktionale

- τ_n, die Zeit zurück zum jüngsten Urahn der n Individuen,
- E_n, die Länge eines zufällig ausgewählten externen Zweiges,

- C_n^{ext}, die Anzahl der Kollisionen im Baum bis zum Ende eines zufällig ausgewählten Zweiges und
- K_n, die Anzahl der verschiedenen Typen in der n-elementigen Stichprobe

betrachtet für einen n-Coalescent $\Pi^{(n)}$ (mit Mutation) beziehungweise für den durch $\Pi^{(n)}$ gegebenen Baum oder für eine Stichprobe von n Individuen, deren Stammbaum durch $\Pi^{(n)}$ (und eine neutrale Mutationstruktur) gegeben ist. Alle vier Funktionale sind für verschiedene Coalescents bereits gut untersucht. Da im Rahmen dieser Arbeit die Asymptotik der Funktionale für $n \to \infty$ im Vordergrund steht, sei insbesondere auf die Ergebnisse für die Asymptotik für $n \to \infty$ von

- τ_n im Kingman-n-Coalescent von Kingman in [57],
- τ_n und E_n im Bolthausen-Sznitman-n-Coalescent von Goldschmidt und Martin in [42],
- τ_n, E_n und C_n^{ext} in (bestimmten) Simple Λ-n-Coalescents von Gnedin, Iksanov und Möhle in [39],
- E_n und C_n^{ext} im Kingman-n-Coalescent von Caliebe, Krawczak, Neininger und Rösler in [22],
- K_n im Kingman-n-Coalescent von Ewens in [31],
- K_n im Bolthausen-Sznitman-n-Coalescent von Basdevant und Goldschmidt in [5] und
- K_n in Λ-n-Coalescents mit $\Lambda = \beta(2-\alpha, \alpha)$ von Berestycki, Berestycki und Schweinsberg in [8] und [9], wobei $\beta(2-\alpha, \alpha)$ die Beta-Verteilung mit Parametern $2-\alpha$ und α bezeichnet und $\alpha \in (1, 2)$ ist,

hingewiesen.
Im Rahmen dieser Arbeit werden für diese vier Funktionale Verteilungsrekursionen in n aufgestellt bzw. zitiert (falls schon bekannt), die in allen Ξ oder zumindest in allen Λ-Coalescents gelten. Dies ist aufgrund der speziellen Struktur der Coalescentprozesse und des Mutations(-Poisson-)prozesses möglich (auf Coalescent-Seite sind hier Natural und Temporal Coupling gemeint). Mit Hilfe dieser Verteilungsrekursionen werden die Aussagen

- $\tau_n - \log\log(n) \xrightarrow{d} G$, G (Standard-)Gumbel-verteilt,
- $\log(n)E_n \xrightarrow{d} Exp(1)$,
- $\frac{\log n}{n} C_n^{ext} \xrightarrow{d} U_{[0,1]}$, $U_{[0,1]}$ Gleichverteilung auf $[0,1]$,

für $n \to \infty$ für τ_n, E_n und C_n^{ext} in einem Bolthausen-Sznitman-n-Coalescent gezeigt (das Konvergenzresultat für $(E_n)_{n\in\mathbb{N}}$ wurde mit anderen Methoden bereits von Goldschmidt und Martin in [42] gezeigt). Dies wird mit einer Methode gezeigt, die von Drmota, Iksanov, Möhle und Rösler in [27] verwendet wurde, um die Asymptotik anderer Funktionale eines Bolthausen-Sznitman-n-Coalescents zu untersuchen. Die Ergebnisse werden auf folgende Art und Weise bewiesen: Aus der Verteilungsrekursion für das betrachtete Funktional wird eine Differentialgleichung einer erzeugenden Funktion aufgestellt, deren n-ter Koeffizient eine Verteilungsgröße von τ_n, E_n oder C_n^{ext} ist, etwa ein Moment oder die wahrscheinlichkeitserzeugende Funktion. Mittels Singularitätsananlyse kann man nun diese Koeffizienten für große $n \in \mathbb{N}$ asymptotisch bestimmen (zur Singularitätsanalyse siehe etwa [33]) und daraus die Verteilungskonvergenz zeigen.
Für Ξ-Coalescents Π mit Staub betrachte für jedes $n \in \mathbb{N}$ die Restriktion $\Pi^{(n)}$ auf die Partitionen $\mathbb{E}_n$ von $\{1, \ldots, n\}$. Sei K_n die Anzahl der verschiedenen Typen in einer n-elementigen Stichprobe, deren Stammbaum durch den n-Coalescent $\Pi^{(n)}$ mit Mutationsprozess (Mutationsrate $r > 0$) gegeben ist. Die Stichproben seien aufsteigend in n ineinander enthalten und die Mutationen seien so gesetzt, dass sie auf den gemeinsamen Zweigen der (Stamm-)Bäume $\Pi^{(n)}$ und $\Pi^{(n+1)}$ für jedes $n \in \mathbb{N}$ pfadweise übereinstimmen. In dieser Arbeit wird die Asymptotik von $(K_n)_{n\in\mathbb{N}}$ für $n \to \infty$ analysiert. Hierzu wird ein verwandtes Funktional, die Anzahl M_n der externen Zweige von $\Pi^{(n)}$ mit mindestens einer Mutation, betrachtet. Für beliebige Ξ-Coalescents Π wird

$$\frac{M_n}{n} \longrightarrow \int_0^\infty re^{-rt} S_t dt$$

fast sicher und in L^p, $p \geq 1$, für $n \to \infty$ gezeigt. Dazu wird der Zusammenhang zwischen den Längen der externen Zweige von $\Pi^{(n)}$ und dem Anteil S_t der nicht kollidierten Blöcke $\{i\}$ in Π_t für $i \in \mathbb{N}$ (hier geht die Kingman-Darstellung für austauschbare Partitionen ein) und das starke Gesetz der großen Zahlen für gewichtete i.i.d. Zufallsvariablen von Cuzick aus [25] verwendet.

Für Coalescents mit Staub wird gezeigt, dass $(K_n/n)_{n\in\mathbb{N}}$ dieselbe L^p-Asymptotik wie $(M_n/n)_{n\in\mathbb{N}}$ besitzt. Für Simple Coalescents wird dies zusätzlich für die fast sichere Asymptotik gezeigt. Dies wird unter Verwendung der für beliebige Ξ-Coalescents und beliebiges $n \in \mathbb{N}$ gültigen Ungleichung

$$0 \leq K_n - M_n \leq C_n$$

für das Funktional C_n, die Anzahl der Kollisionen in $\Pi^{(n)}$, gezeigt. Für $(C_n)_{n\in\mathbb{N}}$ wird dann $C_n/n \to 0$ in L^p, $p \geq 1$ bewiesen, falls Π ein Coalescent mit Staub ist und zusätzlich fast sichere Konvergenz, falls Π sogar ein Simple Coalescent ist. Diese Aussagen über $(C_n)_{n\in\mathbb{N}}$ werden mit Hilfe der Poisson-Konstruktion von Ξ-Coalescents von Schweinsberg aus [79] bewiesen.
Somit gilt für Ξ-Coalescents mit Staub auch

$$\frac{K_n}{n} \longrightarrow \int_0^\infty re^{-rt}S_t dt$$

für $n \to \infty$ in L^p, $p \geq 1$; für Simple Ξ-Coalescents gilt diese Konvergenz zusätzlich fast sicher. Die Grenzvariable $\int_0^\infty re^{-rt}S_t dt$ wird allgemein und in speziellen Coalescents (mit Staub) analysiert. Dazu wird verwendet, dass für einen Coalescent mit Staub der Prozess $(-\log(S_t))_{t\geq 0}$ eine Modifikation besitzt, die ein Subordinator mit exponentieller Vernichtung ist, d.h. ein Subordinator, der nach einer exponentialverteilten Wartezeit nach ∞ springt.
Die Hauptergebnisse und -argumentationen dieser Arbeit finden sich in den gemeinsamen Publikationen [35] und [36] des Autors mit M. Möhle sowie in dem eingereichten Manuskript [34] des Autors. Diese Ergebnisse sind in der vorliegenden Arbeit durch Zitation gekennzeichnet. Kapitel (bzw. Teilkapitel), die zum großen Teil diesen Arbeiten entstammen, sind zusätzlich an ihrem Anfang durch Zitation gekennzeichnet.
Die Gliederung der Arbeit ist die folgende: In Kapitel 1 werden die Grundlagen der Coalescent-Theorie vorgestellt, die in den weiteren Kapiteln benötigt werden. In Kapitel 2 werden die betrachteten Funktionale eingeführt und die jeweiligen Rekursionen aufgestellt. Kapitel 3 enthält die Ergebnisse zum Funktional τ_n, der Zeit zurück zum jüngsten Urahn eines n-Coalescents. In Kapitel 4 werden die Ergebnisse für E_n, die Länge eines zufällig ausgewählten externen Zweiges eines n-Coalescents, gezeigt und in Kapitel 5 diejenigen für C_n^{ext}, die Anzahl der Kollisionen vor dem Ende eines zufällig ausgewählten externen Zweiges eines n-Coalescents. Kapitel 6 beschäftigt sich mit dem

Funktional K_n, der Anzahl der Typen einer Stichprobe der Größe n, deren Genealogie durch einen n-Coalescent mit Mutation gegeben ist.

Notation

- $[n] := \{1, \dots, n\}$ für $n \in \mathbb{N}$, $[0] := \emptyset$, $[\infty] := \mathbb{N}$.
- $a_n \sim b_n$ für $n \to \infty$: $\lim_{n\to\infty} \frac{a_n}{b_n} = 1$ für reelle Folgen $(a_n)_{n\in\mathbb{N}}$, $(b_n)_{n\in\mathbb{N}}$.
- $b_n = O(a_n)$ für $n \to \infty$: $(b_n/a_n)_{n\in\mathbb{N}}$ ist beschränkt.
- $\delta_{ij} = 1_{\{j\}}(i)$: Kronecker-Symbol von i und j.
- $sign(\cdot)$: Signum-Funktion.
- $s(k, j)$: Stirling-Zahlen erster Art zu den Parametern $k, j \in \mathbb{N}$, $k \leq j$.
- $S(k, j)$: Stirling-Zahlen zweiter Art zu den Parametern $k, j \in \mathbb{N}$, $k \leq j$.
- Γ: Gammafunktion.
- $Beta(a, b)$: Betafunktion zu den Parametern $a > 0$, $b > 0$.
- $\mathcal{F} \otimes \mathcal{G}$: die von $\mathcal{F} \times \mathcal{G}$ erzeugte σ-Algebra für σ-Algebren $\mathcal{F}, \mathcal{G}$.
- $\mu \otimes \nu$: Produktmaß zweier Maße μ und ν.
- $\nu(dx) := f(x)\mu(dx)$: ν hat μ-Dichte f.
- δ_a: Dirac-Maß in $a \in \Omega$ für Messraum $(\Omega, \mathcal{F})$.
- $\beta(a, b)$: Beta-Verteilung zu den Parametern $a > 0$ und $b > 0$.
- $Bin(n, p)$: Binomialverteilung zu den Parametern $n \in \mathbb{N}$ und $p \in [0, 1]$.
- $Poiss(\lambda)$: Poisson-Verteilung zum Parameter $\lambda > 0$.
- $\stackrel{d}{=}$, $\stackrel{d}{\to}$: Gleichheit bzw. Konvergenz in Verteilung.
- $\stackrel{f.s.}{\to}$: fast sichere Konvergenz.
- $\stackrel{p}{\to}$: stochastische Konvergenz.
- $\mathbb{E}$: Menge der Partitionen von $\mathbb{N}$.
- $\mathbb{E}_n$: Menge der Partitionen von $\{1, \dots, n\}$, $n \in \mathbb{N}$.

- $\rho_n : \mathbb{E} \to \mathbb{E}_n$: natürliche Restriktion.
- $\rho_{nm} : \mathbb{E}_n \to \mathbb{E}_m$: natürliche Restriktion $(n, m \in \mathbb{N},\ m \leq n)$.
- $\mathsf{Diag}{:=}(\{i\})_{i\in\mathbb{N}}$: triviale Partition aus $\mathbb{E}$.
- $\mathsf{Diag}_n{:=}(\{i\}, \ldots, \{n\})$: triviale Partition aus $\mathbb{E}_n$
- $|\eta|$ für $\eta \in \mathbb{E}$ oder $\eta \in \mathbb{E}_n$: Anzahl der Blöcke von η.
- $\Delta := \left\{(x_i)_{i\in\mathbb{N}} | x_i \geq x_{i+1} \geq 0\ \forall\ i \in \mathbb{N}, \sum_{i\in\mathbb{N}} x_i \leq 1\right\}$: unendlich dimensionaler Simplex.
- $|x| := \sum_{i\in\mathbb{N}} x_i$ für $x \in \Delta$.
- $\Delta^* := \{x \in \Delta \mid |x| = 1\}$.
- $(x, x) := \sum_{i\in\mathbb{N}} x_i^2$ für $x \in \Delta$.
- $\mu_{-1} := \int_{\Delta\setminus\{(0,0,\ldots)\}} |x| \frac{\Xi(dx)}{(x,x)}$ für ein endliches Maß Ξ auf Δ.
- $\mu_{-2} := \int_{\Delta\setminus\{(0,0,\ldots)\}} \frac{\Xi(dx)}{(x,x)}$ für ein endliches Maß Ξ auf Δ.

1 Coalescent-Prozesse

In diesem Kapitel werden die für die Fragestellungen dieser Arbeit nötigen Begriffe und Methoden aus der Coalescent-Theorie vorgestellt.

1.1 Grundlagen

In dieser Arbeit werden stochastische Prozesse betrachtet, die Werte in der Menge $\mathbb{E}$ der Partitionen der natürlichen Zahlen oder in der Menge $\mathbb{E}_n$ der Partitionen von $[n] := \{1, \ldots, n\}$, $n \in \mathbb{N}$ annehmen (es wird $[0] := \emptyset$ und $[\infty] := \mathbb{N}$ gesetzt).

Notation 1.1.1 *Sei $\eta \in \mathbb{E}_n$ ($\eta \in \mathbb{E}$). Setze $\eta = (A_1, \ldots, A_k)$ für $k \in [n]$ (für $k \in \mathbb{N} \cup \{\infty\}$), wobei $A_1, \ldots, A_k$ ($A_1, A_2, \ldots$, falls es unendlich viele Klassen gibt) die Klassen der η erzeugenden Äquivalenzrelation sind. Die Klassen sind in aufsteigender Reihenfolge ihrer kleinsten Elemente sortiert. Die Klassen $A_1, \ldots, A_k$ ($A_1, A_2, \ldots$, falls es unendlich viele Klassen gibt) werden als Blöcke der Partition η bezeichnet.*

Es folgen einige wichtige Definitionen und Eigenschaften von $\mathbb{E}_n$ und $\mathbb{E}$.

Definition 1.1.2 *(Restriktionen von Partitionen) Für $n \in \mathbb{N}$ ist die natürlichen Restriktion $\rho_n : \mathbb{E} \to \mathbb{E}_n$ als Einschränkung der Partitionen aus $\mathbb{E}$ auf die Menge $[n]$ definiert, d.h. für $\eta = (A_1, \ldots, A_k) \in \mathbb{E}$ (k darf den Wert ∞ annehmen) gilt*

$$\rho_n(\eta) = (A_i \cap [n])_{i \in [k], A_i \cap [n] \neq \emptyset}.$$

Des Weiteren betrachte Partitionen $\eta = (A_1, \ldots, A_k) \in \mathbb{E}$, $\eta^{(n)} = (B_1, \ldots, B_{k_n}) \in \mathbb{E}_n$ für $n \in \mathbb{N}$ und eine Menge $M \subseteq [n]$. Sei außerdem $m \in [n]$. Dann sind die Restriktionen ρ_M, $\rho_{n,M}$ und ρ_{nm} definiert durch

$$\rho_M(\eta) := (A_i \cap M)_{i \in [k], A_i \cap M \neq \emptyset}, \quad \rho_{n,M}(\eta^{(n)}) := (B_i \cap M)_{i \in [k_n], B_i \cap M \neq \emptyset}$$

und $\rho_{nm}(\eta^{(n)}) := \rho_{n,[m]}(\eta^{(n)})$.

Notation 1.1.3 *(Diagonale Relation)*
$\mathsf{Diag} := (\{i\} : i \in \mathbb{N}) \in \mathbb{E}$ *und* $\mathsf{Diag}_n := (\{i\} : i \in [n]) \in \mathbb{E}_n$ *für $n \in \mathbb{N}$.*

Bemerkung 1.1.4 *(Topologie und σ-Algebren auf $\mathbb{E}_n$ und $\mathbb{E}$) Für jedes $n \in \mathbb{N}$ wird die Menge $\mathbb{E}_n$ mit der diskreten Topologie versehen, also sind alle ihre Teilmengen offen. Man identifiziert $e \in \mathbb{E}$ mit $(\rho_n(e))_{n\in\mathbb{N}}$ und somit $\mathbb{E}$ als abgeschlossene Teilmenge von $\times_{n\in\mathbb{N}}\mathbb{E}_n$. Die Menge $\times_{n\in\mathbb{N}}\mathbb{E}_n$ wird mit der Produkttopologie versehen. Nun wird $\mathbb{E}$ mit der auf $\mathbb{E}$ zurückgezogenen Spurtopologie $\mathcal{O}$ der Topologie auf $\times_{n\in\mathbb{N}}\mathbb{E}_n$ versehen. Diese ist gerade die gröbste Topologie auf $\mathbb{E}$, bezüglich der alle Restriktionen ρ_n, $n \in \mathbb{N}$, stetig sind. Desweiteren gibt es eine Metrik auf $\mathbb{E}$, die $\mathcal{O}$ erzeugt und bezüglich der E polnisch ist (Metrik siehe [6, S.7]).*
Als σ-Algebren werden die von den offenen Mengen erzeugten σ-Algebren gesetzt, also $\mathcal{E}_n := \wp(\mathbb{E}_n)$ auf $\mathbb{E}_n$ und $\mathcal{E} := \mathcal{B}(\mathcal{O})$ auf $\mathbb{E}$.

Bemerkung 1.1.5 *Im Sinne der Identifikation von $\mathbb{E}$ mit einer Teilmenge von $\times_{n\in\mathbb{N}}\mathbb{E}_n$ kann man die Restriktionen ρ_n, $n \in \mathbb{N}$, auch als kanonische Projektionen von $\mathbb{E}$ nach $\mathbb{E}_n$ bezeichnen.*

Definition 1.1.6 *Sei $|\eta|$ für $\eta \in \mathbb{E}$ bzw. $\eta \in \mathbb{E}_n$ die Anzahl der Blöcke von η. $|\cdot|$ bildet nach $\mathbb{N} \cup \{\infty\}$ bzw. $[n]$ ab und ist messbar.*

Nun lässt sich die Klasse der Coalescent-Prozesse definieren, deren Eigenschaften der Gegenstand dieser Arbeit sind.

Definition 1.1.7 *Ein Coalescent-Prozess $\Pi := (\Pi_t)_{t\geq 0}$, kurz <u>Coalescent</u>, ist ein stochastischer Prozess mit Zustandsraum $(\mathbb{E}, \mathcal{E})$ und Startverteilung δ_{Diag} sowie folgenden Eigenschaften:*

1) *Die Pfade von Π sind càdlàg.*
2) *Übergänge sind nur möglich durch Vereinigung von Blöcken des derzeitigen Zustandes.*
3) *Für jedes $n \in \mathbb{N}$ ist $(\rho_n(\Pi_t))_{t\geq 0}$ ein Markov-Prozess. Jeder Übergang von $\eta \in \mathbb{E}_n$ nach $\xi \in \mathbb{E}_n$, der durch Vereinigen von (disjunkten) Mengen von $k_1 \geq \ldots \geq k_m \geq 2$ der $b := |\eta|$ Blöcken von η zu jeweils einem Block von ξ und dem Beibehalten der restlichen $r := b - \sum_{i\in m} k_i$ Blöcken von η entsteht, besitzt die infinitesimale Rate $\lambda(b; k_1, \ldots, k_m)$. Ein solcher Übergang heißt $(b; k_1, \ldots, k_m)$-Kollision.*

Bemerkung 1.1.8

- *Da für jedes $n \in \mathbb{N}$ der Prozess $(\rho_n(\Pi_t))_{t\geq 0}$ ein Markov-Prozess ist, ist auch $(\Pi_t)_{t\geq 0}$ Markov'sch. Dies folgt aus dem Setzen der Produkttopologie (und der zugehörigen σ-Algebra) auf $\mathbb{E}$ bzw. $\times_{n\in\mathbb{N}}\mathbb{E}_n$.*

- *Bedingung 3 aus dem vorhergehenden Satz impliziert, dass Π_t für jedes $t \geq 0$ eine austauschbare Partition von $\mathbb{N}$ ist, d.h. dass $\rho_n(\Pi_t) \stackrel{d}{=} \sigma \circ \rho_n(\Pi_t)$ für jedes $n \in \mathbb{N}$ und $\sigma \in S_n$ gilt, wobei σ in natürlicher Weise auf $\mathbb{E}_n$ wirkt (siehe etwa Bemerkung nach [79, Definition 19]). Dies bedeutet, dass $\rho_n(\Pi_t)$ für jedes $n \in \mathbb{N}$ eine austauschbare Partition von $[n]$ ist für $t \geq 0$. Es gilt sogar $(\rho_n(\Pi_t))_{t\geq 0} \stackrel{d}{=} (\sigma \circ \rho_n(\Pi_t))_{t\geq 0}$ für alle $n \in \mathbb{N}$. Dies bezeichnet man auch als die Austauschbarkeit des Prozesses Π (bzw. $(\rho_n(\Pi_t)_{t\geq 0})$).*

- *Man bezeichnet in einer $(b; k_1, \ldots, k_m)$-Kollision jede der Vereinigungen von k_i Blöcken, $i \in [m]$, zu einem neuen Block als eine Kollision, d.h. eine $(b; k_1, \ldots, k_m)$-Kollision besteht aus m Kollisionen.*

Die Verteilung einer austauschbaren Partition von $\mathbb{N}$ lässt sich mittels der Kingman'schen Darstellung beschreiben (siehe etwa [55], [56]. Ein schöner Beweis findet sich in [2, Proposition 11.9, S.88]). Es wird die Formulierung der Kingman-Darstellung aus [72, Theorem 2.2, S.43] verwendet.

Satz 1.1.9 *(Kingman'sche Darstellung) Sei η eine austauschbare Partition von $\mathbb{N}$. Sei $n \in \mathbb{N}$ und seien $\mathsf{Bl}_{n,i}$, $i \in \mathbb{N}$, die absteigend geordneten Mächtigkeiten der Blöcke von $\rho_n(\eta)$, wobei $\mathsf{Bl}_{n,i} = 0$ gesetzt wird, falls $\rho_n(\eta)$ weniger als i Blöcke hat. Dann existieren die Grenzwerte $A_i := \lim_{n\to\infty} \mathsf{Bl}_{n,i}/n$ fast sicher für jedes $i \in \mathbb{N}$. Die bedingte Verteilung $P_\eta(\cdot|(A_i)_{i\in\mathbb{N}})$ entsteht bedingt $(A_i)_{i\in\mathbb{N}}$ durch Sampling aus einer Verteilung mit nach Größe geordneten Atomen $(A_i)_{i\in\mathbb{N}}$.*

Bemerkung 1.1.10 *Man gewinnt eine zufällige Partition η von $\mathbb{N}$ durch Sampling aus einer Verteilung Q, indem man eine i.i.d. Folge von Zufallsvariablen $X_1, X_2, \ldots$ mit $X_1 \stackrel{d}{=} Q$ erzeugt und die Partition η bezüglich der zufälligen Äquivalenzrelation*

$$i \sim j \quad \Leftrightarrow \quad X_i = X_j$$

für $i, j \in \mathbb{N}$ bildet. Eine genauere Beschreibung des Samplings aus Satz 1.1.9 findet man in [72, S.43].

Die Klasse aller Coalescent-Prozesse wurde unabhängig voneinander von Möhle und Sagitov [69] und Schweinsberg [79] eingeführt.
Da ein Coalescent càdlàg Pfade auf dem polnischen Raum $\mathbb{E}$ besitzt, bestimmt die Angabe der infinitesimalen Raten die gesamte Verteilung des Prozesses. Die Raten lassen sich mittels endlichen Maßen Ξ auf dem unendlich dimensionalen Simplex

$$\Delta := \left\{ (x_i)_{i\in\mathbb{N}} \in \mathbb{R}^{\mathbb{N}} \middle| x_i \geq x_{i+1} \geq 0 \ \forall\, i \in \mathbb{N}, \sum_{i\in\mathbb{N}} x_i \leq 1 \right\}$$

eindeutig beschreiben.

Bemerkung 1.1.11 *Δ ist eine abgeschlossene Teilmenge der Menge $[0,1]^{\mathbb{N}}$, wenn diese mit der Produkttopologie versehen wird ($[0,1]$ wird mit der Topologie der euklidischen Norm versehen). Somit ist Δ kompakt. Auf Δ wird die Spurtopologie gesetzt und als σ-Algebra die zugehörige Borel'sche σ-Algebra gesetzt. Aus Notationsgründen wird $\Xi_0 := 1_{\{\Delta\setminus\{0\}\}}\Xi$ gesetzt. Der Punkt $(0,0,\ldots) \in \Delta$ wird mit 0 bezeichnet. Desweiteren sei für $x \in \Delta$*

$$|x| := \sum_{i\in\mathbb{N}} x_i, \quad (x,x) := \sum_{i\in\mathbb{N}} x_i^2$$

definiert und $\Delta^* := \{x \in \Delta \,|\, |x| = 1\}$.

Aus [79, Theorem 2 u. Proposition 4] stammt folgende Charakterisierung der Coalescent-Prozesse.

Theorem 1.1.12 *Es gibt genau dann einen Coalescent-Prozess Π (auf einem geeigneten Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, P)$) mit den Eigenschaften aus 1.1.7, falls die Übergangsraten die Form*

$$\lambda(b; k_1, \ldots, k_m) =$$

$$\int_\Delta \Big(\sum_{l=0}^{r} \sum_{i_1 \neq \cdots \neq i_{m+l}} \binom{r}{l} \prod_{j=1}^{m} x_{i_j}^{k_j} \prod_{o=1}^{l} x_{i_{m+o}} (1-|x|)^{r-l} \Big) \frac{\Xi_0(dx)}{(x,x)} + a 1_{\{m=1, k_1=2\}} \quad (1)$$

für ein endliches Maß Ξ auf Δ besitzen, wobei $b \in \mathbb{N}$, $m \in [b]$, $k_1 \geq \ldots \geq k_m \geq 2$, $k_1, \ldots, k_m \in \mathbb{N}$ und $b \geq \sum_{i\in[m]} k_i$ sowie $r := b - \sum_{i\in[m]} k_i$ gilt. Das Maß Ξ ist durch die Raten eindeutig bestimmt. Man nennt einen solchen Prozess auch Ξ-Coalescent oder Coalescent mit simultanen multiplen Kollisionen.

Für die Teilklasse der Coalescentprozesse, die nur Übergänge von $\eta \in \mathbb{E}$ nach $\xi \in \mathbb{E}$ zulassen, die durch Vereinigen einer Menge von $k_1 \geq 2$ Blöcken von η entstehen, kann man die Raten eines Coalescent-Prozesses auch eindeutig durch ein endliches Maß Λ auf $([0,1], \mathcal{B} \cap [0,1])$ beschreiben. Die Klasse dieser Prozesse wurde unabhängig voneinander von Pitman [71] und Sagitov [75] eingeführt. Nach [71, Theorem 1] gilt

Theorem 1.1.13 *Es gibt genau dann einen Coalescent-Prozess* Π *(auf einem geeigneten Wahrscheinlichkeitsraum* $(\Omega, \mathcal{F}, P)$*) mit den Eigenschaften aus 1.1.7 und nur einer möglichen Kollision pro Übergang (d.h.* $\lambda(b; k_1, \ldots, k_m) = 0$ *falls* $m \geq 2$*), falls es ein endliches Maß* Λ *auf* $[0,1]$ *gibt mit*

$$\lambda(b; k_1) = \int_{[0,1]} x^{k_1-2}(1-x)^{b-k_1}\Lambda(dx) \text{ für } k_1, b \in \mathbb{N},\ b \geq k_1 \geq 2. \tag{2}$$

Das Maß Λ *ist eindeutig durch die Raten bestimmt. Man nennt einen solchen Prozess* Λ*-Coalescent oder Coalescent mit multiplen Kollisionen.*

Natürlich lässt sich jeder Λ-Coalescent als Ξ-Coalescent auffassen. Aus der Darstellung der Raten aus Theorem 1.1.12 erhält man folgende Beschreibung der Λ-Coalescents als Ξ-Coalescents.

Bemerkung 1.1.14 *Ein* Ξ*-Coalescent ist genau dann auch ein* Λ*-Coalescent, wenn* Ξ *auf* $([0,1], 0, 0, \ldots)$ *konzentriert ist. Dann gilt* $\Lambda(A) = \Xi(A, 0, 0, \ldots)$*. Dies folgt direkt aus der Darstellung der Raten in Theorem 1.1.12.*

Der bekannteste Λ-Coalescent ist der Kingman-Coalescent mit $\Lambda = \delta_0$, dem Punktmaß in 0, der nur binäre Kollisionen, d.h. (2)-Kollisionen zulässt. Dieser Coalescent ist der Grundstein der Coalescent-Theorie und wurde in [56] eingeführt. Ein weiterer gut untersuchter Λ-Coalescent ist der Bolthausen-Sznitman-Coalescent mit $\Lambda = U_{[0,1]}$, der Gleichverteilung auf $[0,1]$ aus [18]. Der Bolthausen-Sznitman-Coalescent liegt in der Klasse der Beta-Coalescents; dies sind Λ-Coalescents mit $\Lambda = \beta(a,b)$, $a, b > 0$, wobei $\beta(a,b)$ die Beta-Verteilung mit Parametern a, b bezeichnet. Für $a = 1, b = 1$ erhält man einen Bolthausen-Sznitman-Coalescent.
Eine Klasse von Ξ-Coalescent-Prozessen, die keine Λ-Coalescents sind, d.h. die mit positiver Wahrscheinlichkeit Übergänge mit mehreren (simultanen)

Kollisionen besitzen, ist die Familie der (Zwei-Parameter-) Poisson-Dirichlet-Coalescents, d.h. die Familie von Ξ-Coalescents, deren Verteilung charakterisiert wird durch $\Xi(dx) = (x,x)\mathsf{PD}_{\alpha,\theta}(dx)$, wobei $\mathsf{PD}_{\alpha,\theta}$ eine 2-Parameter Poisson-Dirichlet-Verteilung für $0 \leq \alpha < 1$ und $\theta > -\alpha$ ist (Informationen zu dieser Verteilung finden sich etwa in [73] und [44]). Diese Klasse von Coalescents wurde zuerst für die Parameter-Paare $\alpha = 0,\ \theta > 0$ in [76] beschrieben.
Eine weitere Klasse von Coalescents sind die Dirac-Coalescents, die durch $\Xi = \delta_x$ für $x = (x_1, x_2, \ldots) \in \Delta \setminus \{0\}$ charakterisiert werden bzw. für $x \in ((0,1], 0, 0, \ldots)$ auch durch $\Lambda = \delta_{x_1}$. Diese beinhaltet den star-shaped Coalescent mit $\Lambda = \delta_1$, bei dem nach einer $Exp(1)$-verteilten Wartezeit alle Singletons zu einem Block kollidieren. Den mit Abstand einfachsten Coalescent erhält man allerdings, wenn man für Ξ das Nullmaß auf Δ wählt. Ein hierzu gehörender Coalescent Π erfüllt $\Pi_t = \mathsf{Diag}$ fast sicher für alle $t \in [0,\infty)$, es gibt also überhaupt keine Übergänge. Dieser Coalescent ist eigentlich uninteressant, allerdings gelten für ihn einige Eigenschaften nicht, die für alle anderen Coalescents gelten. Daher wird dieser Spezialfall in dieser Arbeit nicht betrachtet, d.h. für alle betrachteten Ξ-Coalescent-Prozesse gilt $\Xi(\Delta) > 0$.

Bemerkung 1.1.15 *Sowohl der Kingman-Coalescent als auch der Bolthausen-Sznitman-Coalescent wurden bereits eingeführt, bevor die gesamte Klasse der Λ-Coalescents und der Ξ-Coalescents durch Sagitov, Pitman (Λ-Coalescents) sowie durch Möhle und Sagitov, Schweinsberg (Ξ-Coalescents) eingeführt wurde.*

Für die in dieser Arbeit diskutierten Fragestellungen lohnt es, die Klasse aller Ξ-Coalescent-Prozesse in zwei Unterklassen aufzuteilen. Diese Unterscheidung hängt vom Anteil der Singletons von Π_t an $\mathbb{N}$ ab. $i \in \mathbb{N}$ heißt Singleton von Π_t, $t \geq 0$, falls $\{i\}$ ein Block von Π_t ist.

Definition/Satz 1.1.16 *(Anteil der Singletons) Sei Π ein Ξ-Coalescent auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, P)$. Dann existiert für beliebiges $t \geq 0$*

$$S_t := \limsup_{n\to\infty} \frac{1}{n} \sum_{i\in[n]} 1_{\{\{i\} \text{ ist Block von } \rho_n(\Pi_t)\}} \tag{3}$$

und es gilt

$$S_t := \lim_{n\to\infty} \frac{1}{n} \sum_{i\in[n]} 1_{\{\{i\} \text{ ist Block von } \rho_n(\Pi_t)\}} \ P\text{-fast sicher.} \tag{4}$$

S_t heißt Anteil der Singletons von Π_t.

Beweis: Sei $t \geq 0$. Da $\frac{1}{n}\sum_{i\in[n]} 1_{\{\{i\} \text{ ist Block von } \rho_n(\Pi_t)\}} \in [0,1]$ für alle $n \in \mathbb{N}$ gilt, existiert $S_t \in [0,1]$. Nach Bemerkung 1.1.8 ist Π_t eine austauschbare Partition von $\mathbb{N}$. Somit gilt mit der Kingman'schen Darstellung aus Satz 1.1.9 mit den dort eingeführten Zufallsvariablen $(A_i)_{i\in\mathbb{N}}$ (diese existieren fast sicher), dass bedingt $(A_i)_{i\in\mathbb{N}}$ die Verteilung von Π_t durch Sampling aus einer Verteilung mit Atomen $(A_i)_{i\in\mathbb{N}}$ entsteht. Daraus folgt, dass bedingt $(A_i)_{i\in\mathbb{N}}$ fast sicher alle $i \in \mathbb{N}$, die nicht in einem Block mit Frequenz $A_j > 0$ für ein $j \in \mathbb{N}$ liegen, Singletons sind. Somit kommt bedingt $(A_i)_{i\in\mathbb{N}}$ fast sicher ein Anteil von $1-\sum_{i\in\mathbb{N}} A_i$ Singletons in Π_t vor. Die Frequenzen $(A_i)_{i\in\mathbb{N}}$ existieren fast sicher, es gilt also mit der Notation aus Satz 1.1.9

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i\in[n]} 1_{\{\{i\} \text{ ist Block von } \rho_n(\Pi_t)\}} = 1 - \sum_{i\in\mathbb{N}} A_i \text{ fast sicher.}$$

Dies zeigt die Behauptung. □

Mehr Informationen zum Prozess $(S_t)_{t\geq 0}$ werden in Abschnitt 2.3 gegeben. Nun kann man folgende Klassen von Coalescentprozessen einführen.

Definition 1.1.17 *(Coalescents mit/ohne Staub) Ein Ξ-Coalescent Π hat Staub bzw. heißt Coalescent mit Staub (ohne proper frequencies), falls*

$$P(S_t > 0) > 0 \text{ für alle } t \geq 0. \tag{5}$$

Ansonsten heißt Π ein Coalescent ohne Staub bzw. staubfrei (mit proper frequencies).

Diese Unterscheidung lässt sich durch Bedingungen an das Maß Ξ (im Falle eines Λ-Coalescents an das Maß Λ) charakterisieren. Diese Charakterisierung stammt aus [71, Theorem] (Λ-Coalescent) und [79, Proposition 30] (Ξ-Coalescent).

Lemma 1.1.18 *Ein Ξ-Coalescent ist genau dann ein Coalescent mit Staub, falls*

$$\Xi(\{0\}) = 0 \text{ und } \mu_{-1} := \int_{\Delta\setminus\{0\}} |x| \frac{\Xi(dx)}{(x,x)} < \infty. \tag{6}$$

Ein Λ-Coalescent ist genau dann ein Coalescent mit Staub, falls

$$\Lambda(\{0\}) = 0 \text{ und } \mu_{-1} := \int_{(0,1]} x^{-1}\Lambda(dx) < \infty. \tag{7}$$

Für einen Ξ- oder Λ-Coalescent ohne Staub setze man $\mu_{-1} = \infty$.

Nun lassen sich die oben eingeführten Coalescent-Prozesse auf diese beiden Klassen aufteilen.

Beispiele 1.1.19

a) Der Kingman-Coalescent und der Bolthausen-Sznitman-Coalescent sind Coalescents ohne Staub.

b) Beta-Coalescents ($\Lambda = \beta(a, b)$) sind für $0 < a \leq 1$, $b > 0$ staubfrei und haben für $a > 1, b > 0$ Staub. Für $a > 1$, $b > 0$ gilt $\mu_{-1} = 1 + \frac{b}{a-1}$.

c) Jeder Ξ-Coalescent mit $\Xi(dx) = (x, x)\nu(dx)$ für ein endliches Maß ν auf Δ hat Staub.

d) Alle Poisson-Dirichlet-Coalescents und Dirac-Coalescents haben Staub. Für den (α, θ)-Poisson-Dirichlet-Coalescent ist $\mu_{-1} = 1$ und für $\Xi = \delta_c$ gilt $\mu_{-1} = |c|/(c, c)$.

Beweis: zu b): Für $\Lambda = \beta(a, b)$ mit $a, b > 0$ gilt $\Lambda(\{0\}) = 0$ und

$$\mu_{-1} = \frac{1}{Beta(a,b)} \int x^{a-2}(1-x)^{b-1}dx = \begin{cases} \infty & \text{falls } a \leq 1, \\ \frac{Beta(a-1,b)}{Beta(a,b)} = 1 + \frac{b}{a-1} < \infty & \text{falls } a > 1, \end{cases}$$

wobei *Beta* die Beta-Funktion bezeichnet.
zu a): Der Bolthausen-Sznitman-Coalescent ist ein $\beta(1, 1)$-Coalescent; somit folgt die Aussage aus b). Für den Kingman-Coalescent ist $\Lambda(\{0\}) = 1 \neq 0$ nicht erfüllt.
zu c): Es gilt $\Xi(\{0\}) = 0 \cdot \nu(\{0\}) = 0$ und $\mu_{-1} := \int_{\Delta\setminus\{0\}} |x|\nu(dx) \leq \nu(\Delta) < \infty$.
zu d): Für einen Poisson-Dirichlet-Coalescent hat Ξ die Dichte $x \to (x, x)$ bzgl. einer Poisson-Dirichlet-Verteilung, somit folgt die Aussage aus c). Da $\mathsf{PD}_{\alpha,\theta}$ ein Wahrscheinlichkeitsmaß ist, dessen Masse auf Δ^* konzentriert ist, folgt

$$\mu_{-1} = \int_{\Delta\setminus\{0\}} |x| \frac{\Xi(dx)}{(x,x)} = \int_{\Delta^*} 1\mathsf{d}PD_{\alpha,\theta} = 1.$$

Im Falle eines Dirac-Coalescents mit $\Xi = \delta_c$, $c \in \Delta \setminus \{0\}$ hat Ξ keine Masse in 0 und es gilt $\mu_{-1} = \frac{|c|}{(c,c)} < \infty$. □

Eine Unterklasse der Coalescents mit Staub sind die Simple Coalescents, erstmals benannt von Bertoin und LeGall in [12]. Diese sind durch folgende Bedingung definiert.

Definition 1.1.20 *(Simple Coalescent) Ein Ξ-Coalescent heißt Simple Coalescent, falls*

$$\Xi(\{0\}) = 0 \textit{ und } \mu_{-2} := \int_{\Delta\setminus\{0\}} \frac{\Xi(dx)}{(x,x)} < \infty. \tag{8}$$

Ein Λ-Coalescent heißt Simple Coalescent, falls

$$\Lambda(\{0\}) = 0 \textit{ und } \mu_{-2} := \int_{(0,1]} x^{-2}\Lambda(dx) < \infty. \tag{9}$$

Aus Satz 1.1.19 c) folgt, dass Simple Coalescents immer Staub haben. Somit sind alle Coalescents ohne Staub auch keine Simple Coalescents, etwa der Kingman-Coalescent oder der Bolthausen-Sznitman-Coalescent. Es folgen einige Beispiele für Simple Coalescents, wobei die Bedingungen (8) bzw. (9) analog zu Beispiel 1.1.19 nachgerechnet werden können.

Beispiele 1.1.21

a) Poisson-Dirichlet-Coalescents und Dirac-Coalescents sind Simple Coalescents. Für $\Xi = (x,x)\mathsf{PD}_{\alpha,\theta}$ gilt $\mu_{-2} = 1$, für $\Xi = \delta_c$, $c \in \Delta\setminus\{0\}$, gilt $\mu_{-2} = (c,c)^{-1}$.

b) Ein Beta-Coalescent ist genau dann ein Simple Coalescent, falls $\Lambda = \beta(a,b)$ mit $a > 2$ und $b > 0$. In diesem Fall gilt $\mu_{-2} = \frac{(a+b-1)(a+b-2)}{(a-1)(a-2)}$.

1.2 n-Coalescents, Natural Coupling und Temporal Coupling

Betrachte für festes $n \in \mathbb{N}$ für einen beliebigen (Ξ-)Coalescent-Prozess $(\Pi_t)_{t\geq 0}$ die Einschränkung

$$\Pi^{(n)} = (\Pi_t^{(n)})_{t\geq 0} := (\rho_n(\Pi_t))_{t\geq 0}$$

auf $[n]$ bzw. $\mathbb{E}_n$. Man nennt jeden càdlàg Prozess mit derselben Verteilung wie $\Pi^{(n)}$ einen (Ξ-)n-Coalescent ($(\rho_n(\Pi_t))_{t\geq 0}$ selbst ist aufgrund der Stetigkeit der

Restriktionen ρ_n càdlàg). n-Coalescents tauchen in natürlicher Weise schon bei der Definition 1.1.7 in Eigenschaft 3 auf. Aus Definition 1.1.7 erhält man sofort, dass ein n-Coalescent ein Markov-Prozess mit Zuständen in $\mathbb{E}_n$ ist, Startverteilung δ_{Diag_n} und die Übergangsraten aus (1) (Ξ-n-Coalescent) bzw. aus (2) (Λ-n-Coalescent) besitzt. Jeder n-Coalescent besitzt folgende Eigenschaft.

Satz 1.2.1 *(Natural Coupling) Sei Ξ ein endliches Maß auf Δ. Seien $n \in \mathbb{N}$ und $m \in [n]$. Betrachte einen Ξ-n-Coalescent $\Pi^{(n)}$ und einen Ξ-m-Coalescent $\Pi^{(m)}$. Sei $M \subseteq [n]$ mit $|M| = m$. Dann gilt*

$$(\rho^*_M(\Pi^{(n)}_t))_{t\geq 0} \stackrel{d}{=} (\Pi^{(m)}_t)_{t\geq 0},$$

mit $\rho^*_M := r \circ \rho_{n,M}$, wobei $r : M \to [m]$ die messbare Abbildung ist, die den i-t größten Wert von M auf i abbildet.

Beweis: Sei $(\Pi_t)_{t\geq 0}$ ein Ξ-Coalescent, es gilt also $(\rho_n(\Pi_t))_{t\geq 0} \stackrel{d}{=} \Pi^{(n)}$ und $(\rho_m(\Pi_t))_{t\geq 0} \stackrel{d}{=} \Pi^{(m)}$. Es gilt auch $(\rho_{n,M'}(\Pi^{(n)}_t))_{t\geq 0} \stackrel{d}{=} (\rho_{M'}(\Pi_t))_{t\geq 0}$ für jedes $M' \subseteq [n]$. Da $(\Pi_t)_{t\geq 0}$ austauschbar ist, gilt außerdem $(r \circ \rho_M(\Pi_t))_{t\geq 0} \stackrel{d}{=} (\rho_m(\Pi_t))_{t\geq 0}$, also insbesondere

$$(r \circ \rho_{n,M}(\Pi^{(n)}_t))_{t\geq 0} \stackrel{d}{=} (r \circ \rho_M(\Pi_t))_{t\geq 0} \stackrel{d}{=} (\Pi^{(m)}_t)_{t\geq 0}. \qquad \square$$

Bemerkung 1.2.2 *Da $\Pi^{(n)}$ jedes $\eta \in \mathbb{E}_n$ nur höchstens einmal als Zustand annehmen kann und $\mathbb{E}_n$ endlich ist, endet (fast) jeder Pfad im absorbierenden Zustand ($[n]$).*

$(\Pi^{(n)}_t)_{t\geq 0}$ ist ein càdlàg Markov-Prozess mit endlichem Zustandsraum $\mathbb{E}_n$. Somit gilt die starke Markov-Eigenschaft für die (fast sicher endliche) Stoppzeit T_n des ersten Sprungs in $\Pi^{(n)}$. Dies führt zu einer weiteren Eigenschaft eines n-Coalescents.

Satz 1.2.3 *(Temporal Coupling) Sei Ξ ein endliches Maß auf Δ. Für jedes $n \in \mathbb{N}$ sei $\Pi^{(n)}$ ein Ξ-n-Coalescent. Sei T_n die Wartezeit auf den ersten Sprung in $\Pi^{(n)}$. Dann gilt*

$$(r' \circ \Pi^{(n)}_{T_n+t})_{t\geq 0} \stackrel{d}{=} (\Pi^{(\mathcal{M})}_t)_{t\geq 0}, \tag{10}$$

wobei

- *$\mathcal{M}$ eine von $(\Pi^{(n)})_{n\in\mathbb{N}}$ unabhängige Zufallsvariable mit $\mathcal{M} \overset{d}{=} |\Pi^{(n)}_{T_n}|$ ist.*
- *r' die Abbildung ist, die eine Partition $\eta = (\cup_{i\in M_j} A_i)_{j\in[l]}$ von $[n]$, die durch Vereinigen von Mengen $M_1, \ldots, M_l$ der Blöcke $(A_1, \ldots, A_k)$ von $\Pi^{(n)}_{T_n}$ entsteht, auf die Partition von $[k]$ mit den Blöcken $(M_1, \ldots, M_l)$ abbildet (in Reihenfolge der kleinsten Elemente).*

Dies bedeutet anschaulich, dass sich der Prozess $(\Pi^{(n)}_{T_n+t})_{t\geq 0}$ wie ein $\mathcal{M}$-Coalescent verhält. Gegeben $\Pi^{(n)}_{T_n}$ sind $(\Pi^{(n)}_{T_n+t})_{t\geq 0}$ und $(\Pi^{(n)}_t)_{0\leq t\leq T_n}$ unabhängig.

Beweis: Nach der starken Markov-Eigenschaft ist $(\Pi^{(n)}_{T_n+t})_{t\geq 0}$ ein Markov-Prozess mit Start in $\Pi^{(n)}_{T_n}$ und denselben Übergangsraten wie $\Pi^{(n)}$. Diese Übergangsraten sind dieselben wie im Prozess $(\Pi^{(\mathcal{M})}_t)_{t\geq 0}$, wenn man die jeweiligen Zustandsräume mittels der Abbildung r' identifiziert, da in beiden Prozessen die Übergangsraten nur von der Anzahl der jeweils zu einem neuen Block zu vereinigenden Blöcke der Partition vor einem Übergang abhängen und bei identischer Anzahl von zu vereinigenden Blöcken gleich sind. Dies folgt aus Eigenschaft 3 aus Definition 1.1.7 und der Definition des n-Coalescents. Des Weiteren stimmen nach Identifizierung der Zustandsräume durch r' auch die jeweiligen Startverteilungen überein. Also gilt (10), da die Verteilungen der Prozesse durch Raten und Startverteilung eindeutig bestimmt sind. Die starke Markov-Eigenschaft bewirkt auch, dass gegeben $\Pi^{(n)}_{T_n}$ die Prozesse $(\Pi^{(n)}_{T_n+t})_{t\geq 0}$ und $(\Pi^{(n)}_t)_{0\leq t\leq T_n}$ unabhängig sind. □

1.3 Poisson-Konstruktion eines Ξ-Coalescent

Nach Schweinsberg [79] ist es möglich, für jedes endliche Maß Ξ auf Δ einen Ξ-Coalescent mittels eines geeigneten Poisson-Punktprozesses zu konstruieren. Dies ist eine Verallgemeinerung der Pitman'schen Poisson-Konstruktion für Λ-Coalescents mit $\Lambda(\{0\}) = 0$ [71]. Die Konstruktion ist im Grunde die Kingman-Paintbox-Konstruktion [57] für austauschbare Partitionen. Auf ähnliche Weise lässt sich auch das Bernoulli-Sieb aus [40] konstruieren. Die Poisson-Konstruktion eines Ξ-Coalescents wird in den folgenden Kapiteln häufig benutzt, weswegen sie hier explizit, wenn auch leicht umformuliert, aus [79, Kapitel 3] angegeben ist:
Sei Ξ ein endliches Maß auf Δ. Sei $\mathcal{L}$ das σ-endliche Maß auf dem Messraum

$([0,\infty)\times\mathbb{N}_0^{\mathbb{N}}, \mathcal{B}([0,\infty))\otimes\bigotimes_{n\in\mathbb{N}}\wp(\mathbb{N}_0))$ definiert durch $\mathcal{L} := \lambda_{[0,\infty)}\otimes L$, wobei

$$L(A) := \int_\Delta \frac{P_x(A)}{(x,x)}\Xi_0(dx) + a\sum_{i\in\mathbb{N}}\sum_{j\in\mathbb{N}, j\geq i} 1_{\{z(i,j)\in A\}}$$

mit

- $A \in \otimes_{n\in\mathbb{N}}\wp(\mathbb{N}_0)$,
- für $x\in\Delta$ ist P_x die gemeinsame Verteilung einer Folge von i.i.d. Zuvallsvariablen $(\xi_i^{(x)})_{i\in\mathbb{N}}$ auf $\mathbb{N}_0$ mit $P(\xi_1^{(x)} = j) = x_j$ für $j\in\mathbb{N}$ und $P(\xi_1^{(x)} = 0) = 1-|x|$,
- für $i,j\in\mathbb{N}$ sei $z(i,j) := (z(i,j)_k)_{k\in\mathbb{N}}$ die durch $z(i,j)_i = z(i,j)_j = 1$ und $z(i,j)_k = 0$ sonst definierte Folge und
- $a := \Xi(\{0\})$.

Betrachte nun einen Poisson-Punktprozess Ψ auf einem geeigneten Wahrscheinlichkeitsraum $(\Omega,\mathcal{F},P)$ mit Intensitätsmaß $\mathcal{L}$. Ψ besteht also aus zufälligen Punkten der Form $(T,X) = (T,(X_j)_{j\in\mathbb{N}})\in[0,\infty)\times\mathbb{N}_0^{\mathbb{N}}$. Definiere zunächst die Menge

$$A_n := \left\{z\in\mathbb{N}_0^{\mathbb{N}} | z_i = z_j \neq 0 \text{ für mindestens ein Paar } \{i,j\}\subseteq[n]\right\}.$$

Es gilt $L(A_n) < \infty$ und somit $\mathcal{L}([0,t]\times A_n) < \infty$ für $t\geq 0$. Somit häufen sich die t-Koordinaten der Punkte von $\Psi\cap([0,\infty)\times A_n)$ fast sicher nicht in $[0,\infty)$ und lassen sich total ordnen. Ordne die Punkte von $\Psi\cap([0,\infty)\times A_n)$ in der t-Koordinate als $\Psi_n' := ((T^{(i)},X^{(i)}))_{i\in\mathbb{N}}$, also derart, dass

$$\left\{(T^{(i)},X^{(i)}) | i\in\mathbb{N}\right\} = \Psi\cap([0,\infty)\times A_n) \text{ und } i\leq j \Leftrightarrow T^{(i)}\leq T^{(j)}.$$

Um aus Ψ einen Ξ-Coalescent $(\Pi_t)_{t\geq 0}$ zu konstruieren, konstruiert man für jedes $n\in\mathbb{N}$ zunächst pfadweise den stochastischen Prozess $\Pi^{(n)} = (\Pi_t^{(n)})_{t\geq 0}$ mit Zustandsraum $\mathbb{E}_n$wie folgt:
Setze $\Pi_0^{(n)} = \mathsf{Diag}_n$. Betrachte nun Ψ_n'. Der Prozess $\Pi^{(n)}$ soll nur zu den Zeitpunkten $T^{(1)}, T^{(2)},\dots$ den Zustand wechseln können. Für $i\in\mathbb{N}$ betrachte den Zustand von $\Pi_{T^{(i)}-}^{(n)}$ bestehend aus den Blöcken $(B_1,\dots,B_k)$. Um die Partition $\Pi_{T^{(i)}}^{(n)}$ zu bestimmen, vereinigt man für jedes $k\in\mathbb{N}$ alle Blöcke B_j, $j\in[k]$, für deren Indizes j $X_i^{(j)} = k$ gilt. Alle Blöcke mit Index j und $X_i^{(j)} = 0$

werden nicht vereinigt.
Hat man nun pfadweise $\Pi^{(n)}$ für jedes $n \in \mathbb{N}$ konstruiert, so ist Π eindeutig bestimmt durch $\rho_n(\Pi_t) = \Pi_t^{(n)}$ für $t \geq 0$ und alle $n \in \mathbb{N}$. Ein derart konstruierter Ξ-Coalescent heißt durch den Poisson-Punktprozess Ψ konstruierter (Ξ-)Coalescent, kurz PPP-(Ξ-)Coalescent.
Im Falle eines Simple Coalescent lässt sich Π auch direkt pfadweise aus dem Poisson-Punkt-Prozess Ψ konstruieren. Da in diesem Fall

$$\nu(dx) := (x, x)^{-1}\Xi(dx)$$

ein endliches Maß auf Δ ist, gilt sogar $L(\mathbb{N}_0^{\mathbb{N}}) < \infty$ und somit $\mathcal{L}([0,t] \times \mathbb{N}_0^{\mathbb{N}}) < \infty$ für jedes $t \in [0, \infty)$. Somit häufen sich sogar die t-Koordinaten aller Punkte von Ψ nicht. Betrachte wiederum $\Psi' = ((T^{(i)}, X^{(i)}))_{i \in \mathbb{N}}$, die in der t-Koordinate geordneten Punkte von Ψ. Dann konstruiere $(\Pi_t)_{t \geq 0}$ pfadweise folgendermaßen:

Konstruktion 1.3.1

1) Starte mit $\Pi_0 = \mathsf{Diag}$.

2) Übergänge sind nur möglich zu den Zeitpunkten $T^{(1)}, T^{(2)}, \ldots$.

3) Um $\Pi_{T^{(i)}}$ *zu konstruieren, vereinige alle Blöcke* B_j *von* $\Pi_{T^{(i)}-}$*, deren Indizes* j $X_j^{(i)} = k$ *für dasselbe* $k \in \mathbb{N}$ *erfüllen. Für Indizes* j *mit* $X_j^{(i)} = 0$ *mache nichts.*

Bemerkung 1.3.2 *Man sieht anhand dieser Poisson-Konstruktion, dass sich ein Simple Coalescent auf folgende Weise verhält: Nach jeweils einer fast sicher positiven Wartezeit springt der Simple Coalescent zum nächsten Zustand. Ist dies der absorbierende Zustand* $\mathbb{N}$*, so bleibt der Prozess für immer dort. Man nennt die Klasse der Simple Coalescents deswegen auch Jump-Hold-Coalescents.*

Da bei der Poisson-Konstruktion eines Simple Coalescents L ein endliches Maß ist und ebenso $\mathcal{L}$ auf $[n, n+1) \times \mathbb{N}_0^{\mathbb{N}}$ endlich ist, kann man den Poisson-Prozess Ψ noch genauer analysieren. Betrachtet man die Konstruktion eines Poisson-Punktprozesses in [58, S.23], so kann man Ψ oBdA als eine Vereinigung unabhängiger Poisson-Punktprozesse $(\Psi_n)_{n \in \mathbb{N}}$ mit jeweiligem Intensitätsmaß $\mathcal{L}_n$ auffassen, wobei $\mathcal{L}_n$ für $n \in \mathbb{N}$ die Einschränkung von $\mathcal{L}$ auf $[n, n+1) \times \mathbb{N}_0^{\mathbb{N}}$ ist. Setze $m := \mathcal{L}([n, n+1) \times \mathbb{N}_0^{\mathbb{N}}) = L(\mathbb{N}_0^{\mathbb{N}})$. Für

jedes $n \in \mathbb{N}$ lässt sich Ψ_n nun wie in [58, S.23] als eine zufällige Menge auffassen, die aus den Werten von Zufallsvariablen $Y_{1,n}, \ldots, Y_{Z_n,n}$ gebildet wird, wobei $Z_n \stackrel{d}{=} Poiss(m)$ und $(Y_{i,n})_{i\in\mathbb{N}}$ eine von Z_n unabhängige i.i.d. Folge von Zufallsvariablen ist mit $Y_{1,n} \stackrel{d}{=} \mathcal{L}/m$. Es gilt also $\Psi = \bigcup_{n\in\mathbb{N}} \Psi_n = \{(Y_{i,n}) | n \in \mathbb{N},\ i \in [Z_n]\}$. Betrachtet man nun die Projektion $\rho : [0,\infty) \times \mathbb{N}_0^{\mathbb{N}} \to \mathbb{N}_0^{\mathbb{N}}$ auf die x-Koordinate definiert durch $\rho(t,x) = x$, so erhält man $\rho(\Psi) = \{X | (T,X) \in \Psi\}$ als die Werte einer i.i.d. Folge von Zufallsvariablen mit jeweiliger Verteilung $\mathcal{L}/m$.
Ordnet man nun die Punkte von Ψ in der t-Koordinate als $\Psi' = ((T^{(i)}, X^{(i)}))_{i\in\mathbb{N}}$, so bleibt aufgrund der Produktgestalt von $\mathcal{L}$ (und damit auch von $\mathcal{L}_n$) zumindest die Folge $(X^{(i)})_{i\in\mathbb{N}} = \rho(\Psi')$ i.i.d.. Die Verteilung von $X^{(1)}$ ist wiederum

$$P_{X^{(1)}} = L/m = \int_\Delta P_x \frac{\nu(dx)}{\nu(\Delta)}.$$

Anhand dieser Integraldarstellung der Verteilung von $X^{(1)} = (X^{(1)})_{j\in\mathbb{N}}$ sieht man, dass für festes $i \in \mathbb{N}$ die Zufallsvariablen $(X_j^{(i)})_{j\in\mathbb{N}}$ austauschbar sind. Es existiert also nach dem Satz von de Finetti (etwa [59, Satz 12.26]) ein zufälliges Maß $Q^{(i)}$ auf $\mathbb{N}_0$ mit

$$\begin{aligned} &P_{(X_j^{(i)})_{j\in\mathbb{N}}}(\cdot | Q^{(i)}) = (Q^{(i)})^{\mathbb{N}}(\cdot) \quad \text{mit} \\ &P((Q^{(i)}(\{j\}))_{j\in\mathbb{N}} \in \cdot) = \nu(\cdot)/\nu(\Delta), \end{aligned} \tag{11}$$

wobei die Verteilung von $Q^{(i)}$ nicht von $i \in \mathbb{N}$ abhängt und die Verteilung von $Q^{(i)}(\{0\})$ eindeutig durch (11) bestimmt ist (via Gegenereignis). Schließlich ist $Q^{(i)}$ messbar bezüglich der terminalen σ-Algebra erzeugt von $(X_j^{(i)})_{j\in\mathbb{N}}$, also ist die Folge $(Q^{(i)})_{i\in\mathbb{N}}$ eine i.i.d. Folge zufälliger Maße auf $\mathbb{N}_0$.
In der Poisson-Konstruktion eines Simple Coalescent wird nun der Poisson-Punkt $(T^{(i)}, X^{(i)})$ für $i \in \mathbb{N}$ betrachtet. Betrachte den zugehörigen Schritt 3 der Poissonkonstruktion 1.3.1. Man kann diesen auch folgendermaßen interpretieren: Für jeden Block A_j von $\Pi_{T^{(i)}-} = (A_1, \ldots, A_b)$ werfe einen Ball $j \in [b]$ auf Kästen $0, 1, \ldots$. Die Bälle werden derart geworfen, dass Ball j Kasten $X_j^{(i)}$ trifft. Vereinige nun jeweils alle Blöcke, deren Bälle zusammen in einen Kasten $1, 2, \ldots$ gefallen sind (Blöcke mit Bällen in Kasten 0 werden nicht vereinigt). Bedingt $Q^{(i)}$ sind die Zufallsvariablen $(X_j^{(i)})_{j\in[b]}$ i.i.d. mit Verteilung $Q^{(i)}$, d.h. die Bälle werden unabhängig geworfen und jeder Ball trifft Kasten $k \in \mathbb{N}_0$ mit Wahrscheinlichkeit $Q^{(i)}(\{k\})$. Somit ist Schritt 3

bedingt auf $Q^{(i)}$ nichts anderes als ein klassisches Bälle-auf-Kästen-Werfen, vergleiche hierzu etwa die Arbeit [53] von Karlin oder den Überblicksartikel [38], bei dem man nach den Würfen all diejenigen Blöcke zusammenfasst, deren Bälle in demselben Kasten $k \neq 0$ gelandet sind. Ohne Bedingen kann man das Bälle-auf-Kästen-Werfen in Schritt 3 auch als Werfen auf Kästen mit zufälligen Trefferwahrscheinlichkeiten auffassen.
Nun betrachtet man für $n \in \mathbb{N}$ nur die Poisson-Konstruktion 1.3.1 für den Simple n-Coalescent $\Pi^{(n)}$. Der einzige Unterschied ist, dass nun höchstens n Bälle geworfen werden. Betrachte nun Schritt 3 der Konstruktion von $\Pi^{(n)}$ bedingt auf $Q^{(i)} = (x_0, x_1, \ldots)$ für $x \in \Delta$ und $x_0 = 1 - |x|$ (fasse $Q^{(i)}$ als Wahrscheinlichkeitsvektor auf). Die Anzahl der Kollisionen, die durch Schritt 3 unter der Bedingung $Q^{(i)} = (x_0, x_1, \ldots)$ erzeugt werden, wird durch die Anzahl $O_n^{(x)}$ der belegten Kästen im Bälle-in-Kästen-Modell mit den $n \geq b$ geworfenen Bällen $X_1^{(i)}, \ldots, X_n^{(i)}$ und der Treffer-Wahrscheinlichkeit x_k für Kasten $k \in \mathbb{N}_0$ nach oben abgeschätzt (b ist die Anzahl der Blöcke direkt vor Schritt 3). Dies ist klar, da nur dann eine Kollision entsteht, wenn mindestens 2 Bälle in einen Kasten mit Index $k \neq 0$ geworfen werden und nur so viele Bälle geworfen werden, wie Blöcke unmittelbar vor $T^{(i)}$ vorhanden sind, somit höchstens die genannten n Bälle. Aus [53, Theorem 8, S.395] erhält man für festes $x \in \Delta$

$$E(O_n^{(x)}) = \sum_{i \in \mathbb{N}_0} (1 - (1 - x_i)^n), \quad O_n^{(x)} \sim E(O_n^{(x)}) \tag{12}$$

P-fast sicher für $n \to \infty$. Somit erhält man folgendes Konvergenzergebnis für die Anzahl der Kollisionen, die durch einen Poisson-Punkt $(T^{(i)}, X^{(i)})$ ausgelöst werden.

Satz 1.3.3 *[34] Sei Π ein Simple PPP-Ξ-Coalescent mit erzeugendem Poisson-Prozess Ψ. Sei $n \in \mathbb{N}$ und sei $C_n^{(i)} (= C_n^{(i)}(T^{(i)}, X^{(i)}))$ die Anzahl der Kollisionen im n-Coalescent in Schritt 3 der Poissonkonstruktion 1.3.1 durchgeführt für einen beliebigen Poisson-Punkt $(T^{(i)}, X^{(i)})$ von Ψ. Des Weiteren sei $V_n^{(i)} := \sum_{k \in \mathbb{N}_0} 1_{\left\{\exists\ j \in [n]:\ X_j^{(i)} = k\right\}}$ die Anzahl der verschiedenen in $(X_j^{(i)})_{j \in [n]}$ vorkommenden Werte. Es gilt $C_n^{(i)} \leq V_n^{(i)}$ und*

$$\lim_{n \to \infty} \frac{C_n^{(i)}}{n} = \lim_{n \to \infty} \frac{V_n^{(i)}}{n} = 0 \ \textit{P-fast sicher.}$$

Beweis: Mit den Überlegungen direkt vor diesem Lemma erhält man $C_n^{(i)} \leq V_n^{(i)}$. Zeige nun $\lim_{n\to\infty} \frac{V_n^{(i)}}{n} = 0$ fast sicher.
Sei $x \in \Delta$ und $x_0 = 1 - |x|$. Bedingt auf $Q^{(i)} = (x_0, x_1, \ldots)$ ist $V_n^{(i)} = \sum_{k\in\mathbb{N}_0} 1_{\left\{\exists\ j\in[n]:\ X_j^{(i)}=k\right\}}$ gerade $O_n^{(x)}$ und man erhält

$$
\begin{aligned}
& P(\lim_{n\to\infty} \frac{V_n^{(i)}}{n} = 0) = E(P(\lim_{n\to\infty} \frac{V_n^{(i)}}{n} = 0|Q^{(i)})) \\
= \quad & \int_{\Delta\setminus\{0\}} P(\lim_{n\to\infty} \frac{O_n^{(x)}}{n} = 0)\frac{\nu(dx)}{\nu(\Delta)} \\
\overset{(12)}{=} \quad & \int_{\delta\setminus\{0\}} P(\lim_{n\to\infty} \frac{1}{n} \sum_{i\in\mathbb{N}_0} (1-(1-x_i)^n) = 0)\frac{\nu(dx)}{\nu(\Delta)}.
\end{aligned} \tag{13}
$$

Stelle $\frac{1}{n}\sum_{i\in\mathbb{N}_0}(1-(1-x_i)^n) = \int_{\mathbb{N}_0} \frac{1-(1-x_i)^n}{n}\mu_{\mathbb{N}_0}(di)$ als Integral bezüglich des Zählmaßes $\mu_{\mathbb{N}_0}$ auf $\mathbb{N}_0$ dar. Dann folgt mit dem Satz für majorisierte Konvergenz

$$
\lim_{n\to\infty} \int_{\mathbb{N}_0} \frac{1-(1-x_i)^n}{n}\mu_{\mathbb{N}_0}(di) = 0,
$$

da $0 \leq 1-(1-x_i)^n \leq nx_i$ für $i \in \mathbb{N}_0$ aufgrund der Bernoulli-Ungleichung und die Folge $(x_i)_{i\in\mathbb{N}_0}$ integrierbar bezüglich $\mu_{\mathbb{N}_0}$ ist. Somit gilt

$$
P(\lim_{n\to\infty} \frac{1}{n} \sum_{i\in\mathbb{N}_0} (1-(1-x_i)^n) = 0) = 1,
$$

da dies eine wahre Aussage ist. Das Ergebnis des Gleichungssystems (13) ist also $P(\lim_{n\to\infty} \frac{V_n^{(i)}}{n} = 0) = 1$, was zu zeigen war. □

Ein Simple PPP-Coalescent besitzt eine gut handhabbare Struktur. Für spätere Analysen von Simple Coalescents stellt sich somit die Frage, ob man jeden (Simple) Coalescent Π als einen PPP-Coalescent auffassen kann, d.h. ob man einen passenden Poisson-Punktprozess konstruieren kann, so dass Π der dazugehörige PPP-Coalescent ist. Dies ist zumindest für bestimmte Simple Coalescents möglich. Um dies zu zeigen, zeigt man zunächst eine weitere Besonderheit bei der Poisson-Konstruktion eines Ξ-Coalescents.

Bemerkung 1.3.4 *Sei* Π *ein* Ξ*-*PPP*-Coalescent mit erzeugendem Poisson-Punktprozess* Ψ*, für den* $\Xi(\{0\}) = 0$ *gilt. Betrachte das Intensitätsmaß* $\mathcal{L}$

von Ψ, genauer das Maß L auf $\mathbb{N}_0^{\mathbb{N}}$. Dieses ist in diesem Fall eine Mischung über das (nicht notwendigerweise endliche) Maß $\nu(dx) = \frac{\Xi(dx)}{(x,x)}$ von gemeinsamen Verteilungen P_x, $x \in \Delta$, von i.i.d. Zufallsvariablen $(\xi_i^{(x)})_{i\in\mathbb{N}}$ mit jeweiligem Zustandsraum $\mathbb{N}_0$. Hat nun $P_{\xi_1^{(x)}}$ Masse $x_i > 0$ in $i \in \mathbb{N}_0$ (setze $x_0 = 1 - |x|$), so kommt dieser Wert nach dem Borel-Cantelli-Lemma fast sicher unendlich oft in $(\xi_j^{(x)})_{j\in\mathbb{N}}$ vor und nach dem starken Gesetz der großen Zahlen gilt $\lim_{n\to\infty} \frac{1}{n} \sum_{j\in[n]} 1_{\left\{\xi_j^{(x)}=i\right\}} = x_i$ fast sicher. Gilt dagegen $P_{\xi_1^{(x)}}(\{i\}) = 0$, so kommt dieser Wert fast sicher nicht in der Folge $(\xi_j^{(x)})_{j\in\mathbb{N}}$ vor. Daraus folgt, dass das Maß L nur Masse auf den Mengen $A_{nosingle} := \left\{a \in \mathbb{N}_0^{\mathbb{N}} | a_j \neq 0 \text{ für alle } j \in \mathbb{N}\right\}$ und $A_{single} := \left\{a \in \mathbb{N}_0^{\mathbb{N}} | \lim_{n\to\infty} \frac{1}{n} \sum_{j\in[n]} 1_{\{a_j=0\}} > 0\right\}$ hat. Dieselbe Aussage gilt auch, wenn man in den beiden Mengen $a_j = 0$ ($a_j \neq 0$) mit $a_j = k$ ($a_j \neq k$) für $k \in \mathbb{N}$ austauscht. $A_{nosingle}$ ist eine messbare Menge. Ist Π ein Simple Coalescent, so gilt sogar

$$L(A_{nosingle}) = \int \frac{P_x(A_{nosingle})}{(x,x)} \Xi(dx) = \int 1_{\{x\in\Delta^*\}} \nu(dx) = \nu(\Delta^*) \qquad (14)$$

für das (im Falle eines Simple Coalescents) endliche Maß $\nu(dx) = \frac{\Xi(dx)}{(x,x)}$.

Lemma 1.3.5 *Betrachte einen Simple Ξ-Coalescent Π mit*

$$P(S_t > 0) = 1 \text{ für alle } t \geq 0. \qquad (15)$$

Dann lässt sich ein Poisson-Punkt-Prozess Ψ' aus Π konstruieren, so dass Π ein durch Ψ' konstruierter PPP-Coalescent ist.

Beweis: Die Anzahl $|\Pi_t|$ der Blöcke von Π_t ist monoton fallend in t. Somit ist (15) äquivalent zu $P(S_t > 0 \; \forall \; t \geq 0) = 1$. Sei zunächst Π schon ein PPP-Coalescent mit zugehörigem Poisson-Punktprozess Ψ auf $(\Omega, \mathcal{F}, P)$. Konstruiere nun Ψ', ohne Ψ zu verwenden.
Notiere alle Sprungzeiten $T_1 \leq T_2 \leq \ldots$ von Π. Sprünge können nur durch Punkte von Ψ ausgelöst werden. Da der Coalescent nach (15) zu jedem Zeitpunkt fast sicher unendlich viele Blöcke hat, führt auch jeder Poisson-Punkt von Ψ nach Bemerkung 1.3.4 zu mindestens einer Kollision. Somit gilt $T_i = T^{(i)}$ für die Variablen $T^{(i)}$, $i \in \mathbb{N}$, von Ψ.
Betrachte nun $\Pi_{T_{i-1}}$ und Π_{T_i} für $i \in \mathbb{N}$ auf der Einsmenge $\Omega' :=$

$\{S_t > 0 \; \forall \; t \geq 0\}$. $\Pi_{T_{i-1}}$ bestehe aus Blöcken $(A_j)_{j\in\mathbb{N}}$ und Π_{T_i} aus Blöcken $(B_k)_{k\in\mathbb{N}}$. Beide Partitionen haben durch Bedingung (15) fast sicher unendlich viele Blöcke. In der Poisson-Konstruktion entstehen die Blöcke $(B_k)_{k\in\mathbb{N}}$ aus den Blöcken $(A_j)_{i\in\mathbb{N}}$, indem man alle Blöcke A_j mit $X_j^{(i)} = l$ für ein $l \in \mathbb{N}$ vereinigt und die entstandenen Blöcke nach den kleinsten Elementen sortiert. Betrachte die Partition η von $\mathbb{N}$ mit den Blöcken $C_k := \{j \in \mathbb{N} | A_j \subseteq B_k\}$. Dann gilt fast sicher für jedes $k \in \mathbb{N}$ entweder $C_k := \{j \in \mathbb{N} | A_j \subsetneq B_k\} = \left\{j \in \mathbb{N} | X_j^{(i)} = l\right\}$ für ein $l \in \mathbb{N}$ oder $C_k = \{k\}$, wobei im zweiten Fall $X_j^{(i)} = 0$ für $A_j = B_k$ gilt. Dies folgt aus Bemerkung 1.3.4, da für fast alle $\omega \in \Omega$ eine Zahl in $X^{(i)}(\omega) = (X_j^{(i)}(\omega))_{j\in\mathbb{N}}$ entweder unendlich oft oder gar nicht vorkommt. Die $(C_k)_{k\in\mathbb{N}}$ lassen sich aus Π konstruieren. $\eta = (C_k)_{k\in\mathbb{N}}$ ist die Partition von $\mathbb{N}$, die durch die zufällige Äquivalenzrelation

$$m \sim m \; \forall \; m \in \mathbb{N} \text{ und } m \sim n, m \neq n \;\Leftrightarrow\; X_m^{(i)} = X_n^{(i)} = l \in \mathbb{N}$$

definiert wird. Da $(X_n^{(i)})_{n\in\mathbb{N}}$ austauschbar ist, ist η eine austauschbare Partition. Somit existieren nach der Kingmandarstellung die asymptotischen Frequenzen $A_k := \lim_{n\to\infty} \frac{|\tilde{C}_k \cap [n]|}{n}$, wobei $(\tilde{C}_k)_{k\in\mathbb{N}}$ die Blöcke von η geordnet nach absteigender Mächtigkeit bezeichnet. Es gilt des Weiteren $(A_k)_{k\in\mathbb{N}} = (Q^{(i)}(\{k\}))_{k\in\mathbb{N}}$ fast sicher für das zufällige Maß $Q^{(i)}$ aus (11), da nach dem starken Gesetz für austauschbare Zufallsvariable aus [30]

$$\lim_{n\to\infty} \frac{|\tilde{C}_k \cap [n]|}{n} = \lim_{n\to\infty} \frac{1}{n} \sum_{j\in[n]} 1_{\left\{X_j^{(i)}=k\right\}} = Q^{(i)}(\{k\})$$

für $k \in \mathbb{N}$ gilt. Hier ist zu beachten, dass sowohl $(A_l)_{l\in\mathbb{N}}$ als auch $(Q^{(i)}(\{l\}))_{l\in\mathbb{N}}$ nach Größe absteigend geordnet sind. Aus $(A_k)_{k\in\mathbb{N}} = (Q^{(i)}(l))_{l\in\mathbb{N}}$ folgt auch $1 - \sum_{i\in\mathbb{N}} A_i = Q^{(i)}(\{0\})$ fast sicher. Somit lässt sich $Q^{(i)}$ fast sicher aus Π rekonstruieren. Definiere nun $Y_j^{(i)} := l$ falls $j \in \tilde{C}_l(\omega)$ mit $|\tilde{C}_l(\omega)| > 1$ und $Y_j^{(i)} := 0$ sonst. Dann wird η nach Konstruktion auch P-fast sicher durch die Relation

$$m \sim m \; \forall \; m \in \mathbb{N} \text{ und } \; m \sim n, m \neq n \;\Leftrightarrow\; Y_m^{(i)} = Y_n^{(i)} = l \in \mathbb{N}$$

beschrieben. Des Weiteren ist $(Y_j^{(i)})_{j\in\mathbb{N}}$ austauschbar, dies folgt aus der Austauschbarkeit von η. Es gilt auch

$$\lim_{n\to\infty} \frac{1}{n} \sum_{j\in[n]} 1_{\left\{Y_j^{(i)}=l\right\}} = Q^{(i)}(\{l\})$$

nach Konstruktion. Somit ist nach dem Satz von de Finetti $Y^{(i)} := (Y_j^{(i)})_{j\in\mathbb{N}}$ bedingt i.i.d. bezüglich $Q^{(i)}$ und somit gilt $X^{(i)} \overset{d}{=} Y^{(i)}$. Im Allgemeinen muss aber nicht $X^{(i)} = Y^{(i)}$ gelten, da auf $\left\{Q^{(i)}(\{j\}) = Q^{(i)}(\{j'\})\right\}$ für $j, j' \in \mathbb{N}$ die Werte nicht übereinstimmen müssen. Des Weiteren ist nach Konstruktion $(Y^{(i)})_{i\in\mathbb{N}}$ eine i.i.d. Folge unabhängig von $(T_i)_{i\in\mathbb{N}}$. Also gilt

$$\Psi' := \left\{(T_i, Y^{(i)}) | i \in \mathbb{N}\right\} \overset{d}{=} \Psi$$

und nach Konstruktion ist Π auch ein durch Ψ' erzeugter PPP-Coalescent. Sei nun Π ein beliebiger Simple Coalescent, der (15) erfüllt. Dann lässt sich Ψ' analog aus Π konstruieren und ist nach obiger Konstruktion (es gehen nur Verteilungs-, keine Pfadeigenschaften ein) ein Poisson-Punktprozess mit den gewünschten Eigenschaften. □

Bemerkung 1.3.6

a) *Betrachtet man im Beweis von Lemma 1.3.5 die Konstruktion einer der Variablen $Y^{(i)}$, $i \in \mathbb{N}$, so sieht man, dass diese Konstruktion funktioniert, da der Coalescent Π vor dem i-ten Sprung ∞-viele Blöcke hat. Die Konstruktion funktioniert insbesondere auch, falls nach dem Sprung nur noch endlich viele Blöcke oder keine Singletons mehr vorhanden sind (bei endlich vielen Blöcken sind nur endlich viele P_k grösser als 0). Dies bedeutet, dass man das Verhalten von Π zum Zeitpunkt T_i (etwa Anzahl Blöcke nach dem Sprung, Anzahl der Kollisionen) genauso bestimmen kann wie in einem PPP-Coalescent, falls vor dem Sprung fast sicher unendlich viele Blöcke vorhanden sind. Bedingung (15) muss dafür nicht erfüllt sein.*

b) *Lemma 1.3.5 ist eine Variante von [71, Proposition 24] für die beschriebenen Simple Ξ-Coalescents. Die Aussage lässt sich auch für Coalescents mit Staub und $P(S_t > 0) = 1$ für alle $t \geq 0$ zeigen, nur wird in diesem Fall die Argumentation etwas unhandlicher (läuft aber analog), da man den „großen" PPP-Coalescent nicht direkt konstruieren kann. Im Rahmen dieser Arbeit ist nur die Frage, ob man einen Simple Coalescent immer als PPP-Coalescent auffassen kann, relevant. Deswegen wird hier darauf verzichtet, mehr zu zeigen.*

Es lohnt sich für spätere Anwendungen, die Zufallsvariable

$$\zeta := \inf\{t \in [0, \infty) | S_t = 0\}$$

zu untersuchen (setze $\inf \emptyset := \infty$). Für Coalescents ohne Staub gilt $\zeta \equiv 0$ fast sicher. Für Simple Coalescents besitzt $(S_t)_{t\geq 0}$ càdlàg Pfade, dies folgt aus Bemerkung 1.3.2. Des Weiteren ist $(S_t)_{t\geq 0}$ monoton fallend. Somit gilt für Simple Coalescents $\{\zeta \leq t\} = \{S_t = 0\}$ für alle $t \geq 0$, also ist ζ eine Stoppzeit bezüglich $(\sigma(S_u, u \leq t))_{t\geq 0}$ (und damit auch bezüglich $(\sigma(\Pi_u, u \leq t))_{t\geq 0}$). Mit Hilfe der Poisson-Konstruktion lassen sich weitere Eigenschaften von ζ zeigen.

Lemma 1.3.7 *Sei $\Pi = (\Pi_t)_{t\geq 0}$ ein Simple Ξ-Coalescent und sei $\zeta := \inf\{t \geq 0 | S_t = 0\}$. Dann gilt*

a) $\zeta \overset{d}{=} Exp(\nu(\Delta^*))$, wobei $Exp(0) := \delta_\infty$,

b) auf $[0, \zeta)$ lässt sich Π als ein Ξ'-PPP-Coalescent $(\Pi'_t)_{0\leq t<\zeta}$ mit $\Xi'(dx) = 1_{\{x \notin \Delta^*\}}\Xi(dx)$ auffassen und

c) für die Anzahl $C^{(\zeta)}$ der Kollisionen von $(\rho_n(\Pi_t))_{t\geq 0}$, die zum Zeitpunkt ζ stattfinden, gilt $C_n^{(\zeta)}/n \to 0$ fast sicher für $n \to \infty$.

Beweis: OBdA sei Π ein PPP-Coalescent erzeugt durch den Poisson-Prozess $\Psi = \left\{(T^{(i)}, X^{(i)}) | i \in \mathbb{N}\right\}$ auf $(\Omega, \mathcal{F}, P)$. Da Π ein Simple Coalescent ist, benutze man die direkte Poisson-Konstruktion von Π aus Ψ mit den in diesem Kapitel beschriebenen Eigenschaften. Betrachte die Menge A_{single} aus Bemerkung 1.3.4. Erfüllt der Coalescent vor dem Zeitpunkt $T^{(i)}$ noch $S_t > 0$ für alle $t \leq T^{(i)}$, so wird er nach Poisson-Konstruktion genau dann P-fast sicher $S_t > 0$ für alle $T^{(i)} \leq t < T^{(i+1)}$ erfüllen, falls $X^{(i)} \in A_{single}$ P-fast sicher gilt. Dies sieht man mit folgender Argumentation ein: $(\Pi_t)_{0\leq t<T^{(i)}}$ ist aufgrund der Struktur von Ψ unabhängig von $X^{(i)}$. Somit bleibt von den Singletons von $\Pi_{T^{(i)}-}$ ein Anteil von $\lim_{n\to\infty} \frac{1}{n} \sum_{j\in[n]} 1_{\left\{X_j^{(i)}=0\right\}}$ in $\Pi_{T^{(i)}}$ übrig, also gilt nach Bemerkung 1.3.4 fast sicher genau dann $S_{T^{(i)}} > 0$, falls $X^{(i)} \in A_{single}$ gilt (und somit $S_t > 0$ mindestens bis zum nächsten Sprung des Prozesses).

Aus dieser Argumentation folgt $\zeta = T^{(H)}$ mit $H := \inf\left\{i \in \mathbb{N} | X^{(i)} \in A_{nosingle}\right\}$. Da Ψ ein Poisson-Punktprozess ist, lässt sich die Verteilung von ζ nun leicht bestimmen. Es gilt mit (14)

$$P(\zeta > t) = P(|\Psi \cap ([0,t] \times A_{nosingle})| = 0) = e^{-t\nu(\Delta^*)},$$

also gilt $\zeta \overset{d}{=} Exp(\nu(\Delta^*))$. Dies gilt als reine Verteilungseigenschaft auch in einem beliebigen Ξ-Coalescent, also gilt a).

Vergleiche nun die zwei eingeschränkten Poisson-Punktprozesse $\Psi_1 := \Psi \cap ([0,\infty) \times A_{nosingle})$ und $\Psi_2 := \Psi \cap ([0,\infty) \times (\mathbb{N}_0^{\mathbb{N}} \setminus A_{nosingle}))$. Ψ_1 und Ψ_2 sind unabhängig, insbesondere ist Ψ_2 ein Poisson-Punktprozess auf $\mathbb{R} \times \mathbb{N}_0^{\mathbb{N}}$ mit Intensitätsmaß $\lambda_{[0,\infty)} \otimes L'$ mit $L'(\cdot) := \int \frac{P_x(\cdot)}{(x,x)} \Xi'(dx)$ (vergleiche mit Bemerkung 1.3.4).
Sei nun Π' der aus Ψ_2 konstruierte PPP-Coalescent. ζ hängt nur von Ψ_1 ab und ist somit von Π' unabhängig. Des Weiteren gilt $\Pi'_t = \Pi_t$ für alle $0 \le t < \zeta$ nach Konstruktion. Also gilt b).
Aus Bemerkung 1.3.6 a) folgt auch, dass sich $C_n^{(\zeta)}$ für jeden Ξ-Coalescent pfadweise wie in einem PPP-Coalescent berechnen lässt. In einem Simple PPP-Coalescent gilt nach Satz 1.3.3 $\lim_{n\to\infty} C_n^{(i)}/n = 0$ P-fast sicher für die Anzahl der Kollisionen beim i-ten Sprung $T^{(i)}$. Da $\left\{\lim_{n\in\mathbb{N}} C_n^{(i)}/n = 0\right\}$ eine Einsmenge ist für alle $i \in \mathbb{N}$, folgt mit dem Satz von der totalen Wahrscheinlichkeit (bedinge auf die Mengen $\left\{\zeta = T^{(i)}\right\}$, $i \in \mathbb{N}$) auch $\lim_{n\in\mathbb{N}} C_n^{(\zeta)}/n = 0$ P-fast sicher. □

Bemerkung 1.3.8

- *Das Lemma zeigt insbesondere, dass jeder Simple Coalescent mit* $\Xi(\Delta^*) = 0 (= \nu(\Delta^*))$ *Bedingung (15) erfüllt.*
- *Die Eigenschaften von* ζ *gelten auch für Coalescents mit Staub. Dies wird später in Satz 2.3.8 gezeigt.*

1.4 Der Block-Zählprozess

Für bestimmte Fragestellungen zum n-Coalescent ist es nicht nötig, den gesamten n-Coalescent zu untersuchen. Oft genügt es, nur den Block-Zählprozess $(|\Pi_t^{(n)}|)_{t\ge 0}$ eines n-Coalescents $\Pi^{(n)}$ zu untersuchen. Dieser Prozess hat die folgenden Eigenschaften.

Satz 1.4.1 *[36] Sei* $n \in \{2,3,\ldots\}$, *und* $\Pi^{(n)}$ *ein* Ξ-n-*Coalescent. Setze* $\Xi_0 := 1_{\{\Delta\setminus\{0\}\}}\Xi$ *und* $a := \Xi(\{0\})$, *also* $\Xi = a\delta_0 + \Xi_0$. *Dann ist der Block-Zählprozess* $(|\Pi_t|)_{t\ge 0}$ *von* $\Pi^{(n)}$ *ein Markov-Prozess mit Zustandsraum* $([n], \wp([n]))$ *und Startverteilung* δ_n. *Der Prozess ist ein reiner Todesprozess mit absorbierendem Zustand 1. Die Übergangsraten*

$$g_{nk} := \lim_{t\searrow 0} \frac{P(|\Pi_t^{(n)}| = k)}{t}, \quad n,k \in \mathbb{N}, k < n,$$

und die totale Rate $g_n := \lim_{t\searrow 0}\frac{P(|\Pi_t^{(n)}|<n)}{t} = \sum_{k\in[n-1]} g_{nk}$ *des Block-Zählprozesses haben die Form*

$$g_{nk} = a\binom{n}{2}1_{\{k=n-1\}} + \int_\Delta \sum_{j\in[k]} f_{nkj}(x)\frac{\Xi_0(dx)}{(x,x)}, \quad n,k\in\mathbb{N}, k<n, \tag{16}$$

$$g_n = a\binom{n}{2} + \int_\Delta \left(1-(1-|x|)^n - \sum_{j\in[n]} h_{nj}(x)\right)\frac{\Xi_0(dx)}{(x,x)} \tag{17}$$

mit $f_{nkj}(x) := \sum_{\substack{i_1,\dots,i_j\in\mathbb{N}\\ i_1<\dots<i_j}} \sum_{\substack{n_1,\dots,n_j\in\mathbb{N}\\ n_1+\dots+n_j=n-k+j}} \frac{n!}{(k-j)!n_1!\cdots n_j!}(1-|x|)^{k-j}x_{i_1}^{n_1}\cdots x_{i_j}^{n_j}$ *und* $h_{nj}(x) := \binom{n}{j}(1-|x|)^{n-j}\sum_{\substack{i_1,\dots,i_j\in\mathbb{N}\\ \text{verschieden}}} x_{i_1}\cdots x_{i_j}$. *Ist* $\Pi^{(n)}$ *ein* Λ*-Coalescent, so lassen sich die Raten auch als*

$$g_{nk} = \binom{n}{k-1}\int_{[0,1]} u^{n-k-1}(1-u)^{k-1}\,\Lambda(du), \tag{18}$$

$$g_n = \int_{[0,1]} \frac{1-(1-u)^n - nu(1-u)^{n-1}}{u^2}\,\Lambda(du) \tag{19}$$

für $k\in[n-1]$ *schreiben.*

Beweis: Es genügt, die Rosenblattbedingung für Raten aus [21] nachzuweisen. Die Rosenblattbedingung ist eine hinreichende Bedingung dafür, dass eine Funktion eines Markov-Prozesses wieder ein Markov-Prozess ist und sie gibt auch die Raten des neuen Prozesses an. Wir betrachten $\Pi^{(n)}$ unter der Abbildung $|\cdot|$. Für die Rosenblattbedingung muss man hier nachweisen, dass für $n,l,k\in\mathbb{N}$, $k\le l\le n$, $\eta\in\mathbb{E}_n$ mit $|\eta|=l$ die Summe $\sum_{\substack{\xi\in\mathbb{E}_n\\|\xi|=k}} q_{\eta\xi}$ von Übergangsraten $q_{\eta\xi}$ von $\Pi^{(n)}$ nur von l, nicht aber von der Wahl von η mit l Blöcken abhängt (für $\xi\in\mathbb{E}_n$ gilt $q_{\xi\xi}=-q_\xi$, wobei q_ξ die totale Rate von $\Pi^{(n)}$ an der Stelle ξ bezeichnet). Dies zeigt man so: Man sieht, dass nur diejenigen Raten $q_{\eta\xi}$ ungleich 0 sein können, bei denen ξ durch Vereinigen von einer oder mehrerer Mengen von Blöcken von η entsteht. Betrachte nun auch die i-Coalescents $\Pi^{(i)} := (\rho_{ni}(\Pi_t^{(n)}))_{t\ge 0}$ für $i\in[n]$. Nach Bedingung 3 aus Definition 1.1.7 ist die Rate für beliebige Übergange von einer Partition mit l Blöcken zu einer Partition mit k Blöcken, bei denen jeweils gleich große Mengen von Blöcken zu jeweils einem Block vereinigt werden, identisch in allen Prozessen $\Pi^{(i)}$ für $i\in\{l,\dots,n\}$. Vergleicht man nun $\eta\in\mathbb{E}_n$ mit $|\eta|=l$

mit $\mathsf{Diag}_l \in \mathbb{E}_l$, so folgt

$$\sum_{\xi\in\mathbb{E}_n,|\xi|=k} q_{\eta\xi} = \sum_{\xi\in\mathbb{E}_n,|\xi|=k,\eta\to\xi} q_{\eta\xi} = \sum_{\xi\in\mathbb{E}_l,|\xi|=l} q_{\mathsf{Diag}_l\xi},$$

wobei $\eta \to \xi$ bedeutet, dass ξ aus η durch Vereinigung von Blöcken von η entsteht (hier werden der Einfachheit die Raten von $\Pi^{(i)}$, $i \in [n]$, immer mit $q_{\zeta\xi}$ für $\zeta, \xi \in \mathbb{E}_l$ bezeichnet). Die Summe $\sum_{\substack{\xi\in\mathbb{E}_l \\ |\xi|=k}} q_{\mathsf{Diag}_l\xi}$ hängt offensichtlich nicht von der Wahl von $\eta \in \mathbb{E}_n$ ab. Also ist nach der Rosenblatt-Bedingung $(|\Pi_t^{(n)}|)_{t\geq 0}$ ein Markovprozess mit Raten $g_{lk} = \sum_{\substack{\xi\in\mathbb{E}_n \\ |\xi|=k}} q_{\eta\xi}$ für ein $\eta \in \mathbb{E}_n$ mit $|\eta| = l$. Alle anderen Aussagen bis auf die expliziten Darstellungen der Raten folgen sofort aus der Definition des Prozesses als Anzahl der Blöcke eines n-Coalescents.
Die Darstellung (17) der totalen Rate des Block-Zählprozesses eines Ξ-Coalescents findet sich in [79, Gleichung 70]. Analog findet man die explizite Darstellung von g_{nk}. Definiere dazu $X_i^{(x)} := \sum_{j\in[n]} 1_{\left\{\xi_j^{(x)}=i\right\}}$ für $i \in \mathbb{N}_0$, wobei $(\xi_j^{(x)})_{j\in\mathbb{N}}$ i.i.d. Zufallsvariablen sind mit $P(\xi_1^{(x)} = i) = x_i$ für $i \in \mathbb{N}$ und $P(\xi_1^{(x)} = 0) = 1 - |x|$ für $x \in \Delta$ ($\xi^{(x)}$ taucht auch in Kapitel 1.7 auf). Sei $(A_1, \ldots, A_l)$ eine Partition von $[n]$. Sei $\mathsf{Part}((\xi_i^{(x)})_{i\in[n]})$ die durch die Äquivalenzrelation

$$i \sim i \ \forall\ i \in \mathbb{N} \text{ und } i \sim j, i \neq j \ \Leftrightarrow \xi_i^{(x)} = \xi_j^{(x)} \neq 0$$

für $i, j \in \mathbb{N}$ definierte Partition von $[n]$. Betrachte die Darstellung der Raten von $\Pi^{(n)}$ aus Theorem 1.1.12 (ab jetzt werden die Raten $q_{\eta\xi}$ für $\eta, \xi \in \mathbb{E}_n$ wieder als $\lambda(n; k_1, \ldots, k_m)$ für geeignete Argumente wie in Definition 1.1.7 geschrieben). Durch Vergleich mit der Definition von $\mathsf{Part}((\xi_i^{(x)})_{i\in[n]})$ erhält man, dass für $n \in \mathbb{N}$, $k_1 \geq \ldots \geq k_m \geq 2$, $r \in \mathbb{N}_0$ mit $n = r + \sum_{i\in[m]} k_i$ für die Rate $\lambda(n; k_1, \ldots, k_m)$

$$\begin{aligned}
&\lambda(n; k_1, \ldots, k_m) \\
= \ & a1_{\{m=1,k_1=2\}} + \int_\Delta P\Big(\mathsf{Part}((\xi_i^{(x)})_{i\in[n]}) = (A_1, \ldots, A_l)\Big)\frac{\Xi_0(dx)}{(x,x)} \qquad (20)
\end{aligned}$$

gilt, wobei eine Partition $(A_1, \ldots, A_l)$ mit $l = m + r$ und $\{|A_i||i \in [l]\} = \{k_1, \ldots, k_m, 1, \ldots 1\}$ gewählt wird. Summiert man (20) über alle $\eta = (A_1, \ldots, A_k) \in \mathbb{E}_n$ mit genau k Blöcken, erhält man auf der linken Seite von (20) nach der Rosenblattbedingung g_{nk}. Auf der rechten Seite von (20) zieht man die Summe als Vereinigung in den Integranden und erhält

$$P\Big(\bigcup_{\eta\in\mathbb{E}_n,|\eta|=k}\Big\{\mathsf{Part}((\xi_i^{(x)})_{i\in[n]})=\eta\Big\}\Big)=P\Big(X_0^{(x)}+\sum_{i=1}^{\infty}1_{\{X_i^{(x)}\geq 1\}}=k\Big),$$

da für jede 0 in $(\xi_i^{(x)})_{i\in[n]}$ jeweils ein Block von $\mathsf{Part}((\xi_i^{(x)})_{i\in[n]})$ entsteht, für jede andere Zahl $i\in\mathbb{N}$ aber nur ein Block entsteht, wenn diese in $(\xi_i^{(x)})_{i\in[n]}$ mindestens einmal vorkommt. Die beschriebene Summation von (20) liefert also für $n\in\mathbb{N}$ und $k\in[n-1]$

$$g_{nk} \;=\; a\binom{n}{2}1_{\{k=n-1\}}+\int_\Delta P\Big(X_0^{(x)}+\sum_{i=1}^{\infty}1_{\{X_i^{(x)}\geq 1\}}=k\Big)\,\frac{\Xi_0(dx)}{(x,x)}, \tag{21}$$

da der erste Summand aus (20) bei der Summation für alle Partitionen vorkommt, die genau zwei Blöcke vereinigen ($k=n-1$ bzw. $m=1,k_1=2$); dies sind $\binom{n}{2}$. Man erhält

$$\begin{aligned}
P\Big(X_0^{(x)}+\sum_{i=1}^{\infty}1_{\{X_i^{(x)}\geq 1\}}=k\Big) &= \sum_{j\in[k]}P\Big(X_0^{(x)}=k-j,\sum_{i=1}^{\infty}1_{\{X_i^{(x)}\geq 1\}}=j\Big)\\
&= \sum_{j\in[k]}\sum_{\substack{I\subseteq\mathbb{N}\\|I|=j}}P(X_0^{(x)}=k-j,X_l^{(x)}\geq 1\;\forall\,l\in I,X_i^{(x)}=0\;\forall\,i\in\mathbb{N}\setminus I)\\
&= \sum_{j\in[k]}\sum_{\substack{i_1,\ldots,i_j\in\mathbb{N}\\i_1<\cdots<i_j}}\sum_{\substack{n_1,\ldots,n_j\in\mathbb{N}\\n_1+\cdots+n_j=n-(k-j)}}\frac{n!}{(k-j)!n_1!\cdots n_j!}(1-|x|)^{k-j}x_{i_1}^{n_1}\cdots x_{i_j}^{n_j},
\end{aligned}$$

wobei $\frac{n!}{(k-j)!n_1!\cdots n_j!}$ die Anzahl der Möglichkeiten ist, dass die Zufallsvariablen $(\xi_1^{(x)},\ldots,\xi_n^{(x)})$ jeweils n_l mal den Wert i_l und $k-j$ mal den Wert 0 annehmen. Dies zeigt Darstellung (16).

Im Fall eines Λ-n-Coalescents finden sich die Formeln für die Raten g_n und g_{nk} in [71, S. 1886]. Sie lassen sich aber auch direkt aus den Raten im Ξ-Coalescent-Fall vereinfachen, da in diesem Fall Ξ auf $([0,1],0,0,\ldots)$ konzentriert ist. Somit trägt für die Raten g_{nk} nur der erste Summand f_{nk1} in (16) zum Integral bei und es gilt $f_{nk1}=\binom{n}{k-1}(1-u)^{k-1}u^{n-k+1}$. Dies zeigt (18), da $(x,x)=x_1^2$ gilt für $x\in([0,1],0,0,\ldots)$. (19) folgt analog aus (17). $\square$

Bemerkung 1.4.2 *Die totalen Raten des Block-Zählprozesses* $(|\Pi_t^{(n)}|)_{t\geq 0}$ *eines* n*-Coalescents* $\Pi^{(n)}$ *sind auch die totalen Raten von* $\Pi^{(n)}$*, da in beiden Fällen einfach alle Übergangsraten* g_{nk}*,* $k \in [n-1]$*, bzw.* $q_{\eta,\xi}$*,* $\eta, \xi \in \mathbb{E}_n$*, addiert werden, um die totale Rate zu erhalten. Da aber die Raten* g_{nk}*,* $k \in [n-1]$*, gerade derart aus Summation der Raten von* $\Pi^{(n)}$ *entstehen, dass jede Rate von* $\Pi^{(n)}$ *in genau einer Berechnung einer Rate* g_{nk} *auftaucht, sind die totalen Raten beider Prozesse gleich. Anschaulich ist das auch klar, da die totale Rate der Parameter der Exponentialverteilung ist, die die Verteilung der Wartezeit auf das erste Ereignis im* n*-Coalescent bzw. im Block-Zählprozess ist. Dies zeigt insbesondere, dass die Wartezeit* T_n *auf den ersten Sprung in* $\Pi^{(n)}$ *auch die Wartezeit auf den ersten Sprung im Block-Zählprozess* $(|\Pi_t^{(n)}|)_{t\geq 0}$ *ist. Um* T_n *zu untersuchen, genügt es also, nur den Block-Zählprozess zu untersuchen.*

Bemerkung 1.4.3 *Mit den Raten* g_{nk}*,* $k \in [n-1]$*, und* g_n *aus Satz 1.4.1 lassen sich die Übergangswahrscheinlichkeiten der Sprungkette* $(\chi_k^{(n)})_{k\in\mathbb{N}_0}$ *des Block-Zählprozesses* $|\Pi^{(n)}| := (|\Pi_t^{(n)}|)_{t\geq 0}$ *angeben. Es gilt*

$$P(\chi_i^{(n)} = l | \chi_{i-1}^{(n)} = k) = \frac{g_{kl}}{g_k}$$

für $k \in [n]$*,* $l \in [k-1]$ *und* $i \in \mathbb{N}$*.*

1.5 Der n-Coalescent als zufälliger Baum

Man kann den n-Coalescent in folgender Weise als einen zufälligen Graphen mit zufälligen Kantenlängen auffassen. Sei $\Pi^{(n)}$ ein n-Coalescent. Betrachte die verschiedenen nacheinander angenommenen Zustände $\pi_1, \ldots, \pi_k \in \mathbb{E}_n$ der Sprungkette von $\Pi^{(n)}$. Dann ist die Menge der verschiedenen Blöcke von $\pi_1, \ldots, \pi_k$ die Menge der Knoten des Graphen. Zwischen zwei dieser Blöcke A und B gibt es genau dann eine Kante, falls $A \subseteq B$ und B als Block des n-Coalescents durch Kollision von mehreren Blöcken entstanden ist, zu denen A gehört, oder wenn dieselbe Bedingung für getauschte Rollen von A und B gilt. Die Länge dieser Kante ist die Aufenthaltszeit, während der $\Pi^{(n)}$ den Block A besitzt, falls $A \subseteq B$ ist; andernfalls ersetze A durch B. Fasse den so konstruierten Graphen als einen Baum mit Wurzel Block $[n]$ auf. Die Blätter des Baumes sind alle Knoten mit Grad 1, dies sind gerade die externen Knoten, also die Blöcke $\{i\}$, $i \in [n]$. Alle Kanten des Baumes werden mit der Richtung ausgestattet, die zur Wurzel zeigt, d.h. dem Zeitverlauf in $(\Pi_t^{(n)})_{t\geq 0}$

folgend. Man kann aus den Pfaden des zufälligen Baumes die Pfade des n-Coalescents rekonstruieren, somit ist die Darstellung als Baum äquivalent zur Darstellung als partitionswertiger Prozess. Im Folgenden werden die Kanten dieses Baumes als Zweige und die externen Knoten als Blätter bezeichnet. Die Kanten, die an einen externen Knoten, also an ein Blatt anschließen, heißen externe Zweige. Der i-te externe Zweig ist derjenige, der an Blatt $i \in \mathbb{N}$ anschließt. Andere Zweige heißen interne Zweige.
Betrachtet man $\Pi^{(n)}$ als zufälligen Graphen, so lässt sich eine anschauliche Reformulierung von Natural Coupling (siehe Satz 1.2.1) und Temporal Coupling (siehe Satz 1.2.3) finden. Sei $M \subseteq [n]$. Natural Coupling lässt sich jetzt folgendermaßen umformulieren: Bildet man den Unterbaum des durch $\Pi^{(n)}$ gegebenen Baumes, der von den Blättern $\{i\}$ für $i \in M$ erzeugt wird (d.h. man führt die oben beschriebene Konstruktion nur für die Zustände $\pi_1 \cap M, \ldots, \pi_k \cap M$ in der Restriktion $\rho_{n,M}(\Pi^{(n)})$ durch), so ist dieser verteilt wie der Baum eines $|M|$-Coalescents mit denselben Raten (nach Umnummerierung). Temporal Coupling lässt sich im durch $\Pi^{(n)}$ gegebenen Baum so formulieren: Schneidet man den Teil des Baumes ab, der der Entwicklung von $(\Pi_t^{(n)})_{t\geq 0}$ in der Zeit von 0 bis einschliesslich des ersten Sprunges entspricht, so ist der verbliebene Teilbaum genauso verteilt wie der Baum eines l-Coalescents, wobei l die Anzahl der Blöcke von $\Pi^{(n)}$ nach dem ersten Sprung ist.

1.6 Der n-Coalescent als stochastisches Modell

Sei $n \in \mathbb{N}$. Im letzten Abschnitt 1.5 wurde gezeigt, dass man den n-Coalescent als einen zufälligen Baum auffassen kann. Dies führt anschaulich zu der für die Anwendung wichtigsten Interpretation der n-Coalescent-Prozesse. Ein n-Coalescent lässt sich als ein (approximatives) Modell für den Stammbaum einer Stichprobe von n Individuen aus einer sehr großen haploiden (= eingeschlechtlichen) Population mit nicht überlappenden Generationen auffassen, deren Größe über die Zeit konstant bleibt (unter bestimmten Bedingungen an die Population und ihrer zu Grunde liegenden Reproduktionsstruktur).
Kingman führte 1982 den Kingman-n-Coalescent in [57] als ein approximatives Modell für den Stammbaum einer n-elementigen Stichprobe einer sehr großen Population (Populationsgröße $N \gg n$) ein, deren Reproduktionsmechanismus durch ein Wright-Fisher-Modell mit fester Populationsgröße N gegeben ist. Präziser formuliert betrachtet man für jedes feste $N \in \mathbb{N}$ mit $N \geq n$ eine Population mit der festen Größe von N Individuen zu den Zeit-

punkten/Generationen $z \in \mathbb{Z}$. Generation $z+1$ besteht aus den Kindern der Individuen aus Generation z, die Individuen aus Generation z werden außerhalb der eigenen Generation nicht betrachtet. Betrachte nun die N Individuen aus Generation $z \in \mathbb{Z}$. Diese reproduzieren sich nach den Regeln eines Wright-Fisher-Modells, d.h. jedes Individuum aus Generation $z+1$ wählt sich zufällig seinen Elter/Vorfahr aus den Individuen der Generation z aus. Äquivalent bedeutet dies, dass die N (nummerierten) Individuen $\{1, \dots, N\}$ der Generation z jeweils eine Anzahl Kinder $\nu^{(z)} = (\nu_1^{(z)}, \dots, \nu_N^{(z)})$ bekommen, wobei die Verteilung von $\nu_N^{(z)}$ eine symmetrische Multinomialverteilung mit Parametern N und N^{-1} ist. Die Art der Nummerierung der Kinder ist im Prinzip unwichtig; ziehe die Nummern etwa einfach aus $[N]$ ohne Zurücklegen. Die Reproduktion jeder Generation z soll unabhängig von den Reproduktionen aller anderen Generationen vonstatten gehen, also ist $(\nu^{(z)})_{z\in\mathbb{Z}}$ i.i.d.. Damit sind die Vorfahren jedes Individuums in jeder vorhergehenden Generation definiert. Nun betrachtet man für eine feste Generation $z \in \mathbb{Z}$, etwa $z = 0$, n Individuen, etwa die Individuen $[n]$. Bilde nun den $\mathbb{E}_n$-wertigen Prozess $(R_n^{(N)})_{n\in\mathbb{N}}$, dessen Zustand $R_n^{(N)}$ durch die Äquivalenzrelation

$$i \sim j \;\Leftrightarrow\; i, j \text{ haben gemeinsamen Vorfahr in Generation } z - n$$

gegeben ist. Einen solchen Prozess nennt man auch diskreten n-Coalescent. Dann gilt (siehe [57, S.31])

$$(R_{\lfloor Nt \rfloor}^{(N)})_{t\geq 0} \xrightarrow{d} \Pi^{(n)}$$

für $N \to \infty$, wobei $\Pi^{(n)}$ einen Kingman-n-Coalescent bezeichnet. Anschaulich fasst man also bei der Approximation die Verwandtschaftsbeziehungen von N Generationen in eine Zeiteinheit des n-Coalescents zusammen.
Das Wright-Fisher-Modell ist ein wichtiges Modell in der Populationsgenetik, auch da es mathematisch sehr gut analysierbar ist (und auch ausführlich analysiert wurde). Es existiert eine Vielzahl von alternativen Populationsmodellen einer haploiden Population mit über die Generationen konstanter Populationsgröße. Betrachte etwa die Klasse der Cannings-Modelle, dies sind Populationsmodelle mit folgenden Eigenschaften:

a) feste Populationsgrösse $N \in \mathbb{N}$,

b) nicht überlappende Generationen,

c) austauschbare Kinderzahlen $\nu^{(z)} := (\nu_1^{(z)}, \ldots, \nu_N^{(z)})$ für jede Generation $z \in \mathbb{Z}$, wobei $(\nu^{(z)})_{z\in\mathbb{Z}}$ eine i.i.d. Folge ist. $\nu_l^{(z)}$ nimmt Werte in $\{0, \ldots, N\}$ an für $l \in [N]$ und $z \in \mathbb{Z}$.

a) bedeutet $\sum_{i\in[N]} \nu_i^{(z)} = N$ für alle $z \in \mathbb{Z}$. Das Wright-Fisher-Modell ist ein spezielles Cannings-Modell.
Wähle $n \in \mathbb{N}$ und betrachte für jedes $N \geq n$ ein Cannings-Modell mit Reproduktionsstruktur $(\nu^{(z,N)})_{z\in\mathbb{Z}}$. Bilde nun zur Stichprobe $[n]$ wie im Falle des Wright-Fisher-Modells für jedes $N \geq n$ den diskreten n-Coalescent $(R_t^{(N)})_{t\geq 0}$. Dann gilt nach [69, Theorem 2.1, S. 1551], dass unter den Voraussetzungen

- $\lim_{N\to\infty} c_N = 0$ mit $c_N := \frac{Var(\nu_1^{(0,N)})}{N-1}$,
- $\lim_{N\to\infty} \frac{E((\nu_1^{(0,N)})_{k_1} \cdots (\nu_j^{(0,N)})_{k_j})}{N^{k_1+\cdots+k_j-j} c_N}$ existiert für alle $j \in \mathbb{N}$, $k_1, \ldots, k_j \in \mathbb{N}$ mit $k_1 \geq \cdots \geq k_j \geq 2$,

wobei $(x)_k := x(x-1)\cdots(x-k+1)$ für $x \in \mathbb{R}$, $k \in \mathbb{N}$ das absteigende faktorielle Produkt ist, für $n \to \infty$

$$(R^{(N)}_{\lfloor t/c_N \rfloor})_{t\geq 0} \xrightarrow{d} \Pi^{(n)} \tag{22}$$

gilt, wobei $\Pi^{(n)}$ ein eindeutig bestimmter Ξ-n-Coalescent ist. Anschaulich fasst man beim Grenzübergang die Verwandtschaftsbeziehungen in $[n]$ von $\approx \lfloor c_N^{-1} \rfloor$ Generationen im Cannings-Modell in einer Zeiteinheit von $\Pi^{(n)}$ zusammmen. Durch spezielle Wahlen der Cannings-Modelle für $N \in \mathbb{N}$ kann jeder Ξ-n-Coalescent $\Pi^{(n)}$ mit $\Xi(\Delta) = 1$ als Grenzprozess in (22) vorkommen (Ändert man in (22) zusätzlich die Zeitskalierung geeignet, kann sogar jeder Ξ-n-Coalescent als Grenzprozess vorkommen).
In vielen Fällen wird der Stammbaum der Stichprobe $[n]$, also der diskrete n-Coalescent, für große N gut durch den Kingman-n-Coalescent approximiert. Kingman zeigt in [57, S. 35], dass falls die Cannings-Modelle $\lim_{N\to\infty} Var(\nu_1^{(0,N)}) = \sigma^2 < \infty$ erfüllen und $\nu_1^{(0,N)}$ nur endliche Momente besitzt, der Stammbaum der Stichprobe immer noch durch den Kingman-Coalescent approximiert wird (anstatt N Generationen muss man allerdings $\sigma^{-2}N$ Generationen zu einem Zeitschritt zusammenfassen). In [63] findet sich folgendes Kriterium (Möhle's lemma) für die Konvergenz von Cannings-Modellen im Sinne von Gleichung (22) gegen den Kingman-n-Coalescent. Falls $\lim_{N\to\infty} E((\nu_1^{(0,N)})_2)/(N^2 c_N) = 0$ gilt, so gilt auch (22) mit

einem Kingman-n-Coalescent $\Pi^{(n)}$ (es gilt für Stichprobengröße $n \geq 3$ sogar Äquivalenz). Dieses Kriterium lässt sich veranschaulichen: c_N ist die Wahrscheinlichkeit, dass im Cannings-Modell für Populationsgröße N zwei zufällig ausgewählte Individuen in einer Generation z einen gemeinsamen Vorfahr in der vorhergehenden Generation $z-1$ besitzen und $E((\nu_1^{(0,N)})_2)/N^2$ ist die Wahrscheinlichkeit, dass drei zufällig ausgewählte Individuen in Generation z einen gemeinsamen Vorfahr in Generation $z-1$ besitzen. Das Kriterium vergleicht nun einfach diese beiden Wahrscheinlichkeiten miteinander.
Folglich wird gerade in biologischen Anwendungen der Kingman-n-Coalescent als Standard-Stammbaum verwendet, da in vielen Reproduktionsmechanismen ein einzelnes Individuum nur mit geringer Wahrscheinlichkeit extrem viele Kinder bekommen kann, wobei extrem viele Kinder etwa $\nu_1^{(z,N)} \sim pN$ für $N \to \infty$ und $p \in (0,1]$ bedeuten kann. Betrachtet man aber Modelle mit solchen ,,extremen" Reproduktionsmechanismen, diese können etwa bei maritimen Lebewesen vorkommen, so ist der Kingman-n-Coalescent keine gute Approximation des diskreten n-Coalescents (vergleiche hierzu [3], [19] oder [45]). In solchen Fällen ist es sinnvoller, den Stammbaum der Stichprobe durch einen geeigneten Ξ-Coalescent zu approximieren, siehe etwa [29] oder für statistische Fragenstellungen [15].
Im Rahmen dieser Arbeit werden meist der Bolthausen-Sznitman-Coalescent oder Coalescents mit Staub betrachtet. Der Bolthausen-Sznitman-n-Coalescent taucht in einigen Anwendungen und mathematischen Modellen auf. Bolthausen und Sznitman führen ihn in [18] im Zusammenhang mit der Spin-Glass-Theorie ein. Der Bolthausen-Sznitman-Coalescent lässt sich auch aus der Genealogie des Continuous-State-Verzweigungsprozesses von Neveu konstruieren (siehe [13]), der ebenfalls im Bereich der Spin-Glass-Theorie eine Rolle spielt. Anschaulich lässt er sich (nach Transformation der Zeitskalen) als ein approximativer Verwandtschaftsbaum einer Stichprobe von Spin-Konfigurationen im CREM (continuous random energy model) von Bovier und Kurkova [20] auffassen. Ein Überblick zu diesen und verwandten (physikalischen) Modellen, in denen der Bolthausen-Sznitman-n-Coalescent in verschiedener Form auftaucht, findet sich in [6, Kapitel 6.2 und 6.3]. Ein weiteres mathematisches Modell, dass ebenso in Verbindung mit Spin-Glass-Modellen steht, ist das System von verzweigenden Brown'schen Bewegungen mit negativem Drift aus [10]. Hier taucht (im fast kritischen Fall) der Bolthausen-Sznitman-n-Coalescent als Grenzwertprozess von Stammbäumen einer Stichprobe von n Brown'schen Bewegungen auf. Der Bolthausen-Sznitman-n-Coalescent taucht in [81] als Grenzprozess $\Pi^{(n)}$ von diskreten n-Coalescents

wie in (22) auf, wobei für $N \geq n$ das Canningsmodell durch N-faches Ziehen ohne Zurücklegen aus den Nachkommen eines geeigneten Galton-Watson-Bienaymé-Verzweigungsprozesses entsteht (hier wird das Canningsmodell nur für die Generationen $0, -1, \ldots$ definiert).
n-Coalescents mit Staub tauchen etwa als Grenzprozesse für diskrete n-Coalescents in Populationen auf, in denen einzelne Individuen mit moderater Wahrscheinlichkeit sehr viele Kinder bekommen können. In [29] werden bestimmte Cannings-Modelle mit Populationsgröße $N \geq n$ vorgestellt, die mit gewisser Wahrscheinlichkeit einem Individuum erlauben, pN Kinder zu bekommen ($p \in (0,1) \cap \mathbb{Q}$, pN muss im Modell ganzzahlig sein, also geht man am besten auf eine geeignete Teilfolge über). Die diskreten n-Coalescents mancher dieser Modelle erfüllen (22) derart, dass $\Pi^{(n)}$ ein Λ-n-Coalescent ist mit $\Lambda = c\delta_d$ mit $c > 0$ und $d \in (0,1]$[1]. Ein Dirac-Ξ-n-Coalescent lässt sich auch als Grenzprozess von verzerrten Canningsmodellen wie in [47, Beispiel 5.2] gewinnen, wenn man diese geeignet mit dem trivialen Cannings-Modell mischt (im trivialen Canningsmodell bekommt jedes Individuum genau ein Kind). Analog zu Beispiel 1.1.19 ist $\Pi^{(n)}$ in diesen Fällen ($\Lambda = c\delta_d$ für $c > 0$, $d \in \Delta$ geeignet) die Restriktion eines Coalescents mit Staub auf $[n]$ (sogar eines Simple Coalescents, vergleiche Beispiel 1.1.21). In [47, S.550] taucht ein weiterer Simple Coalescent auf, der Kingman-Dirichlet-Coalescent, dessen charakterisierendes Maß Ξ die Form $\Xi(dx) = (x,x)\nu(dx)$ für ein Wahrscheinlichkeitsmaß ν auf Δ besitzt (ν kann über die Kingman-Dirichlet-Verteilung definiert werden). Auch in anderen Modellen, etwa dem Modell für Populationen in einem räumlichen Kontinuum von Barton, Etheridge und Véber aus [4] oder den verallgemeinerten Insel-Cannings-Modellen aus [85], tauchen Simple Coalescents als Grenzprozesse von Stammbaumprozessen auf (siehe [4, Theorem 3.7, S.176] und [4, S. 204] sowie [85, Proposition 4.1, S.280] und [85, Beispiel 5, S.281-286]). Die Klasse der $\beta(2-a,a)$-Coalescents für $a \in (0,1)$, dies sind nach Beispiel 1.1.19 Coalescents mit Staub, lassen sich wiederum aus der Genealogie von a-stabilen Continuous-State-Verzweigungsprozessen konstruieren, siehe [16].

Bemerkung 1.6.1

- *Es gibt noch weitere Modelle, in denen der Bolthausen-Sznitman-n-Coalescent oder ein n-Coalescent mit Staub vorkommt. Hier wurden*

[1] Eine Verallgemeinerung der Modelle aus [29], so dass auch Ξ-n-Coalescents mit simultanen multiplen Kollisionen als Grenzwertprozess vorkommen können, findet sich in [77].

nur solche Modelle vorgestellt, für die der betreffende n-Coalescent als ein Stammbaum aufgefasst werden kann, d.h. die Genealogie von Individuen/Elementen aus irgend einer Population beschreibt.

- *Eine Einführung über die biologische Anwendung der n-Coalescents findet sich in [46], [84] oder [87]. Einen Überblick über die mathematische Analyse der Λ-Coalescents liefert [6].*

Da der n-Coalescent als Approximation des Stammbaums der Individuen aus einer n-elementigen Stichprobe in einem Populationsmodell einer sehr großen Population verwendet werden kann, benutzt man auch in rein mathematischen Anwendungen die biologische Terminologie, also bezeichnet man $\mathbb{N}$ als die Population, $[n]$ als Stichprobe und alle $i \in [n]$ respektive $i \in \mathbb{N}$ als Individuen. Die Wurzel des Baumes wird auch als der jüngste gemeinsame Urahn (most recent common ancestor, kurz MRCA) bezeichnet.

1.7 Coalescent mit Mutation

Sei $\Pi^{(n)}$ ein Ξ-n-Coalescent. $\Pi^{(n)}$ wird als Modell für den Stammbaum der Stichprobe $[n]$ einer unendlich großen Population $\mathbb{N}$ angesehen (vgl. Kapitel 1.6). Ein realistisches Modell einer Population, besonders im Hinblick auf genetische Fragestellungen, muss auch einen Mutationsmechanismus beinhalten. Sei dazu $n \in \mathbb{N}$ und $\Pi^{(n)}$ gegeben. Dann setze bedingt $\Pi^{(n)}$ auf den Kanten des durch $\Pi^{(n)}$ gegebenen Baumes Punkte eines auf diesen Kanten definierten Poisson-Prozesses $\mathsf{Mu}^{(n)}$ mit Intensitätsrate $r > 0$, der unabhängig von $\Pi^{(n)}$ ist. Für diese Zwecke betrachten wir jede (gerichtete) Kante des Graphen als halboffenes reelles Intervall der Form $[0, l)$, wobei l die Länge der Kante ist. Jeder Punkt wird als Mutation bezeichnet. Man nennt r auch die Mutationsrate. Im Folgenden wird ein n-Coalescent mit Mutation als $(\Pi^{(n)}, \mathsf{Mu}^{(n)})$ geschrieben.
Sei nun Π ein Ξ-Coalescent und $\Pi^{(n)} := (\rho_n(\Pi_t)_{t\geq 0})$ für $n \in \mathbb{N}$. Betrachtet man die Folge der durch $(\Pi^{(n)})_{n\in\mathbb{N}}$ gegebenen Bäume und auf diesen jeweils durch Poisson-Punktprozesse $\mathsf{Mu}^{(n)}$ gegebene Mutationen, dann sollte auch die Mutationsstruktur (pfadweise) konsistent sein, d.h. dass die durch $\mathsf{Mu}^{(m)}$ gesetzten Mutationen auf dem von $\rho_m(\Pi_t)_{t\geq}$, $m \in [n]$ gegebenen Baum pfadweise identisch sind mit den durch $\mathsf{Mu}^{(n)}$ gesetzten Mutationen auf dem von $\Pi^{(n)}$ gegebenen Unterbaum, der nur von den Blättern $\{i\}$, $i \in [m]$ erzeugt wird (in der auf Seite 35 beschriebenen Weise). Eine solche Mutationsstruktur wird hier pfadweise konsistent genannt. Eine Möglichkeit, eine pfadweise

konsistente Mutationsstruktur auf $(\Pi^{(n)})_{n\in\mathbb{N}}$ zu konstruieren wird im Folgenden beschrieben.
Betrachte einen Ξ-Coalescent Π und eine Folge $(P_i)_{i\in\mathbb{N}}$ von homogenen i.i.d. Poisson-Punktprozessen auf $[0,\infty)$ mit Intensitätsrate $r>0$, die unabhängig von Π sind. Für jede endliche Teilmenge $M\subseteq\mathbb{N}$ mit $m:=\max M$ setzt man auf folgende Weise die (Mutations-)Punkte auf den durch $(\rho_M(\Pi_t))_{t\geq 0}$ gegebenen Baum. Dieser Baum ist pfadweise der durch M erzeugte Unterbaum des von $\Pi^{(m)}$ erzeugten Baums (Konstruktion siehe wieder Seite 35). Auf diesen von $\Pi^{(m)}$ erzeugten Baum setzen wir pfadweise Mutationen wie folgt. Setze zunächst für jedes $i\in[m]$ auf den i-ten externen Zweig mit Länge $E_m^{(i)}$ die Punkte von P_i aus $[0,E_m^{(i)})$ in Richtung des jeweiligen externen Zweiges. Sei nun I eine innere Kante des Baumes. Sei A der Anfangsknoten von I. Die Länge von I beträgt $\beta-\alpha$, wobei $\alpha:=\inf\left\{t\in[0,\infty)|A \text{ ist Block von } \Pi_t^{(m)}\right\}$ und $\beta:=\sup\left\{t\in[0,\infty)|A \text{ ist Block von } \Pi_t^{(m)}\right\}$. Dann setze als Mutationen längs I die Punkte von P_a auf $[\alpha,\beta)$, wobei $a:=\min\{A\}$. a ist also das Individuum mit der kleinsten Nummer aus Block A. Auf diese Weise liefert $(P_i)_{i\in\mathbb{N}}$ eine pfadweise konsistente Mutationsstruktur auf $(\Pi^{(n)})_{n\in\mathbb{N}}$.
Im Rahmen dieser Arbeit werden nur Ξ-Coalescents mit pfadweise konsistenter Mutationsstruktur betrachtet, diese werden kurz (Ξ-)Coalescent mit Mutation genannt.

Bemerkung 1.7.1 *[34] Ist Π ein Coalescent mit Mutation, besitzt Π also eine pfadweise konsistente Mutationsstruktur mit Mutationsrate $r>0$, so lässt sich daraus ein Familie von homogenen i.i.d. Poisson-Punktprozessen $(P_i)_{i\in\mathbb{N}}$ auf $[0,\infty)$ mit Intensitätsrate $r>0$ konstruieren. Dies erreicht man, indem man für alle $n\in\mathbb{N}$ die Mutationen auf den Zweigen der durch $((\rho_n(\Pi_t))_{t\geq 0})_{n\in\mathbb{N}}$ gegebenen Bäume in derselben Weise rückwärts auf $[0,\infty)^{\mathbb{N}}$ anordnet wie in der eben beschriebenen Konstruktion. Dadurch erhält man auf der i-ten Kopie von $[0,\infty)$ aus $[0,\infty)^{\mathbb{N}}$ nur Punkte/Mutationen eines homogenen Poisson-Punktprozesses mit Intensitätsrate r bis zu einem zufälligen Zeitpunkt T_i, nämlich dem Zeitpunkt, an dem das Individuum i zum ersten Mal mit einem Individuum j mit $j<i$ kollidiert oder Π den absorbierenden Zustand $([\infty])$ erreicht. T_i kann auch endliche Werte annehmen. In diesem Fall setzt man auf $[T_i,\infty)$ Punkte nach einem unabhängigen homogenen Poisson-Punktprozess auf $[0,\infty)$ mit der gleichen Intensitätsrate. Aufgrund der Eigenschaft von Poisson-Punktprozessen, auf disjunkten Mengen unabhängige Punkte in diesen Mengen zu generieren, sind die so zusammen-*

geklebten Punktprozesse wiederum unabhängige Poissonprozesse auf $[0,\infty)$. *Man kann also oBdA annehmen, dass ein Coalescent mit Mutation bzw. mit pfadweise konsistenter Mutationsstruktur auf die vor dieser Bemerkung beschriebene Art und Weise konstruiert worden ist.*

Fasse nun den von $\Pi^{(n)}$ gegebenen Baum als Stammbaum der Stichprobe $[n]$ auf. Die Mutationsstruktur kann auf verschiedene Weisen interpretiert werden. Im Rahmen dieser Arbeit wird die Mutation immer als neutral angesehen, d.h. die Mutationen haben keinen Einfluss auf die Reproduktionsstruktur. Mathematisch entspricht dies der Unabhängigkeit von $\Pi^{(n)}$ und $\mathsf{Mu}^{(n)}$. Welchen Einfluss haben nun die Mutationen auf die genetische Struktur der Stichprobe bzw. wie bestimmt man die genetische Struktur der Stichprobe? Grundsätzlich geht man davon aus, dass jedes Individuum einen Typ besitzt. In genetischen Fragestellungen ist dies meist die Struktur des gesamten Genoms oder eine bestimmte Kombinationen von Ausprägungen (Allele) mehrerer Gene/Loci im Genom. Diese Typen werden durch Mutationen verändert. Im Rahmen dieser Arbeit werden zwei Mutationsmodelle betrachtet. In jedem dieser Modelle bestimmt man den Typ eines Individuums aus der Stichprobe $[n]$ zum Beobachtungszeitpunkt $t = 0$, indem man den Weg vom Blatt i zur Wurzel (jüngster Urahn bzw. MRCA) verfolgt und bei jeder Mutation auf diesem Weg den Typ auf die im jeweiligen Modell beschriebene Weise verändert. Liegt eine Mutation auf mehreren Wegen, so verändert sie den Typ jedes Individuums, auf dessen Weg sie liegt, immer auf dieselbe Weise. Interpretiert man die Knoten im Baum als Reproduktionsereignisse in der Vergangenheit und die Zweige als Vorfahren (und die Länge als die Dauer zwischen Geburt und Reproduktion, vgl. Abschnitt 1.5), so ist dieses Vorgehen einleuchtend, wenn man annimmt, dass jedes Individuum seinen Typ an alle Nachkommen vererbt. Hier werden zwei Modelle betrachtet,

- das Infinitely-Many-Sites-Modell (siehe [54]) und
- das Infinitely-Many-Alleles-Modell (siehe [24])

Im ersten Modell nimmt man an, dass jede Mutation einen anderen Abschnitt/Locus des Genoms betrifft und diesen verändert. Der Typ eines Individuums wird dann durch die Angabe der im Genom dieses Individuums mutierten Abschnitte/Loci angegeben. Im zweiten Modell nimmt man an, dass der Typ bei jeder Mutation so verändert wird, dass ein neuer Typ entsteht, der davor nicht in der Population vorgekommen ist. Beide Modelle verändern also bei jeder Mutation den Typ der betroffenen Individuen. Das Infinitely-Many-Sites-Modell beinhaltet mehr Information als das Infinitely-Many-Alleles-Modell. Ignoriert man im Infinitely-Many-Sites-Modell die Angabe, welche Loci mutiert wurden und unterscheidet nur, ob die Individuen sich hinsichtlich mutierter Loci unterscheiden, so erhält man das Infinitely-Many-Alleles-Modell.

2 Funktionale von n-Coalescents

2.1 Funktionale von n-Coalescents

Sei $n \in \mathbb{N}$ und sei $\Pi^{(n)}$ ein Ξ-n-Coalescent. Viele der Fragestellungen über n-Coalescents sind Fragestellungen über Funktionale dieser Prozesse. Aufgrund der Interpretationen des n-Coalescents als partitionswertigen Markov-Prozess, als zufälligen Graphen und/oder als Stammbaum der Stichprobe $[n]$ lassen sich Funktionale definieren und untersuchen, die Fragestellungen aus dem jeweiligen Bereich wiedergeben. Funktionale sind derzeit von starkem Interesse im Gebiet der Coalescent-Theorie. Als Funktional wird hier eine Abbildung bezeichnet, die jedem Pfad von $\Pi^{(n)}$ eine Zahl zuordnet. Einige interessante Funktionale sind in den nächsten zwei Definitionen beschrieben, wobei die Funktionale danach geordnet sind, ob sie von der Mutationsstruktur abhängen. Für die präsentierten Funktionale werden mitunter verschiedene Definitionen gegeben, diese sind synonym.

Definition 2.1.1 *(Funktionale von n-Coalescents)*
Sei $n \in \mathbb{N}$ und $\Pi^{(n)}$ ein Ξ-n-Coalescent. Definiere

1) $J_n := \min\{k \in \mathbb{N} | \mathcal{X}_k = ([n])\}$ *als die Anzahl der Sprünge der Sprungkette $(\mathcal{X}_k)_{k\in\mathbb{N}}$ von $\Pi^{(n)}$ bis zum Erreichen des absorbierenden Zustands $([n])$,*

2) C_n *als die Anzahl der Kollisionen in $\Pi^{(n)}$ bzw. die Anzahl der Nicht-Blätter-Knoten im durch $\Pi^{(n)}$ gegebenen Baum,*

3) L_n *als die Gesamtlänge aller Zweige des durch $\Pi^{(n)}$ gegebenen Baumes bzw. die Gesamtlänge dieses Baumes,*

4) $\tau_n := \inf\left\{t \in [0,\infty) | \Pi_t^{(n)} = ([n])\right\}$ *als die Absorptionszeit von $\Pi^{(n)}$, die Höhe des durch $\Pi^{(n)}$ gegebenen Baumes bzw. die Zeit zurück bis zum jüngsten Urahn (MRCA),*

5) E_n *als die Länge eines zufällig ausgewählten externen Zweiges bzw.*

$$E_n^{(i)} := \inf\left\{t \in [0,\infty) | i \text{ ist kein Singleton von } \Pi_t^{(n)}\right\}$$

als die Länge des i-ten externen Zweiges,

6) C_n^{ext} als die Anzahl der Kollisionen im durch $\Pi^{(n)}$ gegebenen Baum <u>vor</u> dem Ende eines zufällig ausgewählten externen Zweiges und

7) $L_n^{ext} = \sum_{i\in[n]} E_n^{(i)}$ als die Summe der Längen aller externer Zweige.

Definition 2.1.2 *(Funktionale von n-Coalescents mit Mutation)*
Sei $(\Pi^{(n)}, \mathsf{Mu}^{(n)})$ *ein* Ξ*-n-Coalescent mit Mutation. Betrachte das Infinitely-Many-Sites- oder das Infinitely-Many-Alleles-Modell. Definiere*

1) Seg_n *als die Gesamtzahl der Mutationen auf dem durch* $\Pi^{(n)}$ *gegebenen Baum (im Infinitely-Many-Sites-Modell auch die Anzahl der verschiedenen mutierten Loci (segregating sites) in der Stichprobe* $[n]$*),*

2) Seg_n^{ext} *als die Gesamtzahl der Mutationen auf den externen Zweigen des durch* $\Pi^{(n)}$ *gegebenen Baumes,*

3) M_n *als die Anzahl der externen Zweige, die mindestens eine Mutation tragen,*

4) N_n *als die Anzahl der Zweige, die mindestens eine Mutation tragen,*

5) K_n *als die Anzahl der in der Stichprobe* $[n]$ *vorkommenden verschiedenen Typen und*

6) $K_n(i)$ *als die Anzahl der Typen, die von genau* i *Individuen der Stichprobe* $[n]$ *getragen werden* $(i \in [n])$*.*

Es gilt $K_n = \sum_{i\in[n]} K_n(i)$ *und man nennt* $(K_n(1), \ldots, K_n(n))$ *das vollständige Allelfrequenz-Spektrum (complete allele frequency spectrum).*

Bemerkung 2.1.3

- *Die zufällige Auswahl des externen Zweiges für die Funktionale* E_n *und/oder* C_n^{ext} *ist natürlich unabhängig von* $\Pi^{(n)}$*.*

- *Für bestimmte Coalescents können diese Funktionale auch sinnvoll für den Gesamt-Coalescent auf* $\mathbb{N}$ *definiert werden. In diesem Fall wird ein solches Funktional auf dieselbe Weise, nur ohne den Index* n *notiert. Etwa existiert für alle Coalescents die externe Zweiglänge*

$$E^{(i)} := \inf\{t \in [0,\infty) | i \text{ ist kein Singleton von } \Pi_t\}$$

von Individuum i *auch im Gesamt-Coalescent (siehe Kapitel 2.3).*

- *Für Λ-Coalescents gilt $J_n = C_n$, da jeder Sprung im Coalescent genau eine Kollision beinhaltet.*
- *Die Anzahl der Typen ist für die feste Stichprobe $[n]$ definiert. Aufgrund der Austauschbarkeit von Coalescent-Prozessen ist aber die Anzahl der Typen in einer beliebigen n-elementigen Stichprobe identisch wie K_n verteilt. Auch das vollständige Allelfrequenz-Spektrum einer solchen Stichprobe hat dieselbe Verteilung wie $(K_n(1), \ldots, K_n(n))$.*

Viele dieser Funktionale wurden bereits analysiert und es existieren Ergebnisse - seien es Rekursionen in Verteilung, Verteilungseigenschaften oder die Asymptotik der Funktionale für $n \to \infty$. Die Anzahl der Kollisionen C_n wird in [26], [28], [34], [36], [39], [41], [48] und [49] behandelt. Für Poisson-Dirichlet-Coalescents mit Parametern $\alpha = 0$ und $\theta > 0$ wird die Anzahl J_n der Sprünge in [62] behandelt. Die Gesamtlänge L_n des durch $\Pi^{(n)}$ gegebenen Baumes wird in [7], [9], [26] (Analyse einer Teillänge von L_n), [27], [37], [60] (hier implizit, vergleiche [7]), [62], [66], [68], [84, S.21-23] und [87] analysiert und die Zeit τ_n zurück zum jüngsten Urahn in [35], [39], [42], [57], [62], [64] und [80]. Zur externen Zweiglänge E_n finden sich Ergebnisse in [17], [22], [35], [37], [39] und [51]. Das Funktional C_n^{ext} wird in [22], in [35] und in [39] analysiert und das Funktional L_n^{ext} in [37], [68] und [51].
Für Funktionale von n-Coalescents mit Mutation gibt es ebenfalls viele Ergebnisse. Die Gesamtzahl Seg_n der Mutationen wird in [8], [9], [26], [27], [37], [66], [68] und [88] untersucht, die Anzahl K_n der Typen und/oder das vollständige Allel-Frequenzspektrum wird in [5], [8], [9], [31] (Ewens Sampling Formel), [34], [36] und [67] untersucht. Die Funktionale M_n und N_n werden in [34] und [36] untersucht. Seg_n^{ext} wird schließlich in [37] und [68] untersucht. Hier sei noch auf [61] hingewiesen, dort werden verschiedene Ergebnisse vorgestellt, die mit diesen Funktionalen von n-Coalescents mit Mutationen zusammenhängen.
Im nächsten Abschnitt wird deutlich, dass in der biologischen Anwendung insbesondere die Funktionale τ_n, L_n, E_n, S_n und $K_n(1), \ldots, K_n(n)$, K_n sowie S_n^{ext} und L_n^{ext} von Bedeutung sind. In Tabelle 1 sind asymptotische Resultate für einige der Funktionale aus den Definitionen 2.1.1 und 2.1.2 aufgelistet.

Tabelle 1: Asymptotik einiger Funktionale von n-Coalescents für $n \to \infty$

	$\Lambda = \delta_0$	$\Lambda = \beta(2-\alpha,\alpha),\, 1<\alpha<2$	$\Lambda = U_{[0,1]}$	$\mu_{-1} < \infty$	$\mu_{-2} < \infty$
C_n	$C_n/n = 1 - \frac{1}{n} \xrightarrow{f.s.} 1$	$\frac{C_n-(\alpha-1)n}{(\alpha-1)n^{\alpha-1}} \xrightarrow{d} C^{(\alpha)}$ [41]*, [26]*	$\frac{C_n - \frac{n}{\log(n)} + \frac{n\log(\log(n))}{\log^2(n)}}{\frac{n}{\log^2(n)}} \xrightarrow{d} S$ [28]	$\frac{C_n}{\Gamma(2-\alpha)n^\alpha} \xrightarrow{d} \int_0^\infty e^{-U_t}dt$ [49]$^+$ $\frac{C_n-(1/(2m_1))\log^2(n)}{\sqrt{(m_2/(3m_1^3))\log^3(n)}} \xrightarrow{d} N(0,1)$ [48]$^{++}$	$\frac{C_n-a_n}{b_n} \xrightarrow{d} C^{(1)}$ [39]**
τ_n	$\tau_n \xrightarrow{d} \tau^{(1)}$ [57]	—	$\tau_n - \log\log(n) \xrightarrow{d} G$ [42], [35]	—	$\frac{\tau_n-a_n}{b_n} \xrightarrow{d} \tau^{(2)}$ [39]**
L_n	$\frac{L_n}{2} - \log(n) \xrightarrow{d} G$ [84], [87]	$\frac{L_n}{n^{2-\alpha}} \xrightarrow{f.s.} \frac{\Gamma(\alpha)\alpha(\alpha-1)}{2-\alpha}$ [7],[9]	$\frac{L_n - \frac{n}{\log(n)} + \frac{n\log(\log(n))}{log^2(n)}}{\frac{n}{\log^2(n)}} \xrightarrow{d} S$ [27]	$\frac{L_n}{n} \xrightarrow{d} \int_0^\infty S_t dt$ [68]	$\frac{L_n}{n} \xrightarrow{d} \int_0^\infty S_t dt$ [68]
E_n	$nE_n \xrightarrow{d} Z$ [22], [51]	—	$\log(n)E_n \xrightarrow{d} Exp(1)$ [35]	$E_n \xrightarrow{d} Exp(\mu_{-1})$	$E_n \xrightarrow{d} Exp(\mu_{-1})$
Seg_n	$\frac{\mathsf{Seg}_n - E(\mathsf{Seg}_n)}{\sqrt{Var(\mathsf{Seg}_n)}} \xrightarrow{d} N(0,1)$ [88]	$\frac{\mathsf{Seg}_n}{n^{2-\alpha}} \xrightarrow{p} r\frac{\alpha(\alpha-1)\Gamma(\alpha)}{2-\alpha}$ [8], [9]	$\frac{\mathsf{Seg}_n - r(\frac{n}{\log(n)} + \frac{n\log(\log(n))}{log^2(n)})}{r(\frac{n}{\log^2(n)})} \xrightarrow{d} S$ [27]	$\frac{\mathsf{Seg}_n}{n} \xrightarrow{d} \int_0^\infty rS_t dt$ [68]	$\frac{\mathsf{Seg}_n}{n} \xrightarrow{d} \int_0^\infty rS_t dt$ [68]
K_n	$\frac{K_n-E(K_n)}{\sqrt{Var(K_n)}} \xrightarrow{d} N(0,1)$ [31]	$\frac{K_n}{n^{2-\alpha}} \xrightarrow{p} r\frac{\alpha(\alpha-1)\Gamma(\alpha)}{2-\alpha}$ [8], [9]	$\frac{\log(n)}{n}K_n \xrightarrow{p} r$ [5]	$\frac{K_n}{n} \xrightarrow{L^p} \int_0^\infty re^{-rt}S_t dt$ [36], [34]	$\frac{K_n}{n} \xrightarrow{f.s.} \int_0^\infty re^{-rt}S_t dt$ [34]

Legende:

- $C^{(\alpha)}$ besitzt eine stabile Verteilung mit Fouriertransformierter $E(e^{it\tau^{(\alpha)}}) = \exp(-e^{-i\pi sign(t)/2}|t|^{2-\alpha})$.
- $-S$ ist (Standard-)Luria-Delbrück-verteilt, d.h. $E(e^{itS}) = e^{-\frac{1}{2}|t|+itlog(|t|)}$, $t \in \mathbb{R}$.
- a_n, b_n sind für $n \in \mathbb{N}$ geeignet zu setzen. $C^{(1)}$ (und auch $\tau^{(2)}$) kann je nach Λ (siehe **) eine stabile Verteilung, die Normalverteilung oder eine skalierte Mittag-Leffler-Verteilung als Verteilung besitzen (genauere Angaben siehe [39, Theoreme 3.1, 4.1]).
- $\tau^{(1)}$ hat Dichte $t \mapsto \sum_{m=2}^\infty (-1)^m \binom{m}{2}(2m-1)e^{-\binom{m}{2}t}$, $t \geq 0$.
- G ist Standard-Gumbel-verteilt.
- Z hat Dichte $t \mapsto 8/(2+t)^3$, $t \geq 0$.
- r bezeichnet die Mutationsrate im n-Coalescent mit Mutation.
- $p \geq 1$.

Erläuterungen:

*: In [41] und in [26] wird diese Aussage für eine größere Klasse von Λ-Maßen gezeigt.

**: Diese Aussage gilt für bestimmte Λ-Coalescents, die zusätzlich $\int_{(0,1)} \frac{|\log(x)|}{x^2}\Lambda$ erfüllen und für die $supp(\Lambda)$ nicht in einer Menge der Form $\{1-\rho\gamma^n | n \in \mathbb{N}\}$ für ein $\rho > 0$, $\gamma > 0$ liegt.

+: Die Aussage gilt für $\beta(2-\alpha,1)$-Coalescents mit $\alpha \in (0,1)$. $(U_t)_{t\geq 0}$ ist ein Subordinator mit $U_t \stackrel{d}{=} -\alpha\log(S_{\frac{\alpha t}{2-\alpha}})$ für $t \geq 0$.

++: Die Aussage gilt für $\Lambda = \beta(2,b)$, $b > 0$, wobei $m_1 := \sum_{i=0}^\infty (i+b)^{-2}$ und $m_2 := 2\sum_{i=0}^\infty (i+b)^{-3}$.

2.2 Interpretation der Funktionale im Modell

Vom rein mathematischen Standpunkt aus ist die Analyse der Funktionale aus 2.1.1 und 2.1.2 eines n-Coalescents $\Pi^{(n)}$ lohnend, da die Funktionale wichtige Kenngrößen der betrachteten Markov-Prozesse oder Bäume wiedergeben (oder als Hilfsfunktionale eine wichtige Rolle bei der Analyse anderer Funktionale spielen). $C_n + n$ ist etwa die Anzahl der Knoten im von $\Pi^{(n)}$ erzeugten Baum und τ_n die Eintrittszeit in den absorbierenden Zustand $([n])$ des Markov-Prozesses $\Pi^{(n)}$. Fasst man den n-Coalescent wie in Abschnitt 1.6 als Approximation des Stammbaums einer n-elementigen Stichprobe aus einer (sehr großen) Population auf, so lassen sich viele der betrachteteten Funktionale auch im (meist) biologischen Populationsmodell interpretieren. Für diese betrachtet man immer einen n-Coalescent mit neutraler Mutation wie in Abschnitt 1.7. Die Zeit τ_n zurück zum jüngsten Urahn und die Gesamtlänge L_n des durch $\Pi^{(n)}$ gegebenen Baumes lassen Rückschlüsse auf die Größe des gesamten Stammbaums zu, vergleiche etwa [46, 1.9.1 u. 1.9.2] oder [87, S.75]. Dies ist insbesondere für die genetische Struktur der Stichprobe von Bedeutung, da die Struktur des Baumes Auswirkungen auf die Struktur der Mutationen hat (die Position von Mutationen auf den Zweigen ist zwar unabhängig von $\Pi^{(n)}$, dennoch verändert sich etwa die Anzahl der Mutationen, wenn sich die Zweiglängen ändern). τ_n gibt die Zeitspanne an, in der sich die genetische Verwandtschaft der Stichprobe überhaupt verändern kann ($\approx \lfloor \tau_n / c_N \rfloor$ Generationen im zu Grunde liegenden Cannings-Modell für große $N \in \mathbb{N}$, vergleiche Kapitel 1.6). Die Länge E_n eines zufällig ausgewählten externen Zweiges wird dagegen als ein Maß für die genetische Einzigartigkeit eines zufällig ausgewählten Individuums aus der Stichprobe angesehen, vergleiche [17] und [74]. Auch hier ist der Zusammenhang zur Mutationsstruktur entscheidend, denn die externe Zweiglänge E_n ist die Zeit, in der Mutationen den (genetischen) Typ des ausgewählten Individuums im Vergleich zu den Typen der restlichen Stichprobe (genauer: dem Typ des nächsten Verwandten) verändern können. Analog lässt sich L_n^{ext}, die Gesamtlänge aller externer Zweige von $\Pi^{(n)}$, interpretieren als die Zeit, in der Mutationen auftreten können, die die genetische Struktur genau eines Individuums verändern. Aufgrund der Unabhängigkeit von Mutationsprozess und n-Coalescent kann man z.B. Resultate über die externe(n) Zweiglänge(n) nutzen, um Aussagen über erwartete Anzahl von Mutationen auf diesem (diesen) Zweig(en) zu machen. Ein solches Vorgehen wird etwa im Test auf Neutralität von Mutationen

in [37] verwendet (der Stammbaum der Stichprobe wird dort durch den Kingman-n-Coalescent gegeben), um die Eigenschaften der dort verwendeten Teststatistik zu bestimmen.
Betrachte nun die Funktionale S_n, die Anzahl der Mutationen, und S_n^{ext}, die Anzahl der Mutationen auf den externen Zweigen, im n-Coalescent mit Mutation. Im Infinitely-Many-Sites-Modell ist S_n die Anzahl der Loci in der genetischen Struktur der Stichprobe, die durch Mutation verändert wurden (also die segregating sites) und S_n^{ext} die Anzahl der Loci, die nur für jeweils ein Individuum verändert wurden. Diese Funktionale sind auch jenseits der unmittelbaren Interpretierbarkeit interessant. Beispielsweise wird in [37] aus S_n und S_n^{ext} eine Teststatistik für den Test auf genetische Neutralität der Mutationen gebildet, in [83] wird ein anderer Test für Neutralität unter Verwendung von S_n vorgeschlagen.
Eine Sonderstellung nimmt das vollständige Allelfrequenzspektrum $(K_n(1), \ldots, K_n(n))$ und dessen Summe, die Anzahl K_n der verschiedenen Typen in der betrachteten n-elementigen Stichprobe, ein. Die Werte dieser Funktionale lassen sich direkt aus der Stichprobe zum Zeitpunkt 0 bestimmen, sind also beobachtbar. Bei S_n und S_n^{ext} geht dies nur genau dann, wenn man die Genstruktur/den Typ des jüngsten Urahns kennt; die Werte aller anderen vorgestellten Funktionale können nicht aus der Stichprobe zum Zeitpunkt 0 bestimmt werden. Auch das vollständige Allelfrequenzspektrum und die Anzahl der verschiedenen Typen sind zuvorderst gut interpretierbare Kenngrößen der genetischen Struktur der Stichprobe und dadurch der Gesamtpopulation. Da diese gut beobachtbar sind, eignen sie sich besonders gut dazu, durch statistische Verfahren Rückschlüsse auf den zu Grunde liegenden n-Coalescent oder die Mutationsrate ziehen. Wird der Stammbaum der Stichprobe gut durch den Kingman-n-Coalescent approximiert, so wird in [43] eine Methode vorgestellt, mit deren Hilfe man etwa die Mutationsrate oder die Verteilung der Zeit τ_n zurück zum jüngsten Urahn aus den Werten von K_n und S_n schätzen kann (weitere Quellen siehe [15, S. 440]).
Ähnliche statistische Methoden, die zu Grunde liegende Parameter des n-Coalescents und/oder des Mutationsprozesses aus den Werten der Funktionale schätzen, sind noch nicht allzu gut untersucht, wenn der Stammbaum <u>nicht</u> durch den Kingman-n-Coalescent approximiert wird. Einige (bemerkenswerte) Pionierarbeit zu den statistischen Methoden in einem Nicht-Kingman-Setting findet sich in [15] und [82] sowie [29] und [77]. Die Tatsache, dass statistische Modelle für Nicht-Kingman-n-Coalescents

bislang wenig untersucht wurden, liegt hauptsächlich daran, dass die Anwendung von Nicht-Kingman-n-Coalescents als Stammbäume erst vor weniger als 5 Jahren aufkam bzw. für viele Anwendungen der Stammbaum der Stichprobe gut durch einen Kingman-n-Coalescent approximiet wird (vergleiche Abschnitt 1.6). Andererseits ist es natürlich auch schwieriger, anstatt etwa nur die Mutationsrate aus den beobachteten Daten zu schätzen, zusätzlich noch das Maß Ξ des zu Grunde liegenden Coalescent-Modells zu schätzen (vergleiche [15]).

2.3 Der Anteil $(S_t)_{t\geq 0}$ der Singletons als stochastischer Prozess und die externen Zweiglängen

Sei Π ein Ξ-Coalescent und $\Pi^{(n)} := (\rho_n(\Pi_t)_{t\geq 0})$. In Definition 1.1.16 wurde der Prozess $(S_t)_{t\geq 0}$ des Anteils der Singletons an allen Individuen in Π eingeführt. Aus dem Verhalten von $(S_t)_{t\geq 0}$ lassen sich viele Eigenschaften von Π lesen, ebenso Eigenschaften einiger der Funktionale aus den Definitionen 2.1.1 und 2.1.2. Dies gilt insbesondere im Hinblick auf deren Asymptotik für $n \to \infty$. Somit lohnt es sich, die Eigenschaften von $(S_t)_{t\geq 0}$ zu beschreiben und zu analysieren.

Offensichtlich ist $(S_t)_{t\geq 0}$ monoton fallend. Es gilt $S_0 = 1$, da der Coalescent in Diag startet. Der Prozess ist trivial für Coalescents ohne Staub, da für diese $S_t = 0$ fast sicher für alle $t > 0$ gilt. $(S_t)_{t\geq 0}$ hat für jeden Coalescent eine Beziehung zu den externen Zweiglängen $(E_n^{(i)})_{i\in[n]}$ von $\Pi^{(n)}$ und auch zu den externen Zweiglängen $(E^{(i)})_{i\in\mathbb{N}}$, von Π. Die externen Zweiglängen sind in Definition 2.1.1 definiert worden als

$$E_n^{(i)} := \inf\left\{t \in [0,\infty) | i \text{ ist kein Singleton von } \Pi_t^{(n)}\right\} \text{ und}$$

$$E^{(i)} := \inf\left\{t \in [0,\infty) | i \text{ ist kein Singleton von } \Pi_t\right\}.$$

Satz 2.3.1 *Sei Π ein Ξ-Coalescent und für $n \in \mathbb{N}$ sei $\Pi^{(n)} := (\rho_n(\Pi_t)_{t\geq 0})$. Für die externen Zweiglängen $(E_n^{(j)})_{j\in[n]}$ von $\Pi^{(n)}$ respektive $(E^{(j)})_{j\in\mathbb{N}}$, von Π und den Anteil $(S_t)_{t\geq 0}$ der Singletons von Π gilt für $i \in [n]$ respektive $i \in \mathbb{N}$:*

a) $E_n^{(i)} = \sup\left\{t \in [0,\infty) | i \textit{ ist Singleton von } \Pi_t^{(n)}\right\}$,
$E^{(i)} = \sup\left\{t \in [0,\infty) | i \textit{ ist Singleton von } \Pi_t\right\}$ *und* $\lim_{n\to\infty} E_n^{(i)} = E^{(i)}$.

b) $\left\{E_n^{(i)} < t\right\} \subseteq \left\{i \textit{ ist kein Singleton von } \Pi_t^{(n)}\right\} = \left\{E_n^{(i)} \leq t\right\}$,
$\left\{E^{(i)} < t\right\} \subseteq \{i \textit{ ist kein Singleton von } \Pi_t\} \subseteq \left\{E^{(i)} \leq t\right\}$ *für* $t > 0$.

c) $(E_n^{(j)})_{j\in[n]}$ *ist austauschbar,* $(E^{(j)})_{j\in\mathbb{N}}$ *ist austauschbar.*

d) $E^{(i)} \stackrel{d}{=} Exp(\mu_{-1})$, *wobei* $Exp(\infty) := \delta_0$.

e) *[36]* $S_t := \lim_{n\to\infty} \frac{1}{n} \sum_{j\in[n]} 1_{\left\{E_n^{(j)}>t\right\}} = \lim_{n\to\infty} \frac{1}{n} \sum_{j\in[n]} 1_{\left\{E^{(j)}>t\right\}}$
$= \lim_{n\to\infty} \frac{1}{n} \sum_{j\in[n]} 1_{\left\{E^{(j)}\geq t\right\}}$ *fast sicher.*

f) $P(E^{(1)} > t_1, \ldots, E^{(k)} > t_k) = E(\prod_{j\in[k]} S_{t_j})$ *für* $k \in \mathbb{N}$ *und* $t_1, \ldots, t_k \in [0, \infty)$.

Beweis: zu a): Hat $\Pi_t^{(n)}$ oder Π_t i nicht als Singleton, so gilt dies auch für alle $\Pi_{t'}^{(n)}$ respektive $\Pi_{t'}$ mit $t' \geq t$. Also sind $\left\{t \in [0,\infty)|i \text{ kein Singleton von } \Pi_t^{(n)}\right\}$ und $\{t \in [0,\infty)|i \text{ kein Singleton von } \Pi_t\}$ Intervalle in $\mathbb{R}$ mit oberer offener Grenze ∞. Somit folgt

$$\begin{aligned} & \inf\left\{t \in [0,\infty)|i \text{ kein Singleton von } \Pi_t^{(n)}\right\} \\ = \; & \sup\left\{t \in [0,\infty)|i \text{ Singleton von } \Pi_t^{(n)}\right\} \end{aligned}$$

und die analoge Aussage für den Gesamt-Coalescent Π. Es gilt ferner, dass falls i Singleton in Π_t oder in $\Pi_t^{(n+k)}$ ist ($k, n \in \mathbb{N},\ t \in [0,\infty)$), i auch Singleton in $\Pi_t^{(n)}$ ist. Somit gilt $E_n^{(i)} \geq E_{n+k}^{(i)} \geq E^{(i)} \geq 0$. Also existiert $\lim_{n\to\infty} E_n^{(i)}$ für alle $i \in \mathbb{N}$. Sei Π definiert auf einem Wahrscheinlichkeitsraum $(\Omega, \mathcal{F}, P)$ und sei $\omega \in \Omega$. Angenommen $\lim_{n\to\infty} E_n^{(i)}(\omega) > E^{(i)}(\omega)$, dann gibt es ein $\epsilon > 0$, sodass i ein Singleton von $\Pi_{E^{(i)}(\omega)+\epsilon}^{(n)}(\omega)$ für alle $n \in \mathbb{N}$ und damit auch von $\Pi_{E^{(i)}(\omega)+\epsilon}(\omega)$ ist. Dies ergibt den Widerspruch $E^{(i)}(\omega) \geq E^{(i)}(\omega) + \epsilon$, also gilt $\lim_{n\to\infty} E_n^{(i)} = E^{(i)}$.
zu b): folgt direkt aus der Definition der beteiligten Größen bis auf $\left\{i \text{ ist kein Singleton von } \Pi_t^{(n)}\right\} = \left\{E_n^{(i)} \leq t\right\}$. Hier folgt wiederum die Inklusion „$\subseteq$"sofort, zu zeigen ist nur noch $\left\{i \text{ ist kein Singleton von } \Pi_t^{(n)}\right\} \supseteq$

$\left\{E_n^{(i)} = t\right\}$. Um dies zu zeigen, verwende, dass jeder Pfad von $\Pi^{(n)}$ eine càdlàg Treppenfunktion mit endlich vielen Sprüngen ist. Es folgt aus

$$t_0 := E_n^{(i)} = \sup\left\{t \in [0,\infty) | i \text{ ist Singleton von } \Pi_t^{(n)}\right\},$$

dass i kein Singleton von $\Pi_{t_0+\epsilon}^{(n)}$ für jedes $\epsilon > 0$ ist. Die beschriebenen Pfadeigenschaften von Π erzwingen dann, dass i auch kein Singleton von $\Pi_t^{(n)}$ ist.
zu c): folgt direkt aus der Austauschbarkeit von Π bzw. $\Pi^{(n)}$.
zu d): OBdA sei Π ein PPP-Ξ-Coalescent konstruiert durch einen Poisson-Punktprozesses Ψ. Für alle $\epsilon > 0$ gilt

$$\{i \text{ ist kein Singleton von } \Pi_t\} \subseteq \left\{E^{(i)} \leq t\right\} \subseteq \{i \text{ ist kein Singleton von } \Pi_{t+\epsilon}\}.$$

Somit genügt es, $P(i \text{ ist kein Singleton von } \Pi_t) = 1 - e^{-\mu_{-1}t}$ für alle $t \geq 0$ zu zeigen. Da Π austauschbar ist, genügt es, dies für $i = 1$ zu zeigen. Nach der Kingman'schen Darstellung für austauschbare Partitionen (siehe Satz 1.1.9) folgt

$$P(1 \text{ ist Singleton von } \Pi_t | S_t) = S_t$$

fast sicher und somit $P(1 \text{ ist Singleton von } \Pi_t | S_t = 0) = 0$. Also ist $P(\{1 \text{ ist Singleton von } \Pi_t\} \cap \{S_t = 0\}) = 0$. Auf der Menge $\{S_t > 0\}$ sieht man anhand der Poisson-Konstruktion, dass 1 genau dann ein Singleton von Π_t ist, wenn keine Punkte von Ψ in der Menge

$$A_{coll}^{(1)}(t) := [0,t] \times \left\{a \in \mathbb{N}_0^{\mathbb{N}} | a_1 = a_j \neq 0 \text{ für ein } 1 \neq j \in \mathbb{N}\right\}$$

liegen (vergleiche Bemerkung 1.3.4). Falls keine Punkte in $A_{coll}^{(1)}(t)$ liegen, so ist 1 Singleton von Π_t. Es folgt

$$\begin{aligned} & P\Big(\left\{|\Psi \cap A_{coll}^{(1)}(t)| = 0\right\} \cap \{S_t = 0\} \Big) \\ \leq \; & P(\{1 \text{ Singleton von } \Pi_t\} \cap \{S_t = 0\}) = 0, \end{aligned}$$

also insgesamt

$$\begin{aligned} & P(|\Psi \cap A_{coll}^{(1)}(t)| = 0) = P\Big(\left\{|\Psi \cap A_{coll}^{(1)}(t)| = 0\right\} \cap \{S_t > 0\} \Big) \\ = \; & P(1 \text{ ist Singleton von } \Pi_t). \end{aligned}$$

Diese Wahrscheinlichkeit wird im Beweis von [79, Proposition 30] berechnet, es gilt

$$P\Big(|\Psi \cap A_{coll}^{(1)}(t)| = 0\Big) = e^{-\mu_{-1}t}.$$

Daraus folgt die Behauptung.
zu e): Betrachte $S_t = \lim_{n\to\infty} \frac{1}{n} \sum_{j\in[n]} 1_{\left\{\{j\} \text{ ist Block von } \Pi_t^{(n)}\right\}}$. Mit b) folgt das erste Gleichheitszeichen. Das zweite Gleichheitszeichen ist die Bemerkung nach [79, Lemma 40]. Das dritte Gleichheitszeichen folgt dann mit b) und d), da sich die Indikatoren auf beiden Seiten des Gleichheitszeichens nur auf (abzählbar vielen) Nullmengen unterscheiden.
zu f) Nach b) und d) gilt

$$\begin{aligned} & P(E^{(1)} > t_1, \dots, E^{(k)} > t_k | S_{t_1}, \dots, S_{t_k}) \\ = \ & P(\bigcap_{i\in[k]} \{i \text{ ist Singleton von } \Pi_{t_k}\} | S_{t_1}, \dots, S_{t_k}) \text{ fast sicher.} \end{aligned} \tag{23}$$

Gilt nun $S_{t_i} = 0$ für ein t_i, $i \in [k]$, so gilt auch $S_{t_k} = 0$ und es folgt

$$\begin{aligned} 0 \ \leq \ & P(\bigcap_{i\in[k]} \{i \text{ ist Singleton von } \Pi_{t_i}\} | S_{t_1}, \dots, S_{t_k}) \\ \leq \ & P(i \text{ ist Singleton von } \Pi_{t_k} | S_{t_1}, \dots, S_{t_k}) = 0 \end{aligned}$$

fast sicher, da für die fast sicher nicht negativen bedingten Wahrscheinlichkeiten $P(i \text{ ist Singleton von } \Pi_{t_k} | S_{t_1}, \dots, S_{t_k})$ und $P(i \text{ ist Singleton von } \Pi_{t_k} | S_{t_k})$ nach Definition

$$\begin{aligned} & \int_{\{S_{t_k}=0\}} P(i \text{ ist Singleton von } \Pi_{t_k} | S_{t_1}, \dots, S_{t_k}) dP \\ = \ & \int_{\{S_{t_k}=0\}} P(i \text{ ist Singleton von } \Pi_{t_k} | S_{t_k}) dP \\ \overset{(*)}{=} \ & \int_{\{S_{t_k}=0\}} S_{t_k} dP = 0 \end{aligned}$$

gilt, wobei $(*)$ wiederum aus der Kingman-Darstellung für austauschbare Partitionen folgt. Es gilt also

$$P(E^{(1)} > t_1, \dots, E^{(k)} > t_k | S_{t_1}, \dots, S_{t_k}) = 0 = \prod_{i\in[k]} S_{t_i}$$

fast sicher auf $\{S_{t_i} = 0 \text{ für ein } i \in [k]\}$.
Auf der Menge $\{S_{t_1} > 0, \dots, S_{t_k} > 0\}$ lässt sich die bedingte Wahrscheinlichkeit (23) aufgrund der Austauschbarkeit der $(E^{(i)})_{i\in\mathbb{N}}$ fast sicher folgendermaßen beschreiben (dies ist verwandt mit der Kingman-Darstellung von

$\Pi_{t_1}, \ldots, \Pi_{t_k})$:
Ziehe zu jedem Zeitpunkt t_i aus $(t_1, \ldots, t_k)$ eine Kugel aus einer Urne U_i mit abzählbar unendlich vielen Kugeln, von denen ein Anteil von S_{t_i} Kugeln weiß sind (=Singletons) und der Rest schwarz ist. Die Züge sind unabhängig. Die Wahrscheinlichkeit (23) entspricht nun der Wahrscheinlichkeit, zu jedem Zeitpunkt $t_1, \ldots, t_k$ eine weiße Kugel zu ziehen, also $\prod_{i\in[k]} S_{t_i}$. Die Unabhängigkeit der Züge im Urnenmodell entspricht der Eigenschaft, dass auf der Menge $\{S_{t_1} > 0, \ldots, S_{t_k} > 0\}$ zu jedem Zeitpunkt t_i, $i \in [k]$, unendlich viele Singletons vorhanden sind. Somit beeinflusst das Verhalten eines oder mehrerer (endlich vieler) Individuen zum Zeitpunkt t_i nicht den Anteil der Singletons zu den anderen Zeitpunkten. Es gilt also auch auf $\{S_{t_1} > 0, \ldots, S_{t_k} > 0\}$ und somit fast sicher

$$P(E^{(1)} > t_1, \ldots, E^{(k)} > t_k | S_{t_1}, \ldots, S_{t_k}) = \prod_{i\in[k]} S_{t_i}. \tag{24}$$

Integriert man diesen Ausdruck, folgt f). □

Bemerkung 2.3.2 *Das Urnenmodell aus dem vorangegangenen Beweis lässt sich auch mittels des Darstellungssatzes von de Finetti (siehe etwa [59, Satz 12.26]) rechtfertigen: Bedingt auf das zufällige Maß* $Q = \lim_{n\to\infty} \frac{1}{n} \sum_{i\in[n]} \delta_{E^i}$ *(fast sicher existent, siehe [59, Bemerkung 12.27]) sind die Zufallsvariablen* $(E^{(i)})_{i\in\mathbb{N}}$ *i.i.d. mit* $E^{(1)} \stackrel{d}{=} Q$*. Zusammen mit* $S_t = Q((t,\infty))$ *fast sicher für* $t \geq 0$ *(vergleiche Satz 2.3.1 e)) folgt mit der Turmeigenschaft der bedingten Erwartung die Richtigkeit des obigen Urnenmodells.*

Bemerkung 2.3.3 *[39, Proposition 3.1] enthält eine Version von Satz 2.3.1 d) für Simple* Λ*-Coalescents (in [39] wird eine Zusatzbedingung an* Λ *gestellt, der Beweis von Proposition 3.1 funktioniert aber auch ohne diese). Die Aussage für* Ξ*-Coalescents mit Staub findet sich in [68, Proposition 2]. Implizit findet man Satz 2.3.1 d) schon im Beweis von [79, Proposition 30].*

Mit Hilfe dieser Eigenschaften lassen sich für Ξ-Coalescents mit Staub folgende Aussagen über $(S_t)_{t\geq 0}$ zeigen.

Satz 2.3.4 *Sei* Π *ein* Ξ*-Coalescent mit Staub und* $(S_t)_{t\geq 0}$ *der zugehörige Anteil der Singletons. Dann gilt*

a) $E(S_t^\eta) = e^{-t\int_{\Delta\setminus\{0\}} \frac{1-(1-|x|)^\eta}{(x,x)} \Xi(dx)}$ *für* $t \geq 0$, $\eta \in \mathbb{N}_0$*. Insbesondere gilt* $E(S_t) = e^{-\mu_1 t}$.

b) $S_t \to 1$ *in* L^1 *für* $t \to 0$

Beweis: zu a): OBdA sei Π ein PPP-Coalescent mit erzeugendem Poisson-Punktprozess Ψ. Aus Satz 2.3.1 f) erhält man

$$E(S_t^\eta) = P(E^{(1)} > t, \ldots, E^{(k)} > t).$$

Aus (23) erhält man durch Integration

$$\begin{aligned} & P(E^{(1)} > t, \ldots, E^{(k)} > t) = P(\{1\}, \ldots, \{k\} \text{ sind Blöcke von } \Pi_t) \\ = \ & P(\bigcap_{i\in[k]} \left\{|\Psi \cap A_{coll}^{(i)}(t)| = 0\right\}) = P(|\Psi \cap (\bigcup_{i\in[k]} A_{coll}^{(i)}(t))| = 0) \end{aligned}$$

mit $A_{coll}^{(i)}(t) := [0,t] \times \left\{a \in \mathbb{N}_0^{\mathbb{N}} | a_i = a_j \neq 0 \text{ für mindestens ein } i \neq j \in \mathbb{N}\right\}$, $i \in \mathbb{N}$ (vergleiche Beweis von Satz 2.3.1 d)). Nun gilt

$$\bigcup_{i\in[k]} A_{coll}^{(i)}(t) = [0,t] \times \left\{a \in \mathbb{N}_0^{\mathbb{N}} | a_i = a_j \neq 0 \text{ für mindestens ein } i \in [k], i \neq j \in \mathbb{N}\right\}$$

und mit den Maßen $\mathcal{L}$, L aus der Poissonkonstruktion (siehe Kapitel 1.7)

$$\begin{aligned} & P(|\Psi \cap (\bigcup_{i\in[k]} A_{coll}^{(i)}(t))| = 0) = e^{-\mathcal{L}(\bigcup_{i\in[k]} A_{coll}^{(i)}(t))} \\ = \ & e^{-tL(\left\{a\in\mathbb{N}_0^{\mathbb{N}} | a_i = a_j \neq 0 \text{ für mindestens ein } i\in[k], i\neq j\in\mathbb{N}\right\})} = e^{-t\int_{\Delta\setminus\{0\}} \frac{1-(1-|x|)^k}{(x,x)} \Xi(dx)}, \end{aligned}$$

da für das Maß P_x aus der Poisson-Konstruktion

$$\begin{aligned} & P_x(\left\{a \in \mathbb{N}_0^{\mathbb{N}} | a_i = a_j \neq 0 \text{ für mindestens ein } i \in [k], i \neq j \in \mathbb{N}\right\}) \\ = \ & 1 - P_x(\left\{a \in \mathbb{N}_0^{\mathbb{N}} | a_i \neq a_j \text{ für alle } i \in [k], i \neq j \in \mathbb{N} \text{ mit } a_j \neq 0\right\}) \\ = \ & 1 - (1 - |x|)^k \end{aligned}$$

gilt. Die Gültigkeit dieses Gleichungssystems sieht man leicht ein, wenn man P_x als gemeinsame Verteilung von i.i.d. Zufallsvariablen $(\xi_j^{(x)})_{j\in\mathbb{N}}$ (vergleiche Kapitel 1.3) auffasst. Nach dem Borel-Cantelli-Lemma folgt, dass jeder Wert $j \in \mathbb{N}$ mit $P(\xi_1^{(x)} = j) > 0$ von unendlich vielen $\xi_j^{(x)}$'s angenommen wird (vergleiche Bemerkung 1.3.4). Somit kann $a_i \neq a_j$ für alle $j \neq i$ mit $a_j \neq 0$ nur dann gelten, falls $a_i = 0$ ist. Anschaulich lässt sich die eben berechnete Wahrscheinlichkeit auch in der Poisson-Konstruktion des Coalescents

wiederfinden. Hat der Coalescent unendlich viele Blöcke, dann kollidiert der j-te Block fast sicher nur dann nicht für einen Poisson-Punkt (T, X), wenn $X_j = 0$ gilt.
zu b): Es gilt mit Satz 2.3.1 d) und f) mit $k = 1$ (oder mit Teil a) dieses Satzes 2.3.4)

$$E(|S_t - 1|) = E(1 - S_t) = 1 - e^{-\mu_{-1} t} \to 0 \text{ für } t \to 0. \qquad \square$$

Bemerkung 2.3.5 *Die Aussage $E(S_t) = e^{-\mu_{-1} t}$ aus Satz 2.3.4 a) folgt somit auch direkt aus Satz 2.3.1 d) und f).*

Für Coalescents mit Staub ist $(-\log(S_t))_{t\geq 0}$ (bzw. eine zugehörige càdlàg-Modifikation) ein Subordinator mit exponentieller Vernichtung. Ein solcher Prozess ist auf folgende Art definiert.

Definition 2.3.6 *(Subordinator mit exponentieller Vernichtung) Ein numerischer stochastischer Prozess $X := (X_t)_{t\geq 0}$ heißt Subordinator mit exponentieller Vernichtung, falls $X_t = Y_t + \infty \cdot 1_{\{T\leq t\}}$ für $t \geq 0$, wobei $Y := (Y_t)_{t\geq 0}$ ein Subordinator (mit càdlàg Pfaden) und T eine Exp(α)-verteilte Zufallsvariable unabhängig von Y ist mit $\alpha \geq 0$. Hier wird als Konvention $\infty \cdot 0 = 0$ gesetzt und $T \equiv \infty$ für $\alpha = 0$. Für $\alpha = 0$ ist $X \equiv Y$ ein ,,normaler"Subordinator. α heißt die Vernichtungsrate von X. Ist d die Drift und ρ das Lévy-Maß von Y, so sind die Verteilungen von X durch das Triple (d, ρ, α) eindeutig bestimmt (da càdlàg). Man nennt X einen (d, ρ, α)-Subordinator. Für $d = 0$ heißt Y driftfrei, in diesem Fall nennt man auch X driftfrei.*

Bemerkung 2.3.7 *Auch die Verteilung einer nicht negativen, numerischen Zufallsvariablen X ist eindeutig durch ihre Laplace-Transformierte $\psi : \eta \mapsto E(e^{-\eta X})$, $\eta \in [0, \infty)$, bestimmt, da die Zufallsvariable e^{-X} nur Werte in $[0, 1]$ annimmt und somit in Verteilung eindeutig durch ihre Momente bestimmt ist. Ferner wird aus Stetigkeitsgründen $\psi(0) = \lim_{\eta\searrow 0} \psi(\eta)$ gesetzt.*

Satz 2.3.8 *[36] Sei Π ein Ξ-Coalescent mit Staub. Es existiert eine Modifikation $X := (X_t)_{t\geq 0}$ von $(-\log(S_t))_{t\geq 0}$, die ein (d, ρ, α)-Subordinator ist mit*

- *$d = 0$ (also X driftfrei),*
- *das Lévy-Maß ρ ist die Einschränkung auf $(0, \infty)$ des Bildmaßes von ν unter der Abbildung $x \mapsto -\log(1 - |x|)$ und*

- $\alpha = \nu(\Delta^*)$, *wobei* $\nu(dx) := x^{-2}\Lambda(dx)$.

Der Laplace-Exponent des zu X gehörenden Subordinators ist

$$\Phi(\eta) \;=\; \int_{\Delta\setminus(\Delta^*\cup\{0\})} \frac{1-(1-|x|)^\eta}{(x,x)}\,\Xi(dx), \qquad \eta \geq 0. \tag{25}$$

Desweiteren gilt für die Laplace-Transformierte ψ_t *von* X_t, $t \geq 0$,

$$\psi_t(\eta) = E(S_t^\eta) = e^{-t(\Phi(\eta)+\alpha)} = \lim_{n\to\infty} g_n E(1-(\frac{I_n}{n})^\eta) \text{ für } t,\eta \geq 0, \tag{26}$$

wobei g_n *die totale Rate im Zustand* n *des Block-Zählprozesses* $(|\rho_n(\Pi_t)|)_{t\geq 0}$ *ist und* I_n *verteilt ist wie der Zustand nach dem ersten Sprung im Block-Zählprozess.*

Bemerkung 2.3.9 *Die Darstellung* $\psi_t(\eta) = \lim_{n\to\infty} g_n E(1-(\frac{I_n}{n})^\eta)$ *für* $t,\eta \geq 0$ *stammt aus [68, Lemma 6].*

Beweis:(von Satz 2.3.8) Dies ist eine Verallgemeinerung von [71, Proposition 26], dort wird die Aussage für Λ-Coalescents gezeigt. Der Beweis der Verallgemeinerung verläuft analog zu [71, Proposition 26].
Sei zunächst $Y := (Y_t)_{t\geq 0}$ ein driftfreier Subordinator mit Lévy-Maß ρ definiert wie in obigem Satz und T eine von Y unabhängige Zufallsvariable mit $T \overset{d}{=} Exp(\nu(\Delta^*))$. Setze $X_t := Y_t + \infty \cdot 1_{\{T\leq t\}}$ für $t \geq 0$. Sei Φ' der Laplace-Exponent von Y. Dann gilt für die Laplacetransformierte ψ_t von X_t

$$\psi_t(\eta) = E(e^{-\eta X_t}) = e^{-t\nu(\Delta^*)}E(e^{-\eta Y_t}) = e^{-t(\nu(\Delta^*)+\Phi'(\eta))} \tag{27}$$

für $\eta > 0$. Für $\eta = 0$ gilt $\psi_t(0) = \lim_{\eta\searrow 0}\psi_t = e^{-t(\Delta^*)}$. Rechne nun den Laplace-Exponenten von Φ' explizit aus. Es gilt nach dem Transformationssatz für Maße und mit dem Lévy-Maß $\rho = \nu_{x\mapsto -\log(1-|x|)}$ von Y

$$\Phi'(\eta) = \int_{(0,\infty)} 1-e^{-\eta u}\nu_{x\mapsto -\log(1-|x|)}(du) = \int_{\Delta\setminus(\Delta^*\cup\{0\})} 1-(1-|x|)^\eta \nu(dx)$$

für $\eta \geq 0$. Daraus erhält man

$$\begin{aligned}\psi_t(\eta) &= e^{-t(\nu(\Delta^*)+\int_{\Delta\setminus(\Delta^*\cup\{0\})}1-(1-|x|)^\eta\nu(dx))}\\ &= \begin{cases} e^{-t\int_{\Delta\setminus\{0\}}1-(1-|x|)^\eta\nu(dx)} & \text{für } \eta > 0.\\ e^{-t\nu(\Delta^*)} & \text{für } \eta = 0.\end{cases}\end{aligned}$$

Durch Vergleich von $\psi_t(\eta)$ mit dem η-ten Moment von S_t fällt

$$\psi_t(\eta) = E((e^{-X_t})^\eta) = E(S_t^\eta) = E(e^{-\eta \log(S_t)}) \tag{28}$$

für $\eta \in \mathbb{N}$ und $t \geq 0$ auf. Da S_t und e^{-X_t} für alle $t \geq 0$ Zufallsvariablen mit Werten in $[0,1]$ sind, sind auch die Verteilungen dieser Zufallsvariablen (und damit auch die Verteilung von X_t) eindeutig durch ihre Momente bestimmt. Insbesondere gilt deshalb

$$-\log(S_t) \stackrel{d}{=} X_t \text{ für alle } t \geq 0, \tag{29}$$

$$P(-\log(S_t) = \infty) = P(X_t = \infty) = 1 - e^{-t\nu(\Delta^*)}. \tag{30}$$

Betrachte den Prozess $(Z_t)_{t\geq 0} = (-\log(S_t))_{t\geq 0}$ für einen Ξ-Coalescent mit Staub. Es wird nun gezeigt, dass eine Modifikation dieses Prozesses existiert, die ein Subordinator ist. Dazu kann man analog zu [52, Lemma 15.11] vorgehen. Zunächst zeigt man für $t_2 \geq t_1 \geq 0$

$$P(Z_{t_2} - Z_{t_1} \in \cdot | \mathcal{F}_{t_1}) = P(Z_{t_2 - t_1} \in \cdot) \; P\text{-fast sicher auf } \{Z_{t_1} < \infty\} \tag{31}$$

für $\mathcal{F}_s := \sigma(Z_t, t \leq s)$, also die bedingte Unabhängigkeit und die Stationarität der Zuwächse von $(Z_t)_{t\geq 0}$. Betrachte den Zuwachs $-\log(S_{t_2}) + \log(S_{t_1})$ für $t_2 \geq t_1 \geq 0$ auf der Menge $A := \{Z_{t_1} < \infty\}$. Auf dieser gilt $S_{t_1} > 0$, somit gilt dort $\sum_{i\in[n]} 1_{\{\{i\} \text{ ist Block von } \rho_n(\Pi_{t_1})\}} \sim S_{t_1} \cdot n$ fast sicher für $n \to \infty$. Also hat Π_{t_1} unendlich viele Singletons auf A (fast sicher). Des Weiteren gilt $-\log(S_{t_2}) + \log(S_{t_1}) = -\log(S_{t_2}/S_{t_1})$ auf A. Alle folgenden Überlegungen betreffen nur die Menge A.

S_{t_2}/S_{t_1} hängt nur vom weiteren Verhalten der (unendlich vielen) Singletons von Π_{t_1} zum Zeitpunkt t_2 ab. Betrachte $(\rho_{\mathcal{M}_n}(\Pi_{t_1+t}))_{t\geq 0}$ für jedes $n \in \mathbb{N}$, wobei $\mathcal{M}_n := \{i \in [n] | \{i\} \text{ ist Block von } \rho_n(\Pi_{t_1})\}$ die Menge der Singletons von $\rho_n(\Pi_{t_1})$ ist. $(\rho_{\mathcal{M}_n}(\Pi_{t_1+t}))_{t\geq 0}$ ist nach Natural Coupling und der Markov-Eigenschaft von Π wieder ein Ξ-$|\mathcal{M}_n|$-Coalescent (nach Umnummerierung) auf A, der bedingt $\mathcal{M}_n$ unabhängig von $(\Pi_u)_{0\leq u\leq t_1}$ ist. Sei nun $(\mathcal{M}_{n_k})_{k\in\mathbb{N}}$ eine streng isotone Teilfolge von $(\mathcal{M}_n)_{n\in\mathbb{N}}$ mit $|\mathcal{M}_{n_k}| = k$ für alle $k \in \mathbb{N}$. Dann ist $(\Pi'_t)_{t\geq 0}$, definiert über die Werte der Restriktionen $(\rho_k(\Pi'_t))_{t\geq 0} := r'(\rho_{\mathcal{M}_{n_k}}(\Pi_{t_1+t})_{t\geq 0}$ für alle $k \in \mathbb{N}$, ein Ξ-Coalescent, wobei r' eine endliche Menge $M \subseteq \mathbb{N}$ ordnungserhaltend auf $[|M|]$ abbildet. Es gilt für den Anteil S'_t der Singletons in Π'_t nach Konstruktion

$$S'_t \cdot S_{t_1} = S_{t_1+t} \;\; P\text{-fast sicher (auf } A) \; \forall \; t \geq 0,$$

somit auch

$$S'_{t_2-t_1} = \frac{S_{t_2}}{S_{t_1}} \text{ } P\text{-fast sicher (auf } A). \tag{32}$$

Nach Konstruktion ist Π' unabhängig von $(\Pi_u)_{0 \leq u \leq t_1}$, also ist der Zuwachs $Z_{t_2} - Z_{t_1} = -\log(S_{t_2}) + \log(S_{t_1})$ unabhängig von $(\Pi_u)_{0 \leq u \leq t_1}$, also insbesondere von $\mathcal{F}_{t_1}$. (32) liefert die gewünschte bedingte Stationarität der Zuwächse. Insgesamt folgt (31).
Bisher wurde gezeigt, dass $(Z_t)_{t \geq 0}$ dieselben eindimensionalen Verteilungen wie ein $(0, \rho, \alpha)$-Subordinator besitzt und die Zuwachsbedingung (31) erfüllt. Des Weiteren ist $(Z_t)_{t \geq 0}$ monoton steigend, da $(S_t)_{t \geq 0}$ monoton fallend ist. Betrachte nun $\zeta := \inf\{t | Z_t = \infty\}$. Da

$$\{Z_{t+\epsilon} < \infty\} \subseteq \{\zeta > t\} \subseteq \{Z_t < \infty\}$$

für alle $\epsilon > 0$ aufgrund der Monotonie von Z, folgt $\zeta \stackrel{d}{=} Exp(\alpha)$ aus (30). Definiere für $t \geq 0$ das Wahrscheinlichkeitsmaß

$$P_t(\cdot) := P(Z_t \in \cdot | Z_t < \infty).$$

Alle Maße P_t, $t \geq 0$, haben nur auf $[0, \infty)$ Masse. Mit Hilfe der Gedächtnislosigkeit der Exponentialverteilung und der Eigenschaft (31) von Z folgt $P_s * P_t = P_{s+t}$, d.h. $(P_t)_{t \geq 0}$ ist eine Faltungshalbgruppe. Somit existiert nach dem Kolmogoroff'schen Fortsetzungssatz ein monoton wachsender Markov-Prozess $V := (V_t)_{t \geq 0}$ mit unabhängigen stationären Zuwächsen, Startzustand 0 und eindimensionalen Verteilungen $(P_t)_{t \geq 0}$. Da $(Z_t)_{t \geq 0}$ stochastisch stetig ist, überträgt sich die stochastische Stetigkeit auf V via

$$P(V_t > \epsilon) = P(Z_t > \epsilon | Z_t < \infty) \leq \frac{P(Z_t > \epsilon)}{P(Z_t < \infty)} \to 0$$

für $t \to 0$. Es genügt die stochastische Stetigkeit in 0 zu zeigen, da V unabhängige und stationäre Zuwächse besitzt (etwa [78, S.4]). Sei nun T' eine $Exp(\alpha)$-verteilte Zufallsvariable, die unabhängig von Y ist. Betrachte den Prozess $(V_t + \infty \cdot 1_{\{T' \leq t\}})_{t \geq 0}$. Vergleicht man nun

$$Z = (Z_t 1_{\{Z_t < \infty\}} + \infty \cdot 1_{\{Z_t = \infty\}})_{t \geq 0} \text{ mit } (V_t + \infty \cdot 1_{\{T \leq t\}})_{t \geq 0},$$

so sieht man mit (31), dass beide Prozesse dieselben endlich dimensionalen Verteilungen haben. Somit kann man Z auffassen als einen stochastisch stetigen Prozess V mit unabhängigen stationären Zuwächsen, der zu einem exponentialverteilten Zeitpunkt T', der von V unabhängig ist, auf ∞ gesetzt wird (vergleiche etwa [52, Theorem 6.10]). V besitzt nach [78, Theorem 11.5] eine càdlàg-Modifikation, die ein Subordinator ist. Ersetze V durch diese. Dann ist $(V_t + \infty \cdot 1_{\{T \leq t\}})_{t \geq 0}$ die gewünschte Modifikation von Z, also ein Subordinator mit exponentieller Vernichtung. Das Tripel (d, ρ, α), das die Verteilung der càdlàg-Modifikation von Z charakterisiert, ist eindeutig durch die Verteilung von Z_1 bestimmt. Diese wurde aber schon berechnet und es gilt $Z_1 \stackrel{d}{=} X_1$, wobei X_1 aus dem am Anfang des Beweises betrachteten Subordinators X mit exponentieller Vernichtung stammt, der genau die im Satz geforderten Verteilungseigenschaften besitzt. Schließlich ist in (26) der erste Teil die schon in (27) berechnete Laplace-Transformierte von X_t und die zweite Gleichung ist [68, Lemma 6]. $\square$

Bemerkung 2.3.10 *[36] Für Simple Coalescents ist das Lévy-Maß ρ der Subordinator-Modifikation X von $(-\log(S_t))_{t\geq 0}$ endlich, da nach dem vorangegangenen Satz 2.3.8*

$$\begin{aligned} \rho((0,\infty)) &= \nu(\{x \in \Delta | -\log(1-|x|) \in (0,\infty)\}) \\ &= \nu(\Delta \setminus (\Delta^* \cup \{0\})) \leq \mu_{-2} < \infty \end{aligned}$$

gilt. Sei Y der zu X gehörige driftfreie Subordinator mit Lévy-Maß ρ. Y ist ein zusammengesetzter Poisson-Prozess (compound Poisson process) mit

$$Y_t := \sum_{i \in [N'_t]} \eta'_i,$$

wobei $N' := (N'_t)_{t\geq 0}$ ein homogener Poisson-Prozess mit Intensität $\beta := \nu(\Delta \setminus (\Delta^* \cup \{0\}))$ und $\eta'_1, \eta'_2, \ldots$ von N' unabhängige i.i.d. Zufallsvariable mit Verteilung $\eta'_1 \stackrel{d}{=} \frac{1}{\beta}\rho$ sind, dies folgt etwa aus [11, Proposition 2 u. Korollar 3]. Man kann eine analoge Darstellung für X angeben. Sei dazu $\alpha = \nu(\Delta^*)$ die Vernichtungsrate von X. Es gilt

$$X_t := \sum_{i \in [N_t]} \eta_i, \tag{33}$$

wobei $N := (N_t)_{t\geq 0}$ ein homogener Poisson-Prozess mit Intensität $\alpha + \beta = \mu_{-2}$ ist und $\eta_1, \eta_2, \ldots$ von N unabhängige i.i.d. Zufallsvariable mit Verteilung $\eta_1 \stackrel{d}{=} \frac{1}{\mu_{-2}}\rho'$ sind, wobei ρ' das durch $\rho' := \rho + \alpha\delta_\infty$ definierte Maß auf $(0,\infty]$ ist. Um dies einzusehen, fasse zunächst die Menge

$\Psi_1 := \{(T'_i, \eta'_i) | i \in \mathbb{N},\ T_i$ *i-ter Sprungzeitpunkt von* $N'\}$ *als einen Poisson-Punktprozess auf* $[0,\infty) \times (0,\infty]$ *mit Intensitätsmaß* $\lambda_{[0,\infty)} \otimes \rho$ *auf (*ρ *wird hier als Maß auf* $(0,\infty]$ *aufgefasst). Den Sprungzeitpunkt* T*, an dem* X *nach* ∞ *springt, kann man als das Infimum aller Werte der ersten (t-)Koordinate eines von* Ψ_1 *unabhängigen Poisson-Punktprozesses* Ψ_2 *auf* $[0,\infty) \times (0,\infty]$ *mit Intensitätsmaß* $\lambda_{[0,\infty)} \otimes \alpha\delta_\infty$ *auffassen. Dann ist* $\Psi_1 \cup \Psi_2$ *nach dem Superpositionstheorem [58, S.16] wiederum ein Poisson-Punktprozess auf* $[0,\infty) \times (0,\infty]$ *mit Intensitätsmaß* $\lambda_{[0,\infty)} \otimes \rho'$ *und es gilt*

$$X = \sum_{(T,\eta)\in\Psi_1\cup\Psi_2, T\leq t} \eta. \tag{34}$$

Man kann analog zur Argumentation auf Seite 22 $(\eta)_{(T,\eta)\in\Psi_1\cup\Psi_2}$ *als Folge von i.i.d. Zufallsvariablen auffassen. Ordnet man nun* $(\eta)_{(T,\eta)\in\Psi_1\cup\Psi_2}$ *in* T *aufsteigend als* $(T_i, \eta_i)_{i\in\mathbb{N}}$*, so bleibt aufgrund der Produktstruktur des Intensitätsmaßes von* $\Psi_1 \cup \Psi_2$ *die Folge* $(\eta_i)_{i\in\mathbb{N}}$ *i.i.d. mit* $\eta_1 = \eta$ *für ein beliebiges* $(T,\eta) \in \Psi_1 \cup \Psi_2$*. Nach dem Abbildungssatz [58, S.18] für Poisson-Punktprozesse ist die Projektion* $\{T | (T,\eta) \in \Psi_1 \cup \Psi_2\}$ *auf die erste Koordinate von* $\Psi_1 \cup \Psi_2$ *ein Poisson-Punktprozess auf* $[0,\infty)$ *mit Intensitätsmaß* $\rho'((0,\infty]) \cdot \lambda_{[0,\infty)} = \mu_{-2} \cdot \lambda_{[0,\infty)}$ $(\mu_{-2} = \alpha + \beta)$*. Mit anderen Worten ist also* $N := (N_t)_{t\geq 0}$ *definiert durch* $N_t := |\{T | (T,\eta) \in \Psi_1 \cup \Psi_2, T \leq t\}|$ *für* $t \geq 0$ *ein homogener Poisson-Prozess mit Intensität* μ_{-2}*.* $(N_t)_{t\geq 0}$ *und* $(\eta_i)_{i\in\mathbb{N}}$ *sind nach Konstruktion (Produktstruktur des Intensitätsmaßes) unabhängig. Schreibt man nun (34) in Termen von* $(N_t)_{t\geq 0}$ *und* $(\eta_i)_{i\in\mathbb{N}}$ *um, erhält man Darstellung (33) von* X*.*
In diesem Fall gilt außerdem

$$\begin{aligned} & \rho'(A) = \rho(A \cap \mathbb{R}) + \alpha 1_{\{\infty\in A\}} \\ = \ & \nu(\{x \in \Delta | -\log(1-|x|) \in A \cap (0,\infty)\}) + \nu(\Delta^*) 1_{\{\infty\in A\}} \\ = \ & \nu(\{x \in \Delta | -\log(1-|x|) \in A\}) \end{aligned}$$

für $A \in \mathcal{B}((0,\infty])$*, also ist* ρ' *die Einschränkung auf* $(0,\infty]$ *des Bildmaßes von* $\nu(dx)$ *unter der Abbildung* $x \mapsto -\log(1-|x|)$*.*
Somit ist auch X *ein zusammengesetzter Poisson-Prozess, falls man für die zu summierende i.i.d. Folge numerische Zufallsvariablen zulässt. In diesem Fall wird der Prozess* X *(im Rahmen dieser Arbeit) auch ein numerischer zusammengesetzter Poisson-Prozess genannt.*

Für einige Ξ-Coalescents mit Staub wird nun mit Hife von Satz 2.3.8 und Bemerkung 2.3.10 die càdlàg-Modifikationen X von $(-\log(S_t))_{t\geq 0}$ analysiert, die Subordinatoren mit (oder ohne) exponentieller Vernichtung sind.

Beispiel 2.3.11 *(Dirac-Coalescents)[36] Aus Beispiel 1.1.21 ist bekannt, dass jeder Dirac-Coalescent, also ein* Ξ*-Coalescent mit* $\Xi = \delta_c$ *für* $c \in \Delta \setminus \{0\}$*, ein Simple Coalescent ist.* X *ist in diesem Fall ein zusammengesetzter Poisson-Prozess, der genau dann* ∞ *annehmen kann, falls* $\alpha := \nu(\Delta^*) = (c,c)^{-1}1_{\{c\in\Delta^*\}} > 0$ *ist, also falls* $c \in \Delta^*$*. Das Lévy-Maß auf* $(0,\infty)$ *des zugehörigen Subordinators ohne exponentielle Vernichtung ist* $\rho = (c,c)^{-1}\delta_{-\log(1-|c|)}$*, falls* $c \notin \Delta^*$*; ansonsten ist* ρ *das Nullmaß. Somit ist* X *ein* $(0,\rho,\alpha)$*-Subordinator. Der Laplace-Exponent des zugehörigen Subordinators ohne exponentielle Vernichtung ist*

$$\Phi(\eta) = (c,c)^{-1}(1-(1-|c|)^\eta)1_{\{c\notin\Delta^*\}}, \ \eta > 0. \tag{35}$$

Liegt $c \in \Delta^*$*, so nimmt* X *nur die Werte 0 und* ∞ *an und springt nach einer* $Exp(\alpha)$*-verteilten Wartezeit von 0 nach* ∞*. Gilt* $c \notin \Delta^*$*, so ist* X *selbst ein Subordinator ohne exponentielle Vernichtung mit dem oben angegebenen Lévy-Maß.*

Beispiel 2.3.12 *(Beta-Coalescents)[36] Sei* Π *ein Beta-Coalescent mit* $\Lambda = \beta(a,b)$ *für* $a > 1, b > 0$*, also ein* Ξ*-Coalescent mit* $\Xi(A,0,0\ldots) = \Lambda(A)$ *für alle* $A \in \mathcal{B}([0,1])$*. Aus den Beispielen 1.1.19 und 1.1.21 folgt, dass* Π *ein Coalescent mit Staub und für* $a > 2$ *sogar ein Simple Coalescent ist. Da* $\Xi(\Delta^*) = 0$ *(und damit auch* $\nu(\Delta^*) = 0$*) ist, ist* X *ein gewöhnlicher Subordinator, es gibt also keine exponentielle Vernichtung (bzw. die Vernichtungsrate ist 0). Somit ist* X *ein* $(0,\rho,0)$*-Subordinator. Das Lévy-Maß* ρ *hat die Dichte* $y \mapsto (Beta(a,b))^{-1}(1-e^{-y})^{a-3}(e^{-y})^b$*,* $y \in (0,\infty)$*, bezüglich* $\lambda_{(0,\infty)}$*, dies folgt aus*

$$\rho((0,y]) = \nu((0,1-e^{-y}]) = \int_{(0,1-e^{-y}]} u^{-2}\frac{u^{a-1}(1-u)^{b-1}}{Beta(a,b)}\lambda(du),$$

wenn man im Integral die Substitution $u = 1-e^{-x}$ *vornimmt. Der Laplace-Exponent von* X *ist*

$$\Phi(\eta) \ = \ \frac{1}{Beta(a,b)}\int_0^1 \frac{1-(1-u)^\eta}{u^2}u^{a-1}(1-u)^{b-1}\,du, \qquad \eta \geq 0, \tag{36}$$

wobei man dies mit Hilfe von $1-(1-u)^\eta = \sum_{i=1}^\infty \binom{\eta}{i}(-1)^{i+1}u^i$ *und den Rechenregeln für Beta- und Gammafunktionen auch als*

$$\begin{aligned}\Phi(\eta) &= \frac{1}{b(a,b)}\sum_{i=1}^{\infty}\binom{\eta}{i}(-1)^{i+1}Beta(a+i-2,b)\\ &= \frac{a+b-1}{a-1}\sum_{i=1}^{\infty}\binom{\eta}{i}(-1)^{i+1}\prod_{j=1}^{i-1}\frac{a-2+j}{a+b-2+j}, \qquad \eta\geq 0.\end{aligned}$$

darstellen kann. Insbesondere gilt

$$\Phi(1)=(a+b-1)/(a-1),\ \Phi(2)=(a+2b-1)/(a-1). \tag{37}$$

Für spezielle Wahl der Parameter kann man die Darstellung von $\Phi(\eta)$, $\eta\geq 0$, *stark vereinfachen. Es gilt etwa für einen* $\beta(2-p,p)$*-Coalescent* ($p\in(0,1)$)

$$\Phi(\eta) = \frac{1}{1-p}\sum_{i=1}^{\infty}\binom{\eta}{i}\binom{p-1}{i-1} = \frac{\eta\Gamma(\eta+p)}{(1-p)\Gamma(p+1)\Gamma(\eta+1)}, \qquad \eta\geq 0.$$

Beispiel 2.3.13 *[36] Sei* Ξ *konzentriert auf* Δ^* *(d.h.* $\Xi(\Delta)=\Xi(\Delta^*)$*) mit* $\mu_{-1}<\infty$. *Beispiele für Coalescents zu solchen Maßen sind die in Beispiel 2.3.11 behandelten Dirac-Coalescents mit Punktmasse von* Ξ *in* Δ^*, *insbesondere der sternförmige Coalescent, oder die Poisson-Dirichlet-Coalescents. Unter dieser Bedingung ist*

$$\alpha=\nu(\Delta^*)=\int_{\Delta\setminus\{0\}}\frac{1}{(x,x)}\Xi(dx)=\int_{\Delta\setminus\{0\}}\frac{|x|}{(x,x)}\Xi(dx),$$

d.h. $\alpha=\mu_{-2}=\mu_{-1}<\infty$, *also ist* Π *ein Simple Coalescent.* X *ist somit ein Subordinator mit exponentieller Vernichtung mit Vernichtungsrate* α *und ein zusammengesetzter numerischer Poisson-Prozess. Das zugehörige Lévy-Maß* ρ *ist das Nullmaß (vergleiche Bemerkung 2.3.10) und der zugehörige Laplace-Exponent* Φ *erfüllt* $\Phi(\eta)=0$ *für alle* $\eta\geq 0$. *Also gilt wie im Falle der Dirac-Coalescents mit Masse in* Δ^*, *dass* X *nur die Werte 0 und 1 annimmt und nach einer Exp(*α*)-verteilten Wartezeit von 0 nach* ∞ *springt.*

2.4 Rekursionen für Funktionale [35], [36]

Die Eigenschaften Natural Coupling (Satz 1.2.1) und Temporal Coupling (Satz 1.2.3) eines n-Coalescents $\Pi^{(n)}$ zeigen, dass man in dem von $\Pi^{(n)}$ gegebenen Baum Teilbäume finden kann, die selbst m-Coalescents sind ($m\in[n]$).

Dies kann man ausnutzen, um Verteilungsrekursionen für die Funktionale von $\Pi^{(n)}$ zu finden.

Satz 2.4.1 *[35],[36] Für $n \in \mathbb{N}$ sei $\Pi^{(n)}$ ein Ξ-n-Coalescent. Sei I_n eine Zufallsvariable, die verteilt ist wie der Block-Zählprozess von $\Pi^{(n)}$ direkt nach dem ersten Sprung, d.h. $P(I_n = k) = g_{nk}/g_n$ für $k \in [n-1]$ und die Raten g_{nk} und g_n des Block-Zählprozesses. I_n habe jeweils die unten beschriebenen Unabhängigkeiten. Dann gilt für die Funktionale τ_n und C_n*

a) $\tau_1 = 0$ *und* $\tau_n \stackrel{d}{=} T_n + \tau_{I_n}$,

b) $C_1 = 0$ *und* $C_n \stackrel{d}{=} V_n + C_{I_n}$

für $n \in \{2, 3, \ldots\}$, wobei in

a) die Zufallsvariablen I_n, T_n unabhängig von $\tau_2, \ldots, \tau_{n-1}$ sind, I_n und T_n unabhängig sind und $T_n \stackrel{d}{=} Exp(g_n)$.

b) I_n, V_n unabhängig von $C_1, \ldots, C_{n-1}$ sind sowie V_n verteilt ist wie die Anzahl der Kollisionen in $\Pi^{(n)}$, die im ersten Sprung von $\Pi^{(n)}$ stattfinden.

Beweis: Beide Rekursionen werden dadurch bewiesen, dass man den Coalescent am ersten Sprung stoppt und jeweils die Beiträge zum Wert des betreffenden Funktionals des n-Coalescent-Baumes vor und nach dem ersten Sprung bestimmt. Hierzu verwendet man Natural und Temporal Coupling. Sei immer $\Pi^{(n)}$ ein Ξ-n-Coalescent.
zu a): Die Wartezeit T_n bis zum ersten Sprung von $\Pi^{(n)}$ ist $Exp(g_n)$-verteilt, da dies auch die Wartezeit auf den ersten Sprung im zugehörigen Block-Zählprozess ist. Es gilt $\tau_n = T_n + \tau'$, wobei τ' die Absorptionszeit in $([n])$ des Prozesses $(\Pi^{(n)}_{T_n+t})_{t\geq 0}$ ist. Nach Temporal Coupling gilt $\tau' \stackrel{d}{=} \tau_{I_n}$, da der in T_n gestartete Block-Zählprozess $(|\Pi^{(n)}_{T_n+t}|)_{t\geq 0}$ genauso verteilt ist wie der Block-Zählprozess eines I_n-Coalescents. Des Weiteren zeigt die starke Markov-Eigenschaft, dass T_n sowohl von der Sprungkette von $\Pi^{(n)}$ als auch von allen weiteren Aufenthaltszeiten in $\Pi^{(n)}$ unabhängig ist. Also sind T_n, τ' unabhängig und a) ist gezeigt.
zu b): Analog zu a) erhält man $C_n = V_n + C'$, wobei C' die Anzahl der Kollisionen in $(\Pi^{(n)}_{T_n+t})_{t\geq 0}$ ist. Nach Temporal Coupling sind V_n, $(\Pi^{(n)}_{T_n+t})_{t\geq 0}$ und damit auch V_n, C' unabhängig bedingt $\Pi^{(n)}_{T_n}$, wobei $\Pi^{(n)}_{T_n} \stackrel{d}{=} I_n$ gilt. Es gilt $C' \stackrel{d}{=} C_{I_n}$ analog zu a) wiederum nach Temporal Coupling. Dies zeigt b). □

Bemerkung 2.4.2 *Für Λ-Coalescents gilt $V_n \equiv 1$ für $n \in \mathbb{N} \setminus \{1\}$, somit vereinfacht sich dort die Rekursion aus dem vorigen Satz 2.4.1 zu $C_1 = 0$ und $C_n = 1 + C_{I_n}$, wie sie etwa in [27, Gl. 29, S. 1415] zu finden ist. Die Verteilungsrekursion für $(\tau_n)_{n\in\mathbb{N}}$ findet sich auch in [62, S. 9].*

Für Λ-n-Coalescents betrachten wir nun die Länge E_n eines zufällig ausgewählten externen Zweiges und die Kollisionen C_n^{ext} bis zum Ende dieses ausgewählten Zweiges.

Satz 2.4.3 *[35] Sei $\Pi^{(n)}$ ein Λ-n-Coalescent für $n \in \mathbb{N}$. Sei $\chi^{(n)} := (\chi_k^{(n)})_{k\in\mathbb{N}_0}$ die Sprungkette des Block-Zählprozesses $(|\Pi_t^{(n)}|)_{t\geq 0}$ und sei $r_{nk} := P(\chi_1^{(n)} = k) = \frac{g_{nk}}{g_n}$ für $k \in [n]$, $n \in \mathbb{N}$ und die Raten g_{nk}, g_n des Block-Zählprozesses. Es gilt*

a) $C_1^{ext} = 0$ *und* $C_n^{ext} \overset{d}{=} \min\{i \in \{0,\dots,n-2\} | B_{n-i} = 0\}$,

b) $E_1 = 0$ *und* $E_n \overset{d}{=} \sum_{k=0}^{C_n^{ext}} T_{\chi_k^{(n)}}, \quad n \in \{2,3,\dots\}$,

c) $P(C_n^{ext} = 0) \;=\; 1 - \sum_{j=2}^{n-1} \frac{j-1}{n} r_{nj}$ *und*
$P(C_n^{ext} = k) \;=\; \sum_{j=2}^{n-1} \frac{j-1}{n} P(C_j^{ext} = k-1)\, r_{nj}, \qquad k \in [n-2],\ n \in \{2,3,\dots\}$.

wobei $T_2, \dots, T_n$ unabhängige Zufallsvariablen mit $T_l \overset{d}{=} Exp(g_l)$ für $l \in \mathbb{N}$ sind, die unabhängig von $(\chi_k^{(n)})_{k\in\mathbb{N}_0}$ und C_n^{ext} sind und $B_2, \dots, B_n$ Bernoulli-Zufallsvariablen, die bedingt auf $\chi^{(n)}$ unabhängig mit bedingten Erfolgswahrscheinlichkeiten

$$P(B_{n-i} = 1 \,|\, \chi_0^{(n)}, \chi_1^{(n)}, \dots) \;=\; \frac{\chi_{i+1}^{(n)} - 1}{\chi_i^{(n)}}, \quad i \in \{0, \dots, n-2\}.$$

sind.

Beweis: Nach Definition gilt $C_1^{ext} = 0$ und für $n \geq 2$ nimmt C_n^{ext} Werte in $\{0, \dots, n-2\}$ an. Betrachte $\Pi^{(n)}$. Es gilt $C_n^{ext} > i$ für $i \in \{0, \dots, n-3\}$ genau dann, wenn der zufällig ausgewählte externe Zweig nicht an den ersten i Kollisionen in $\Pi^{(n)}$ teilnimmt. Nach Bemerkung 1.1.8 ist $\Pi_t^{(n)}$ für jedes $t \geq 0$ eine austauschbare Partition von $[n]$, also $\sigma \circ \Pi^{(n)} \overset{d}{=} \Pi^{(n)}$, wobei $\sigma \in S_n$ in natürlicher Weise auf $\mathbb{E}_n$ wirkt. Somit lässt sich bedingt $\chi^{(n)}$ die Wahrscheinlichkeit von $\{C_n^{ext} > i\}$ in folgendem Urnenmodell berechnen:

1. Wähle aus einer Urne mit n weißen (unnummerierten) Kugeln (die Blöcke $\{i\}$, $i \in [n]$) eine Kugel aus und nummeriere diese mit 0 (das Individuum, dessen externen Zweig zufällig ausgewählt wurde).

2. Ziehe $n - \chi_1^{(n)} + 1$ Kugeln aus der Urne (die Anzahl der an der ersten Kollision beteiligten Blöcke) und lege eine weiße Kugel wieder zurück, die mit 1 nummeriert ist (der neu entstandene Block). Es befinden sich nun $\chi_1^{(n)}$ Kugeln in der Urne (die Anzahl der Blöcke nach dem ersten Sprung in $\Pi^{(n)}$).

3. Ziehe $\chi_1^{(n)} - \chi_2^{(n)} + 1$ Kugeln aus der Urne (die Anzahl der an der zweiten Kollision beteiligten Blöcke) und lege eine weiße Kugel wieder zurück, die mit 2 nummeriert ist (der neu entstandene Block). Es befinden sich nun $\chi_2^{(n)}$ Kugeln in der Urne (die Anzahl der Blöcke nach dem zweiten Sprung in $\Pi^{(n)}$).

$\vdots$

Die Wahrscheinlichkeit $P(C_n^{ext} > i|\chi^{(n)})$ entspricht der Wahrscheinlichkeit, dass sich nach $i+1$ Ziehungen aus der Urne die Kugel mit Nummer 0 noch in der Urne befindet. Sei B_{n-i} für $i \in \{0, \ldots, n-2\}$ die Bernoulli-Zufallsvariable, die genau dann 1 ist, wenn die mit i nummerierte Kugel nach dem $(i+1)$-ten Zug noch in der Urne ist. Diese sind nach Konstruktion bedingt $\chi^{(n)}$ unabhängig. Aus dem Urnenmodell liest man

$$P(B_{n-i} = 1 \,|\, \chi_0^{(n)}, \chi_1^{(n)}, \ldots) = \frac{\binom{\chi_i^{(n)} - 1}{\chi_i^{(n)} - \chi_{i+1}^{(n)} + 1}}{\binom{\chi_i^{(n)}}{\chi_i^{(n)} - \chi_{i+1}^{(n)} + 1}} = \frac{\chi_{i+1}^{(n)} - 1}{\chi_i^{(n)}}$$

für $i \in \{0, \ldots, n-2\}$ ab. Betrachte nun die Bernoulli-Variablen $\tilde{B}_{n-i}$ für $i \in \{0, \ldots, n-2\}$, die jeweils genau dann 1 sind, wenn die mit 0 nummerierte Kugel nach dem $(i+1)$-ten Zug noch in der Urne ist. Diese Zufallsvariablen sind zwar nicht mehr unabhängig, es gilt aber

$$\min \left\{ i \in \{0, \ldots, n-2\} \,|\, \tilde{B}_{n-i} = 0 \right\} \stackrel{d}{=} \min \left\{ i \in \{0, \ldots, n-2\} \,|\, B_{n-i} = 0 \right\}.$$

Man erhält nun aus dem Urnenmodell

$$\begin{aligned} C_n^{ext} &\overset{d}{=} \min\Big\{i \in \{0,\dots,n-2\} \,|\tilde{B}_{n-i} = 0\Big\} \\ &\overset{d}{=} \min\{i \in \{0,\dots,n-2\} : B_{n-i} = 0\}. \end{aligned}$$

Dies zeigt a), denn $C_1^{ext} = 0$ folgt nach Definition.
Betrachtet man E_n die Länge des zufällig ausgewählten externen Zweiges, so gilt nach Definition

$$E_n = \sum_{k=0}^{C_n^{ext}} \tilde{T}_k \text{ für } n \in \{2,3,\dots\}\,, \tag{38}$$

wobei bedingt $\chi^{(n)}$ für die Wartezeit $\tilde{T}_k$ vom k-ten bis zum $(k+1)$-ten Sprung des n-Coalescents $\tilde{T}_k \overset{d}{=} T_{\chi_k^{(n)}}$ gilt mit unabhängigen $T_i \overset{d}{=} Exp(g_i)$ für $i \in \{2,\dots,n\}$, die auch unabhängig von $\chi^{(n)}$ sind. Die Wartezeiten $(\tilde{T}_i)_{i\in\{0,\dots,C_n^{ext}\}}$ sind bedingt $\chi^{(n)}$ unabhängig. Dies zeigt b), $E_1 = 0$ gilt nach Definition.
Zeige nun die Verteilungs-Rekursion für $(C_n^{ext})_{n\in\mathbb{N}}$. Für $k \in \{0,\dots,n-2\}$ erhält man mit a)

$$\begin{aligned} P(C_n^{ext} = k) &= P(B_n = B_{n-1} = \dots = B_{n-k+1} = 1, B_{n-k} = 0) \\ &= \mathrm{E}\Big(\Big(1 - \frac{\chi_{k+1}^{(n)} - 1}{\chi_k^{(n)}}\Big) \prod_{i=0}^{k-1} \frac{\chi_{i+1}^{(n)} - 1}{\chi_i^{(n)}}\Big) \\ &= \mathrm{E}\Big(\prod_{i=0}^{k-1} \frac{\chi_{i+1}^{(n)} - 1}{\chi_i^{(n)}}\Big) - \mathrm{E}\Big(\prod_{i=0}^{k} \frac{\chi_{i+1}^{(n)} - 1}{\chi_i^{(n)}}\Big). \end{aligned}$$

Somit hat C_n^{ext} die Verteilungsfunktion

$$P(C_n^{ext} \le x) \;=\; \sum_{k=0}^{\lfloor x \rfloor} P(C_n^{ext} = k) \;=\; 1 - \mathrm{E}\Big(\prod_{i=0}^{\lfloor x \rfloor} \frac{J_{i+1}^{(n)} - 1}{J_i^{(n)}}\Big) \tag{39}$$

für $x \in [0, n-2]$. Aus (39) folgt mit Temporal Coupling und den Sprungketten $(\chi^{(j)})_{j\in\{2,\ldots,n\}}$ von Λ-j-Coalescents, $j \in \{2, \ldots, n\}$,

$$\begin{aligned} P(C_n^{ext} > k) &= \sum_{j=2}^{n-1} \mathrm{E}\Big(\frac{j-1}{n} \prod_{i=1}^{k} \frac{\chi_{i+1}^{(n)} - 1}{\chi_i^{(n)}} \,\Big|\, \chi_1^{(n)} = j \Big) \, P(\chi_1^{(n)} = j) \\ &= \sum_{j=2}^{n-1} \frac{j-1}{n} \mathrm{E}\Big(\prod_{i=0}^{k-1} \frac{\chi_{i+1}^{(j)} - 1}{\chi_i^{(j)}} \Big) \, P(\chi_1^{(n)} = j) \\ &= \sum_{j=2}^{n-1} \frac{j-1}{n} \, P(C_j^{ext} > k-1) \, r_{nj}, \quad k \in \{0, \ldots, n-2\}. \end{aligned}$$

Also gilt

$$P(C_n^{ext} = 0) \;=\; 1 - P(C_n^{ext} > 0) \;=\; 1 - \sum_{j=2}^{n-1} \frac{j-1}{n} \, r_{nj} \tag{40}$$

und

$$P(C_n^{ext} = k) \;=\; \sum_{j=2}^{n-1} \frac{j-1}{n} \, P(C_j^{ext} = k-1) \, r_{nj}, \qquad k \in \{1, \ldots, n-2\}. \tag{41}$$

Korollar 2.4.4 *[35] Für $n \in \mathbb{N}$ sei $\Pi^{(n)}$ ein Λ-n-Coalescent. Betrachte E_n, die Länge eines zufällig ausgewählten externen Zweiges des durch $\Pi^{(n)}$ gegebenen Baumes (die Auswahl erfolgt unabhängig von $\Pi^{(n)}$). Sei I_n eine Zufallsvariable für $n \in \mathbb{N}$, die verteilt ist wie der Block-Zählprozess von $\Pi^{(n)}$ direkt nach dem ersten Sprung, d.h. $P(I_n = k) = g_{nk}/g_n =: r_{nk}$ für $k \in [n-1]$ für die Raten g_n, g_{nk} des Block-Zählprozesses von $\Pi^{(n)}$. Dann gilt*

$$E_1 \;=\; 0 \quad \mathit{und} \quad E_n \;\overset{d}{=}\; T_n + B_n E_{I_n}, \quad n \in \{2, 3, \ldots\}, \tag{42}$$

wobei $T_n \overset{d}{=} Exp(g_n)$, T_n unabhängig von $B_n E_{I_n}$ ist, B_n bernoulliverteilt ist mit $P(B_n = 1|I_n) = (I_n - 1)/n$ und, bedingt auf I_n, die Zufallsvariablen B_n und E_{I_n} unabhängig sind.

Beweis: Es wird die Notation aus dem Beweis von Satz 2.4.3 benutzt. Aus Satz 2.4.3, dem Urnenmodell aus dem Beweis von Satz 2.4.3 sowie (38) erhält man für $n \in \{2, 3, \ldots\}$

$$E_n = \sum_{k=0}^{C_n^{ext}} \tilde{T}_k = T_n + B_n \sum_{k\in[C_n^{ext}]} \tilde{T}_k \tag{43}$$

mit $B_n = 1_{\{C_n^{ext} \neq 0\}}$ Bernoulli-verteilt mit bedingter Erfolgswahrscheinlichkeit $P(B_n = 1 | \chi_1^{(n)}) = \frac{\chi_1^{(n)}-1}{n}$ (hängt nur vom ersten Sprung ab) und der Wartezeit T_n auf den ersten Sprung in $\Pi^{(n)}$ mit $T_n \stackrel{d}{=} Exp(g_n)$. Wie in Satz 2.4.1 folgt mit der starken Markov-Eigenschaft, dass T_n unabhängig von $\chi^{(n)}$ und $\tilde{T}_1, \tilde{T}_2, \ldots$ ist. Dadurch ist T_n auch unabhängig von $B_n \sum_{k \in [C_n^{ext}]} \tilde{T}_k$.
Betrachte nun die Zufallsvariable $B_n \sum_{k \in [C^{ext}]} \tilde{T}_k$. Zeige, dass diese verteilt ist wie $B_n E_{\chi_1^{(n)}}$ mit B_n und $E_{\chi_1^{(n)}}$ unabhängig bedingt $\chi_1^{(n)}$. Zeige zunächst $B_n \sum_{k \in [C_n^{ext}]} \tilde{T}_k \stackrel{d}{=} B_n E_{\chi_1^{(n)}}$. Bedingt $B_n = 0$ ist dies offensichtlich. Es bleibt zu zeigen, dass bedingt $\{B_n = 1\}$ die Zufallsvariable $\sum_{k \in [C_n^{ext}]} \tilde{T}_k$ verteilt ist wie $E_{\chi_1^{(n)}}$. Betrachte dazu den Prozess $(\Pi_{T_n+t}^{(n)})_{t \geq 0}$, der nach Temporal Coupling ein $\chi_1^{(n)}$-Coalescent ist. Nun ist $\sum_{k \in [C_n^{ext}]} \tilde{T}_k$ bedingt $\{B_n = 1\}$ gerade der Anteil der Länge des ausgewählten externen Zweiges E_n, der im durch $(\Pi_{T_n+t}^{(n)})_{t \geq 0}$ gegebenen Teilbaum des durch $\Pi^{(n)}$ gegebenen Baumes liegt. Dieser Anteil ist bedingt $\{B_n = 1\}$ auch die Länge eines externen Zweiges von $(\Pi_{T_n+t}^{(n)})_{t \geq 0}$, da die Sprungkette dieses Prozesses gerade $(\chi_{k+1}^{(n)})_{i \in \mathbb{N}_0}$ ist und die Wartezeiten von Sprung zu Sprung $(\tilde{T}_{k+1})_{k \in \mathbb{N}_0}$ sind. Bedingt auf $\chi_1^{(n)}$ und $\{B_n = 1\}$ sieht man, dass dieser Anteil sogar die Länge eines zufällig und unabhängig von $(\Pi_{T_n+t}^{(n)})_{t \geq 0}$ ausgewählten externen Zweiges des $\chi_1^{(n)}$-Coalescents $(\Pi_{T_n+t}^{(n)})_{t \geq 0}$ ist; dies kann man dem Urnenmodell aus dem Beweis von Satz 2.4.3 entnehmen. Somit gilt nach Satz 2.4.3 bedingt auf $\chi_1^{(n)}$ und $\{B_n = 1\}$, dass $\sum_{k \in [C_n^{ext}]} \tilde{T}_k$ identisch verteilt ist wie $E_{\chi_1^{(n)}}$. Integriert man über beide Bedingungen $\{B_n = 1\}$ und $\{B_n = 0\}$, so folgt $B_n \sum_{k \in [C_n^{ext}]} \tilde{T}_k \stackrel{d}{=} B_n E_{\chi_1^{(n)}}$, falls B_n und $E_{\chi_1^{(n)}}$ unabhängig bedingt $\chi_n^{(1)}$ definiert sind. Durch Vergleichen der Verteilungen und (Un)Abhängigkeiten in (43) mit (42) folgt (42). $\square$

Bemerkung 2.4.5 *Für E_n im Kingman-n-Coalescent taucht Rekursion (42) schon in [37, S. 694, Gl. 6] und in [22, S. 246, Gl. 1] auf.*

Aus den Rekursionen aus Satz 2.4.1 und Korollar 2.4.4 lassen sich auch Rekursionen für andere Verteilungsgrößen wie z.B. erzeugende Funktionen oder Momente aufstellen. Diese sind für die spätere Analyse der Asymptotik von C_n^{ext}, E_n und τ_n von entscheidender Bedeutung.
In [27, S. 1406, Gl. 2] (oder auch [66, S. 753, Gl. 2.5]) wird für $n \in \{2, 3, \ldots\}$

die Rekursion

$$L_n \stackrel{d}{=} nT_n + L_{I_n}$$

für die Gesamt-Baumlänge L_n betrachtet, wobei T_n und I_n wie in Satz 2.4.1 definiert sind (für den Kingman-n-Coalescent findet sich die Rekursion bereits in [37, S. 695]). Aus dieser Rekursion wird in [27] eine Rekursion der Momente von $(L_n)_{n\in\mathbb{N}}$ hergeleitet. Analog lässt sich auch eine Rekursion für die Momente von τ_n aufstellen. Für $n \in \mathbb{N}$ und $j \in \mathbb{N}_0$ sei $\mu_n^{(j)} := \mathrm{E}(\tau_n^j)$ das j-te Moment von τ_n. Analog zu [27, S. 1407] erhält man aus der Rekursion aus Satz 2.4.1 a) die Startbedingung $\mu_1^{(j)} = 0$ und die Rekursion

$$\mu_n^{(j)} \;=\; \sum_{k\in[n-1]} r_{nk}\mu_k^{(j)} + r_n^{(j)}, \quad n \geq 2, j \in \mathbb{N}_0, \tag{44}$$

wobei

$$r_n^{(j)} \;:=\; \sum_{i\in[j]} \binom{j}{i} \mathrm{E}(T_n^i) \sum_{k\in[n-1]} r_{nk}\mu_k^{(j-i)}$$

und $r_{nk} = P(I_n = k)$ mit I_n verteilt wie in Satz 2.4.1 gilt. Diese gilt (wie auch die zu Grunde liegende Verteilungsrekursion für τ_n) sogar für die Zeit τ_n zurück zum Urahn in beliebigen Ξ-n-Coalescents.
Auch für die Rekursion aus Satz (42) für die zufällig ausgewählte externe Zweiglänge E_n in einem Λ-n-Coalescent ist es sinnvoll, eine Rekursion für die Momente von E_n aufzustellen. Für die Momente $\nu_n^{(j)} := \mathrm{E}(E_n^j)$ von E_n, $j \in \mathbb{N}_0$, in einem Λ-n-Coalescent gilt

$$\begin{aligned}
\nu_n^{(j)} \;&=\; \mathrm{E}(E_n^j) \;=\; \mathrm{E}((T_n + B_nE_{I_n})^j) \\
&=\; \mathrm{E}(T_n^j) + \sum_{i=0}^{j-1} \binom{j}{i} \mathrm{E}(T_n^i)\mathrm{E}((B_nE_{I_n})^{j-i}) \\
&=\; \mathrm{E}(T_n^j) + \sum_{i=0}^{j-1} \binom{j}{i} \mathrm{E}(T_n^i) \sum_{k\in[n-1]} P(I_n = k)\mathrm{E}((B_nE_k)^{j-i} \,|\, I_n = k) \\
&=\; \mathrm{E}(T_n^j) + \sum_{i=0}^{j-1} \binom{j}{i} \mathrm{E}(T_n^i) \sum_{k\in[n-1]} r_{nk}\frac{k-1}{n}\mathrm{E}(E_k^{j-i}),
\end{aligned}$$

da B_n bedingt auf $\{I_n = k\}$ Bernoulli-verteilt ist mit Erfolgswahrscheinlichkeit $(k-1)/n$ und von E_{I_n} unabhängig ist. Man kann ebenso für die j-ten

Momente $\nu_n^{(j)}$ von E_n in einem Λ-n-Coalescent die Rekursion

$$\nu_n^{(j)} \;=\; \sum_{k\in[n-1]} r_{nk}\frac{k-1}{n}\nu_k^{(j)} + \tilde{r}_n^{(j)}, \qquad n\geq 2, j\in\mathbb{N}_0, \tag{45}$$

aufstellen, wobei

$$\tilde{r}_n^{(j)} \;:=\; \mathrm{E}(T_n^j) + \sum_{i\in[j-1]}\binom{j}{i}\mathrm{E}(T_n^i)\sum_{k\in[n-1]} r_{nk}\frac{k-1}{n}\nu_k^{(j-i)}.$$

Man kann die Restterme $r_n^{(j)}$ und $\tilde{r}_n^{(j)}$ auch wieder in Termen der Momente von τ_n respektive E_n ausdrücken, analog zum Vorgehen in [27]. Indem man [27, Lemma 3.1] für $\alpha_n = g_n$ (α_n in dortiger Notation) anwendet, erhält man

$$r_n^{(j)} = \frac{j}{g_n}\mu_n^{(j-1)} \tag{46}$$

für $n\geq 2$ und $j\in\mathbb{N}$. Für die Restterme $\tilde{r}_n^{(j)}$, $n\geq 2$, gilt folgendes Analogon zu [27, Lemma 3.1].

Lemma 2.4.6 *Für den Restterm $\tilde{r}_n^{(j)}$ für $n\in\{2,3,\ldots\}$ und $j\in\mathbb{N}$ aus (45) gilt*

$$\tilde{r}_n^{(j)} = \frac{j}{g_n}\nu^{(j-1)} \tag{47}$$

Beweis: Vergleicht man Rekursion (44) mit Rekursion (45), so wird in letzterer nur an zwei Stellen der zusätzliche Faktor $\frac{k-1}{n}$ eingefügt. Zusammen mit $T_n \stackrel{d}{=} Exp(g_n)$ ermöglicht dies, den Beweis von [27, Lemma 3.1] einfach Schritt für Schritt analog für das hier zu beweisende Lemma durchzuführen. $\square$

Bemerkung 2.4.7 *Auf den ersten Blick erscheint es nicht sinnvoll, die Restterme $r_n^{(j)}$ und $\tilde{r}_n^{(j)}$ in (44) und (45) abzuspalten. Da die beiden Rekursionen aber bis auf die Restterme sehr einfach sind, wird es in späteren Analysen möglich sein, die Restterme zunächst zu ignorieren, die Rekursionen ohne die Restterme zu verwenden und aus den dortigen Ergebnissen analoge Ergebnisse für die ,,richtigen" Rekursionen zu finden.*

Schließlich ist es auch für das Funktional C_n^{ext} ratsam, eine weitere Rekursion aufzustellen, diesmal allerdings für die erzeugenden Funktionen $s \mapsto E(s^{C_n^{ext}})$ für $n \in \{2, 3, \ldots\}$ mit $s \in \mathbb{C}$. Für diese folgt mit der Verteilungsrekursion für C_n^{ext} in einem Λ-n-Coalescent aus 2.4.3 c) $\mathrm{E}(s^{C_2^{ext}}) = 1$ und

$$\begin{aligned}
\mathrm{E}(s^{C_n^{ext}}) &= P(C_n^{ext} = 0) + \sum_{k \in [n-2]} P(C_n^{ext} = k)\, s^k \\
&= P(C_n^{ext} = 0) + \sum_{k \in [n-2]} \sum_{j=k+1}^{n-1} \frac{j-1}{n} P(C_j^{ext} = k-1)\, r_{nj}\, s^k \\
&= P(C_n^{ext} = 0) + s \sum_{j=2}^{n-1} \frac{j-1}{n} r_{nj} \sum_{k \in [j-1]} P(C_j^{ext} = k-1)\, s^{k-1} \\
&= P(C_n^{ext} = 0) + s \sum_{j=2}^{n-1} \frac{j-1}{n} r_{nj}\, \mathrm{E}(s^{C_j^{ext}}) \qquad (48)
\end{aligned}$$

für $s \in \mathbb{C}$ und $n \in \{3, 4, \ldots\}$, wobei $P(C_n = 0)$ aus (40) berechnet wird.
Es lassen sich noch weitere Rekursionen für andere Kenngrößen der Verteilungen der betrachteten Funktionale aufstellen. Da solche aber in dieser Arbeit nicht benötigt werden und analog hergeleitet werden können, wird hier auf weitere Rekursionen verzichtet. Es findet sich etwa in [35, Bemerkung 4, S.5] (hierzu auch [35, Anhang, S.24-26]) eine Rekursion für die charakteristische Funktion von τ_n.
Betrachtet man nun n-Coalescents mit Mutationen, so besitzt auch der die Mutationen steuernde Poisson-Punktprozess durch die Unabhängigkeit der Punkte auf verschiedenen Zweigen eine zur Coupling-Struktur des Coalescents passende Struktur. Somit kann man ebenso Rekursionen für Funktionale von n-Coalescents mit Mutation finden. In dieser Arbeit wird die Anzahl K_n der Typen in der Stichprobe $[n]$ untersucht, die folgende Verteilungsrekursion erfüllt.

Satz 2.4.8 *[36] Für $n \in \mathbb{N}$ sei $\Pi^{(n)}$ ein Ξ-n-Coalescent mit Mutation. Sei $r > 0$ die Mutationsrate und g_n, g_{nk} seien die Raten des zu $\Pi^{(n)}$ gehörenden Block-Zählprozesses aus Satz 1.4.1 ($k \in [n-1]$). Dann gilt für die Anzahl K_n der Typen in der Stichprobe $[n]$ die Verteilungsrekursion*

$$K_1 = 1 \text{ und } K_n \stackrel{d}{=} B_n(K_{n-1} + 1) + (1 - B_n) K_{I_n}, \quad n \in \{2, 3, \ldots\}, \quad (49)$$

wobei B_n eine Bernoulli-verteilte Zufallsvariable ist, die unabhängig von $(K_2,\ldots,K_{n-1},I_n)$ ist mit Verteilung

$$P(B_n=1) \;=\; 1-P(B_n=0) \;=\; \frac{nr}{g_n+nr}, \quad n\in\mathbb{N},$$

und I_n eine von $(K_2,\ldots,K_{n-1})$ unabhängige Zufallsvariable ist mit Verteilung

$$r_{nk} \;:=\; P(I_n=k) \;=\; \frac{g_{nk}}{g_n}, \qquad n,k\in\mathbb{N}, k\in[n-1],$$

d.h. I_n ist verteilt wie $(|\Pi_t^{(n)}|)_{t\geq 0}$ nach dem ersten Sprung.

Beweis: Rekursion (49) ist äquivalent zu $P(K_1=1)=1$ und

$$P(K_n=k) \;=\; \frac{nr}{g_n+nr}P(K_{n-1}=k-1)+\frac{g_n}{g_n+nr}\sum_{i=k}^{n-1}r_{ni}P(K_i=k) \quad (50)$$

für $n\in\{1,2,\ldots\}$ und $k\in[n]$ (die zweite Summe läuft erst ab k, da $P(K_i=k)=0$ für $i\in[k-1]$). Zeige nun (50).

Im n-Coalescent mit Mutation kommen zwei Arten von zustandsändernden Ereignissen vor: Mutationen und Kollisionen. Für $n\in\mathbb{N}$ sei oBdA $\Pi^{(n)}:=(\rho_n(\Pi_t))_{t\geq 0}$ für einen Ξ-Coalescent Π, dessen Mutationsstruktur definiert ist über eine Familie $(P_i)_{i\in\mathbb{N}}$ von homogenen Poisson-Punktprozessen auf $[0,\infty)$ mit Intensistätsrate r wie in Kapitel 1.7. Betrachte nun das zeitlich erste Ereignis, dass im Verlauf des n-Coalescent mit Mutation eintritt. Die Wartezeit $W_n^{(i)}$ auf die erste Mutation, die auf den externen Zweig $i\in[n]$ fallen kann, also auf den ersten Punkt in P_i, ist $Exp(r)$ verteilt. Somit ist die Wartezeit W_n auf die erste Mutation, die auf irgendeinen externen Zweig fallen kann, das Minimum dieser n unabhängigen Wartezeiten, also $W_n\stackrel{d}{=}Exp(nr)$. Die Wartezeit auf die erste Kollision T_n (es kann mehrere simultane Kollisionen geben) ist unabhängig von W_n und es gilt $T_n\stackrel{d}{=}Exp(g_n)$, da dies gerade die Wartezeit auf den ersten Sprung in $(|\Pi_t^{(n)}|)_{t\geq 0}$ ist. Somit ist das erste Ereignis im Verlauf des Prozesses mit Wahrscheinlichkeit $P(W_n<T_n)=nr/(g_n+nr)$ eine Mutation und mit Wahrscheinlichkeit $P(W_n>T_n)=g_n/(g_n+nr)$ eine Kollision. Es gilt also

$$P(K_n=k)=(P(K_n=k|W_n<T_n)nr+P(K_n=k|T_n>W_n)g_n)/(g_n+nr),$$

da $\{T_n = W_n\}$ eine Nullmenge ist.
Berechne zunächst $P(K_n = k|W_n < T_n)$. Mit Hilfe der Zufallsvariablen $(W_n^{(i)})_{i\in[n]}$ lässt sich dies auch als

$$\begin{aligned} & P(K_n = k|W_n < T_n) \\ = & \sum_{i\in[n]} P(K_n = k|W_n^{(i)} < T_n)P(W_n^{(i)} < T_n|W_n < T_n) \end{aligned} \tag{51}$$

schreiben. Für $i \in [n]$ betrachte das Ereignis $\{K_n = k\}$ unter der Bedingung $\left\{W_n^{(i)} < T_n\right\}$. Dann tritt die erste Mutation auf dem zu $i \in [n]$ gehörenden externen Zweig vor der ersten Kollision ein. Das Individuum i in der Stichprobe $[n]$ hat also einen von allen anderen Individuen in $[n]$ verschiedenen Typ. Um dann $K_n = k$ zu erhalten, muss die Anzahl der verschiedenen Typen unter den restlichen $n-1$ Individuen genau $k-1$ betragen. Streicht man nun den i-ten externen Zweig aus dem von $\Pi^{(n)}$ gegebenen Baum, so verhält sich nach Natural Coupling der restliche Baum wie ein $(n-1)$-Coalescent, somit $P(K_n = k|W_n^{(i)} < T_n) = P(K_{n-1} = k-1)$. Da dies für jedes $i \in [n]$ gilt, folgt aus (51) $P(K_n = k|W_n < T_n) = P(K_{n-1} = k-1)$.
Betrachte nun das Ereignis $\{K_n = k\}$ unter der Bedingung $\{W_n > T_n\}$. Das erste Ereignis ist nun eine Kollision. Da hier keine Mutation vor dem Kollisionsereignis stattfindet, muss für das Eintreten von $\{K_n = k\}$ die Anzahl der Typen unter den Blöcken von $\Pi_{T_n}^{(n)}$ genau k betragen, also die Anzahl der verschiedenen Typen in der Stichprobe $[n]$ (der Typ eines Blockes wird genauso wie der Typ eines Individuums bestimmt, es wird aber nur der von $\Pi^{(n)}$ erzeugte Baum ab Zeitpunkt T_n betrachtet). $(\Pi_{T_n+t}^{(n)})_{t\geq 0}$ ist nach Temporal Coupling ein $\chi_1^{(n)}$-Coalescent (wenn man die Blöcke von $\Pi^{(n)}$ zum Zeitpunkt T_n mit $[\chi_1^{(n)}]$ identifiziert), wobei $\chi_1^{(n)}$ die Anzahl der Blöcke nach dem ersten Sprung von $\Pi^{(n)}$ ist. In diesem Sinne ist die Anzahl der Typen unter den Blöcken von $\Pi_{T_n}^{(n)}$ gerade die Anzahl der Typen in $[\chi_1^{(n)}]$ von $(\Pi_{T_n+t}^{(n)})_{t\geq 0}$. Es gilt also

$$P(K_n = k|T_n < W_n, \chi_1^{(n)} = i) = P(K_i = k).$$

$\chi_1^{(n)}$ ist unabhängig von T_n und W_n (starke Markov-Eigenschaft und Unabhängigkeit von W_n und $\Pi^{(n)}$). Also gilt

$$\begin{aligned}
& P(K_n = k | T_n < W_n) \\
= & \sum_{i=k}^{n-1} P(K_i = k | T_n < W_n,\ \chi_1^{(n)} = i) P(\chi_1^{(n)} = i | T_n < W_n) \\
= & \sum_{i=k}^{n-1} P(K_i = k) P(\chi_1^{(n)} = i) = \sum_{i=k}^{n-1} P(K_i = k) r_{ni},
\end{aligned}$$

wobei es genügt, nur von k bis $n-1$ zu summieren, da wiederum $P(K_i = k) = 0$ für $i \in [k-1]$ gilt.

Bemerkung 2.4.9 *[36] Die Beweisführung des vorangegangenen Satzes verläuft analog zum Beweis der Verteilungsrekursion für das vollständige Allelfrequenz-Spektrum für n-Coalescents aus [67, Theoreme 3.1, 5.1] (Möhle's recursion). Eine zu dieser Rekursion für das vollständige Allelfrequenz-Spektrum äquivalente Rekursion für die zugehörigen erzeugenden Funktionen findet sich in [65, Gl. 4], dort wird die Rekursion allerdings nur für Λ-n-Coalescents betrachtet. Betrachte die erzeugenden Funktionen* $f_n(s) = E(s^{K_n})$, *$n \in \mathbb{N}$ und $s \in \mathbb{C}$, mit $f_1(s) = s$. Aus Rekursion (50) erhält man für $n \in \{2, 3, \ldots\}$*

$$\begin{aligned}
& (g_n + nr) f_n(s) = (g_n + nr) \sum_{k \in [n]} P(K_n = k) s^k \\
= & \sum_{k \in [n]} nr P(K_{n-1} = k-1) s^k + g_n \sum_{k \in [n]} \sum_{i=k}^{n-1} P(K_i = k) r_{ni} s^k \\
= & \; nrs f_{n-1}(s) + g_n \sum_{i \in [n-1]} r_{ni} \sum_{k \in [i]} P(K_i = k) s^k \\
= & \; nrs f_{n-1}(s) + \sum_{i \in [n-1]} g_{ni} f_i(s),
\end{aligned}$$

also die zu (50) äquivalente Rekursion

$$(g_n + nr) f_n(s) \;=\; nrs f_{n-1}(s) + \sum_{k \in [n-1]} g_{nk} f_k(s), \quad n \in \{2, 3, \ldots\}, s \in \mathbb{C}. \quad (52)$$

(52) gilt für beliebige Ξ-n-Coalescents. Für Λ-n-Coalescents lässt sich (52) auch leicht aus Möhle's recursion für die erzeugenden Funktionen des vollständigen Allel-Frequenz-Spektrums aus [65, Gleichung 4] gewinnen, indem man dort $s_1 = \cdots = s_n = s$ setzt.

Mit Hilfe der Rekursion (50) lassen sich die Wahrscheinlichkeiten $P(K_n = k)$ für $k \in [n]$ rekursiv berechnen. Es gilt z.B. für die Wahrscheinlichkeit $P(K_n = n)$, dass alle Individuen in $[n]$ verschiedene Typen haben, die Rekursion

$$(g_n + nr)P(K_n = n) = nrP(K_{n-1} = n-1), \quad n \in \{2, 3, \ldots\}$$

und damit

$$P(K_n = n) \;=\; \prod_{i=2}^{n} \frac{ir}{g_i + ir} \;=\; \frac{r^{n-1} n!}{\prod_{i=2}^{n}(g_i + ir)}, \quad n \in \mathbb{N}. \tag{53}$$

Auch aus der Rekursion (52) für die erzeugenden Funktionen $f_n(s) = E(s^{K_n})$ für $n \in \{2, 3, \ldots\}$ und $f_1(s) = s$ lässt sich eine weitere Rekursion bilden. Bildet man in (52) die j-te Ableitung $f_n^{(j)}$ nach s und wendet man die Leibniz-Regel an, erhält man die Rekursion

$$(g_n + nr)f_n^{(j)}(s) \;=\; nr\big(sf_{n-1}^{(j)}(s) + jf_{n-1}^{(j-1)}(s)\big) + \sum_{k \in [n-1]} g_{nk} f_k^{(j)}(s) \tag{54}$$

für $n \in \{2, 3, \ldots\}$, $j \in \mathbb{N}$ und $s \in \mathbb{C}$ (für $j > n$ ist die Rekursion trivial, nämlich $0 = 0$). Nun ist $f_n^{(j)}(s) = \sum_{k=j}^{n} (k)_j P(K_n = k) s^{k-j}$, wobei $(k)_j = k(k-1)\cdots(k-j+1)$ das absteigende faktorielle Produkt von k bezüglich $j \in \mathbb{N}$ bezeichnet. Es gilt $\lim_{s \to 1} f_n^{(j)}(s) = E((K_n)_j)$ und somit folgt aus (54) für $s \to 1$ die Rekursion

$$(g_n + nr)E((K_n)_j) = nr\big(E((K_{n-1})_j) + jE((K_{n-1})_{j-1})\big) + \sum_{k \in [n-1]} g_{nk} E((K_k)_j) \tag{55}$$

für $n \in \{2, 3, \ldots\}$ und $j \in \mathbb{N}$, wobei $E((K_1)_j) = \delta_{j1}$ (Kronecker-Symbol) ist. Insbesondere erhält man für $j = 1$, also für $E(K_n)$, die Rekursion

$$(g_n + nr)E(K_n) \;=\; nr(E(K_{n-1}) + 1) + \sum_{k \in [n-1]} g_{nk} E(K_k) \tag{56}$$

für $n \in \{2, 3, \ldots\}$.
Auch Rekursion (55) lässt sich für bestimmte Werte leicht lösen. Für $j = n$

erhält man $(g_n+nr)E((K_n)_n) = n^2 r E((K_{n-1})_{n-1})$, woraus man die geschlossene Formel

$$E((K_n)_n) \;=\; \prod_{i=2}^{n} \frac{i^2 r}{g_i + ir} \;=\; \frac{r^{n-1}(n!)^2}{\prod_{i=2}^{n}(g_i+ir)}, \quad n \in \mathbb{N},$$

folgern kann. Diese folgt allerdings auch direkt aus $E((K_n)_n) = n!P(K_n = n)$. Aus allen in diesem Kapitel vorgestellten Rekursionen lassen sich für kleine bis moderate $n \in \mathbb{N}$ die Verteilungen/Momente der zugehörigen Funktionale direkt und explizit berechnen. Für manche Ξ-Coalescents (bzw. Λ-Coalescents) lässt sich aus den vorgestellten Rekursionen die Asymptotik der betreffenden Funktionale für $n \to \infty$ bestimmen. Dies wird in den folgenden Kapiteln beschrieben.

Bemerkung 2.4.10 *Auch für andere Funktionale eines n-Coalescents $\Pi^{(n)}$ lassen sich mit den in diesem Kapitel vorgestellten Methoden Rekursionen aufstellen. Die Rekursionen für das vollständige Allel-Frequenzspektrum $(K_n(1), \ldots, K_n(n))$ und für die Gesamt-Baumlänge L_n wurden bereits erwähnt, für die Anzahl Seg_n der Mutationen auf dem durch $\Pi^{(n)}$ gegebenen Baum findet sich eine Verteilungsrekursion in [66, Gleichung 2.3]. Die Methoden, mit denen alle vorgestellten Rekursionen aufgestellt werden, wurden bereits in [43, Kapitel 3] beschrieben.*

3 τ_n, die Zeit zurück zum jüngsten Urahn [35]

In diesem Abschnitt wird τ_n, die Zeit zurück zum jüngsten Urahn der Stichprobe $[n]$ bzw. die Absorptionszeit in $([n])$ für einen Ξ-n-Coalescent betrachtet ($n \in \mathbb{N}$). Für dieses Funktional sind viele Ergebnisse bereits bekannt. Für τ_n im Kingman-n-Coalescent hat schon Kingman selbst in [57, Gl. (5.9)] bewiesen, dass $\tau_n \to \tau$ in Verteilung konvergiert für $n \to \infty$, wobei τ die Lebesgue-Dichte

$$t \;\mapsto\; \sum_{m=2}^{\infty}(-1)^m g_m(2m-1)e^{-g_m t}, \quad t \geq 0,$$

besitzt mit $g_m = \binom{m}{2}$ die totale Rate im Zustand m des Blockzählprozesses des Kingman-m-Coalescent. τ hat die Laplace-Transformierte

$$\mathrm{E}(e^{-\theta\tau}) \;=\; 1 - \sum_{m=2}^{\infty}(-1)^m(2m-1)\frac{\theta}{g_m+\theta}, \quad \theta \geq 0,$$

dies korrigiert einen Fehler auf [57, S.37] (die Reihe in [57, Zeile , S.37] konvergiert nicht). Dieses Ergebnis lässt sich auch leicht aus der Rekursion für τ_n aus Satz 2.4.1 a) folgern. Mit der Notation aus Satz 2.4.1 gilt im Kingman-n-Coalescent $I_n \equiv n-1$ für $n \in \mathbb{N}$, da jede Kollision binär ist und es keine simultanen Kollisionen gibt. Somit erhält man aus der Rekursion die Darstellung $\tau_n = \sum_{i=2}^n T_n$ für $n \in \mathbb{N}$, wobei $T_2, \ldots, T_n$ unabhängige Zufallsvariablen mit $T_i \stackrel{d}{=} Exp(d_i)$ sind, da nach Formel (18) $g_i = \binom{i}{2}$ die totale Rate im Zustand i des Block-Zählprozesses des Kingman-n-Coalescents ist. Das Konvergenzresultat kann nun mit Standardmethoden gezeigt werden.

Im sternförmigen n-Coalescent ($\Lambda = \delta_1$) gilt $\tau_n \stackrel{d}{=} Exp(1)$ für alle $n \in \mathbb{N}$, da $g_n = \Lambda([0,1]) = 1$ nach (18) gilt. Insbesondere gilt $\tau_n \stackrel{d}{\to} Exp(1)$ für $n \to \infty$. Für andere Coalescent-Prozesse scheint die Analyse der Asymptotik von τ_n schwieriger zu sein, da sich die Rekursion für $(\tau_n)_{n\in\mathbb{N}}$ nicht soweit vereinfachen lässt. Man weiß allerdings für $\Pi^{(n)} = (\rho_n(\Pi_t))_{t\geq 0}$ für einen Ξ-Coalescent Π, dass τ_n monoton in n steigt und dass $\tau_n \to \tau := \inf\{t > 0 : |\Pi_t| = 1\} \in [0,\infty]$ fast sicher für $n \to \infty$ (etwa [79, S. 40] für Ξ-Coalescents oder [71, Gl. 31] für Λ-Coalescents). Es gilt $\mathrm{E}(\tau) < \infty$, falls der Coalescent nicht im Unendlichen bleibt (comes down from infinity), siehe etwa den Beweis von [79, Proposition 32].

Für verschiedene Unterklassen der Simple Λ-n-Coalescents ist τ_n nach

[39] asymptotisch normal, asymptotisch stabil oder asymptotisch (skaliert) Mittag-Leffler verteilt (nach geeigneter Normierung). Ein Klasse von Λ-n-Coalescents, in der τ_n asymptotisch normal ist, sind die Beta-n-Coalescents mit $a > 2, b > 0$ (siehe [39, Beispiel 4.1]). Ein schwaches Gesetz der großen Zahlen für $(\tau_n)_{n\in\mathbb{N}}$ in Poisson-Dirichlet-n-Coalescents findet sich in [62, Theorem 4.1]. Im folgenden Kapitel wird die Asymptotik für $n \to \infty$ von τ_n für den Bolthausen-Sznitman-n-Coalescent untersucht ($\Lambda = \lambda_{[0,1]}$). Aus [42, Proposition 3.4] ist bekannt, dass im Bolthausen-Sznitman-Coalescent

$$\tau_n - \log\log n \xrightarrow{d} \tau \text{ für } n \to \infty \tag{57}$$

gilt, wobei τ (Standard-)Gumbel-verteilt ist. Dieses Resultat wurde in [42] mit Hilfe des dort analysierten Zusammenhangs zwischen dem Bolthausen-Sznitman-Coalescent und zufälligen rekursiven Bäumen bewiesen. Im nächsten Kapitel wird dieses Resultat auf andere Weise, nämlich unter Verwendung der Rekursion (44) für die Momente von $(\tau_n)_{n\in\mathbb{N}}$ bewiesen.

3.1 τ_n im Bolthausen-Sznitman-n-Coalescent [35]

Um die Rekursion (44) für τ_n für den Bolthausen-Sznitman-n-Coalescent verwenden zu können, müssen wir zunächst die totale Rate g_n im Zustand n des Block-Zählprozesses und die Verteilung des ersten Sprungs des Block-Zählprozesses, also die Verteilung von $\chi_1^{(n)}$ im Bolthausen-Sznitman-n-Coalescent bestimmen. Nach (18) gilt für die Raten des Block-Zählprozesses des Bolthausen-Sznitman-Coalescents

$$\begin{aligned} g_{nk} &= \binom{n}{k-1} \int_{[0,1]} u^{n-k-1}(1-u)^{k-1} du \\ &= \binom{n}{k-1} Beta(n-k,k) = \frac{n}{(n-k)(n-k+1)} \end{aligned}$$

für $n \in \mathbb{N},\ k \in [n]$, somit

$$g_n = \sum_{k\in[n-1]} g_{nk} = \sum_{k\in[n-1]} \frac{n}{(n-k)(n-k+1)} = n \sum_{k\in[n-1]} \frac{1}{k(k+1)} = n-1 \tag{58}$$

und damit (siehe Korollar 1.4.3)

$$r_{nk} = P(\chi_1^{(n)} = k) = \frac{g_{nk}}{g_n} = \frac{n}{(n-1)(n-k)(n-k+1)} \tag{59}$$

für $n \in \mathbb{N}$, $k \in [n]$. Betrachte für die j-ten Momente $(\mu_n^{(j)})_{n\in\mathbb{N}}$ von $(\tau_n)_{n\in\mathbb{N}}$ und die Restterme $r_n^{(j)}$ aus (44), $j \in \mathbb{N}_0$, die erzeugende Funktionen

$$\mu_j(s) \ := \ \sum_{n=2}^{\infty} \mu_n^{(j)} s^n \quad \text{und} \quad r_j(s) \ := \ \sum_{n=2}^{\infty} g_n r_n^{(j)} s^n.$$

Die beiden Funktionen μ_j and r_j sind analytisch auf

$$D := \{s \in \mathbb{C} \, : \, |s| < 1\},$$

dies wird in diesem Kapitel bewiesen. Es wird sogar gezeigt (vergleiche etwa die folgenden Gleichungen (64) und (69)), dass μ_j und r_j analytisch fortsetzbar auf $\mathbb{C} \setminus [1, \infty)$ sind, dies wird allerdings für die Zwecke dieser Arbeit nicht benötigt.

Bemerkung 3.1.1 *Im Folgenden werden gewöhnliche Differentialgleichungen mit komplexen Argumenten betrachtet. Alle vorkommenden Funktionen sind aber holomorph auf dem jeweils betrachteten Gebiet, somit kann man diese Differentialgleichungen wie reelle Differentialgleichungen behandeln bzw. sie als Differentialgleichungen von Potenzreihen auffassen.*
Es werden auch komplexe (Weg-)Integrale von holomorphen Funktionen verwendet. Der Einfachheit halber werden nur die Grenzen des Weges, längs dessen integriert wird, angegeben (Der Weg muss natürlich vollständig in einem Holomorphiegebiet des Integranden liegen).

Lemma 3.1.2 *(Rekursion für die erzeugende Funktion μ_j) [35] Die erzeugenden Funktionen $\mu_0, \mu_1, \ldots$ genügen der Rekursion*

$$\mu_j(s) \ = \ \frac{js}{1-s} \int_0^s \frac{\mu_{j-1}(t)}{t(-\log(1-t))} \, dt, \qquad j \in \mathbb{N}, s \in D, \tag{60}$$

mit Startbedingung $\mu_0(s) = s^2/(1-s)$, $s \in \mathbb{C} \setminus \{1\}$. Insbesondere gilt

$$\mu_1(s) \ = \ \frac{s}{1-s} \int_0^s \frac{t}{(1-t)(-\log(1-t))} \, dt, \qquad s \in D.$$

Beweis: Der Beweis ist nahezu identisch zum Beweis von [27, Lemma 4.1] und wird hier mit geringfügigen Modifikationen ausgeführt.
Sei $j \in \mathbb{N}$ und $s \in D$. Definiere zunächst die Hilfsfunktion

$$a(s) := \sum_{k=1}^{\infty} \frac{s^k}{k(k+1)} = 1 + \frac{\log(1-s)}{s} - \log(1-s) \tag{61}$$

für $s \in D$. Multipliziert man Rekursion (44) mit $\frac{n-1}{n}$ und setzt (59) ein, erhält man

$$\begin{aligned}
\frac{n-1}{n}\mu_n^{(j)} &= \sum_{k\in[n-1]} r_{nk}\frac{n-1}{n}\mu_k^{(j)} + \frac{n-1}{n}r_n^{(j)} \\
= \quad \frac{n-1}{n}r_n^{(j)} + \sum_{k\in[n-1]} \frac{\mu_k^{(j)}}{(n-k)(n-k+1)} &= \frac{n-1}{n}r_n^{(j)} + \sum_{k\in[n-1]} \frac{\mu_{n-k}^{(j)}}{k(k+1)}.
\end{aligned}$$

Multipliziert man diese Rekursion mit s^n und summiert über $n \in \{2, 3, \ldots\}$, so erhält man

$$\begin{aligned}
\mu_j(s) - \int_0^s \frac{\mu_j(t)}{t}dt &= \sum_{n=2}^{\infty} \frac{n-1}{n}\mu^{(j)}s^n \\
&= \sum_{n=2}^{\infty} \frac{n-1}{n}r_n^{(j)}s^n + \sum_{n=2}^{\infty} s^n \sum_{k=1}^{n-1} \frac{\mu_{n-k}^{(j)}}{k(k+1)} \\
&= \int_0^s \frac{r_j(t)}{t}dt + \sum_{k=1}^{\infty} \frac{s^k}{k(k+1)} \sum_{n=k+1}^{\infty} \mu_{n-k}^{(j)}s^{n-k} \\
&= \int_0^s \frac{r_j(t)}{t}dt + a(s)\mu_j(s), \qquad (62)
\end{aligned}$$

wobei man den ersten Summanden in der letzten Zeile aus

$$\sum_{n=2}^{\infty} \frac{n-1}{n}r_n^{(j)}s^n \;=\; \sum_{n=2}^{\infty} g_n r_n^{(j)}\frac{s^n}{n} \;=\; \int_0^s \sum_{n=2}^{\infty} g_n r_n^{(j)}t^{n-1}\,dt \;=\; \int_0^s \frac{r_j(t)}{t}\,dt$$

erhält. Differenziert man nun Gleichung (62) nach s erhält man die Differentialgleichung

$$\mu_j'(s) - \frac{\mu_j(s)}{s} = \frac{r_j(s)}{s} + a'(s)\mu_j(s) + a(s)\mu_j'(s)$$

oder die äquivalente Darstellung

$$\mu_j'(s) = \frac{(1 + s\cdot a'(s))\mu_j(s) + r_j(s)}{s(1-a(s))}.$$

Setzt man in diese Gleichung $a(s)$ und $a'(s) = -s^{-1} - s^{-2}\log(1-s)$ ein, so erhält man

$$\mu_j'(s) \;=\; \frac{\mu_j(s)}{s(1-s)} - \frac{r_j(s)}{(1-s)\log(1-s)}. \tag{63}$$

Alle Lösungen der homogenen Differentialgleichung $f'(s) = f(s)/(s(1-s))$ haben die Form $f(s) = cs/(1-s)$ mit $c \in \mathbb{C}$. Somit erhält man durch Variation der Konstanten für die inhomogene Differentialgleichung (63) mit Startwert $\mu_j(0) = 0$ die Lösung

$$\mu_j(s) \;=\; c_j(s)\frac{s}{1-s}, \tag{64}$$

wobei

$$c_j(s) \;:=\; \int_0^s \frac{r_j(t)}{t(-\log(1-t))}\,dt. \tag{65}$$

(60) folgt nun mit $r_j(t) = j\mu_{j-1}(t)$, dies ist eine Anwendung von [27, Lemma 3.1, Gl. (7)] mit $\alpha_n := g_n = n-1$ (bzw. (46)). Schließlich gilt $\mu_0(s) = \sum_{n=2}^{\infty} s^n = s^2/(1-s)$. □

Um die erzeugende Funktion μ_j, $j \in \mathbb{N}$, genauer zu analysieren ist es wegen (64) ratsam, zunächst die Funktion c_j zu analysieren. Aus Lemma 3.1.2 und (64) erhält man die Rekursion

$$c_j(s) \;=\; j\int_0^s \frac{c_{j-1}(t)}{(1-t)(-\log(1-t))}\,dt, \qquad j \in \mathbb{N}, s \in D, \tag{66}$$

für die Funktionen $c_1, c_2, \ldots$ mit Anfangsbedingung $c_0(s) = s$, $s \in \mathbb{C}$. Substituiert man $t = 1 - e^{-u}$ in (66) (dies ist entlang jedes Weges in D möglich), so erhält man eine weitere Rekursion, nämlich

$$c_j(s) \;=\; j\int_0^{-\log(1-s)} \frac{c_{j-1}(1-e^{-u})}{u}\,du, \quad j \in \mathbb{N}, s \in D. \tag{67}$$

Für die Funktion c_1 erhält man insbesondere

$$c_1(s) \;=\; \int_0^{-\log(1-s)} \frac{1-e^{-u}}{u}\,du, \qquad s \in D. \tag{68}$$

Daraus lässt sich eine Reihendarstellung für c_j herleiten.

Lemma 3.1.3 *[35] Die Funktion c_j hat die Reihendarstellung*

$$c_j(s) \;=\; -j!\sum_{k=1}^{\infty} \frac{(\log(1-s))^k}{k!k^j}, \quad j \in \mathbb{N}_0, s \in D. \tag{69}$$

Beweis: Dies wird mittels Induktion über $j \in \mathbb{N}_0$ gezeigt. Da $c_0(s) = s$, ist (69) für $j = 0$ erfüllt. Sei nun (69) für $j-1$ mit $j \in \mathbb{N}$ erfüllt. Mit (67) und der Induktionsvoraussetzung folgt dann

$$\begin{aligned}
c_j(s) &= j\int_0^{-\log(1-s)} \frac{c_{j-1}(1-e^{-u})}{u}\,du \\
&= j\int_0^{-\log(1-s)} -(j-1)!\sum_{k=1}^{\infty}\frac{(-u)^k}{k!k^{j-1}}\frac{1}{u}\,du \\
&\overset{(*)}{=} -j!\sum_{k=1}^{\infty}\frac{(-1)^k}{k!k^{j-1}}\int_0^{-\log(1-s)} u^{k-1}\,du \\
&= -j!\sum_{k=1}^{\infty}\frac{(-1)^k}{k!k^{j-1}}\frac{(-\log(1-s))^k}{k} \;=\; -j!\sum_{k=1}^{\infty}\frac{(\log(1-s))^k}{k!k^j},
\end{aligned}$$

wobei für $(*)$ die vorkommende Potenzreihe gliedweise integriert wird. □

Bemerkung 3.1.4 *Die Reihendarstellung (69) zeigt insbesondere, dass c_j für jedes $j \in \mathbb{N}$ in D holomorph ist und auf $\mathbb{C} \setminus [1,\infty)$ holomorph fortsetzbar ist. Somit sind auch μ_j und r_j in D holomorph und auf $\mathbb{C} \setminus [1,\infty)$ analytisch fortsetzbar. Die Rekursion ist dazu geeignet, $c_j(s)$ mit moderatem Zeitaufwand numerisch zu approximieren.*

Mit Hilfe der eben bewiesenen Reihendarstellung ist es möglich, eine explizite Formel für die Momente von τ_j zu finden.

Satz 3.1.5 *[35] Sei $n \in \mathbb{N}$ und $j \in \mathbb{N}_0$. Dann gilt für das j-te Moment von τ_n im Bolthausen-Sznitman-n-Coalescent*

$$\mathrm{E}(\tau_n^j) \;=\; j!\sum_{i\in[n-1]}\frac{1}{i!}\sum_{k\in[i]}\frac{(-1)^{k+1}}{k^j}s(i,k),$$

wobei $s(i,k)$ die absolute Stirling-Zahl erster Art mit Parametern i,k ist, also die Anzahl der Permutationen von $[i]$, die genau k Zykel haben ($i,k \in \mathbb{N},\ k \in [i]$).

Beweis: Aus (64) und (69) erhält man

$$\mu_j(s) \;=\; \frac{j!s}{s-1}\sum_{k=1}^{\infty}\frac{(\log(1-s))^k}{k!k^j}.$$

Setzt man die Darstellung

$$\frac{(-\log(1-s))^k}{k!} = \sum_{i=k}^{\infty} \frac{s^i}{i!} s(i,k) \tag{70}$$

aus [1, S. 824] in obige Formel ein, so erhält man für $s \in D$

$$\begin{aligned} \mu_j(s) &= \frac{j!s}{s-1} \sum_{k=1}^{\infty} \frac{(-1)^k}{k^j} \sum_{i=k}^{\infty} \frac{s^i}{i!} s(i,k) = \frac{j!s}{1-s} \sum_{i=1}^{\infty} \frac{s^i}{i!} \sum_{k\in[i]} \frac{(-1)^{k+1}}{k^j} s(i,k) \\ &= \Big(\sum_{l=1}^{\infty} s^l\Big) \sum_{i=1}^{\infty} \frac{s^i}{i!} d_{ij} = \sum_{n=2}^{\infty} \Big(\sum_{i\in[n-1]} \frac{d_{ij}}{i!} \Big) s^n \end{aligned}$$

mit

$$d_{ij} := j! \sum_{k\in[i]} \frac{(-1)^{k+1}}{k^j} s(i,k).$$

Durch Koeffizientenvergleich mit der Reihendarstellung $\mu_j(s) = \sum_{n=2}^{\infty} \mu_n^{(j)} s^n$ erhält man die explizite Darstellung $\mu_n^{(j)} = \sum_{i\in[n-1]} d_{ij}/i!$, also die Behauptung. $\square$

Bemerkung 3.1.6 *Satz 3.1.5 ermöglicht es,* $\mathrm{E}(\tau_n^j)$ *für kleine oder mittelgroße Werte von* n *in nicht allzu großer Rechenzeit zu errechnen. Für große* $n \in \mathbb{N}$ *ist diese Formel aufgrund der Verwendung von Stirling-Zahlen sehr rechenintensiv und auch numerisch nicht allzu stabil.*

Um das asymptotische Verhalten von τ_n zu analysieren, kann man eine weitere, auf den ersten Blick unzugänglichere Rekursion für die Funktionen $(c_j)_{j\in\mathbb{N}_0}$ verwenden. Diese ermöglicht es, das Verhalten von $c_j(s)$ nahe der Singularität $s = 1$ genauer zu analysieren, welches entscheidend für das asymptotische Verhalten der j-ten Momente ist.

Lemma 3.1.7 *[35] Die Funktionen* $c_0, c_1, \ldots$ *erfüllen die Rekursion*

$$\begin{aligned} c_j(s) &= \sum_{k\in[j]} \binom{j}{k} (-1)^{k+1} c_{j-k}(s) \Big(\log\log \frac{1}{1-s} \Big)^k \\ &\quad + \int_0^{-\log(1-s)} e^{-u} (-\log u)^j \, du, \quad j \in \mathbb{N}, s \in D, \end{aligned} \tag{71}$$

mit Anfangswert $c_0(s) = s$, $s \in \mathbb{C} \setminus \{1\}$.

Beweis: Partielle Integration von Gleichung (68) ergibt

$$\begin{aligned}
c_1(s) &= \int_0^{-\log(1-s)} \frac{1-e^{-u}}{u}\,du \\
&= [(1-e^{-u})\log u]_0^{-\log(1-s)} - \int_0^{-\log(1-s)} e^{-u}\log u\,du \\
&= s\log\log\frac{1}{1-s} - \int_0^{-\log(1-s)} e^{-u}\log u\,du,
\end{aligned}$$

also ist (71) erfüllt für $j=1$. Für beliebiges $j \in \{2,3,\ldots\}$ erhält man (wieder) durch partielle Integration von (67)

$$\begin{aligned}
c_j(s) &= j\int_0^{-\log(1-s)} \frac{c_{j-1}(1-e^{-u})}{u}\,du \\
&= [jc_{j-1}(1-e^{-u})\log u]_0^{-\log(1-s)} - j\int_0^{-\log(1-s)} c'_{j-1}(1-e^{-u})e^{-u}\log u\,du.
\end{aligned}$$

Aus (69) folgt

$$\lim_{u\searrow 0} c_{j-1}(1-e^{-u})(\log(u))^k = \lim_{u\searrow 0}(-j!\sum_{k=1}^{\infty}\frac{(-u)^k}{k!k^j}(\log(u))^k) = 0$$

für alle $k \in \mathbb{N}$. Aus (66) erhält man zusätzlich

$$c'_{j-1}(s) = (j-1)c_{j-2}(s)/((1-s)(-\log(1-s))),$$

also insbesondere $c'_{j-1}(1-e^{-u}) = (j-1)c_{j-2}(1-e^{-u})/(e^{-u}u)$. Mit beiden Überlegungen ergibt sich

$$c_j(s) = jc_{j-1}(s)\log\log\frac{1}{1-s} - j(j-1)\int_0^{-\log(1-s)}\frac{\log u}{u}c_{j-2}(1-e^{-u})\,du.$$

Eine weitere partielle Integration des letzten Integrals zusammen mit $c'_{j-2}(1-e^{-u}) = (j-2)c_{j-3}(1-e^{-u})/(e^{-u}u)$ liefert analog

$$\begin{aligned}
c_j(s) &= jc_{j-1}(s)\log\log\frac{1}{1-s} - \binom{j}{2}c_{j-2}(s)\Big(\log\log\frac{1}{1-s}\Big)^2 \\
&\quad + \binom{j}{2}(j-2)\int_0^{-\log(1-s)}\frac{\log^2 u}{u}c_{j-3}(1-e^{-u})\,du.
\end{aligned}$$

Durch weiteres $(j-3)$-faches partielles Integrieren des jeweils auftretetenden Integrals erhält man schließlich

$$\begin{aligned} c_j(s) &= \sum_{k=1}^{j-1} \binom{j}{k} (-1)^{k+1} c_{j-k}(s) \Big(\log\log \frac{1}{1-s} \Big)^k \\ &\quad + j \int_0^{-\log(1-s)} \frac{(-\log u)^{j-1}}{u} c_0(1-e^{-u})\, du. \end{aligned}$$

und das Lemma folgt mit $c_0(s) = s$ durch eine weitere partielle Integration des letzten Integrals. □

Lemma 3.1.7 ermöglicht es nun, c_j in der Nähe der Singularität $s = 1$ asymptotisch zu entwickeln.

Lemma 3.1.8 *[35] Sei $j \in \mathbb{N}_0$. Für $s \to 1$ in D gilt*

$$c_j(s) = \sum_{i=0}^{j} m_i \binom{j}{i} \Big(\log\log \frac{1}{1-s} \Big)^{j-i} + O\Big((1-s) \Big(\log\log \frac{1}{1-s} \Big)^j \Big),$$

wobei $(m_i)_{i \in \mathbb{N}_0}$ definiert ist durch

$$m_i := \int_0^{\infty} (-\log u)^i e^{-u}\, du, \quad i \in \mathbb{N}_0. \tag{72}$$

Insbesondere gilt $c_1(s) - \log(-\log(1-s)) \to m_1 = \gamma$ für $s \to 1$ in D.

Beweis: Für $j = 0$ stimmt die Aussage, da $c_0(s) = s$ und $m_0 = 1$. Zur Vereinfachung der kommenden Terme setze $x := -\log(1-s)$ und zeige die äquivalente Darstellung

$$c_j(s) = \sum_{i=0}^{j} m_i \binom{j}{i} \log^{j-i} x + O\Big(\frac{\log^j x}{e^x} \Big), \qquad s \to 1 \text{ in } D,$$

der Behauptung durch Induktion nach j. Nach Lemma 3.1.7 gilt

$$c_j(s) = \sum_{k \in [j]} \binom{j}{k} (-1)^{k+1} c_{j-k}(s) \log^k x + \int_0^x (-\log u)^j e^{-u}\, du. \tag{73}$$

Es gilt

$$\int_0^x e^{-u}(-\log u)^j\,du \;=\; m_j - \int_x^\infty e^{-u}(-\log u)^j\,du \;=\; m_j + O\Big(\frac{\log^j x}{e^x}\Big). \quad (74)$$

Um dies zu sehen, zeigt man mit partieller Integration zunächst für die Stammfunktion

$$\int e^{-u}(-\log u)^j\,du = -e^{-u}(-\log u)^j + j\int e^{-u}(-\log(u))^{j-1}u^{-1}du$$

und stellt dann fest, dass für Wege γ im Holomorphiegebiet des Integranden, die zusätzlich in der Halbebene $\{z\in\mathbb{C}|\Re(z)\geq\Re(x)\}$ verlaufen ($\Re(z)$ bezeichnet den Realteil von $z\in\mathbb{C}$),

$$\Big|\int_\gamma e^{-u}(-\log(u))^{j-1}u^{-1}du\Big| \leq |e^{-x}|\int_\gamma |(-\log(u))^{j-1}u^{-1}|du$$

und somit $\int_x^\infty e^{-u}(-\log u)^j\,du = O\Big(\frac{\log^j x}{e^x}\Big)$ gilt.
Setzt man nun (74) in (73) ein, erhält man

$$\begin{aligned} c_j(s) &= \sum_{k\in[j]}\binom{j}{k}(-1)^{k+1}c_{j-k}(s)\log^k x + m_j + O\Big(\frac{\log^j x}{e^x}\Big) \\ &= \sum_{k\in[j-1]}\binom{j}{k}(-1)^{k+1}c_{j-k}(s)\log^k x + (-1)^{j+1}\log^j x + m_j + O\Big(\frac{\log^j x}{e^x}\Big). \end{aligned}$$

Sei die Aussage für $j-1$ mit $j\in\mathbb{N}$ gezeigt. Dann gilt

$$\begin{aligned} & c_j(s) \\ &= \sum_{k\in[j-1]}\binom{j}{k}(-1)^{k+1}\left(\sum_{i=0}^{j-k} m_i\binom{j-k}{i}\log^{j-k-i}x + O\Big(\frac{\log^{j-k}x}{e^x}\Big)\right)\log^k x \\ &\qquad\qquad + (-1)^{j+1}\log^j x + m_j + O\Big(\frac{\log^j x}{e^x}\Big) \\ &= \sum_{i=0}^{j-1} m_i\binom{j}{i}\log^{j-i}x \underbrace{\sum_{k\in[j-i]}\binom{j-i}{k}(-1)^{k+1}}_{=1} + m_j + O\Big(\frac{\log^j x}{e^x}\Big) \\ &= \sum_{i=0}^{j} m_i\binom{j}{i}\log^{j-i}x + O\Big(\frac{\log^j x}{e^x}\Big). \qquad \square \end{aligned}$$

Bemerkung 3.1.9 *Die Konstante* $m_i = \int_{-\infty}^{\infty} x^i \exp(-x - \exp(-x))\, dx$ *ist gerade das* i*-te Moment der (Standard-)Gumbelverteilung mit Dichte* $x \mapsto \exp(-x - \exp(-x))$, $x \in \mathbb{R}$. *Insbesondere ist* $m_0 = 1$, $m_1 = \gamma$ *(wobei* γ *die Eulerkonstante ist) und* $m_2 = \pi^2/6 + \gamma^2$.

Die Funktion $f(s) := \log\log\frac{1}{1-s}$, die in Lemma 3.1.8 auftaucht, unterscheidet sich von $g(s) := \log(\frac{1}{s}\log\frac{1}{1-s})$ nur um $f(s) - g(s) = \log s$. $\log s$ verschwindet für $s \to 1$ in D. Also gilt $g(s) = O(f(s))$ für $s \to 1$ in D. Verwendet man dies und $\log s = O(1-s)$ für $s \to 1$ in D, ergibt sich folgendes Korollar aus Lemma 3.1.8.

Korollar 3.1.10 *[35] Für* $j \in \mathbb{N}_0$ *und* $s \to 1$ *in* D *gilt*

$$c_j(s) \;=\; \sum_{i=0}^{j} m_i \binom{j}{i} \Big(\log\Big(\frac{1}{s}\log\frac{1}{1-s}\Big)\Big)^{j-i} + O\Big((1-s)\Big(\log\log\frac{1}{1-s}\Big)^j\Big),$$

wobei $(m_i)_{i\in\mathbb{N}_0}$ *wie in (72) definiert ist.*

Bemerkung 3.1.11 *Diese asymptotische Darstellung von* $c_j(s)$ *für* $s \to 1$ *in* D *hat gegenüber der Darstellung aus Lemma 3.1.8 unter anderem den Vorteil, dass die Funktion* $g(s) := \log(\frac{1}{s}\log\frac{1}{1-s})$ *eine Taylor-Entwicklung um 0 besitzt, nämlich* $g(s) = \frac{1}{2}s + \frac{5}{24}s^2 + \frac{1}{8}s^3 + \cdots$.

Mit Hilfe der Darstellung von c_j aus Korollar 3.1.10 kann man eine asymptotische Darstellung der j-ten Momente von τ_n für $n \to \infty$ aufstellen, indem man die Koeffizienten der erzeugende Funktion μ_j (also die zu berechnenden Momente) mittels den Methoden der Singularitätsanalyse aus [32] näherungsweise bestimmt. Hier wird analog zur Bestimmung der Asymptotik der Momente der Gesamtlänge L_n des durch einen Bolthausen-Sznitman-n-Coalescent gegebenen Baumes in [27] vorgegangen.

Bemerkung 3.1.12 *Grob gesprochen versucht die Singularitätenanalyse, die Koeffizienten einer erzeugenden Funktion (aufgefasst als holomorphe Funktion) mittels Cauchy-Integralformel zumindest approximativ zu berechnen, nur dass das Cauchy-Integral anstatt über einen Kreis über eine sogenannte Hankel-Kurve gebildet wird. Eine umfangreiche Beschreibung dieser Methode findet sich in [33, Kapitel VI].*

Satz 3.1.13 *(Momente von τ_n im Bolthausen-Sznitman-Coalescent)[35] Betrachte τ_n in einem Bolthausen-Sznitman-n-Coalescent für $n \in \mathbb{N}$. Das j-te Moment von τ_n erfüllt*

$$\mathrm{E}(\tau_n^j) \;=\; \sum_{i=0}^{j} m_i \binom{j}{i} (\log\log n)^{j-i} + O\Big(\frac{(\log\log n)^j}{\log n}\Big), \qquad n \to \infty,$$

wobei m_i, $i \in \mathbb{N}_0$, wie in (72) definiert ist. Insbesondere gilt

$$\mathrm{E}(\tau_n^j) \sim (\log\log n)^j \text{ und } Var(\tau_n) = \frac{\pi^2}{6} + O\Big(\frac{(\log\log n)^2}{\log n}\Big) \tag{75}$$

für $n \to \infty$.

Bemerkung 3.1.14 *Hier und in allen folgenden Analysen von erzeugenden Funktionen bzw. Potenzreihen verwende $[s^n]f(s)$ als Schreibweise für den Koeffizienten von s^n in einer Potenzreihe f, d.h. für $f(s) = \sum_n f_n s^n$ gilt $[s^n]f(s) = f_n$.*

Beweis:(von Satz 3.1.13) Nach Korollar 3.1.10 und (64) erhält man für $j \in \mathbb{N}_0$

$$\begin{aligned}\frac{\mu_j(s)}{s} \;&=\; \frac{c_j(s)}{1-s} \\ =\; \frac{1}{1-s}\sum_{i=0}^{j} m_i \binom{j}{i} \Big(\log\Big(\frac{1}{s}\log\frac{1}{1-s}\Big)\Big)^{j-i} &+ O\Big(\Big(\log\log\frac{1}{1-s}\Big)^j\Big)\end{aligned} \tag{76}$$

für $s \to 1$ in D. Wendet man [32, S. 225, Theorem 2] mit den Parametern $\alpha = \gamma := 0$ und $\delta := j$ (in dortiger Notation) auf den $O(\cdot)$-Term auf der rechten Seite von (76) an, so sieht man, dass der n-te Koeffizient h_n der Potenzreihendarstellung des $O(\cdot)$-Terms $h_n = O((\log\log n)^j/n)$ für $n \to \infty$ erfüllt. Des Weiteren erhält man durch Anwendung der Erweiterung von [32, Theorem 3B, S. 230] für die modifizierten Parameter $\alpha := -1$, $\gamma := 0$ und $\delta \in \mathbb{N}_0$ aus Satz 7.0.1

$$[s^n]\frac{1}{1-s}\Big(\log\Big(\frac{1}{s}\log\frac{1}{1-s}\Big)\Big)^{\delta} \;=\; (\log\log n)^{\delta} + O\Big(\frac{(\log\log n)^{\delta}}{\log n}\Big), \qquad n \to \infty.$$

Somit erhält man aus (76), dass der n-te Koeffizient der Potenzreihe $\mu_j(s)$

$$[s^n]\mu_j(s) \;=\; \sum_{i=0}^{j} m_i \binom{j}{i} (\log\log n)^{j-i} + O\Big(\frac{(\log\log n)^j}{\log n}\Big), \qquad n \to \infty, \tag{77}$$

erfüllt, was die gewünschte Momentenformel zeigt. Die Darstellung des Erwartungswertes folgt sofort, und mit den Werten für m_0, m_1 und m_2 aus Bemerkung 3.1.9 gilt

$$\begin{aligned}
& Var(\tau_n) = [s^n]\mu_2(s) - ([s^n]\mu_1(s))^2 \\
= \quad & (\log\log n)^2 + 2\gamma \log\log n + \gamma^2 + \frac{\pi^2}{6} + O\Big(\frac{(\log\log n)^2}{\log n}\Big) \\
& -\big(\log\log n + \gamma + O\big(\frac{\log\log n}{\log n}\big)\big)^2 \\
= \quad & \frac{\pi^2}{6} + O(\frac{(\log\log n)^2}{\log n}),
\end{aligned}$$

wobei γ die Euler-Konstante ist. □

Bemerkung 3.1.15 *Gleichung (77) in obigem Beweis ist faktisch [32, Korollar 6, S. 230], wiederum mit modifizierten Parametern $\alpha := -1$ und $\gamma := 0$.*

Die nachfolgende Tabelle zeigt Erwartungswert, zweites Moment und Varianz für τ_n für einige $n \in \mathbb{N}$, ausgerechnet mit Hilfe der expliziten Momentenformel aus Satz 3.1.5. Zum Vergleich sind auch die asymptotischen Näherungen für diese Größen aus Satz 3.1.13 angegeben.

Tabelle 2: Erwartungswert, zweites Moment und Varianz von τ_n

n	$\mathrm{E}(\tau_n)$	$\mathrm{E}(\tau_n^2)$	$\mathrm{Var}(\tau_n)$
1	0	0	0
2	1	2	1
3	$\frac{5}{4} = 1.25$	$\frac{11}{4} = 2.75$	$\frac{19}{16} = 1.1875$
4	$\frac{25}{18} \approx 1.388889$	$\frac{173}{54} \approx 3.203704$	$\frac{413}{324} \approx 1.274691$
5	$\frac{427}{288} \approx 1.482639$	$\frac{6\,091}{1\,728} \approx 3.524884$	$\frac{110\,039}{82\,944} \approx 1.326666$
6	$\frac{11\,177}{7\,200} \approx 1.552361$	$\frac{814\,669}{216\,000} \approx 3.771616$	$\frac{70\,595\,231}{51\,840\,000} \approx 1.361791$
10	≈ 1.723214	≈ 4.405107	≈ 1.435641
100	≈ 2.256059	≈ 6.662353	≈ 1.572549
200	≈ 2.374378	≈ 7.226326	≈ 1.588656
300	≈ 2.437855	≈ 7.538955	≈ 1.595816
400	≈ 2.480663	≈ 7.753821	≈ 1.600134
500	≈ 2.512697	≈ 7.916765	≈ 1.603119
∞	$\approx \log\log n + \gamma$	$\approx (\log\log n + \gamma)^2 + \frac{\pi^2}{6}$	$\sim \frac{\pi^2}{6} \approx 1.644934$

Aus Theorem 3.1.13 wird zunächst ein starkes Gesetz der großen Zahlen für $(\tau_n)_{n\in\mathbb{N}}$ gefolgert.

Satz 3.1.16 *(Starkes Gesetz der großen Zahlen für τ_n)[35] Sei Π ein Bolthausen-Sznitman-Coalescent und $\Pi^{(n)} := (\rho_n(\Pi_t))_{t\geq 0}$. Für τ_n in $\Pi^{(n)}$ gilt $\tau_n/\log\log n \to 1$ fast sicher.*

Beweis: Zeige zunächst $\tau_n/E(\tau_n) \xrightarrow{p} 1$ für $n \to \infty$. Aus (75) erhält man $E(\tau_n) \sim \log\log n$ und $Var(\tau_n) = \frac{\pi^2}{6} + O(\frac{(\log\log n)^2}{\log n})$ für $n \to \infty$. Mit der Tschebyscheff-Ungleichung folgt für $\epsilon > 0$

$$P(|\tau_n - E(\tau_n)| \geq \epsilon E(\tau_n)) \leq (\epsilon E(\tau_n))^{-2}(\frac{\pi^2}{6} + O(\frac{(\log\log n)^2}{\log n})) \to 0 \qquad (78)$$

für $n \to \infty$, also $\tau_n/E(\tau_n) \to 1$ stochastisch für $n \to \infty$.
Betrachte nun die Teilfolge $n_k := \lfloor \exp(\exp(k)) \rfloor$, $k \in \mathbb{N}$, von $\mathbb{N}$. Hier gilt analog zu (78)

$$\begin{aligned} & \sum_{k\in\mathbb{N}} P(|\tau_{n_k} - E(\tau_{n_k})| \geq \epsilon E(\tau_{n_k})) \leq \sum_{k\in\mathbb{N}} (\epsilon E(\tau_{n_k})^{-2} C_1 \frac{\pi^2}{6}) \\ \leq \quad & C_2 C_1 \epsilon \frac{\pi^2}{6} \sum_{k\in\mathbb{N}} \frac{1}{k^2} < \infty \end{aligned}$$

für geeignete Konstanten $C_1, C_2 > 0$, da $E(\tau_{n_k}) \sim k$ für $k \to \infty$ nach (75). Nach dem Borel-Cantelli-Lemma gilt somit $\tau_{n_k}/E(\tau_{n_k}) \to 1$ sogar P-fast sicher für $k \to \infty$. Nun ist $(\tau_n)_{n\in\mathbb{N}}$ pfadweise monoton steigend, also auch $(E(\tau_n))_{n\in\mathbb{N}}$. Man erhält die obere und untere Grenze

$$\frac{\tau_{n_k}}{E(\tau_{n_k})}\frac{E(\tau_{n_k})}{E(\tau_{n_{k+1}})} \;\leq\; \frac{\tau_n}{E(\tau_n)} \;\leq\; \frac{\tau_{n_{k+1}}}{E(\tau_{n_{k+1}})}\frac{E(\tau_{n_{k+1}})}{E(\tau_{n_k})}, \qquad n \in \{n_k, \ldots, n_{k+1}\},$$

von $\tau_n/E(\tau_n)$. Aus Satz 3.1.13 weiß man aber $E(\tau_{n_{k+1}})/E(\tau_{n_k}) \sim (k+1)/k \sim 1$, also gilt $\tau_n/E(\tau_n) \to 1$ fast sicher auch für $n \to \infty$. Wieder mit $E(\tau_n) \sim \log\log n$ für $n \to \infty$ folgt die Behauptung. □

Skaliert man nun τ_n gar nicht, sondern zentriert es nur, so erhält man sogar Konvergenz gegen eine nicht degenerierte Verteilung.

Bemerkung 3.1.17 *Das folgende Resultat wurde schon in [42, Proposition 3.4, S. 729] bewiesen, dort allerdings mit anderen Methoden, nämlich der Konstruktion des Bolthausen-Sznitman-Coalescents über zufällige rekursive Bäume.*

Satz 3.1.18 *(Verteilungskonvergenz von zentriertem τ_n) [35] Für τ_n im Bolthausen-Sznitman-n-Coalescent gilt*

$$\tau_n - \log\log n \xrightarrow{d} \tau$$

für $n \to \infty$, wobei τ eine (Standard-)Gumbel-verteilte Zufallsvariable mit Verteilungsfunktion $x \mapsto \exp(-\exp(-x))$, $x \in \mathbb{R}$, ist.

Beweis: Aus Bemerkung 3.1.9 ist bekannt, dass $\mathrm{E}(\tau^i) = m_i$ für $i \in \mathbb{N}$ ist mit m_i aus (72). Aus Satz 3.1.13 folgert man für $k \in \mathbb{N}_0$

$$\begin{aligned}
\mathrm{E}((\tau_n - \log\log n)^k) &= \sum_{j=0}^{k} \binom{k}{j} \mathrm{E}(\tau_n^j)(-\log\log n)^{k-j} \\
&= \sum_{j=0}^{k} \binom{k}{j} \left(\sum_{i=0}^{j} m_i \binom{j}{i} (\log\log n)^{j-i} + O\Big(\frac{(\log\log n)^j}{\log n}\Big) \right) (-\log\log n)^{k-j} \\
&= \sum_{i=0}^{k} m_i \binom{k}{i} (\log\log n)^{k-i} \sum_{j=i}^{k} \binom{k-i}{k-j} (-1)^{k-j} + O\Big(\frac{(\log\log n)^k}{\log n}\Big) \\
&= m_k + O\Big(\frac{(\log\log n)^k}{\log n}\Big),
\end{aligned}$$

also insbesondere $\lim_{n\to\infty} \mathrm{E}((\tau_n - \log\log n)^k) = m_k$ für $k \in \mathbb{N}_0$. Somit konvergieren alle Momente von $(\tau_n - \log\log n)_{n\in\mathbb{N}}$ gegen die Momente der Gumbel-Verteilung. Nun sind die Momente m_i, $i \in \mathbb{N}_0$, der Gumbel-Verteilung nicht negativ und es gilt $\sum_{i=0}^{\infty} m_i t^i/i! = \mathrm{E}(e^{t\tau}) = \Gamma(1-t) < \infty$ für $t \in \mathbb{R}$ mit $t < 1$. Also folgt mit [14, Theoreme 30.1, 30.2] aus der Konvergenz der Momente die Konvergenz in Verteilung. □

Bemerkung 3.1.19 *Die asymptotischen Resultate für τ_n sind potentiell auch für Biologen interessant (siehe Kapitel 2.2). Da der Bolthausen-Sznitman-Coalescent im Unendlichen bleibt (stays infinite, siehe etwa [80, Proposition 11]), war es zu erwarten, dass $\tau_n \to \infty$ fast sicher für $n \to \infty$*

gilt. Allerdings führt für Fragen in der Praxis die sehr langsame Divergenzgeschwindigkeit von $\log\log n$ *dazu, dass der Bolthausen-Sznitman-Coalescent ein Grenzfall ist, der „fast" nicht im Unendlichen bleibt. Es gilt für alle praktischen Fragen* $\log\log n < 6$ *(Für* $n = 10^{78}$*, die ungefähre Anzahl der Atome im Universum, erhält man* $\log\log n \approx 5.2$*). In gewisser Weise legen also die asymptotischen Ergebnisse für* $(\tau_n)_{n\in\mathbb{N}}$ *nahe, dass man für praktische Fragen annehmen kann, dass der Bolthausen-Sznitman-Coalescent* ***doch*** *nicht im Unendlichen bleibt.*

4 E_n, die Länge eines zufällig ausgewählten externen Zweiges

In diesem Kapitel wird das Funktional E_n, also die Länge eines zufällig ausgewählten externen Zweiges eines n-Coalescents $\Pi^{(n)}$ genauer analysiert, insbesondere im Hinblick auf die Asymptotik. Für einige Maße Ξ sind asymptotische Resultate bekannt. Aus Satz 2.3.1 d) ist bekannt, dass die Länge $E^{(i)}$ des externen Zweiges des Individuums $i \in \mathbb{N}$ in einem Ξ-Coalescent Π exponentialverteilt mit Parameter μ_{-1} ist. Des Weiteren besagt Satz 2.3.1 a), dass für die Länge $E_n^{(i)}$ des externen Zweiges von Individuum i im zu Π gehörenden n-Coalescent $(\rho_n(\Pi_t))_{t\geq 0}$

$$\lim_{n\to\infty} E_n^{(i)} = E^{(i)}$$

gilt. Aufgrund der Austauschbarkeit der $(E^{(i)}_{i\in\mathbb{N}})$ (Satz 2.3.1 c)) gilt $E_n^{(i)} \stackrel{d}{=} E_n$, da das zufällige Auswählen unabhängig vom Coalescent Π stattfindet. Somit erhält man das Konvergenzresultat

$$E_n \xrightarrow{d} E^{(1)} \stackrel{d}{=} Exp(\mu_{-1})$$

für $n \to \infty$. Dieses Konvergenzresultat ist nicht degeneriert, falls $\mu_{-1} < \infty$ gilt, also für Ξ-Coalescents mit Staub. Das Konvergenzresultat findet sich etwa in [68, Proposition 2], vergleiche dazu Bemerkung 2.3.3.

Für die biologische Anwendung (vergleiche Kapitel 1.6) ist meist der Kingman-Coalescent von Bedeutung. Deswegen verwundert es nicht, dass die Asymptotik der externen Zweiglängen für den Kingman-Coalescent bereits ausgiebig analysiert wurde. Für die mathematische Behandlung sei hier insbesondere auf [22] hingewiesen. Dort wird für den Kingman-Coalescent gezeigt, dass

$$nE_n \xrightarrow{d} E,\ E \text{ hat Dichte } x \mapsto 8/(2+x)^3,\ x > 0.$$

Zum Beweis dieser Aussage wird in [22] die Darstellung $E_n \stackrel{d}{=} \sum_{k=0}^{C_n^{ext}} T_{n-k}$ in einer Version für charakteristische Funktionen verwendet, wobei C_n^{ext} die Kollisionen im n-Coalescent vor dem Ende des ausgewählten externen Zweiges sind und $T_2, \ldots, T_n$ unabhängige, auch von C_n^{ext} unabhängige Zufallsvariable sind mit $T_i \stackrel{d}{=} Exp(\binom{i}{2})$ für $i \in \{2, \ldots, n\}$. Diese Darstellung ist die Darstellung von E_n aus Satz 2.4.3 b) für den Kingman-n-Coalescent. Hier sei daran erinnert, dass die Sprungkette $(\chi_k^{(n)})_{k\in\mathbb{N}}$ des Block-Zählprozesses des Kingman-n-Coalescents $\chi_k^{(n)} = n-k$ für $k \in \{0,\ldots,n-1\}$ erfüllt und die zugehörigen totalen Raten $g_i \overset{d}{=} \binom{i}{2}$ für $i \in [n]$ sind.

4.1 E_n im Bolthausen-Sznitman-n-Coalescent [35]

Betrachtet man für $n \in \mathbb{N}$ die Länge E_n in einem Bolthausen-Sznitman-n-Coalescent $\Pi^{(n)}$, so ist die Darstellung $E_n \overset{d}{=} \sum_{k=0}^{C_n^{ext}} T_{\chi_k^{(n)}}$ aus Satz 2.4.3 a) durch die Abhängigkeiten zwischen der Sprungkette $(\chi_k^{(n)})_{k\in\mathbb{N}_0}$ von $\Pi^{(n)}$ und dem Funktional C_n^{ext} nicht mehr gut dazu geeignet, das asymptotische Verhalten von E_n für $n \to \infty$ zu analysieren. Es bietet sich an, analog zur vorangegangenen Analyse von τ_n, der Zeit zurück zum ersten Urahn, vorzugehen. Aus Rekursion (45) für die j-ten Momente von E_n wird eine Differentialgleichung für die erzeugende Funktion der j-ten Momente gebildet. Aus der erzeugenden Funktion wird mittels Singularitätsanalyse das j-te Moment asyptotisch bestimmt und dann aus der Momentenkonvergenz für alle $j \in \mathbb{N}$ auf die Verteilungskonvergenz der $(E_n)_{n\in\mathbb{N}}$ für $n \to \infty$ geschlossen.
Setzt man in die Rekursion (45) für $\nu_n^{(j)} = E(E_n^j)$, $n, j \in \mathbb{N}$, die passenden Werte für den Bolthausen-Sznitman-Coalescent ein, also die totale Rate $g_n = n-1$ im Zustand $n \in \mathbb{N}$ des Block-Zählprozesses eines Bolthausen-Sznitman-n-Coalescent und die Verteilung $(r_{nk})_{k\in[n-1]} = (\frac{n}{(n-1)(n-k)(n-k+1)})_{k\in[n-1]}$ des ersten Sprungs dieses Prozesses (siehe (58) und (59)), so erhält man

$$\nu_n^{(j)} = \sum_{k\in[n-1]} \frac{k-1}{(n-1)(n-k)(n-k+1)} \nu_k^{(j)} + \tilde{r}_n^{(j)}, \tag{79}$$

für $n \in \{2,3\ldots\}$ und $j \in \mathbb{N}_0$ mit Restterm

$$\tilde{r}_n^{(j)} := \mathrm{E}(T_n^j) + \sum_{i\in[j-1]} \binom{j}{i} \mathrm{E}(T_n^i) \sum_{k\in[n-1]} \frac{k-1}{(n-1)(n-k)(n-k+1)} \nu_k^{(j-i)},$$

wobei $T_n \overset{d}{=} Exp(n-1)$. Für $j \in \mathbb{N}_0$ betrachte die erzeugenden Funktionen

$$\nu_j(s) := \sum_{n=2}^{\infty} \nu_n^{(j)} s^n \quad \text{und} \quad \tilde{r}_j(s) := \sum_{n=2}^{\infty} (n-1)\tilde{r}_n^{(j)} s^n. \tag{80}$$

Aus Gleichung (47) erhält man mit $g_n = n-1$ die Formel $\tilde{r}_n^{(j)} = j(n-1)^{-1}\nu_n^{(j-1)}$ und damit $\tilde{r}_j(s) = j\nu_{j-1}(s)$, $j \in \mathbb{N}$.

Lemma 4.1.1 *(Rekursion für die erzeugenden Funktionen ν_j)[35] Für den Bolthausen-Sznitman-Coalescent erfüllen die erzeugenden Funktionen $\nu_0, \nu_1, \ldots$ die Rekursion*

$$\nu_j(s) \;=\; js \int_0^s \frac{\nu_{j-1}(t)}{t(1-t)(-\log(1-t))}\,dt, \qquad j \in \mathbb{N}, s \in D, \tag{81}$$

mit Anfangsbedingung $\nu_0(s) = s^2/(1-s)$, $s \in \mathbb{C} \setminus \{1\}$. Insbesondere gilt

$$\nu_1(s) \;=\; s \int_0^s \frac{t}{(1-t)^2(-\log(1-t))}\,dt, \qquad s \in D. \tag{82}$$

Bemerkung 4.1.2 *Auch hier werden Differentialgleichungen mit komplexen Argumenten betrachtet (und komplexer Integration). Von Integrationswegen in Wegintegralen werden wieder nur Anfangs- und Endpunkt angegeben.*

Beweis: Es wird wieder die Funktion

$$a(s) \;:=\; \sum_{n=1}^{\infty} \frac{s^n}{n(n+1)} \;=\; 1 - \log(1-s) + \frac{\log(1-s)}{s}.$$

aus (61) verwendet. Man erhält unter Verwendung von (79)

$$\begin{aligned}
s^2 \frac{\partial}{\partial s}\frac{\nu_j(s)}{s} \;&=\; \sum_{n=2}^{\infty} (n-1)\nu_n^{(j)} s^n \\
&=\; \sum_{n=2}^{\infty}(n-1)\Big(\sum_{k=1}^{n-1} \frac{k-1}{(n-1)(n-k)(n-k+1)}\nu_k^{(j)} + \tilde{r}_n^{(j)}\Big)s^n \\
&=\; \sum_{n=2}^{\infty}(n-1)\tilde{r}_n^{(j)} s^n + \sum_{k=1}^{\infty}(k-1)\nu_k^{(j)} s^k \sum_{n=k+1}^{\infty} \frac{s^{n-k}}{(n-k)(n-k+1)} \\
&=\; \tilde{r}_j(s) + a(s)s^2 \frac{\partial}{\partial s}\frac{\nu_j(s)}{s},
\end{aligned}$$

also

$$\tilde{r}_j(s) \;=\; (1-a(s))s^2 \frac{\partial}{\partial s}\frac{\nu_j(s)}{s} \;=\; (1-a(s))(s\nu_j'(s) - \nu_j(s)),$$

oder umgeformt

$$\nu_j'(s) - \frac{\nu_j(s)}{s} \;=\; \frac{\tilde{r}_j(s)}{s(1-a(s))}. \tag{83}$$

Die Lösungen der homogenen Differentialgleichung $f'(s) = f(s)/s$ haben die Form $f(s) = ds$, $d \in \mathbb{C}$. Somit hat die Lösung der inhomogenen Differentialgleichung (83) mit Anfangswert $\nu_j(0) = 0$ nach Variation der Konstanten die Gestalt $\nu_j(s) = s d_j(s)$ mit

$$d_j(s) \;=\; \int_0^s \frac{\tilde{r}_j(t)}{t^2(1-a(t))}\,dt.$$

Um das gewünschte Ergebnis zu bekommen, setze $1-a(t) = (1-t)(-\log(1-t))/t$ und $\tilde{r}_j(t) = j\nu_{j-1}(t)$ ein. □

Bemerkung 4.1.3 *[35] Aus obigem Lemma erhält man, dass d_j die Rekursion*

$$d_j(s) \;=\; j\int_0^s \frac{d_{j-1}(t)}{(1-t)(-\log(1-t))}\,dt, \quad j \in \mathbb{N}, s \in D, \tag{84}$$

mit Anfangsbedingung $d_0(s) = s/(1-s)$, $s \in \mathbb{C} \setminus \{1\}$ erfüllt. Per Induktion nach j weist man analog zu Lemma 3.1.3 die Reihenentwicklung

$$d_j(s) \;=\; j!\sum_{k=1}^{\infty} \frac{(-\log(1-s))^k}{k!k^j}, \qquad j \in \mathbb{N}_0, s \in D,$$

nach. Somit ist d_j und damit auch ν_j für $j \in \mathbb{N}$ holomorph auf D und analytisch fortsetzbar auf $\mathbb{C} \setminus [1,\infty)$.

Aus der erzeugenden Funktion ν_j lässt sich leicht eine explizite Darstellung der j-ten Momente von E_n zeigen, wie sie schon in (3.1.5) für die j-ten Momente von τ_n aufgestellt wurde.

Satz 4.1.4 *(explizite Darstellung der Momente von E_n) [35] Sei $n \in \mathbb{N}$ und $j \in \mathbb{N}_0$. Betrachte E_n, die Länge eines zufällig ausgewählten externen Zweiges in einem Bolthausen-Sznitman-n-Coalescent. Dann gilt für das j-te Moment von E_n*

$$\mathrm{E}(E_n^j) \;=\; \frac{j!}{(n-1)!}\sum_{k\in[n-1]} \frac{s(n-1,k)}{k^j},$$

wobei $s(i,k)$ die absolute Stirlingzahl erster Art mit Parametern i,k bezeichnet.

Beweis: Der Beweis läuft analog zum Beweis von 3.1.5 und ist sogar einfacher. Setzt man in

$$\nu_j(s) \ = \ sd_j(s) \ = \ j!s\sum_{k=1}^{\infty}\frac{(-\log(1-s))^k}{k!k^j}$$

die Darstellung (70) aus [1, S. 824] ein, so erhält man

$$\begin{aligned}\nu_j(s) \ &= \ j!s\sum_{k=1}^{\infty}\frac{1}{k^j}\sum_{i=k}^{\infty}\frac{s^i}{i!}s(i,k) \ = \ j!s\sum_{i=1}^{\infty}\frac{s^i}{i!}\sum_{k\in[i]}\frac{1}{k^j}s(i,k)\\ &= \ \sum_{n=2}^{\infty}\Big(\frac{j!}{(n-1)!}\sum_{k\in[n-1]}\frac{s(n-1,k)}{k^j}\Big)s^n.\end{aligned}$$

Ein Koeffizientenvergleich mit der Darstellung $\nu_j(s) = \sum_{n=2}^{\infty}\nu_n^{(j)}s^n$ liefert die gewünschte Darstellung. □

Satz 4.1.4 liefert für kleine bis moderate $n \in \mathbb{N}$ eine gut benutzbare/berechenbare Darstellung der Momente von E_n. Um das asymptotische Verhalten der Verteilungen von $(E_n)_{n\in\mathbb{N}}$ zu analysieren, ist die Momentenformel allerdings nicht gut geeignet, da die Stirling-Zahlen für große $n \in \mathbb{N}$ schlecht handhabbar sind. Somit ist es wieder ratsam, die Koeffizienten der erzeugenden Funktion ν_j asymptotisch zu entwickeln und dann das asymptotische Verhalten dieser Koeffizienten, also der Momente, zu betrachten. Dazu wird wiederum Singularitätsanalyse verwendet, aber in einer schon auf die spezielle Form von ν_j angepassten Version aus [70] (die aber auf [32] zurück geht).

Satz 4.1.5 *(Asymptotik der Momente von E_n) [35] Sei $j \in \mathbb{N}_0$. Im Bolthausen-Sznitman-n-Coalescent hat das j-te Moment von E_n die asymptotische Entwicklung*

$$\mathrm{E}(E_n^j) \ = \ \frac{j!}{\log^j n}\Big(1+\frac{\kappa_j}{\log n}+O\Big(\frac{1}{\log^2 n}\Big)\Big), \qquad n\to\infty,$$

wobei $\kappa_j := j((j+1)/2-\gamma)$, $j \in \mathbb{N}_0$ und γ die Eulerkonstante bezeichnet. Insbesondere gilt

$$E(E_n) \sim \frac{1}{\log n} \text{ und } Var(E_n) = \frac{1}{\log^2 n} + O(\frac{1}{\log^3 n}).$$

Bemerkung 4.1.6 *Es gilt $\kappa_0 = 0$, $\kappa_1 = 1 - \gamma = \Psi^{(0)}(2)$ und $\kappa_j = j + \Psi^{(0)}(1) + \kappa_{j-1}$ für $j \in \mathbb{N}$, wobei $\Psi^{(0)} := \Gamma'/\Gamma$ die Psi-Funktion, also die logarithmische Ableitung der Gammafunktion bezeichnet. Einige Eigenschaften der Psi-Funktion finden sich in [1, S. 258/259].*

Beweis:(von Theorem 4.1.5) Der Beweis verläuft analog zu [70, Theorem 2.1]. Alle vorkommenden Funktionen werden als Potenzreihen aufgefasst und alle O-Terme gelten für $n \to \infty$. Es werden für $\alpha, p > 0$ die Formel [70, Gl. (19)]

$$[t^n]\frac{1}{(1-t)^\alpha(-\log(1-t))^p} = \frac{n^{\alpha-1}}{\Gamma(\alpha)\log^p n}\Big(1+\frac{p\Psi^{(0)}(\alpha)}{\log n}+O\Big(\frac{1}{\log^2 n}\Big)\Big) \quad (85)$$

für das asymptotische Wachstum der Koeffizienten sowie die Formel [70, Gl. (20)]

$$[s^n]\int_0^s F(t)\,dt = \frac{n^{\alpha-1}}{\log^p n}\Big(1+O\Big(\frac{1}{n}\Big)\Big) \quad (86)$$

für das asymptotische Wachstum der Koeffizienten bei Integration der erzeugenden Funktion $F(s) = \sum_{n=2}^\infty s^n n^\alpha/(\log^p n)$, $\alpha, p > 0$ verwendet. Dazu wird noch die asymptotische Entwicklung der Summe aus [70, Gl. (16)] verwendet, wobei man $\Psi^{(0)}(1) - \Psi^{(0)}(2) = 1$ verwendet (etwa [1, Stichpunkt 6.3.2]): Für $p, q \geq 0$ gilt

$$\sum_{k=2}^{n-2}\frac{1}{\log^p k\log^q(n-k)} = \frac{n}{\log^{p+q} n}\Big(1+\frac{p+q}{\log n}+O\Big(\frac{1}{\log^2 n}\Big)\Big). \quad (87)$$

Nun zum Beweis des Satzes. Wegen $\nu_j(s) = sd_j(s)$ genügt es,

$$[s^n]d_j(s) = \frac{j!}{\log^j n}\Big(1+\frac{\kappa_j}{\log n}+O\Big(\frac{1}{\log^2 n}\Big)\Big), \quad j \in \mathbb{N}, \quad (88)$$

zu zeigen, da Multiplikation mit s nur die Indizes der Koeffizienten der Potenzreihe d_j um 1 verschiebt. Die Gleichung (88) wird per Induktion über $j \in \mathbb{N}$ gezeigt. Mit (85) erhält man

$$[t^n]\frac{t}{(1-t)^2(-\log(1-t))} = \frac{n}{\log n} + \Psi^{(0)}(2)\frac{n}{\log^2 n} + O\Big(\frac{n}{\log^3 n}\Big).$$

Integriert man $\frac{t}{(1-t)^2(-\log(1-t))}$ über t und verwendet man die Integraldarstellung (84) für d_1, so erhält man

$$\begin{aligned}[s^n]d_1(s) &= [s^n]\int_0^s \frac{t}{(1-t)^2(-\log(1-t))}\,dt \\ &= \frac{1}{\log n} + \frac{\Psi^{(0)}(2)}{\log^2 n} + O\Big(\frac{1}{\log^3 n}\Big),\end{aligned}$$

da Integrieren einer Potenzreihe jeden Koeffizient $[s^n]$ an die Stelle von $[s^{n+1}]$ verschiebt und einen Vorfaktor $(n+1)^{-1} \sim n^{-1}$ hinzufügt. Wegen $\Psi^{(0)}(2) = 1-\gamma = \kappa_1$ gilt (88) für $j=1$.
Sei (88) für $j-1$ gezeigt. Betrachte (85), genauer

$$[t^n]\frac{1}{(1-t)(-\log(1-t))} = \frac{1}{\log n} + \frac{\Psi^{(0)}(1)}{\log^2 n} + O\Big(\frac{1}{\log^3 n}\Big).$$

Es gilt

$$\begin{aligned}[t^n]\frac{jd_{j-1}(t)}{(1-t)(-\log(1-t))} &= \sum_{k\in[n-2]} j[t^k]d_{j-1}(t)[t^{n-k}]\frac{1}{(1-t)(-\log(1-t))} \\ &= \sum_{k\in[n-2]} j\Big(\frac{(j-1)!}{\log^{j-1} k} + \frac{(j-1)!\kappa_{j-1}}{\log^j k} + O\Big(\frac{1}{\log^{j+1} k}\Big)\Big)\cdot \\ &\qquad \cdot\Big(\frac{1}{\log(n-k)} + \frac{\Psi^{(0)}(1)}{\log^2(n-k)} + O\Big(\frac{1}{\log^3(n-k)}\Big)\Big) \\ &= \sum_{k=2}^{n-2}\frac{j!}{\log^{j-1} k\log(n-k)} + \sum_{k=2}^{n-2}\frac{j!\Psi^{(0)}(1)}{\log^{j-1} k\log^2(n-k)} + \\ &\quad + \sum_{k=2}^{n-2}\frac{j!\kappa_{j-1}}{\log^{j} k\log(n-k)} + \sum_{k=2}^{n-2}\frac{j!\kappa_{j-1}\Psi^{(0)}(1)}{\log^{j} k\log^2(n-k)} + O\Big(\frac{n}{\log^{j+2} n}\Big),\end{aligned}$$

wobei man $jd_{j-1}(t)\cdot(\frac{1}{1-t}(-\log(1-t)))$ als Cauchyprodukt zweier Reihen der Form $b_1t^1+b_2t^2+\ldots$ und $a_2t^2+a_3t^3+\ldots$ auffasst und die Induktionsvoraussetzung und obigen Spezialfall von (85) einsetzt. Wendet man (87) auf jede dieser vier Summen an, so ergibt sich

$$\begin{aligned}
&[t^n]\frac{jd_{j-1}(t)}{(1-t)(-\log(1-t))} \\
&= j!\frac{n}{\log^j n}\Big(1+\frac{j}{\log n}\Big) + j!\Psi^{(0)}(1)\frac{n}{\log^{j+1} n} + j!\kappa_{j-1}\frac{n}{\log^{j+1} n} + O\Big(\frac{n}{\log^{j+2} n}\Big) \\
&= j!\frac{n}{\log^j n}\Big(1+\frac{j+\Psi^{(0)}(1)+\kappa_{j-1}}{\log n} + O\Big(\frac{1}{\log^2 n}\Big)\Big) \\
&= j!\frac{n}{\log^j n}\Big(1+\frac{\kappa_j}{\log n} + O\Big(\frac{1}{\log^2 n}\Big)\Big).
\end{aligned}$$

Analog zum Induktionsanfang integriert man nun $\frac{jd_{j-1}(t)}{(1-t)(-\log(1-t))}$ nach t und verwendet (84). Dies ergibt

$$[s^n]d_j(s) \;=\; [s^n]\int_0^s \frac{jd_{j-1}(t)}{(1-t)(-\log(1-t))}\,dt \;=\; \frac{j!}{\log^j n}\Big(1+\frac{\kappa_j}{\log n}+O\Big(\frac{1}{\log^2 n}\Big)\Big),$$

da die Integration asymptotisch nur für einen Faktor n^{-1} im n-ten Koeffizienten der Potenzreihe sorgt (siehe Induktionsanfang). Also gilt (88) für alle $j \in \mathbb{N}$ und dies zeigt die gewünschte Darstellung der Momente. Insbesondere gilt

$$\begin{aligned}
&Var(E_n) \\
&= \frac{2}{\log^2 n}\big(1+\frac{\kappa_2}{\log n}+O(\frac{1}{\log^2 n})\big) - \big(\frac{1}{\log n}\big(1+\frac{\kappa_1}{\log n}+O(\frac{1}{\log^2 n})\big)\big)^2 \\
&= \frac{2}{\log^2 n}\big(1+\frac{\kappa_2}{\log n}+O(\frac{1}{\log^2 n})\big) - \frac{1}{\log^2 n}\big(1+\frac{2\kappa_1}{\log n}+O(\frac{1}{\log^2 n})\big) \\
&= \frac{1}{\log^2 n}+O\big(\frac{1}{\log^3 n}\big). \qquad \square
\end{aligned}$$

Die folgende Tabelle wurde mittels der Momentenformel aus Satz 4.1.4 berechnet; zum Vergleich sind die asymptotischen Werte aus Satz 4.1.5 angegeben.

Tabelle 3: Erwartungswert, zweites Moment und Varianz von E_n

n	$\mathrm{E}(E_n)$	$\mathrm{E}(E_n^2)$	$\mathrm{Var}(E_n)$
1	0	0	0
2	1	2	1
3	$\frac{3}{4}= 0.75$	$\frac{5}{4}= 1.25$	$\frac{11}{16}= 0.6875$
4	$\frac{23}{36}\approx 0.638889$	$\frac{103}{108}\approx 0.953704$	$\frac{707}{1\,296}\approx 0.545525$
5	$\frac{55}{96}\approx 0.572917$	$\frac{455}{576}\approx 0.789931$	$\frac{4\,255}{9\,216}\approx 0.461697$
6	$\frac{1901}{3600}\approx 0.528056$	$\frac{73\,897}{108\,000}\approx 0.684231$	$\frac{5\,253\,839}{12\,960\,000}\approx 0.405389$
10	≈ 0.431647	≈ 0.474437	≈ 0.288112
100	≈ 0.228368	≈ 0.133230	≈ 0.081078
200	≈ 0.198537	≈ 0.098752	≈ 0.059335
300	≈ 0.184283	≈ 0.084057	≈ 0.050097
400	≈ 0.175300	≈ 0.075413	≈ 0.044683
500	≈ 0.168891	≈ 0.069543	≈ 0.041019
∞	$\sim 1/(\log n)$	$\sim 2/\log^2 n$	$\sim 1/\log^2 n$

Aus der Asymptotik der Momente von E_n aus Theorem 4.1.5 lässt sich leicht folgende Verteilungskonvergenz zeigen.

Satz 4.1.7 *(Verteilungskonvergenz von E_n) [35] Für $n \in \mathbb{N}$ sei E_n die Länge eines zufällig ausgewählten externen Zweiges in einem Bolthausen-Sznitman-n-Coalescent. Dann gilt*

$$E_n \log n \to Exp(1)$$

in Verteilung für $n \to \infty$.

Beweis: Nach Theorem 4.1.5 gilt $\lim_{n\to\infty} \mathrm{E}((E_n \log n)^j) = j!$ für $j \in \mathbb{N}_0$. Aus der Konvergenz der Momente folgt etwa mit [14, Theoreme 30.1,30.2] die Konvergenz $E_n \log n \to Exp(1)$ in Verteilung mit $n \to \infty$, da das j-te Moment einer $Exp(1)$-verteilten Zufallsvariable gerade $j!$ ist . □

Satz 4.1.7 lässt sich auch ohne Analyse erzeugender Funktionen probabilistisch zeigen, indem man den Zusammenhang zwischen dem Bolthausen-Sznitman-Coalescent und zufälligen rekursiven Bäumen ausnutzt (siehe [42]). Der folgende Beweis stammt von C. Goldschmidt und M. Möhle (und findet sich in [35]).

Beweis:(alternativer Beweis für Satz 4.1.7) Sei Π ein Bolthausen-Sznitman-Coalescent und E_n die externe Zweiglänge eines zufällig ausgewählten Individuums im zugehörigen n-Coalescent $(\rho_n(\Pi_t))_{t\geq 0}$. Aus der Austauschbarkeit

von $(E_n^{(i)})_{i\in[n]}$ aus Satz 2.3.1 d) folgt $E_n \overset{d}{=} E_n^{(1)}$, da die Auswahl des betrachteten Zweiges unabhängig von Π ist. Betrachte $P(E_n > t) = P(E_n^{(1)} > t)$ für $t \geq 0$. Dies ist die Wahrscheinlichkeit, dass das Individuum 1 zur Zeit t noch Singleton von $\rho_n(\Pi_t)$ ist (vgl. Satz 2.3.1 b)).
Betrachte nun einen zufälligen rekursiven Baum mit n nummerierten Knoten. Setze Knoten 1 als Wurzel. Ordne jeder Kante dieses Baumes eine vom Baum unabhängige $Exp(1)$-verteilte Zufallsvariable zu, diese Zufallsvariablen sollen auch unabhängig voneinander sein. Bilde nun auf folgende Art und Weise einen $\mathbb{E}_n$-wertigen Prozess $(\mathsf{BS}_t)_{t\geq 0}$:

1) Nach Ablauf der ersten exponentialverteilten Zufallsvariable schneide die zugehörige Kante vom Baum ab und füge alle Nummern des abgeschnittenen Unterbaums zu dem Knoten hinzu, von dem dieser Unterbaum getrennt wurde und der noch an der Wurzel hängt.

2) Betrachte nur noch den Baum, der an der Wurzel hängt.

3) Nach Ablauf der nächsten exponentialverteilten Zufallsvariablen, die dem (verkleinerten) Baum zugeordnet ist, verfahre analog zu Schritt 1 (und Schritt 2).

4) Fahre mit dieser Prozedur fort, bis die Wurzel isoliert ist, d.h. bis alle Knoten zur Wurzel hinzugefügt wurden.

5) Setze BS_t für jedes $t \geq 0$ als die Partition, deren Blöcke aus den Zahlen bestehen, die der zum Zeitpunkt t betrachtete Baum aus den vorangegangenen Schritten an jeweils einem Knoten trägt.

[42, Proposition 2.2] besagt, dass $(\mathsf{BS}_t)_{t\geq 0}$ ein Bolthausen-Sznitman-n-Coalescent ist. Also ist $E_n^{(1)} \overset{d}{=} E_n$ verteilt wie die Zeit bis zum ersten Schnitt, der im zufälligen rekursiven Baum ein Kind von der Wurzel trennt. Somit ist E_n bedingt auf die Anzahl $K(1,n)$ der Kinder der Wurzel des rekursiven Baumes verteilt wie das Minimum von $K(1,n)$ unabhängigen, $Exp(1)$-verteilten Zufallsvariablen, also $Exp(K(1,n))$-verteilt. Als Formel heißt dies $E_n \overset{d}{=} T_{K(1,n)}$, $n \in \{2,3,\ldots\}$, wobei $T_1, T_2, \ldots$ Zufallsvariable mit $E_i \overset{d}{=} \mathrm{Exp}(i)$, $i \in \mathbb{N}$, sind und $K(1,n)$ eine von $T_1, \ldots, T_{n-1}$ unabhängige Zufallsvariable ist, die wie die Anzahl der Kinder der Wurzel in einem zufälligen rekursiven

Baum verteilt ist. Aus der rekursiven Konstruktion eines zufälligen rekursiven Baumes erhält man $K(1,1) \equiv 0$ und

$$K(1,n) \stackrel{d}{=} I_1 + \cdots + I_{n-1}, \quad n \in \{2,3,\ldots\},$$

wobei $I_1, I_2, \ldots$ unabhängige Bernoulli-Variable mit $P(I_k = 1) = 1/k$, $k \in \mathbb{N}$ sind. Somit hat man $\mathrm{E}(K(1,n)) = \sum_{k\in[n-1]} 1/k \sim \log n$ und $\mathrm{Var}(K(1,n)) = \sum_{k\in[n-1]}(1/k)(1-1/k) \sim \log n$ für $n \to \infty$. Die Tschebyscheff-Ungleichung zeigt

$$\begin{aligned}
& P(|K(1,n) - \log(n)| \geq \epsilon \log(n)) \\
\leq\ & \frac{1}{\log^2(n)\epsilon^2} E((K(1,n) - \log(n))^2) \\
=\ & \frac{1}{\log^2(n)\epsilon^2} (Var(K(1,n)) + (E(K(1,n)) - \log(n))^2) \\
\sim\ & \frac{1}{\log^2(n)\epsilon^2} Var(K(1,n)) \to 0
\end{aligned}$$

für $n \to \infty$, also gilt $(\log(n))^{-1}K(1,n) \to 1$ stochastisch. Benutzt man nun die Eigenschaft, dass für eine $Exp(1)$-verteilte Zufallsvariable X und $a > 0$ die Zufallsvariable $X/a \stackrel{d}{=} Exp(a)$ erfüllt, so sieht man

$$E_n \stackrel{d}{=} T_{K(1,n)} \stackrel{d}{=} \frac{T_1}{K(1,n)}.$$

Also gilt nach dem Satz von Slutsky $(\log n)E_n \stackrel{d}{=} (\log n)K(1,n)^{-1}T_1 \stackrel{d}{\to} T_1$. Dies zeigt die gewünschte Aussage. □

5 C_n^{ext}, die Anzahl der Kollisionen bis zum Ende eines zufällig ausgewählten externen Zweiges

Sei $\Pi^{(n)}$ ein n-Coalescent. Betrachte das Funktional C_n^{ext}, also die Anzahl der Kollisionen in $\Pi^{(n)}$, die bis vor dem Ende eines zufällig ausgewählten externen Zweiges stattfinden. Dieses Funktional steht in engem Zusammenhang mit der Länge E_n eines zufällig ausgewählten externen Zweiges (bei Auswahl desselben Zweiges), betrachte hierzu die Gleichung $E_n \overset{d}{=} \sum_{k=0}^{C_n^{ext}} T_{\chi_k^{(n)}}$ für $n \in \{2, 3, \ldots\}$ aus Satz 2.4.3 b). Wie in Kapitel 4.1 beschrieben, wird in [22] diese Formel benutzt, um für E_n im Kingman-n-Coalescent die Verteilungskonvergenz von nE_n für $n \to \infty$ nachzuweisen. Um dies zu erreichen wird in [22, Gl. 5] auch ein Konvergenzresultat für C_n^{ext} beweisen, im Kingman-Coalescent gilt

$$C_n^{ext}/n \xrightarrow{d} \beta(1,2)$$

für $n \to \infty$. Für den sternförmigen Coalescent dagegen ist das Funktional $C_n^{ext} \equiv 0$ trivial für alle $n \in \mathbb{N}$, da dort alle externen Zweige nach derselben $Exp(1)$-verteilten Wartezeit in einer einzigen Kollision enden. Für Simple Λ-n-Coalescents wurde in [39, Proposition 3.1] mit Hilfe der Poisson-Konstruktion $C_n^{ext} \to G(\frac{\mu_{-1}}{\mu_{-2}})$ in Verteilung für $n \to \infty$ gezeigt, wobei $G(\frac{\mu_{-1}}{\mu_{-2}})$ die geometrische Verteilung auf $\mathbb{N}_0$ mit Parameter $\frac{\mu_{-1}}{\mu_{-2}}$ ist. Der dortige Beweis wird zwar nur für eine Unterklasse der Simple Λ-Coalescents geführt, funktioniert aber auch für die gesamte Klasse dieser Λ-Coalescents.

5.1 C_n^{ext} im Bolthausen-Sznitman-n-Coalescent [35]

Um C_n^{ext} im Bolthausen-Sznitman-n-Coalescent zu analysieren, geht man analog zu den Kapiteln 3 und 4.1 vor, d.h. über einen Erzeugende-Funktionen-Ansatz. Hier wird allerdings nicht die erzeugende Funktion der j-ten Momente, sondern die erzeugende Funktion $f(s,t) := \sum_{n=2}^{\infty} \mathrm{E}(s^{C_n^{ext}})t^{n-1}$ für $s \in [0,1]$ und $t \in [0,1)$ benutzt, also eine erzeugende Funktion erzeugender Funktionen. Für diese gilt folgendes Lemma.

Lemma 5.1.8 *[35] Sei $s \in [0,1]$ und $t \in [0,1)$. Es gilt*

$$f(s,t) \;=\; \int_0^t \frac{1}{(1-u)^2}\frac{1-a(u)}{1-sa(u)}\,du, \tag{89}$$

wobei $a(\cdot)$ die in (61) definierte Funktion ist.

Beweis: Aus der Rekursion (48) für die erzeugende Funktion $s \mapsto \mathrm{E}(s^{C_n^{ext}})$ und mit $r_{nj} = n/((n-1)(n-j)(n-j+1))$, $n \in \mathbb{N}$, $j \in [n-1]$, erhält man bei Differentiation von f nach t

$$\begin{aligned}
f_t(s,t) &= \sum_{n=2}^{\infty}(n-1)\mathrm{E}(s^{C_n^{ext}})t^{n-2} \\
&= \sum_{n=2}^{\infty}(n-1)\Big(P(C_n^{ext}=0) + s\sum_{j=2}^{n-1}\frac{j-1}{n}r_{nj}\mathrm{E}(s^{C_j^{ext}})\Big)t^{n-2} \\
&= \sum_{n=2}^{\infty}(n-1)P(C_n^{ext}=0)t^{n-2} \\
&\qquad\qquad + s\sum_{j=2}^{\infty}(j-1)\mathrm{E}(s^{C_j^{ext}})t^{j-2}\sum_{n=j+1}^{\infty}\frac{t^{n-j}}{(n-j)(n-j+1)} \\
&= \sum_{n=2}^{\infty}(n-1)P(C_n^{ext}=0)t^{n-2} + sa(t)\sum_{j=2}^{\infty}(j-1)\mathrm{E}(s^{C_j^{ext}})t^{j-2} \\
&= \frac{1-a(t)}{(1-t)^2} + sa(t)f_t(s,t),
\end{aligned}$$

wobei das letzte Gleichheitszeichen aus

$$\begin{aligned}
&\sum_{n=2}^{\infty}(n-1)P(C_n^{ext}=0)t^{n-2} \\
&\qquad= \sum_{n=2}^{\infty}(n-1)\Big(1-\sum_{j=2}^{n-1}\frac{j-1}{n}r_{nj}\Big)t^{n-2} \\
&\qquad= \sum_{n=2}^{\infty}(n-1)t^{n-2} - \sum_{j=2}^{\infty}(j-1)t^{j-2}\sum_{n=j+1}^{\infty}\frac{t^{n-j}}{(n-j)(n-j+1)} \\
&\qquad= \frac{1}{(1-t)^2} - \frac{a(t)}{(1-t)^2} \;=\; \frac{1-a(t)}{(1-t)^2}
\end{aligned}$$

folgt, hier wird die Formel für $P(C_n^{ext} = 0)$ aus Satz 2.4.3 c) verwendet (und die Potenzreihendarstellungen von $(1-t)^{-2}$ und $a(t)$). Löst man das erste Gleichungssystem im Beweis nach $f_t(s,t)$ auf, so erhält man

$$f_t(s,t) \;=\; \frac{1}{(1-t)^2}\frac{1-a(t)}{1-sa(t)}.$$

Durch Integration und mit der Anfangsbedingung $f(s,0) = 0$ folgt die gewünschte Aussage. □

Aus der erzeugenden Funktion f lässt sich wiederum die Asymptotik der Momente folgern. Im Gegensatz zur Analyse der Funktionale τ_n und E_n wird hier auf eine explizite Darstellung der Momente von C_n^{ext} verzichtet, da C_n^{ext} in der Anwendung keine große Rolle spielt (siehe Kapitel 2.2).

Satz 5.1.9 *(Asymptotik der Momente von C_n^{ext}) [35] Betrachte C_n^{ext} im Bolthausen-Sznitman-n-Coalescent für $n \geq 2$. Die Momente von C_n^{ext} besitzen die asymptotische Darstellung*

$$\mathrm{E}((C_n^{ext})^k) \;=\; \frac{1}{k+1}\frac{n^k}{\log^k n}\Big(1+\frac{k\Psi^{(0)}(k+2)}{\log n}+O\Big(\frac{1}{\log^2 n}\Big)\Big), \quad k \in \mathbb{N},$$

wobei $\Psi^{(0)}$ die Psi-Funktion, also die logarithmische Ableitung der Gammafunktion ist. Insbesondere gilt $\mathrm{E}(C_n^{ext}) \sim n/(2\log n)$ und $\mathrm{Var}(C_n^{ext}) \sim n^2/(12\log^2 n)$.

Beweis: Differenziert man Darstellung (89) der erzeugenden Funktion $f = f(s,t)$ aus dem vorigen Lemma k-fach nach s und vertauscht (k-fache) Differentiation und Integration, so erhält man

$$\begin{aligned}\Big(\frac{\partial}{\partial s}\Big)^k f(s,t) \;&=\; \int_0^t \frac{1-a(u)}{(1-u)^2}\Big(\frac{\partial}{\partial s}\Big)^k \frac{1}{1-sa(u)}\,du \\ &=\; k!\int_0^t \frac{1-a(u)}{(1-u)^2}\frac{(a(u))^k}{(1-sa(u))^{k+1}}\,du, \quad 0 < s,t < 1. \qquad (90)\end{aligned}$$

Das Vertauschen von Integration und Differentiation ist nach dem Lebesgue'schen Differentiationslemma erlaubt, da für festes $k \in \mathbb{N}$ und $0 < t < 1$ die Funktion $f_k : (0,1) \times (0,t) \to \mathbb{R}$, gegeben durch

$$f_k(s,u) \;:=\; \frac{1-a(u)}{(1-u)^2}\frac{(a(u))^k}{(1-sa(u))^{k+1}},$$

die Bedingung $|f_k(s,u)| \leq 1/((1-u)^2(1-s)^{k+1})$ für $0 < s < 1$ und $0 < u < t$ erfüllt (es gilt $0 < a(u) < 1$ für $0 < u < 1$) und die majorisierende Funktion $u \mapsto c/(1-u)^2$ (für $c > 0$ geeignet) integrierbar bezüglich des Lebesgue-Maßes auf $(0,t)$ ist. Da sich $f(s,t)$ für festes t auch als Potenzreihe in s auffassen lässt, kann man $\left(\frac{\partial}{\partial s}\right)^k f(s,t)$ auch durch gliedweises k-faches Differenzieren berechnen. Macht man dies und verwendet die Differentiationregeln für wahrscheinlichkeitserzeugende Funktionen, so erhält man

$$\left(\frac{\partial}{\partial s}\right)^k f(s,t) + \sum_{n=2}^{\infty} E((C_n^{ext})_k \cdot s^{C_n^{ext}-k})t^{n-1}, \tag{91}$$

wobei $(x)_k = x(x-1)\cdots(x-k+1)$ das absteigende faktorielle Produkt für $x \in \mathbb{R}$ und $k \in \mathbb{N}$ ist. Lässt man $s \nearrow 1$ in (90) und (91), so erhält man

$$\sum_{n=2}^{\infty} \mathrm{E}((C_n^{ext})_k)t^{n-1} \;=\; k! \int_0^t \frac{1}{(1-u)^2}\left(\frac{a(u)}{1-a(u)}\right)^k du,$$

wobei man Summe und Grenzwert hier nach dem Satz von Lebesgue für Integrale/Summen vertauschen darf. Differenziert man diese Gleichung in t und multipliziert dann mit t^2, so sieht man

$$f_k(t) \;:=\; \sum_{n=2}^{\infty}(n-1)\mathrm{E}((C_n^{ext})_k)t^n \;=\; k!\left(\frac{t}{1-t}\right)^2\left(\frac{a(t)}{1-a(t)}\right)^k.$$

Setzt man hier

$$\frac{a(t)}{1-a(t)} \;=\; \frac{t}{(1-t)(-\log(1-t))} - 1$$

ein und verwendet den binomischen Lehrsatz, so folgt

$$\begin{aligned} f_k(t) \;&=\; k!\left(\frac{t}{1-t}\right)^2 \sum_{i=0}^{k} \binom{k}{i}\left(\frac{t}{(1-t)(-\log(1-t))}\right)^i (-1)^{k-i} \\ &=\; k! \sum_{i=0}^{k} \binom{k}{i}(-1)^{k-i}\frac{t^{i+2}}{(1-t)^{i+2}(-\log(1-t))^i}. \end{aligned}$$

Um aus f_k die (faktoriellen) Momente, also die Koeffizienten zu bestimmen, werden wie im Beweis von Satz 4.1.5 die Formeln aus [70] verwendet. Wende nun (85) an, dies zeigt

$$\begin{aligned}
(n-1)\mathrm{E}((C_n^{ext})_k) &= [t^n]f_k(t) \\
&= k!\sum_{i=0}^{k}\binom{k}{i}(-1)^{k-i}[t^{n-i-2}]\frac{1}{(1-t)^{i+2}(-\log(1-t))^i} \\
&= k!\sum_{i=0}^{k}\binom{k}{i}(-1)^{k-i}\frac{n^{i+1}}{\Gamma(i+2)\log^i n}\Big(1+\frac{i\Psi^{(0)}(i+2)}{\log n}+O\Big(\frac{1}{\log^2 n}\Big)\Big) \\
&= \frac{1}{k+1}\frac{n^{k+1}}{\log^k n}\Big(1+\frac{k\Psi^{(0)}(k+2)}{\log n}+O\Big(\frac{1}{\log^2 n}\Big)\Big),
\end{aligned}$$

für $n \to \infty$, wobei man für das letzte Gleichheitszeichen alle „störenden" Terme in den O-Term packt. Das Ergebnis wird mit $1/(n-1)$ multipliziert, also

$$\mathrm{E}((C_n^{ext})_k) = \frac{1}{k+1}\frac{n^k}{\log^k n}\Big(1+\frac{k\Psi^{(0)}(k+2)}{\log n}+O\Big(\frac{1}{\log^2 n}\Big)\Big)$$

für $n \to \infty$. Es gilt

$$\mathrm{E}((C_n^{ext})^k) = \sum_{j=0}^{k} S(k,j)\mathrm{E}((C_n^{ext})_j), \tag{92}$$

wobei $S(k,j)$ die Stirling-Zahl zweiter Art mit Parametern k,j bezeichnet. Die Aussage des Satzes folgt nun aus (92), wenn man alle Summanden aus (92) bis auf den k-ten Summanden in den O-Term zusammenfasst und $S(k,k) = 1$ einsetzt. Die Darstellung der Varianz erhält man analog zum Beweis von Satz 4.1.5 aus der asymptotischen Darstellung der Momente. □

Aus der asymptotischen Darstellung der Momente lässt sich ein Konvergenzresultat in Verteilung folgern, da bei entsprechender Skalierung die Momente von C_n^{ext} für $n \to \infty$ konvergieren.

Satz 5.1.10 *(Verteilungskonvergenz von skaliertem C_n^{ext})[35] Betrachte C_n^{ext} in einem Bolthausen-Sznitman-n-Coalescent für $n \in \mathbb{N}$. Für $n \to \infty$ gilt*

$$\frac{\log n}{n}C_n^{ext} \xrightarrow{d} U_{[0,1]},$$

wobei $U_{[0,1]}$ die Gleichverteilung auf $[0,1]$ ist.

Beweis: Aus Satz 5.1.9 erhält man

$$\lim_{n\to\infty} \mathrm{E}\Big(\Big(\frac{\log n}{n} C_n^{ext}\Big)^k\Big) \;=\; \lim_{n\to\infty} \frac{\log^k n}{n^k} \mathrm{E}((C_n^{ext})^k) \;=\; \frac{1}{k+1} \;=\; \mathrm{E}(U^k), \quad k \in \mathbb{N},$$

falls U eine Zufallsvariable mit $U \stackrel{d}{=} U_{[0,1]}$ ist. Aus der Konvergenz der Momente folgt wieder mit [14, Theoreme 30.1, 30.2] auch die Konvergenz in Verteilung. □

6 K_n, die Anzahl der Typen in einer n-elementigen Stichprobe

Für $n \in \mathbb{N}$ betrachte eine Stichprobe von n Individuen, deren Genealogie durch einen Ξ-n-Coalescent mit Mutation unter dem Infinitely-Many-Alleles-Modell beschrieben wird. In diesem Kapitel wird das Funktional K_n, die Anzahl der verschiedenen Typen, die in der Stichprobe vorkommen, untersucht. Hier ist besonders die Asymptotik von $(K_n)_{n\in\mathbb{N}}$ für $n \to \infty$ von Interesse. In Abschnitt 2.4 wurde die Verteilungsrekursion (49) für $(K_n)_{n\in\mathbb{N}}$ bewiesen. Diese Rekursion ermöglicht es, für festes $n \in \mathbb{N}$ die Verteilung von K_n explizit zu bestimmen. Da die Rekursion recht kompliziert ist, lässt sich nur in Spezialfällen eine einfache, geschlossene Formel für die Verteilung von K_n angeben. Im Falle des Kingman-Coalescents ($\Lambda = \delta_0$) und des sternförmigen Coalescents (star-shaped coalescent, $\Lambda = \delta_1$) lassen sich solche Formeln finden. Diese ermöglichen es auch, die Asymptotik von $(K_n)_{n\in\mathbb{N}}$ für $n \to \infty$ zu bestimmen.

Im gesamten Kapitel 6 wird die feste Stichprobe $[n]$ betrachtet, deren Genealogie durch einen n-Coalescent mit Mutation gegeben ist.

6.1 K_n im Kingman-/sternförmigen n-Coalescent mit Mutation [36]

Betrachte zunächst die Anzahl K_n der Typen in der Stichprobe $[n]$, wenn der Stammbaum/die Genealogie der Stichprobe durch einen Kingman-n-Coalescent mit Mutation gegeben ist. In Bemerkung 2.4.9 wurde bereits darauf hingewiesen, dass sich die Rekursion (49) aus der Rekursion für das vollständige Allelfrequenzspektrum $(K_n(1), \ldots, K_n(n))$ (moehle's recursion) gewinnen lässt. Im Falle des Kingman-n-Coalescents kann man sogar aus der Rekursion für $(K_n(1), \ldots, K_n(n))$ eine geschlossene Formel für die Verteilung von $(K_n(1), \ldots, K_n(n))$ für festes $n \in \mathbb{N}$ folgern. Diese explizite Formel ist die Ewens-Sampling-Formel (siehe [31]), ein Grundpfeiler der mathematischen Populationsgenetik. Sie besagt für $(K_n(1), \ldots, K_n(n))$ im Kingman-n-Coalescent mit Mutation und Mutationsrate $r > 0$, dass

$$P(K_n(1) = k_1, \ldots, K_n(n) = k_n) = \frac{n!}{[2r]_n} \prod_{i\in[n]} \left(\frac{2r}{i}\right)^{k_i} \frac{1}{k_i!}$$

für $k_1, \ldots, k_n \in \mathbb{N}_0$ mit $\sum_{i\in[n]} ik_i = n$ und $[x]_l := x(x+1)\cdots(x+l-1)$ das aufsteigende faktorielle Produkt mit Parametern $x \in \mathbb{R}$ und $l \in \mathbb{N}$. Aus der Ewens-Sampling-Formel lässt sich über den Zusammenhang $K_n = \sum_{i\in[n]} K_n(i)$ die Verteilung von K_n für alle $n \in \mathbb{N}$ bestimmen und daraus die Verteilungsasymptotik von $(K_n)_{n\in\infty}$ für $n \to \infty$. Allerdings ermöglicht es Rekursion (49), diese Asymptotik auch direkt ohne den Zusammenhang zum vollständigen Allelfrequenzspektrum zu bestimmen.

Satz 6.1.1 *(Anzahl der Typen im Kingman-n-Coalescent)[36] Für $n \in \mathbb{N}$ betrachte einen Kingman n-Coalescent mit Mutation (Mutationsrate $r > 0$). Für die Anzahl K_n der Typen in der Stichprobe $[n]$ gilt*

a) $K_n \stackrel{d}{=} \sum_{i\in[n]} B_i$ *für unabhängige* $B_1, B_2, \ldots$ *mit* $B_i \stackrel{d}{=} Bin(1, \frac{2r}{2r+i-1})$, $i \in \mathbb{N}$.

b) $P(K_n = k) = (2r)^k s(n,k)/[2r]_n$, $k \in [n]$, *wobei* $[2r]_n := 2r(2r+1)\cdots(2r+n-1)$ *das aufsteigende Produkt und* $s(n,k)$ *die absolute Stirlingzahl erster Art mit Parametern* k *und* n *bezeichnet, also die Anzahl der Permutationen von* $[n]$ *mit genau* k *Zykeln.*

c)
$$\begin{aligned} \mathrm{E}(K_n) &= 2r\sum_{i=0}^{n-1}(2r+i)^{-1} \sim 2r\log n \text{ und} \\ \mathrm{Var}(K_n) &= 2r\sum_{i=1}^{n-1} i(2r+i)^{-2} \sim 2r\log n. \end{aligned}$$

d) $(K_n - 2r\log n)/\sqrt{2r\log n} \stackrel{d}{\to} N(0,1)$.

Beweis: Im Kingman-n-Coalescent kollidieren bei jeder Kollision genau 2 Blöcke und es gibt keine simultanen Kollisionen. Somit gilt in der Rekursion (49) $P(I_n = n-1) = 1$, $g_n = g_{n,n-1} = \binom{n}{2}$ und $g_{ni} = 0$ für $i \in [n-2]$. Dies vereinfacht Rekursion (49) zu

$$K_n \stackrel{d}{=} B_n(K_{n-1}+1) + (1-B_n)K_{n-1} \stackrel{d}{=} B_n + K_{n-1}$$

für $n \in \mathbb{N}$, wobei B_n von K_{n-1} unabhängig ist und $B_n \stackrel{d}{=} Bin(1, \frac{nr}{g_n+nr}) = Bin(1, \frac{2r}{2r+n-1})$. Aus dieser Rekursion folgt a). Für $k \in [n]$ erhält man aus a)

$$\begin{aligned} P(K_n = k) &= P(\sum_{i\in[n]} B_i = k) = \sum_{T\subseteq[n],|T|=k} \prod_{i\in T} P(B_i = 1) \prod_{i\notin T} P(B_i = 0) \\ &= \sum_{T\subseteq[n],|T|=k} \frac{(2r)^k \prod_{i\notin T}(i-1)}{[2r]_n} \\ &= \frac{(2r)^k}{[2r]_n} \sum_{T\subseteq[n],|T|=k} \prod_{i\notin T}(i-1) = \frac{(2r)^k}{[2r]_n} s(n,k), \end{aligned}$$

da der Koeffizient von x^k in

$$(x)_n = \prod_{i\in[n]}(x-(i-1)) = \sum_{l=0}^{n}(-1)^{n-l} s(n,l) x^l$$

einerseits $(-1)^{n-k}s(n,k)$, andererseits

$$\sum_{T\subseteq[n],|T|=k} \prod_{i\notin T}(-(i-1)) = (-1)^{n-k} \sum_{T\subseteq[n],|T|=k} \prod_{i\notin T}(i-1)$$

ist. c) folgt sofort aus a), die Approximationen für $n \to \infty$ sind Standard. Da K_n eine Summe unabhängiger Bernoulli-Variablen ist und nach c) $Var(K_n) \to \infty$ für $n \to \infty$ gilt, ist die Lindebergbedingung erfüllt und es gilt

$$(K_n - E(K_n))/\sqrt{Var(K_n)} \xrightarrow{d} N(0,1)$$

für $n \to \infty$. Daraus folgt d) mit Hilfe der asymptotischen Abschätzungen für Erwartungswert und Varianz aus c). □

Bemerkung 6.1.2 *Die Aussagen über $(K_n)_{n\in\mathbb{N}}$ aus dem vorangegangenen Satz 6.1.1 sind bekannt und lassen sich auch in der klassischen Arbeit [31] von Ewens finden.*

Für den sternförmigen n-Coalescent, $n \geq 2$, erhält man für die in Rekursion (49) (und allen abgeleiteten Rekursionen) vorkommenden Größen I_n, g_n und g_{nk} ($k \in [n-1]$) die Werte $P(I_n = 1) = 1$ und somit $g_{n1} = g_n = 1$ und $g_{nk} = 0$ für $k \in \{2, \ldots, n-1\}$. Somit vereinfacht sich z.B. Rekursion (52) zu

$$(1+nr)f_n(s) = nrsf_{n-1}(s) + s, \; n \in \{2,3,\ldots\}, \; s \in \mathbb{C},$$

und die Rekursion (56) zu $E(K_1) = 1$ und

$$E(K_n) = \frac{nr}{1+nr}(E(K_{n-1}) + 1) + \frac{E(K_1)}{1+nr} = 1 + \frac{nrE(K_{n-1})}{1+nr},\ n \in \{2, 3, \ldots\}.$$

Aus letzterer Rekursion lässt sich per Induktion die geschlossene Formel $E(K_n) = nr/(1+r)+1-\prod_{i\in[n]} ir/(1+ir)$ für $n \in \mathbb{N}$ ableiten. Diese und weitere Ergebnisse finden sich in [67, Abschnitt 4]. Insbesondere wird dort unter Benutzung eines geeigneten Martingals gezeigt, dass K_n/n im sternförmigen n-Coalescent für $n \to \infty$ fast sicher gegen eine $\beta(1, 1/r)$-verteilte Zufallsvariable konvergiert. Diese Konvergenz lässt sich auch mittels anderer Methoden beweisen, wie in Abschnitt 6.5 gezeigt wird.

6.2 K_n in Beta-n-Coalescents

Für bestimmte Beta-n-Coalescents mit Mutation (Mutationsrate $r > 0$), nämlich die $\beta(2-\alpha, \alpha)$-n-Coalescents für $1 < \alpha < 2$, wurde in [8] (und [9]) das vollständige Allel-Frequenzspektrum $(K_n(1), \ldots, K_n(n))$ untersucht. Die Klasse der $\beta(2-\alpha, \alpha)$-Coalescents mit $1 < \alpha < 2$ ist staubfrei (Beispiel 1.1.19). Ein solcher $\beta(2-\alpha, \alpha)$-Coalescent lässt sich in der Genealogie eines α-stabilen Continuous-State-Verzweigungsprozesses (CSBP) finden (siehe [16, Theorem 2.1, Bemerkung 2.2]) und ist auch in einen α-stabilen Continuous Random Tree (CRT) einbettbar (siehe [9, Theorem 1]). Diese Zusammenhänge ermöglichen es, ein schwaches Gesetz der großen Zahlen für $(K_n)_{n\in\mathbb{N}}$ aufzustellen. Nach [9, Theorem 9] gilt

$$\frac{K_n(i)}{n^{2-\alpha}} \xrightarrow{p} r\frac{\alpha(\alpha-1)^2\Gamma(i+\alpha-2)}{i!}$$

für $k \in \mathbb{N}$ und $n \to \infty$. Im Beweis dieses Theorems ergibt sich auch

$$\frac{K_n}{n^{2-\alpha}} \xrightarrow{p} r\frac{\alpha(\alpha-1)\Gamma(\alpha)}{2-\alpha}$$

für $n \to \infty$ (vergleiche hier auch [8, Theorem 1.9] und die diesem Theorem vorangestellte Erläuterung).

Der ,,Grenzfall" der hier betrachteten Familie von β-Verteilungen für $\alpha \to 1$ ist $\beta(1, 1)$, also die Gleichverteilung auf $[0, 1]$. Der zugehörige Λ-n-Coalescent ist der Bolthausen-Sznitman-n-Coalescent. Für diesen muss K_n (und das vollständige Allel-Frequenz-Spektrum) mit anderen Methoden untersucht werden; dies wird in Kapitel 6.6 beschrieben.

6.3 Mutierte Zweige in Coalescents mit Staub und Mutation [34], [36]

In der Klasse der Coalescents mit Staub und Mutation ist es abgesehen vom sternförmigen Coalescent weitaus komplizierter, mit der Rekursion (49) umzugehen. Um das asymptotische Verhalten von K_n für $n \to \infty$ zu untersuchen, sind Methoden, die die Rekursion (49) nicht verwenden, angenehmer zu handhaben.

Im Folgendem werden zwei Möglichkeiten vorgestellt, die Asymptotik von K_n zu untersuchen. Beide analysieren das Funktional M_n des n-Coalescents aus Definiton 2.1.2, das einfacher zu handhaben ist und „nahe" an K_n liegt für große $n \in \mathbb{N}$. Hier sei nochmals die Definition der Funktionale M_n (und N_n) gegeben.

Definition 6.3.1 *(Anzahl der mutierten (externen) Zweige) Sei* $(\Pi^{(n)}, \mathsf{Mu}^{(n)})$ *ein n-Coalescent mit Mutation. Betrachte das Infinitely-Many-Sites- oder das Infinitely-Many-Alleles-Modell. Ein Zweig des von* $\Pi^{(n)}$ *gegebenen Baumes heißt mutiert, falls mindestens eine Mutation von* $\mathsf{Mu}^{(n)}$ *auf diesem Zweig liegt. Setze*

- N_n *als die Anzahl der mutierten Zweige des von* $\Pi^{(n)}$ *gegebenen Baumes.*

- M_n *als die Anzahl der externen mutierten Zweige des von* $\Pi^{(n)}$ *gegebenen Baumes.*

Bemerkung 6.3.2 *Sei* Π *ein* Ξ*-Coalescent mit pfadweise konsistenter Mutationsstruktur gegeben durch eine Familie von Poisson-Punktprozessen* $(P_i)_{i \in \mathbb{N}}$ *auf* $[0, \infty)$*. Dann gilt für* M_n*, die Anzahl der mutierten externen Zweige des durch* $(\rho_n(\Pi_t))_{t \geq 0}$ *gegebenen Baumes,*

$$M_n = \sum_{i \in [n]} 1_{\left\{|P_i \cap [0, E_n^{(i)})| > 0\right\}}. \tag{93}$$

Sei Ξ ein endliches Maß auf dem unendlichdimensionalen Simplex Δ. Für $n \in \mathbb{N}$ sei zunächst $(\Pi^{(n)}, \mathsf{Mu}^{(n)})$ ein Ξ-n-Coalescent mit Mutation und Mutationsrate $r > 0$. Die Mutationen werden hinsichtlich des Infinitely-Many-Alleles-Modell interpretiert. Betrachte nun M_n im durch $(\Pi^{(n)}, \mathsf{Mu}^{(n)})$ gegebenen Baum mit Mutation. Erfüllt Ξ Bedingung (6), ist also $\Pi^{(n)}$ ein Ξ-n-Coalescent mit Staub, so kann man die Asymptotik der Verteilungen von $(M_n)_{n \in \mathbb{N}}$ für $n \to \infty$ über die Momentenmethode bestimmen.

Satz 6.3.3 *[36] Sei Ξ ein endliches Maß auf Δ, das (6) erfüllt. Für $n \in \mathbb{N}$ sei $(\Pi^{(n)}, \mathsf{Mu}^{(n)})$ ein Ξ-n-Coalescent mit Mutation und Mutationsrate $r > 0$. Dann gilt $M_n/n \xrightarrow{d} M$ für $n \to \infty$ für eine reelle Zufallsvariable M. M ist eindeutig durch die Momente*

$$\mathrm{E}(M^k) \;=\; \mathrm{E}\Big(\prod_{i=1}^{k}(1 - e^{-rE^{(i)}})\Big), \quad k \in \mathbb{N},$$

bestimmt, wobei $E^{(i)}$ für $i \in \mathbb{N}$ die Länge des i-ten externen Zweiges in einem Ξ-Coalescent Π ist.

Beweis: Sei $n \in \mathbb{N}$. Die Mutationen auf den externen Zweigen des durch $\Pi^{(n)}$ gegebenen Baumes werden bedingt $\Pi^{(n)}$ durch den von $\Pi^{(n)}$ unabhängigen homogenen Poisson-Punktprozess $\mathsf{Mu}^{(n)}$ mit Rate $r > 0$ gesetzt. Bedingt auf die Längen $(E_n^{(i)})_{i\in[n]}$ der externen Zweige der Individuen $i \in [n]$ sind die Mutationen auf verschiedenen Zweigen somit unabhängig und auf einem Zweig der Länge $E_n^{(i)}$ befindet sich mit Wahrscheinlichkeit $e^{-rE_n^{(i)}}$ keine Mutation. Für festes $j \in \mathbb{N}$ gilt für die (Gegen-)Ereignisse $B_{n,i}$:=„externer Zweig i in $\Pi^{(n)}$ ist mutiert"

$$\begin{aligned} P(B_{n,1} \cap \cdots \cap B_{n,j}) &= E(P(B_{n,1} \cap \cdots \cap B_{n,j} | E_n^{(1)}, \ldots, E_n^{(j)})) \\ &= E(P(B_{n,1}|E_n^{(1)}) \cdots P(B_{n,j}|E_n^{(j)})) \\ &= E((1 - e^{-rE_n^{(1)}}) \cdots (1 - e^{-rE_n^{(j)}})) \\ &\to E((1 - e^{-rE^{(1)}}) \cdots (1 - e^{-rE^{(j)}})) \end{aligned} \tag{94}$$

für $n \to \infty$, da nach Satz 2.3.1 a) die externen Zweiglängen in $\Pi^{(n)}$ gegen die externen Zweiglängen in Π konvergieren (Grenzwertbildung und Erwartungswert dürfen nach dem Satz von der majorisierten Konvergenz mit Majorante 1 vertauscht werden). Mit $M_n = \sum_{i\in[n]} 1_{B_{n,i}}$ erhält man

$$\mathrm{E}(M_n^k) \;=\; \mathrm{E}\Big(\Big(\sum_{i\in[n]} 1_{B_{n,i}}\Big)^k\Big) \;=\; \sum_{i_1,\ldots,i_k\in[n]} \mathrm{E}(1_{B_{n,i_1}} \cdots 1_{B_{n,i_k}}). \tag{95}$$

Für festes $n \in \mathbb{N}$ sind die Ereignisse $(B_{n,i})_{i\in[n]}$ austauschbar, dies folgt aus der Austauschbarkeit von $(E_n^{(i)})_{i\in[n]}$ (siehe Satz 2.3.1 c)). Somit sind all diejenigen Summanden $E(1_{B_{n,i_1}} \cdots 1_{B_{n,i_k}})$ aus (95) gleich $P(B_{n,1}\cap\cdots\cap B_{n,j})$, die als

Erwartungswert eines (n-fachen) Produkts von genau j verschiedenen Indikatoren entstehen. Zähle diese gleichen Summanden ab. Es gibt genau $S(n,j)$ Möglichkeiten, welche j nicht leeren Teilmengen von Indikatoren im n-fachen Produkt jeweils identisch sein sollen ($S(n,j)$ ist die Stirling Zahl zweiter Art zu den Parametern n und j) und es gibt $(n)_j := n(n-1)\cdots(n-j+1)$ Möglichkeiten, welche verschiedenen Indikatoren aus $(1_{B_{n,i}})_{i\in[n]}$ als Faktor jeweils einer dieser j Teilmengen zugeordnet werden. Fasst man nun die identischen Summanden aus (95) zusammen, so ergibt sich

$$\mathrm{E}(M_n^k) \;=\; \sum_{j\in[k]} S(k,j)(n)_j P(B_{n,1}\cap\cdots\cap B_{n,j}).$$

Teilt man diesen Ausdruck durch n^k und betrachtet man den Grenzübergang $n\to\infty$, erhält man für alle $k\in\mathbb{N}_0$

$$\begin{aligned}\lim_{n\to\infty}\mathrm{E}\Big(\Big(\frac{M_n}{n}\Big)^k\Big) &= \sum_{j\in[k]} S(k,j)\lim_{n\to\infty}\frac{(n)_j}{n^k}P(B_{n,1}\cap\cdots\cap B_{n,j})\\ &= \lim_{n\to\infty} P(B_{n,1}\cap\cdots\cap B_{n,k})\\ &= E((1-e^{-rE^{(1)}})\cdots(1-e^{-rE^{(k)}})) \;=:\; \mu_k.\end{aligned}$$

Für die Folge $(\mu_k)_{k\in\mathbb{N}_0}$ gilt für jedes $m\in\mathbb{N}_0$

$$\begin{aligned}\sum_{j=0}^{m}\binom{m}{j}(-1)^j\mu_{k+j} &= \lim_{n\to\infty}\mathrm{E}\left(\sum_{j=0}^{m}\binom{m}{j}(-1)^j\Big(\frac{M_n}{n}\Big)^{k+j}\right)\\ &= \lim_{n\to\infty}\mathrm{E}\left(\Big(\frac{M_n}{n}\Big)^k\Big(1-\frac{M_n}{n}\Big)^m\right) \;\geq\; 0.\end{aligned}$$

Somit ist hier das Hausdorff'sche Momentenproblem lösbar, d.h. es gibt eine Zufallsvariable M mit eindeutig bestimmter Verteilung, $M\in[0,1]$ fast sicher und Momenten $E(M^k)=\mu_k$, $k\in\mathbb{N}_0$. Es gilt $E((M_n/n)^j)\to E(M^j)$ für $n\to\infty$ für alle $j\in\mathbb{N}_0$. Aus der Momentenkonvergenz folgt $M_n/n\xrightarrow{d} M$, da auch $M_n/n\in[0,1]$ fast sicher für alle $n\in\mathbb{N}$ gilt. □

Nach Satz 2.3.1 gibt es Zusammenhänge zwischen dem Anteil S_t der Singletons in Π_t für $t\geq 0$ und den externen Zweiglängen. Dies ermöglicht es, die Grenzvariable M aus Satz 6.3.3 genauer zu beschreiben.

Korollar 6.3.4 *[36] Die Zufallsvariable M aus Satz 6.3.3 erfüllt*

$$M \stackrel{d}{=} \int_0^\infty re^{-rt} S_t \, dt.$$

Beweis: Sei $k \in \mathbb{N}$ und $g : \mathbb{R}^k \to \mathbb{R}$ definiert durch $g(t) := (-1)^k e^{-r(t_1+\cdots+t_k)}$ für $t = (t_1, \ldots, t_k) \in \mathbb{R}^k$. Für $x, y \in \mathbb{R}^k$ wird die Notation $x \leq y$ ($x < y$) für $x_i \leq y_i$ ($x_i < y_i$) $\forall\ i \in [k]$ verwendet. Setze

$$h(t) := \frac{\partial^k}{\partial t_1 \cdots \partial t_k} g(t) = r^k e^{-r(t_1+\cdots+t_k)} > 0.$$

Für $x = (x_1, \ldots, x_k), y = (y_1, \ldots, y_k) \in \mathbb{R}^k$ mit $x \leq y$ setze

$$\Delta_x^y g \ := \sum_{\varepsilon_1,\ldots,\varepsilon_k \in \{0,1\}} (-1)^{\varepsilon_1+\cdots+\varepsilon_k} g(\varepsilon_1 x_1 + (1-\varepsilon_1)y_1, \ldots, \varepsilon_k x_k + (1-\varepsilon_k)y_k).$$

Mit $E := (E^{(1)}, \ldots, E^{(k)})$ erhält man

$$\begin{aligned} \mathrm{E}(M^k) &= \mathrm{E}((1-e^{-rE^{(1)}}) \cdots (1-e^{-rE^{(k)}})) \ = \ \mathrm{E}(\Delta_0^E g) \\ &= \int_{\mathbb{R}_+^k} \Delta_0^y g \, P_E(dy) \ = \ \int_{\mathbb{R}_+^k} \int_{\mathbb{R}_+^k} 1_{[0,y)}(t) h(t) \, \lambda^k(dt) \, P_E(dy). \end{aligned}$$

Durch mehrfaches Anwenden des Satzes von Fubini folgt

$$\begin{aligned} \mathrm{E}(M^k) &= \int_{\mathbb{R}_+^k} h(t) \int_{\mathbb{R}_+^k} 1_{(t,\infty)}(y) P_E(dy) \, \lambda^k(dt) \ = \ \int_{\mathbb{R}_+^k} h(t) P(E > t) \, \lambda^k(dt) \\ &\stackrel{(*)}{=} \int_{\mathbb{R}_+^k} r^k e^{-r(t_1+\cdots+t_k)} \mathrm{E}(S_{t_1} \cdots S_{t_k}) \, \lambda^k(dt_1, \ldots, dt_k) \\ &= \mathrm{E}\left(\left(\int_0^\infty re^{-rt} S_t \, dt \right)^k \right), \end{aligned}$$

wobei $(*)$ mit Hilfe von Satz 2.3.1 f) folgt. Somit stimmen die Momente der Zufallsvariablen M und $\int_0^\infty re^{-rt} S_t \, dt$ überein. Da beide Zufallsvariablen fast sicher nur Werte in $[0,1]$ annehmen, sind ihre Verteilungen gleich. $\square$

Bemerkung 6.3.5 *Analog zu Satz 6.3.3 lassen sich auch die Asymptotik der Länge L_n^{ext} aller externen Zweige und die der Anzahl Seg_n^{ext} aller Mutationen auf allen externen Zweigen für Ξ-Coalescents mit Staub bestimmen. Man erhält analog für diese Funktionale, dass bei Skalierung mit n^{-1} schwache Konvergenz gegen $\int_0^\infty f(t) S_t dt$ vorliegt, wobei man $f(t) = 1$ respektive $f(t) = r$ setzt. Diese Resultate finden sich in [68, S. 2166, S. 2168].*

Das Konvergenzresultat aus Korollar 6.3.4 lässt sich durch andere Methoden sogar für beliebige Ξ-Coalescents in einer stärkeren Form beweisen. Sei dazu Π ein Ξ-Coalescent mit Staub und Mutation. Für jedes $n \in \mathbb{N}$ betrachte $\Pi^{(n)} := (\rho_n(\Pi_t))_{t\geq 0}$, die Mutationen auf dem durch $\Pi^{(n)}$ gegebenen Baum werden wie in Kapitel 1.7 pfadweise konsistent wie in Π gesetzt. Betrachte nun M_n in diesem n-Coalescent mit Mutation.

Satz 6.3.6 *[34] Sei Π ein Ξ-Coalescent mit Staub und Mutation mit Mutationsrate $r > 0$. Für $n \in \mathbb{N}$ betrachte M_n im n-Coalescent $\Pi^{(n)} = (\rho_n(\Pi_t))_{t\geq 0}$ mit Mutation. Es gilt*

$$\frac{M_n}{n} \longrightarrow \int_0^\infty re^{-rt}S_t dt \quad \textit{fast sicher und in } L^p \ (p \geq 1) \textit{ für } n \to \infty. \tag{96}$$

Um diesen Satz zu beweisen verwendet man das starke Gesetz der großen Zahlen für gewichtete i.i.d. Zufallsvariablen aus [25].

Satz 6.3.7 *Seien $(A_{in})_{i\in[n],n\in\mathbb{N}}$, S reelle Zufallsvariablen mit $\sup_{i\in[n],n\in\mathbb{N}} |A_{in}| < \infty$ fast sicher und $\sum_{i\in[n]} A_{in}/n \to S$ fast sicher für $n \to \infty$. Seien $X_1, X_2, \ldots$ i.i.d. integrierbare Zufallsvariablen unabhängig von $(A_{in})_{i\in[n],n\in\mathbb{N}}$. Dann gilt*

$$\frac{1}{n}\sum_{i\in[n]} A_{in}X_i \longrightarrow S \cdot E(X_1) \quad \textit{fast sicher für } n \to \infty.$$

Beweis: Theorem 1.1 und Bemerkung (v) aus [25] zeigen

$$\frac{1}{n}\sum_{i\in[n]} A_{in}(X_i - E(X_i)) \to 0 \text{ fast sicher für } n \to \infty.$$

Die Aussage folgt, da

$$\frac{1}{n}\sum_{i\in[n]} A_{in}E(X_i) = E(X_1)\frac{1}{n}\sum_{i\in[n]} A_{in} \to E(X_1)S \quad \text{fast sicher für } n \to \infty. \quad \square$$

Beweis:(von Satz 6.3.6) Nach Bemerkung 1.7.1 kann man Π als Coalescent mit pfadweise konsistenter Mutationsstruktur gegeben durch homogene i.i.d. Poisson-Punktprozess $(P_i)_{i\in\mathbb{N}}$ auf $[0, \infty)$ mit Intensitätsrate $r > 0$ auffassen (siehe Bemerkung 1.7.1). Somit haben die Mutationen auf dem i-ten externen Zweig von $\Pi^{(n)}$ genau die Positionen der Punkte von P_i auf $[0, E_n^{(i)})$.

Betrachte die Anzahl $M_n^{(t)}$ der externen Zweige, die bis zum Zeitpunkt $t > 0$ (ausschließlich t) mindestens eine Mutation tragen. Zeige zunächst $M_n^{(t)}/n \overset{f.s.}{\to} \int_0^t re^{-ru}S_u du$ für $n \to \infty$. Sei dazu

- $t_j := jt/k$ für $0 \leq j \leq k$,
- $A_{inj} := 1_{\left\{t_{j-1} \leq E_i^{(n)} < t_j\right\}}$ für $n \in \mathbb{N}, i \in [n], j \in [k-1]$,
- $A_{ink} := 1_{\left\{t_{k-1} \leq E_i^{(n)}\right\}}$ für $n \in \mathbb{N}, i \in [n]$ und
- $Y_i^{(j)} := 1_{\{|P_i \cap [0,t_j)|>0\}}$ für $i \in \mathbb{N}, 0 \leq j \leq k$.

$Y_i^{(j)}$ zeigt an, ob sich mindestens ein Punkt (eine Mutation) von P_i auf $[0, t_j)$ befindet. Für jedes feste $j \in \{1, \ldots, k\}$ ist $(Y_i^{(j)})_{i\in\mathbb{N}}$ i.i.d. mit

$$E(Y_i^{(j)}) = P(|P_i \cap [0, t_j)| > 0) = 1 - e^{-rt_j}.$$

Betrachte $M_n^{(t)} = \sum_{i\in[n]} 1_{\left\{|P_i \cap [0, \min\left\{E_n^{(i)}, t\right\})|>0\right\}}$, die Anzahl der externen Zweige von $\Pi^{(n)}$, die mindestens eine Mutation vor dem Zeitpunkt t tragen. Für jeden externen Zweig i von $\Pi^{(n)}$, der die Länge s mit $t_{j-1} \leq s < t_j$ für ein $j \in [k-1]$ hat, ist $Y_i^{(j)}$ eine obere Schranke und $Y_i^{(j-1)}$ eine untere Schranke für $1_{\{|P_i \cap [0,s)|>0\}}$. Für einen externen Zweig i der Länge s mit $t_{k-1} \leq s$ ist $Y_i^{(k)}$ eine obere und $Y_i^{(k-1)}$ eine untere Schranke für $1_{\{|P_i \cap [0, \min\{s,t\})|>0\}}$ (nur Mutationen bis zum Zeitpunkt t spielen für $M_n^{(t)}$ eine Rolle). Summiert man jeweils die oberen und unteren Schranken auf, so erhält man eine obere und eine untere Schranke für $M_n^{(t)}$. Diese Schranken sind

$$\sum_{j\in[k]} \sum_{i\in[n]} \frac{A_{inj} Y_i^{(j)}}{n} \geq \frac{M_n^{(t)}}{n} \geq \sum_{j\in[k]} \sum_{i\in[n]} \frac{A_{inj} Y_i^{(j-1)}}{n}. \tag{97}$$

Für $j \in [k-1]$ ist $(A_{inj})_{i\in[n], n\in\mathbb{N}}$ unabhängig von $(Y_i^{(j)})_{i\in\mathbb{N}}$ und durch 1 beschränkt. Für $n \to \infty$ liefert somit Satz 6.3.7 für $j \in [k-1]$

$$\begin{aligned} & \sum_{i\in[n]} \frac{A_{inj} Y_i^{(j)}}{n} \overset{f.s.}{\to} E(Y_1^{(j)})(S_{t_{j-1}} - S_{t_j}) \\ = \ & (1 - e^{-rt_j})(S_{t_{j-1}} - S_{t_j}) = \int_0^{t_j} re^{-ru}(S_{t_{j-1}} - S_{t_j}) du \text{ und} \end{aligned}$$

$$\sum_{i\in[n]} \frac{A_{inj}Y_i^{(j-1)}}{n} \stackrel{f.s.}{\to} E(Y_1^{(j-1)})(S_{t_{j-1}} - S_{t_j})$$

$$= (1 - e^{rt_{j-1}})(S_{t_{j-1}} - S_{t_j}) = \int_0^{t_{j-1}} re^{-ru}(S_{t_{j-1}} - S_{t_j})du,$$

da nach Satz 2.3.1 e) $S_{t_{j-1}} - S_{t_j}$ fast sicher der Anteil derjenigen externen Zweige $(E^{(i)})_{i\in\mathbb{N}}$ von Π ist, die mindestens Länge t_{j-1} besitzen, aber kürzer als t_j sind. Für $j = k$ liefert dasselbe Argument

$$\sum_{i\in[n]} \frac{A_{ink}Y_i^{(k)}}{n} \stackrel{a.s.}{\to} \int_0^{t_k} re^{-ru}S_{t_{k-1}}du \text{ und}$$

$$\sum_{i\in[n]} \frac{A_{ink}Y_i^{(k-1)}}{n} \stackrel{a.s.}{\to} \int_0^{t_{k-1}} re^{-ru}S_{t_{k-1}}du.$$

Aus (97) erhält man so für $n \to \infty$ nach Umsortieren bezüglich S_{t_j}, $j \in \{0, \ldots, k\}$,

$$\int_0^t re^{-ru} \sum_{j\in[k]} S_{t_{j-1}} 1_{\{t_{j-1}\le u<t_j\}} du \ge \limsup_{n\to\infty} \frac{M_n^{(t)}}{n}$$

$$\ge \liminf_{n\to\infty} \frac{M_n^{(t)}}{n} \ge \int_0^{t_{k-1}} re^{-ru} \sum_{j\in[k]} S_{t_j} 1_{\{t_{j-1}\le u<t_j\}} du \text{ fast sicher.} \quad (98)$$

Jeder Pfad von $(S_t)_{t\ge0}$ ist monoton fallend in t und nimmt nur Werte in $[0, 1]$ an, hat also höchstens abzählbar viele Sprungstellen/Unstetigkeitsstellen. Man erhält somit

$$\sum_{j\in[k]} S_{t_{j-1}} 1_{\{t_{j-1}\le u<t_j\}}, \sum_{j\in[k]} S_{t_j} 1_{\{t_{j-1}\le u<t_j\}} \to S_u$$

für λ-fast alle $u \in [0, t)$ für $k \to \infty$, da diese Konvergenz zumindest für alle Stetigkeitsstellen t von $(S_t)_{t\ge0}$ gilt. Für $k \to \infty$ erhält man mit dem Satz der majorisierten Konvergenz aus (98)

$$\int_0^t re^{-ru}S_u du \ge \limsup_{n\to\infty} \frac{M_n^{(t)}}{n} \ge \liminf_{n\to\infty} \frac{M_n^{(t)}}{n} \ge \int_0^t re^{-ru}S_u du \quad (99)$$

fast sicher mit Majorante re^{-ru}. (99) gilt fast sicher, da außerhalb der Vereinigung der (abzählbar vielen) Ausnahmemengen der Ungleichungen (98) für verschiedene Werte von k diese Ungleichungen für alle Parameterwahlen stimmen, somit erhalten sie sich auch bei der Grenzwertbildung. Dies zeigt für $t > 0$ die fast sichere Konvergenz von $M_n^{(t)}/n$ gegen $\int_0^t re^{-ru} S_u du$ für $n \to \infty$.
Betrachte nun $\frac{M_n}{n}$. Für jedes $t > 0$ zerlege

$$\frac{M_n}{n} = \frac{M_n^{(t)}}{n} + \frac{M_n - M_n^{(t)}}{n}. \tag{100}$$

$M_n - M_n^{(t)}$ ist die Anzahl der mutierten externen Zweige des durch $(\Pi_t^{(n)})_{t\geq 0}$ gegebenen Baumes, die keine Mutation bis zum Zeitpunkt t (t ausschließlich) aufweisen, somit gilt $0 \leq (M_n - M_n^{(t)})/n \leq S_t^{(n)}$, wobei $S_t^{(n)} := \frac{1}{n}\sum_{i\in[n]} 1_{\left\{E_n^{(i)}>t\right\}}$ der Anteil der Singletons von $\Pi_t^{(n)}$ an allen n Individuen ist. Für jedes $t \geq 0$ gilt nach der Kingman-Darstellung (siehe Satz 1.1.9) $\lim_{n\to\infty} S_t^{(n)} = S_t$ P-fast sicher. Satz 2.3.4 a) liefert

$$E(S_t) = e^{-\mu_{-1}t} \to 0$$

für $t \to \infty$. Also gilt $S_t \to 0$ in L^1 für $t \to \infty$. Da $(S_t)_{t\geq 0}$ monoton fallend ist, gilt diese Konvergenz auch fast sicher (es gibt eine fast sicher konvergente Teilfolge, benutze dann Monotonie). Dies zeigt

$$0 \leq \liminf_{n\to\infty} \frac{M_n - M_n^{(t)}}{n} \leq \limsup_{n\to\infty} \frac{M_n - M_n^{(t)}}{n} \leq S_t \to 0 \ P\text{-fast sicher}$$

für $t \to \infty$. Die gewünschte fast sichere Konvergenz $M_n/n \to \int_0^\infty re^{-ru} S_u du$ für $n \to \infty$ folgt damit durch Grenzübergang in (100) für $n, t \to \infty$, da nach (99) für jedes $t > 0$ $M_n^{(t)}/n \to \int_0^t re^{-ru} S_u du$ P-fast sicher für $n \to \infty$ gilt und $\int_0^t re^{-ru} S_u du \to \int_0^\infty re^{-ru} S_u du$ für $t \to \infty$ gilt. Da $(M_n/n)_{n\in\mathbb{N}}$ wegen $0 \leq M_n/n \leq 1$, $n \in \mathbb{N}$, beschränkt und damit gleichgradig integrierbar ist, gilt diese Konvergenz auch in L^p ($p \geq 1$). □

Bemerkung 6.3.8 *Für Coalescents ohne Staub gilt $S_t = 0$ fast sicher für alle $t > 0$. Somit gilt in diesem Fall nach Satz 6.3.6 $M_n/n \to 0$ fast sicher und in L^p ($p \geq 1$) für $n \to \infty$.*

Für Coalescents mit Staub lässt sich ebenfalls die Verteilung der Grenzvariablen $M := \int_0^\infty re^{-rt}S_t$ bestimmen. Aus Satz 2.3.8 ist bekannt, dass für einen Coalescent mit Staub eine Modifikation $(X_t)_{t\geq 0}$ von $(-\log(S_t))_{t\geq 0}$ existiert, die ein Subordinator mit exponentieller Vernichtung ist. Somit gilt $M \stackrel{d}{=} r\int_0^\infty e^{-rt-X_t}dr$ für den Prozess $(X_t)_{t\geq 0}$, da die verwendete càdlàg-Modifikation (vgl. [78, Theorem 11.1, Theorem 11.5]) jeden Pfad an höchstens abzählbar vielen Stellen, nämlich den Unstetigkeitsstellen von $(S_t)_{t\geq 0}$ verändert, der Übergang zur Modifikation somit keinen Unterschied bei der Integration bewirkt. Ist X ein $(0,\rho,\alpha)$-Subordinator, so ist der Prozess $\tilde{X}$ definiert durch $\tilde{X}_t := rt + X_t$ für $t \geq 0$ ein (r,ρ,α)-Subordinator, d.h. ein Subordinator mit exponentieller Vernichtung, exponentieller Vernichtungsrate α und Drift $r > 0$. Der Laplace-Exponent $\tilde{\Phi}$ des zu $\tilde{X}$ gehörenden Subordinators (ohne exponentielle Vernichtung) ist definiert durch $\tilde{\Phi}(\eta) = \Phi(\eta) + r\eta$ für $\eta \geq 0$. Somit ist auch $M \stackrel{d}{=} r\int_0^\infty e^{-\tilde{X}_t}dt$ für den (r,ρ,α)-Subordinator $\tilde{X}$. Solche exponentiellen Integrale von Subordinatoren mit Drift und exponentieller Vernichtung werden in [23] analysiert. [23, Proposition 3.1] liefert folgende Formel für die Momente von $M \stackrel{d}{=} \int_0^\infty re^{-rt-X_t}dt \stackrel{d}{=} r\int_0^\infty e^{-\tilde{X}_t}dt$.

Bemerkung 6.3.9 *[36] Betrachte für einen Coalescent mit Staub den Anteil $(S_t)_{t\geq 0}$ der Singletons. Sei $r > 0$. Für die Zufallsvariable $M = \int_0^\infty re^{-rt}S_t$ (siehe Satz 6.3.3 und Satz 6.3.6) gilt*

$$\mathrm{E}(M^k) = \frac{r^k k!}{(r+\alpha+\Phi(1))(2r+\alpha+\Phi(2))\cdots(kr+\alpha+\Phi(k))}, \quad k \in \mathbb{N}, \tag{101}$$

wobei Φ der Laplace-Exponent des Subordinatoranteils der càdlàg-Modifikation von $(-\log(S_t))_{t\geq 0}$, α die exponentielle Vernichtungsrate und $r > 0$ die Mutationsrate ist. Insbesondere erhält man

$$\begin{aligned}
\mathrm{Var}(M) &= \mathrm{E}(M^2) - (\mathrm{E}(M))^2 \\
&= \frac{2r^2}{(r+\alpha+\Phi(1))(2r+\alpha+\Phi(2))} - \frac{r^2}{(r+\alpha+\Phi(1))^2} \\
&= \frac{2r^2(r+\alpha+\Phi(1)) - r^2(2r+\alpha+\Phi(2))}{(r+\alpha+\Phi(1))^2(2r+\alpha+\Phi(2))} \\
&= \frac{r^2}{(r+\alpha+\Phi(1))^2(2r+\alpha+\Phi(2))}\int_{\Delta\setminus\{0\}} \frac{|x|^2}{(x,x)}\,\Xi(dx), \qquad (102)
\end{aligned}$$

da $\alpha + 2\Phi(1) - \Phi(2) = \int_{\Delta\setminus\{0\}} |x|^2/(x,x)\Xi(dx)$ *(vergleiche Satz 2.3.8). Für* Λ*-Coalescents gilt* $\int_{\Delta\setminus\{0\}} \frac{|x|^2}{(x,x)}\Xi(dx) = \Lambda(\Delta\setminus\{0\})$*, da in diesem Fall die Masse von* Ξ *auf* $([0,1],0,0,\ldots)$ *liegt und* $(x,x) = x_1^2$ *sowie* $|x| = x_1$ *für* $x \in ([0,1],0,0,\ldots)$ *gilt. Somit gilt*

$$\mathrm{Var}(M) = \frac{r^2}{(r+\alpha+\Phi(1))^2(2r+\alpha+\Phi(2))}\Lambda(\Delta\setminus\{0\}). \tag{103}$$

Für Simple Coalescents kann man die Verteilung von M zusätzlich über eine stochastische Fixpunktgleichung charakterisieren.

Bemerkung 6.3.10 *[36] Für Simple Coalescents mit Mutation (Mutationsrate* $r > 0$*) gilt für die Zufallsvariable* $M = \int_0^\infty re^{-rt}S_t dt$ *(*$(S_t)_{t\geq 0}$ *ist der Anteil der Singletons von* Π*)*

$$M \stackrel{d}{=} B + A(1-B)M, \tag{104}$$

wobei A *und* B *unabhängig sind, beide unabhängig von* M *sind,* $B \stackrel{d}{=} \beta(1,\mu_{-2}/r)$ *und* $1-A$ *verteilt ist wie die Einschränkung auf* $(0,1]$ *des Bildmaßes von* $\nu_0 := \nu/\mu_{-2}$ *unter der Abbildung* $|\cdot| : \Delta\setminus\{0\} \to (0,1]$*,* $x \mapsto |x|$*. Die Verteilung von* M *ist durch (104) eindeutig bestimmt.*

Beweis: Für Simple Coalescents ist nach Bemerkung 2.3.10 $(X_t)_{t\geq 0} = (-\log(S_t))_{t\geq 0}$ ein numerischer zusammengesetzter Poisson-Prozess. Es gilt also

$$X_t = \sum_{i=1}^{N_t} \eta_i,\ t \geq 0,$$

wobei $N := (N_t)_{t\geq 0}$ ein homogener Poisson-Prozess mit Intensität μ_{-2} und $\eta_1, \eta_2, \ldots$ von N unabhängige i.i.d. Zufallsvariable mit Verteilung $\eta_1 \stackrel{d}{=} \frac{\rho'}{\mu_{-2}}$ sind mit $\rho' = \rho + \alpha\cdot\delta_\infty$, wobei ρ das Lévy-Maß des zu X gehörenden Subordinators ohne exponentielle Vernichtung ist und α die Vernichtungsrate von X ist. Seien außerdem $T_1 < T_2 < \ldots$ die Sprungzeiten von N (OBdA wird angenommen, dass diese verschieden sind, dies stimmt auch immer außerhalb einer Nullmenge). Dann gilt $T_{i+1} - T_i \stackrel{d}{=} Exp(\mu_{-2})$ für $i \in \mathbb{N}_0$ und $(T_{i+1} - T_i)_{i\in\mathbb{N}_0}$ sind unabhängig ($T_0 \equiv 0$). Für $S_t = e^{-X_t}$ gilt somit $S_t = e^{-\sum_{i\in[n]}\eta_i}$ für alle t mit $N_t = n$ ($n \in \mathbb{N}_0$), also mit $t \in [T_n, T_{n+1})$. Es folgt

$$
\begin{aligned}
M &\stackrel{d}{=} \int_0^\infty re^{-rt}S_t\,dt \;=\; \sum_{i=0}^{\infty}\int_{T_i}^{T_{i+1}} re^{-rt}S_t\,dt \\
&= \int_0^{T_1} re^{-rt}\,dt + e^{-\eta_1}\int_{T_1}^{T_2} re^{-rt}\,dt + e^{-\eta_1-\eta_2}\int_{T_2}^{T_3} re^{-rt}\,dt + \cdots \\
&= (1-e^{-rT_1}) + e^{-\eta_1}(e^{-rT_1}-e^{-rT_2}) + e^{-\eta_1-\eta_2}(e^{-rT_2}-e^{-rT_3}) + \cdots \\
&= (1-e^{-rT_1}) + e^{-\eta_1}e^{-rT_1}\,\cdot \\
&\qquad \cdot\Big((1-e^{-r(T_2-T_1)}) + e^{-\eta_2}(e^{-r(T_2-T_1)}-e^{-r(T_3-T_1)}) + \cdots\Big) \\
&= B + A(1-B)\tilde{M},
\end{aligned}
$$

mit $A := e^{-\eta_1}$, $B := 1-e^{-rT_1}$ und

$$
\tilde{M} := (1-e^{-r(T_2-T_1)}) + e^{-\eta_2}(e^{-r(T_2-T_1)}-e^{-r(T_3-T_1)}) + \cdots .
$$

A und B sind unabhängig, da η_1 und T_1 unabhängig sind. Ebenso sind A und B von $\tilde{M}$ unabhängig, da (η_1, T_1) von $(T_{i+1}-T_i)_{i\in\mathbb{N}}$ und $\eta_2, \eta_3, \ldots$ unabhängig ist.
Bestimme die Verteilungen der Variablen A, B und $\tilde{M}$. Betrachtet man den Prozess N', der in 0 startet und an den Zeitpunkten $T_i' := T_{i+1}-T_1$, $i\in\mathbb{N}$, um $+1$ springt, so ist dieser wieder ein homogener Poisson-Prozess mit Intensität μ_{-2}. Somit ist $X' := (X_t')_{t\geq 0}$ definiert durch $X_t' := \sum_{i\in[N_t']}\eta_{i+1}$ identisch verteilt wie X. Führt man die zu obigem Gleichungssystem gehörenden Umformungen analog für $\int_0^\infty re^{-rt-X_t'}\,dt$ durch, so ergibt die dritte Umformung gerade $\tilde{M}$, also folgt $\tilde{M} \stackrel{d}{=} M$. Für B ergibt sich für $x\in(0,1)$

$$
P(B>x) = P(T_1 > -\frac{\log(1-x)}{r}) = (1-x)^{\frac{\mu_{-2}}{r}},
$$

somit $B \stackrel{d}{=} \beta(1, \mu_{-2}/r)$. Für die Verteilung von A erhält man für $z\in(0,1]$

$$
\begin{aligned}
&P(1-A\leq z) = P(\eta_1 \leq -\log(1-z)) = \frac{1}{\mu_{-2}}\rho'((0,-\log(1-z)]) \\
&= \frac{1}{\mu_{-2}}\nu\Big(\,\{x\in\Delta \mid -\log(1-|x|)\in(0,-\log(1-z)]\}\,\Big) \\
&= \frac{1}{\mu_{-2}}\nu\Big(\,\{x\in\Delta \mid |x|\in(0,z]\}\,\Big),
\end{aligned}
$$

da ρ' nach Bemerkung 2.3.10 die Einschränkung auf $(0,\infty]$ des Bildmaßes von ν unter $x \mapsto -\log(1-|x|)$ ist.
In [86, Theorem 1.5] wird gezeigt, dass durch Gleichung (104) die Verteilung von M eindeutig bestimmt ist, da $A(1-B)$ fast sicher nur Werte in $(0,1)$ annimmt (dies entspricht Fall I auf [86, S. 754]). $\square$

Bemerkung 6.3.11 *Aus [86, Theorem 1.5] erhält man auch, dass die Verteilung von M für einen Simple Coalescent mit Mutation gerade die stationäre Verteilung des Prozesses $(Y_n)_{n\in\mathbb{N}_0}$ ist, der rekursiv durch $Y_0 := 0$ und $Y_{n+1} := A_n(1-B_n)Y_n + B_n$ definiert wird, wobei $((A_n,B_n))_{n\in\mathbb{N}_0}$ eine Folge von i.i.d. Zufallsvariablen mit $(A_1,B_1) \stackrel{d}{=} (A,B)$ ist (A,B definiert wie in der vorangegangenen Bemerkung). Durch Iterieren der Rekursion erhält man*

$$Y_n = \sum_{i=0}^{n-1} B_{n-i-1} \prod_{j=n-i}^{n-1} A_j(1-B_j) \stackrel{d}{=} \sum_{i=0}^{n-1} B_i \prod_{j=0}^{i-1} A_j(1-B_j), \quad n \in \mathbb{N}_0,$$

und somit $M \stackrel{d}{=} \sum_{i=0}^{\infty} B_i \prod_{j=0}^{i-1} A_j(1-B_j)$. $\square$

Einige Beispiele, in denen die Verteilung von M für bestimmte Coalescents mit Staub explizit ausgerechnet wird, finden sich in Kapitel 6.5.

6.4 Anzahl der Kollisionen und deren Beziehung zu anderen Funktionalen in Coalescents mit Staub und Mutation [34], [36]

Sei wiederum Π ein Ξ-Coalescent mit Mutation, dessen pfadweise konsistente Mutationsstruktur durch homogene i.i.d. Poisson-Punktprozesse $(P_i)_{i\in\mathbb{N}}$ auf $[0,\infty)$ mit Intensitätsrate $r > 0$ gegeben ist. Betrachte für $n \in \mathbb{N}$ die Funktionale M_n, N_n und K_n in dem durch $\Pi^{(n)} := (\rho_n(\Pi_t))_{t\geq 0}$ gegebenen Baum mit der durch $(P_i)_{i\in\mathbb{N}}$ induzierten Mutationsstruktur. Es sei daran erinnert, dass M_n die Anzahl der mutierten externen Zweige, N_n die Anzahl der mutierten Zweige in diesem Baum sind und dass K_n die Anzahl der Typen in der Stichprobe $[n]$ ist, wenn deren Genealogie durch $\Pi^{(n)}$ und $(P_i)_{i\in\mathbb{N}}$ gegeben ist. Es gilt für alle $n \in \mathbb{N}$ die Ungleichung

$$M_n \leq K_n \leq N_n + 1. \tag{105}$$

Die erste Ungleichung folgt daraus, dass ein Individuum $i \in [n]$ auf jeden Fall einen Typ trägt, den kein anderes Individuum trägt, falls auf den i-ten externen Zweig eine oder mehrere Mutationen fallen. Die zweite Ungleichung folgt daraus, dass einerseits jeder Typ in der Stichprobe $[n]$, der nicht der ursprüngliche Typ des jüngsten Urahns (MRCA) der Stichprobe ist, durch eine Mutation entsteht, andererseits mehrere Mutationen auf demselben Zweig genau denselben Effekt auf die Anzahl der Typen haben wie eine einzelne Mutation auf diesem Zweig. Man kann aus Ungleichung (105) die Ungleichungen

$$\begin{aligned} 0 \;\leq\; K_n - M_n \;&\leq\; N_n + 1 - M_n \\ &= \text{ Anzahl interner mutierter Zweige} + 1 \\ &\leq \text{ Anzahl interner Zweige} + 1 \;=\; C_n \end{aligned}$$

folgern, da jede Kollision im n-Coalescent bis auf die letzte Kollision, die dem jüngsten Urahn (MRCA) der Individuen aus $[n]$ entspricht, genau einen internen Zweig erzeugt (interne Zweige entstehen immer auf diese Weise). Will man die Asymptotik von $(K_n)_{n\in\mathbb{N}}$ in einem Coalescent mit Staub und Mutation untersuchen, erscheint es mit den Ergebnissen aus dem vorangegangenen Kapitel und obigen Ungleichungen sinnvoll, zunächst die Asymptotik von $(C_n)_{n\in\mathbb{N}}$ für $n \to \infty$ zu untersuchen. Um Aussagen über die Asymptotik von $(C_n)_{n\in\mathbb{N}}$ für $n \to \infty$ machen zu können, benötigt man einige technische Lemmata. Dafür benötigen wir die in (16) und (17) eingeführten Raten g_n, g_{nk} $(k \in [n-1])$ des Block-Zählprozesses $(|\Pi_t^{(n)}|)_{t\geq 0}$.

Lemma 6.4.1 *[36] Für $n \in \{2, 3, \ldots\}$ sei $\Pi^{(n)}$ ein Ξ-n-Coalescent und $\Xi = a\delta_0 + \Xi_0$, wobei $\Xi_0(\{0\}) = 0$. Für $n \in \mathbb{N}$ sei I_n verteilt wie die Anzahl der Blöcke von $\Pi^{(n)}$ nach dem ersten Sprung. Dann gilt*

$$g_n \mathrm{E}(n - I_n) \;=\; a\binom{n}{2} + \int_\Delta \left(n|x| - \sum_{i=1}^{\infty} \big(1 - (1-x_i)^n\big)\right) \frac{\Xi_0(dx)}{(x,x)}. \qquad (106)$$

Beweis: Wähle $n \in \{2, 3, \ldots\}$ fest. Da der Beweis ähnlich wie der Beweis von Satz 1.4.1 verläuft, wird die dortige Notation verwendet. Sei also $X_i^{(x)} := \sum_{j\in[n]} 1_{\{\xi_j^{(x)}=i\}}$ für $i \in \mathbb{N}_0$, wobei $(\xi_j^{(x)})_{j\in\mathbb{N}}$ i.i.d. Zufallsvariablen sind mit $P(\xi_1^{(x)} = i) = x_i$ für $i \in \mathbb{N}$ und $P(\xi_1^{(x)} = 0) = 1 - |x|$ für $x \in \Delta$.
Den ersten Summanden in (106) erhält man, wenn man in $g_n E(n - I_n) = \sum_{k\in[n-1]}(n-k)g_{nk}$ Darstellung (21), nämlich

$$g_{nk} \;=\; a\binom{n}{2} 1_{\{k=n-1\}} + \int_\Delta P\Big(X_0^{(x)} + \sum_{i=1}^{\infty} 1_{\{X_i^{(x)}\geq 1\}} = k\Big) \frac{\Xi_0(dx)}{(x,x)}$$

einsetzt und nur den ersten Summanden aus (21) betrachtet. Dies lässt sich auch anschaulich deuten, da sich $\Pi^{(n)}$ mit Wahrscheinlichkeit $a = \Xi(\{0\})$ wie ein Kingman-n-Coalescent verhält, also in diesem Fall $I_n = n - 1$ und $g_n = \binom{n}{2}$ ist. Setzt man den zweiten Summanden aus (21) in $g_n E(n - I_n)$ ein, erhält man

$$\sum_{k\in[n-1]} (n-k) \int_\Delta P\Big(X_0^{(x)} + \sum_{i=1}^\infty 1_{\{X_i^{(x)}\geq 1\}} = k\Big) \frac{\Xi_0(dx)}{(x,x)}.$$

Dies ist aber (vergleiche wieder (21)) genau $g_n E(n - I_n)$, falls man die Größen statt in einem Ξ-n-Coalescent in einem Ξ_0-n-Coalescent betrachtet. Um (106) zu zeigen, kann man also $a = 0$ annehmen; dies wird auch getan.
Zeige nun (106). Für ein gegebenes $x \in \Delta$ betrachte die Zufallsvariablen $(X_i^{(x)})_{i\in\mathbb{N}_0}$. Es gilt $X_i^{(x)} \stackrel{d}{=} Bin(n, x_i)$ für alle $i \in \mathbb{N}_0$ und $\sum_{i=0}^\infty X_i^{(x)} = n$. Bemerke

$$\begin{aligned}
& P\Big(\bigcap_{i\in\mathbb{N}} \{X_i^{(x)} \leq 1\}\Big) \\
= \;& P(X_0^{(x)} = n) + \sum_{l\in[n]} \sum_{\substack{i_1,\dots,i_l\in\mathbb{N}\\ i_1<\dots<i_l}} P(X_0^{(x)} = n-l, X_{i_1} = 1, \dots, X_{i_l}^{(x)} = 1) \\
= \;& x_0^n + \sum_{l\in[n]} \binom{n}{l} x_0^{n-l} \sum_{\substack{i_1,\dots,i_l\in\mathbb{N}\\ \text{alle verschieden}}} x_{i_1}\cdots x_{i_l}, \qquad (107)
\end{aligned}$$

wobei für das letzte Gleichheitszeichen verwendet wird, dass das Ereignis $\Big\{X_0^{(x)} = n-l, X_{i_1} = 1, \dots, X_{i_l}^{(x)} = 1\Big\}$ bedeutet, dass aus den Werten der i.i.d. Zufallsvariablen $(\xi_i^{(x)})_{i\in\mathbb{N}}$ genau $n-l$ Werte 0 sind und die Werte $i_1, \dots, i_l$ von jeweils genau einer Zufallsvariable angenommen werden.

Es gilt wieder mit (106)

$$
\begin{aligned}
& g_n \mathrm{E}(I_n) \\
= & \sum_{k\in[n-1]} k g_{nk} \;=\; \sum_{k\in[n-1]} k \int_\Delta P\Big(X_0^{(x)} + \sum_{i=1}^\infty 1_{\{X_i^{(x)}\geq 1\}} = k\Big) \frac{\Xi_0(dx)}{(x,x)} \\
= & \int_\Delta \sum_{k\in[n-1]} kP\Big(X_0^{(x)} + \sum_{i=1}^\infty 1_{\{X_i^{(x)}\geq 1\}} = k\Big) \frac{\Xi_0(dx)}{(x,x)} \\
= & \int_\Delta \Big(E\Big(X_0^{(x)} + \sum_{i=1}^\infty 1_{\{X_i^{(x)}\geq 1\}}\Big) - nP\Big(X_0^{(x)} + \sum_{i=1}^\infty 1_{\{X_i^{(x)}\geq 1\}} = n\Big)\Big) \frac{\Xi_0(dx)}{(x,x)} \\
= & \int_\Delta \Big(\mathrm{E}(X_0^{(x)}) + \sum_{i=1}^\infty P(X_i^{(x)} \geq 1) - nP\Big(\bigcap_{i=1}^\infty \{X_i^{(x)} \leq 1\}\Big)\Big) \frac{\Xi_0(dx)}{(x,x)}. \quad (108)
\end{aligned}
$$

Zieht man diesen Term von (siehe [79, Gl. (70), S.36] mit $a = 0$)

$$
\begin{aligned}
n g_n \quad &= \quad n \int_\Delta \Bigg(1 - x_0^n - \sum_{l\in[n]} \binom{n}{l} x_0^{n-l} \sum_{\substack{i_1,\ldots,i_l\in\mathbb{N} \\ \text{alle verschieden}}} x_{i_1}\cdots x_{i_l}\Bigg) \frac{\Xi_0(dx)}{(x,x)} \\
&\overset{(107)}{=} \quad \int_\Delta \Bigg(n - nP\Big(\bigcap_{i=1}^\infty \{X_i^{(x)} \leq 1\}\Big)\Bigg) \frac{\Xi_0(dx)}{(x,x)}
\end{aligned}
$$

ab, erhält man

$$
g_n E(n - I_n) = \int_\Delta \Big(n - E(X_0^{(x)}) - \sum_{i=1}^\infty P(X_i^{(x)} \geq 1)\Big) \frac{\Xi_0}{(x,x)}.
$$

Setzt man dann $\mathrm{E}(X_0^{(x)}) = n(1 - |x|)$ und $P(X_i^{(x)} \geq 1) = 1 - (1 - x_i)^n$ ein, so erhält man das gewünschte Ergebnis. □

Bemerkung 6.4.2 *Für Λ-Coalescents wurde Lemma 6.4.1 bereits in [80, Lemma 3] bewiesen.*

Für $n \in \{2, 3, \ldots\}$ sei nun V_n die Anzahl der Kollisionen beim ersten Sprung von $\Pi^{(n)}$ und $V_1 \equiv 0$. Ist $\Pi^{(n)}$ ein Λ-n-Coalescent, so gilt $V_n \equiv 1$ für alle $n \geq 2$, da bei jedem Sprung genau eine Kollision stattfindet. Für jeden Ξ-n-Coalescent $\Pi^{(n)}$ gilt

$$
V_n = \chi_1^{(n)} - S_1^{(n)}, \tag{109}
$$

wobei $\chi_1^{(n)}$ die Anzahl der Blöcke und $S_1^{(n)}$ die Anzahl der Singletons in $\Pi^{(n)}$ nach dem ersten Sprung ist. Für $E(V_n)$ gilt folgendes Lemma.

Lemma 6.4.3 *[36] Für $n \in \{2,3,\dots\}$ sei $\Pi^{(n)}$ ein Ξ-Coalescent mit $\Xi = a\delta_0 + \Xi_0$, wobei $\Xi_0(\{0\}) = 0$. Betrachte V_n in $\Pi^{(n)}$. Dann gilt*

$$g_n \mathrm{E}(V_n) \;=\; a\binom{n}{2} + \int_\Delta \sum_{i=1}^{\infty} \bigl(1-(1-x_i)^n - nx_i(1-x_i)^{n-1}\bigr)\frac{\Xi_0(dx)}{(x,x)}. \quad (110)$$

Beweis: Sei $n \in \{2,3,\dots\}$. Die Variablen $(\xi_i^{(x)})_{i\in\mathbb{N}}$ und $(X_i^{(x)})_{i\in\mathbb{N}_0}$ seien definiert wie im Beweis von Lemma 6.4.1. Analog zum Beweis von Lemma 6.4.1 erklärt sich der erste Summand der rechten Seite von (110) dadurch, dass sich $\Pi^{(n)}$ mit Wahrscheinlichkeit a wie ein Kingman-n-Coalescent verhält und dort $V_n \equiv 1$ sowie $g_n = \binom{n}{2}$ ist. Nehme nun wiederum $a = 0$ an.
Aus der Darstellung der Raten $\lambda(n, k_1, \dots, k_m)$ für $m \in [n]$, $k_1 \geq k_2 \geq \dots \geq k_m \geq 2$ mit $\sum_{i\in[m]} k_i \leq n$ aus (1) bzw. ihrer Reformulierung (20) mit Hilfe der von $(\xi_i^{(x)})_{i\in[n]}$ erzeugten Partition $\mathsf{Part}((\xi_i^{(x)})_{i\in[n]})$ (siehe Beweis von Satz 1.4.1) folgt

$$\begin{aligned} P(S_1^{(n)} = s) &= \frac{P\Bigl(\bigcup_{\eta\in\mathbb{E}_n,\eta \text{ hat } s \text{ Singletons}} \Bigl\{\mathsf{Part}((\xi_i^{(x)})_{i\in[n]}) = \eta\Bigr\}\Bigr)}{g_n} \\ &= \frac{P\Bigl(X_0^{(x)} + \sum_{i=1}^\infty 1_{\bigl\{X_i^{(x)}=1\bigr\}} = s\Bigr)}{g_n} \end{aligned}$$

für $s \in \{0,\dots,n-2\}$, da die totalen Raten von $\Pi^{(n)}$ gerade den totalen Raten des zugehörigen Block-Zählprozesses entsprechen und $X_0^{(x)} + \sum_{i=1}^\infty 1_{\bigl\{X_i^{(x)}=1\bigr\}}$ nicht den Wert $n-1$ annehmen kann. Damit erhält man

$$\begin{aligned} g_n \mathrm{E}(S_1^{(n)}) &= \sum_{s=0}^{n-1} s \int_\Delta P\Bigl(X_0^{(x)} + \sum_{i=1}^\infty 1_{\{X_i^{(x)}=1\}} = s\Bigr)\frac{\Xi_0(dx)}{(x,x)} \\ &= \int_\Delta \sum_{s=0}^{n-1} sP\Bigl(X_0^{(x)} + \sum_{i=1}^\infty 1_{\{X_i^{(x)}=1\}} = s\Bigr)\frac{\Xi_0(dx)}{(x,x)} \\ &= \int_\Delta \Biggl(\mathrm{E}(X_0^{(x)}) + \sum_{i=1}^\infty P(X_i^{(x)} = 1) - nP\Bigl(\bigcap_{i=1}^\infty \{X_i^{(x)} \leq 1\}\Bigr)\Biggr)\frac{\Xi_0(dx)}{(x,x)}. \end{aligned}$$

Zieht man diese Größe von der Darstellung von $g_n E(\chi_1^{(n)}) = g_n \mathrm{E}(I_n)$ aus (108) ab (die vorkommenden Summen konvergieren absolut, man kann also gliedweise subtrahieren), erhält man via (109)

$$g_n \mathrm{E}(V_n) \;=\; \int_\Delta \sum_{i=1}^\infty P(X_i^{(x)} \geq 2) \frac{\Xi_0(dx)}{(x,x)}$$

und das Lemma folgt aus $P(X_i^{(x)} \geq 2) = 1 - (1-x_i)^n - nx_i(1-x_i)^{n-1}$, da $X_i \stackrel{d}{=} Bin(n, x_i)$. $\square$

Bemerkung 6.4.4 *Für Λ-n-Coalescents reduziert sich (110) für $n \in \mathbb{N}\backslash\{1\}$ auf*

$$g_n \mathrm{E}(V_n) \;=\; \int_{[0,1]} (1-(1-x)^n - nx(1-x)^{n-1}) \frac{\Lambda(dx)}{x^2} \;=\; g_n,$$

also $E(V_n) = 1$, was natürlich schon aus $V_n \equiv 1$ für Λ-Coalescents folgt.

Korollar 6.4.5 *[36] Sei Π ein Ξ-Coalescent mit Staub. Betrachte die Variablen I_n und V_n für $\Pi^{(n)} := (\rho_n(\Pi_t))_{t\geq 0}$. Dann gilt*

$$\lim_{n\to\infty} \mathrm{E}(n - I_n)/\mathrm{E}(V_n) = \infty.$$

Beweis: Definiere die Hilfsfunktion $H : \mathbb{N} \to \mathbb{R}$ durch

$$H(n) \;:=\; \int_{\Delta\backslash\{0\}} \sum_{i=1}^\infty \big(1 - (1-x_i)^n\big) \frac{\Xi(dx)}{(x,x)}, \quad n \in \mathbb{N}.$$

Mit $0 \leq 1-(1-x_i)^n \leq nx_i$ für $n \in \mathbb{N}$ und $x_i \in [0,1]$ (Bernoulli-Ungleichung) erhält man

$$0 \;<\; H(n) \;\leq n \int_{\Delta\backslash\{0\}} |x| \frac{\Xi(dx)}{(x,x)} \;=\; nH(1) \;<\; \infty,$$

da $H(1) = \mu_{-1} < \infty$ in einem (n-)Coalescent mit Staub ist. Die linke Seite von (106) lässt sich bezüglich der Hilfsfunktion H umschreiben als $g_n \mathrm{E}(n - I_n) = nH(1) - H(n)$. Aus (110) erhält man $0 \leq g_n \mathrm{E}(V_n) \leq H(n)$ (Ξ hat keine Masse in 0, also $a = 0$). Somit gilt

$$\frac{\mathrm{E}(n-I_n)}{\mathrm{E}(V_n)} \;\geq\; \frac{nH(1) - H(n)}{H(n)} \;=\; \frac{nH(1)}{H(n)} - 1.$$

Zu zeigen ist also nur noch $\lim_{n\to\infty} H(n)/n = 0$. Für einen Ξ-Coalescent mit Staub ist das Maß μ definiert durch $\mu(dx) := (|x|/(x,x))\Xi(dx)$ endlich und besitzt keine Masse in 0. Man erhält

$$\frac{H(n)}{n} \;=\; \int_{\Delta\setminus\{0\}} f_n(x)\mu(dx),$$

wobei $f_n(x) := \sum_{i=1}^{\infty}(1-(1-x_i)^n)/(n|x|)$ für $n \in \mathbb{N}$ und $x \in \Delta \setminus \{0\}$. Wiederum mit $0 \leq 1-(1-x_i)^n \leq nx_i$ für $x_i \in [0,1]$ gilt $0 \leq f_n \leq 1$ für alle $n \in \mathbb{N}$. Konvergiert nun f_n punktweise für $n \to \infty$ gegen die Nullfunktion auf $\Delta \setminus \{0\}$, so gilt $H(n)/n \to 0$ für $n \to \infty$ nach majorisierter Konvergenz, was die gewünschte Aussage zeigt. Die punktweise Konvergenz von f_n gegen 0 bleibt zu zeigen. Sei dazu $x \in \Delta \setminus \{0\}$ und $\mu_{\mathbb{N}}$ das Zählmaß auf $\mathbb{N}$. Es gilt

$$|x|f_n(x) \;=\; \sum_{i=1}^{\infty}\frac{1-(1-x_i)^n}{n} \;=\; \int h_n d\mu_{\mathbb{N}},$$

wobei $h_n : \mathbb{N} \to \mathbb{R}$ durch $h_n(i) := (1-(1-x_i)^n)/n$ definiert ist. Offensichtlich gilt $h_n \to 0$ punktweise für $n \to \infty$, da $0 \leq h_n \leq 1/n$ für alle $n \in \mathbb{N}$. Des Weiteren ist $h_n(i) \leq x_i =: h(i)$ für alle $n \in \mathbb{N}$. h ist integrierbar bezüglich des Zählmaßes $\mu_{\mathbb{N}}$ ($\int h d\mu_{\mathbb{N}} = \sum_{i=1}^{\infty} x_i \leq 1$). Somit folgt $f_n(x) \to 0$ für $n \to \infty$ aufgrund majorisierter Konvergenz. $\square$

Diese Vorarbeiten ermöglichen es nun, folgenden Satz zu beweisen.

Satz 6.4.6 *[36] Sei Π ein Coalescent mit Staub. Sei $\Pi^{(n)} := (\rho_n(\Pi_t))_{t\geq 0}$. Für $n \in \mathbb{N}$ betrachte C_n, die Anzahl der Kollisionen in $\Pi^{(n)}$. Es gilt $C_n/n \to 0$ in L^p $(p \geq 1)$.*

Beweis: Sei Π ein Ξ-Coalescent und $\Pi^{(n)} := (\rho_n(\Pi_t))_{t\geq 0}$. Aus der Verteilungsrekursion für $(C_n)_{n\in\mathbb{N}}$ aus Satz 2.4.1 b) erhält man die Rekursion

$$E(C_1) = 0 \text{ und } E(C_n) = E(V_n) + \sum_{k\in[n-1]} E(C_k)r_{nk} \text{ für } n \geq 2, \qquad (111)$$

wobei $r_{nk} = g_{nk}/g_n$, $n,k \in \mathbb{N}$, $k < n$ und V_n die Anzahl der Kollisionen beim ersten Sprung von $\Pi^{(n)}$ ist. Zeige nun $C_n/n \to 0$ in L^1 durch Widerspruchsbeweis in Analogie zum Beweis von [40, Proposition 3]. Ein ähnliches Argument wird auch in [50, S. 219] verwendet.
Angenommen es gäbe ein $\varepsilon > 0$ mit $E(C_n) > n\varepsilon$ für unendlich viele $n \in \mathbb{N}$.

Wählt man dann ε noch kleiner, so gilt für jedes feste $c > 0$ sogar die Ungleichung $E(C_n) > \varepsilon n + c$ für unendlich viele $n \in \mathbb{N}$ (für dasselbe ε). Sei n_c das kleinste n, für das diese Ungleichung gilt. n_c erfüllt $n_c \to \infty$ für $c \to \infty$. Für $k < n_c$ gilt $E(C_k) \leq \varepsilon k + c$, woraus mit Rekursion (111)

$$\begin{aligned}\varepsilon n_c + c &< E(C_{n_c}) = E(V_{n_c}) + \sum_{k \in [n_c - 1]} r_{n_c,k} E(C_k) \\ &\leq E(V_{n_c}) + c + \varepsilon \sum_{k \in [n_c-1]} k r_{n_c,k} = E(V_{n_c}) + c + \varepsilon \mathrm{E}(I_{n_c})\end{aligned}$$

folgt. Die Konstante c kürzt sich auf beiden Seiten und man erhält

$$\varepsilon \mathrm{E}(n_c - I_{n_c}) < E(V_{n_c}).$$

Lässt man $c \to \infty$, erhält man den gewünschten Widerspruch, da $\mathrm{E}(n - I_n)/E(V_n) \to \infty$ für $n \to \infty$ nach Korollar 6.4.5. Es gibt also für alle $\varepsilon > 0$ ein $n_0 = n_0(\varepsilon) \in \mathbb{N}$, sodass $E(C_n)/n \leq \varepsilon$ für alle $n \geq n_0$ und somit $C_n/n \to 0$ in L^1 für $n \to \infty$. Da $\frac{C_n}{n} \in [0, 1]$ für alle $n \in \mathbb{N}$ ist, folgt daraus sogar $C_n/n \to 0$ in L^p für alle $p \geq 1$. $\square$

Betrachtet man nur die Klasse der Simple Coalescents, so kann man zeigen, dass diese Konvergenz zusätzlich sogar fast sicher gilt. Um dies zu tun, benutzt man die Poisson-Konstruktion des Coalescents, die im Fall eines Simple Coalescents besonders gut handhabbar ist (vgl. Abschnitt 1.3).

Satz 6.4.7 *[34] Sei Π ein Simple Ξ-PPP-Coalescent. Sei $\Pi^{(n)} := (\rho_n(\Pi_t)_{t \geq 0})$. Für $n \in \mathbb{N}$ betrachte C_n, die Anzahl der Kollisionen in $\Pi^{(n)}$. Dann gilt*

$$\frac{C_n}{n} \to 0 \quad \textit{fast sicher für } n \to \infty.$$

Beweis: Sei Ψ der Poisson-Punktprozess, durch den Π konstruiert wird. Jede Kollision in Π wird durch einen Poisson-Punkt $(T^{(l)}, X^{(l)})$, $l \in \mathbb{N}$, von Ψ (bzw. der in der ersten Koordinate geordneten Menge Ψ') erzeugt. Für jedes $j \in \mathbb{N}$ teile C_n auf in die Anzahl der Kollisionen, die durch die ersten j Punkte $(T^{(1)}, X^{(1)}), \ldots, (T^{(j)}, X^{(j)})$ von Ψ erzeugt werden und in diejenigen, die von anderen Punkten erzeugt werden. Um die zweite Anzahl zu untersuchen, ist es hilfreich, die Zufallsvariablen

$$B_i^{(j)} := 1_{\left\{X_i^{(1)} = \cdots = X_i^{(j)} = 0\right\}}$$

für $i,j \in \mathbb{N}$ zu definieren. Um $(B_i^{(j)})_{i\in\mathbb{N}}$ zu analysieren benötigt man die Eigenschaften von $(X^{(i)})_{i\in\mathbb{N}}$. Aus (11) in Kapitel 1.3 ist bekannt, dass für einen Poisson-Punktprozess Ψ, der einen Simple PPP-Coalescent erzeugt, $(X^{(i)})_{i\in\mathbb{N}}$ eine i.i.d. Folge ist. Zusätzlich ist $(X_j^{(i)})_{j\in\mathbb{N}}$ für jedes $i \in \mathbb{N}$ austauschbar, also bedingt i.i.d. bezüglich eines zufälligen Maßes $Q^{(i)}$ auf $\mathbb{N}_0$ mit bedingter gemeinsamer Verteilung $(Q^{(i)})^{\mathbb{N}}$. Des Weiteren ist $(Q^{(i)})_{i\in\mathbb{N}}$ i.i.d. und $Q^{(i)}$ ist messbar bezüglich der von $X^{(i)}$ erzeugten σ-Algebra. Durch mehrfaches Anwenden des Satzes von Fubini folgt, dass für festes $j \in \mathbb{N}$ auch $(B_i^{(j)})_{i\in\mathbb{N}} = \left(1_{\left\{X_i^{(1)}=\cdots=X_i^{(j)}=0\right\}}\right)_{i\in\mathbb{N}}$ austauschbar und bedingt i.i.d. bezüglich $Q^{(1)},\ldots,Q^{(j)}$ mit $B_1^{(j)}$ Bernoulli-verteilt mit Parameter $\prod_{k\in[j]} P(X_0^{(k)} = 0|Q^{(k)}) = \prod_{k\in[j]} Q^{(k)}(\{0\})$ bedingt auf $Q^{(1)},\ldots,Q^{(j)}$ ist. Wendet man nun das starke Gesetz der großen Zahlen für austauschbare Zufallsvariable aus [30] an, so erhält man

$$\frac{\sum_{l\in[n]} B_l^{(j)}}{n} \to \prod_{k\in[j]} Q^{(k)}(\{0\}) \text{ fast sicher für } n\to\infty. \tag{112}$$

Teile nun C_n auf in die Anzahl der Kollisionen, die durch die ersten j Punkte von Ψ erzeugt werden und in diejenigen, die von anderen Punkten erzeugt werden. Sei $C_n^{(i)}$ die Anzahl der Kollisionen in $\Pi^{(n)}$, die durch den i-ten Poisson-Punkt verursacht werden. Aus Satz 1.3.3 ist $0 \le C_n^{(i)} \le V_n^{(i)}$ für $n \in \mathbb{N}$ bekannt, wobei $V_n^{(i)}$ die Anzahl der verschiedenen von $(X_k^{(i)})_{k\in[n]}$ erzeugten Zahlen aus $\mathbb{N}_0$ ist. Nach j Poisson-Punkten besitzt $\Pi_{t_j}^{(n)}$ höchstens $V_n^{(1)}+\cdots+V_n^{(j)}+\sum_{k=1}^n B_k^{(j)}$ Blöcke, da jeder Block von $\Pi_{T^{(j)}}^{(n)}$ entweder

a) durch eine Kollision entsteht, die durch einen der ersten j Poisson-Punkte verursacht wird oder

b) ein Singleton(-Block) $\{i\}$ ist mit $i \in \mathbb{N}$, für den $X_i^{(l)} \neq 0$ ist für mindestens ein $l \in [j]$ oder

c) ein Singleton(-Block) $\{i\}$ ist mit $i \in \mathbb{N}$, für den $X_i^{(l)} = 0$ für alle $l \in [j]$ gilt,

wobei sich die Anzahl der Blöcke, die a) oder b) erfüllen, nach oben durch $V_n^{(1)}+\cdots+V_n^{(j)}$ abschätzen lässt und die Anzahl der Blöcke, die c) erfüllen, gerade $\sum_{k=1}^{n} B_k^{(j)}$ ist. C_n ist somit nach oben beschränkt durch

$$\begin{aligned} C_n &\leq C_n^{(1)} + \cdots + C_n^{(j)} + V_n^{(1)} + \cdots + V_n^{(j)} + \sum_{k \in [n]} B_k^{(j)} - 1 \\ &\leq 2(V_n^{(1)} + \cdots + V_n^{(j)}) + \sum_{k \in [n]} B_k^{(j)} - 1, \end{aligned}$$

da $l \in \mathbb{N}$ Blöcke höchstens $l-1$ Kollisionen haben können. Aus Satz 1.3.3 erhält man $\lim_{n\to\infty}(V_n^{(i)}/n) = 0$ P-fast sicher für jedes $i \in \mathbb{N}$. Zusammen mit (112) folgt

$$0 \leq \liminf_{n\to\infty} \frac{C_n}{n} \leq \limsup_{n\to\infty} \frac{C_n}{n} \leq \prod_{k\in[j]} Q^{(k)}(\{0\})$$

fast sicher für alle $j \in \mathbb{N}$. $(Q^{(i)})_{i\in\mathbb{N}}$ ist eine i.i.d. Folge zufälliger Maße mit $P(Q^{(1)}(\{0\}) = 1) = 0$ (siehe (11), Ξ erfüllt (8), also haben Ξ und somit auch ν keine Masse in $0 \in \Delta$). Dies zeigt, dass es ein $\epsilon > 0$ gibt mit $P(Q^{(i)}(\{0\}) \leq 1-\epsilon) > 0$. Somit zeigt das Lemma von Borel-Cantelli, dass $P(\limsup_{n\to\infty} \{Q^{(i)}(\{0\}) \leq 1-\epsilon\}) = 1$ für ein $\epsilon > 0$. Es gilt folglich $\prod_{k=1}^{\infty} Q^{(k)}(\{0\}) = 0$ fast sicher, dies zeigt die Behauptung. □

Dieses Ergebnis gilt auch für beliebige Simple Ξ-Coalescents.

Korollar 6.4.8 *Sei Π ein Simple Coalescent und $\Pi^{(n)} := (\rho_n(\Pi_t))_{t\geq 0}$. Betrachte C_n im n-Coalescent $\Pi^{(n)}$. Dann gilt*

$$\frac{C_n}{n} \to 0 \quad \textit{fast sicher für } n \to \infty$$

Beweis: Ist Π ein Simple Ξ-Coalescent, der die Bedingung (15), also $P(S_t > 0) = 1$ für alle $t \geq 0$ erfüllt, so folgt aus Satz 1.3.5, dass Π schon ein PPP-Coalescent ist. Dann folgt die Behauptung direkt aus dem vorangegangenen Satz. Sei also (15) nicht erfüllt. Betrachte die Zufallsvariable $\zeta = \inf\{t \in \mathbb{R} | S_t = 0\}$. Lemma 1.3.7 b) besagt, dass man $(\Pi_t)_{0\leq t<\zeta}$ als einen Ξ'-PPP-Coalescent Π' mit $\Xi'(dx) := 1_{\{x\notin\Delta^*\}}\Xi(dx)$ auffassen kann. Sei C_n' die Anzahl der Kollisionen in $(\rho_n(\Pi_t'))_{0\leq t<\zeta}$. Dann gilt

$$C_n \leq C_n' + C_n^{(\zeta)} + |\Pi_\zeta^{(n)}| - 1$$

mit $C_n^{(\zeta)}$ definiert als die Anzahl der Kollisionen zum Zeitpunkt ζ. Dies gilt, da $C_n' + C_n^{(\zeta)}$ gerade die Anzahl der Kollisionen bis zum Zeitpunkt ζ in Π

ist und sich nach ζ höchstens $|\Pi_\zeta^{(n)}| - 1$ Kollisionen ereignen können, da ein Coalescent mit l Blöcken höchstens $l-1$ Kollisionen haben kann. Es gilt zusätzlich

$$|\Pi_\zeta^{(n)}| \leq C'_n + C_n^{(\zeta)} + \sum_{i \in [n]} 1_{\left\{i \text{ Singleton von } \Pi_\zeta^{(n)}\right\}},$$

da jeder Nicht-Singleton-Block durch eine Kollision entstehen muss. Aus den Sätzen 6.4.7 und 1.3.7 erhält man

$$\frac{C'_n}{n} \to 0, \ \frac{C_n^{(\zeta)}}{n} \to 0 \ \text{ fast sicher für } n \to \infty,$$

da die Anzahl der Kollisionen in einem Teil-Coalescent höchstens so groß ist wie die Anzahl der Kollisionen in einem ,,kompletten" Coalescent. Nach Definition gilt $\lim_{n\to\infty} \frac{1}{n} \sum_{i\in[n]} 1_{\left\{i \text{ Singleton von } \Pi_\zeta^{(n)}\right\}} = S_\zeta$ fast sicher. Es gilt $S_\zeta = 0$, da für Simple Coalescents $(S_t)_{t\geq 0}$ rechtsstetig ist. Also gilt mit obigen Abschätzungen

$$\frac{C_n}{n} \leq \frac{2(C'_n + C_n^{(\zeta)}) + \sum_{i\in[n]} 1_{\left\{i \text{ Singleton von } \Pi_\zeta^{(n)}\right\}}}{n} \to 0$$

fast sicher für $n \to \infty$. □

6.5 K_n in n-Coalescents mit Staub und Mutation [34], [36]

Im letzten Abschnitt wurde die Gleichung

$$0 \leq \frac{K_n - M_n}{n} \leq \frac{C_n}{n} \tag{113}$$

für die betreffenden Funktionale eines beliebigen Coalescents mit Mutation hergeleitet. Diese ermöglicht es, mit Hilfe der Konvergenzresultate für $(M_n/n)_{n\in\mathbb{N}}$ (Satz 6.3.6) und $(C_n/n)_{n\in\mathbb{N}}$ (Satz 6.4.6 bzw. Korollar 6.4.8) für $n \to \infty$ die Aymptotik von $(K_n)_{n\in\mathbb{N}}$ für $n \to \infty$ zu bestimmen. In diesem Abschnitt sei, falls nicht explizit anders angegeben, Π ein Ξ-Coalescent mit pfadweiser Mutationsstruktur gegeben durch eine Familie von Poisson-Punktprozessen $(P_i)_{i\in\mathbb{N}}$ auf $[0,\infty)$. Die Funktionale M_n, K_n und C_n werden im durch $\Pi^{(n)} = (\rho_n(\Pi))_{t\geq 0}$ gegebenen Baum mit der durch $(P_i)_{i\in\mathbb{N}}$ induzierten Mutationsstruktur definiert.

Satz 6.5.1 *[34], [36] Sei* Π *ein* Ξ*-Coalescent mit Staub und Mutation (Mutationsrate* $r > 0$*). Für* $n \in \mathbb{N}$ *betrachte* K_n*, die Anzahl der Typen in der Stichprobe* $[n]$*, wobei die Genealogie von* $[n]$ *durch* Π *(bzw.* $(\rho_n(\Pi))_{t\geq 0}$*) und die zugehörige Mutationsstruktur gegeben ist. Es gilt*

$$\frac{K_n}{n} \longrightarrow \int_0^\infty re^{-rt} S_t dt \quad \textit{in } L^p \ (p \geq 1) \textit{ für } n \to \infty. \tag{114}$$

Ist Π *sogar ein Simple Coalescent, so gilt diese Konvergenz zusätzlich fast sicher.*

Beweis: Aus Satz 6.3.6 erhält man $\frac{M_n}{n} \longrightarrow \int_0^\infty re^{-rt} S_t dt$ in L^p $(p \geq 1)$ und fast sicher für $n \to \infty$. Des Weiteren gilt $0 \leq (K_n - M_n)/n \leq C_n/n \to 0$ für $n \to \infty$ in L^p, $p \geq 1$, nach Satz 6.4.6 und für Simple Coalescents auch fast sicher nach Korollar 6.4.8. Also hat $(K_n/n)_{n\in\mathbb{N}}$ dasselbe asymptotische Verhalten in L^p, $p \geq 1$, wie $(M_n/n)_{n\in\mathbb{N}}$ für $n \to \infty$; für Simple Coalescents gilt dies auch für die fast sichere Asymptotik. □

Verzichtet man auf die oben genannten Voraussetzungen, d.h. darauf, dass K_n in $(\rho_n(\Pi_t))_{t\geq 0}$ für $n \in \mathbb{N}$ definiert wird und betrachtet K_n in beliebigen Ξ-n-Coalescents mit Mutation, so erhält man zumindest noch schwache Konvergenz.

Korollar 6.5.2 *[36] Für* $n \in \mathbb{N}$ *sei* $(\Pi^{(n)}, \mathsf{Mu}^{(n)})$ *ein* Ξ*-*n*-Coalescent mit Staub und Mutation mit Mutationsrate* $r > 0$*. Sei* K_n *die Anzahl der Typen in der Stichprobe* $[n]$*, dere Genealogie durch* $(\Pi^{(n)}, \mathsf{Mu}^{(n)})$ *bestimmt ist. Dann gilt*

$$\frac{K_n}{n} \longrightarrow \int_0^\infty re^{-rt} S_t dt \quad \textit{in Verteilung für } n \to \infty. \tag{115}$$

Beweis: Folgt entweder durch Coupling direkt aus (114) oder analog zum Beweis von (114) aus den Sätzen 6.3.3, 6.3.4, 6.4.6 und dem Satz von Slutsky.

Bemerkung 6.5.3 *[36] Wegen* $K_n - M_n \leq N_n - M_n \leq C_n$ *für alle* $n \in \mathbb{N}$ *gelten die Resultate (114) und (115) auch für die Funktionale* $(N_n)_{n\in\mathbb{N}}$*, wobei* N_n *die Anzahl der mutierten Zweige des von* $\Pi^{(n)}$ *erzeugten Baumes ist. Für* $K_n(1)$*, die Anzahl der verschiedenen Typen, die genau einmal in der Stichprobe* $[n]$ *vorkommen, gilt* $M_n \leq K_n(1) \leq K_n$ *für alle* $n \in \mathbb{N}$*, da jeder mutierte externe Zweig einen Typ erzeugt, der genau einmal in der Stichprobe vorkommt (die zweite Ungleichung ist offensichtlich). Somit gelten (114) und (115) ebenso für* $(K_n(1))_{n\in\mathbb{N}}$*.*

Bemerkung 6.5.4 *Die Momente der Grenzvariablen* $K = \int_0^\infty re^{-rt}S_t dt$ *werden in der Bemerkung 6.3.9 charakterisiert. Für Simple Coalescents erfüllt die Verteilung von* K *eine Fixpunktgleichung, siehe dazu Bemerkung 6.3.10.*

Bemerkung 6.5.5 *Analog zum Beweis von Korollar 6.5.2 (Variante ohne Coupling) wird in [68, Kapitel 4 u. 5] gezeigt, dass sich die Funktionale* $(L_n^{ext})_{n\in\mathbb{N}}$ *(Gesamtlänge aller externer Zweige) und* $(\mathsf{Seg}_n^{ext})_{n\in\mathbb{N}}$ *(Anzahl der Mutationen auf allen externen Zweigen) bei Normierung mit* n^{-1} *asymptotisch gleich (in Verteilung) verhalten wie die ebenso normierten* $(L_n)_{n\in\mathbb{N}}$ *bzw.* $(\mathsf{Seg}_n)_{n\in\mathbb{N}}$ *(vergleiche Bemerkung 6.3.5, dort sind auch die Grenzvariablen explizit angegeben).*

Für einige Maße Ξ wird nun die Grenzvariable $K := \int_0^\infty re^{-rt}S_t dt$ genauer analysiert. Dies ist eine Weiterführung der Beispiele 2.3.11, 2.3.12 und 2.3.13. Sei dazu X die Subordinator-Modifikation von $(-\log(S_t))_{t\geq 0}$ (mit exponentieller Vernichtung) und sei Φ der zugehörige Laplace-Exponent und α die Vernichtungsrate von X.

Beispiel 6.5.6 *(Dirac-Coalescents)[36] Für* $c \in \Delta \setminus \{0\}$ *definiere* $\Xi = \delta_c$ *und betrachte einen* Ξ*-Coalescent* Π. Π *ist ein Simple Coalescent mit* $\mu_{-1} = (c,c)^{-1}$*, somit konvergieren* $\frac{K_n}{n}$, $\frac{M_n}{n}$ *und* $\frac{N_n}{n}$ *P-fast sicher und in* L^p, $p \geq 1$, *für* $n \to \infty$ *gegen* K*. Nach (35) sind* $\Phi(1) = |c|/(c,c)1_{\{|c|\neq 1\}}$ *und* $\Phi(2) = |c|(2-|c|)/(c,c)1_{\{|c|\neq 1\}}$ *und somit gilt mit der exponentiellen Vernichtungsrate* $\alpha = (c,c)^{-1}1_{\{|c|=1\}}$ *nach Bemerkung 6.3.9*

$$\mathrm{E}(K) = \frac{r}{r+\alpha+\Phi(1)} = \frac{r}{r+\frac{|c|}{(c,c)}} \tag{116}$$

und analog

$$\mathrm{Var}(K) = \frac{r^2|c|^2/(c,c)}{(r+\frac{|c|}{(c,c)})^2(2r+\frac{|c|(2-|c|)}{(c,c)})}. \tag{117}$$

Da Π *ein Simple Coalescent ist, lässt sich die Verteilung von* K *auch wie in Bemerkung 6.3.10 über die Fixpunktgleichung* $K \stackrel{d}{=} B + (1-|c|)(1-B)K$ *eindeutig beschreiben, wobei* B *eine von* K *unabhängige,* $\beta(1,((c,c)r)^{-1})$*-verteilte Zufallsvariable ist. Für den Fall* $c \in \Delta^*$ *lässt sich* K *noch genauer beschreiben, dies wird im übernächsten Beispiel 6.5.9 behandelt.*

Beispiel 6.5.7 *(Beta-coalescents)[36] Sei $\Lambda = \beta(a,b)$ mit $a > 1$, $b > 1$ und Π ein Λ-Coalescent. Nach Beispiel 1.1.19 ist Π ein Coalescent mit Staub. Also konvergieren M_n/n, K_n/n und N_n/n für $n \to \infty$ in L^p, $p \geq 1$, gegen K, M_n/n auch fast sicher. Im Fall $a > 2$ ist Π nach Beispiel 1.1.21 sogar ein Simple Coalescent, also konvergieren in diesem Fall M_n/n, K_n/n und N_n/n auch fast sicher gegen K. In Beispiel 2.3.12 wurde gezeigt, dass die Vernichtungsrate $\alpha = 0$ erfüllt und somit X selbst ein Subordinator (ohne exponentielle Vernichtung) ist. Aus (37) erhält man $\Phi(1) = (a+b-1)/(a-1)$ und $\Phi(2) = (a+2b-1)/(a-1)$ für den zugehörigen Laplace-Exponenten Φ, somit nach Bemerkung 6.3.9 und insbesondere Gleichung (103)*

$$\mathrm{E}(K) \;=\; \frac{r}{r+\frac{a+b-1}{a-1}} \quad \textit{und} \quad \mathrm{Var}(K) \;=\; \frac{r^2}{(r+\frac{a+b-1}{a-1})^2(2r+\frac{a+2b-1}{a-1})},$$

da $\Lambda(\Delta \setminus \{0\}) = 1$. Falls Π ein Simple Coalescent ist, erfüllt K nach (104) die Verteilungs-Fixpunktgleichung $K \stackrel{d}{=} B + A(1-B)K$, wobei A und B unabhängige Zufallsvariable sind, die unabhängig von K sind. $(1-A)$ ist $\beta(a-2,b)$-verteilt und B ist $\beta(1,\mu_{-2}/r)$-verteilt mit $\mu_{-2} = \frac{(a+b-1)(a+b-2)}{(a-1)(a-2)}$ (siehe Beispiel 1.1.21).

Bemerkung 6.5.8 *Falls die Behauptung von Basdevant und Goldschmidt auf [5, S. 495] stimmt, dann zeigt das vorangegangene Beispiel (hier wird die Notation aus [5] verwendet), dass die dort definierte Zufallsvariable C_1 gerade wie K verteilt ist.*

Beispiel 6.5.9 *[36] Sei Ξ konzentriert auf Δ^*, d.h. $\Xi(\Delta) = \Xi(\Delta^*)$ und sei $\mu_{-1} < \infty$. Sei Π ein Ξ-Coalescent. Beispiele von Coalescents zu solchen Maßen sind die in Beispiel 2.3.11 behandelten Dirac-Coalescents mit Punktmasse von Ξ in Δ^*, insbesondere der sternförmige Coalescent, oder die Poisson-Dirichlet-Coalescents. Aus Beispiel 2.3.13 ist bekannt, dass $0 < \alpha = \mu_{-1} = \mu_{-2} < \infty$ und dass der zum Subordinator mit exponentieller Vernichtung X gehörige Subordinator der konstante Nullprozess ist, also auch $\phi(\eta) = 0$ für alle $\eta > 0$ gilt. Also konvergieren M_n/n, K_n/n and N_n/n fast sicher und in L^p, $p \geq 1$, für $n \to \infty$ gegen K. K hat die Momente*

$$\mathrm{E}(K^j) = r^j j!/((r+\alpha+0)\cdots(jr+\alpha+0)) = j!/((1+(\mu_{-2}/r))\cdots(j+(\mu_{-2}/r))),$$

$j \in \mathbb{N}$, und ist durch diese eindeutig bestimmt. Durch Momentenvergleich erhält man, dass $K \stackrel{d}{=} \beta(1,\mu_{-2}/r)$ ist.

Bemerkung 6.5.10 *Beispiel 6.5.9 verallgemeinert die Konvergenzaussage für K_n/n für $n \to \infty$ aus [67, Abschnitt 4], dort wurde der Fall des sternförmigen Coalescents $\Lambda = \delta_1$ behandelt. Für diesen gilt $K \stackrel{d}{=} \beta(1, \frac{1}{r})$.*

6.6 K_n im Bolthausen-Sznitman-n-Coalescent

In diesem Abschnitt wird das Funktional K_n, die Anzahl der Typen in der Stichprobe $[n]$, deren Genealogie durch einen Bolthausen-Sznitman-n-Coalescent mit Mutation gegeben wird, betrachtet. In einem Bolthausen-Sznitman-Coalescent Π gilt für den Anteil $(S_t)_{t\geq 0}$ der Singletons von Π fast sicher $S_t = 0$ für alle $t > 0$, da ein Bolthausen-Sznitman-Coalescent nach Beispiel 1.1.19 keinen Staub besitzt. Geht man zur Analyse von $(K_n)_{n\in\mathbb{N}}$ im Bolthausen-Sznitman-Coalescent analog zur Analyse im Falle der Coalescents mit Staub aus den vorangegangenen Kapiteln 6.3, 6.4 und 6.5 vor, liefert die Normierung mit $\frac{1}{n}$ nur das folgende triviale Konvergenzresultat für $(K_n)_{n\in\mathbb{N}}$.

Satz 6.6.1 *Für K_n, die Anzahl der Typen in $[n]$ im Bolthausen-Sznitman-n-Coalescent mit Mutation gilt*

$$\frac{K_n}{n} \stackrel{d}{\to} 0$$

für $n \to \infty$.

Beweis: OBdA sei Π ein Bolthausen-Sznitman-Coalescent und $\Pi^{(n)} := (\rho_n(\Pi_t))_{t\geq 0}$. Alle Funktionale werden in diesen n-Coalescents betrachtet.
Es wird wieder $K_n = M_n + (K_n - M_n)$ zerlegt und $0 \leq K_n - M_n \leq C_n$ aus (113) für alle $n \in \mathbb{N}$ verwendet, wobei M_n die Anzahl der mutierten externen Zweige und C_n die Anzahl der Kollisionen in $\Pi^{(n)}$ ist. Es genügt nun, $M_n/n \stackrel{d}{\to} 0$ und $\frac{C_n}{n} \stackrel{d}{\to} 0$ für $n \to \infty$ zu zeigen, da mit obiger Abschätzung aus $\frac{C_n}{n} \stackrel{d}{\to} 0$ auch $\frac{K_n - M_n}{n} \stackrel{d}{\to} 0$ folgt und damit nach dem Satz von Slutsky die Behauptung.
Nach Bemerkung 6.3.8 gilt $M_n/n \to 0$ fast sicher für $n \to \infty$, da der Bolthausen-Sznitman-Coalescent keinen Staub hat. Somit gilt insbesondere

$$\frac{M_n}{n} \stackrel{d}{\to} 0 \text{ für } n \to \infty. \tag{118}$$

Zeige nun $C_n/n \xrightarrow{d} 0$ für $n \to \infty$. Es gilt nach [28, Theorem 1.1] oder [49, Theorem 1], dass

$$\frac{(\log n)^2}{n} C_n - \log n - \log\log n \xrightarrow{d} X \tag{119}$$

für $n \to \infty$, wobei $-X$ die (Standard-) stetige Luria-Delbrück-Verteilung hat, also die charakteristische Funktion $E(e^{itX}) = \exp(-\pi|t|/2 + it\log|t|)$ für $t \in \mathbb{R}$ besitzt. Durch Skorohod-Coupling (siehe etwa [52, Theorem 4.30, S.79]) existieren auf einem geeigneten Wahrscheinlichkeitsraum verteilungsgleiche Kopien $(\tilde{C}_n)_{n\in\mathbb{N}}$, $\tilde{X}$ von $(C_n)_{n\in\mathbb{N}}$, X, für die diese Konvergenz sogar fast sicher gilt. Für diese fast sicher konvergente Folge $(\tilde{C}_n)_{n\in\mathbb{N}}$ gilt somit insbesondere $\frac{\tilde{C}_n}{n} \to 0$ fast sicher für $n \to \infty$. Somit folgt durch das beschriebene Coupling auch

$$\frac{C_n}{n} \xrightarrow{d} 0 \text{ für } n \to \infty. \tag{120}$$

Dies zeigt die Behauptung. □

Bemerkung 6.6.2 *Aus dem Skorohod-Coupling für $(C_n)_{n\in\mathbb{N}}$ und dem Konvergenzresultat (119) folgt auch*

$$\frac{\log(n)}{n} C_n \to 1 \tag{121}$$

stochastisch für $n \to \infty$.

Satz 6.6.1 folgt auch direkt aus dem Resultat von Basdevant und Goldschmidt aus [5, Theorem 1.1], welches besagt, dass

$$\frac{\log(n)}{n} K_n \to r \text{ stochastisch für } n \to \infty \tag{122}$$

gilt, wobei r die Mutationsrate ist. Mit dem Satz von Slutsky folgt Satz 6.6.1 sofort aus (122).

Bemerkung 6.6.3 *In [5] wird nicht nur die Asymptotik von $(K_n)_{n\in\mathbb{N}}$, sondern auch die Asymptotik des vollständigen Allel-Frequenzspektrums $(K_n(1), \dots, K_n(n))$ für $n \to \infty$ mit Hilfe der „fluid limit"-Technik untersucht. Es gilt*

$$\frac{\log^2(n)}{n} K_n(i) \xrightarrow{p} \frac{r}{k(k-1)}$$

für $k \geq 2$ und $n \to \infty$ sowie

$$\frac{\log(n)}{n} K_n(1) \xrightarrow{p} r$$

für $n \to \infty$.

Man kann natürlich auch versuchen, trotzdem einige der in den vorangegangenen Kapiteln verwendeten Methoden und Ergebnisse auf die Analyse von $(K_n)_{n\in\mathbb{N}}$ im Bolthausen-Sznitman-Coalescent zu übertragen, mit dem Unterschied, dass anstatt der Normierung n^{-1} eine schwächere Normierung verwendet wird. Etwa lässt sich aus den Ergebnissen für die externen Zweiglängen im Bolthausen-Sznitman-Coalescent sehr leicht eine asymptotische Darstellung für den Erwartungswert der Anzahl M_n der mutierten externen Zweige für $n \to \infty$ ableiten.

Korollar 6.6.4 *[35] Für $n \in \mathbb{N}$ betrachte einen Bolthausen-Sznitman-n-Coalescent mit Mutation und Mutationsrate $r > 0$. Dann gilt für die Anzahl M_n der mutierten externen Zweige*

$$\mathrm{E}(M_n) \;=\; r\frac{n}{\log n} + r(1-\gamma-r)\frac{n}{\log^2 n} + O\Big(\frac{n}{\log^3 n}\Big), \qquad n \to \infty,$$

wobei γ die Euler-Konstante ist.

Beweis: Betrachte $M_n = \sum_{i\in[n]} 1_{B_{n,i}}$, wobei $B_{n,i}$, $i \in [n]$, das Ereignis ist, dass der i-te externe Zweig im n-Coalescent mutiert ist. Es gilt $P(B_{n,i}|E_n^{(i)}) = 1 - e^{-rE_n^{(i)}}$ P-fast sicher, wobei $E_n^{(i)}$ die Länge des i-ten externen Zweiges ist (vergleiche mit dem Beweis von Satz 6.3.3). Damit erhält man aufgrund der Austauschbarkeit von $(E_n^{(i)})_{i\in[n]}$ und $E_n \stackrel{d}{=} E_n^{(1)}$ für die Länge E_n eines zufällig ausgewälten externen Zweiges des betrachteten n-Coalescents

$$\mathrm{E}(M_n) \;=\; \sum_{i\in[n]} P(B_{n,i}) \;=\; \sum_{i\in[n]} \mathrm{E}(1-e^{-rE_n^{(i)}}) \;=\; n\mathrm{E}(1-e^{-rE_n}).$$

Mit $x - \frac{1}{2}x^2 \leq 1 - e^{-x} \leq x - \frac{1}{2}x^2 + \frac{1}{3}x^3$ für $x \in [0,\infty)$ folgert man

$$r\mathrm{E}(E_n) - \frac{r^2}{2}\mathrm{E}(E_n^2) \;\leq\; \frac{\mathrm{E}(M_n)}{n} \;\leq\; r\mathrm{E}(E_n) - \frac{r^2}{2}\mathrm{E}(E_n^2) + \frac{r^3}{3}\mathrm{E}(E_n^3).$$

Aus Theorem 4.1.5 erhält man $\mathrm{E}(E_n) = 1/\log n + (1-\gamma)/\log^2 n + O(1/\log^3 n)$, $\mathrm{E}(E_n^2) = 2/\log^2 n + O(1/\log^3 n)$ und $\mathrm{E}(E_n^3) = O(1/\log^3 n)$, woraus das Korollar folgt. □

Für die Analyse der Asymptotik von $(K_n)_{n\in\mathbb{N}}$ im Bolthausen-Sznitman-Coalescent ist dieses Ergebnis bzw. die weitere Analyse der Asymptotik von $(M_n)_{n\in\mathbb{N}}$ nicht unmittelbar lohnend. Aus (122) geht hervor, dass man schwächer als $n^{-1}\log(n)$ normieren muss, um sinnvolle Aussagen über die schwache Asymptotik von $(K_n)_{n\in\mathbb{N}}$ im Bolthausen-Sznitman-Coalescent zu erhalten. Allerdings gilt nach (121), dass $n^{-1}\log(n)C_n$ für $n\to\infty$ nicht verschwindet (in Verteilung). Somit funktioniert es hier eben nicht, nur die Asymptotik von $(M_n)_{n\in\mathbb{N}}$ mit entsprechender Normierung zu bestimmen und über die Zerlegung $K_n = M_n + (K_n - M_n)$ und die Abschätzung $0 \le K_n - M_n \le C_n$ zu erhalten, dass $(K_n)_{n\in\mathbb{N}}$ dieselbe Asymptotik besitzt. Man müsste in diesem Fall die Asymptotik der Differenzen $(K_n - M_n)_{n\in\mathbb{N}}$ für $n\to\infty$ direkt analysieren anstatt nur $(C_n)_{n\in\mathbb{N}}$ zu analysieren. Somit erscheint es sinnvoller, zunächst andere Methoden zu verwenden, um die Asymptotik von $(K_n)_{n\in\mathbb{N}}$ zu analysieren. Man kann etwa versuchen, aus der Rekursion (49) (oder aus einer daraus abgeleiteten Rekursion) analog zum Vorgehen in den Kapiteln 3.1, 4.1 und 5.1 direkt Informationen über das asymptotische Verhalten der Verteilungen von $(K_n)_{n\in\mathbb{N}}$ zu gewinnen. Hier bietet es sich wieder an, geeignete erzeugende Funktionen zu verwenden. Betrachte die erzeugende Funktion f_K der Erwartungswerte $(E(K_n))_{n\in\mathbb{N}}$, also

$$f_K \ := \ \sum_{n=1}^{\infty} E(K_n)t^{n-1}, \qquad t\in\mathbb{C}, |t|<1. \tag{123}$$

Es gilt $f_K(0) = E(K_1) = 1$ und

$$f_K'(0) = E(K_2) \overset{(56)}{=} (4r+1)/(2r+1)$$

(die zugehörigen Raten des Block-Zählprozesses eines Bolthausen-Sznitman-n-Coalescents sind $g_n = n-1$ und $g_{nk} = \frac{n}{(n-k)(n-k+1)}$ für $n\in\mathbb{N}, k\in[n-1]$). Aus $0\le E(K_n)\le n$ folgt $|f_{K.}(t)| \le \sum_{n=1}^{\infty} n|t|^{n-1} = 1/(1-|t|)^2 < \infty$ für $|t|<1$. f_K hat also mindestens Konvergenzradius 1.

Lemma 6.6.5 *Betrachte für jedes $n\in\mathbb{N}$ das Funktional K_n in einem Bolthausen-Sznitman-n-Coalescent mit Mutation und Mutationsrate $r>0$. Die erzeugende Funktion f_K ist Lösung der gewöhnlichen Differentialgleichung*

$$f_K'(t) \ = \ b_1(t)f_K(t) + b_2(t), \qquad t\in\mathbb{C},\ 0<|t|<1 \tag{124}$$

mit

$$b_1(t) \ := \ \frac{1}{1-t} - \frac{r}{rt-\log(1-t)} \quad \textit{und} \quad b_2(t) \ := \ \frac{r}{(1-t)^3(rt-\log(1-t))}$$

für $t \in \mathbb{C}$ *mit* $0 < |t| < 1$.

Beweis: Sei $n \in \mathbb{N}$, $t \in \mathbb{C}$ mit $0 < |t| < 1$ und seien $g_n = n-1$ die totale Rate in Zustand n und $g_{ni} = n/((n-i)(n-i+1))$, $i \in [n-1]$, die Übergangsraten von n nach i des Block-Zählprozesses eines Bolthausen-Sznitman-n-Coalescents. Es wird wieder die Hilfsfunktion a aus (61) verwendet, also

$$a(t) \;=\; \sum_{k=1}^{\infty} \frac{t^k}{k(k+1)} \;=\; 1 - \log(1-t) + \frac{\log(1-t)}{t}.$$

Es gilt einerseits

$$\begin{aligned}
& \sum_{n=1}^{\infty} \frac{g_n + nr}{n} E(K_n) t^n \\
=\;& \sum_{n=1}^{\infty} \Big(1 + r - \frac{1}{n}\Big) E(K_n) t^n \\
=\;& (1+r) \sum_{n=1}^{\infty} E(K_n) t^n - \int_0^t \sum_{n=1}^{\infty} E(K_n) x^{n-1}\, dx \\
=\;& (1+r) t f_K(t) - \int_0^t f_K(x)\, dx.
\end{aligned}$$

Verwendet man andererseits Rekursion (56), also

$$(g_n + nr) E(K_n) \;=\; nr(1 + E(K_n - 1)) + \sum_{i \in [n-1]} g_{ni} E(K_i), \qquad n \in \{2, 3, \ldots\},$$

so erhält man mit $E(K_1)=1$

$$\begin{aligned}
&\sum_{n=1}^{\infty}\frac{g_n+nr}{n}E(K_n)t^n\\
= \quad & rt+\sum_{n=2}^{\infty}\Big(r(1+E(K_{n-1}))+\frac{1}{n}\sum_{i\in[n-1]}g_{ni}E(K_i)\Big)t^n\\
= \quad & r\sum_{n=1}^{\infty}t^n+r\sum_{n=2}^{\infty}E(K_{n-1})t^n+\sum_{n=2}^{\infty}\sum_{i\in[n-1]}\frac{E(K_i)t^n}{(n-i)(n-i+1)}\\
= \quad & r\sum_{n=1}^{\infty}t^n+r\sum_{n=2}^{\infty}E(K_{n-1})t^n+\sum_{i=1}^{\infty}E(K_i)t^i\sum_{n=i+1}^{\infty}\frac{t^{n-i}}{(n-i)(n-i+1)}\\
= \quad & r\frac{t}{1-t}+rt^2f_K(t)+tf_K(t)a(t)\\
= \quad & r\frac{t}{1-t}+f_K(t)\left(rt^2+ta(t)\right).
\end{aligned}$$

Somit gilt

$$(1+r)tf_K(t)-\int_0^t f_K(x)\,dx \;=\; r\frac{t}{1-t}+f_K(t)\left(rt^2+ta(t)\right).$$

Differenziert man diese Gleichung in t, so erhält man

$$\begin{aligned}
& (1+r)\left(f_K(t)+tf_K'(t)\right)-f_K(t)\\
= \quad & \frac{r}{(1-t)^2}+f_K'(t)(rt^2+ta(t))+f_K(t)(2rt+a(t)+ta'(t)),
\end{aligned}$$

bzw. dazu äquivalent

$$(rt(1-t)+t-ta(t))f_K'(t) \;=\; (2rt-r+a(t)+ta'(t))f_K(t)+\frac{r}{(1-t)^2}. \quad (125)$$

Es gilt $a(t)+ta'(t)=-\log(1-t)$ and $t(1-a(t))=-(1-t)\log(1-t)$. Für $t\in\mathbb{C}$, $0<|t|<1$, ist der Faktor $rt(1-t)+t-ta(t)=(1-t)(rt-\log(1-t))$ auf der rechten Seite nicht Null. Man kann also Gleichung (125) durch diesen Faktor dividieren und erhält dadurch die zu zeigende Differentialgleichung (124) mit

$$\begin{aligned}
b_1(t) \quad := \quad & \frac{2rt-r+a(t)+ta'(t)}{rt(1-t)+t(1-a(t))}\\
= \quad & \frac{2rt-r-\log(1-t)}{(1-t)(rt-\log(1-t))} \;=\; \frac{1}{1-t}-\frac{r}{rt-\log(1-t)}
\end{aligned}$$

und $b_2(t) \;:=\; \dfrac{r/(1-t)^2}{rt(1-t)+t(1-a(t))} \;=\; \dfrac{r}{(1-t)^3(rt-\log(1-t))}$. □

Bemerkung 6.6.6 *(Lösung der Differentialgleichung für f_K) Definiere für $0<|t|<1$ die Funktion $B(t) := \int_0^t b_1(x)\,dx$. Die Differentialgleichung (124) mit der Anfangsbedingung $f_K(0)=1$ besitzt die eindeutige Lösung*

$$f_K(t) \;=\; e^{B(t)}\int_0^t b_2(x)e^{-B(x)}\,dx, \quad 0<|t|<1.$$

Dies ist einfach die Standard-Lösung der inhomogenen linearen Differentialgleichung in Abhängigkeit der Funktionen b_1 und b_2 (Lösung der homogenen Gleichung und Variation der Konstanten). Leider ist diese Lösung nicht ,,explizit" genug, um daraus direkt eine Formel für die Asymptotik der Koeffizienten der Potenzreihendarstellung von f_K herleiten zu können.

Um mehr Informationen über die Verteilung von K_n für große $n\in\mathbb{N}$ als nur den Erwartungswert zu erhalten, lohnt es sich, die erzeugende Funktion $f_n(s) := \mathrm{E}(s^{K_n})$, $s\in\mathbb{C}$, zu betrachten. Aus (52) ist bekannt, dass f_n die Rekursion $f_1(s)=s$ und

$$(g_n+nr)f_n(s) \;=\; nrsf_{n-1}(s) + \sum_{k\in[n-1]} g_{nk}f_k(s), \qquad n\in\{2,3,\ldots\}, s\in\mathbb{C},$$

erfüllt. Betrachte nun die erzeugende Funktion

$$e(s,t) \;:=\; \sum_{n=1}^{\infty} f_n(s)t^n,\;\; s,t\in\mathbb{C},\;\; |s|\le 1,\;\; |t|<1. \tag{126}$$

Es gilt $e(s,0)=0$. Aus $|f_n(s)|\le \mathrm{E}(|s|^{K_n})\le |s|$ für $|s|\le 1$ erhält man, dass die erzeugende Funktion (126) als Reihe zumindest für $s,t\in\mathbb{N}$ mit $|s|\le 1$ und $|t|<1$ konvergiert. Als nächstes wird eine Differentialgleichung für e aufgestellt.

Lemma 6.6.7 *Betrachte die Funktionale $(K_n)_{n\in\mathbb{N}}$ für jedes $n\in\mathbb{N}$ in einem Bolthausen-Sznitman-n-Coalescent mit Mutation und Mutationsrate $r>0$. Die zugehörige erzeugende Funktion e aus (126) erfüllt die inhomogene Differentialgleichung*

$$e_t(s,t) \;=\; c_1(s,t)e(s,t)+c_2(s,t), \qquad s,t\in\mathbb{C}, |s|\le 1, 0<|t|<1, \tag{127}$$

mit

$$c_1(s,t) \; := \; \frac{1}{t(1-t)} - \frac{r(1-2st+st^2)}{(1-t)(rt(1-st)-(1-t)\log(1-t))}$$

sowie

$$c_2(s,t) \; := \; \frac{rst}{rt(1-st)-(1-t)\log(1-t)}$$

für $s,t \in \mathbb{C}$ *mit* $|s| \leq 1$ *und* $0 < |t| < 1$.

Beweis: Dieser Beweis verläuft analog zum Beweis von Lemma 6.6.5. Wie dort beschrieben gilt für $n \in \mathbb{N}$ im Bolthausen-Sznitman-n-Coalescent für die totale Rate g_n im Zustand n des Block-Zählprozesses $g_n = n-1$ und für die Raten g_{nk}, $k \in [n-1]$, gilt $g_{nk} = n/((n-k)(n-k+1))$. Seien $s,t \in \mathbb{C}$ mit $|s| \leq 1$ und $0 < |t| < 1$. Es gilt einerseits

$$\begin{aligned}
\sum_{n=1}^{\infty} \frac{1}{n}(g_n+nr)f_n(s)t^n &= \sum_{n=1}^{\infty} \Big(1+r-\frac{1}{n}\Big) f_n(s)t^n \\
&= (1+r)\sum_{n=1}^{\infty} f_n(s)t^n - \int_0^t \sum_{n=1}^{\infty} f_n(s)u^{n-1}\,du \\
&= (1+r)e(s,t) - \int_0^t \frac{e(s,u)}{u}\,du.
\end{aligned}$$

Andererseits liefert hier Rekursion (52)

$$\begin{aligned}
\sum_{n=1}^{\infty} \frac{1}{n}(g_n+nr)f_n(s)t^n &= rst + \sum_{n=2}^{\infty} \frac{1}{n}\Big(nrsf_{n-1}(s) + \sum_{i\in[n-1]} g_{ni}f_i(s)\Big)t^n \\
&= rst + rs\sum_{n=2}^{\infty} f_{n-1}(s)t^n + \sum_{i=1}^{\infty} f_i(s)t^i \sum_{n=i+1}^{\infty} \frac{t^{n-i}}{(n-i)(n-i+1)} \\
&= rst + rste(s,t) + e(s,t)a(t)
\end{aligned}$$

mit der Hilfsfunktion a aus (61). Man erhält aus diesen Gleichungen also

$$(1+r)e(s,t) - \int_0^t \frac{e(s,u)}{u}\,du \; = \; rst + rste(s,t) + e(s,t)a(t).$$

Differentation nach t ergibt

$$\begin{aligned}
&(1+r)e_t(s,t) - \frac{e(s,t)}{t} \\
&\qquad = \; rs + rse(s,t) + rste_t(s,t) + e_t(s,t)a(t) + e(s,t)a'(t),
\end{aligned}$$

bzw. äquivalent dazu

$$(1+r-rst-a(t))e_t(s,t) \;=\; \Big(rs+a'(t)+\frac{1}{t}\Big)e(s,t)+rs.$$

Der Vorfaktor von $e_t(s,t)$ ist $\neq 0$, da $1-a(t)\in(0,1)$ für t mit $|t|\in(0,1)$. Teile die Gleichung durch diesen. Mit $a'(t)+1/t=-t^{-2}\log(1-t)$ folgt dann, dass e wie gewünscht die Differentialgleichung (127) erfüllt mit

$$\begin{aligned}
c_1(s,t) &:= \frac{rs+a'(t)+1/t}{1+r-rst-a(t)}\\
&= \frac{rs-t^{-2}\log(1-t)}{r(1-st)+\log(1-t)(1-1/t)}\\
&= \frac{rst^2-\log(1-t)}{rt^2(1-st)-t(1-t)\log(1-t)}\\
&= \frac{1}{t(1-t)}-\frac{r(1-2st+st^2)}{(1-t)(rt(1-st)-(1-t)\log(1-t))}
\end{aligned}$$

und

$$\begin{aligned}
c_2(s,t) &:= \frac{rs}{1+r-rst-a(t)}\\
&= \frac{rs}{r(1-st)+\log(1-t)(1-1/t)}\\
&= \frac{rst}{rt(1-st)-(1-t)\log(1-t)}.
\end{aligned}$$

□

Bemerkung 6.6.8 *(Lösung der Differentialgleichung für e) Sei* $C(s,t):=\int_0^t c_1(s,x)\,dx$. *Die Lösung der inhomogenen linearen Differentialgleichung (127) ist*

$$e(s,t) \;=\; e^{C(s,t)}\int_0^t c_2(s,x)e^{-C(s,x)}\,dx. \tag{128}$$

Auch hier hilft die generische Form der Lösung nicht sofort dabei, die Koeffizienten der Potenzreihendarstellung von e zu bestimmen (bzw. deren Asymptotik).

Es sollte prinzipiell möglich sein, aus der Differentialgleichung (127) Aussagen über das asymptotische Verhalten von $(K_n)_{n\in\mathbb{N}}$ im Bolthausen-Sznitman-Coalescent zu treffen, wie es in den Kapiteln 3.1, 4.1 und 5.1 für die Funktionale $(\tau_n)_{n\in\mathbb{N}}$, $(E_n)_{n\in\mathbb{N}}$ und $(C_n^{ext})_{n\in\mathbb{N}}$ getan wurde. Allerdings lässt sich die

dort verwendete Technik der Singularitätsanalyse hier nicht direkt verwenden, da die Darstellung (128) der erzeugenden Funktion e nicht die geeignete Form hat bzw. die Form nicht explizit genug ist, um die Technik analog anzuwenden.
Als nächsten Schritt zur Bestimmung der Verteilungsasymptotik von $(K_n)_{n\in\mathbb{N}}$ für $n \to \infty$ könnte man versuchen, die Darstellung (128) explizit auszurechnen bzw. andere Formen der Singularitätsanalyse zu suchen, die direkt auf (128) anwendbar sind. Eine weitere Möglichkeit könnte es sein, ähnlich wie im Beweis von Satz 5.1.9 aus der erzeugenden Funktion e andere Verteilungsgrößen (etwa Momente) abzuleiten. Diese haben vielleicht angenehmere Eigenschaften wie die erzeugende Funktion e. Es ist geplant, die Verteilungsasymptotik von $(K_n)_{n\in\mathbb{N}}$ im Rahmen einer Publikation weiter zu untersuchen (falls dies gelingt...).

A Anhang

Satz 7.0.1 *(Degenerierter Fall von Theorem 3B aus [32]) Für den n-ten Koeffizienten der Potenzreihendarstellung von*

$$f(z) := \frac{1}{1-z}\log(\frac{1}{z}\log(\frac{1}{1-z}))^\delta,\ z \in D,\ \delta \in \mathbb{N}_0,$$

gilt

$$[z^n]f(z) = (\log\log n)^\delta + O\Big(\frac{(\log\log n)^\delta}{\log n}\Big)$$

für $n \to \infty$.

Beweis: Man geht analog zum Beweis der Theoreme [32, Theorem 3A, 3B] vor. f ist holomorph auf der geschlitzten Ebene $\mathbb{C} \setminus [1, \infty)$. Sei $n \in \mathbb{N}$. Alle im Beweis vorkommenden O-Terme sind für $n \to \infty$ bestimmt. Benutze nun die Cauchy-Integralformel

$$[z^n]f(z) = \frac{1}{2\pi i}\int_{\gamma_n} \frac{f(z)}{z^{n+1}}dz$$

für die zu einem Kreis um 0 mit Radius kleiner 1 homotope Hankelkurve $\gamma_n = \bigcup_{i=1}^4 \gamma_{n,i}$ gegeben durch

$$\begin{aligned}
\gamma_{n,1} = &\ \left\{z = 1 - \tfrac{t}{n} \mid t = e^{i\theta},\ \theta \in [-\tfrac{\pi}{2}, \tfrac{\pi}{2}]\right\},\\
\gamma_{n,2} = &\ \left\{z = 1 + \tfrac{t+i}{n} \mid t \in [0, n]\right\},\\
\gamma_{n,3} = &\ \left\{z \mid |z| = \sqrt{4 + n^{-2}},\ \Re(z) \le 2\right\},\\
\gamma_{n,4} = &\ \left\{z = 1 + \tfrac{t-i}{n} \mid t \in [0, n]\right\},
\end{aligned}$$

wobei $\Re(z)$ den Realteil von $z \in \mathbb{C}$ bezeichnet. Anschaulich ist $\gamma_{n,3}$ ein Kreisbogenstück des Kreises mit Radius $\sqrt{4 + n^{-2}}$ um 0. Der restliche Weg $\bigcup_{i \in \{1,2,4\}} \gamma_{n,i}$ ist eine rechts offene Schleife um $[1, 2]$ mit Abstand $\frac{1}{n}$. Ein Bild der Kurve γ_n findet sich in [32, Abbildung (2a)]. Es gilt

$$\frac{1}{2\pi i}\int_{\gamma_{n,3}} \frac{f(z)}{z^{n+1}}dz = O(2^{-n}),$$

da der Term $|z|^{n+1}$ für große n das Verhalten des Integrals bestimmt. Betrachte nun $r_n := \frac{1}{2\pi i}\int_{\gamma_n \setminus \gamma_{n,3}} \frac{f(z)}{z^{n+1}}dz$. Substituiere $z = 1 + \frac{t}{n}$ im Integral r_n

und setze ω_n als die nach rechts offene Schlaufe im Abstand 1 um $[0,n]$, also $\omega_n = \bigcup_{i\in\{1,2,4\}} \omega_{n,i}$ mit

$$\begin{aligned} \omega_{n,1} = & \left\{t = -e^{i\theta} \mid \theta \in [-\tfrac{\pi}{2}, \tfrac{\pi}{2}]\right\}, \\ \omega_{n,2} = & \left\{t = u + i \mid u \in [0,n]\right\}, \\ \omega_{n,4} = & \left\{t = u - i \mid u \in [0,n].\right\} \end{aligned}$$

Dann gilt durch diesen Integrationsvariablen- und damit verbundenem Integrationswegwechsel

$$\begin{aligned} r_n = \frac{1}{2\pi i n} \int_{\omega_n} \frac{f(1+\frac{t}{n})}{(1+\frac{t}{n})^{n+1}} dt \\ = \frac{1}{2\pi i n} \int_{\omega_n} \frac{-n}{t} \log\Big(\frac{1}{1+\frac{t}{n}} \log\big(-\frac{n}{t}\big)\Big)^{\delta} \big(1+\frac{t}{n}\big)^{-n-1} dt \\ = \frac{1}{2\pi i} \int_{\omega_n} (-t)^{-1} \big(1+\frac{t}{n}\big)^{-n-1} \sum_{k=0}^{\delta} \binom{\delta}{k} \big(-\log\big(1+\frac{t}{n}\big)\big)^{k} \big(\log\log\big(-\frac{n}{t}\big)\big)^{\delta-k} dt \\ = \frac{1}{2\pi i} \int_{\omega_n} \frac{(\log\log(-\frac{n}{t}))^{\delta}}{(-t)\big(1+\frac{t}{n}\big)^{n+1}} \Big(1 + \sum_{k\in[\delta]} \binom{\delta}{k} \Big(\frac{-\log(1+\frac{t}{n})}{\log\log(-\frac{n}{t})}\Big)^{k}\Big) dt. \end{aligned}$$

Definiere den Integranden dieses Integrals als

$$g_n(t) = \frac{(\log\log(-\frac{n}{t}))^{\delta}}{(-t)(1+\frac{t}{n})^{n+1}} \Big(1 + \sum_{k\in[\delta]} \binom{\delta}{k} \Big(\frac{-\log(1+\frac{t}{n})}{\log\log(-\frac{n}{t})}\Big)^{k}\Big)$$

und ω'_n als den Teil des Weges ω_n, auf dem $|t| < (\log n)^2$ gilt, also

$$\omega'_n := \left\{t \in \omega_n \mid |t| < (\log n)^2\right\}.$$

Auf $\omega_n \backslash \omega'_n$ beinhaltet der Integrand g_n den Faktor $((1+\frac{t}{n})^{-n-1})$ der Ordnung $O(e^{-c\log^2 n})$ für $c > 0$ geeignet. Alle anderen Faktoren und die Weglänge von ω_n haben höchstens Ordnung $O(n)$, deswegen macht man nur einen Fehler der (exponentiellen) Ordnung $O(e^{-(c\log^2 n - \log n)})$, wenn man r_n durch $\int_{\omega'_n} g_n(t)dt$ approximiert. Dieser kann vernachlässigt werden, approximiere also r_n durch $\int_{\omega'_n} g_n(t)dt$. Entlang der Kurve ω'_n, also für $t \in \omega'_n$, gilt nun mit $-t = re^{i\Phi}$ für $r > 0$ und $\Phi \in [-\pi, \pi)$

a1) $1 \leq |t| \leq \log^2(n)$, $|t^{-1}| \leq 1$, $1 - \frac{1}{n} \leq |1 + \frac{t}{n}| \leq 1 + \frac{|t|}{n} \leq 1 + \frac{\log^2(n)}{n}$, a2) $|e^{-t}| = e^{-\Re(t)} \leq e$

a3) $|\log(-t)| = |\log(r) + i\Phi| \leq (\log(r)^2 + \pi^2)^{\frac{1}{2}} \leq_n C \log\log(n)$ für $C > 0$ geeignet, also auch

a4) $\frac{\log(-t)}{\log(n)} \in D$ für n groß genug,

a5) mit a1)

$$\Big|\log\Big(1+\frac{t}{n}\Big)\Big| \leq \Big(\max\Big\{|\log(1-\frac{1}{n})|, \log(1+\frac{\log^2(n)}{n})\Big\}^2 + \pi^2\Big)^{\frac{1}{2}} = O(1).$$

Hierbei bedeutet $\leq_n$, dass die Ungleichung für alle $n \geq n_0$ für ein geeignetes $n_0 \in \mathbb{N}$ gilt. Des Weiteren gilt entlang ω'_n

$$\Big(1+\frac{t}{n}\Big)^{-n-1} = e^{-t}\Big(1+O\Big(\frac{\log^4(n)}{n}\Big)\Big), \tag{129}$$

siehe etwa [33, Gl. 19, S. 382], der $O(\cdot)$-Term auf der rechten Seite hängt nicht von t ab (den $O(\cdot)$-Term kann man noch genauer bestimmen, hier genügt aber diese Abschätzung). Betrachte nun

$$\Big(\log\log\Big(-\frac{n}{t}\Big)\Big)^{\delta} = (\log\log(n))^{\delta}\Big(1+\frac{1}{\log\log(n)}\log\Big(1-\frac{\log(-t)}{\log(n)}\Big)\Big)^{\delta}. \tag{130}$$

Für n so groß, dass $\frac{\log(-t)}{\log(n)} \in D$ gilt, existiert die Potenzreihendarstellung

$$\begin{aligned}\log\Big(1-\frac{\log(-t)}{\log(n)}\Big) &= \sum_{k=1}^{\infty}\frac{(-1)^{k-1}\log(-t)^k}{\log(n)^k k}\\ &= \frac{\log(-t)}{\log(n)}\Big(1+\sum_{k=2}^{\infty}\frac{(-1)^{k-1}\log(-t)^{k-1}}{\log(n)^{k-1}k}\Big).\end{aligned} \tag{131}$$

Hier lässt sich mit den Abschätzungen $a3$) und $a4$)

$$\begin{aligned}\Big|\frac{\log(-t)}{\log(n)}\Big(1+\sum_{k=2}^{\infty}\frac{(-1)^{k-1}\log(-t)^{k-1}}{\log(n)^{k-1}k}\Big)\Big| &\leq \frac{C\log\log(n)}{\log n}(1+O(1))\\ &= O\Big(\frac{\log\log(n)}{\log n}\Big)\end{aligned} \tag{132}$$

gewinnen, wobei der erste $O(\cdot)$-Term via Abschätzung mit der geometrischen Reihe bestimmt wird. Aus (131) und (132) folgt

$$\frac{1}{\log\log(n)}\log\Big(1-\frac{\log(-t)}{\log(n)}\Big)=O\Big(\frac{1}{\log(n)}\Big). \tag{133}$$

Setzt man nun Gleichung (133) in (130) ein, so erhält man

$$\begin{aligned}\Big(\log\log\Big(-\frac{n}{t}\Big)\Big)^{\delta} &= (\log\log(n))^{\delta}\Big(1+O\Big(\frac{1}{\log(n)}\Big)\Big)^{\delta}\\ &= (\log\log(n))^{\delta}+O\Big(\frac{(\log\log(n))^{\delta}}{\log(n)}\Big). \qquad (134)\end{aligned}$$

Da die Länge der Kurve ω_n' gerade $2\log^2(n)+\pi$ ist, folgt durch die vorhergehenden Abschätzungen

$$\begin{aligned}&\int_{\omega_n'} g_n(t)dt\\ =\ &\int_{\omega_n'}\frac{(\log\log(-\frac{n}{t}))^{\delta}}{(-t)}\Big(e^{-t}+O\Big(\frac{(\log(n))^4}{n}\Big)\Big)\cdot\\ &\qquad\cdot\Big(1+\sum_{k=1}^{\delta}\Big(\frac{\binom{\delta}{k}O(1)}{(\log\log(n))^k+O(\frac{(\log\log(n))^k}{\log(n)})}\Big)\Big)dt\\ =\ &\int_{\omega_n'}(-t)^{-1}e^{-t}\Big(\log\log\Big(-\frac{n}{t}\Big)\Big)^{\delta}dt+O\Big(\frac{(\log(n))^6(\log\log(n))^{\delta}}{n}\Big),\end{aligned}$$

wobei der $O(\cdot)$-Term durch das Abschätzen des Integrals durch den maximalen Integranden und die Länge des Integrationsweges ω_n' folgt. r_n weicht von diesem Ergebnis nach obigen Überlegungen nur um einen exponentiell kleinen Fehler ab, also gilt ebenso

$$r_n=\frac{1}{2\pi i}\int_{\omega_n'}(-t)^{-1}e^{-t}\Big(\log\log\Big(-\frac{n}{t}\Big)\Big)^{\delta}dt+O\Big(\frac{(\log(n))^6(\log\log(n))^{\delta}}{n}\Big). \tag{135}$$

Setzt man nun $(\log\log(n))^{\delta}+O(\frac{(\log\log(n))^{\delta}}{\log(n)})$ in (135) ein, wobei der $O(\cdot)$-Term unabhängig von t in ω_n' ist, so erhält man

$$\begin{aligned}r_n &= \frac{(\log\log(n))^{\delta}+O(\frac{(\log\log(n))^{\delta}}{\log(n)})}{2\pi i}\int_{\omega_n'}\frac{1}{(-t)e^t}dt+\\ &\qquad +O\Big(\frac{(\log(n))^6(\log\log(n))^{\delta}}{n}\Big). \qquad (136)\end{aligned}$$

Integriert man nun $\frac{1}{(-t)e^t}$ nicht nur über ω_n', sondern über die Hankelkurve $\omega' = \omega_{1,1} \cup \omega_2 \cup \omega_3$ mit

$$\begin{aligned} \omega_2 &:= \ \{t = u + i | u \in \mathbb{R}, \ \geq 0\} \\ \omega_3 &:= \ \{t = u - i | u \in \mathbb{R}, \ \geq 0\}, \end{aligned}$$

so unterscheiden sich die Werte der beiden Integrale nur um einen exponentiell kleinen Fehler (vergleiche den Übergang von ω_n zu ω_n'). Somit kann man in (136) ω_n' einfach durch ω' ersetzen, der zusätzliche Fehler verändert den $O(\cdot)$-Term nicht. Schließlich folgt durch Einsetzen von $\frac{1}{2\pi i}\int_{\omega'}(-t)^{-1}e^{-t}dt = \Gamma(1) = 1$ (siehe etwa [89, S. 245]) in (136) die Darstellung

$$\begin{aligned} r_n &= (\log\log(n))^\delta + O\Big(\frac{(\log\log(n))^\delta}{\log(n)}\Big) + O\Big(\frac{(\log(n))^6(\log\log(n))^\delta}{n}\Big) \\ &= (\log\log(n))^\delta + O\Big(\frac{(\log\log(n))^\delta}{\log(n)}\Big). \end{aligned}$$

Somit hat man die Behauptung

$$\begin{aligned} [z^n]f(z) &= r_n + \frac{1}{2\pi i}\int_{\gamma_{n,3}} \frac{f(z)}{z^{n+1}} \\ &= (\log\log(n))^\delta + O\Big(\frac{(\log\log(n))^\delta}{\log(n)}\Big) + O(2^{-n}) \\ &= (\log\log(n))^\delta + O\Big(\frac{(\log\log(n))^\delta}{\log(n)}\Big) \end{aligned}$$

gezeigt. □

Literatur

[1] Abramowitz, M. und Stegun, I.A. (1972) *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables.* 9. Auflage, Dover, New York.

[2] Aldous, D.J. (1985) Exchangeability and related topics. In: *Ecole d'Eté de probabilités de Saint-Flour, XIII–1983*, *Lecture Notes in Mathematics* **1117**, Springer, 1–198. MR0883646.

[3] Árnason, E. (2004) Mitochondrial cytochrome b variation in the high-fecundity Atlantic cod: trans-Atlantic clines and shallow gene genealogy. *Genetics* **166**, 1871–1885.

[4] Barton, N.H., Etheridge, A.M. und Véber, A. (2010) A new model for evolution in a spatial continuum. *Electron. J. Probab.* **15**, 162–216. MR2594876.

[5] Basdevant, A.-L. und Goldschmidt, C. (2008) Asymptotics of the allele frequency spectrum associated with the Bolthausen-Sznitman coalescent. *Electron. J. Probab.* **13**, 486–512. MR2386740.

[6] Berestycki, N. (2009) Recent progress in coalescent theory. *Ensaios Matemáticos* **16**. Sociedade Brasileira de Matemática, Rio de Janeiro. MR2574323.

[7] Berestycki, J., Berestycki, N., Limic, V. (2010) The Λ-coalescent speed of coming down from infinity. *Ann. Probab.* **38**, 207–233. MR2599198.

[8] Berestycki, J., Berestycki, N. und Schweinsberg J. (2007) Beta-coalescents and continuous stable random trees. *Ann. Probab.* **35**, 1835–1887. MR2349577.

[9] Berestycki, J., Berestycki, N. und Schweinsberg, J. (2008) Small-time behavior of beta-coalescents. *Ann. Inst. H. Poincaré Probab. Statist.* **44**, 214–238. MR2446321.

[10] Berestycki, J., Berestycki, N. und Schweinsberg, J. (2010) The genealogy of branching Brownian motion with absorption. *arXiv*: 1001.2337, wird in *Ann. Probab.* erscheinen.

[11] Bertoin, J. (1996) *Lévy Processes*. Cambridge Tracts in Mathematics, **121**. Cambridge University Press, Cambridge. MR1406564.

[12] Bertoin, J. und LeGall, J.-F. (2003) Stochastic flows associated to coalescent processes. *Probab. Theory Relat. Fields* **126**, 261–288. MR1990057.

[13] Bertoin, J. und LeGall, J.-F. (2000) The Bolthausen-Sznitman coalescent and the genealogy of continuous-state branching. *Probab. Theory Relat. Fields* **117**, 249–266. MR1771663.

[14] Billingsley, P. (1995) *Probability and Measure*. 3. Auflage. Wiley Series in Probability and Mathematical Statistics, John Wiley & Sons, Inc., New York. MR1324786.

[15] Birkner, M. und Blath, J. (2008) Computing likelihoods for coalescents with multiple collisions in the infinitely many sites model. *J. Math. Biol.* **57**, 435–465. MR2411228.

[16] Birkner, M., Blath, J., Capaldo, M., Etheridge, A., Möhle, M., Schweinsberg, J. und Wakolbinger, A. (2005) Alpha-stable branching and beta-coalescents. *Electron. J. Probab.* **10**, 303–325. MR2120246.

[17] Blum, M.G.B. und François, O. (2005) Minimal clade size and external branch length under the neutral coalescent. *Adv. Appl. Probab.* **37**, 647–662. MR2156553.

[18] Bolthausen, E. und Sznitman, A.-S. (1998) On Ruelle's probability cascades and an abstract cavity method. *Commun. Math. Phys.* **197**, 247–276. MR1652734.

[19] Boom, J. D. G., Boulding, E. G. und Beckenbach, A. T. (1994) Mitochondrial DNA variation in introduced populations of Pacific oyster, Crassostrea gigas, in British Columbia. *Can. J. Fish. Aquat.Sci.* **51**, 1608–1614.

[20] Bovier, A. und Kurkova, I. (2007) Much ado about Derrida's GREM. In: *Spin Glasses*, *Lecture Notes in Math.*, **1900**, Springer, Berlin, 81–115. MR2309599.

[21] Burke, C.J. und Rosenblatt, M. (1958) A Markovian function of a Markov chain. *Ann. Math. Stat.* **29**, 1112-1122. MR0101557.

[22] Caliebe, A., Neininger, R., Krawczak, M. und Rösler, U. (2007) On the length distribution of external branches in coalescence trees: genetic diversity within species. *Theor. Popul. Biol.* **72**, 245–252.

[23] Carmona, P., Petit, F. und Yor, M. (1997) On the distribution and asymptotic results for exponential integrals of Lévy processes. In: *Exponential Functionals and Principal Values Related to Brownian Motion*, Editor M. Yor, Biblioteca de la Revista Matematica Iberoamericana, Madrid, 73–121. MR1648657.

[24] Crow, J.F. und Kimura, M. (1964) The number of alleles that can be maintained in a finite population. *Genetics* **49**, 725–738.

[25] Cuzick, J. (1995) A strong law for weighted sums of i.i.d. random variables. *J. Theor. Probab.* **8**, 625–640. MR1340830.

[26] Delmas, J.-F., Dhersin, J.-S. und Siri-Jegousse, A. (2008) Asymptotic results on the length of coalescent trees. *Ann. Appl. Probab.* **18**, 997–1025. MR2418236.

[27] Drmota, M., Iksanov, A., Möhle, M. und Rösler, U. (2007) Asymptotic results concerning the total branch length of the Bolthausen-Sznitman coalescent. *Stoch. Process. Appl.* **117**, 1404–1421. MR2353033.

[28] Drmota, M., Iksanov, A., Möhle, M. und Rösler, U. (2009) A limiting distribution for the number of cuts needed to isolate the root of a random recursive tree. *Random Struct. Algorithms* **34**, 319–336. MR2504401.

[29] Eldon, B. und Wakeley, J. (2006) Coalescent processes when the distribution of offspring number among individuals is highly skewed. *Genetics* **192**, 2621–2633.

[30] Etemadi, N. und Kaminski, M. (1996) Strong law of large numbers for 2-exchangeable random variables. *Stat. Probab. Letters* **28**, 245–250. MR1406997.

[31] Ewens, W.J. (1972) The sampling theory of selectively neutral alleles. *Theoret. Popul. Biol.* **3**, 87–112. MR0325177.

[32] Flajolet, P. und Odlyzko, A. (1990) Singularity analysis of generating functions. *SIAM J. Disc. Math.* **3**, 216–240. MR1039294.

[33] Flajolet, P. und Sedgewick, R. (2009) *Analytic Combinatorics.* Cambridge University Press, Cambridge. MR2483235.

[34] Freund, F. (2012) Almost sure asymptotics for the number of types for simple Ξ-coalescents. *Electron. Comm. Probab.* **17**, 3. Artikel.

[35] Freund, F. und Möhle, M. (2009) On the time back to the most recent common ancestor and the external branch length of the Bolthausen-Sznitman coalescent. *Markov Process. Related Fields* **15**, 387–416. MR2554368.

[36] Freund, F. und Möhle, M. (2009) On the number of allelic types for samples taken from exchangeable coalescents with mutation. *Adv. Appl. Probab.* **41**, 1082–1101. MR2663237.

[37] Fu, Y.-X. und Li, W.-H. (1993) Statistical tests of neutrality of mutations. *Genetics* **133**, 693–709.

[38] Gnedin, A., Hansen, B. und Pitman, J. (2007) Notes on the occupancy problem with infinitely many boxes: general asymptotics and power laws. *Probab. Surv.* **4**, 146–171. MR2318403.

[39] Gnedin, A., Iksanov, A. und Möhle, M. (2008) On asymptotics of exchangeable coalescents with multiple collisions. *J. Appl. Probab.* **45**, 1186–1195. MR2484170.

[40] Gnedin, A. (2004) The Bernoulli sieve. *Bernoulli* **10**, 79–96. MR2044594.

[41] Gnedin, A. und Yakubovich, Y. (2007) On the number of collisions in Λ-coalescents. *Electron. J. Probab.* **12**, 1547–1567. MR2365877.

[42] Goldschmidt, C. und Martin, J.B. (2005) Random recursive trees and the Bolthausen-Sznitman coalescent. *Electron. J. Probab.* **10**, 718–745. MR2164028.

[43] Griffiths, R. und Tavaré, S. (1996) Monte Carlo inference methods in population genetics. *Math. Comput. Modelling* **23**, S. 141–158. MR1398007.

[44] Handa, K. (2009) The two-parameter Poisson-Dirichlet point process. *Bernoulli* **15**, 1082–1116. MR2597584.

[45] Hedgecock, D. (1994) Does variance in reproductive success limit effective population sizes of marine organisms? In: *Genetics and Evolution of Aquatic Organisms*, Editor Beaumont, A., Chapman and Hall, London, 1222–1344.

[46] Hein, J., Schierup, M.H. und Wiuf, C.(2005) *Gene genealogies, variation and evolution - a primer in coalescent theory.* Oxford University Press, New York. MR2120677.

[47] Huillet, T. und Möhle, M. (2011) Population genetics models with skewed fertilities: a backward and forward analysis. *Stoch. Models* **27**, 521–554. MR2827443.

[48] Iksanov, A., Marynych, A. und Möhle, M. (2009) On the number of collisions in beta(2,b)-coalescents. *Bernoulli* **15**, 829–845. MR2555201.

[49] Iksanov, A. und Möhle, M. (2007) A probabilistic proof of a weak limit law for the number of cuts needed to isolate the root of a random recursive tree. *Electron. Comm. Probab.* **12**, 28–35. MR2407414.

[50] Iksanov, A. und Möhle, M. (2008) On the number of jumps of random walks with a barrier. *Adv. Appl. Probab.* **40**, 206–228. MR2411821.

[51] Janson, S. und Kersting, G. (2011) On the total external length of the Kingman coalescent. *Electron. J. Probab.* **16**, 2203–2218.

[52] Kallenberg, O. (2002) *Foundations of Modern Probability Theory.* 2. Auflage, Springer, New York. MR1876169.

[53] Karlin, S. (1967) Central limit theorems for certain infinite urn schemes. *J. Math. Mech.* **17**, 373–401. MR0216548.

[54] Kimura, M. (1969) The number of heterozygous nucleotide sites maintained in a finite population due to a steady flux of mutations. *Genetics* **61**, 893–903.

[55] Kingman, J.F.C. (1978) The representation of partition structures. *J. London Math. Soc.* **18**, 374–380. MR0509954.

[56] Kingman, J.F.C. (1982) The coalescent. *Stoch. Proc. Appl.* **13**, 235–248. MR0671034.

[57] Kingman, J.F.C. (1982) On the genealogy of large populations. *J. Appl. Probab.* **19A**, 27–43. MR0633178.

[58] Kingman, J.F.C. (1993) *Poisson Processes.* Oxford Studies in Probability, Oxford Science Publications. MR1207584.

[59] Klenke, A. (2008) *Wahrscheinlichkeitstheorie.* 2. Auflage Springer, Berlin Heidelberg.

[60] Limic, V. (1993) On the speed of coming down from infinity for Ξ-coalescent processes. *Electron. J. Probab.* **15**, 217–240. MR2594877.

[61] Limic, V. (2010) Genealogies of regular exchangeable coalescents with applications to sampling. *arXiv*: 1004.3897, akzeptiert bei *Ann. IHP*.

[62] Marynych, A. (2010) On the asymptotics of moments of linear random recurrences. *Theory Stoch. Proc.* **16**, 106–119. MR2779988.

[63] Möhle, M. (2000) Total variation distances and rates of convergence for ancestral coalescent processes in exchangeable population models. *Adv. Appl. Probab.* **32**, 983–993. MR1808909.

[64] Möhle, M. (2004) The time back to the most recent common ancestor in exchangeable population models. *Adv. Appl. Probab.* **36**, 78–97. MR2035775.

[65] Möhle, M. (2005) Coalescent theory - simultaneous multiple collisions and sampling distributions. *Oberwolfach Reports* **40**, 2279–2282.

[66] Möhle, M. (2006) On the number of segregating sites for populations with large family sizes. *Adv. Appl. Probab.* **38**, 750–767. MR2256876.

[67] Möhle, M. (2006) On sampling distributions for coalescent processes with simultaneous multiple collisions. *Bernoulli* **12**, 35–53. MR2202319.

[68] Möhle, M. (2010) Asymptotic results for coalescent processes without proper frequencies and applications to the two-parameter Poisson-Dirichlet coalescent. *Stoch. Process. Appl.* **120**, 2159–2173. MR2684740.

[69] Möhle, M. und Sagitov, S. (2001) A classification of coalescent processes for haploid exchangeable population models. *Ann. Probab.* **29**, 1547–1562. MR1880231.

[70] Panholzer, A. (2004) Destruction of recursive trees. In: *Mathematics and Computer Science III*, Birkhäuser, Basel, 267–280. MR2090518.

[71] Pitman, J. (1999) Coalescents with multiple collisions. *Ann. Probab.* **27**, 1870–1902. MR1742892.

[72] Pitman, J. (2005) Combinatorial Stochastic Processes. In: *Ecole d'Eté der Probabilités de Saint-Flour XXXII–2002*, Editor: Picard, J., *Lecture Notes in Mathematics*, **1875**, Springer. MR2245368.

[73] Pitman, J. und Yor, M. (1997) The two-parameter Poisson-Dirichlet distribution derived from a stable subordinator. *Ann. Probab.* **25**, 855–900. MR1434129.

[74] Rauch, I.M. und Bar-Yam, Y.(2004) Theory predicts the uneven distribution of genetic diversity within species. *Nature* **431**, 449–452.

[75] Sagitov, S. (1999) The general coalescent with asynchronous mergers of ancestral lines. *J. Appl. Probab.* **36**, 1116–1125. MR1742154.

[76] Sagitov, S. (2003) Convergence to the coalescent with simultaneous multiple mergers. *J. Appl. Probab.* **40**, 839–854. MR2012671.

[77] Sargsyan, O. und Wakeley, J. (2008) A coalescent process with simultaneous multiple mergers for approximating the gene genealogies of many marine organisms. *Theor. Pop. Biol.* **74**, 104–114.

[78] Sato, K.-I. (1999) *Lévy processes and infinitely divisible distributions.* Cambridge Studies in Advanced Mathematics, **68**. Cambridge University Press, Cambridge. MR1739520.

[79] SCHWEINSBERG, J. (2000) Coalescents with simultaneous multiple collisions. *Electron. J. Probab.* **5**, 1-50. MR1781024.

[80] SCHWEINSBERG, J. (2000) A necessary and sufficient condition for the Λ-coalescent to come down from infinity. *Electron. Commun. Probab.* **5**, 1–11. MR1736720.

[81] SCHWEINSBERG, J. (2003) Coalescent processes obtained from supercritical Galton-Watson processes. *Stoch. Proc. Appl.* **106**, 107–139. MR1983046

[82] STEINRÜCKEN, M.(2009) Multiple merger coalescents and population genetic inference. Promotionsarbeit, erhältlich via

`http://opus.kobv.de/tuberlin/volltexte/2009/2373/`

[83] TAJIMA, F. (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**, 585–595.

[84] TAVARÉ, S. (2004) Ancestral inference in population genetics. *Lecture Notes in Mathematics*, **1837**, Springer, Berlin, S. 1–188. MR2071630.

[85] TAYLOR, J. UND VÉBER, A.(2009) Coalescent processes in subdivided populations subject to recurrent mass extinctions. *Electron. J. Probab.* **14**, 242–288. MR2471665.

[86] VERVAAT, W. (1979) On a stochastic difference equation and a representation of non-negative infinitely divisible random variables. *Adv. Appl. Probab.* **11**, 750–783. MR0544194.

[87] WAKELEY, J. (2007) *Coalescent Theory: An Introduction.* Roberts and Company Publishers, Greenwood Village.

[88] WATTERSON, J. (1975) On the number of segregating sites in genetic models without recombination. *Theor. Popul. Biol.* *7*, 256–276.

[89] WHITTAKER, E.T. UND WATSON, G.N. (1927) *A course of modern anlysis.* Vierte Auflage, Cambridge University Press, Cambridge. MR1424469.